Jones, CM, Liyanapathirana, M, Agossa, FR, Weetman, D, Ranson, H, Donnelly, MJ and Wilding, CS

Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae

http://researchonline.ljmu.ac.uk/2641/

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Jones, CM, Liyanapathirana, M, Agossa, FR, Weetman, D, Ranson, H, Donnelly, MJ and Wilding, CS (2012) Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. PROCEEDINGS OF THE NATIONAL ACADEMY OF
Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of *Anopheles gambiae*

Christopher M. Jones, Milindu Liyanapathirana, Fiacre R. Agossa, David Weetman, Hilary Ranson, Martin James Donnelly, and Craig S. Wilding

Insecticide resistance is an ideal model to study the emergence and spread of adaptive variants. In the African malaria mosquito, *Anopheles gambiae*, this is complemented by a strong public health rationale. In this insect, resistance to pyrethroids and DDT insecticides is strongly associated with the mutations L1014F and L1014S within the para voltage-gated sodium channel (VGSC). Across much of West Africa, 1014F frequency approaches fixation. Here, we document the emergence of a mutation, N1575Y, within the linker between domains III-IV of the VGSC. In data extending over 40 kbp of the VGSC 1575Y occurs on only a single long-range haplotype, also bearing 1014F. The 1014F-1575Y haplotype was found in both M and S molecular forms of *An. gambiae* in West/Central African sample sites separated by up to 2,000 km. In Burkina Faso M form, 1575Y allele frequency rose significantly from 0.053 to 0.172 between 2008 and 2010. Extended haplotype homozygosity analysis of the wild-type 1575N allele showed rapid decay of linkage disequilibrium (LD), in sharp contrast to the extended LD exhibited by 1575Y. A haplotype with long-range LD and high/increasing frequency is a classical sign of strong positive selection. Several mutations within the sodium channel have been conclusively linked to reduced insecticide resistance (9). Several mutations within the sodium channel have been identified in an array of insects and cause varying degrees of resistance (reviewed in ref. 10). Many of these mutations occur at key residues within the so-called binding pocket enclosed by the IIS4-S5 linker and IIS5/III6 helices (9). In *An. gambiae* s.s., two single-base-pair substitutions occur at codon 1014 within segment 6 of domain II (numbering according to the housefly para sequence, GenBank accession no. X96668) resulting in substitution of leucine with either phenylalanine or serine (11, 12). The ready availability of assays for the kdr 1014 mutations has led to their routine screening as partial resistance diagnostics in *An. gambiae*. However, insecticide resistance is a rapidly evolving trait (13, 14) and, particularly in large populations, new mutations can arise frequently (15).

Identifying adaptive mutations in natural populations poses a significant challenge. Alleles positively selected from standing variation are difficult to detect owing to the time it usually takes for new variants to reach sufficient frequencies, during which period recombination will break down linkage disequilibrium (LD) with marker loci (16). This is particularly problematic for organisms such as *An. gambiae*, where LD is very short (17–19). For de novo mutations, reduced heterozygosity at linked sites represents a signature of strong selection and can be used as a means to identify variants at selective advantage. In human genetics, for example, strong evidence exists for signatures of adaptation associated with malaria resistance genes, including the Duffy antigen protein (20) and Glucose-6-phosphate dehydrogenase (21). Insecticide resistance provides an ideal model to study the adaptation of newly emerged alleles. First, resistance emerges over a relatively short period, and second, because we know when synthetic insecticides were introduced for insect control, we can estimate when positively selected alleles may have arisen in a population and trace their ancestry. For the *An. gambiae* s.s. VGSC, we have shown previously that two mutations at codon 1014 have been subject to strong selection, have risen to high frequency, and show extended long range LD indicative of a selective sweep (7). However, the observation of putative recombination at the telomeric end of 1014F carrying haplotypes...
Y substitution at codon 1575, located in the DIII-IV linker. The version AgamP3.6 was identified-T substitution at position 2L:2432975 (An. gambiae s.s. from Burkina Faso) and the insecticide-susceptible Kisumu colony. An- of a fragment of exon 30 of the VGSC was sequenced in a subsample of coexisting alleles provides evidence for a secondary selective sweep of the 1014F mutation associated with N1575Y. The functional significance of these findings is discussed.

Results

Detection of N1575Y from Field Collected An. gambiae s.s. A 330-bp fragment of exon 30 of the VGSC was sequenced in a subsample of An. gambiae s.s. S form from Soumouso and Kuinima (Burkina Faso) and the insecticide-susceptible Kisumu colony. An- T substitution at position 2L:2432975 (An. gambiae assembly version AgamP3.6) was identified; this substitution results in a N-Y substitution at codon 1575, located in the DIII-IV linker. The N1575Y mutation was segregating in Burkina Faso, but was not found in the Kisumu samples. Pyrosequencing and TaqMan assays were developed to genotype samples at position 1575 and proved 100% concordant (n = 39).

Frequencies and Distribution of 1014F and 1575Y. The 1014F mutation was at high frequency in 2008 collections of An. gambiae s.s. S form from Burkina Faso [0.846; 95% confidence interval (CI) = 0.787–0.890] and approached fixation in 2009–2010 collections (0.968–0.992) (Table 1). 1014F frequency was significantly lower in M form over the 2008–2010 period (0.368–0.562). The 1575Y mutation was found only in An. gambiae s.s. and not An. arabiensis (n = 191). 1575Y was found at low frequencies in 2008 M form collections from Burkina Faso (0.053; 95% CI = 0.015–0.173); by 2010, the frequency had risen significantly (χ² P = 0.021) to 0.172 (95% CI = 0.143–0.206) (Fig. 2). In S form, 1575Y frequency was highest in 2009 (0.35; 95% CI = 0.271–0.439) but dropped significantly to 0.224 (95% CI = 0.2–0.249) in 2010 (Fig. 2), during which 1014F was almost at fixation (0.992–0.968).

L1014F and N1575Y frequencies in collections from Ghana and Benin are provided in Table 1. 1575Y was not found in 2006 M form An. gambiae s.s. collections from Ghana, whereas 1014F is

Table 1. Allele frequencies (with 95% CI) of L1014F and N1575Y in annual collections (2008-2010) of An. gambiae s.s. and An. arabiensis populations from across West/Central Africa

<table>
<thead>
<tr>
<th>Species/collection site</th>
<th>Year</th>
<th>L1014F (95% CI)</th>
<th>n (alleles)</th>
<th>N1575Y (95% CI)</th>
<th>n (alleles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>An. gambiae M form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2008</td>
<td>0.053 (0.015–0.173)</td>
<td>38</td>
<td>0.368 (0.234–0.527)</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>0.075 (0.037–0.146)</td>
<td>94</td>
<td>0.404 (0.311–0.505)</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>0.172 (0.143–0.206)</td>
<td>552</td>
<td>0.562 (0.520–0.602)</td>
<td>552</td>
</tr>
<tr>
<td>Benin</td>
<td>2011</td>
<td>0.146 (0.096–0.217)</td>
<td>130</td>
<td>0.838 (0.767–0.891)</td>
<td>136</td>
</tr>
<tr>
<td>Yaoundé, Cameroon</td>
<td>2006</td>
<td>0</td>
<td>40</td>
<td>0.225 (0.123–0.375)</td>
<td>40</td>
</tr>
<tr>
<td>Accla, Ghana</td>
<td>2006</td>
<td>0</td>
<td>26</td>
<td>0.179 (0.079–0.356)</td>
<td>28</td>
</tr>
<tr>
<td>An. gambiae S form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2008</td>
<td>0.303 (0.242–0.372)</td>
<td>188</td>
<td>0.846 (0.787–0.890)</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>2009</td>
<td>0.35 (0.271–0.439)</td>
<td>120</td>
<td>0.992 (0.954–0.999)</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>0.224 (0.200–0.249)</td>
<td>1114</td>
<td>0.968 (0.956–0.977)</td>
<td>1114</td>
</tr>
<tr>
<td>Benin</td>
<td>2011</td>
<td>0.262 (0.180–0.365)</td>
<td>84</td>
<td>0.941 (0.868–0.974)</td>
<td>84</td>
</tr>
<tr>
<td>Yaoundé, Cameroon</td>
<td>2006</td>
<td>0.075 (0.026–0.199)</td>
<td>40</td>
<td>0.950 (0.835–0.986)</td>
<td>40</td>
</tr>
<tr>
<td>Accla, Ghana</td>
<td>2006</td>
<td>0.162 (0.112–0.230)</td>
<td>148</td>
<td>0.938 (0.892–0.965)</td>
<td>176</td>
</tr>
<tr>
<td>An. arabiensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>2009</td>
<td>0</td>
<td>164</td>
<td>0</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>0</td>
<td>380</td>
<td>0.011 (0.003–0.029)</td>
<td>380</td>
</tr>
</tbody>
</table>

For 2008 and 2009 Burkina Faso collections, allele frequencies are based on data from nonphenotyped (control) samples. Note that Cameroonian samples were chosen to preferentially include 1014F carriers to increase the likelihood of detecting 1575Y genotypes; therefore, allele frequencies may be unrepresentative of this population. See Fig. 1 for sample site locations.
Haplotype Analysis. The 1575Y variant was inextricably linked to 1014F, and 1575Y homozygotes were detected exclusively in 1014F homozygotes (Table S1) suggesting these mutations occur on the same haplotype. Confirmation came through sequencing 55 individuals for exons 7–10, 21, 28–30 and 32–33. The haplotype network was based on 49 individuals with six excluded due to missing data or uncertainty in haplotype construction (n = 98 haplotypes). Seventeen haplotypes (An. gambiae n = 15; An. arabiensis n = 2) were recovered using PHASE with only a single 1575Y bearing haplotype regardless of molecular form (M or S) or sampling location (Fig. 3). The same 15 An. gambiae haplotypes were recovered in 10 of 10 runs of PHASE. Seven haplotypes carried 1014F (including the 1575Y-bearing haplotype) and form a distinct cluster in the statistical parsimony network (Fig. 3) separated from the other eight An. gambiae haplotypes carrying 1014L by a minimum of seven mutations.

Sequencing revealed a further 31 VGSC SNPs (Dataset S1), of which 16 were intronic, nine synonymous and six noncoding. Of the nonsynonymous SNPs, one (I1532T) encodes an isoleucine to threonine change at position 1532. The remaining five are found within the variable intracellular carboxyl terminus loop 3 of the IVS6 helix, a region where no resistance-associated mutations have been detected previously in insects (23). For I1532T, 3 of 40 M form samples from Goundry were C/T heterozygotes (frequency = 0.038; 95% CI = 0.013–0.105).

Sequencing of Exon 20, the Site of Superkdr. The two copies of duplicated exon 20 (20c and 20d) were sequenced from both An. gambiae s.s. M and S forms carrying 1014-FF and 1575–YY to screen for additional mutations within the sodium channel binding pocket. No additional variation was detected.

Extended Haplotype Homozygosity (EHH). EHH analysis was used to compare patterns of LD of 1575N, 1014F, and 1575Y, assessed across 17 SNPs. For 1014F, LD is complete in the centromeric direction but EHH decays in the telomeric direction (Fig. 4A), consistent with previous observations (7). For 1575N, EHH decays rapidly in both directions (Fig. 4B), whereas there is complete LD for 1575Y. It should be noted that the 95% CIs of EHH are entirely nonoverlapping just a few base pairs away from the core for wild type versus mutant comparisons at both 1014F (telomeric) and 1575Y (both telomeric and centromeric) (Fig. 4A and B). The single haplotype of 1575Y and decay of LD for both 1014F and 1575N suggest that 1575Y is currently undergoing a selective sweep in An. gambiae s.s. populations from West Africa.

Association of N1575Y with Resistance. Because 1575Y occurs solely on a 1014F-bearing haplotype, we applied haplotypic association tests (Fig. 5) to investigate whether there is an additive benefit of 1575Y over and above the resistance benefit conferred by 1014F. Additive benefits of 1575Y were detected for M form samples phenotyped with DDT (OR = 2.6; 1.05–6.48) and S form samples phenotyped with permethrin (OR = 1.93; 1.24–3.0).

At a low frequency in the subsample from this population (0.179; 95% CI = 0.079–0.356). 1575Y was present in 2006 S form populations from Ghana with an allele frequency of 0.162 (95% CI = 0.112–0.23), whereas 1014F was at high frequency (0.938; 95% CI = 0.892–0.965).

In M form samples from Benin, 1575Y was found at a frequency of 0.146 and was highest in the northern site of Malanville (0.321, n = 56 alleles), whereas in the south, 1575Y was detected only from Houeyliho (1 of 54 alleles) (Table 1). 1575Y frequency was somewhat higher in S form (0.262), but not significantly so (P = 0.053).
alytical power (resistance to DDT in M form samples and per-
ci with the resistance phenotype (8). However, in samples
L1014F
lotype to replace
L1575N
codon 1014.

sweep detected and described by Lynd et al. (7) for mutations at
acted on a

detected. In S form samples, there was no signi-
cant increase in

Because
N1575Y
occurs on a 1014F background, disentangling the fitness benefits gained as a result of
N1575Y
from those of
L1014F
is problematic given that
L1014F
is itself strongly asso-
ci with the resistance phenotype (8). However, in samples
where we had sufficient phenotyped samples for reasonable an
talytical power (resistance to DDT in M form samples and per-
methrin in S form
An. gambiae
), a significant additive benefit of
N1575Y
was detectable. 1014F and 1014S, although not directly in
the binding pocket, are thought to produce their resistance phenotype through altering the confirmation of the VGSC,

Discussion
In this study, we describe a mutation (N1575Y) in the
An. gambiae
s.s. VGSC. The mutation, identified originally in
An. gambiae
s.s. S form from Burkina Faso, was shown to be present in both M and S forms collected throughout West/Central Africa over a range of some 2,000 km at frequencies of up to 30%. Although the mutation is found over such a large swath of the continent, only a single
N1575Y haplotype was detected following sequencing of amplicons distributed over 40 kbp of the sodium channel, and occurred upon a 1014F haplotypic background. In a temporal series of collections, a significant increase in
N1575Y frequency from 2008 to 2010 in M form samples has also been detected. In S form samples, there was no significant increase in
N1575Y; however, in these samples, 1014F approached fixation over this time period. Because
1014F alone confers such a strong selective benefit, it may take longer for the
1014F-1575Y hap-
lotype to replace
1014F-1575N than it does when the
1014L-1575Y haplotype dominates, as is it does in M form mosquitoes.

The long-range haplotype bearing
N1575Y, the moderate but increasing frequencies of
N1575Y in M form samples, and the widespread occurrence throughout West/Central Africa, all bear the hallmarks of a mutation under recent and strong selection pressure. Indeed, the rapid rise in frequency of such a beneficial mutation has produced an associated reduction in haplotype diversity at linked loci through a recent positive “hard sweep” (16). EHH analysis demonstrated that, compared with the marked LD associated with
N1575Y (EHH = 1; Fig. 3), LD sur-
drounding coexisting
1014F (telomeric) and
N1575N (telomeric and centromeric) alleles decays rapidly. As this selective sweep has acted on a
1014F bearing haplotype, it overlays the selective sweep detected and described by Lynd et al. (7) for mutations at
codon 1014.

Because
N1575Y occurs on a 1014F background, disentangling the fitness benefits gained as a result of
N1575Y from those of
L1014F is problematic given that
L1014F is itself strongly asso-
ci with the resistance phenotype (8). However, in samples
where we had sufficient phenotyped samples for reasonable an
talytical power (resistance to DDT in M form samples and per-
methrin in S form
An. gambiae
), a significant additive benefit of
N1575Y was detectable. 1014F and 1014S, although not directly in
the binding pocket, are thought to produce their resistance phenotype through altering the confirmation of the VGSC,

preventing binding of insecticide (9, 24). By contrast,
N1575Y occurs within the linker between domains III and IV, the site of the inactivation particle, a sequence of three amino acids (MFM in mammals and IFM in insects), which closes the sodium channel pore following activation, stopping influx of sodium ions into the cell so permitting restoration of the membrane resting potential.

Mutations within the DIII-DIV linker have been identified previously in resistant insects: In pyrethroid resistant tobacco budworm (Heliothis virescens) and cotton bollworm (Helicoverpa armigera) (25), although their role in resistance was not confirmed. In varroa mites, an L-P change at 1770 (I1599P hously numbering) has been demonstrated to effect sensitivity to fluvalinate (a pyrethroid) when expressed in a Xenopus system (26). A G-R mutation at position 1575 in the inactivation particle (position 1559 house numbering) also underlies a cold-sensi-
tive phenotype in Drosophila (27). Interestingly, the G1575R mutation occurs as a double mutation with I1545S (1533 in
Musca
) and the Drosophila Ocd strain carrying these two mutations are 1000-fold more resistant to DDT than the Oregon-R strain, the progenitor of
Ocd (27) suggesting that one or both of these mutations are strongly involved in DDT resistance. I1545S lies in exon 30 and this was covered by our sequencing for the haplotype analysis. No mutation close to this region was detected but the Drosophila story indicates that mutations in the in-
activation particle can interact with mutations elsewhere in the sodium channel to alter the resistance profile. To fully un-
derstand the role of
N1575Y in the physiological response to pyrethroids and DDT in
An. gambiae
this mutation will require expression in a Xenopus system (for example, refs. 24 and 28). Such electrophysiological studies have not only demonstrated the impact of
1014F and 1014S on neuronal response in the presence of insecticide but, additionally, indicated that, in the absence of insecticide, knockdown resistance mutations may exhibit different response profiles in comparison with wild-type alleles (24), which may manifest as a slight fitness detriment. In resistant bacteria, the costs of antibiotic resistance in the absence of selection pressure are known to be ameliorated by compensatory mutations co-occurring in the antibiotic target site (29). No such compensatory mutations have been described to date within insects resistant to pyrethroids/DDT conferred by
kdr. However, in the rat brain sodium channel a mutation at position

Fig. 4. The pattern of LD decay and recombination with increasing distance from the core in the centromeric (left) and telomeric (right) direction. Bifurcation plots (inset) show patterns of recombination for
1014F (A) and
1575N (B) over the 17 SNPs analyzed. The core SNP is represented by the dark blue circle and each of the additional SNPs is represented by a node from which bifurcation indicates a recombination event. (main plot) EHH analysis of
1014F (A) and
1575N (B). LD decay is shown with increasing distance relative to the core (k = 0). The EHH value at each SNP for both alleles is shown with a small black circle (+).

1329 (equivalent to 1410 in housefly numbering) in domain III SIV-V disrupts fast inactivation, but normal inactivation is restored by a compensatory opposing charge mutation within the linker at position 1489 (1565 housefly numbering and part of the IFM inactivation particle motif) (30). Although we provide evidence for an additive benefit of 1575Y, we cannot preclude the possibility that in wild populations, N1575Y may compensate for fitness costs of 1014F in the absence of insecticide exposure which would also result in the signals of positive selection that we have detected.

An additional I-T mutation at position 1552 was also detected at low frequency in M form samples. Position 1552 is located within the III S6 helix which forms one side of the pyrethroid/DDT binding site. Three nearby residues (F1534, G1535 and F1538) have already been implicated in resistance in other insect species (23). In Aedes aegypti, F1534C is correlated with both permethrin and DDT resistance (22). However, the I1532 side-chain actually points away from the binding pocket toward the channel pore and so at present the role of this mutation in re-

channel actually points away from the binding pocket toward the
permethrin and DDT resistance (22). However, the I1532 side-

chain actually points away from the binding pocket toward the channel pore and so at present the role of this mutation in re-

resistance awaits further investigation.

Given the recognized role of L1014F and L1014S in confer-

ring resistance phenotypes in An. gambiae (8) it is understand-

able that studies have focused overwhelmingly on genotyping these kdr markers in studies of insecticide resistance. However, this approach neglects other resistance mutations which may be present with the sodium channel. The identification of a mu-

tation involved in resistance suggests that there is merit in ex-

ploring the sodium channel for additional resistance mutations. Detection of the mutation at an early stage presents an ideal opportunity for modeling studies predicting spread and esti-
mating selection coefficients (31). The N1575Y TaqMan assay will facilitate this.

Materials and Methods

Sample Collections. An. gambiae s.s. females were collected from 10 sites across West/Central Africa, and An. arabiensis females were collected from 4 sites in Burkina Faso (Fig. 1). Burkina samples (342 M form, 711 S form and 272 An. arabiensis) were collected in 2008–2010 from four locations (de-
tailed sample site information in ref. 32). Mosquitoes from Burkina Faso were phenotyped for permethrin (0.75%), deltamethrin (0.05%), or DDT (4%) in WHO susceptibility tests (33).

Both M and S molecular forms of An. gambiae were collected in 2006 from Accra, Ghana, and Yaoundé, Cameroon (18). At the time of collection, the 1014F allele frequency in these populations was high in An. gambiae s.s. S form (0.98 and 0.87, respectively) and rare in M form (~0.01) (18). More recent collections in Ghana show a marked increase in 1014F allele fre-

quency in M form samples (7). A subset of the 2006 samples were included in this study; 88 S form and 16 M form from Ghana, and 20 S form and 20 M form from Cameroon were screened.

A total of 112 An. gambiae s.s. from four sites in Benin were included in the analysis. M and S forms of An. gambiae s.s. were collected from three sites in the south: Houeyiho (M form; n = 30), Bohicon (S form; n = 20; M form n = 4), and Tori Bossito (S form; n = 23, M form n = 6) and one site in the north, Malanville (29 M form) (Fig. 1).

All samples were distinguished using SINE PCR (34).

Targeted Sequencing of Exon 30 of the Voltage-Gated Sodium Channel. A total of 330 bp of exon 30 was amplified (for primers, see Table S2) from S form An. gambiae s.s. from Kuinima (n = 12) and Soumouso (n = 9) in Burkina Faso where DDT and pyrethroid resistance is widespread, and two samples from the insecticide susceptible Kismu laboratory-colony. One nonsynonymous mutation was observed in the resistant mosquitoes, an asparagine to tyro-

sine at position 1575 (N1575Y).

Pyrosequencing of N1575Y. Two genotyping assays for N1575Y were de-

veloped. The first, a pyrosequencing assay, interrogated the mutation con-
taining sequence 5′-ATATGATGAATGAA-3′. PCRs were performed using the primers listed in Table S2. Reactions (20 μL) contained 0.4 μM each primer, 0.2 mM dNTPs, 1.3 mM MgCl2, 0.4 U of KapaTaq (KAPA Biosystems) and 1–5 ng of template DNA using cycling conditions of 95 °C for 2 min followed by 40 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 10 s. PCR products (made up to 40 μL with dH2O) were used as templates for pyrosequencing. Reactions were performed using PyroMark Gold Reagents (Qiagen) according to the manu-

facturer’s instructions with the sequencing primer AG_VGSC_EX23Seq (Table S2) and run on the PyroMark Q96 System (Qiagen).

TaqMan Genotyping of the N1575Y Mutation. To complement the suite of TaqMan assays already developed for Anopheles research (35), a custom TaqMan assay was developed for N1575Y. Primer and probe sequences are provided in Table S2. TaqMan reactions were undertaken in 10-μL volumes containing 1× SensiMix (Bioline), 800 nM each primer, and 200 nM each probe on an Agilent MX3005P with cycling conditions of 10 min at 95 °C followed by 40 cycles of 15 s at 92 °C and 1 min at 60 °C.

All specimens were genotyped for L1014F using a TaqMan assay (36). A subset of samples was also genotyped for L1014S; however, no An. gambiae s.s. individuals were found to carry this allele (n = 119).

Identification of Haplotypes. Long-range sodium channel haplotypes were determined for 55 individuals through amplification and direct sequencing of four regions of the sodium channel gene covering exons 7–10, 21 (site of the 1014 codon), 28–30, and 32–33 (numbered following Davies et al.; ref. 37). Amplification primers are given in Table S2. To facilitate phasing of hap-

lotypes, only individuals homozygous at the 1014 and 1575 codons were included (with the exception of three N1575Y heterozygotes from Came-

eroon where 1575Y homozygotes were not found). Representative samples of 1014-FF 1575-N, 1014-FF 1575-N, and 1014-FF 1575-Y from both M and S form individuals were sequenced. Genotypes at variable positions served as input for PHASE (38). Haplotypes were identified following 10 runs of PHASE with seed values altered for each run. Ambiguous positions were resolved through cloning. Unambiguous haplotypes were used to create a haplotype genealogy in TCS (39) with default conditions (95% connection limit; gaps treated as a fifth state).

EHH Analysis. The pattern of LD for the 1575 and 1014 haplotypes was inferred using EHH analysis (40) on all unequivocal haplotypes from An. gambiae (n = 69 haplotypes for 1014F core and n = 61 for 1575N core). The significance of differences in EHH values between 1014 and 1575 haplotypes
were determined by nonoverlapping 95% CI at each SNP position calculated using a bootstrapping approach (n = 1,000). EHH analysis was performed in R (41). To visualize the breakdown of LD extending from the core region, haplotype bifurcation plots were created using SWEEP (40).

Sequencing of Exon 20 Mutations. Within the putative binding pocket of the insect VGSC, a series of key amino acid residues that interact with DDT and pyrethroids, have been identified and are associated with resistance to DDT and pyrethroids (9, 37). These include the so-called super kdr mutation at codon 918, which exists as a double mutation with 1014F in houseflies and confers additive resistance to permethrin (42). This region is therefore a prime candidate for the emergence of resistant mutations at codons other than 1014. To confirm that 10157S is not a marker of an additional mutation within this site, a ~400-bp region of exon 20 was sequenced from one An. gambiae s.s. M form individual from all four sites in Burkina Faso carrying 1014FF and 1575YY. Exon 20 is present in two alternatively spliced copies (Exon 20c/d) in An. gambiae s.s. (37). Therefore, primers specific to each copy were designed in flanking introns.

15. Karasov T, Messer PM, Petrov DA (2010) Evidence that adaptation in insect VGSC, a series of key amino acid residues that interact with DDT and pyrethroids, have been identified and are associated with resistance to DDT and pyrethroids (9, 37). These include the so-called super kdr mutation at codon 918, which exists as a double mutation with 1014F in houseflies and confers additive resistance to permethrin (42). This region is therefore a prime candidate for the emergence of resistant mutations at codons other than 1014. To confirm that 10157S is not a marker of an additional mutation within this site, a ~400-bp region of exon 20 was sequenced from one An. gambiae s.s. M form individual from all four sites in Burkina Faso carrying 1014FF and 1575YY. Exon 20 is present in two alternatively spliced copies (Exon 20c/d) in An. gambiae s.s. (37). Therefore, primers specific to each copy were designed in flanking introns.

15.7327 Frequency in Gondry M Form. An isoleucine to threonine substitution was detected at position 1532 in the VGSC (2L-242B617). Subsequently, 40 M form samples from Gondry were amplified with Ex28F and Ex28R and sequenced with Ex28R to ascertain the frequency of this mutation.

Data Analysis. Haplotypic association tests for 1014L-1575N, 1014F-1575N, and 1014F-1575Y and insecticide resistance were conducted in Haplovie v4.1 (43).

ACKNOWLEDGMENTS. We thank N’Falé Sagnon (Centre National de Recherche sur la Formation à la Moustiquiculture, Ouagadougou) and Vincent Corbel (CREC) and their respective teams for samples from Burkina Faso and Benin. We also thank Martin Williamson, Emry Davies and Andy O’Reilly ( Rothamsted Research) for assistance, helpful information, and discussions on the nature of these mutations; and Stu Field (Colorado State University) for assistance with R scripts. The research leading to these results has received partial funding from the European Union Seventh Framework Programme FP7 (2007–2013) under Grant Agreement 265560 AveNet, partial financial support from the United Nations Children’s Fund/United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (WHO/TDR), and support from Award R01AI082734 from the National Institute of Allergy and Infectious Diseases.