

1           **The genetic architecture of target-site**  
2           **resistance to pyrethroid insecticides in the**  
3           **African malaria vectors *Anopheles gambiae***  
4           **and *Anopheles coluzzii***

5           Chris S. Clarkson<sup>1</sup>, Alistair Miles<sup>2,1</sup>, Nicholas J. Harding<sup>2</sup>,  
6           @@TODO<sup>1</sup>, Dominic Kwiatkowski<sup>1,2</sup>, Martin Donnelly<sup>3,1</sup>, and The  
7           *Anopheles gambiae* 1000 Genomes Consortium<sup>4</sup>

8           <sup>1</sup>Sanger @@TODO

9           <sup>2</sup>Oxford @@TODO

10           <sup>3</sup>Liverpool @@TODO

11           <sup>4</sup>MalariaGEN @@TODO

12           Work in progress

13           **Abstract**

14           Resistance to pyrethroid insecticides is a major concern for malaria vector control,  
15           because these are the only compounds approved for use in insecticide-treated bed-nets  
16           (ITNs) and are also widely used for indoor residual spraying (IRS). Pyrethroids target  
17           the voltage-gated sodium channel (VGSC), an essential component of the mosquito  
18           nervous system, but mutations in the *Vgsc* gene can disrupt the activity of these  
19           insecticides, inducing a “knock-down resistance” phenotype. Here we use Illumina  
20           whole-genome sequence data from phase 1 of the *Anopheles gambiae* 1000 Genomes

21 Project (Ag1000G) to provide a comprehensive account of genetic variation at the  
22 *Vgsc* locus in mosquito populations from 8 African countries. In addition to three  
23 known resistance variants that alter the protein-coding sequence of the *Vgsc* gene, we  
24 describe 19 previously unknown non-synonymous variants at appreciable frequency in  
25 one or more populations. For each variant we predict a resistance phenotype based on  
26 genetic evidence for recent selection, patterns of linkage between variants, the posi-  
27 tion of the variant within the protein structure, and experimental evidence from other  
28 species. We use analyses of haplotype structure to refine our understanding of the  
29 origins and spread of these resistance variants between species and geographical loca-  
30 tions. These analyses identify 10 distinct lineages, each of which carries one or more  
31 resistance alleles and appears to be undergoing rapid and recent expansion in one or  
32 more populations. The most successful and widespread resistance lineage (F1) origin-  
33 ates in West Africa and has subsequently spread to countries in Central and Southern  
34 Africa. We also reconstruct a putative ancestral haplotype for each lineage, and ana-  
35 lyse patterns of recombination to show that lineages are unrelated and thus represent  
36 independent outbreaks of resistance. Our data demonstrate that the molecular basis  
37 of pyrethroid resistance in African malaria vectors is more complex than previously  
38 appreciated, and provide a foundation for the development of new genetic tools to  
39 inform insecticide resistance management and track the further spread of resistance.

## 40 **Introduction**

41 An estimated 663 million cases of malaria were averted in Africa between 2000 and 2015  
42 due to public health interventions, of which 68% were prevented by insecticide-treated bed-  
43 nets (ITNs) and @@N% through indoor residual spraying of insecticides (IRS). However,  
44 over this same period, insecticide resistance has become increasingly prevalent in malaria  
45 vector populations. Four chemical classes of insecticides – organophosphates, carbamates,  
46 pyrethroids and organochlorines – are licensed for use in public health, but only pyrethroids  
47 are approved by the World Health Organisation (WHO) for use in ITNs. Pyrethroids  
48 are also commonly used for IRS and in agriculture, and mosquito populations are under  
49 pressure to evolve molecular mechanisms of pyrethroid resistance. There is evidence that  
50 pyrethroid resistance has a direct impact on the effectiveness of ITNs and IRS, although  
51 assessing the impact on disease prevalence is difficult and has been hampered by the fact

52 that pyrethroid resistance is now so pervasive that it is nearly impossible to find fully  
53 susceptible mosquito populations to serve as controls. Nevertheless, the position of the  
54 WHO remains that insecticide resistance poses a grave threat to the future of malaria  
55 control in Africa (@@REF GPIRM). Improvements are needed in our ability to monitor  
56 resistance, and gaps must be filled in our knowledge of the molecular mechanisms of  
57 resistance.

58 The voltage-gated sodium channel (VGSC) is the physiological target of pyrethroids  
59 and of the organochlorine DDT. The VGSC protein is integral to the insect nervous sys-  
60 tem, involved in the transmission of nerve impulses. Both pyrethroids and DDT have a  
61 similar mode of action, binding to sites within the protein channel and preventing nor-  
62 mal nerve function, causing paralysis (“knock-down”) and then death. However, amino  
63 acid substitutions at key positions within the channel can alter the interaction between  
64 the channel and the insecticide molecule, and thereby substantially increase the dosage of  
65 insecticide required for knock-down. If this tolerance exceeds the dosage present in ITNs  
66 or on indoor surfaces following IRS, these interventions may be rendered ineffective. In  
67 the African malaria vectors *Anopheles gambiae* and *Anopheles coluzzii*, three substitutions  
68 have been found in natural populations and shown to cause pyrethroid and DDT resist-  
69 ance. Two of these substitutions occur in codon 995<sup>1</sup>, with the Leucine → Phenylalanine  
70 (L995F) substitution prevalent in West and Central Africa, and the Leucine → Serine  
71 (L995S) substitution found in Central and East Africa. A third variant N1570Y has been  
72 found in association with L995F in Central Africa and shown to increase resistance above  
73 L995F alone.

74 Target-site resistance to pyrethroids and DDT has also been studied in a range of other  
75 insect species, including disease vectors as well as domestic and crop pests. Because of  
76 its essential function, the VGSC protein is highly conserved across insect species, and  
77 knowledge gained from one species is relevant to another. Many resistance-associated  
78 variants have been described in these other species, and thus there are many possible  
79 amino acid substitutions that could induce a resistance phenotype in malaria vectors,  
80 other than the known variants in codons 995 and 1570. Some of these variants are within

---

<sup>1</sup>Codon numbering is given here relative to transcript @@TODO as defined in the AgamP4.@@N gene annotations. A mapping of codon numbers from @@TRANSCRIPT to *Musca domestica* @@TRAN-  
SCRIPT is given in Table ??.

81 the trans-membrane channel, and thus may directly interact with insecticide molecules.  
82 However, functional studies have also demonstrated that variants within internal linker  
83 domains can substantially enhance the level of resistance, when present in combination  
84 with channel modifications. Most previous studies of *An. gambiae* and/or *An. coluzzii*  
85 have performed targeted sequencing of small regions within the gene, and there has been  
86 no comprehensive survey of variation across the entire gene in multiple populations.

87 Insecticide resistance monitoring in malaria vector populations now often incorporates  
88 some form of genetic assay to detect the allele present at *Vgsc* codon 995. Both alleles  
89 are present at high frequency in multiple geographical locations, and the L995F allele  
90 is present in both *An. gambiae* and *An. coluzzii*. The extent of mosquito migration  
91 remains an open question, however mosquitoes do travel between different locations and  
92 have the potential to spread resistance alleles from one population to another (adaptive  
93 gene flow). Hybridization between mosquito species also occurs and has the potential  
94 to transfer resistance alleles between species (adaptive introgression). Studies in West  
95 African have shown that the L995F allele has been transferred from *An. gambiae* into  
96 *An. coluzzii* populations. A resistance allele may also arise independently in multiple  
97 populations, either because of multiple mutational events occurring after insecticides are  
98 introduced (selection on new mutations), or because resistance alleles were already present  
99 at low frequency in mosquito populations prior to insecticide use (selection on standing  
100 variation). Previous studies have found evidence that the L995F allele occurs on several  
101 different genetic backgrounds, suggesting multiple origins of resistance. However, these  
102 studies have used information from only a small region of the gene, and have limited  
103 resolution to make inferences about geographical origins or history of spread. Better  
104 information about the origins and spread of resistance could improve insecticide resistance  
105 monitoring and inform strategies for insecticide resistance management.

106 Here we provide a detailed and comprehensive account of genetic variation within the  
107 *Vgsc* gene using data from phase 1 of the *Anopheles gambiae* 1000 Genomes Project  
108 (Ag1000G). We use genotype and haplotype data derived from whole-genome Illumina  
109 sequencing of 765 individual mosquitoes collected from natural populations in 8 African  
110 countries to survey genetic diversity and study the evolutionary and demographic history  
111 of insecticide resistance at the *Vgsc* locus. Our results reveal an unexpected diversity

112 of molecular mechanisms of resistance, and shed new light on the evolutionary processes  
113 underlying the rapid increase in the prevalence of resistance across multiple mosquito  
114 populations.

115 **Results**

116 **Functional variation (identification of resistance alleles)**

117 To identify single nucleotide polymorphisms (SNPs) with a potentially functional role in  
118 pyrethroid resistance, we extracted SNPs from the Ag1000G phase 1 data resource that  
119 alter the amino acid sequence of the VGSC protein, and computed their allele frequencies  
120 among 9 populations defined by species and country of origin. SNPs that confer resistance  
121 are expected to increase in frequency under selective pressure, and we refined the list of  
122 potentially functional SNPs to retain only those at an appreciable frequency (>5%) in one  
123 or more populations (Table ??). The resulting list comprises 20 SNPs, including the known  
124 L995F, L995S and N1570Y variants, and a further 17 SNPs not previously described in these  
125 species. We reported 15 of these novel SNPs in our initial analysis of the Ag1000G phase  
126 1 data (@@REF Ag1000G), and we extend the analyses here to incorporate two tri-allelic  
127 SNPs affecting codons 402 and 410.

128 The two alleles in codon 995 are clearly the main drivers of resistance at this locus.  
129 The L995F allele at high frequency in populations of both species from West, Central and  
130 Southern Africa, and the L995S allele at high frequency among *An. gambiae* populations  
131 from Central and East Africa (Table ??; @@REF Ag1000G). All haplotypes carrying  
132 L995F or L995S have evidence for strong recent positive selection (@@REF Ag1000G).  
133 Both alleles were present in populations sampled from Cameroon and Gabon, including  
134 some individuals with a hybrid L995F/S genotype. Within these populations, the L995F  
135 and L995S alleles were (@@TODO were not?) in Hardy-Weinberg equilibrium (P=@@),  
136 thus there does not (@@does?) appear to be selection against hybrids.

137 The I1527T allele is present in *An. coluzzii* from Burkina Faso at 14% frequency,  
138 and there is evidence that haplotypes carrying this allele have been positively selected  
139 (@@REF Ag1000G). Codon 1527 occurs within trans-membrane domain segment III.S6,  
140 immediately adjacent to a second predicted binding pocket for pyrethroid molecules, thus

**Table 1.** Non-synonymous nucleotide variation in the voltage-gated sodium channel gene. AO=Angola; BF=Burkina Faso; GN=Guinea; CM=Cameroon; GA=Gabon; UG=Uganda; KE=Kenya; GW=Guinea-Bissau; *Ac*=*An. coluzzii*; *Ag*=*An. gambiae*. All variants are at 5% frequency or above in one or more of the 9 Ag1000G phase 1 populations, with the exception of 2,400,071 G>T which is only found in the CMAg population at 0.4% frequency but is included because another mutation (2,400,071 G>A) is found at the same position causing the same amino acid substitution (M490I).

Variant			Population allele frequency (%)									Function	
Position <sup>1</sup>	<i>Ag</i> <sup>2</sup>	<i>Md</i> <sup>3</sup>	AO <i>Ac</i>	BF <i>Ac</i>	GN <i>Ag</i>	BF <i>Ag</i>	CMA <i>g</i>	GAA <i>g</i>	UGA <i>g</i>	KE	GW	Domain <sup>4</sup>	Resistance phenotype <sup>5</sup>
2,390,177 G>A	R254K	R261	0	0	0	0	32	21	0	0	0	IN (I.S4-I.S5)	L995F enhancer (predicted)
2,391,228 G>C	V402L	V410	0	7	0	0	0	0	0	0	0	TM (I.S6)	I1527T enhancer (predicted)
2,391,228 G>T	V402L	V410	0	7	0	0	0	0	0	0	0	TM (I.S6)	I1527T enhancer (predicted)
2,399,997 G>C	D466H	-	0	0	0	0	7	0	0	0	0	IN (I.S6-II.S1)	L995F enhancer (predicted)
2,400,071 G>A	M490I	M508	0	0	0	0	0	0	0	18	0	IN (I.S6-II.S1)	none (predicted)
2,400,071 G>T	M490I	M508	0	0	0	0	0	0	0	0	0	IN (I.S6-II.S1)	none (predicted)
2,416,980 C>T	T791M	T810	0	1	13	14	0	0	0	0	0	TM (II.S1)	L995F enhancer (predicted)
2,422,651 T>C	L995S	L1014	0	0	0	0	15	64	100	76	0	TM (II.S6)	driver
2,422,652 A>T	L995F	L1014	86	85	100	100	53	36	0	0	0	TM (II.S6)	driver
2,424,384 C>T	A1125V	K1133	9	0	0	0	0	0	0	0	0	IN (II.S6-III.S1)	none (predicted)
2,425,077 G>A	V1254I	I1262	0	0	0	0	0	0	0	0	5	IN (II.S6-III.S1)	none (predicted)
2,429,617 T>C	I1527T	I1532	0	14	0	0	0	0	0	0	0	TM (III.S6)	driver (predicted)
2,429,745 A>T*	N1570Y	N1575	0	26	10	22	6	0	0	0	0	IN (III.S6-IV.S1)	L995F enhancer
2,429,897 A>G	E1597G	E1602	0	0	6	4	0	0	0	0	0	IN (III.S6-IV.S1)	L995F enhancer (predicted)
2,429,915 A>C	K1603T	K1608	0	5	0	0	0	0	0	0	0	TM (IV.S1)	L995F enhancer (predicted)
2,430,424 G>T	A1746S	A1751	0	0	11	13	0	0	0	0	0	TM (IV.S5)	L995F enhancer (predicted)
2,430,817 G>A	V1853I	V1858	0	0	8	5	0	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)
2,430,863 T>C	I1868T	I1873	0	0	18	25	0	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)
2,430,880 C>T	P1874S	P1879	0	21	0	0	0	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)
2,430,881 C>T	P1874L	P1879	0	7	45	26	0	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)
2,431,061 C>T	A1934V	A1939	0	12	0	0	0	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)
2,431,079 T>C	I1940T	I1945	0	4	0	0	7	0	0	0	0	IN (IV.S6-)	L995F enhancer (predicted)

<sup>1</sup> Position relative to the AgamP3 reference sequence, chromosome arm 2L. Variants marked with an asterisk (\*) failed conservative variant filters applied genome-wide in the Ag1000G phase 1 AR3 callset, but appeared sound on manual inspection of read alignments.

<sup>2</sup> Codon numbering according to *Anopheles gambiae* transcript AGAP004707-RA in geneset AgamP4.4.

<sup>3</sup> Codon numbering according to *Musca domestica* EMBL accession X966668 [1].

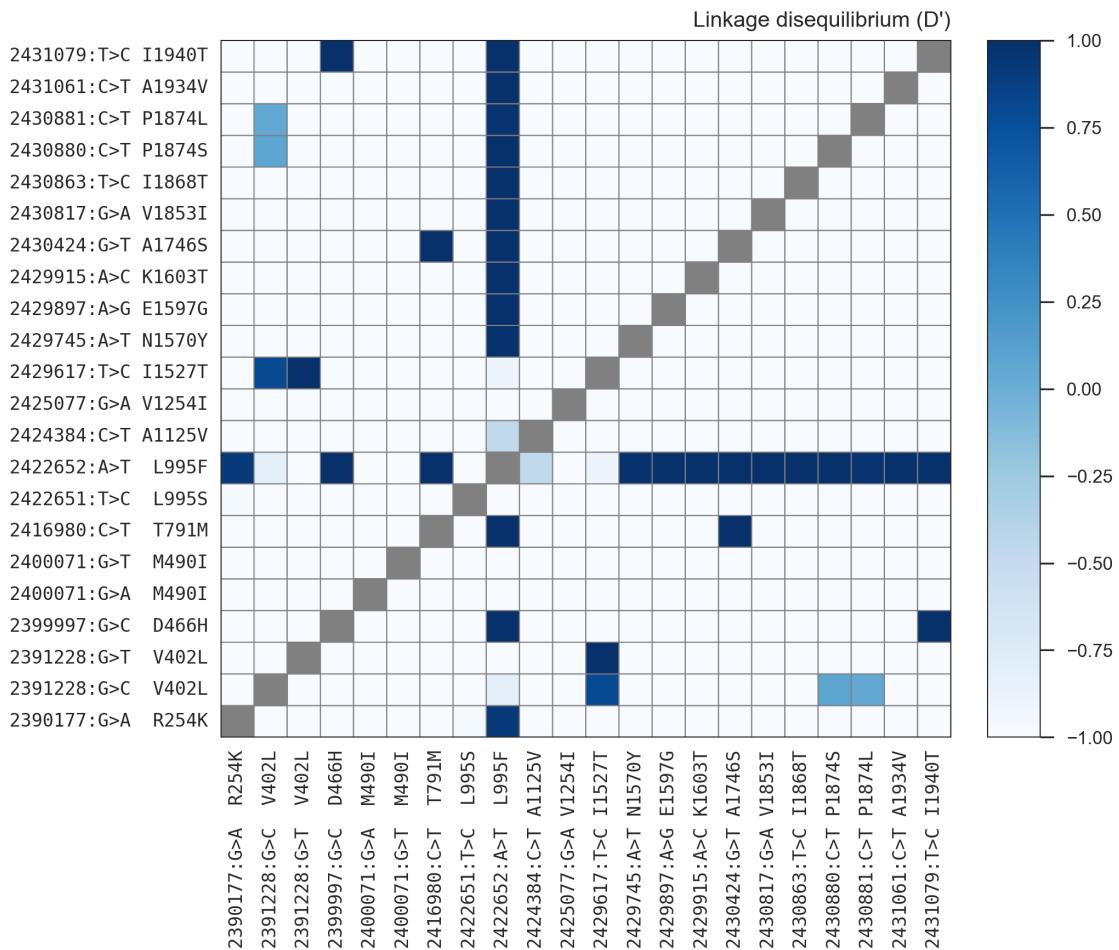
<sup>4</sup> Position of the variant within the protein. IN=internal domain; TM=trans-membrane domain. The protein contains four homologous repeats (I-IV), each having six transmembrane segments (1-6). Codes in parentheses identify the specific domain, e.g., “I.S4” refers to trans-membrane segment 4 in repeat I, and “IS4-IS5” refers to the linker segment between I.S4 and I.S5.

<sup>5</sup> Phenotype predictions are based on population genetic evidence and have not been confirmed experimentally.

141 it is plausible that I1527T could alter insecticide binding (@@REF Dong). We also found  
142 that the two variant alleles affecting codon 402, both of which induce a V402L substi-  
143 tution, were in strong linkage with I1527T (D'>@@N; Figure @@LD), and almost all  
144 haplotypes carrying I1527T also carried a V402L substitution. The most parsimonious  
145 explanation for this pattern of linkage is that the I1527T mutation occurred first, and  
146 mutations in codon 402 subsequently arose on this genetic background. Codon 402 also  
147 occurs within a trans-membrane segment (I.S6), and the V402L substitution has by itself  
148 been shown experimentally to increase pyrethroid resistance in @@species and *Xenopus*  
149 oocytes (@@REFs). However, because V402L appears secondary to I1527T in our cohort,  
150 we classify I1527T as a putative resistance driver and V402L as a putative enhancer. Be-  
151 cause of the limited geographical distribution of these alleles, we hypothesize that the  
152 I1527T+V402L combination represents a pyrethroid resistance allele that arose in West  
153 African *An. coluzzii* populations; however, the L995F allele is at higher frequency (85%)  
154 in our Burkina Faso *An. coluzzii* population, and is known to be increasing in frequency  
155 (@@REFs), therefore L995F may provide a stronger resistance phenotype and is replacing  
156 I1527T+V402L in these populations.

157 Of the other 16 SNPs, 13 occurred almost exclusively in combination with L995F (Figure  
158 @@; @@REF Ag1000G). These include the N1570Y allele, known to enhance pyrethroid  
159 resistance in *An. gambiae* in combination with L995F. These also include two variants  
160 in codon 1874 (P1874S, P1874L). P1874S has previously been found in a colony of the  
161 crop pest *Plutoblah blahdiblah* with a pyrethroid resistance phenotype, but has not been  
162 shown to confer resistance experimentally. 10 of these variants, including N1570Y and  
163 P1874S/L, occur within internal linker domains of the protein, and so fit the model of  
164 variants that may enhance or compensate for the driver phenotype by modifying channel  
165 gating behaviour (@@CHECK; @@REFs). The remaining 3 variants are within trans-  
166 membrane domains, and so may enhance resistance by @@TODO how. Because of the  
167 tight linkage between these 13 SNPs and the L995F allele, we classify all as putative L995F  
168 enhancers, although experimental work is required to confirm a resistance phenotype.

169 The remaining 3 variants (M490I, A1125V, V1254I) do not occur in combination with any  
170 known resistance allele, and do not appear to be associated with haplotypes under selection  
171 (@@REF Ag1000G). A possible exception is the M490I allele found at 18% frequency in



**Figure 1. Linkage disequilibrium between non-synonymous variants.** A value of 1 indicates that the two variants always occur in combination, and conversely a value of -1 indicates that the two variants never occur in combination. @TODO nuance this?

the Kenyan population, although the fact that this population has experienced a recent population crash makes it difficult to test for evidence of selection at this locus. All 3 variants occur in internal linker domains, and so do not fit the model of a resistance driver, although experimental work is required to rule out a resistance phenotype.

### Origins and spread of resistance alleles

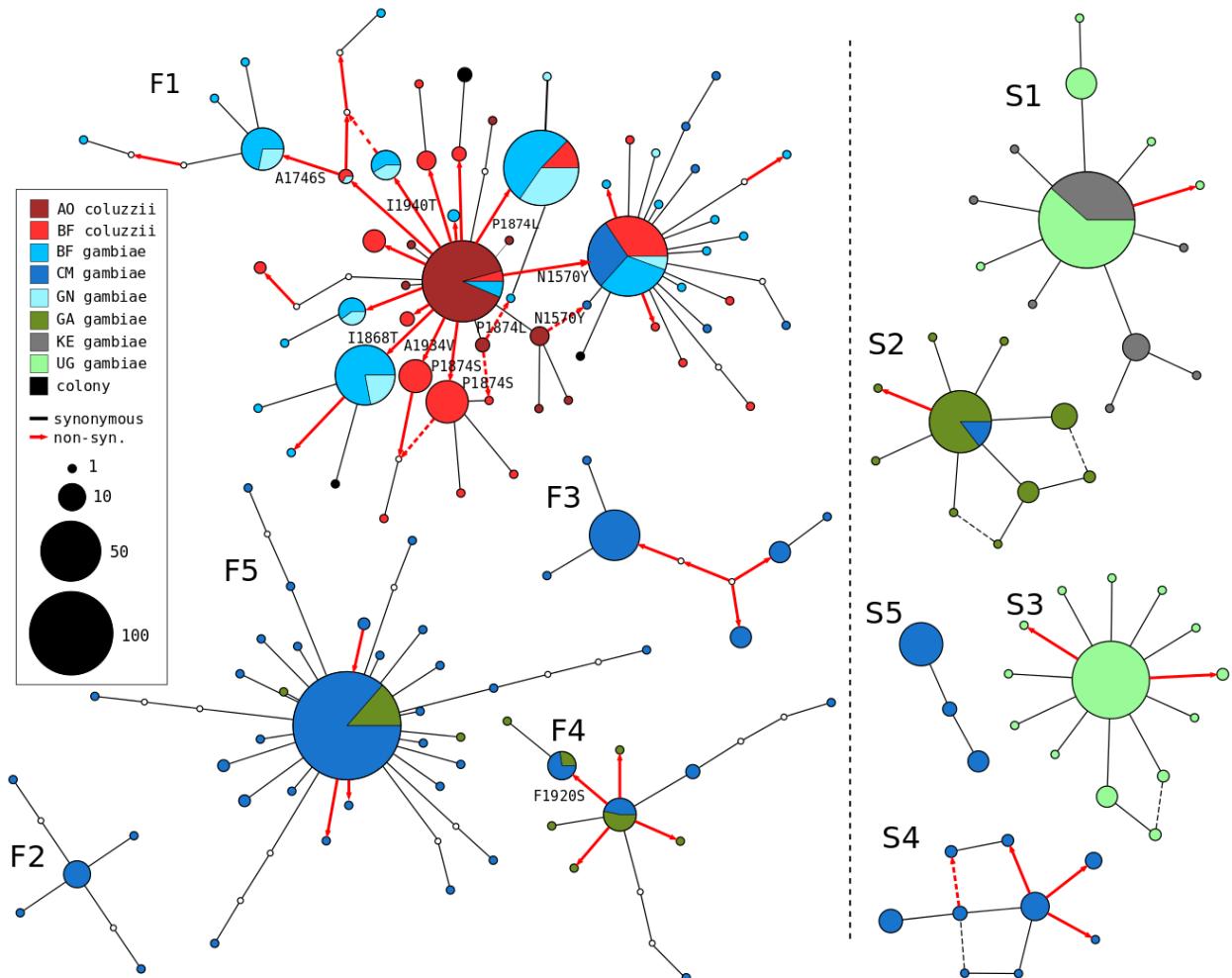
Although it is well-known that pyrethroid resistance is becoming increasingly prevalent in malaria vector populations across Africa, it has not previously been clear whether this increase in prevalence is being driven primarily by the spread of resistance alleles from one location to another via mosquito migration, or by resistance alleles emerging independently and simultaneously in multiple locations, or by a combination of both processes. In

182 our initial analyses of haplotype data from Ag1000G phase 1 (@@REF Ag1000G), we used  
183 a clustering approach based on genetic distance to identify 10 haplotype clusters at the  
184 *Vgsc* locus carrying a known resistance driver allele, of which five clusters carried L995F  
185 (labelled F1-F5) and a further five clusters carried L995S (labelled S1-S5). Within each  
186 cluster, haplotypes were nearly identical across the entire @@70 kbp span of the *Vgsc* gene,  
187 and therefore represent a collection of haplotypes with a very recent common ancestor.  
188 Within some of these clusters, we found haplotypes from mosquitoes collected from differ-  
189 ent geographical locations, demonstrating that adaptive gene flow has occurred between  
190 these locations. Specifically, cluster F1 contained haplotypes from Guinea, Burkina Faso,  
191 Cameroon and Angola; clusters @@ each contained haplotypes from both Cameroon and  
192 Gabon; and cluster @@ contained haplotypes from both Uganda and Kenya. The F1  
193 cluster also contained haplotypes from both *An. gambiae* and *An. coluzzii*, demonstrating  
194 that adaptive introgression had occurred. We also used analyses of haplotype clustering  
195 on the flanks of the *Vgsc* gene to provide evidence that each of these haplotype clusters  
196 represents an independent outbreak of resistance, with the exception of clusters S4 and  
197 S5 which appear to be recently derived from the same ancestor. In this section we present  
198 several new analyses to confirm and extend our initial findings regarding the origins and  
199 spread of *Vgsc* resistance alleles.

200 To provide an alternative view of the genetic similarity between haplotypes carry-  
201 ing known or predicted resistance driver alleles, we used haplotype data across all 1710  
202 (@@CHECK) SNPs within the *Vgsc* open reading frame to construct median-joining net-  
203 works (Figure 2). We constructed these networks up to a maximum distance of @@4 SNP  
204 differences, to ensure that each connected component in the resulting networks represents  
205 a collection of haplotypes with a recent common ancestor, and thus which is also likely to  
206 be minimally affected by recombination within the gene. For L995F, the resulting network  
207 confirms the presence of five distinct clusters of closely-related haplotypes, with close cor-  
208 respondance to the clusters F1-F5 identified previously (@@REF Ag1000G). The L995S  
209 network also confirms five distinct clusters, in concordance with our previous analysis.

210 The haplotype network for cluster F1 brings into sharp relief the explosive evolution of  
211 secondary amino acid substitutions within the F1 lineage, providing further evidence that  
212 they play a functional role in enhancing the L995F resistance phenotype. The distribution

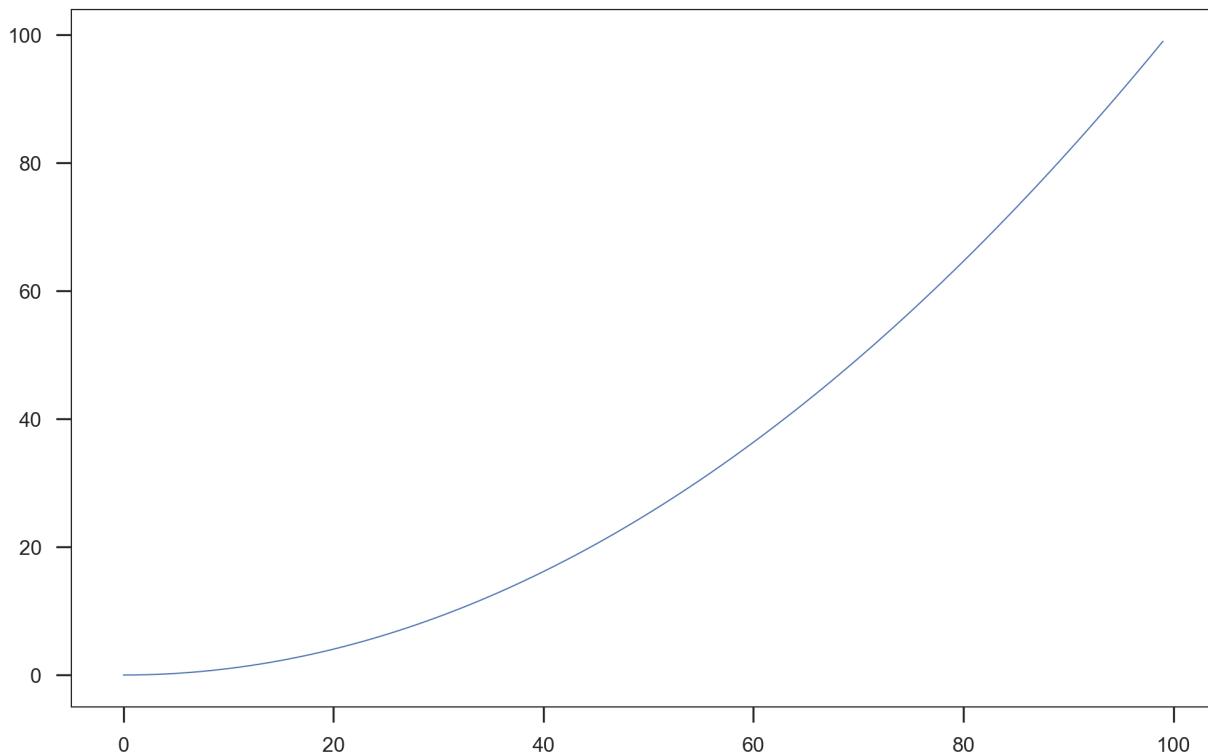
213 of secondary variants within the F1 network also allows us to infer multiple introgression events between the two species. @@TODO details of which variants are found in  
 214 both species. The contrast between the haplotype networks for the L995F and L995S alleles is striking, because of the near-total absence of any non-synonymous variants within  
 215 the L995S networks. As we reported previously, there is a highly significant enrichment  
 216 for non-synonymous variation among haplotypes carrying L995F relative to haplotypes  
 217 with L995S. This difference is the basis for our prediction that all of the secondary non-  
 218 synonymous variants found at appreciable frequency among L995F haplotypes are likely to  
 219 enhance the resistance phenotype, rather than being deleterious or neutral variants that  
 220 are hitch-hiking on selective sweeps for the driver allele, because hitch-hiking should be  
 221



**Figure 2. Haplotype networks.** @@TODO redo the figure. @@TODO annotate non-syn edges in cluster F3. @@TODO mention if any clusters fixed for non-syn variants so not shown. @@TODO annotate other non-syn edges, e.g., in S4?

223 similar for both L995F and L995S.

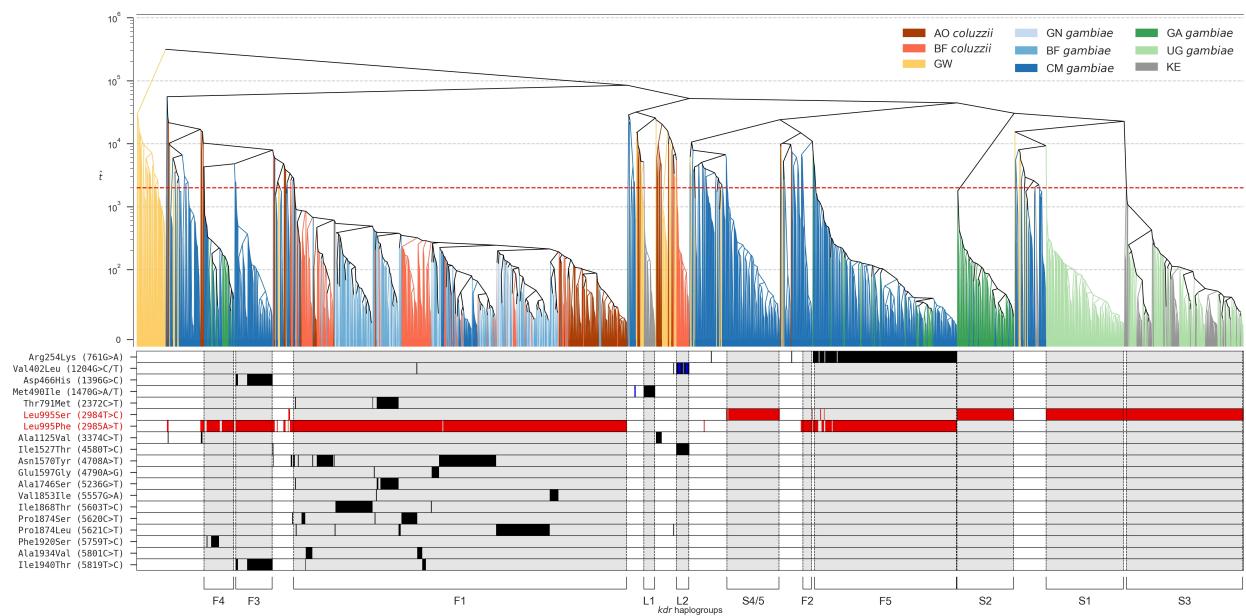
224 A limitation of both the hierarchical clustering and network analyses is that they only  
225 leverage information from within the *Vgsc* gene. Because they are relying on estimates of  
226 genetic distance from within a relatively small genome region, they have limited resolution  
227 to infer very recent events, because insufficient time has elapsed for mutations to occur.  
228 This means that, within each of the lineages we have identified where adaptive gene  
229 flow has occurred, the analyses provide little information about the direction or relative  
230 timing of gene flow events. To improve our resolution to infer recent events, we used  
231 information about the extent of haplotype sharing on both flanks of the gene. For each  
232 pair of haplotypes, we used a heuristic approach to estimate the length of the region  
233 shared identical by descent (IBD) between two haplotypes, extending both upstream and  
234 downstream of the gene, and to estimate the number of mutations that have accumulated  
235 within IBD regions since the most recent common ancestor (MRCA). We then combined  
236 the information about IBD length and number of mutations to estimate the time to MRCA  
237 or "age" for each pair of haplotypes. We then studied the distribution of haplotype ages



**Figure 3. Haplotype age distribution.** @@TODO real figure.

238 (Figure 3), and used hierarchical clustering to visualise the overall age structure (Figure 4).  
 239 We caution that these analyses are relatively crude, and could be improved in a number  
 240 of ways. However, estimating haplotype age and local genealogy is a very active research  
 241 area, and we use a heuristic approach here to provide an initial view into the data. We  
 242 also caution that although the estimated ages are in units of generations, these estimates  
 243 have not been calibrated, and so cannot be taken as accurate absolute values. They can,  
 244 however, be compared with each other, to explore the relative age of different events.

245 A key feature of the overall distribution of haplotype ages is that it is bimodal, with  
 246 a minor mode of haplotypes coalescing recently, and a major mode coalescing further in  
 247 the past (Figure 3). This is expected at a locus experiencing recent positive selection  
 248 and multiple selective sweeps. Within each sweep, all haplotypes share a very recent  
 249 common ancestor, but between sweeps and among haplotypes without any resistance allele,  
 250 genealogies reflect a more even distribution of ancestry. This overall bimodal distribution  
 251 is also not simply a reflection of geographical population structure, because the same  
 252 bimodality is observed within several populations (Figure @@REF). We take the midpoint  
 253 between these two modes as an estimate for the maximum age of selective sweeps at  
 254 this locus. When we then cut the haplotype tree at this age, and take the 11 largest  
 255 clades, we find that the resulting clades are highly concordant with the clustering and



**Figure 4. Clustering of haplotypes by age.** @@TODO bigger font. @@TODO change "kdr haplogroups" to something else. @@TODO yticks to show number of haplotypes.

network analyses based on genetic distance within the gene (Figure 4). Clusters F1-F5 are recapitulated, as are clusters S1-S4, with S4 and S5 merged into a single cluster reflecting their shared ancestry as discovered from previous analyses. We also label a new cluster "L@@" representing the lineage of haplotypes carrying the I1527T allele in combination with one or the other V402L allele. @@TODO what to say about the other "L@@" cluster?

Using these estimates for haplotype age, we draw some tentative conclusions regarding the history of these resistance outbreaks. @@TODO describe evidence for direction of spread within spreading haplogroups.

@@TODO describe how we put together the map (Figure 6).

## Recombination and independent outbreaks of resistance

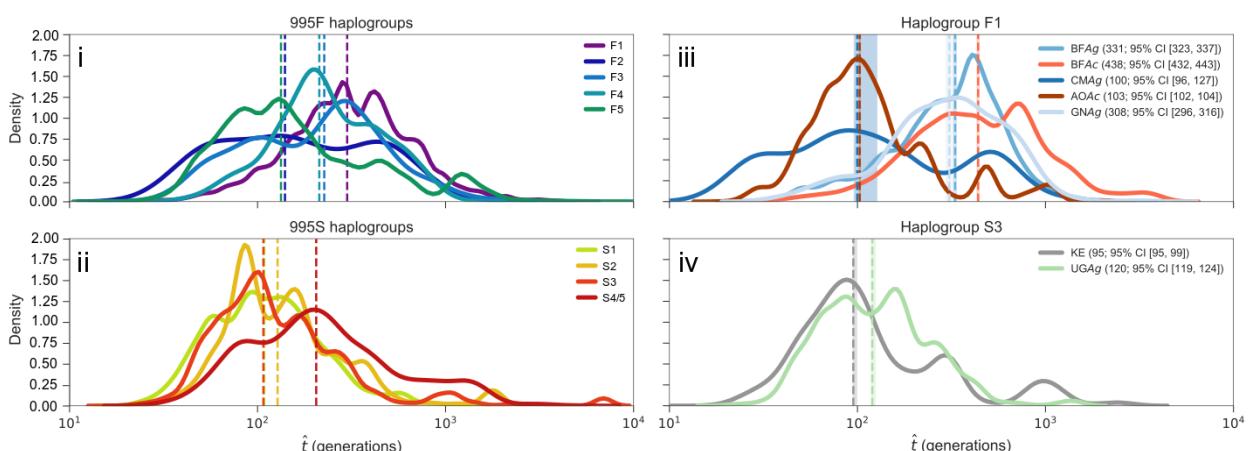
@@TODO

@@TODO

@@TODO

## Discussion

@@TODO



**Figure 5. Haplotype age distributions.** @@TODO rethink what goes in here, also if this needs to be here or can go to supplementary.

271 **Methods**

272 @@TODO

273 **References**

- 274 [1] Martin S Williamson et al. 'Identification of mutations in the houseflypara-type so-  
275 dium channel gene associated with knockdown resistance (kdr) to pyrethroid insect-  
276 icides'. In: *Molecular and General Genetics MGG* 252.1 (1996), pp. 51–60.

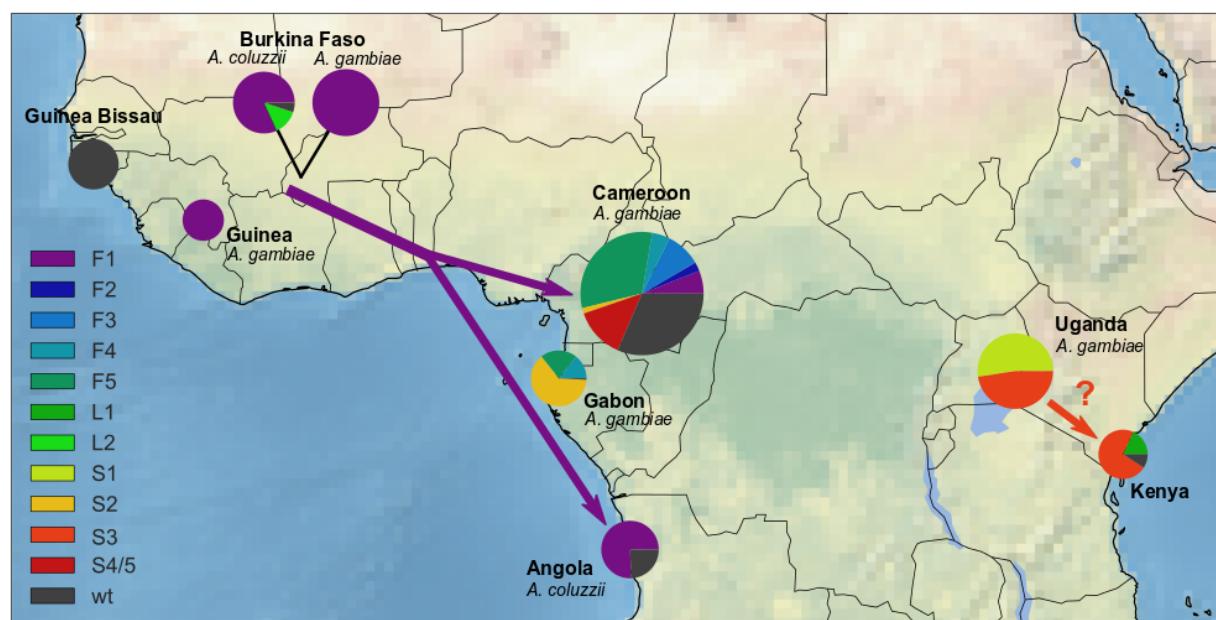
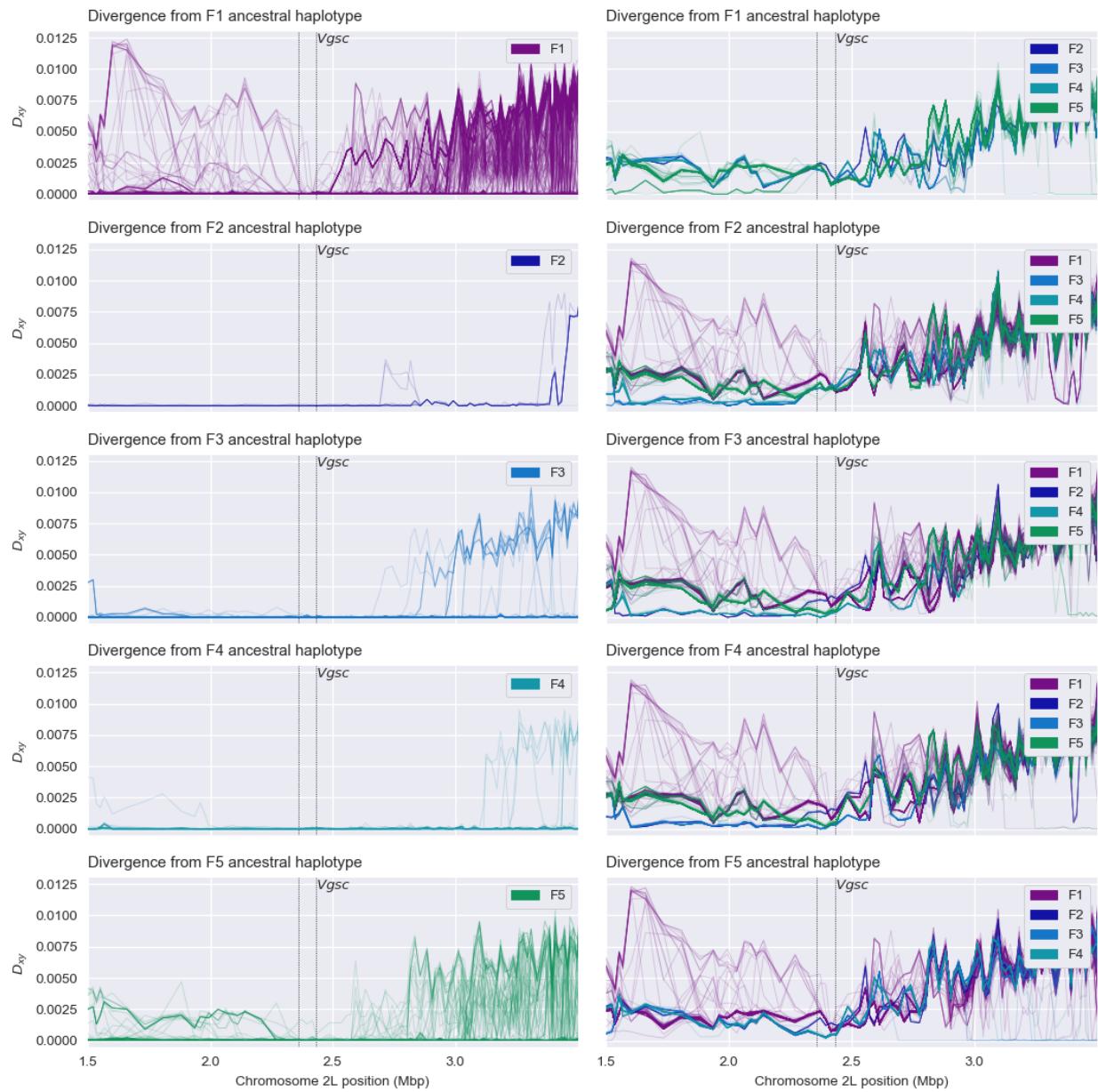
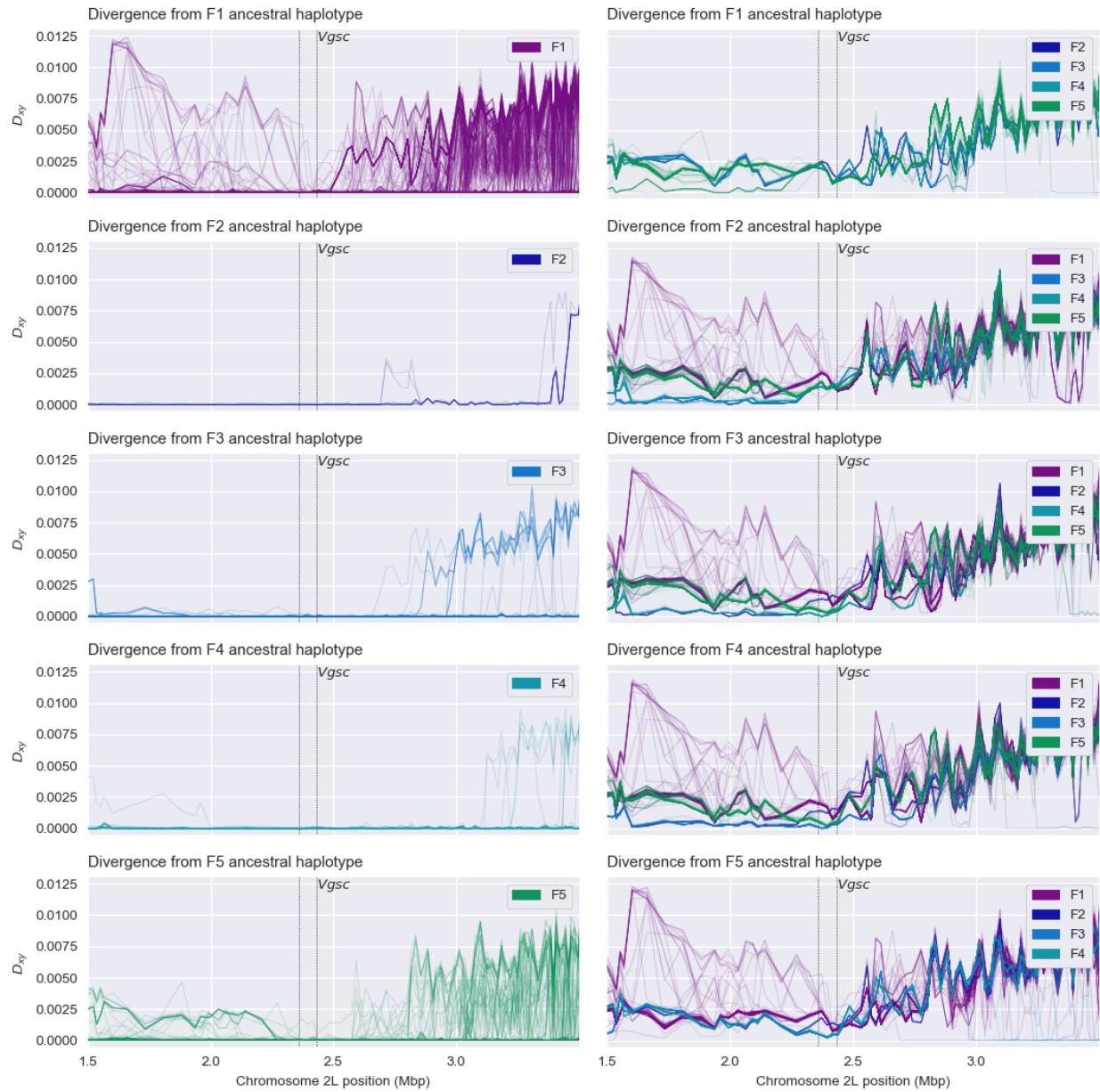


Figure 6. Geographical distribution of resistance haplogroups.



**Figure 7. Recombination and ancestral haplotypes for L995F.** @@TODO legend



**Figure 8. Recombination and ancestral haplotypes for L995S.** @@TODO legend