Whole-genome sequencing of major malaria vectors revealed an evolution of new insecticide resistance variants in a longitudinal study in Burkina Faso

Mahamadi Kientega[1,3,*], Chris S Clarkson[2], Nouhoun Traoré[1,3], Hui Tin-Yu J[4], Samantha O'Loughlin[4], Abdoul-Azize Millogo[1,5], Patric S Epopa[1], Franck A Yao[1], Adrien M G Belem[3], Jon Brenas[2], Alistair Miles[2], Austin Burt[4], Abdoulaye Diabaté[1,*]

[1]Institut de Recherche en Sciences de la Santé (IRSS), Bobo-Dioulasso 01 BP 545, Burkina Faso ;

[2]Vector Surveillance Programme, Genomic Surveillance Unit, Wellcome Sanger Institute, Hinxton, Cambridge, UK ;

[3]Université Nazi Boni, Bobo-Dioulasso 01 BP 1091, Burkina Faso ;

[4]Department of Life Sciences, Imperial College, London SW7 2AZ, UK;

[5]Institut des Sciences des Sociétés, Ouagadougou 03 BP 7047, Burkina Faso.

[*]Corresponding authors: npiediab@gmail.com (A.D.), mkient54@gmail.com (M.K.).

Email addresses : *mkient54@gmail.com (Mahamadi Kientega); cc28@sanger.ac.uk (Chris S Clarkson); nouhoun89@gmail.com (Nouhoun Traoré); tin-yu.hui11@imperial.ac.uk (Hui Tin-Yu J); s.oloughlin@imperial.ac.uk (Samantha O'Loughlin); belemamg@hotmail.fr (Adrien M G Belem); azizemillogo@gmail.com (Abdoul-Azize Millogo); epopastef@yahoo.fr (Patric S Epopa); afrayao@gmail.com (Franck A. Yao); jon.brenas@bdi.ox.ac.uk (Jon Brenas); alistair.miles@sanger.ac.uk (Alistair Miles); a.burt@imperial.ac.uk (Burt Austin); npiediab@gmail.com (Abdoulaye Diabaté).*

Abstract

Intensive deployment of insecticide based malaria vector control tools results in the rapid evolution of phenotypes resistant to these chemicals. Understanding this process on the genomic level is essential for the deployment of successful interventions. Using whole genome sequencing data of 1409 individual *An. gambiae* s.l. collected from 2012 to 2017, we investigated the change in genetic structure and the evolution of the insecticide resistance variants in natural populations over time and space. The results showed similar and constant nucleotide diversity and negative Tajima's D between *An. gambiae* s.s. and *An. coluzzii*. PCA and F_{ST} showed a clear genetic structure in the *An. gambiae* s.l. species. Genome-wide F_{ST} and H12 scans identified genomic regions under divergent selection and also having an implication in the adaptations to ecological changes. Novel voltage-gated sodium channel pyrethroid resistance target-site alleles (*V402L, I1527T*) were identified at increasing frequencies alongside the established *kdr* alleles (*Vgsc-L995F*, *Vgsc-L995S* and *N1570Y*) within the *An. gambiae* s.l. populations. Organophosphate metabolic resistance markers were also identified, at increasing frequencies, within the *An. gambiae* s.s. populations from 2012 to 2017, including the SNP *Ace1-G280S* and its associated duplication. Variants simultaneously identified in the same vector populations raise concerns about the long-term efficacy of new-generation bednets and the recently introduced organophosphate pirimiphos-methyl indoor residual spraying. These findings highlighted the benefit of genomic malaria vector surveillance for the detection of new insecticide resistance variants, the monitoring of the existing resistance variants, and also to get insights into the evolutionary processes driving insecticide resistance.

Author Summary

Genomic surveillance of malaria vectors is crucial for understanding the genetic variation in natural vector populations and also guiding the implementation of novel and innovative vector control tools. Application of sequencing technologies in vector studies provide insights on the genetic and evolutionary phenomena of vectors that could have impact on vector control strategies. By analyzing the whole genome data of 1409 wild *An. gambiae* s.l. mosquito collected between 2012 and 2017, we showed an emergence of novel insecticide resistance markers alongside increasing frequencies of existing insecticide resistance variants over time in Burkina Faso. We showed the benefit of genomic surveillance of malaria vectors for the monitoring of the insecticide resistance variants and also providing insights into the evolutionary processes driving insecticide resistance.

Introduction

Malaria remains a major public health problem in Africa [1]. *Anopheles gambiae* s.l., the main malaria vector in Africa, is phenotypically and genotypically diverse, allowing rapid adaptation to environmental changes, such as introductions of new insecticidal control methods [2]. In Burkina Faso, major malaria vectors *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* can be found living in sympatry, with diverging resting and feeding behaviours [3,4]. These species, through their ability to colonize different ecological settings, their preference for human blood feeding, and their high susceptibility to a malaria parasite (*Plasmodium falciparum*) infection, are responsible for around 90% of the malaria burden in Burkina Faso and the whole sub-Saharan Africa [5,6].

The National Malaria Control Programme (NMCP) of Burkina Faso and their partners have invested much to increase the coverage of long-lasting insecticide treated bednets (LLINs) and other control tools such as indoor residual spraying (IRS) of insecticides to drop the malaria transmission line in the country. The coverage of LLINs in households increased significantly in the country over the years (5.6% in 2003 to about 80% in 2014) reaching 83% in 2021 [7,8]. The recent 2019 and 2022 bednet campaigns involved the free distribution of three types of LLINs, especially the piperonyl butoxide (PBO)-synergist, the dual-AI Interceptor G2 and the standard LLINs [9]. Despite these achievements, malaria remains a countrywide health problem responsible for about 12.2 million cases and 4355 deaths during the year 2021 [10].

In Burkina Faso and most sub-Saharan countries, the primary vector surveillance strategies relied on the routine tracking of the vector species density and distribution and the monitoring of insecticide resistance (IR) status, unable to properly explain the genetic and evolutionary processes driving the increasing spread of IR variants and

vector behavioral changes [11]. Recent advances in genomic sequencing technologies, both price and throughput, and new cloud native analysis tools have made genomic surveillance of malaria vectors possible. Genomic surveillance has advantages to provide insights on the genetic and evolutionary phenomena of vectors that could have any impact on vector control strategies. Towards this aim, the *Anopheles gambiae* 1000 genomes (Ag1000G) project driven by the MalariaGEN Vector Observatory sequenced more than 15000 *Anopheles* mosquitoes collected in 25 African countries [12,13]. A major goal of this project was to support malaria elimination efforts by providing a high quality open access data resource on natural genetic variation within *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* populations, both for the public health and research communities [14].

Previous studies investigating insecticide resistance (IR) using these genomic data have highlighted associated genetic variants distributed throughout the mosquito genome [15–18]. The phenotypic effect of many of these variants has been demonstrated through in-vivo and in-vitro experiments and their effects are evident in low mortality rate observed during the IR bioassays, and in the reduction of the efficacy of the insecticide-based control tools [19–21]. The strength of these resistance phenotypes and the wide geographical distribution of these variants raise concerns about the durability of the new insecticide-based vector control tools being used to overcome resistance, including indoor residual spraying (IRS) with organophosphate insecticides such as Actellic (pirimiphos-methyl) and pyrethroid bed-nets treated with metabolic resistance blocking piperonyl butoxide (PBO) synergists.

There is an increasingly urgent need for sustainable strategies and effective tools to support malaria elimination efforts in the face of an ever evolving vector. Research of alternative and innovative tools such as genetic engineering, new insecticidal

compounds as well as natural symbionts has increased in recent years, giving a glimmer of hope to people living in endemic areas [22,23]. However, given the genome complexity of the malaria vectors, highlighted by many studies [12,13,24], any implementation of vector control tools requires an understanding of the genomic variation of vector species, particularly an understanding of the evolutionary processes, emergence and the spread of the variants associated with IR in natural populations over time. Thus, longitudinal sampling of vectors, followed by whole genome population genomics becomes an essential tool upstream and downstream of the implementation of the control tools to provide insights on the genetic and evolutionary phenomena of vectors that could have any impact on vector control strategies. Here we leverage the 1409 *Anopheles gambiae* s.l. samples, collected in Burkina Faso between 2012-2017 and sequenced as part of the *Anopheles gambiae* 1000 Genomes Project (2017, 2020), to characterize the spread of known IR variants over space and time, while analyzing these natural populations for evidence of the evolution of novel IR mechanisms which could threaten current or future vector control methods.

Results

Population Sampling

Following the *Anopheles gambiae* 1000 Genomes project, further mosquito collections were sequenced as part of the follow-up project, the MalariaGEN Vector Observatory. In total, this resulted in 1409, deep, whole-genome sequenced, mosquitoes collected from three villages surrounding Bobo-Dioulasso city in western Burkina Faso. Mosquitoes were collected longitudinally, across the years 2012, 2014, 2015, 2016 and 2017 (Fig 1), though species representation varies between collection sites and time-points (Table 1). *An. coluz*zii was the most predominant species found, accounting for 52.73% [743/1409] of samples, followed by *An. gambiae s.s*. with 38.96% (549/1409), while *An. arabiensis* only comprised 8.16% (115/1409) of samples collected. Two *An. gambiae* s.s. and *An. coluzzii* hybrids were also identified. As shown by previous work, *An. coluzzii* remains the most predominant species in Bana (83.87% [489/583]) and Souroukoudinga (60.90% [243/399]) while *An. gambiae* s.s. is prevalent in Pala (70.72% [302/427]) [6].

Fig 1. Map of the sampling sites around Bobo-Dioulasso city

Table 1. Species composition of *An. gambiae* complex in the three villages surrounding Bobo-Dioulasso city over the five collection years

Species	*Periods*	*Sampling sites*		
		Bana	*Souroukoudinga*	*Pala*
An. gambiae s.s.	*2012*	*22 [33.85%]*	*28 [49.12%]*	*48 [81.36%]*
	2014	*32 [24.81%]*	*49 [49.49%]*	*89 [89.%]*
	2015	*25 [20.66%]*	*40 [31.5%]*	*76 [46.34%]*
	2016	*13 [9.77%]*	*38 [44.19%]*	*53 [79.1%]*
	2017	*0*	*0*	*36 [97.3%]*
An. coluzzii	*2012*	*42 [64.62%]*	*29 [50.88%]*	*11 [18.64%]*
	2014	*96 [74.42%]*	*50 [50.51%]*	*0*
	2015	*96 [79.34%]*	*87 [68.5%]*	*0*
	2016	*120 [90.23%]*	*48 [55.81%]*	*0*
	2017	*135 [100.%]*	*29 [96.67%]*	*0*
An. arabiensis	*2014*	*1 [0.78%]*	*0*	*11 [11.%]*
	2015	*0*	*0*	*88 [53.66%]*
	2016	*0*	*0*	*14 [20.9%]*
	2017	*0*	*1 [3.33%]*	*0*
intermediate_gambiae_coluzzii	*2012*	*1 [1.54%]*	*0*	*0*
	2017	*0*	*0*	*1 [2.7%]*

Genetic Diversity

Understanding genetic diversity allows us to compare demography of species over time and space. Analyses of these data showed around 27.94 million of segregating sites (~ 91% of biallelic SNPs) in the *An. gambiae* s.l. genome. Approximately 25.68 million of SNPs were identified in the autosomes and 2.262 million of SNPs in the X chromosome. Analysis to evaluate the genetic diversity of this variation (see Methods), demonstrated that nucleotide diversity (θ_π) and Watterson's theta (θ_w), were high and stable over all collection years for *An. gambiae* and *An. coluzzii*, similar to the majority of natural populations across the species range (Ag1000g, 2017) (Fig 2), potentially suggesting similar demographic histories between these two species. In contrast, *An. arabiensis* showed lower diversity, though with apparent increase over collection years. The diversity findings were reflected in a genome-wide negative Tajima's D, caused by an excess of rare variants due to either positive selection and/or rapid demographic changes through population expansion. In addition, the allelic frequency spectrum (Fig S1 & S2) also showed the same trend highlighting the excess of low frequency variants in the *An. gambiae* s.l. genome.

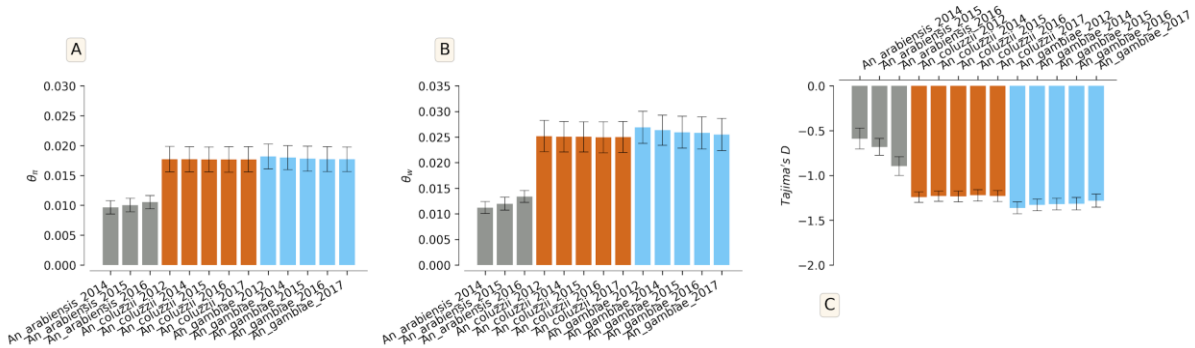

Fig 2. Average of the diversity statistic (A: nucleotide diversity (θ_π), B: Watterson theta (θ_w), C: Tajima's D) in the *An. gambiae* s.l. populations

Population Structure

Understanding population structure between, and within, vector species is not only important to understand due to its implications in gene flow of medically relevant alleles or role out of gene drive control methods, but hidden structure will confound downstream population genomic analyses. Consequently, the genetic structure and the patterns of gene flow were investigated using SNPs on the 3L chromosome (see methods). Pairwise F_{ST} was estimated between species collected in different villages and was found to be low between *An. gambiae* and *An. coluzzii* scaling from 0 to 0.05 (Fig 3A). These results suggest gene flow between *An. coluzzii* and *An. gambiae* s.s.. *An. arabiensis*, however, showed relatively high F_{ST} between both *An. coluzzii* and *An. gambiae*, ranging from 0.26 to 0.30 (Fig 3A), potentially suggesting weaker gene-flow between species. No differentiation (F_{ST} ~ 0) was observed between populations of the same species collected in different villages, suggesting no population structure on this scale.

Principal component analysis (PCA) showed a clear divergence between the *An. gambiae* s.l. species. Examining the first two principal components, samples fell into three clusters, where individuals of the same species were more genetically related than to others. No sub-clusters were observed within each species (Fig 3B). No change was observed in the population structure of each species throughout the years, and the sampling sites (Fig S3); however, some other genomic regions were shown to be under divergent selection (see Organophosphate and carbamate resistance: increasing threats to control plus a potential novel IR mechanism, Fig S4). The lack of genetic structure (low F_{ST} ~ *0.05*) between *An. coluzzii* and *An. gambiae* s.s. showed that these species are capable of sharing advantageous variants in this region Burkina Faso, such those involved in insecticide resistance.

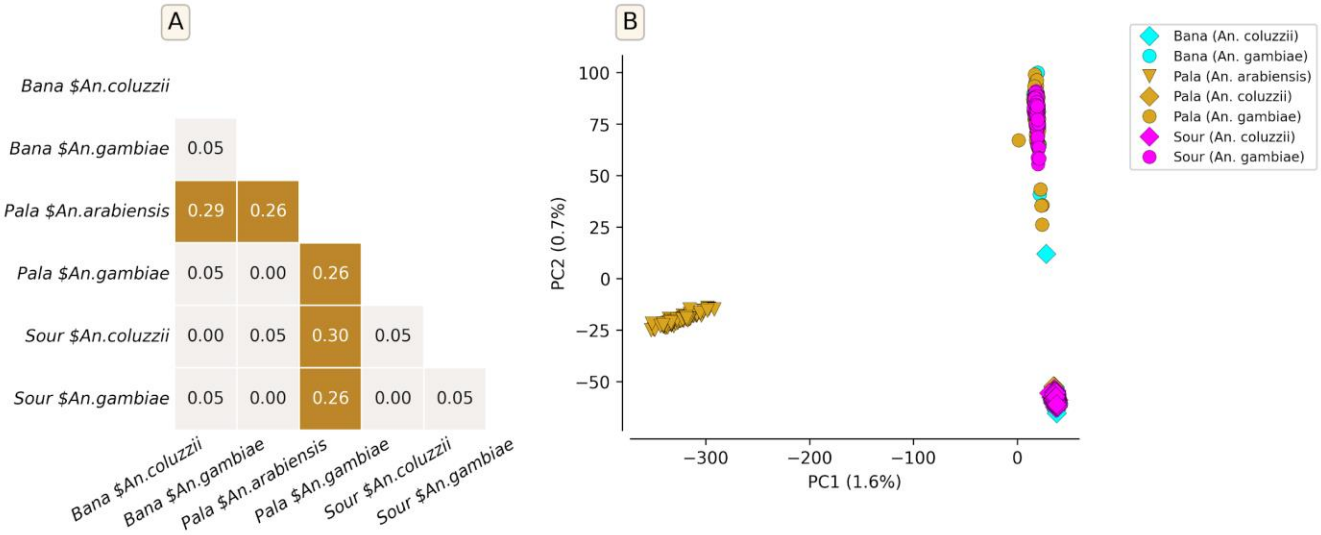

Fig 3. Genetic differentiation and population structure of *An. gambiae* s.l. using the SNPs identified in the 3L chromosome. A: Pairwise F_{ST} between *An. gambiae* s.l. populations collected in different villages. No differentiation between populations of the same species collected in two different villages. B: PCA showing the genetic structure of *An. gambiae* s.l. populations.

Insecticide Resistance

Pyrethroid Target-site Resistance: concerning increase of new mutations

We investigated the distribution of genetic variants in the voltage gated sodium channel gene (*Vgsc)*, which codes for the target-site protein of the pyrethroid insecticides used in bednets). Non-synonymous SNPs in insecticide target site genes, can lead to insecticide resistance via physically altering the conformation of the target-site protein, causing reduced insecticide binding [25].

Pyrethroid target-site resistance in *Anopheles gambiae* s.l. has, historically, been primarily driven by two main "*kdr"* SNPs, causing the amino acid changes *Vgsc-L995F* and *Vgsc-L995S* [17]. These mutations are widespread, and are at high frequencies in many regions in Africa [12]. Improvements in sequencing technologies have allowed

the identification of many additional SNPs in the *Vgsc* gene that could enhance the level of pyrethroid resistance over the time [26,27]. This study identified 634 non-synonymous SNPs in all the *An. gambiae* s.l. populations (Table S1). The *kdr* allele *Vgsc-L995F* was identified at high frequencies (> 30%) in all the populations, and was found at fixation in *An. gambiae* s.s. collections from 2012. While *An. arabiensis* populations also showed high frequencies of *Vgsc-L995S* (Fig 4, Fig S5). The *Vgsc-N1570Y* allele, previously shown to increase the level of pyrethroid resistance when in concert with *Vgsc-L995F* [26], was also present in both *An. gambiae s.s.* and *An. coluzzii* populations at around 30% frequency. Some SNPs were found to be species specific over the years. We identified the alleles *Vgsc-V402L, N1345I, I1527T, K1603T, L1667M, P1874S, A1934V* and *I1940T* specifically in *An. coluzzii* populations whereas *Vgsc-V1853I, I1868T, E1597G* were specific to *An. gambiae* s.s. populations (Fig S5). Most of these species specific alleles were identified at low frequencies.

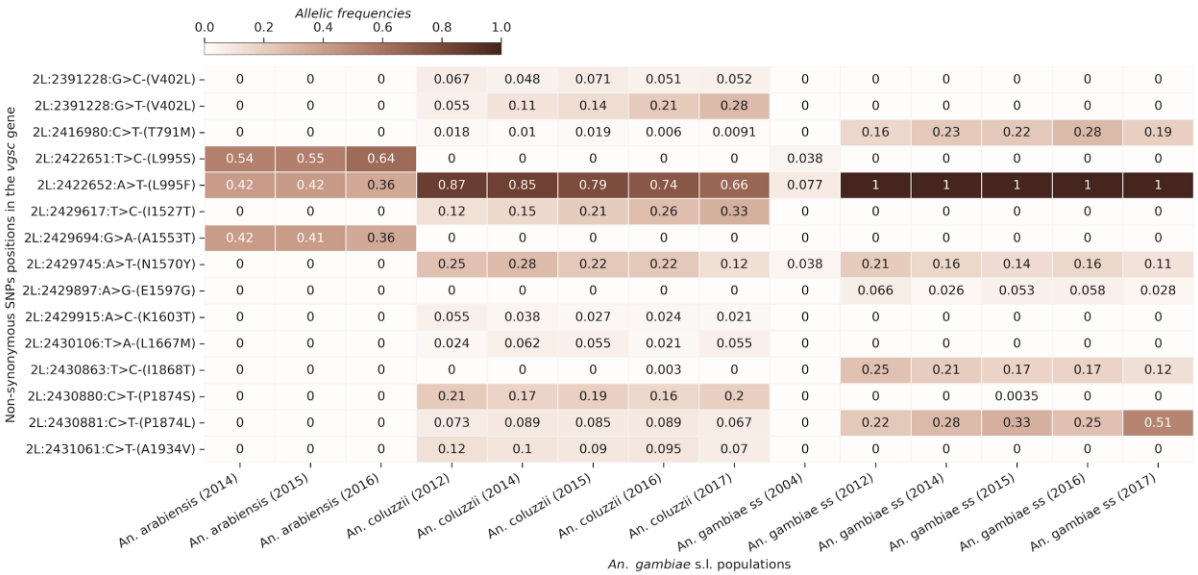

Fig 4. Heat map showing the frequencies of the non-synonymous SNPs frequencies (max freq > 0.05) in the *Vgsc* gene in the *An. gambiae s.l.* populations over time. The X axis shows the *An. gambiae s.l.* populations and the years of collection. The Y axis

shows the non-synonymous SNPs positions in the chromosome 2L. The gradient color bar shows the distribution of the allelic frequencies.

A *Vgsc* haplotype carrying the apparently linked amino acid substitutions *Vgsc-V402L+I1527T* was also identified in Burkina Faso *An. coluzzii* (Fig 4 and 5, Appendix 5). A recent *in-vivo* investigation has shown that the *Vgsc-V402L* substitution results in a pyrethroid resistance phenotype with lower fitness costs than the classic *Vgsc-L995F kdr* mutation [27]. Upon discovery of *Vgsc-V402L+I1527T,* it was suggested that this haplotype may be a “relic” resistance haplotype, pre-dating and being overtaken by a more effective *Vgsc-L995F* carrying haplotype [17]. However, our time series data shows that the inverse appears to be the case. *Vgsc-V402L+I1527T* increased from 12.19% in 2012 to 33.12% in 2017 (Fig 4 and 5, Fig S5). The increase of these alleles coincides with the drop in the frequency of the *Vgsc-L995F* mutation in *An. coluzzii* populations from 86.58% in 2012 to 66.46% in 2017 (Fig 4 and 5). Our results and other works [17,27] suggest these two alleles are increasing in frequency and replacing *Vgsc-L995F* in *An. coluzzii*, but in 2017 had not spread to *An. gambiae* s.s. The increased emergence of new *kdr* alleles alongside the existing *kdr* mutations (*Vgsc-L995F* and *Vgsc-L995S*) at high frequencies may increase the resistance level of vector populations to pyrethroid compounds.

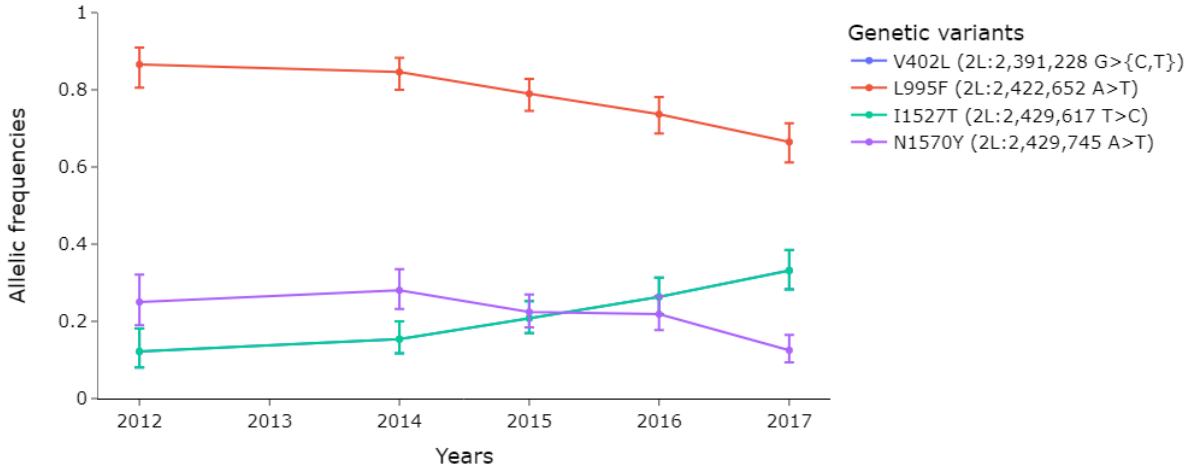

Fig 5. Dynamics of *Vgsc-L995F* and double mutant *Vgsc-V402L+I1527T* allelic frequencies in the *An. coluzzii* populations.

Pyrethroid Metabolic Resistance: increasing CNV frequencies.

Gene duplications and deletions (copy number variants - CNVs) are known to influence the expression level of genes by providing additional copies of the original gene. Consequently, CNVs of genes involved in xenobiotic detoxification can increase insecticide metabolism, producing a resistance phenotype [15]. To help overcome pyrethroid resistance, now commonplace in many natural populations, PBO synergist bednets have recently been introduced. These bednets enhance the pyrethroid efficacy by inhibiting the cytochrome P450 enzymes that can be responsible for metabolic pyrethroid resistance [28]. Often, mechanisms driving the resistance in a given area remain unclear [20], however, whole genome sequencing allows CNV calling, and the detection of resistance associated alleles [16].

CNVs were identified in 136 P450 genes distributed across the whole genome and most of these (~ 63.97 %) were gene amplifications (Table S2). The majority of the CNVs were species-specific, and some were found at increasing frequencies over time

in the three species. Some genes showing low CNVs frequencies in 2012 displayed increasing CNVs frequencies over the time (Fig 6). This situation was observed in the *Cyp6z* cluster of genes (*Cyp6z1, Cyp6z2 and Cyp6z3*), known to be associated with pyrethroid resistance [29], whose CNVs frequencies increased from 17% in 2012 to up to 60% in 2017 in *An. gambiae* s.s. populations. Additionally, the *Cyp9k1* gene, also previously shown to be involved in pyrethroid resistance, showed CNVs at high frequencies up to 80% from 2012 to 2017. In the *An. coluzzii* populations, CNVs were identified at frequencies ranging from 50% in 2012 to up to 90% in 2017 within the *Cyp6aa1, Cyp6aa2, Cyp6p15p* and *Cyp12f2* genes. Moreover, the *Cyp6af* (*Cyp6af1* and *Cyp6af2*) and *Cyp9m* cluster genes (*Cyp9m1 and Cyp9m2*) exhibited high CNVs frequencies (50 – 100%) in all the populations (Fig 6).

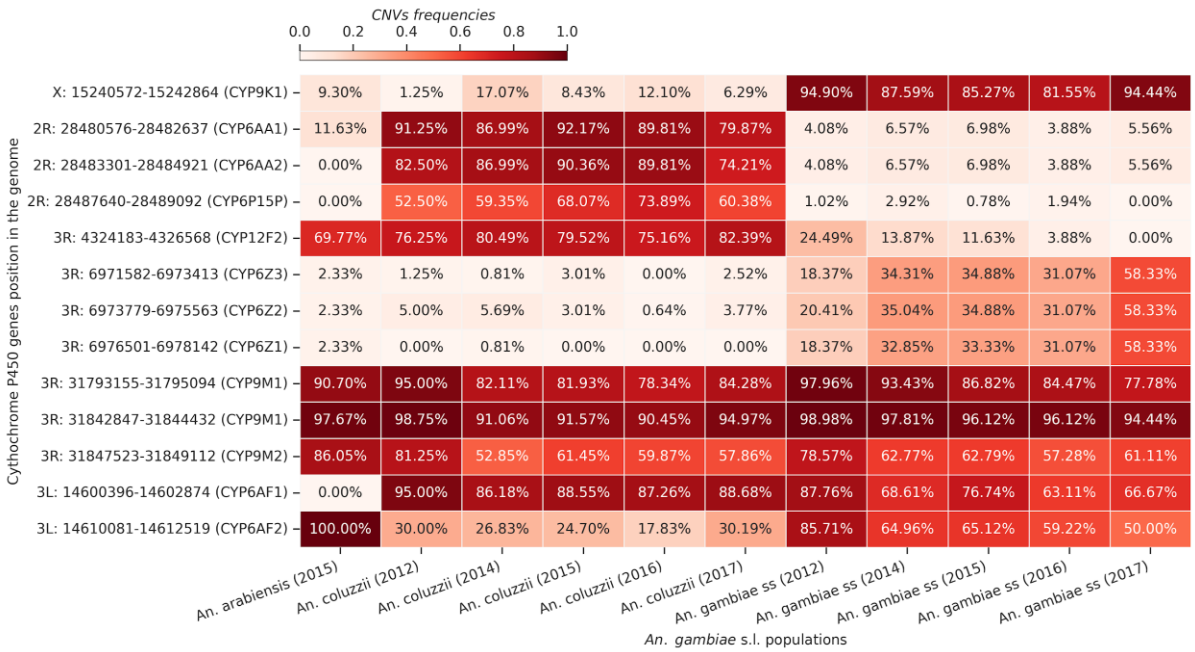

Fig 6. Heat map of the cytochrome P450 genes showing high CNVs frequencies in the *An. gambiae* s.l. populations. The X axis shows the *An. gambiae* s.l. populations and the years of collection. The Y axis shows the positions of the P450 genes in the genome. The gradient color bar shows the distribution of the allelic frequencies.

Organophosphate and carbamate resistance: increasing threats to control plus a potential novel IR mechanism

Indoor residual spraying (IRS) remains another major vector control strategy in Africa [30]. However, the coverage of this strategy dropped considerably over the years because of the emergence of resistance to the pyrethroid compounds and also the low residual efficacy of carbamates compounds [31]. Recently, IRS with Actellic 300 CS containing pirimiphos-methyl, an organophosphate as the active ingredient, showed interesting results including a prolonged residual efficacy up to seven months and a reduction of vector density in many sprayed areas in Africa [32–34]. This new formulation demonstrated better ability to target pyrethroid resistant and indoor resting mosquitoes and showed additional benefit when used in combination with LLINs [35,36].

Using the time series collection, we investigated the magnitude of resistance to organophosphates and carbamates resistance in Burkina Faso that could threaten the long term efficacy of pirimiphos-methyl-based IRS. A sliding window genome-wide F_{ST} scan showed a year to year (from 2012 to 2017) genetic difference in some regions of the genome especially on chromosomes 2 and X in *An. gambiae* s.s. and *An. coluzzii* samples. Most of these changes were observed in genomic regions previously shown to be associated with insecticide-resistance suggesting a signal of positive selection in these genes (Fig 8, 9 and 10).

Both *Ace1* and Diacylglycerol Kinase (DGK) loci showed no selection signal in 2012 in any vector populations we sampled, but by 2017 selection signals were detected at both loci in *An. gambiae* s.s. populations. The increasing gene amplification signal in the *Ace1* locus, from 32.65% in 2012 to up to 91.67% in 2017, was corroborated by the increase in the insecticide resistance associated *Ace1-G280S* variant previously

shown to be involved in organophosphate and carbamate resistance [37] (from 15.81% in 2012 to 65.27% in 2017), but only in *An. gambiae* s.s. populations (Fig 7). We also identified 243 other non-synonymous SNPs in the *Ace1* gene (Table S3). CNV amplifications were also found in *An. coluzzii* populations, but at low and constant frequencies (~ 3%), but not in *An. arabiensis* populations.

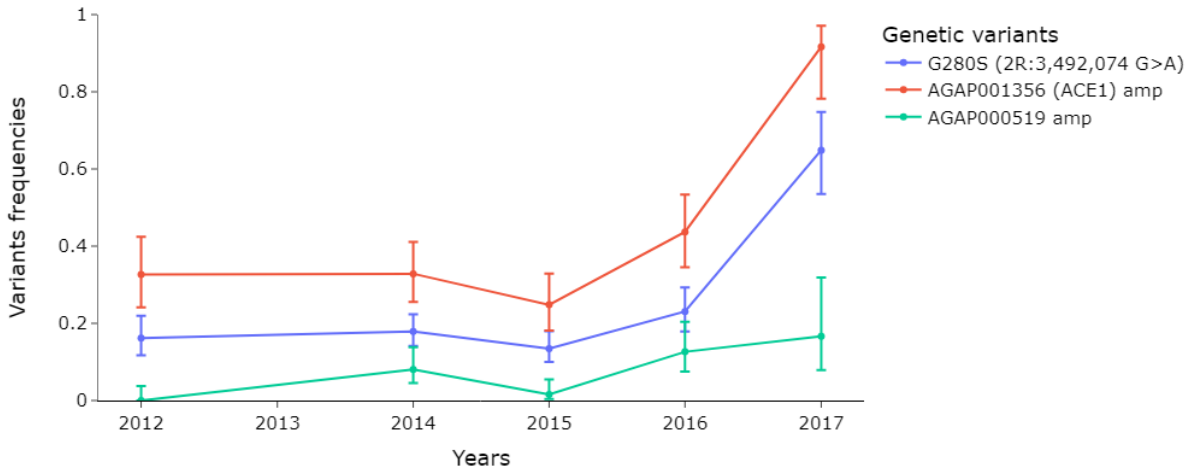

Fig 7. Evolution of the *Ace1* and Diacylglycerol Kinase (AGAP000519) CNV frequencies in the *An. gambiae* s.s. populations.

Genetic variants of the *Gste* genes have also been reported to be involved in resistance to organophosphates [38]. In this study, we identified CNVs in 41 genes (24 genes with amplifications and 17 with deletions) of the *Gste* genes. These CNVs are distributed across the genome (Table S4). Most of these CNVs were found at low frequencies (< 50%) except the *GstD5* which showed high CNV frequencies (60% to 100%) in all the populations (Fig S6, Table S4).

The carboxylesterases (COEs) enzymes have been reported to be involved in the rapid metabolization of both pyrethroid and organophosphate compounds [39] and consequently could have an impact on the benefit of the combined organophosphate-

based IRS and pyrethroid-based LLINs approach. In this study, CNVs were identified in 39 genes across the genome, most of these CNVs were gene amplifications, only 25.64% were deletions (Table S5). The frequencies of these CNVs were relatively low, except those of the *Coeae*60 gene which were identified at frequencies above 60% in *An. coluzzii* populations from 2012 to 2017 (Fig S7). Additionally, a positive selection signal was also detected in the *Coeae*60 gene (Fig 8). The *Coeae*G* cluster genes (*Coeae2G, Coeae3G, Coeae5G, Coeae6G* and *Coeae7G*) and the *Coeae3H* gene showed CNVs at increasing frequencies from 11% in 2012 to up to 25% in 2017 in the *An. gambiae* s.s. populations. The *Coejhe4E* gene showed CNVs at 32.56% frequencies in *An. arabiensis*, and also at frequencies higher than 30% in *An. gambiae* s.s. populations (Fig S7).

The emergence of new forms of insecticide resistance may impact vector control strategy in unexpected ways. Recent studies reported new forms of resistance, especially behavioural changes (early and/or outdoor biting) in *An. gambiae* mosquitoes, in response to the intense use of LLINs causing residual malaria transmission [40,41]. Here, we detected a signal of positive selection in the DGK locus and also an increased frequency of CNVs of this gene within the *An. gambiae s.s.* populations (Fig 9 and 10). No selection signal was detected in vector populations in 2012, with no CNVs identified either. However, in 2017, a signal of positive selection was detected in the DGK locus, with an associated increasing frequency (from 0.00 in 2012 to 16.67% in 2017) of CNVs in vector populations (Fig 7, 9 and 10). Considering this evolution pattern of the DGK gene and also its clear role in the adaptation of *C. elegans* and *Drosophila sp* to environmental change [42,43], this gene may play an important role in the modulation of *An. gambiae* biology in response to the presence of vector control tools.

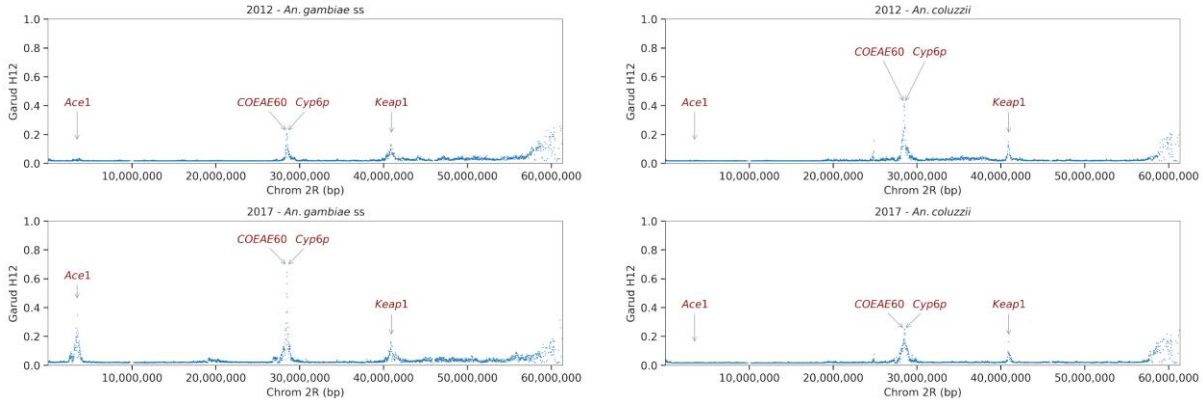

Fig 8. Evolution of the Garud H12 in the chromosome arm 2R of the *An. gambiae s.l.* populations from 2012 to 2017. High values of H12 indicate a signal of positive selection in the corresponding genomic region within the populations: *Ace1: Acetylcholinesterase gene, Cyp6p: Cytochrome P450 gene, COEAE60: Carboxyl-esterase gene, Keap1: Kelch-like ECH-Associated Protein 1.*

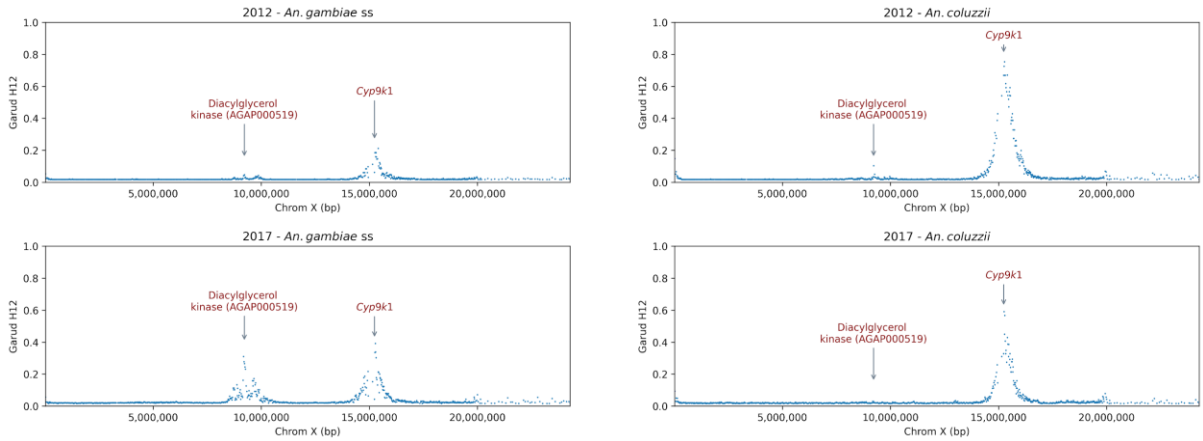

Fig 9. Garud H12 statistic showing a signal of positive selection in the *DGK* and the *Cyp9k1* genes within the *An. gambiae* s.l. populations.

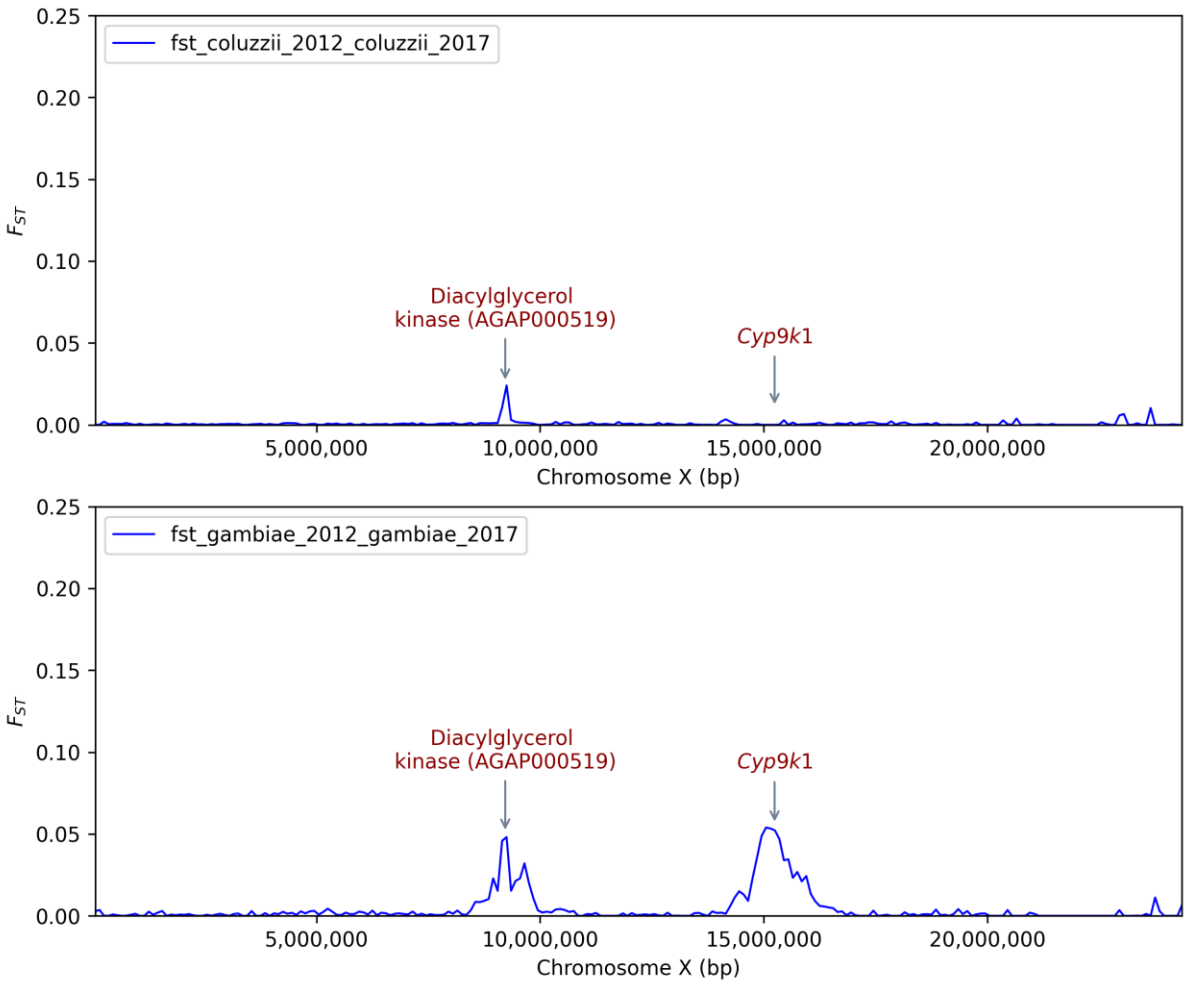

Fig 10. Year to year variation of the genetic differentiation between populations of the same species of *An. gambiae* complex collected in 2012 and in 2017. Signals of differentiation were observed in some regions involved in insecticides resistance including *Cyp9k1* and DGK gene.

Discussion

Significant advances have been made in the reduction of malaria burden using LLINs and IRS, but the efficacy of these tools is at risk due to the high adaptability of the vector populations to the anthropogenic ecological changes [30]. A better understanding of the molecular, ecological and evolutionary processes allowing vector populations to adapt to the environmental changes would help in the development of innovative tools to support the current control tools in malaria elimination. In this study, we used whole genome sequencing data to study the genetic diversity, the population structure and the evolution of insecticide resistance variants in the *An. gambiae* s.l. populations in the western part of Burkina Faso.

Bednets – evidence and impacts of target-site and metabolic pyrethroid resistance

The insecticides pressure on malaria vectors lead to an emergence of new insecticide resistance variants that confer fitness benefits to vector populations [44]. This seems to be consistent and evident for the SNPs *Vgsc-V402L+I1527T* which were identified at increasing frequencies in *An. coluzzii* populations. The rapid rise of their frequencies in Burkina Faso *An. coluzzii* populations may be expected given the results of a recent study which revealed the *Vgsc-V402L* produces a pyrethroid resistance phenotype with reduced fitness costs, when compared with *Vgsc-L995F* [27]. The simultaneous emergence of different types of genetic variants around genes and gene clusters known to be associated with pyrethroid metabolism and target site mutations within the *An. gambiae* s.l. populations raise major concerns in the long term efficacy of new pyrethroid insecticide-based control strategies such as LLINs.

Insecticides resistance has been widely studied in Africa, and recent works have reported multiple resistance variants spreading in vectors populations, jeopardizing

NMCP efforts in control and elimination [17,18,45]. The implementation of LLINs next-generation bed nets incorporating piperonyl butoxide (PBO) aims to strengthen the efficacy of LLINs by inhibiting the cytochromes P450 enzymes responsible for the rapid detoxifying of pyrethroid insecticide [46]. The high expression of the P450 enzymes was shown to increase the pyrethroid resistance and pre-exposure to PBO restores the susceptibility of malaria vector to pyrethroid [28]. In most cases, the main genetic variants responsible for the insecticide resistance in a given area remain unclear because of the high heterogeneity of the genetic markers mediating metabolic resistance. Although PBO inhibits the cytochrome genes, the emergence of additional new variants could enhance the level of pyrethroid resistance impacting the efficacy of the new vector control tools.

IRS - evidence and impacts of organophosphates resistance

Another major concern related to this multi-modal resistance is its threat on the effectiveness of the indoor residual spraying (IRS). This strategy is efficient when it is used in combination with LLINs because of its ability to target indoor resting pyrethroid resistant vectors [47]. Although the SNP *Ace1-G280S* has long been identified in *An. gambiae* populations, its frequencies remained low in the populations probably due to its fitness cost in association with the kdr mutation [48]. Indeed, the duplication of the *Ace1* gene (*ace1*^{D}) seems to have completely solved the problem of genetic cost yielded by the SNP *Ace1-G280S* [49]. Recent reports showed a wide distribution of this duplication in West African *An. coluzzii* and *An. gambiae* s.s. populations [18,49] probably in response to the use of organophosphate and carbamates compounds in vector control. Similarly, the duplication of *Ace1* gene follows the same trend as the SNP *Ace1-G280S* raising up in frequencies in *An. gambiae* s.s. populations in our study. This clearly showed the close related link between these two variants in

conferring high fitness to the mosquito populations. In addition, the CNVs identified at high frequencies in some of the *Gste* and COE genes could increase the expression level of these genes. The simultaneous presence of these variants (CNVs and SNPs) in *Ace1*, *Gste* and esterases genes within the same vector populations strongly contribute to increase the survival capacity of the vector against organophosphates and carbamates. This situation threatens the long-term efficacy of vector control interventions including the recently introduced pirimiphos-methyl-based IRS.

Whole genome surveillance enables detection of potential novel IR loci

Genome-wide scans for positive selection have proved their usefulness in the identification of genomic regions under divergent selection. In malaria vectors, under high insecticide mediated evolutionary pressure, these regions often have implications in insecticide resistance [50]. We detected a signal of positive selection in a genomic region of the X chromosome corresponding to the DGK gene. Though there is currently no evidence of this gene's association with insecticide resistant phenotypes in malaria vectors, studies in *Drosophila melanogaster* and *Caenorhabditis elegans* showed its potential involvement in the adaptation to environmental changings [42,43]. In *D. melanogaster,* the rdgA gene, an orthologue of the DGK gene, is eye-specific, strongly involved in light signaling. Any genetic variant influencing the functioning of this gene would probably affect its sensitivity to light, as reported previously [42,51]. Consequently, if light signaling is regulated the same way in mosquito species, the DGK gene may also be involved in the regulation of *An. gambiae* vision. In *An. gambiae*, a previous study showed a rhythmic expression of the genes involved in light signaling under the control of the circadian clock [52]. The expression pattern of these genes suggests their role in calibration of the mosquito visual system to start the nocturnal activities such as flight and biting. In response to the high coverage of LLINs,

major changes were reported in vector biting behavior particularly early biting and later biting activities in many areas of Africa [31,41]. Even though no evidence was highlighted, these changes could be related to alterations in the expression of genes involved in the mosquito visual system.

Another possible implication of the DGK gene in the adaptation of vector populations is its indirect implication to insecticide resistance via the regulation of the acetylcholine release at synaptic junctions. In *C. elegans*, the DGK gene regulates the amount of acetylcholine in synaptic junctions and its inactivation causes a high susceptibility to aldicarb, a carbamate insecticide [43]. This report clearly showed the potential implication of the DGK gene in organophosphates and carbamates resistance *C. elegans*. If synaptic junctions are regulated in the same way in mosquitoes, any selective variant in the DGK gene would affect the acetylcholine release in synapses. These two hypotheses of the potential implication of the DGK gene in adaptations, either insecticide resistance or biting time change, were previously drawn up to explain the selection signal observed in the DGK gene [50]. These hypotheses seem different in both organisms, but the molecular mechanisms underlying the DGK gene function remain similar. Given this intriguing role of DGK, further studies are needed to deeply investigate the role of the DGK gene (AGAP000519) in modulating malaria vector behavior and its potential involvement in behavioral and genetic resistance.

Conclusion

Investigation of genome variation in vector populations is crucial in understanding dynamics of populations in a given area and guiding strategies of malaria control programs. This study showed a weak differentiation and strong genetic connection between *An. gambiae* and *An. coluzzii* populations in the western area of Burkina Faso. The emergence of new variants together with the already existing pyrethroids resistance variants at high frequencies raise concerns about the long-term efficacy of new-generation nets impregnated with pyrethroids compounds. The long-term efficacy of pirimiphos-methyl-based IRS is also threatened because of the increasing frequency of the SNP *Ace1-G280S* and the duplication of the *Ace1* gene, which confer high fitness to vector populations carrying them. Genome-wide selection scans detected strong signals of positive selection in the genomic regions having an implication in insecticide resistance. This study highlighted the benefit of the molecular surveillance of malaria vectors in the detection of new insecticide resistance variants, the monitoring of the existing resistance variants and also to get insights in evolutionary processes underlying these variants.

Methods

Sample collection

Mosquito samples were collected by the Target Malaria project in three sites, Bana (11.233, -4.472), Souroukoudinga (11.235, -4.535) and Pala (11.150, -4.235), located in the western part of Burkina Faso (see Results - Figure 1). These villages are located in the humid savannah zone surrounding the Bobo-Dioulasso city and are characterized by two seasons: a rainy period from May to October and a dry season from November to April. The average annual rainfall in this area is above 800mm (average annual temperature ~ 27 °C and humidity ~ 60%) [6]. The main agricultural practices are cotton and cereal (maize, rice, sorghum etc.) cultivation utilizing a large amount of pesticides [53].

Malaria remains endemic in the area surrounding the Bobo-dioulasso city, causing 338 deaths in the three health districts of the city [10]. *An. gambiae* complex species remains the major malaria vector in the sampling area [6]. Malaria control relies on the antimalarial drugs and prevention of mosquito bites using long-lasting insecticidal nets (LLINs), the most widespread vector control tool, and to a lesser extent indoor residual spraying (IRS), sprays and mosquito-repellent coils.

Mosquitoes were collected during the rainy and dry seasons from 2012 to 2017 using three collection methods: Pyrethroid spray catches (PSC), human landing catches (HLC) and swarm collection [54]. PSC and HLC were used for female mosquito collection, whereas males were collected in swarms. After collection, mosquito samples were morphologically identified using the morphological keys [55] and *An. gambiae s.l.* specimens were stored in 80% ethanol and then shipped to Wellcome Sanger Institute for the whole genome sequencing.

Whole genome sequencing and genomic data management

An. gambiae s.l. mosquitoes were sequenced individually at high coverage (30x) using Illumina technology at the Wellcome Sanger Institute. Genomic data were analyzed at the MalariaGEN Resource Centre Team using a wide range of genome analysis software. Reads were aligned to the AgamP4 reference genome using BWA version 0.7.15. Indel realignment was performed using GATK version 3.7-0 RealignerTargetCreator and IndelRealigner. Single nucleotide polymorphisms were called using GATK version 3.7-0 UnifiedGenotyper. Genotypes were called for each sample independently, in genotyping mode, given all possible alleles at all genomic sites where the reference base was not "N". Coverage was capped at 250× by random down-sampling. Complete specifications of the alignment and genotyping pipelines are available from the malariagen/pipelines GitHub repository [56]. After variant calling, the raw sequences in FASTQ format and the aligned sequences in BAM format were stored in the European Nucleotide Archive (ENA, Study Accession n° PRJEB42254). The genetic variants in VCF and zarr formats, including the samples' metadata, were stored on Google Cloud and are accessible via the malariagen-data package or are directly downloadable. More details regarding the sequencing technology, the genomic data storage and management including the SNPs variants calling, haplotypes phasing, and copy number variants identification are available on the homepage of MalariaGEN (https://malariagen.github.io/vector-data/ag3/methods.html).

Data analyses

Genetic diversity and population structure were analyzed in the python programming language on the Google Colab platform. The genetic variants (SNPs, haplotypes and

CNVs) and the reference genome of *An. gambiae* s.l. were accessed using the malariagen-data package for direct analyses of the genomic data in the cloud. Specific python packages including malariagen-data [57], scikit-allel [58] and other python standard data-management packages were used for the data analyses. The diversity statistics (segregating sites, nucleotide diversity (θ_π), Tajima's D (D), Watterson theta (θ_w), allele frequency spectrum) were computed in the whole chromosome in all the mosquito populations. Tukey multiple comparison test was applied to compare the diversity statistic between vector populations over the year. Genome wide selection scan was performed using the Garud h12 [59] statistic to detect the genomic region under positive selection. F_{ST} and PCA analyses were performed in each chromosome to investigate the genetic structure and the differentiation of *An. gambiae* s.l. populations. The extent and the spread of insecticide resistance variants were also investigated in the *An. gambiae* s.l. populations using the malariagen_data python package. Time series (from 2012 to 2017) SNPs frequencies were analyzed in known target-site resistance associated genes, *Vgsc, Rdl* and *Ace1*. Also copy number variation (CNV) frequencies were investigated in genomic regions associated with metabolic resistance, cytochrome P450s, glutathione-s-transferases, acetylcholinesterase, diacylglycerol kinase and carboxylesterases.

Ethics approval and consent to participate

All methods in this paper have been implemented in accordance with the relevant guidelines/regulations/legislation in Burkina Faso. No ethics approval was required to run all the activities related to this paper.

Consent for publication

Not applicable.

Availability of data and materials

Jupyter Notebooks and scripts to reproduce all the analyses, tables and figures are available in the GitHub repository: https://github.com/mkient/BF_Ag1000G_analyses. The SNPs and haplotypes data are available on the homepage of MalariaGEN and can be accessed using the malariagen-data package.

Competing interests

The authors declare no competing interests.

Funding

The sample collection is supported by Target Malaria, which receives core funding from the Bill and Melinda Gates Foundation and Open Philanthropy. Whole genome sequencing and genomics data management were supported by MalariaGEN at the Wellcome Sanger Institute.

Authors' contributions

MK, CSC and AM conceived the study. AB and AD provided the resources. AB, AMGB and AD supervised the study. AAM, NT, PSA, SOL and FYA carried out samples collection in the

field. AM, CSC and JB produced the genomic data. MK, CSC, HTY and JB carried out data analysis and visualization. MK and CSC drafted the manuscript. All authors have read and approved this version of the manuscript.

Acknowledgements

The authors would like to acknowledge the Institut de Recherche en Sciences de la Santé, its funders and partners for supporting the sample collection. We are grateful to the Ag1000G project, the MalariaGEN, Sanger Core and GSU Vector Surveillance Programme at the Wellcome Sanger Institute, and their partners in Africa for the production of the *An. gambiae* genomic data. The authors also thank the populations of the sampling sites for their sincere cooperation during the mosquito sampling.

References

1. WHO. World malaria report 2022. Geneva; 2022. Available: https://apps.who.int/iris/rest/bitstreams/1484818/retrieve
2. Hemming-Schroeder E, Zhong D, Machani M, Nguyen H, Thong S, Kahindi S, et al. Ecological drivers of genetic connectivity for African malaria vectors Anopheles gambiae and An. arabiensis. Sci Rep. 2020;10: 19946. doi:10.1038/s41598-020-76248-2
3. Costantini C, Ayala D, Guelbeogo WM, Pombi M, Some CY, Bassole IHN, et al. Living at the edge: Biogeographic patterns of habitat segregation conform to speciation by niche expansion in anopheles gambiae. BMC Ecol. 2009;9: 1–27. doi:10.1186/1472-6785-9-16
4. Fournet F, Cussac M, Ouari A, Meyer PE, Toé HK, Gouagna LC, et al. Diversity in anopheline larval habitats and adult composition during the dry and wet seasons in Ouagadougou (Burkina Faso). Malar J. 2010;9: 1–10. doi:10.1186/1475-2875-9-78
5. Soma DD, Zogo BM, Somé A, Tchiekoi BN, Hien DF de S, Pooda HS, et al. Anopheles bionomics, insecticide resistance and malaria transmission in southwest Burkina Faso: A pre-intervention study. Chitnis N, editor. PLoS One. 2020;15: e0236920. doi:10.1371/journal.pone.0236920
6. Epopa PS, Collins CM, North A, Millogo AA, Benedict MQ, Tripet F, et al. Seasonal malaria vector and transmission dynamics in western Burkina Faso. Malar J. 2019;18: 113. doi:10.1186/s12936-019-2747-5
7. Ministere de la Santé. Rapport De Revue De Performances Du Programme National De Lutte 2011-2015- Burkina Fasso. Ouagadougou; 2016.
8. INSD, ICF. Enquête Démographique et de Santé 2021 : Rapport de synthèse.

Rockville, Maryland; 2023. Available: https://dhsprogram.com/pubs/pdf/SR279/SR279.pdf

9. U.S. President's Malaria Initiative. Burkina Faso Malaria Operational Plan FY 2023. 2023. Available: www.pmi.gov

10. Ministère de la Santé. Annuaire statistique 2021. Ouagadougou; 2022. Available: http://cns.bf/IMG/pdf/annuaire_mshp_2021_signe.pdf

11. Coulibaly ZI, Gowelo S, Traore I, Mbewe RB, Ngulube W, Olanga EA, et al. Strengthening adult mosquito surveillance in Africa for disease control: learning from the present. Curr Opin Insect Sci. 2023;60: 101110. doi:10.1016/j.cois.2023.101110

12. The Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector Anopheles gambiae. Nature. 2017;552: 96–100. doi:10.1038/nature24995

13. The Anopheles gambiae 1000 Genomes Consortium. Genome variation and population structure among 1142 mosquitoes of the African malaria vector species Anopheles gambiae and Anopheles coluzzii. Genome Res. 2020;30: 1533–1546. doi:10.1101/gr.262790.120

14. MalariaGEN. Ag1000G | MalariaGEN. MalariaGEN. 2014. Available: https://www.malariagen.net/mosquito/ag1000g

15. Weetman D, Djogbenou LS, Lucas E. Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? Curr Opin Insect Sci. 2018;27: 82–88. doi:10.1016/j.cois.2018.04.005

16. Lucas ER, Miles A, Harding NJ, Clarkson CS, Lawniczak MKN, Kwiatkowski DP, et al. Whole-genome sequencing reveals high complexity of copy number variation at insecticide resistance loci in malaria mosquitoes. Genome Res.

2019;29: 1250–1261. doi:10.1101/gr.245795.118

17. Clarkson CS, Miles A, Harding NJ, O'Reilly AO, Weetman D, Kwiatkowski D, et al. The genetic architecture of target-site resistance to pyrethroid insecticides in the African malaria vectors Anopheles gambiae and Anopheles coluzzii. Mol Ecol. 2021;30: 5303–5317. doi:10.1111/mec.15845

18. Grau-Bové X, Lucas E, Pipini D, Rippon E, van 't Hof AE, Constant E, et al. Resistance to pirimiphos-methyl in West African Anopheles is spreading via duplication and introgression of the Ace1 locus. ffrench-Constant RH, editor. PLOS Genet. 2021;17: e1009253. doi:10.1371/journal.pgen.1009253

19. Amoudji AD, Ahadji-Dabla KM, Hien AS, Apétogbo YG, Yaméogo B, Soma DD, et al. Insecticide resistance profiles of Anopheles gambiae s.l. in Togo and genetic mechanisms involved, during 3-year survey: is there any need for resistance management? Malar J. 2019;18: 177. doi:10.1186/s12936-019-2813-z

20. Moyes CL, Wiebe A, Gleave K, Trett A, Hancock PA, Padonou GG, et al. Analysis-ready datasets for insecticide resistance phenotype and genotype frequency in African malaria vectors. Sci Data. 2019;6. doi:10.1038/s41597-019-0134-2

21. Sovi A, Keita C, Sinaba Y, Dicko A, Traore I, Cisse MBM, et al. Anopheles gambiae (s.l.) exhibit high intensity pyrethroid resistance throughout Southern and Central Mali (2016–2018): PBO or next generation LLINs may provide greater control. Parasit Vectors. 2020;13: 239. doi:10.1186/s13071-020-04100-7

22. Sougoufara S, Ottih EC, Tripet F. The need for new vector control approaches targeting outdoor biting anopheline malaria vector communities. Parasit Vectors.

2020;13: 295. doi:10.1186/s13071-020-04170-7

23. Wilson AL, Courtenay O, Kelly-Hope LA, Scott TW, Takken W, Torr SJ, et al. The importance of vector control for the control and elimination of vector-borne diseases. Barrera R, editor. PLoS Negl Trop Dis. 2020;14: e0007831. doi:10.1371/journal.pntd.0007831

24. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. Science (80-). 2015;347. doi:10.1126/science.1258522

25. Hemingway J, Ranson H. Insecticide resistance in insect vectors of human disease. Annu Rev Entomol. 2000;45: 371–91. doi:10.1146/annurev.ento.45.1.371

26. Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ, et al. Footprints of positive selection associated with a mutation (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae. Proc Natl Acad Sci. 2012;109: 6614–6619. doi:10.1073/pnas.1201475109

27. Williams J, Cowlishaw R, Sanou A, Ranson H, Grigoraki L. In vivo functional validation of the V402L voltage gated sodium channel mutation in the malaria vector An. gambiae. Pest Manag Sci. 2022;78: 1155–1163. doi:10.1002/ps.6731

28. Protopopoff N, Mosha JF, Lukole E, Charlwood JD, Wright A, Mwalimu CD, et al. Effectiveness of a long-lasting piperonyl butoxide-treated insecticidal net and indoor residual spray interventions, separately and together, against malaria transmitted by pyrethroid-resistant mosquitoes: a cluster, randomised controlled, two-by-two fact. Lancet. 2018;391: 1577–1588. doi:10.1016/S0140-6736(18)30427-6

29. Chandor-Proust A, Bibby J, Régent-Kloeckner M, Roux J, Guittard-Crilat E, Poupardin R, et al. The central role of mosquito cytochrome P450 CYP6Zs in insecticide detoxification revealed by functional expression and structural modelling. Biochem J. 2013;455: 75–85. doi:10.1042/BJ20130577

30. Ranson H. Current and Future Prospects for Preventing Malaria Transmission via the Use of Insecticides. Cold Spring Harb Perspect Med. 2017;7: a026823. doi:10.1101/cshperspect.a026823

31. Sherrard-Smith E, Skarp JE, Beale AD, Fornadel C, Norris LC, Moore SJ, et al. Mosquito feeding behavior and how it influences residual malaria transmission across Africa. Proc Natl Acad Sci. 2019;116: 15086–15095. doi:10.1073/pnas.1820646116

32. Kenea O, Balkew M, Tekie H, Deressa W, Loha E, Lindtjørn B, et al. Impact of combining indoor residual spraying and long-lasting insecticidal nets on Anopheles arabiensis in Ethiopia: results from a cluster randomized controlled trial. Malar J. 2019;18: 182. doi:10.1186/s12936-019-2811-1

33. Loha E, Deressa W, Gari T, Balkew M, Kenea O, Solomon T, et al. Long-lasting insecticidal nets and indoor residual spraying may not be sufficient to eliminate malaria in a low malaria incidence area: results from a cluster randomized controlled trial in Ethiopia. Malar J. 2019;18: 141. doi:10.1186/s12936-019-2775-1

34. Soma DD, Zogo B, Hien DF de S, Hien AS, Kaboré DA, Kientega M, et al. Insecticide resistance status of malaria vectors Anopheles gambiae (s.l.) of southwest Burkina Faso and residual efficacy of indoor residual spraying with microencapsulated pirimiphos-methyl insecticide. Parasit Vectors. 2021;14: 58. doi:10.1186/s13071-020-04563-8

35. Rowland M, Boko P, Odjo A, Asidi A, Akogbeto M, N'Guessan R. A New Long-Lasting Indoor Residual Formulation of the Organophosphate Insecticide Pirimiphos Methyl for Prolonged Control of Pyrethroid-Resistant Mosquitoes: An Experimental Hut Trial in Benin. Vontas J, editor. PLoS One. 2013;8: e69516. doi:10.1371/journal.pone.0069516

36. Hamainza B, Sikaala CH, Moonga HB, Chanda J, Chinula D, Mwenda M, et al. Incremental impact upon malaria transmission of supplementing pyrethroid-impregnated long-lasting insecticidal nets with indoor residual spraying using pyrethroids or the organophosphate, pirimiphos methyl. Malar J. 2016;15: 100. doi:10.1186/s12936-016-1143-7

37. Weill M, Malcolm C, Chandre F, Mogensen K, Berthomieu A, Marquine M, et al. The unique mutation in ace-1 giving high insecticide resistance is easily detectable in mosquito vectors. Insect Mol Biol. 2004;13: 1–7. doi:10.1111/j.1365-2583.2004.00452.x

38. Riveron JM, Yunta C, Ibrahim SS, Djouaka R, Irving H, Menze BD, et al. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol. 2014;15: R27. doi:10.1186/gb-2014-15-2-r27

39. Cui F, Li M-X, Chang H-J, Mao Y, Zhang H-Y, Lu L-X, et al. Carboxylesterase-mediated insecticide resistance: Quantitative increase induces broader metabolic resistance than qualitative change. Pestic Biochem Physiol. 2015;121: 88–96. doi:10.1016/j.pestbp.2014.12.016

40. Moiroux N, Damien GB, Egrot M, Djenontin A, Chandre F, Corbel V, et al. Human Exposure to Early Morning Anopheles funestus Biting Behavior and Personal Protection Provided by Long-Lasting Insecticidal Nets. Vontas J,

editor. PLoS One. 2014;9: e104967. doi:10.1371/journal.pone.0104967

41. Finda MF, Moshi IR, Monroe A, Limwagu AJ, Nyoni AP, Swai JK, et al. Linking human behaviours and malaria vector biting risk in south-eastern Tanzania. Carvalho LH, editor. PLoS One. 2019;14: e0217414. doi:10.1371/journal.pone.0217414

42. Masai I, Okazaki A, Hosoya T, Hotta Y. Drosophila retinal degeneration A gene encodes an eye-specific diacylglycerol kinase with cysteine-rich zinc-finger motifs and ankyrin repeats. Proc Natl Acad Sci. 1993;90: 11157–11161. doi:10.1073/pnas.90.23.11157

43. Miller KG, Emerson MD, Rand JB. Goα and Diacylglycerol Kinase Negatively Regulate the Gqα Pathway in C. elegans. Neuron. 1999;24: 323–333. doi:10.1016/S0896-6273(00)80847-8

44. Benelli G, Beier JC. Current vector control challenges in the fight against malaria. Acta Trop. 2017;174: 91–96. doi:10.1016/J.ACTATROPICA.2017.06.028

45. Sougoufara S, Doucouré S, Sembéne PMB, Harry M, Sokhna C. Challenges for malaria vector control in sub-Saharan Africa: Resistance and behavioral adaptations in Anopheles populations. J Vector Borne Dis. 2017;54: 4–15. Available: http://www.ncbi.nlm.nih.gov/pubmed/28352041

46. Feyereisen R. Insect P450 inhibitors and insecticides: challenges and opportunities. Pest Manag Sci. 2015;71: 793–800. doi:10.1002/ps.3895

47. Okumu FO, Moore SJ. Combining indoor residual spraying and insecticide-treated nets for malaria control in Africa: a review of possible outcomes and an outline of suggestions for the future. Malar J. 2011;10: 208. doi:10.1186/1475-2875-10-208

48. Djogbénou L, Noel V, Agnew P. Costs of insensitive acetylcholinesterase insecticide resistance for the malaria vector Anopheles gambiae homozygous for the G119S mutation. Malar J. 2010;9: 12. doi:10.1186/1475-2875-9-12

49. Assogba BS, Djogbénou LS, Milesi P, Berthomieu A, Perez J, Ayala D, et al. An ace-1 gene duplication resorbs the fitness cost associated with resistance in Anopheles gambiae, the main malaria mosquito. Sci Rep. 2015;5: 14529. doi:10.1038/srep14529

50. Miles A. Genomic epidemiology of malaria vectors in the Anopheles gambiae species complex. University of Oxford. 2021. Available: https://ora.ox.ac.uk/objects/uuid:d0fc0f47-e24f-4804-9022-26c4e3cf1428/download_file?safe_filename=thesis.pdf&type_of_work=Thesis

51. Delgado R, Muñoz Y, Peña-Cortés H, Giavalisco P, Bacigalupo J. Diacylglycerol Activates the Light-Dependent Channel TRP in the Photosensitive Microvilli of Drosophila melanogaster Photoreceptors. J Neurosci. 2014;34: 6679–6686. doi:10.1523/JNEUROSCI.0513-14.2014

52. Rund SSC, Hou TY, Ward SM, Collins FH, Duffield GE. Genome-wide profiling of diel and circadian gene expression in the malaria vector Anopheles gambiae. Proc Natl Acad Sci. 2011;108. doi:10.1073/pnas.1100584108

53. Naré RWA, Savadogo PW, Gnankambary Z, Nacro HB, Sedogo MP. Analyzing Risks Related to the Use of Pesticides in Vegetable Gardens in Burkina Faso. Agric For Fish. 2015;4: 165. doi:10.11648/j.aff.20150404.13

54. Epopa PS, Millogo AA, Collins CM, North AR, Benedict MQ, Tripet F, et al. Anopheles gambiae (s.l.) is found where few are looking: assessing mosquito diversity and density outside inhabited areas using diverse sampling methods. Parasit Vectors. 2020;13: 516. doi:10.1186/s13071-020-04403-9

55. Gillies MT, Coetzee M. A SUPPLEMENT TO THE ANOPHELINAE OF AFRICA SOUTH OF THE SAHARA (AFROTROPICAL REGION). The South African Institute for Medical Research. Johannesburg; 1987.

56. Malariagen. MalariaGEN pipelines - Pipelines for processing malaria parasite and mosquito genome sequence data. In: MalariaGEN pipelines - releases 9 v0.0.9 [Internet]. 2022. Available: https://github.com/malariagen/pipelines

57. Miles A. GitHub - malariagen/malariagen-data-python: A Python package providing functions for accessing and analysing MalariaGEN data. In: malariagen / malariagen-data-python [Internet]. 2022 [cited 6 May 2022]. Available: https://github.com/malariagen/malariagen-data-python

58. Miles A, Bot P i., R. M, Ralph P, Harding N, Pisupati R, et al. Scikit-allel - Explore and analyse genetic variation. Zenodo. 2021 [cited 6 May 2022]. doi:10.5281/ZENODO.4759368

59. Garud NR, Messer PW, Buzbas EO, Petrov DA. Recent Selective Sweeps in North American Drosophila melanogaster Show Signatures of Soft Sweeps. Copenhaver GP, editor. PLOS Genet. 2015;11: e1005004. doi:10.1371/journal.pgen.1005004

Supplementary information

Table S1. Non-synonymous SNPs frequencies of the *Vgsc* gene (2L: 2358158 - 2431617) within *An. gambiae* s.l. populations.

Table S2. Copy number variations of the cytochrome p450 genes and their frequencies within *An. gambiae* s.l. populations.

Table S3. Non-synonymous SNPs frequencies of the *Ace1* gene (2R: 3484107 - 3495790) within An. gambiae s.l. populations.

Table S4. Copy number variations of the glutathione-s-transferase genes and their frequencies within *An. gambiae* s.l. populations.

Table S5. Copy number variations of the carboxylesterase genes and their frequencies within *An. gambiae* s.l. populations.

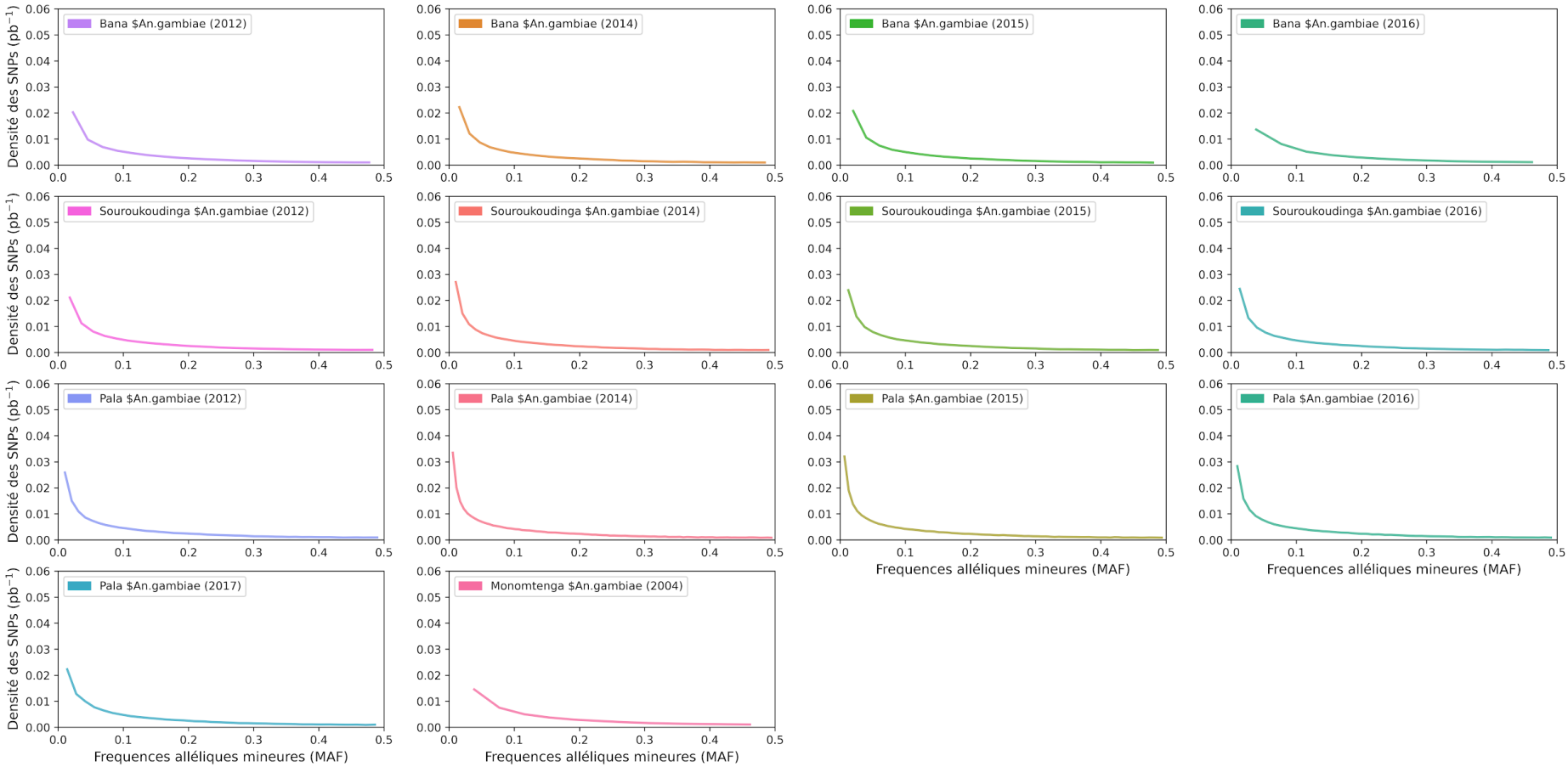

Fig. S1. Site frequency spectra in the 3L chromosome of the *An. gambiae* s.s. populations from 2012 to 2016 in the three villages.

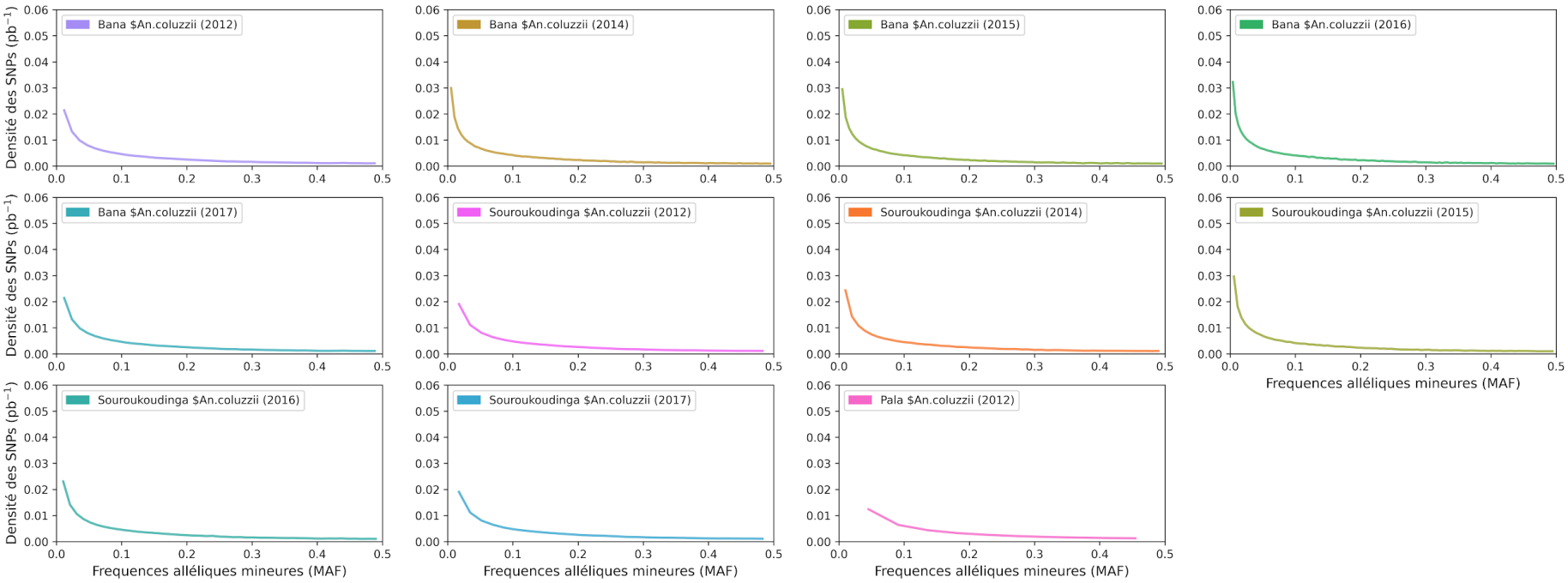

Fig.S2. Site frequency spectra in the 3L chromosome of the *An. coluzzii* populations from 2012 to 2017 in the three villages.

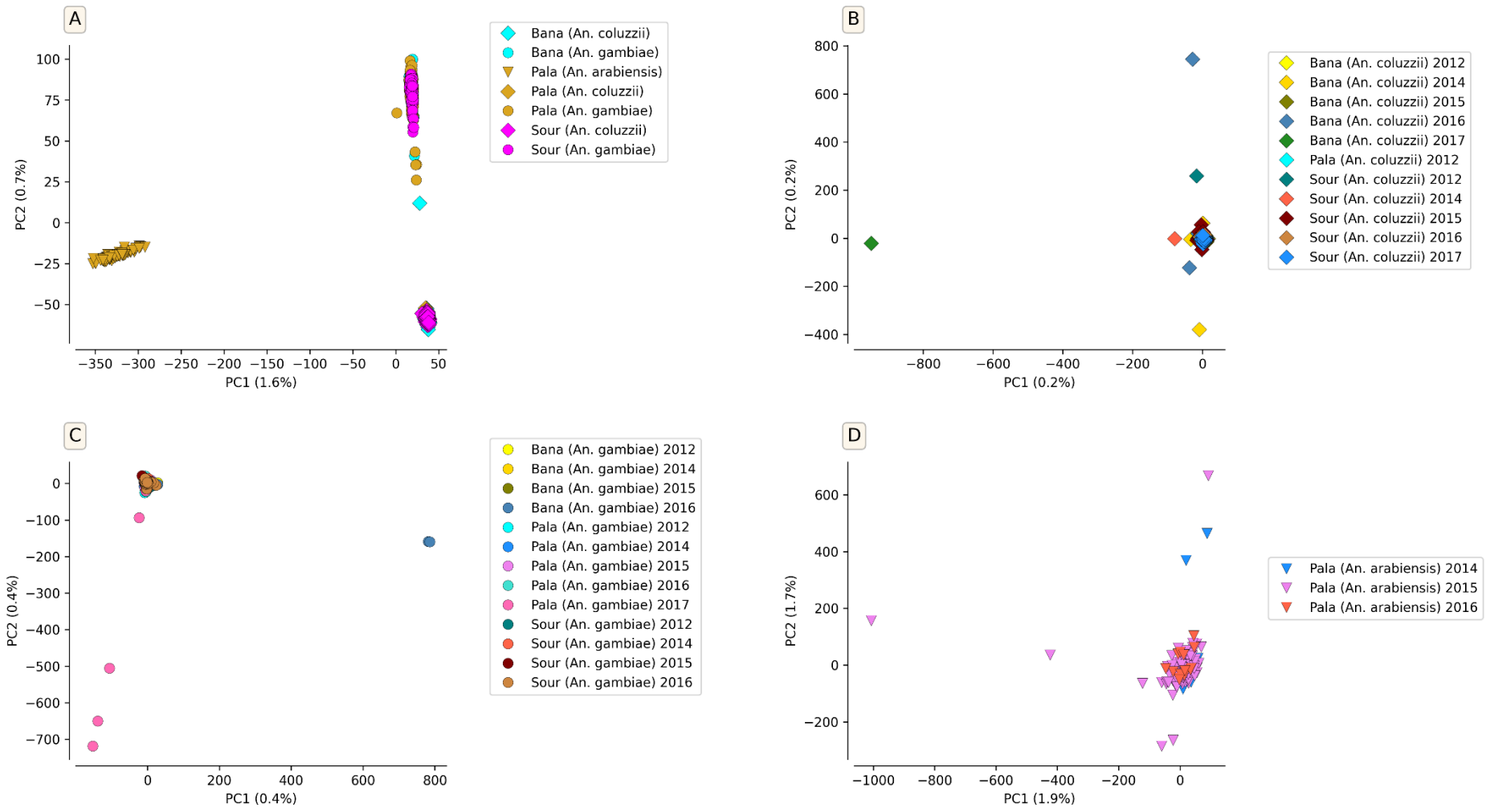
A
PC2 (0.7%)
PC1 (1.6%)
Bana (An. coluzzii)
Bana (An. gambiae)
Pala (An. arabiensis)
Pala (An. coluzzii)
Pala (An. gambiae)
Sour (An. coluzzii)
Sour (An. gambiae)
B
PC2 (0.2%)
PC1 (0.2%)
Bana (An. coluzzii) 2012
Bana (An. coluzzii) 2014
Bana (An. coluzzii) 2015
Bana (An. coluzzii) 2016
Bana (An. coluzzii) 2017
Pala (An. coluzzii) 2012
Sour (An. coluzzii) 2012
Sour (An. coluzzii) 2014
Sour (An. coluzzii) 2015
Sour (An. coluzzii) 2016
Sour (An. coluzzii) 2017
C
PC2 (0.4%)
PC1 (0.4%)
Bana (An. gambiae) 2012
Bana (An. gambiae) 2014
Bana (An. gambiae) 2015
Bana (An. gambiae) 2016
Pala (An. gambiae) 2012
Pala (An. gambiae) 2014
Pala (An. gambiae) 2015
Pala (An. gambiae) 2016
Pala (An. gambiae) 2017
Sour (An. gambiae) 2012
Sour (An. gambiae) 2014
Sour (An. gambiae) 2015
Sour (An. gambiae) 2016
D
PC2 (1.7%)
PC1 (1.9%)
Pala (An. arabiensis) 2014
Pala (An. arabiensis) 2015
Pala (An. arabiensis) 2016

Fig. S3. PCA showing the year-to-year (from 2012 to 2017) genetic structure of *An. gambiae* s.l. populations using the whole SNPs identified in the 3L chromosome. B. Year-to-year genetic structure of *An. coluzzii* populations; C. Year-to-year genetic structure of *An. gambiae* s.s. populations; D. Year-to-year genetic structure of *An. arabiensis* populations; These figures demonstrated the lack of geographic substructure within each species of the *An. gambiae* s.l. populations collected in different villages over the years.

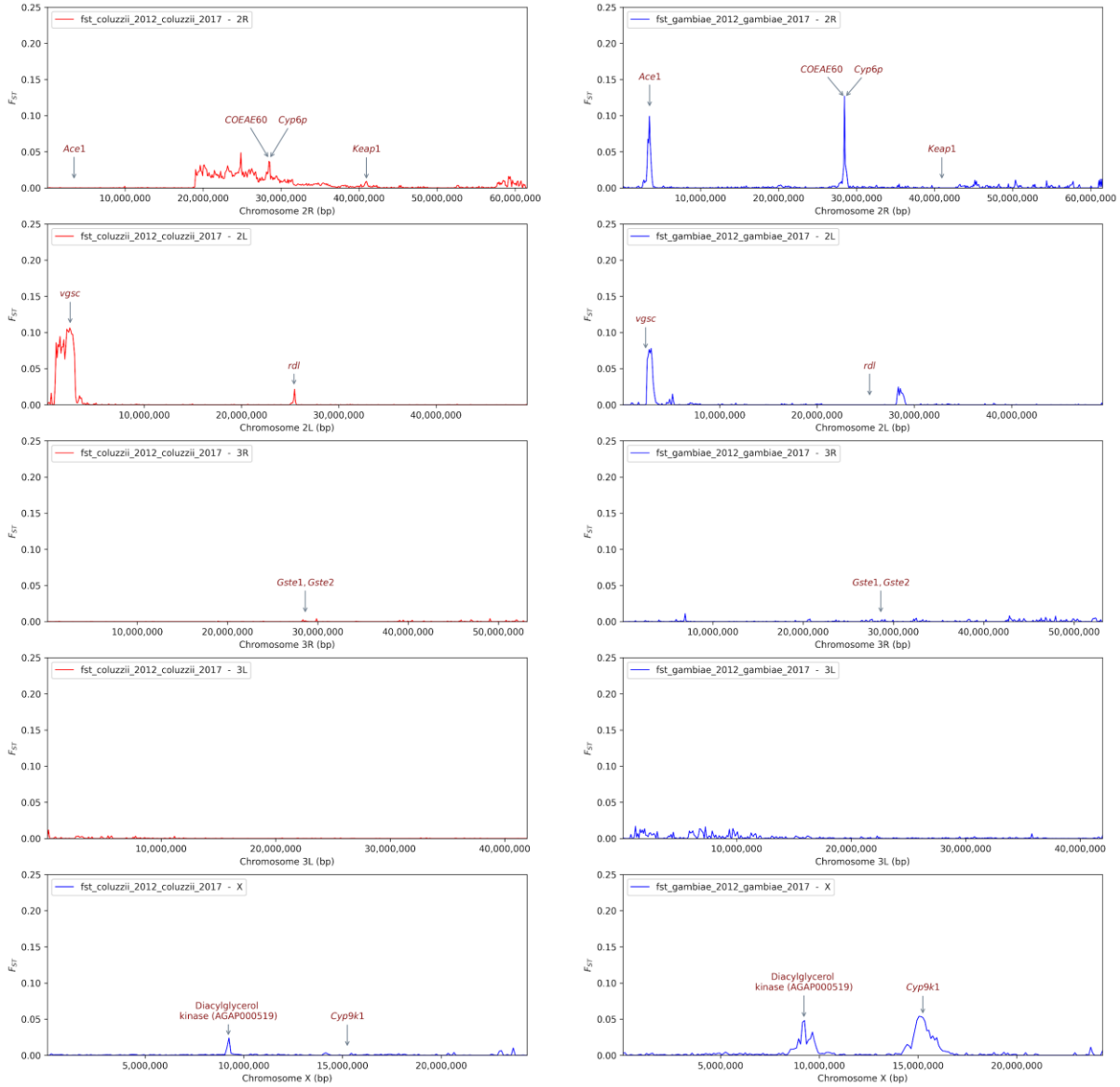

Fig. S4. Year to year variation of the genetic differentiation between populations of the same species of *An. gambiae* complex collected in 2012 and in 2017. Signals of differentiation were observed in some regions involved in insecticides resistance. *Ace1*: acetylcholinesterase gene, *Cyp6p*: Cytochrome P450 gene, *Keap1*: Kelch-like ECH-associated protein 1, *Vgsc*: voltage-gated sodium channel, *rdl*: resistance to dieldrin gene, *Gste1*: glutathione S-transferases epsilon 1, *Gste2*: glutathione-S-transferases epsilon 2, *Cyp9k1*: Cytochrome P450 gene.

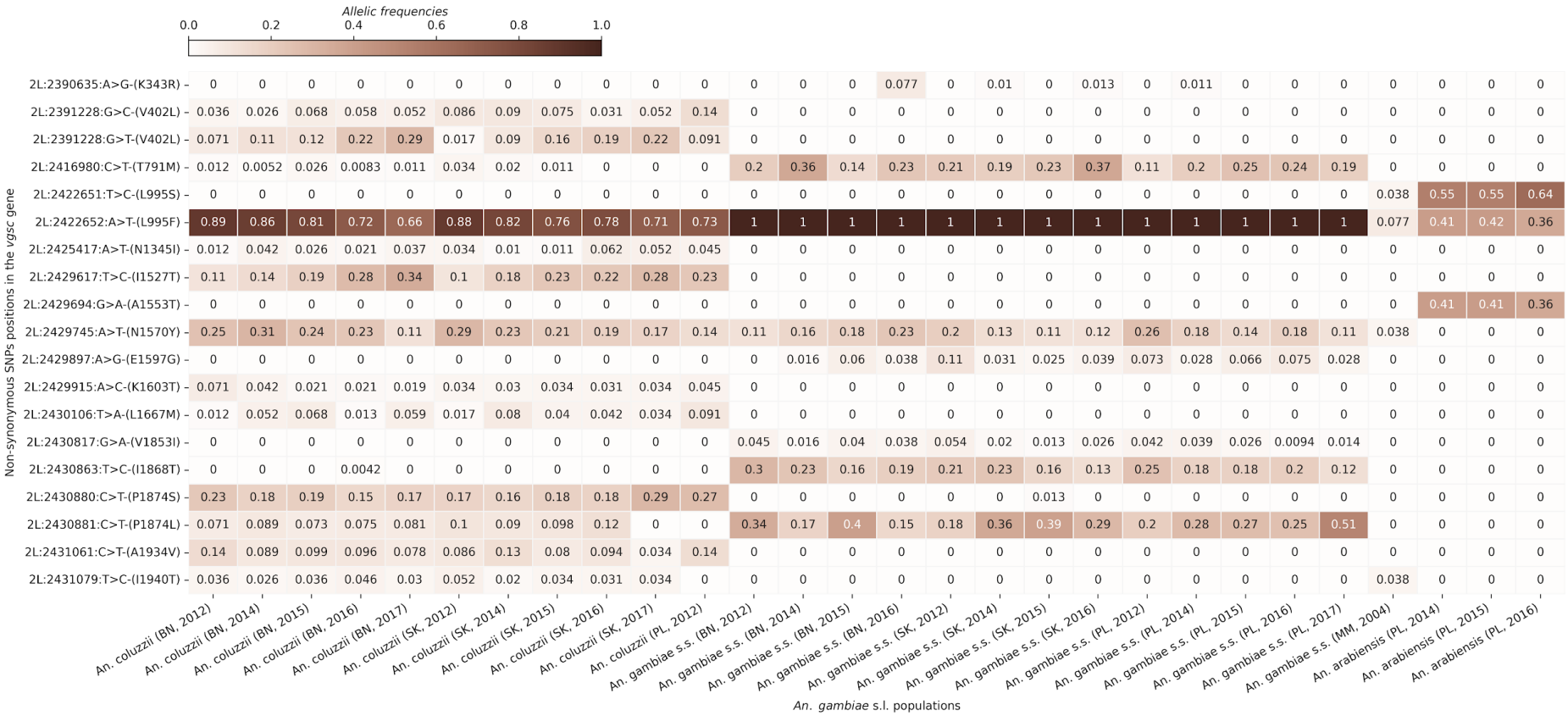

Fig. S5. Heat map showing the evolution of the non-synonymous SNPs frequencies (max freq > 0.05) in the *Vgsc* gene within the *An. gambiae* s.l. populations and the sampling sites. The X axis shows the *An. gambiae* s.l. populations and the years of collection. The Y axis shows the non-synonymous SNPs positions in the chromosome 2L. The gradient color bar shows the distribution of the allelic frequencies. **BN:** Bana, **SK**: Souroukoudinga, **PL**: Pala, **MM**: Monomtenga.

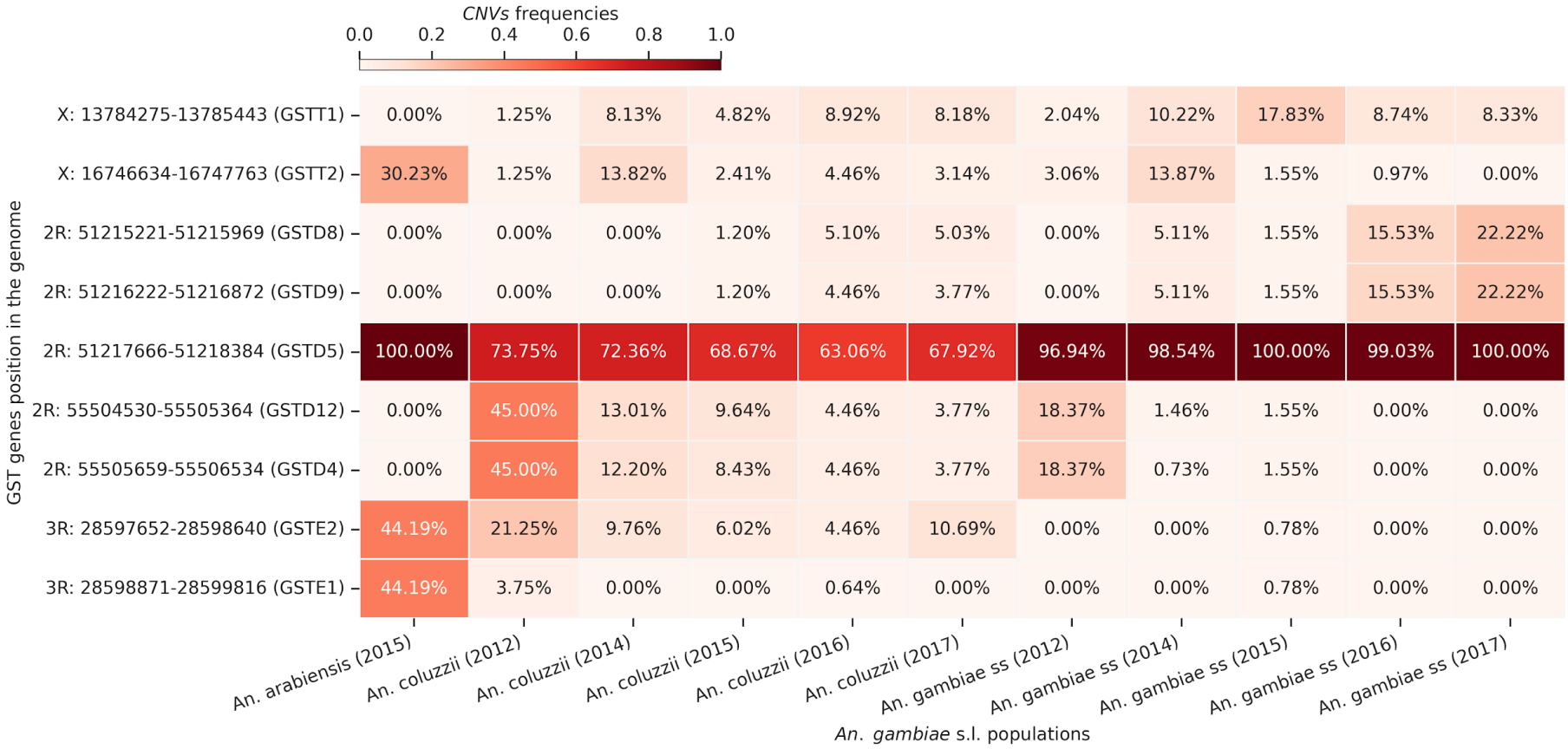

Fig. S6. Heat map showing the CNVs frequencies (max freq > 0.05) of the glutathione-s-transferase genes in the *An. gambiae* s.l. populations. The X axis shows the *An. gambiae* s.l. populations and the years of collection. The Y axis shows the positions of the glutathione-s-transferase genes in the genome. The gradient color bar shows the distribution of the allelic frequencies.

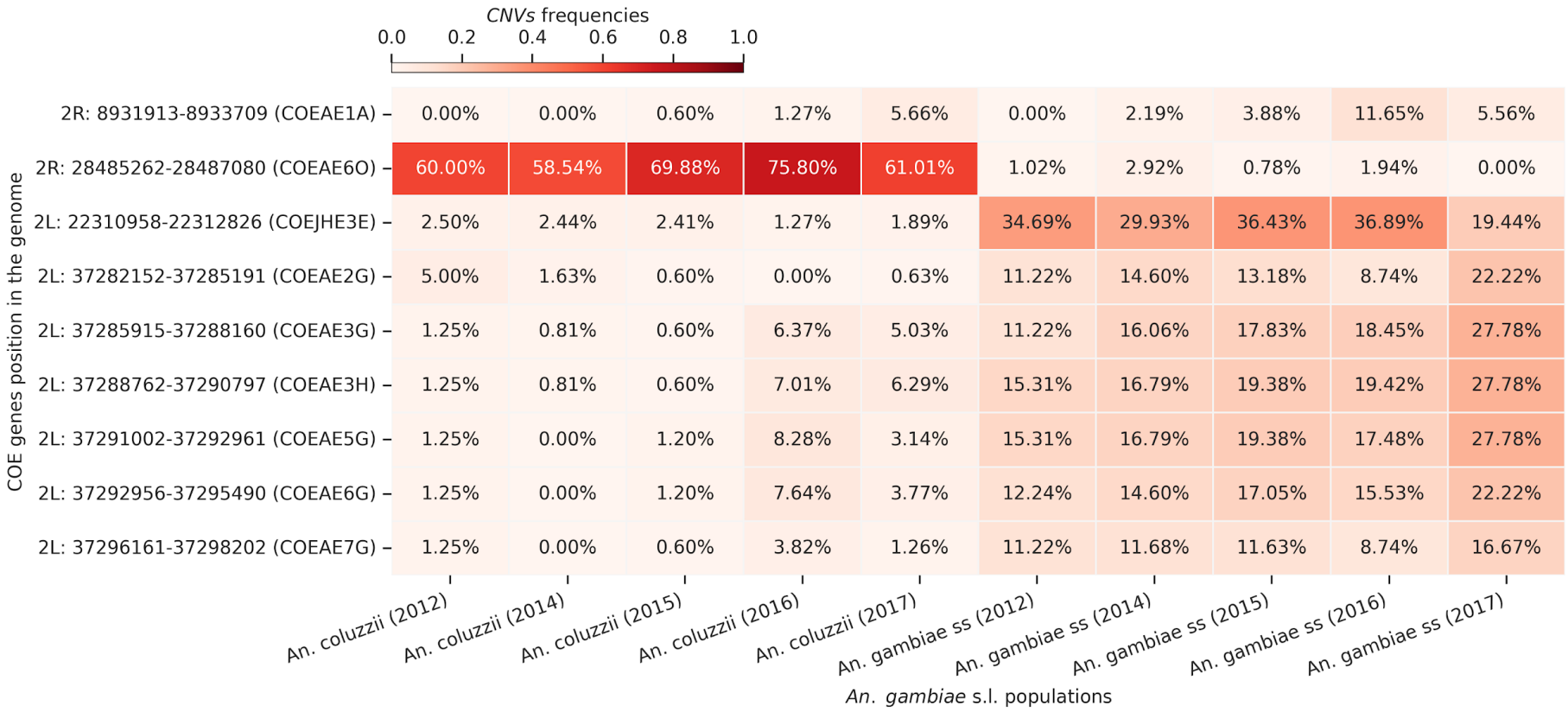

Fig. S7. Heat map showing the CNVs frequencies (max freq > 0.05) of the carboxylesterase genes in the *An. gambiae* s.l. populations. The X axis shows the *An. gambiae* s.l. populations and the years of collection. The Y axis shows the positions of the carboxylesterase genes in the genomes. The gradient color bar shows the distribution of the allelic frequencies.