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 (2011) The ”far-west” of Anopheles gambiae molecular forms. PLoS One, 6 (2). e16415. ISSN 1932-6203 DOI: https://doi.org/10.1371/journal.pone.0016415

Downloaded from: http://researchonline.lshtm.ac.uk/779/

DOI: 10.1371/journal.pone.0016415

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
The “Far-West” of *Anopheles gambiae* Molecular Forms

Beniamino Caputo¹, Federica Santolamazza¹, José L. Vicente², Davis C. Nwakanma³, Musa Jawara³, Katinka Palsson⁴, Thomas Jaenson⁴, Bradley J. White⁵, Emiliano Mancini¹, Vincenzo Petrarcha⁶, David J. Conway⁵, Nora J. Besansky², João Pinto², Alessandra della Torre¹*

¹Istituto Pasteur-Fondazione Cenci-Bolognetti, Dipartimento di Sanità Pubblica e Malattie Infettive, Università di Roma “Sapienza”, Rome, Italy, ²Centro de Malária e outras Doenças Tropicais, UEM Malária e UEI Entomologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal, ³Medical Research Council Laboratories, Banjul, The Gambia, ⁴Medical Entomology Unit, Department of Systematic Biology, Uppsala University, Uppsala, Sweden, ⁵Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana, United States of America, ⁶Istituto Pasteur-Fondazione Cenci-Bolognetti, Dipartimento di Biologia e Biotecnologie “Charles Darwin”, Università di Roma “Sapienza”, Rome, Italy

**Introduction**

*Anopheles gambiae* sensu stricto (hereafter *A. gambiae*) is the major mosquito vector responsible for malaria transmission throughout Sub-Saharan Africa. This species is undergoing a process of sympatric ecological diversification and lineage splitting which makes it a model to study divergent selection and heterogeneous genomic divergence mechanisms [1,2]. Two morphologically indistinguishable incipient species (provisionally named M and S molecular forms) have been described within *A. gambiae*, recognized by form-specific SNPs on the IGS and ITS regions of multicopy rDNA located on the X-chromosome [3,4]. The S-form is distributed across sub-Saharan Africa and breeds mostly in association with rain-dependent pools and temporary puddles. M-form populations overlap with the S-form in West and Central Africa, but are apparently absent to the east of the Great Rift Valley. M-form shows a greater ability to exploit breeding sites that exist across seasons and are more closely associated with human activities, such as those created by irrigation, rice cultivation and urbanisation [5,6,7,3]. This adaptation allows the M-form to breed throughout the year, thus causing a shift from seasonal to year-round malaria transmission. Importantly, genetic traits conferring resistance to insecticides commonly used against these vectors are differently distributed between the two forms[9]. Therefore, beyond its intrinsic interest, the ecological speciation process occurring within *A. gambiae* also has practical consequences for malaria transmission and vector control in Africa [7,8,10].

The evolutionary and ecological forces that generated divergence between M- and S-forms are not yet fully understood. Lehmann & Diabaté [6] suggest that selection by larval predation and inter-form competition drove divergence between temporary and permanent freshwater habitats, possibly explaining the ecological discontinuity of the molecular forms (e.g. rice fields vs. surrounding savannas). The same authors also suggest that quantitative differences in adult body size, reproductive output, and longevity may have contributed to differential adaptations to distinct niches. Costantini et al. [7] showed that ecological segregation between M and S forms in Burkina Faso is consistent with a niche expansion of the M-form into marginal habitats. Altogether, these observations lead to the hypothesis that M and S forms are the products of divergent selection acting on ecologically important traits allowing optimal exploitation of permanent vs.
temporary habitats for larval breeding [7,8,10,11]. Theory predicts that such ecological diversification should lead to selection for reproductive isolation. Indeed, different pre-mating reproductive mechanisms between the two forms are reported, such as complete or almost complete swamp segregation based on visual landscape markers, as observed in Mali and Burkina Faso, respectively [6,12,13], and mating-recognition via matching of male-female flight-tone harmonic frequencies [14]. Field studies in Mali have shown that strictly sympatric and synchronously breeding populations of M and S forms cross-mate at a rate of only ca. 1% [15]. M/S hybrids (as detected by a SNP on the IGS-rDNA X-linked region) are exceedingly rare in the interior of west Africa (52 M/S hybrids among ~18,000 A. gambiæ identified [5,7]) and absent from west-central Africa (none among >12,000 specimens identified [8,16,17]). Contrary to these findings, putative M/S hybrids have been recorded at much higher rates in west-central west Africa (up to 3% in Senegal, [18]; 7% in The Gambia, [19]; and >20% in Guinea Bissau, [20]). Notably, crossing experiments reveal no detectable intrinsic postzygotic reproductive isolation between M and S in F1 or backcross individuals raised in the laboratory [21].

Whole genome scans of M and S divergence using a gene-based microarray have been performed on samples from Central and West Africa [22,23]. High differentiation was detected almost exclusively in pericentromeric regions. These data were initially interpreted in the context of a speciation-with-gene-flow model that assumed introgression and homogenization of genetic variation outside of centromeric “speciation islands” [22]. However, the (nearly) complete genetic association of fixed allelic differences at centromeric markers on all three independently assorting chromosomes sampled from West, Central and East African populations is consistent with an alternative hypothesis, in which realized gene flow between forms is rare or nil across most of their range [23]. Importantly, this “hybridization-without-gene-flow” hypothesis is supported by more recent analyses based on whole genome sequencing and genotyping of M and S from Mali, indicating that heterosexual divergence is genome-wide [24,25].

The availability of diagnostic markers from the differentiated centromeric regions on all three chromosomes, and the expectation of their genetic association in the absence of realized gene flow, provide useful tools to study A. gambiæ population structure in its extreme western range, where unusually high frequencies of putative hybrids between M and S-forms are found. We genotyped one marker in each centromeric island of genetic divergence (plus the IGS-marker defining M and S forms) in A. gambiæ samples from The Gambia and Guinea Bissau, with the aim to evaluate the degree of reproductive isolation between M and S in this area, and to assess whether these unusual populations represent a recent breakdown in the M and S speciation process or whether they have simply retained ancestral polymorphisms. Persistent (though incomplete) genetic association between the unlinked pericentromeric markers, as well as discordant patterns observed for the two X-linked markers, seem to support the first hypothesis, opening new perspectives in the understanding of the molecular form speciation process.

Materials and Methods

*Anopheles gambiæ* samples, genotyping and sequencing

We have analysed samples of indoor-resting *A. gambiae* females collected in three sites along the Gambia river (Mandina Ba, MB; Sare Samba Sowe, SR; Wellingara, WE, The Gambia) in 2006 [19] and in Antula district of Bissau City, Guinea Bissau, in 1995 (A-1995) and 1996 (A-1996) [20]. Each sample comprised specimens collected in a single catch (*i.e.* in the same site within 1-5 subsequent days), with the exception of samples from SR, which were collected in two catches in August and September. An additional sample analysed was collected in Antula district in 2007 (A-2007) (Figure 1).

Individual mosquito DNA was extracted from single legs or other parts of the carcasses not including the abdomen, to avoid the risk of contamination with DNA from sperm deposited in spermathecae. Samples were identified to species and molecular form by the PCR-RFLP approach recognising two form-specific SNPs in the IGS rDNA region on chromosome-X [26]. A subsample was also genotyped by the PCR-RFLP approach recognising an additional form-specific SNP in the IGS region, mapping only 109 bp apart from the former ones [27].

One molecular form-specific marker was analysed in each centromeric region, as follows: i) the M-form specific insertion of a SINE transposon in division 6 of the X-chromosome, detected by the PCR-approach developed by Santolamazza *et al.* [28] (hereafter termed SINE-X); ii) a SNP near the 2L-centromere in the fourth exon of AGAP004679 at position 2L:209 536 in the AgamP3 assembly (VectorBase; www.vectorbase.org), detected by a PCR-RFLP approach in which the M-form specific PCR-band is cleaved [23] (hereafter termed 2L); iii) a SNP near the 3L-centromere in the third exon of AGAP010313 at position 3L:296 923 in the AgamP3 assembly, detected by a PCR-RFLP approach in which the M-form specific PCR-band is cleaved [23] (hereafter termed 3L).

In addition, PCR amplicons from the IGS [29], 2L- and 3L-centromeric PCR assays [23] were sequenced from selected specimens, using ABI BigDye™ Terminator v2.0 chemistry and an ABI Prism 3700 DNA Analyser. Multiple alignments were performed using ClustalX [30]. Electropherograms were visually inspected for heterozygous SNPs. Sequences are available from the Authors upon request.

Figure 1. Map of the four collection sites in the Gambia and Guinea Bissau. [Footnote: MB = Mandina Ba; SR = Sare Samba Sowe; WE = Wellingara; A = Antula district of Bissau City].

doi:10.1371/journal.pone.0016415.g001
PCR and sequencing were carried out in Rome, Lisbon and Banjul. Selected samples were analysed in two different laboratories to validate the results.

Statistical analyses

Estimates of inbreeding coefficient ($F_{IS}$) were obtained according to Weir & Cockerham [31] using Genepop version 4.0 [32]. Departures from Hardy-Weinberg (HW) proportions were tested by exact probability tests implemented in ARLEQUIN v.3.5 with 1 million forecasted chain lengths and 10,000 dememorization steps [33]. The same software was used to perform tests of linkage disequilibrium (LD) between the three markers genotyped; both genetic phase unknown and known procedures were performed, despite the possible bias due to the location of the markers on physically unlinked chromosomes. For the first approach, we used a procedure that is based on a likelihood ratio test, where the likelihood of the sample evaluated under the hypothesis of no association (no LD) between loci is compared to the likelihood of the sample when the association is allowed [34]. In the second approach, we used the EM algorithm to estimate the maximum likelihood haplotype frequencies (see Arlequin manual 3.5) and calculated standard deviations through bootstrap followed by exact tests of LD based on the Markov chain approach (No. of steps in Markov chain = 100,000 and No. of dememorization steps = 1,000). Chi-square values were calculated using the Vassar-Stat website for statistical computation (http://faculty.vassar.edu/lowry/VassarStats.html).

QSVanalyzer software - which was developed to facilitate the extraction of quantitative sequence variant (QSV) information from sequence electropherograms – was applied to estimate the relative proportions of the double peaks (i.e., copy number proportions: CNP) (http://dna.leeds.ac.uk/qsv; [35]) observed in electropherograms of IGS amplicon at positions 581 (M-form = T; S-form = C; [26], hereafter CNP(M)) and 690 (M-form = A; S-form = T; [27], hereafter CNP(A)) in sequences of the IGS locus from single $A. gambiae$ specimens. The program analyses each trace and adjusts it in relation to the peak heights of upstream/downstream nucleotides, allowing rapid batch wise analysis of DNA sequence traces for estimation of the relative proportions of two QSVs at a given site. The CNP score (i.e., the proportion between the heights of the C/T peaks at site 581 and 690) was calculated by dividing the S-form specific QSV by the sum of M- and S-form QSVs. Note that QSVanalyzer has never been applied before to analyse multiplex rDNA sequences. However, direct sequencing experiments revealed that the measurement of relative peak height of SNPs in the ITS regions represents an appropriate tool for studying hybridization in plants [36].

Results

We genotyped one marker per centromeric region on each of the three independently assorting chromosomes (i.e., the M-form-specific SINE-X insertion on chromosome-X and two form-specific SNPs on chromosome-2 and -3) in $A. gambiae$ populations from The Gambia [19] and Guinea Bissau [20].

Chromosome-X

The results of the SINE-PCR analysis revealed the presence of M/S heterozygotes at this locus (hereafter, SINE-XMS), 1.7% in The Gambia (N = 304) and 22.9% in Guinea Bissau (N = 332). However, a significant SINE-XMS deficit was detected in all samples analyzed (P<0.001), with the exception of Wellingara, where the SINE-X insertion was fixed (i.e., SINE-XMM) (Table 1).

In contrast to what was observed elsewhere in Africa [28], results from SINE-X genotyping were not fully consistent with the identifications based on the IGS genotype that defines molecular forms. All specimens in which the IGS-based definition did not match the expected SINE-X genotype were confirmed at least twice by PCRs carried out in different labs. The final results verified mismatched genotypes in 4.7% (16/336) and 12% (40/332) of the Gambian and Guinean samples, respectively. The mismatches were of two types: (1) specimens defined by IGS as pure M (21.4%) and pure S (7.1%) showing a heterozygous SINE-XMM genotype, or (2) specimens defined by IGS as M/S hybrids showing a homozygous SINE-XMM (1.8%) or SINE-XSS (67.9%) genotype. Only one specimen carried opposing homozygous genotypes (i.e. an IGS-based pure M-form specimen with a SINE-XSS genotype); this was later shown to be characterised by IGS mixed array by the PCR-RFLP of a different IGS-SNP [27] and by direct sequencing of the IGS amplicon (see below).

To evaluate the occurrence of possible technical biases or, alternatively, the biological significance of the above-reported mismatched genotypes, we further PCR-RFLP genotyped a second form-specific SNP at position 690 in the IGS-locus [27] in a subsample of IGS/SINE discordant and concordant specimens. The results did not show a full agreement between the two PCR-RFLP approaches for IGS genotyping. Since the IGS region is known to be constituted by an array of tandem repeats subject to gene conversion and since the two IGS-SNPs co-segregate, we hypothesize that this lack of agreement is a product of a technical bias due to the fact that some individuals may contain an unequal number of tandem copies of the M- and S-form specific IGS-arrays (see Text S1 for details).

| Table 1. Frequencies of SINE-X genotypes in Anopheles gambiae adult female samples collected in The Gambia and in Guinea Bissau. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Country         | Samples         | SINE$^{MM}$    | SINE$^{MS}$    | SINE$^{SS}$    | N               | $F_{IS}$        | Exp. Heter.     | P               |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| The Gambia      | MB              | 0.57            | 0.03            | 0.40            | 101             | 0.94            | 0.49            | <0.001          |
|                 | SR              | 0.54            | 0.01            | 0.44            | 153             | 0.97            | 0.50            | <0.001          |
|                 | WE              | 1.00            | 0.00            | 0.00            | 50              | --              | --              | --              |
| Guinea Bissau   | A-1995          | 0.28            | 0.22            | 0.50            | 100             | 0.54            | 0.48            | <0.001          |
|                 | A-1996          | 0.43            | 0.25            | 0.32            | 79              | 0.49            | 0.49            | <0.001          |
|                 | A-2007          | 0.08            | 0.22            | 0.69            | 153             | 0.30            | 0.30            | <0.001          |

Footnotes:

MB = Mandina Ba; SR = Sare Samba Sow; WE = Wellingara; A = Antuila district of Bissau City; N = sample size; $F_{IS}$ = inbreeding coefficient; Exp.Heter. = heterozygote frequencies as expected by Hardy Weinberg (HW) equilibrium; P = significance of deviation from HW equilibrium.

doi:10.1371/journal.pone.0016415.t001
To confirm the above hypothesis, we sequenced the IGS amplicon of 67 and 46 specimens in which the SINE-X and IGS genotypes did or did not match, respectively. High levels of C/T and A/T polymorphisms were observed in position 581 and 690 respectively, not only in heterozygous SINE-X<sup>MM</sup> genotypes, but also in SINE-X<sup>SS</sup> and SINE-X<sup>MS</sup> homozygotes. The sequence electropherograms were further scored by QSV analyser [35] to quantify the proportion of sequences containing C versus T, based on relative peak heights at position 581 (CNP<sup>581</sup>) and 690 (CNP<sup>690</sup>), respectively. M and S individuals from other geographic areas (Mali, Burkina Faso, Cameroon) by White et al. [23] in eastward geographic areas was not found in our populations (Table S1). Results from sequence analyses showed that most pair wise associations among X, 2L and 3L markers were present in the overall sample: 7 of 9 possible genotype associations between X and 2L, 2L and 3L, and 8 of 9 possible associations between 2L and 3L (Table S2).

The results of the 2L-RFLP analysis revealed heterozygotes (hereafter 2L<sup>MS</sup>, i.e. individuals characterised by both the M and the S 2L-specific alleles) in all populations analyzed (113/298 specimens in The Gambia and 65/173 in Guinea Bissau). However, sequence analysis of the uncleaved 2L PCR products (N = 35) uncovered the presence of a second SNP in the restriction enzyme recognition sequence. This mutation, already described at centromeric regions [23], is present in both forms in the study area, even in the exception of that collected in Guinea Bissau in 1995. In Wellingara (where all individuals were SINE-X<sup>MM</sup> homozygotes), only 3L<sup>MM</sup> and 3L<sup>MS</sup> genotypes were detected, at a frequency of 52% and 48%, respectively.

**Association between chromosome-X, -2 and -3 centromeric regions**

The complete association between X-, 2L- and 3L-centromeres observed by White et al. [23] in eastward geographic areas was not found in our populations (Table S1). Results from sequence analyses showed that most pair wise associations among X, 2L and 3L markers were present in the overall sample: 7 of 9 possible genotype associations between X and 2L, all possible associations between X and 3L, and 8 of 9 possible associations between 2L and 3L (Table S2).

Due to the high frequency of “null alleles” in both M and S at the 2L marker, only SINE-X and 3L markers were scored in the

---

**Table 2. Frequencies of 3L genotypes in Anopheles gambiae adult female samples collected in The Gambia and in Guinea Bissau.**

<table>
<thead>
<tr>
<th>Country</th>
<th>Samples</th>
<th>3L&lt;sup&gt;MM&lt;/sup&gt;</th>
<th>3L&lt;sup&gt;MS&lt;/sup&gt;</th>
<th>3L&lt;sup&gt;SS&lt;/sup&gt;</th>
<th>N</th>
<th>F&lt;sub&gt;i&lt;/sub&gt;</th>
<th>Exp. Heter.</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Gambia</td>
<td>MB</td>
<td>0.54</td>
<td>0.15</td>
<td>0.31</td>
<td>97</td>
<td>0.68</td>
<td>0.48</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>SR</td>
<td>0.57</td>
<td>0.20</td>
<td>0.23</td>
<td>151</td>
<td>0.55</td>
<td>0.44</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>WE</td>
<td>0.52</td>
<td>0.48</td>
<td>0.00</td>
<td>44</td>
<td>~0.30</td>
<td>0.37</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>A-1995</td>
<td>0.67</td>
<td>0.28</td>
<td>0.05</td>
<td>99</td>
<td>0.09</td>
<td>0.31</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>A-1996</td>
<td>0.60</td>
<td>0.24</td>
<td>0.15</td>
<td>78</td>
<td>0.40</td>
<td>0.40</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>A-2007</td>
<td>0.37</td>
<td>0.31</td>
<td>0.32</td>
<td>145</td>
<td>0.38</td>
<td>0.50</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Footnotes:

MB = Mandina Ba; SR = Sare Samba Sowe; WE = Wellingara; A = Antula district of Bissau City; N = sample size; F<sub>i</sub> = inbreeding coefficient; Exp.Heter. = heterozygote frequencies as expected by Hardy Weinberg (HW) equilibrium; P = significance of deviation from HW equilibrium; n.s. = not significant.

doi:10.1371/journal.pone.0016415.t002
full set of samples (Table 3). The results show non-random associations of X and 3L genotypes in the whole Guinean sample ($\chi^2 = 40.6; df = 4; P < 0.0001$) and in the Gambian one ($\chi^2 = 179.4; df = 2; P < 0.0001$); NB: $\chi^2$ for the Gambian sample was calculated based on SINE-X-MM and SINE-X-SS genotypes only, due to the small size of SINE-X-MS sub-sample (Table S3). Frequencies of putative “parental” genotypes (SINE-X-MM/3LMM and SINE-X-SS/3LSS) were significantly higher than expected, while “assorted” genotypes (X-MM/3LSS and X-SS/3LMM) were lower than expected. In particular, putative “parental” genotype frequencies were 54% in Guinea Bissau and 73% in The Gambia, while the frequencies of “congruent” (X-MM/3LMM, X-SS/3LSS and X-MS/3LMS) genotypes were 41% in Guinea Bissau and 74% in The Gambia ($\chi^2 = 67.1; df = 1; p < 0.0001$). Moreover, in Guinea Bissau the ratio of “parental” M SINE-X-MM/3LMM genotypes to “assorted” SINE-X-MM/3LSS or SINE-X-SS/3LMM genotypes was 6.3 to 1 (63/10), whereas the ratio of “parental” S SINE-X-SS/3LSS genotypes to the corresponding “assorted” genotypes was 0.5 to 1 (45/131) ($\chi^2 = 75; df = 1; p < 0.0001$). In The Gambia, the above ratios were 4.8:1 (150/31) and 1.5:1 (64/43), respectively ($\chi^2 = 17.5; df = 1; p < 0.0001$).

A significant association, as measured by the gametic phase unknown procedure, was observed between X and 3L loci in all samples ($P < 0.001$), with the exception of that from Antula-2007. When measured by the gametic phase known procedure on inferred haplotypes (see Materials & Methods), the association was stronger in Gambian samples (MB: $r^2 = 0.62$; SR: $r^2 = 0.40$), than in Guinean ones ($\chi^2 = 0.07; \Lambda = 0.96; \Lambda = 0.19$). Results from inference analyses used to estimate the maximum likelihood of X/3L haplotypes, revealed frequencies of X-MS/3LSS and X-SS/3LS haplotypes higher than expected under the hypothesis of linkage equilibrium (Table S4).

**Discussion**

Previous studies have shown that in *A. gambiae* M and S molecular form populations from geographic areas of no or low (~1%) inter-form crosses, such as and Cameroon, Burkina Faso and Mali, the physically unlinked centromeric regions of all three chromosomes contain fixed differences, which have been found in nearly complete linkage disequilibrium [22,23,37]. To date, only one marker in each centromeric region has been tested on a wide geographical scale: the SINE-X insertion unique to and fixed in the M-form [28] and two form-specific SNPs on chromosome-2L and -3L [23].

The high frequency of null alleles at the 2L marker precluded the scoring of markers on all three chromosomes in the full sample set, thus hindering one approach to addressing the above question, as F1 and backcross progeny could not be reliably distinguished. However, a more detailed consideration of patterns at the X-linked markers (i.e. the IGS SNP which defines the M- and S-forms and the SINE-X insertion) from a molecular evolution perspective provides some support for the secondary contact hypothesis. First, the long-term maintenance of ancestral polymorphism would be unexpected near centromeres, given that centromere-proximal regions experience sharply reduced levels of recombination [38,39]. Additionally, the rDNA locus consists of ~1000 tandemly repeated genes, and concerted evolution is believed to be responsible for relatively rapid homogenization of sequence variation among genes in the array and between individuals in populations that are connected by sufficient gene flow. Thus, the presence of mixed M+S IGS arrays is expected to be transitory, and would be unlikely to persist at the high levels observed in this study, in an ancestral population. This is further supported by the virtual absence of mixed IGS in the rest of Africa. On the other hand, mixed IGS arrays in single individuals are a plausible outcome of recombination in M/S hybrids. With respect to SINE-X, no polymorphism has been observed at this locus during extensive surveys of M and S in other parts of Africa [28]: the insertion is fixed in all M form populations and absent in

**Table 3.** Frequencies of diploid SINE-X/3L genotypes in *Anopheles gambiae* adult female samples collected in The Gambia and in Guinea Bissau.

<table>
<thead>
<tr>
<th>Countries</th>
<th>Samples</th>
<th>N</th>
<th>MM/SS</th>
<th>SS/SS</th>
<th>MS/MS</th>
<th>MM/MM</th>
<th>SS/MS</th>
<th>MS/MM</th>
<th>MS/SS</th>
<th>SS/MS</th>
<th>SS/MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Gambia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MB</td>
<td>97</td>
<td>0.515</td>
<td>0.309</td>
<td>0.010</td>
<td>0.052</td>
<td>0</td>
<td>0.010</td>
<td>0.000</td>
<td>0.093</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>SR</td>
<td>151</td>
<td>0.510</td>
<td>0.225</td>
<td>0.007</td>
<td>0.033</td>
<td>0</td>
<td>0</td>
<td>0.007</td>
<td>0.159</td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td>WE</td>
<td>44</td>
<td>0.523</td>
<td>0</td>
<td>0</td>
<td>0.477</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tot</td>
<td>292</td>
<td>0.514</td>
<td>0.219</td>
<td>0.007</td>
<td>0.106</td>
<td>0</td>
<td>0.003</td>
<td>0.003</td>
<td>0.113</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-1995</td>
<td>99</td>
<td>0.263</td>
<td>0.030</td>
<td>0.040</td>
<td>0.020</td>
<td>0</td>
<td>0.152</td>
<td>0.020</td>
<td>0.222</td>
<td>0.253</td>
<td></td>
</tr>
<tr>
<td>A-1996</td>
<td>78</td>
<td>0.397</td>
<td>0.090</td>
<td>0.115</td>
<td>0.013</td>
<td>0.026</td>
<td>0.103</td>
<td>0.038</td>
<td>0.115</td>
<td>0.103</td>
<td></td>
</tr>
<tr>
<td>A-2007</td>
<td>144</td>
<td>0.042</td>
<td>0.243</td>
<td>0.069</td>
<td>0.014</td>
<td>0.021</td>
<td>0.090</td>
<td>0.056</td>
<td>0.229</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>Tot</td>
<td>321</td>
<td>0.196</td>
<td>0.140</td>
<td>0.072</td>
<td>0.016</td>
<td>0.016</td>
<td>0.112</td>
<td>0.040</td>
<td>0.199</td>
<td>0.209</td>
<td></td>
</tr>
</tbody>
</table>

Footnotes:

MB = Mandina Ba; SR = Sare Samba Sowe; WE = Wellingara; A = Antula district of Bissau City; N = sample size.

doi:10.1371/journal.pone.0016415.t003
S. Because SINE elements cannot excise once inserted, the most parsimonious explanation for the absence of the SINE-X insertion in the S-form is that the element inserted into M after its divergence from S. Accordingly, the SINE-X polymorphism observed in both M and S in the study area is most likely the result of inter-form hybridization. Considering both loci jointly, the absence of individuals characterized by “opposite” IGS/SINE genotypes (i.e. M-form/SINE-XSS or S-form/SINE-XMM) and the observation that SINE-XSS individuals are more often characterized by M and S arrays at 50:50 proportion (suggestive of F1 hybrids) than SINE-XMM and SINE-XES ones (Fig. 2) further support the secondary contact hypothesis. Finally, the finding of SINE-XMM and SINE-XSS homozygous individuals characterized by mixed MS IGS-arrays suggests that crossing-over among the IGS-arrays allowed recombination within the X-centromere, despite the low rates reported for this region [38,39].

The lower frequencies of putative parental genotypes and correspondingly higher hybrid frequencies observed in Guinea Bissau as opposed to The Gambia suggest that the former region could be the core (or be closer to the core) of the secondary contact zone. Intriguingly, we observed a lack of genetic association between chromosome-X and -3 centromeres only in the sample collected in Guinea Bissau in 2007. Genetic association between these centromeres was detected in earlier Guinean samples (collected in 1995 and 1996), and is very strong in Gambian samples. Although further investigation is required to confirm this speculation, these data seem to suggest both a geographic and a temporal trend, in which the lack of association observed in 2007 Guinean samples might be due to the weakening of inter-form reproductive barriers in the core of the secondary contact zone. Moreover, the data are consistent with the hypothesis that secondary contact could be the result of a (recent) invasion/colonization by S-form of the study area, where a long-established secondary contact could be the result of a (recent) invasion/colonization by M and S arrays at 50:50 proportion (suggestive of F1 hybrids) than SINE-XMM and SINE-XES ones (Fig. 2) further support the secondary contact hypothesis. Finally, the finding of SINE-XMM and SINE-XSS homozygous individuals characterized by mixed MS IGS-arrays suggests that crossing-over among the IGS-arrays allowed recombination within the X-centromere, despite the low rates reported for this region [38,39].

The relative contribution of reduced pre- and post-mating barriers to inter-form gene flow in the study area is entirely unknown. In fact, this aspect of A. gambiae biology is still very poorly understood anywhere in the M and S range [6]. It may be possible that in the study area hybridization is not more frequent than elsewhere, but that negative selection is weaker against heterozygous genotypes and some genotype associations at the larval stage, resulting in greater survival to the adult stage (the stage most commonly sampled). Alternatively, reproductive barriers between the two molecular forms could be severely broken down, but negative selection could still be acting against heterozygous genotypes and favour some genotype associations. Field studies aimed at associating the centromere genotypes with biological parameters (e.g. assortative mating, inter-form insemi-

Acknowledgments

We thank the entomological teams and the local residents for assistance during mosquito collections in The Gambia and in Guinea Bissau.

Author Contributions

Conceived and designed the experiments: BC FS VP DJC NJB JP AdT. Performed the experiments: BC FS JLV DCN MJ KP TJ BJW EM. Analyzed the data: BC FS EM VP NJB JP AdT. Contributed reagents/ materials/analysis tools: DCN MJ KP TJ BJW. Wrote the paper: BC NJB AdT.
References