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

Christopher M. Jones[a], Milindu Liyanapathirana[a], Fiacre R. Agossa[b], David Weetman[a], Hilary Ranson[a], Martin James Donnelly[a], and Craig S. Wilding[a,1]

[a]Vector Group, Liverpool School of Tropical Medicine, Liverpool L3 5QA, United Kingdom; and [b]Centre de Recherche Entomologique de Cotonou (CREC), 06 BP 2604, Cotonou, Bénin

Edited by Barry J. Beaty, Colorado State University, Fort Collins, CO, and approved March 20, 2012 (received for review January 26, 2012)

Insecticide resistance is an ideal model to study the emergence and spread of adaptative variants. In the African malaria mosquito, *Anopheles gambiae*, this is complemented by a strong public health rationale. In this insect, resistance to pyrethroid 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 acting on a recent mutant. *1575Y* occurs ubiquitously on a *1014F* haplotypic background, suggesting that the *N1575Y* mutation compensates for deleterious fitness effects of *1014F* and/or confers additional resistance to insecticides. Haplotypic tests of association suggest the latter: The *1014F-1575Y* haplotype confers a significant additive benefit above *1014F-1575N* for survival to DDT (M form $P = 0.03$) and permethrin (S form $P = 0.003$).

kdr | selective sweep | inactivation particle

The impetus to eradicate malaria has yielded a significant reduction of malaria mortality and morbidity via antiparasite artemisinin combination therapies (1) and scaling up of coverage with insecticide-based interventions (2). Vector control interventions are based primarily around either provision of long-lasting insecticide-treated nets (LLINs) or indoor residual spraying of insecticide onto surfaces where mosquitoes rest after blood feeding. A major threat to the success of these interventions is the development of insecticide resistance in malaria vectors (3, 4). Resistance is a particular threat to LLINs, as there is currently only one class of insecticides, the pyrethroids, approved by WHO for impregnation of bednets. Resistance is typically evaluated through phenotypic bioassays, although DNA-based diagnostics are supplementing such assays as part of resistance monitoring strategies (3).

In Sub-Saharan Africa, the primary vectors of malaria are *Anopheles gambiae sensu stricto* and *Anopheles arabiensis*. In *An. gambiae s.s.*, two molecular forms, M and S, are recognized, occurring in sympatry throughout West and Central Africa (5). Interform gene flow is geographically variable (6) but sufficient to allow introgression and spread of selected alleles (7). A series of knockdown resistance (*kdr*) mutations in the sodium channel are the best characterized resistance mechanisms in *Anopheles*. The presence of *kdr* has been conclusively linked to reduced mortality following exposure to both DDT and pyrethroids in a large number of studies (3, 8). Pyrethroids and DDT target the insect voltage-gated sodium channel (VGSC), binding to the open (activated) sodium channel pore and preventing inactivation (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

Author contributions: C.M.J., H.R., and C.S.W. designed research; C.M.J., M.L., F.R.A., and C.S.W. performed research; C.M.J., D.W., M.J.D., and C.S.W. analyzed data; and C.M.J., D.W., H.R., M.J.D., and C.S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The SNPs reported in this paper have been deposited in the dbSNP Short Genetic Variations database, www.ncbi.nlm.nih.gov/projects/SNP (submitter SNP accession nos. are listed in Dataset S1).

[1]To whom correspondence should be addressed. E-mail: c.s.wilding@liverpool.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201475109/-/DCSupplemental.

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

Species/collection site	Year	(f) *1575Y* (95% CI)	*n* (alleles)	(f) *1014F* (95% CI)	*n* (alleles)
An. gambiae M form					
Burkina Faso	2008	0.053 (0.015–0.173)	38	0.368 (0.234–0.527)	38
	2009	0.075 (0.037–0.146)	94	0.404 (0.311–0.505)	94
	2010	0.172 (0.143–0.206)	552	0.562 (0.520–0.602)	552
Benin	2011	0.146 (0.096–0.217)	130	0.838 (0.767–0.891)	136
Yaoundé, Cameroon	2006	0	40	0.225 (0.123–0.375)	40
Accra, Ghana	2006	0	26	0.179 (0.079–0.356)	28
An. gambiae S form					
Burkina Faso	2008	0.303 (0.242–0.372)	188	0.846 (0.787–0.890)	188
	2009	0.35 (0.271–0.439)	120	0.992 (0.954–0.999)	120
	2010	0.224 (0.200–0.249)	1114	0.968 (0.956–0.977)	1114
Benin	2011	0.262 (0.180–0.365)	84	0.941 (0.868–0.974)	84
Yaoundé, Cameroon	2006	0.075 (0.026–0.199)	40	0.950 (0.835–0.986)	40
Accra, Ghana	2006	0.162 (0.112–0.230)	148	0.938 (0.892–0.965)	176
An. arabiensis					
Burkina Faso	2009	0	164	0	164
	2010	0	380	0.011 (0.003–0.029)	380

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.

(7) and identification of additional mutations 3′ of 1014 in *Aedes aegypti* (22) prompted us to search for additional functional variants in the VGSC of *An. gambiae*.

Here, we identify an asparagine-to-tyrosine mutation at position 1575 (*N1575Y*) on the domain III-IV linker of the *An. gambiae s.s.* VGSC. We show that this mutation occurs only on a *1014F* haplotype, has a single origin, and has spread widely across West/Central Africa. Extended haplotype homozygosity (EHH) analysis of coexisting *1014F* 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 Soumousso and Kuinima (Burkina Faso) and the insecticide-susceptible Kisumu colony. An A-to-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 (χ^2 $P = 0.021$) 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

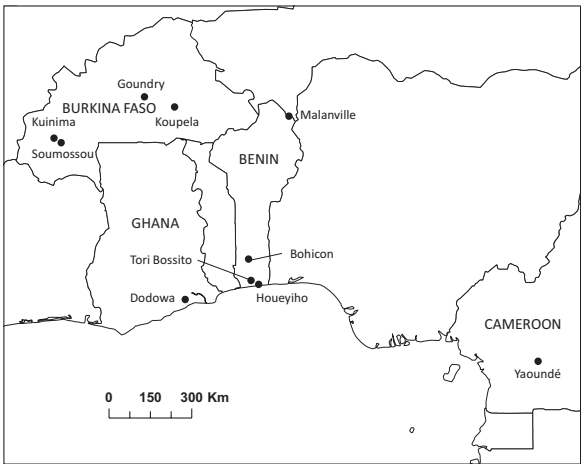

Fig. 1. Map of sample sites from West Africa used in this study.

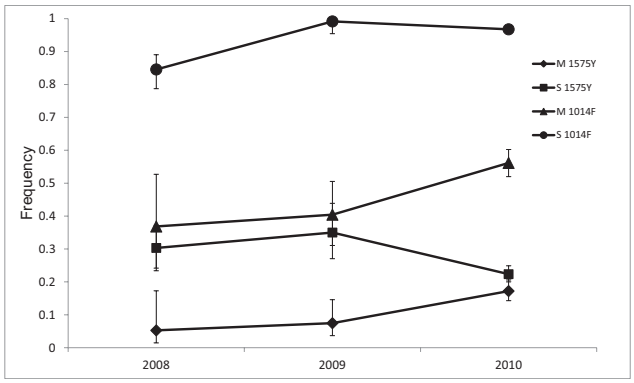

Fig. 2. The pattern of *1014F* and *1575Y* frequency in *An. gambiae s.s.* from Burkina Faso between 2008 and 2010. Error bars show 95% CI.

EVOLUTION

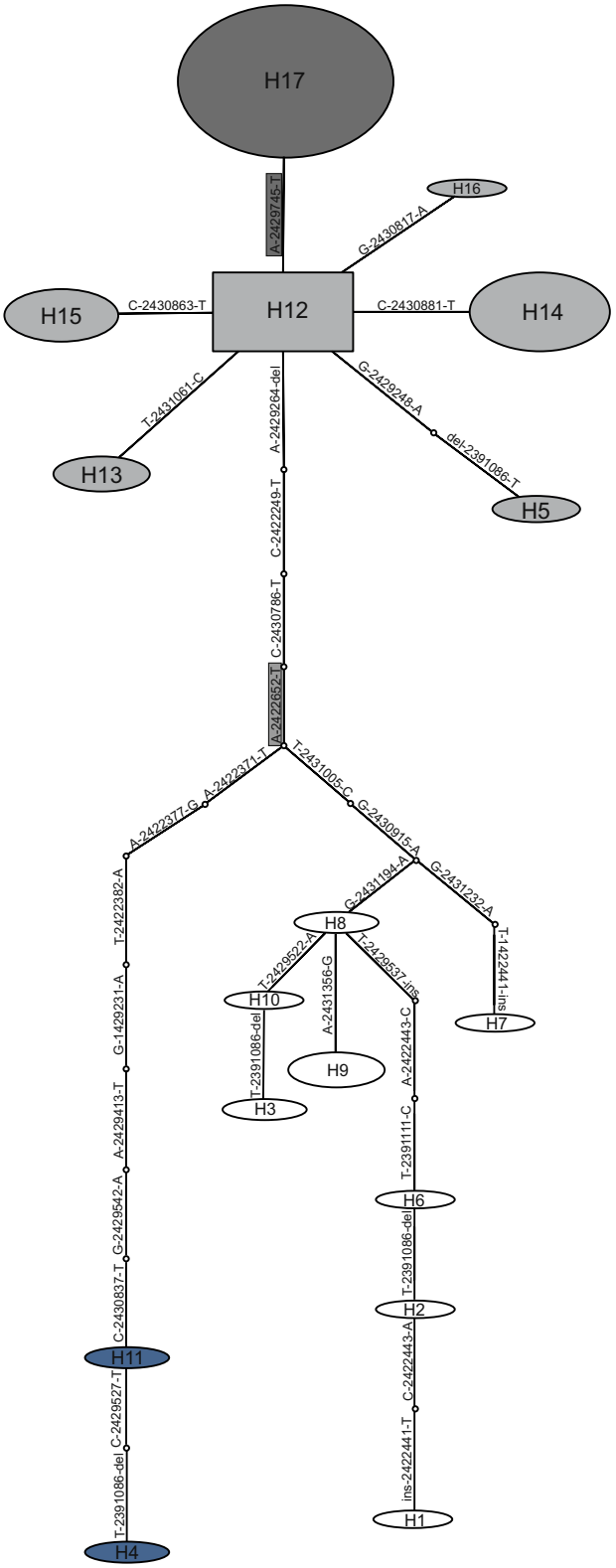

Fig. 3. Network of *An. gambiae sensu lato* VGSC haplotypes. Haplotypes bearing *1014F* are shaded gray with the *1575Y* haplotype (H17) shaded dark gray. The core *1014F*-containing haplotype is depicted in a rectangle. Two haplotypes from *An. arabiensis* (H4 and H11 shaded blue) were included, but their position is equivocal: An alternative network exists in which the positions of all *An. gambiae* haplotypes are identical, but the *An. arabiensis* haplotypes are joined to a node between haplotype 12 and the *1014L An. gambiae* sequences by a reticulation. Polygon size corresponds to the sample number with that haplotype, although samples for haplotype analysis were biased to preferentially choose *1014F* carriers. Each node represents a segregating mutation (base change given above branches).

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 Houeyiho (1 of 54 alleles) (Table 1). *1575Y* frequency was somewhat higher in S form (0.262), but not significantly so ($P = 0.053$).

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 nonsynonymous. 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 Super *kdr*. 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. 4*A*), consistent with previous observations (7). For *1575N*, EHH decays rapidly in both directions (Fig. 4*B*), 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 *1014* (telomeric) and *1575* (both telomeric and centromeric) (Fig. 4 *A* 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).

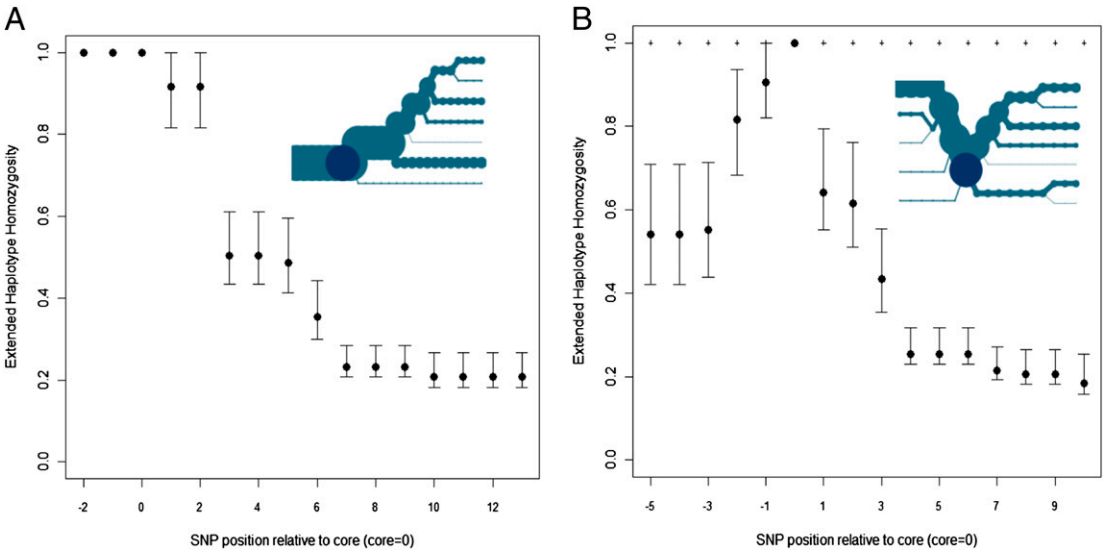

Fig. 4. The pattern of LD decay and recombination with increasing distance from the core in the centromeric (left) and telomeric (right) direction. Bifurcation plots (*Inset*) show patterns of recombination for *1014F* (*A*) and *1575N* (*B*) over the 17 SNPs analyzed. The core SNP is represented by 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 ($x = 0$). The EHH value at each SNP for both alleles is shown with a small black circle (•) and with associated 95% CIs estimated by bootstrapping. The single haplotype (EHH = 1) of *1575Y* in *B* is indicated with the symbol +.

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 swathe of the continent, only a single *1575Y* 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 *1575Y* frequency from 2008 to 2010 in M form samples has also been detected. In S form samples, there was no significant increase in *1575Y*; 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* haplotype to replace *1014F-1575N* than it does when the *1014L-1575N* haplotype dominates, as is it does in M form mosquitoes.

The long-range haplotype bearing *1575Y*, the moderate but increasing frequencies of *1575Y* 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 *1575Y* (EHH = 1; Fig. 3), LD surrounding coexisting *1014F* (telomeric) and *1575N* (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 *1575Y* 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 associated with the resistance phenotype (8). However, in samples where we had sufficient phenotyped samples for reasonable analytical power (resistance to DDT in M form samples and permethrin in S form *An. gambiae*), a significant additive benefit of *1575Y* 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 (*L1596P* housefly 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 housefly numbering) also underlies a cold-sensitive phenotype in *Drosophila* (27). Interestingly, the *G1575R* mutation occurs as a double mutation with *I1545M* (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. *I1545M* 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 inactivation particle can interact with mutations elsewhere in the sodium channel to alter the resistance profile. To fully understand 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

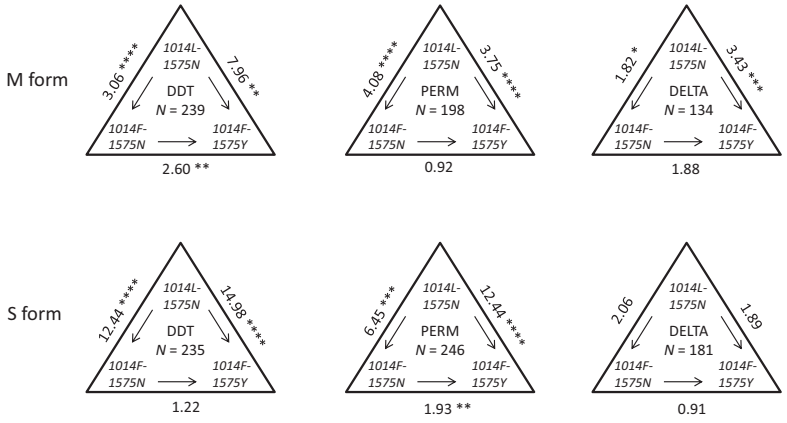

Fig. 5. Results of haplotypic association tests for the three 1014–1575 haplotypes encountered in this study (*1014L-1575N*, *1014F-1575N*, *1014F-1575Y*) with resistance phenotype to DDT, permethrin (perm), and deltamethrin (delta). Phenotype (alive/dead) was determined following 1-h exposure to 4% DDT, 0.75% permethrin, or 0.05% deltamethrin followed by 24-h recovery. Population mortality data are provided in Table S3. Odds ratios (ORs) are given with significance indicated with asterisks. The arrow within the triangle indicates the direction of OR calculation (e.g., M form DDT phenotyped individuals carrying the *1014F-1575Y* haplotype are 2.60 times more likely to be resistant than individuals carrying the *1014F-1575N* haplotype).

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 1532 was also detected at low frequency in M form samples. Position 1532 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 sidechain actually points away from the binding pocket toward the channel pore and so at present the role of this mutation in resistance awaits further investigation.

Given the recognized role of *L1014F* and *L1014S* in conferring resistance phenotypes in *An. gambiae* (8) it is understandable 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 mutation involved in resistance suggests that there is merit in exploring 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 estimating 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). Burkinabe samples (342 M form, 711 S form and 272 *An. arabiensis*) were collected in 2008–2010 from four locations (detailed 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 frequency 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 Soumousso ($n = 9$) in Burkina Faso where DDT and pyrethroid resistance is widespread, and two samples from the insecticide susceptible Kisumu laboratory-colony. One nonsynoymous mutation was observed in the resistant mosquitoes, an asparagine to tyrosine at position 1575 (*N1575Y*).

Pyrosequencing of *N1575Y*. Two genotyping assays for *N1575Y* were developed. The first, a pyrosequencing assay, interrogated the mutation containing sequence 5′-AT[A/T]ATGCAATGAA-3′. PCRs were performed using the primers listed in Table S2. Reactions (20 μL) contained 0.4 μM each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.4 U 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 dH_2O) were used as templates for pyrosequencing. Reactions were performed using PyroMark Gold Reagents (Qiagen) according to the manufacturer's instructions with the sequencing primer AG_VGSC_EX29seq (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 haplotypes, only individuals homozygous at the 1014 and 1575 codons were included (with the exception of three *N1575Y* heterozygotes from Cameroon where *1575Y* homozygotes were not found). Representative samples of *1014-LL 1575-NN*, *1014-FF 1575-NN*, and *1014-FF 1575-YY* 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 *N1575Y* 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.* S and M form individual from all four sites in Burkina Faso carrying *1014-FF* and *1575-YY*. 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.

***I1532T* Frequency in Goundry M Form.** An isoleucine to threonine substitution was detected at position 1532 in the VGSC (2L:2428617). Subsequently, 40 M form samples from Goundry 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 Haploview v4.1 (43).

ACKNOWLEDGMENTS. We thank N'Falé Sagnon (Centre National de Recherche et de Formation sur le Paludisme CNRP, Ouagadougou) and Vincent Corbel (CREC) and their respective teams for samples from Burkina Faso and Benin. We also thank Martin Williamson, Emyr 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 265660 AvecNet, 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.

1. Hastings I (2011) How artemisinin-containing combination therapies slow the spread of antimalarial drug resistance. *Trends Parasitol* 27:67–72.
2. Lim SS, et al. (2011) Net benefits: A multicountry analysis of observational data examining associations between insecticide-treated mosquito nets and health outcomes. *PLoS Med* 8:e1001091.
3. Ranson H, et al. (2011) Pyrethroid resistance in African anopheline mosquitoes: What are the implications for malaria control? *Trends Parasitol* 27:91–98.
4. malERA Consultative Group on Vector Control (2011) A research agenda for malaria eradication: Vector control. *PLoS Med* 8:e1000401.
5. della Torre A, Tu Z, Petrarca V (2005) On the distribution and genetic differentiation of *Anopheles gambiae s.s.* molecular forms. *Insect Biochem Mol Biol* 35:755–769.
6. Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Mol Biol Evol* 29: 279–291.
7. Lynd A, et al. (2010) Field, genetic, and modeling approaches show strong positive selection acting upon an insecticide resistance mutation in *Anopheles gambiae s.s. Mol Biol Evol* 27:1117–1125.
8. Donnelly MJ, et al. (2009) Does *kdr* genotype predict insecticide-resistance phenotype in mosquitoes? *Trends Parasitol* 25:213–219.
9. O'Reilly AO, et al. (2006) Modelling insecticide-binding sites in the voltage-gated sodium channel. *Biochem J* 396:255–263.
10. Davies TG, Field LM, Usherwood PN, Williamson MS (2007) DDT, pyrethrins, pyrethroids and insect sodium channels. *IUBMB Life* 59:151–162.
11. Martinez-Torres D, et al. (1998) Molecular characterization of pyrethroid knockdown resistance (*kdr*) in the major malaria vector *Anopheles gambiae s.s. Insect Mol Biol* 7:179–184.
12. Ranson H, et al. (2000) Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol* 9:491–497.
13. ffrench-Constant RH, Daborn PJ, Le Goff G (2004) The genetics and genomics of insecticide resistance. *Trends Genet* 20:163–170.
14. Labbé P, et al. (2007) Forty years of erratic insecticide resistance evolution in the mosquito *Culex pipiens. PLoS Genet* 3:e205.
15. Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet* 6:e1000924.
16. Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20:R208–R215.
17. Neafsey DE, et al. (2010) SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330:514–517.
18. Weetman D, et al. (2010) Association mapping of insecticide resistance in wild *Anopheles gambiae* populations: Major variants identified in a low-linkage disequilbrium genome. *PLoS ONE* 5:e13140.
19. Harris C, Rousset F, Morlais I, Fontenille D, Cohuet A (2010) Low linkage disequilibrium in wild *Anopheles gambiae s.l.* populations. *BMC Genet* 11:81.
20. Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: Evidence from the Duffy blood group locus. *Am J Hum Genet* 66:1669–1679.
21. Tishkoff SA, et al. (2001) Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. *Science* 293:455–462.
22. Harris AF, Rajatileka S, Ranson H (2010) Pyrethroid resistance in *Aedes aegypti* from Grand Cayman. *Am J Trop Med Hyg* 83:277–284.
23. Davies TGE, Williamson MS (2009) Interactions of pyrethroids with the voltage-gated sodium channel. *Bayer CropScience J* 62:159–178.
24. Burton MJ, et al. (2011) Differential resistance of insect sodium channels with *kdr* mutations to deltamethrin, permethrin and DDT. *Insect Biochem Mol Biol* 41:723–732.
25. Head DJ, McCaffery AR, Callaghan A (1998) Novel mutations in the *para*-homologous sodium channel gene associated with phenotypic expression of nerve insensitivity resistance to pyrethroids in Heliothine lepidoptera. *Insect Mol Biol* 7:191–196.
26. Liu Z, Tan J, Huang ZY, Dong K (2006) Effect of a fluvalinate-resistance-associated sodium channel mutation from varroa mites on cockroach sodium channel sensitivity to fluvalinate, a pyrethroid insecticide. *Insect Biochem Mol Biol* 36:885–889.
27. Lindsay HA, et al. (2008) The dominant cold-sensitive Out-cold mutants of *Drosophila melanogaster* have novel missense mutations in the voltage-gated sodium channel gene *paralytic. Genetics* 180:873–884.
28. Vais H, Williamson MS, Devonshire AL, Usherwood PNR (2001) The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. *Pest Manag Sci* 57:877–888.
29. Handel A, Regoes RR, Antia R (2006) The role of compensatory mutations in the emergence of drug resistance. *PLOS Comput Biol* 2:e137.
30. Smith MR, Goldin AL (1997) Interaction between the sodium channel inactivation linker and domain III S4-S5. *Biophys J* 73:1885–1895.
31. Barbosa S, Black WCIV, Hastings I (2011) Challenges in estimating insecticide selection pressures from mosquito field data. *PLoS Negl Trop Dis* 5:e1387.
32. Ranson H, et al. (2009) Insecticide resistance in *Anopheles gambiae*: Data from the first year of a multi-country study highlight the extent of the problem. *Malar J* 8:299.
33. World Health Organization (WHO). (1998) *Test procedures for insecticide resistance monitoring in malaria vectors, bio-efficacy and persistence of insecticide on treated surfaces* (WHO, Geneva).
34. Santolamazza F, et al. (2008) Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malar J* 7:163.
35. Bass C, et al. (2010) The Vector Population Monitoring Tool (VPMT): High-throughput DNA-based diagnostics for the monitoring of mosquito vector populations. *Malar Res Treat* 2010:190434.
36. Bass C, et al. (2007) Detection of knockdown resistance (*kdr*) mutations in *Anopheles gambiae*: A comparison of two new high-throughput assays with existing methods. *Malar J* 6:111.
37. Davies TGE, Field LM, Usherwood PNR, Williamson MS (2007) A comparative study of voltage-gated sodium channels in the Insecta: Implications for pyrethroid resistance in Anopheline and other Neopteran species. *Insect Mol Biol* 16:361–375.
38. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989.
39. Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659.
40. Sabeti PC, et al. (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419:832–837.
41. R Development Core Team (2011) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
42. Williamson MS, Martinez-Torres D, Hick CA, Devonshire AL (1996) Identification of mutations in the housefly *para*-type sodium channel gene associated with knockdown resistance (*kdr*) to pyrethroid insecticides. *Mol Gen Genet* 252:51–60.
43. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.