Positional cloning of rp2 QTL associates the P450 genes CYP6Z1, CYP6Z3 and CYP6M7 with pyrethroid resistance in the malaria vector Anopheles funestus

H Irving¹, JM Riveron¹, SS Ibrahim¹, NF Lobo² and CS Wondji¹

Pyrethroid resistance in Anopheles funestus is threatening malaria control in Africa. Elucidation of underlying resistance mechanisms is crucial to improve the success of future control programs. A positional cloning approach was used to identify genes conferring resistance in the uncharacterised rp2 quantitative trait locus (QTL) previously detected in this vector using F6 advanced intercross lines (AIL). A 113 kb BAC clone spanning rp2 was identified and sequenced revealing a cluster of 15 P450 genes and one salivary protein gene (SG7-2). Contrary to A. gambiae, AFcYP6M1 is triplicated in A. funestus, while AgCYP6Z2 orthologue is absent. Five hundred and sixty-five new single nucleotide polymorphisms (SNPs) were identified for genetic mapping from rp2 P450s and other genes revealing high genetic polymorphisms with one SNP every 36 bp. A significant genotype/phenotype association was detected for rp2 P450s but not for a cluster of cuticular protein genes previously associated with resistance in A. gambiae. QTL mapping using F6 AIL confirms the rp2 QTL with an increase logarithm of odds score of 5. Multiplex gene expression profiling of 15 P450s and other genes around rp2 followed by individual validation using qRT-PCR indicated a significant overexpression in the resistant FUMOZ-R strain of AFcYP6Z1, AFcYP6Z3, AFcYP6M7 and the glutathione-s-transferase GSTe2 with respective fold change of 11.2, 6.3, 5.5 and 2.8. Polymorphisms analysis of AFcYP6Z1 and AFcYP6Z3 identified amino acid changes potentially associated with resistance further indicating that these genes are controlling the pyrethroid resistance explained by the rp2 QTL. The characterisation of this rp2 QTL significantly improves our understanding of resistance mechanisms in A. funestus. Heredity (2012) 109, 383–392; doi:10.1038/hdy.2012.53; published online 5 September 2012

Keywords: Anopheles funestus; malaria; insecticide resistance; QTL mapping; P450s

BACKGROUND

Malaria control in Africa relies heavily on vector control through the use of insecticide treated nets, long-lasting insecticide nets and indoor residual spraying. However, resistance to the main insecticides such as pyrethroids is threatening the success of these control methods.

Anopheles funestus, a major vector of malaria throughout much of sub-Saharan Africa (Gillies and De Meillon 1968), is increasingly developing resistance to different classes of insecticides used in public health, such as pyrethroids, carbamates and DDT with the fear that this could disrupt control programs against this vector. Indeed, resistance to pyrethroids, DDT and carbamates has been detected in different regions of Africa, such as Southern Africa (Mozambique (Hargreaves et al., 2000; Casimiro et al., 2006; Cuamba et al., 2010) and Malawi (Hunt et al., 2010)), East Africa (Uganda (Morgan et al., 2010)), West Africa (Ghana (Okoye et al., 2008) and Benin (Djouaka et al., 2011)) and Central Africa (Cameroon (Wondji et al., 2011)). To improve the management of these resistances, it is imperative to thoroughly characterise the underlying mechanisms in order to design appropriate control strategies.

Efforts to characterise resistance mechanisms in A. funestus have benefited from recent progress made in the study of this species notably the colonisation of two strains, one resistant to pyrethroids named FUMOZ-R originally from Mozambique and the other FANG fully susceptible to all insecticides and originally from Angola (Hunt et al., 2005). Other progress that have facilitated genetic studies in A. funestus include the construction of a map (Sharakhov et al., 2004), an integrated genetic and physical map (Wondji et al., 2005), identification of a set of genome-wide distributed single nucleotide polymorphisms (SNPs) (Wondji et al., 2007a) and recent sequencing of A. funestus transcriptome (Crawford et al., 2010; Gregory et al., 2011).

Using the resistant (FUMOZ-R) and the susceptible (FANG) laboratory strains, previous studies have identified a major quantitative trait locus (QTL) associated with pyrethroid, named rp1 (for resistance to permethrin) located on chromosome 2R. This rp1 QTL detected with F2 mapping (Wondji et al., 2007b) and also with F6 and F8 advanced intercross lines (AIL) (Wondji et al., 2009) explains 87% of the genetic variance in pyrethroid susceptibility in two families from reciprocal crosses between susceptible and resistant strains. Two additional QTLs named rp2 and rp3 were also detected in chromosomes 2L and 3L, respectively. A positional cloning approach was used to identify the genes conferring pyrethroid resistance in rp1 using AIL at F6 and F8 generations. This involved the sequencing of a 120-kb BAC clone spanning the rp1 QTL, which identified fourteen
protein coding genes and one putative pseudogene (Wondji et al., 2009). Ten of the fourteen genes encoded cytochrome P450s and expression analysis indicated that three of these P450s (CYP6P9, AfCYP6P4 and CYP6AA4) were upregulated in the resistant strain. Furthermore, CYP6P9 and AfCYP6P4, respectively, 25 and 51 times overexpressed in resistant females, were tandemly duplicated compared with A. gambiae.

If rp1 QTL has been well characterised and the genes involved in pyrethroid resistance detected, this is not the case for the other two QTLs notably rp2 the second most important QTL. Given the importance to fully characterise the mechanisms of pyrethroid resistance in this species, it is fundamental to also identify genes associated with pyrethroid resistance in rp2 QTL. Recent observations that CYP6P9 duplicated genes in rp1 are also associated at various degrees in pyrethroid resistance in field populations of An. funestus in Africa (Morgan et al., 2010; Djouaka et al., 2011) indicates that further analysis of rp2 and rp3 QTLs in FUMOZ-R can help to better characterise the resistance in field populations. In A. gambiae, the other major malaria vector, genes associated with pyrethroid resistance such as AgCYP6M2 (Djouaka et al., 2008) and AgCYP6Z1 (David et al., 2005) are found in the chromosome 3R, which is the equivalent to chromosome 2L in An. funestus were rp2 is located. It remains to be established whether the orthologues of these genes could be associated with pyrethroid resistance in An. funestus.

Here, we report the positional cloning of genes in the rp2 QTL associated with pyrethroid resistance in An. funestus, using the AIL approach with F1 generation progeny from reciprocal crosses between the FUMOZ-R resistant and the FANG susceptible strains.

**MATERIALS AND METHODS**

**Isolation and sequencing of the BAC clone containing the rp2 QTL**

An An. funestus BAC library from The Institute for Genomic Research, Notre Dame University, was screened by PCR using primers from nine P450 genes located within the boundaries of the rp2 QTL using the synteny projection with the A. gambiae chromosomal map: these genes are CYP6Z3, CYP6Z1, CYP6Y1, CYP6M8, CYP6M1, CYP6H18, CYP6N2 and CYP6M7. The primers used are listed in Supplementary Table S1. DNA of whole 384-well plates was pooled and a PCR carried out for each plate. The positive plates were then subdivided into six column pools and 4 row pools and the PCR screen repeated. Finally, individual colonies from the set of 16 identified from the pooled column and row screen were used as template to identify the individual clone containing the markers of interest. The BAC clone was grown at 37°C overnight and harvested in a glycerol solution and stored at −80°C. The size of the BAC clone was estimated after a restriction digestion using the BamHI restriction enzyme to separate the insert from the vector. The BAC clone was then fragmented by sonication into shorter, random sequences of around 2–5 kb. These small fragments were cloned into the plasmid vector (pC31) and sequenced. The reads were trimmed into the plasmid vector (pC31) and sequenced. The reads were trimmed using Sequencher (http://www.dnastar.com/web/ir13.php) and assembled using Phrap (P. Green unpublished; http://www.genome.washington.edu/UWGCGFanalysistools/Phrap.cfm).

Analysis of BAC clone sequence and annotation of the P450 genes

The programme Genemarker.hmm version 2.2 was used to locate genes in the assembled sequence of the BAC clone. Putative genes were annotated by using BlastX (http://www.ncbi.nlm.nih.gov/blast/) and the predicted transcripts were also compared with A. gambiae transcripts in Vectorbase (www.vectorbase.org). Further detailed annotation of the P450 genes was aided by the P450 site (http://p450.sophia.inra.fr/). Sequence alignments of An. funestus and A. gambiae genes were carried out using ClustalW (Thompson et al., 1994). Exact boundaries of the exon/intron were manually inspected and compared with that of A. gambiae using the DNAStar sequence analysis package. MEGA 4.0 (Tamura et al., 2007) was used to construct a Neighbour-Joining tree of the An. funestus genes in comparison with A. gambiae.

**Mosquito samples**

Two AIL families at F6 generations were used for this study. These families (Family 1 and Family 10) were generated from reciprocal crosses between the susceptible and the resistant strains as previously described (Wondji et al., 2009). Briefly, a two exposure times approach of 30 min and 2 h was used to select the most susceptible and the most resistant mosquitoes, respectively, as described previously in order to minimise the level of phenotype misclassification and increases the power of QTL detection (Lander and Botstein, 1989).

**SNP identification and genotyping in the reciprocal isofemale lines**

In order to carry out a fine-scale mapping of the rp2 QTL, SNPs were identified in all the genes detected in the BAC clone and in other genes spanning the rp2 QTL boundaries in the 2L chromosome such as the glutathione-s-transferase GST2, the cuticular protein genes CFC, CFC9 and CFC8, the P450s CYP9M1, CYP303A1 and CYP4H18, and AGAP007980 and AGAP009073. The cuticular protein genes were amplified using primers listed in Supplementary Table S2. Genomic DNA was extracted using the LiVAK method (Livak, 1984) and amplified using the primers listed in Supplementary Table S3. For each family, the F2 parental female was amplified with 3F and in order to detect the informative SNPs to be used for genetic mapping. The PCR were carried out using 10 pmol of each primers and 30 ng of genomic DNA as template in 25 ml reactions. PCR products were purified using the QiAquick PCR purification kit (Qiagen, Hilden, Germany) and directly sequenced on both strands. Sequences for each family were analysed to detect the polymorphic sites manually using BioEdit and as sequence differences in multiple alignments using ClustalW (Thompson et al., 1994).

SNPs were identified as transitions or transversions in coding and non-coding regions. Genetic diversity analysis was performed using DnaSP 5.1 (Rozas et al., 2003). The nucleotide diversity π was calculated for each gene as well as the haplotype diversity. The average number of synonymous substitutions per synonymous site (Ks) and non-synonymous substitutions per non-synonymous site (Ka) was computed.

**SNP genotyping.** Equal numbers of surviving and dead F1 progeny were genotyped for each of the two reciprocal families (75 dead and 75 alive for family 1, 48 dead and 48 alive for family 10) with the informative SNP using pyrosequencing method according to the manufacturer’s instructions using the PSQ 96 SNP Reagent Kit (Biotage AB, Uppsala, Sweden) and as previously described (Wondji et al., 2007b). Primers details are given in Supplementary Table S4.

**QTL mapping.** The JoinMap 2.0 package (Stam and Van Ooijen, 1995) was used to build a genetic linkage map for each of the two families using the same parameters as described previously (Wondji et al., 2007b). Windows QTL Cartographer 2.5 (Wang et al., 2005) software was used to plot the genetic map. Associations between genotypes at each locus and the resistance phenotype were assessed using a contingency χ2-analysis. The null hypothesis was that susceptibility to permethrin is equal in each genotype class. For loci with a significant χ2, we analysed the inheritance of the alleles at these loci. The a priori hypothesis was that a higher mortality rate would occur among F1 individuals with one or both alleles inherited from the susceptible parent. The JoinMap linkage map and the genotype/phenotype data were entered into Windows QTL Cartographer 2.5 (Wang et al., 2005). Interval mapping, composite-interval mapping and multiple-interval mapping procedures were implemented for each family as previously described (Wondji et al., 2007b).

**Multiplex GeXP expression profiling of rp2 and 2L chromosome genes**

The expression pattern of 28 genes located in the rp2 BAC clone or in 2L chromosome was compared between the resistant strain FUMOZ-R and the susceptible strain FANG using the GeXP multiplex gene expression profiling method from Beckman Coulter as previously described (Wondji et al., 2009). Total RNA was extracted from three batches of ten (1–3-day old) female and male adult mosquitoes using a PicoPure RNA isolation kit (Arcturus, Mountain View, CA, USA) according to manufacturer’s instructions. Total RNA quantity and quality were assessed using Nanodrop spectrophotometer (Nanodrop Technologies, Oxfordshire, UK) and Bioanalyzer (Agilent, Santa
Clara, CA, USA), respectively. The GenomeLab eXpress Profiler programme was used for automated primer design, calculation of relative gene expression values, data checking and data analysis. The quantitative PCR reaction was carried out using the GenomeLab GeXP Start Kit (Beckman and Coulter, Brea, CA, USA) according to the protocol provided and using the primers listed in Supplementary Table S5. The amplified qPCR products were diluted 10 times and added in a 96-well microplate with the DNA size standard-400 and ran on the GenomeLab GeXP genetic analysis system. The expression level of the RSP7 ribosomal gene was used to normalise for variation in total cDNA concentration. A two-sample t-test was used to compare the results between the two strains.

Expression analysis of potential candidate genes
Genes with a significant overexpression in the FUMOZ-R strain from the GeXP multiplex expression profiling was further analysis using individual quantitative reverse transcription PCR (qRT–PCR). To increase the specificity of the amplification, reverse primers were designed in the 3’UTR region for all the genes. Primers are listed in Supplementary Table S6. One microgram of total RNA from each of the three biological replicates for FUMOZ-R and FANG was used as template for cDNA synthesis using the Superscript III (Invitrogen, Carlsbad, CA, USA) with oligo-dT20 and RNase H, according to the manufacturer’s instructions. A serial dilution of cDNA was used to establish standard curves for each gene in order to assess PCR efficiency and quantitative differences between samples. The qRT–PCR amplification was carried out in a MX 3000 real-time PCR system (Agilent) using Brilliant III Ultra-Fast SYBR Green QPCR Master Mix (Agilent). Ten nanogram of cDNA from each sample was used as template in a three-steps programme involving a denaturation at 95°C for 3 min followed by 40 cycles of 10 s at 95°C and 10 s at 60°C and a last step of 1 min at 95°C, 30 s at 55°C and 95°C at 30 s. The relative expression and fold change of each target gene in FUMOZ-R relative to FANG was calculated according to the 2^{-ΔΔCT} method incorporating PCR efficiency normalisation with the housekeeping RSP7 ribosomal protein gene S7 (AGAP010592) and the actin SC (AGAP000651) genes. A two-sample t-test was used to compare the results between the two strains.

Analysis of polymorphism patterns of candidate genes
In order to detect any potential polymorphism associated with resistance, the full length of the candidate genes detected after qRT–PCR was amplified using the cDNA synthesised from total RNA as template between the resistant and susceptible strains. Amplification was performed with the Phusion polymerase with the following conditions: 1 cycle at 95°C for 1 min; 1 cycle at 95°C for 30 s at 30°C and 95°C at 30 s. The relative expression and fold change of each target gene in FUMOZ-R relative to FANG was calculated according to the 2^{-ΔΔCT} method incorporating PCR efficiency normalisation with the housekeeping RSP7 ribosomal protein gene S7 (AGAP010592) and the actin SC (AGAP000651) genes. Ten nanogram of cDNA from each sample was used as template in a three-steps programme involving a denaturation at 95°C for 3 min followed by 40 cycles of 10 s at 95°C and 10 s at 60°C and a last step of 1 min at 95°C, 30 s at 55°C and 95°C at 30 s. The relative expression and fold change of each target gene in FUMOZ-R relative to FANG was calculated according to the 2^{-ΔΔCT} method incorporating PCR efficiency normalisation with the housekeeping RSP7 ribosomal protein gene S7 (AGAP010592) and the actin SC (AGAP000651) genes. A two-sample t-test was used to compare the results between the two strains.

RESULTS

BAC clone isolation and sequencing
Nine positive 384-well plates containing one or more of the loci spanning rp2 were identified from the A. funestus BAC library. A single individual BAC clone presented a positive PCR result for eight of the nine genes tested (CYP4H18 PCR was negative). The estimation of the size of this clone indicated that it was above 100 kb and likely to span the rp2 QTL.

BAC clone sequence analysis
A total of 1152 reads were generated from the sequencing of the BAC clone from which 91 were excluded either for poor base quality or contamination with vector sequence. The remaining 1061 reads (844 ± 141 bp) were assembled in six contigs of a final size of 113.59 kb with an average coverage of 6.57. The summary of these sequencing statistics is presented in Supplementary Table S7 and the full sequence is submitted to Genbank (accession number: JQ711183). An analysis of the BAC clone sequence for gene prediction indicated the presence of sixteen open reading frames. Annotation using BlastX revealed that fifteen of these open reading frames correspond to cytochrome P450 genes organised in a cluster as seen in A. gambiae on Chromosome 3R (equivalent to 2L chromosome in A. funestus), while one salivary protein named gSG7-2 in A. gambiae was also identified. All the 15 P450 genes belong to the CYP6 family of cytochrome P450 genes. A direct comparison between this P450 cluster in A. funestus and that observed in A. gambiae reveals first that there is a triplication of the CYP6M1 gene with three different copies in A. funestus compared with just one in A. gambiae. Second, the AgCYP6Z2 gene present in A. gambiae is missing in A. funestus. The AgCYP6Z2 gene is located between AgCYP6Z1 and AgCYP6Z3 in A. gambiae but no open reading frame was detected between these two genes in this rp2 BAC clone. The P450 gene called AgCYP6Z3 in this rp2 clone exhibited a same similarity to both AgCYP6Z3 and AgCYP6Z2 in A. gambiae when compared by ClustalW alignment with 79% similarity for DNA sequence and 85% for the protein sequences. However, when using Blastx (NCBI), it indicates an 86% similarity of amino acid sequence to AgCYP6Z3 in A. gambiae and 85% to AgCYP6Z2. Therefore, it is called AgCYP6Z3 in this study, although it is very close to both genes. Third, only one salivary protein gene was detected in the rp2 clone, while three of these SG7 genes are found in this cluster in A. gambiae.

Analysis of the P450 gene cluster in rp2 BAC clone
A complete genomic organisation of this P450 cluster is presented in Figure 1. The genomic arrangement of the fifteen P450 genes is exactly the same as in A. gambiae. The size of the intergenic spaces is also similar between the two species with the lowest intergenic space in this cluster being 78 bp between AgCYP6S1 and AgCYP6R1 (This is also the lowest in A. gambiae with 71bp), while the longest intergenic space is between AgCYP6M4 and the salivary protein gSG7-2 (24739 bp), which is similar in A. gambiae.

All the 15 P450 genes possess two exons separated by short introns from 58 bp for AgCYP6S1 to 87 bp for AgCYP6M7. These intron sizes are also similar to that of A. gambiae (Supplementary Table S8). The amino acid sequences of the fifteen P450 genes varied from 492 amino acids for AgCYP6Z1 and AgCYP6Z2 to 514 for AgCYP6Y2, similar to A. gambiae with the only difference that AgCYP6Y2 is 520 amino acids long in this species. The highly conserved motifs C-helix WxxR, ExLR and PERF are all present as indicated in the multiple alignment. Similarly, the haem-binding loop (the P450 signature motif WxxxR, ExLR and PERF) are all present as indicated in the multiple alignment. Similarly, the haem-binding loop (the P450 signature motif WxxxR, ExLR and PERF) are all present as indicated in the multiple alignment.
AfcYP6M1a and AfcYP6M1c have 45 variant amino acids (90% similarity) among which 8 replacements between amino acid of opposite physicochemical properties (Supplementary Figure S3). AfcYP6M1b and AfcYP6M1c have 57 variant amino acids (88% similarity) among which 12 replacements between amino acid of opposite physicochemical properties (Supplementary Figure S3).

**Sequencing of cuticular protein genes**

Four CPLCG genes were successfully amplified and sequenced in *A. funestus* samples. As some of these genes have been previously associated with insecticide resistance in *A. gambiae* (Vontas et al., 2005; Djouaka et al., 2008), the aim was to assess the correlation between these genes and rp2 QTL. Detailed information about these genes is presented in Supplementary Table S9. Overall, high similarity is observed between the four CPLCG genes and their orthologues in *A. gambiae* (Supplementary Table S9) with percentage of protein sequence similarity between 67 and 86%. The CPLCG3 protein in *A. funestus*, previously associated with possible reduced penetration resistance mechanism in *A. gambiae* (Vontas et al., 2005; Djouaka et al., 2008), had seven amino acid differences with the *A. gambiae* orthologue.

**SNP identification in rp2 genes**

In total, 565 SNPs from 28 genes screened were identified after analysing 20 431 bp of sequencing traces indicating that one SNP is present every 36 bp. The patterns of polymorphism observed in this study is similar to the patterns previously described in this species from 50 genes (Wondji et al., 2007a). This is the case for the ratio of transition/tranversion substitutions with transition SNPs significantly predominant in the total (60.9%) over transversion substitutions (39.1%) compared with a ratio of 62% vs 38% in the previous study (Wondji et al., 2007a). This bias towards transition substitutions was also more pronounced in the coding region with 62.7% (271 SNPs out of 432) of transition and 37.3% (161 SNPs) of transversion slightly lower (but not significantly) than the 66.3% vs 33.7% seen previously (Wondji et al., 2007a). More transversion SNPs were significantly observed in non-coding regions (47.5%) than in coding region (37.3%) as also observed in other species such as *A. gambiae* (Morlais et al., 2004) and *Aedes aegypti* (Morlais and Severson, 2003), confirming that SNPs occur more frequently as transition in coding regions than in non-coding regions. This difference between coding and non-coding regions is also reflected in the frequency of SNPs, which is one SNP every 39 bp in coding regions vs one SNP every 24 bp in non-coding regions (5’UTR, 3’UTR and intron combined).

A higher frequency of SNPs was observed at the third codon position (71.9%) than at the first or second position (Table 1). More than 2/3 (308) of the 432 coding SNPs were synonymous, whereas <1/3 (124) were non-synonymous or replacement SNPs. Indels were also detected (1–6 bp long) or inferred owing to overlapping peaks in the sequencing traces. Twenty-five triallelic and six tetraallelic SNPs were detected confirming that multiallelic SNPs are present in this mosquito species although at low frequency.

The average nucleotide diversity per gene (π = 0.0105) was higher than the estimate of 0.0072 in the previous study (Wondji et al., 2007a). The average nucleotide diversity in non-coding DNA (0.013) was lower than in synonymous sites of the coding regions (0.028). This is an indication that non-coding regions are under greater purifying selection than synonymous sites within coding regions. The Ka/Ks ratio was 0.11, which is lower than the estimate of 0.18 seen in the previous study.

**Fine-scale linkage and QTL mapping at F6**

*Linkage mapping.* The mapping of rp2 BAC clone SNP markers and the other additional SNPs selected from chromosome 2L indicated that they all genetically map to a single linkage group corresponding to chromosome 2L. This was the case for each family and when they were both combined in a single map (Supplementary Figure S4). The genetic distance between the rp2 BAC clone loci are very low around 0.2–0.8 cM, indicating a very limited recombination rate between these P450 SNPs as expected from markers located in such a small
Carried out identified loci significantly associated (equivalent of 3R in A. gambiae) with pyrethroid resistance in Anopheles funestus. The markers order on the linkage map is different to that seen in A. gambiae but is in accordance with the synteny pattern observed between A. funestus and A. gambiae for chromosome 2L (equivalent of 3R in A. gambiae) as indicated previously (Sharakhov et al., 2004).

Genotype/phenotype correlation for rp2 loci. \( \chi^2 \) goodness-of-fit tests carried out identified loci significantly associated (\( P < 0.05 \)) with pyrethroid resistance in rp2 in both reciprocal families (Figure 2). In family 1, all the seven informative SNPs genotyped from the rp2 clone showed a significant correlation with the phenotype with a higher mortality rate observed in individuals homozygous for the allele of the susceptible parent and the heterozygotes had a trend between the mortality rate and the resistance phenotype. For Family 10, although a similar trend was observed (Figure 2b), no significant difference was recorded. Contrary to Family 1, the range of mortality rates between the three genotypes was not observed for these three loci as individuals with no allele of the susceptible parent and the heterozygotes had a similar mortality rate. No correlation was observed between the two SNPs in the cuticular protein genes and the resistance phenotype.

When analysing other loci of the 2L chromosome away from the BAC clone, a significant correlation was also detected for three other markers, the P450 CYP9M1, the cytochrome c BU10 (orthologue of AGAP009537-RA in A. gambiae) and BU29, a putative sensory appendage protein orthologue of SAP-3 in A. gambiae (AGAP008054-RA). However, the additive effect seen for the rp2 BAC clone genes was not observed for these three loci as individuals with no allele of the susceptible parent and the heterozygotes had a similar mortality rate. No correlation was observed between the two SNPs in the cuticular protein genes and the resistance phenotype.

For Family 10, although a similar trend was observed (Figure 2b), no significant difference was recorded. Contrary to Family 1, the range of mortality rates between the three genotypes was not observed in this family. In average, mortality rates ranged from 42–47% when the individuals have no allele of the susceptible parent, 42–47% for heterozygotes and 58–77% for individuals with two alleles of the susceptible parent indicating an additive effect in the resistance associated with the rp2 QTL as the mortality increases with the number of alleles from the susceptible parent.

When analysing other loci of the 2L chromosome away from the BAC clone, a significant correlation was also detected for three other markers, the P450 CYP9M1, the cytochrome c BU10 (orthologue of AGAP009537-RA in A. gambiae) and BU29, a putative sensory appendage protein orthologue of SAP-3 in A. gambiae (AGAP008054-RA). However, the additive effect seen for the rp2 BAC clone genes was not observed for these three loci as individuals with no allele of the susceptible parent and the heterozygotes had a similar mortality rate. No correlation was observed between the two SNPs in the cuticular protein genes and the resistance phenotype.
**QTL mapping.** The fine-scale mapping of the markers on the 2L chromosome confirms the presence of the rp2 QTL in Family 1 with a small increase of the logarithm of the odd from 4 in the previous chromosome confirms the presence of the QTL mapping QTL in Family 1 using F6 generation for Figure 3.

Figure 3 Plot of logarithm of odd scores for the pyrethroid resistance rp2 QTL in Family 1 using F6 progeny. The y axis indicates logarithm of odd ratio scores, and the x axis indicates chromosome positions. Solid lines represent logarithm of odd estimated by composite-interval mapping and dashed lines represent logarithm of odd estimated by interval mapping. The straight line along the top of the graph represents the threshold value for logarithm of odd as determined by permutations. Names of markers are listed around QTL locations.

Table 2 Multiple-interval mapping estimates of QTL position and associated genetic, environmental, phenotypic, additive and dominance effects associated with pyrethroid resistance in Family 1 A. funestus using F6 progeny

<table>
<thead>
<tr>
<th>Nearest Marker</th>
<th>Genetic marker</th>
<th>QTL distance (cM)</th>
<th>Effect</th>
<th>% ( \sigma_g^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP6P9 rp1</td>
<td>g</td>
<td>13.7</td>
<td>A, –4.77</td>
<td>65.8</td>
</tr>
<tr>
<td>(88.7%)</td>
<td>s</td>
<td>(7.3–14.3)</td>
<td>19.6</td>
<td></td>
</tr>
<tr>
<td>CYP6M4 rp2</td>
<td>p</td>
<td>65.9</td>
<td>A, –0.76</td>
<td>4.1</td>
</tr>
<tr>
<td>(7.3–14.3)</td>
<td>s</td>
<td>(60.5–69.5)</td>
<td>–0.8</td>
<td></td>
</tr>
<tr>
<td>BU92 rp3</td>
<td>e</td>
<td>92</td>
<td>A, –0.10</td>
<td>0.2</td>
</tr>
<tr>
<td>(85.9–100.9)</td>
<td>s</td>
<td>(85.9–100.9)</td>
<td>–0.3</td>
<td></td>
</tr>
</tbody>
</table>

\( \sigma_g^2, \sigma_s^2, \sigma_d^2, \) respectively, for genetic, phenotypic (in parentheses) and environmental variance; A for additive; D for dominance; confidence intervals for QTL position are in parentheses near the position estimate.

(Djouaka et al., 2008) probably have no role in the resistance associated with rp2 QTL. Multiple-interval mapping indicated that rp2 QTL explains about 3.3% of the genetic variance with rp1 still the major QTL explaining 85.4% of the genetic variance (Table 2). The genetic variance (\( \sigma_g^2 \)) explained by the three QTLs now accounts for 88.7% of the phenotypic variance (\( \sigma_p^2 \)), whereas 11.3% are explained by environmental variance (\( \sigma_e^2 \)) (Table 2).

The rp2 QTL was not detected in family 10 in line with the lack of significant genotype/phenotype correlation on chromosome 2L in this family. This absence of rp2 in Family 10 is probably due to the heterozygote profile of the parental female of this family, which originated from the resistant strain FUMOZ-R. The combined genetic map of both families did not change much for rp2 as a similar logarithm of odd was observed.

**Multiplex gene expression profiling of genes around rp2 QTL.** In order to detect the candidate genes associated with the rp2 QTL, the expression profile of the 15 P450 genes from the rp2 BAC clone was assessed in comparison with the reference ribosomal protein S7 gene (RSP7) between the resistant strain FUMOZ-R and the susceptible FANG strain. This multiplex qPCR experiment detected no expression signal for AfCYP6M1b, AfCYP6S1 and AfCYP6S2 either in resistant or in susceptible samples. The gene with the highest expression level was AfCYP6M8 orthologue of AgCYP6M2
gene regularly associated with pyrethroid resistance in *A. gambiae* (Djouaka et al., 2008; Muller et al., 2008; Stevenson et al., 2011). However, this high expression of *AfCYP6M8* is observed in both resistant and susceptible samples indicating that contrary to its *A. gambiae* AgCYP6M2 orthologue, *AfCYP6M8* is not associated with pyrethroid resistance in this FUMOZ-R strain. A significant differential expression (*P* < 0.01; *t*-test) was observed at two genes, *AfCYP6M7* and *AfCYP6Z1*, between the resistant FUMOZ-R and the susceptible FANG, with, respectively, a 2.4- and 2.82-fold-change overexpression in the resistant strain (Figure 4a). No significant difference was observed for the remaining genes, although a 1.7-fold change was also observed for *AfCYP6Z3* but at *P* = 0.065 (*t*-test).

The analysis of six other genes of interest on the 2L chromosome away from the *rp2* BAC clone indicated a significant downexpression of the cuticular protein CPLCG3 (CPLC8) gene in the resistant strain contrary to previous observation made in *A. gambiae* associating this gene with pyrethroid resistance in Djouaka et al., 2008. A significant overexpression of the *GSTe2* gene (*P* = 0.017) was also observed in the resistant FUMOZ-R strain. No informative SNP was identified in this gene and therefore was not included in the QTL mapping to assess its correlation with resistance phenotype. No differential expression was observed for the cytochrome c oxidase subunit I gene (orthologue of the *A. gambiae* AGAP009537 gene), which exhibited a significant genotype/phenotype association.

The expression profile of *AfCYP6Z1*, *AfCYP6M7* and *GSTe2* upregulated in FUMOZ-R through the multiplex GeXP expression profiling was further assessed individually using the qRT–PCR. This was also done for *AfCYP6Z3*, which showed a 1.7-fold change although not significant. All genes tested had PCR efficiency between 90 and 102%. A significant overexpression was observed for all the four genes when their (2^−ΔΔCt) relative expression was compared with that of the two housekeeping genes RSP7 and Actin either individually or when combined (Figure 4b). The highest fold change is observed for *CYP6Z1* with an 11.16-fold change upregulation in FUMOZ-R compared with the susceptible FANG. *AfCYP6Z3* is 6.3-fold overexpressed in FUMOZ-R, while *AfCYP6M7* is also upregulated in this resistant strain at 5.5-fold. *GSTe2* also showed a significant upregulation in FUMOZ-R but at a lower fold change of 2.82.

### Table 3 Summary statistics for polymorphism of *AfCYP6Z1* and *AfCYP6Z3* between the FANG susceptible and FUMOZ-R resistant strains

<table>
<thead>
<tr>
<th>Samples</th>
<th>N</th>
<th>S</th>
<th>Syn</th>
<th>Nonsyn</th>
<th>π (k)</th>
<th>D (Tajima)</th>
<th>D* (Fu and Li)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FANG-S</td>
<td>4</td>
<td>32</td>
<td>19</td>
<td>13</td>
<td>0.0114 (16.8)</td>
<td>−0.37^ns</td>
<td>−0.27^ns</td>
</tr>
<tr>
<td>FUMOZ-R</td>
<td>4</td>
<td>28</td>
<td>18</td>
<td>10</td>
<td>0.0094 (14.0)</td>
<td>−0.86^ns</td>
<td>−0.86^ns</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>49</td>
<td>30</td>
<td>19</td>
<td>0.0115 (17.1)</td>
<td>−0.52^ns</td>
<td>−0.56^ns</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Samples</th>
<th>N</th>
<th>S</th>
<th>Syn</th>
<th>Nonsyn</th>
<th>π (k)</th>
<th>D (Tajima)</th>
<th>D* (Fu and Li)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FANG-S</td>
<td>5</td>
<td>39</td>
<td>33</td>
<td>6</td>
<td>0.0116 (17.2)</td>
<td>−0.61^ns</td>
<td>−0.61^ns</td>
</tr>
<tr>
<td>FUMOZ-R</td>
<td>6</td>
<td>34</td>
<td>29</td>
<td>5</td>
<td>0.0117 (17.3)</td>
<td>1.04^ns</td>
<td>1.34^ns</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>59</td>
<td>50</td>
<td>9</td>
<td>0.015 (22.7)</td>
<td>0.61^ns</td>
<td>0.38^ns</td>
</tr>
</tbody>
</table>

N, number of sequences; ns, non-synonymous mutations; ns, not significant; S, number of polymorphic sites; Syn, Synonymous mutations; Tajima’s D and Fu and Li’s D statistics; π, nucleotide diversity (k = mean number of nucleotide differences).

### Polymorphism analysis of *AfCYP6Z1* and *AfCYP6Z3*

The polymorphism of the *AfCYP6Z1* and *AfCYP6Z3* was analysed in order to identify potential mutations associated with pyrethroid resistance. The full length of the coding region of *AfCYP6Z1* was sequenced for eight clones (four FANG and four FUMOZ-R), whereas 11 clones (five FANG and six FUMOZ-R) were sequenced for *AfCYP6Z3*. The two genes exhibited a high polymorphism with a total of 49 and 59 polymorphic sites observed, respectively, for *AfCYP6Z1* and *AfCYP6Z3*, while 19 amino acid changes were observed for *AfCYP6Z1* (Supplementary Figure S5) and 9 for *AfCYP6Z3* (Supplementary Figure S6). This level of polymorphism is similar between the FANG and the FUMOZ-R strain with, nevertheless, a slightly higher number of mutations in the susceptible strain for the two genes. Other parameters of this polymorphism are summarised in Table 3. A Neighbour-joining tree of the haplotypes of both genes indicated a cluster of haplotypes specific to each strain for both genes (Figure 5) but there are also haplotypes similar between the two strains.
DISCUSSION

The characterisation of mechanisms of insecticide resistance is a key prerequisite for the successful management of vector control programmes. In order to fully elucidate the underlying mechanisms conferring pyrethroid resistance in a resistant strain of the major malaria vector *A. funestus*, we successfully carried out a fine-scale mapping of the *rp2* QTL associated with this resistance.

The successful isolation and sequencing of the 113 kb BAC clone was a significant step in allowing the positional mapping of *rp2* QTL. The main observation from this *rp2* clone sequencing was the identification of another copy number increase of a P450 gene in *A. funestus* compared with *A. gambiae* with three genes related to *AgCYP6M1* detected in *A. funestus*. Two other duplication events were already previously reported for *AfCYP6P9* and *AfCYP6P4* genes in a 120 kb BAC clone spanning *rp1* QTL (Wondji et al., 2009). However, the triplication of *AfCYP6M1* is not associated with the resistance mechanism as this event is ancient as indicated by the high number of amino acid changes and the even lower similarity between the three copies (88 to 90%) than that seen between copies of the two duplicated P450 genes *AfCYP6P9a* and *AfCYP6P9b* in *rp1* (95% similarity) and *AfCYP6P4a* and *AfCYP6P4b* (94% similarity).

Another noticeable difference between *A. funestus* and *A. gambiae* is the absence in this *rp2* BAC clone of the *AgCYP6Z2* gene previously associated with insecticide resistance (Muller et al., 2007) although not a pyrethroid metabolizer (McLaughlin et al., 2008). In *A. gambiae*, the *AgCYP6Z1*, *AgCYP6Z2* and *AgCYP6Z3* genes are found together in a 7-kb-long genomic region on the 3R chromosome. In this *rp2* clone, only *AfCYP6Z1* and *AfCYP6Z3* were detected and both are separated by an intergenic space of 897 bp and no other gene was detected within the next 20 kb indicating that *AgCYP6Z2* orthologue is probably missing in *A. funestus* either through a deletion event in *A. funestus* or through a duplication event in *A. gambiae*. The 92% protein similarity observed between *AgCYP6Z2* and *AgCYP6Z3* in *A. gambiae* indicates that these two genes are probably duplicates of an ancestral gene orthologue to the *AfCYP6Z3* found in *A. funestus*. This is further supported by the fact that the similarity between *AfCYP6Z3* and *AfCYP6Z1* in *A. funestus* is only 68%. It is not impossible that this *AgCYP6Z2* orthologue may still be found in *A. funestus* on the 2L chromosome but beyond the *rp2* BAC clone. However, this is unlikely because of the strong conservation observed in the gene organisation pattern of the P450 clusters in *rp1* and in *rp2* between *A. gambiae* and *A. funestus*.

SNP identification for fine-scale mapping indicated a ratio of one SNP every 36 bp, which is slightly higher than the estimate of one SNP every 41 bp found previously in a diverse set of 50 genes in *A. funestus* (Wondji et al., 2007a). This increase and other differences between the two studies could be due to the higher number of P450 genes sequenced (20 out of the 28) in this study contrary to just 10 out of 50 in the previous one. P450 genes have been shown to be a highly polymorphic gene family in *A. funestus* and *A. gambiae* (Wilding et al., 2009). Overall, the similarity of the polymorphism pattern observed in this study to previous reports indicates.

---

**Figure 5** Schematic representation of haplotypes of *AfCYP6Z3* and *AfCYP6Z1* genes between the susceptible FANG and the resistant FUMOZ-R strains. **a, b** show the nucleotide polymorphic sites for both *AfCYP6Z3* and *AfCYP6Z1*, respectively, while **c, d** are for the amino acid polymorphism. Below the list of sequences, R or S indicates the positions that are polymorphic in the resistant or susceptible mosquitoes, respectively, while an asterisk (*) marks a position polymorphic in both phenotypes. **e** Neighbour-joining tree of *AfCYP6Z3* showing the clades specific to each phenotype while **f** is for *AfCYP6Z1*.
the robustness of the SNP identification process in this study. This is further confirmed by a successful genotyping of a subset of these SNPs in the isofemale lines progeny. These 565 SNPs located in genes of interest such as P450s constitute an additionally genomic tool for this species for which less genomic resources are available compared with *A. gambiae*.

Although the *rp2* QTL spans a broad genomic region, it can be concluded that the P450 cluster from the BAC clone exhibits the strongest association with the resistance phenotype and therefore contains the likely genes involved in the pyrethroid resistance. This is seen through the linear and additive correlation observed between SNPs in the P450 cluster and the resistance phenotype contrary to the partial association seen for the cytochrome c and *AfCYP9M1* genes or no correlation for the cuticular genes. This is further confirmed by the significantly overexpression of *AfCYP6Z1*, *AfCYP6Z3* and *AfCYP6M7* genes in the resistant FUMOZ-R strain after qRT–PCR, while no such overexpression was seen for either cuticular protein genes or the cytochrome c.

The association of *AfCYP6Z1* with pyrethroid resistance is not surprising as its orthologue in *A. gambiae* has previously been associated with pyrethroid (David et al., 2005) and is able to metabolise DDT (Chiu et al., 2008). *AfCYP6M8* was not associated with resistance in this *A. funestus* strain despite having a 92% similarity with its *A. gambiae* orthologue *AgCYP6M2*, a metaboliser of pyrethroids (Stevenson et al., 2011).

The 6.3-fold change observed for *AfCYP6Z3* in FUMOZ-R indicates that this gene is probably also associated with the pyrethroid resistance explained by the *rp2* QTL. This will not be surprising as *AgCYP6Z3* have also been linked with insecticide resistance in *A. gambiae* (Muller et al., 2007).

The overexpression of *AfCYP6M7* observed in this study confirmed a recent report that this gene was overexpressed in this FUMOZ-R strain after microarray (Christian et al., 2011). The orthologue of *AfCYP6M7* in *A. gambiae*, *AgCYP6M3*, is not commonly associated with insecticide resistance. It was found to be marginally overexpressed in a field permethrin *A. gambiae* population by microarray but not confirmed by qPCR (Muller et al., 2008).

The absence of upregulation of the cuticular protein gene in FUMOZ-R correlates with the lack of significant correlation between genotypes of markers at these genes and phenotype. This is an indication that this cluster of cuticular protein genes on 2L is not associated with the cuticle thickening observed recently in resistant mosquitoes of this FUMOZ-R strain (Wood et al., 2010).

An overexpression of *GSte2* also was observed in this study, although at a lower fold change of 2.82 compared with the three P450 genes. Unfortunately, no informative SNP was observed in this gene for both family 1 and 10, which prevented its mapping and to assess its association with *rp2* QTL. However, when taking in consideration the synteny between *A. gambiae* and *A. funestus* chromosomal map, *GSte2* appears to be located within the boundaries of *rp2*. *GSte2* has been validated as DDT metabolizer in various mosquito species (Ding et al., 2005; Lumjuan et al., 2005). Because the FUMOZ-R strain is fully susceptible to DDT, the overexpression of *GSte2* in this study is probably associated with pyrethroid resistance as previously observed in *A. gambiae* (David et al., 2005) and in other insects, such as *Nilaparvata lugens* (Vontas et al., 2001). As GSTs are known to conjugate insecticides or their metabolites (phase II), it is possible that *GSte2* might be involved in the conjugation of permethrin metabolites.

The allelic variation of *AfCYP6Z1* and *AfCYP6Z3* between the susceptible and resistant was assessed to establish a possible correlation with the resistance phenotype. Such correlation has been previously established for the CYP6AB3 P450 gene in the insect *Depressaria pastinacella* (parsnip webworm). Five amino acid changes observed in this gene enhance the metabolism of plant allelochemicals by altering a proximal surface residue and potential interactions with cytochrome P450 reductase (Mao et al., 2007). Although the polymorphisms observed between the two FANG and FUMOZ-R strains for each gene could reflect their different genetic background, it could also be a consequence of a correlation with the resistance phenotype. This is supported by the location and possible impact on catalytic properties of some amino acid changes between the two strains observed in the two genes. For example, in *AfCYP6Z1*, three amino acid changes exclusively present in the resistant FUMOZ-R strain could impact the function of the gene. These are an Y175C which is the first residue in the highly conserved E helix; the F265I located within the highly variable H helix and the F416L located within the conserved PERF region. These changes could affect the catalytic activity of *AfCYP6Z1* in FUMOZ-R. For *AfCYP6Z3*, several of the amino acid predominantly found in FUMOZ-R could also enhance the catalytic activity of this gene in association with resistance to pyrethroid. Among these amino acid substitutions are: an A12V and H27Y both located within the hydrophobic region that anchors the protein to the membrane; a S219P mutation within the hydrophobic region located between F and G helices, thought to penetrate lipid bilayer, making contact with environment from which many substrates can enter the active site; a E239G located within the SRS-3 and may enhance catalysis; and a L471F, the first residue in the SRS-6, which may enhance catalysis of *AfCYP6Z3* in FUMOZ-R. Further analysis of such mutations will confirm their importance and role in conferring pyrethroid resistance in the FUMOZ-R strain. Because mutations associated with resistance are also likely to be found in the regulatory regions of these over-transcribed genes such as the upstream region and UTRs, future work will also investigate these regions. Additionally, the sequencing of other P450 genes in this *rp2* QTL should also be envisaged in order to identify non-synonymous mutations, which, although not necessarily associated with an increase in transcription of the gene, could still modify the conformation of the enzyme and its substrate specificity (higher affinity for the chemical in the resistant strain).

The positional cloning of *rp2* QTL in this study is a significant contribution to the full elucidation of mechanisms of pyrethroid resistance in this important malaria vector. Although *rp2* is not the major QTL in the FUMOZ-R, it is possible that its relative importance may vary in field populations as it may have a bigger role in other field populations of this species in Africa as seen in *A. gambiae* for which the main genes conferring metabolic resistance to pyrethroid varied from one region to another. Therefore, this elucidation is significant and will help to improve the understanding of resistance mechanisms in this species. Further work is needed to confirm the role had by each of the overexpressed P450 genes and *GSte2* (such as their ability to metabolise or interact with pyrethroids) and to see whether they are also associated with the resistance in field populations of *A. funestus* across Africa.

**DATA ARCHIVING**

The sequence sets containing all the SNPs have been deposited at Genbank (accession number: JX431304-JX431487). Genotype data for genetic and QTL mapping, qRT–PCR and DNA sequences have been deposited at Dryad: doi:10.5061/dryad.v218g.
CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
This work was supported by a Wellcome Trust Research Career Development Fellowship to CSW. We thank Professor Maureen Coetzee for earlier work done on crossing the FUMOZ and FANG strains. We also thank Professor Janet Hemingway and Professor Hilary Ramsey for initial contribution in this work. JMR was supported by a fellowship from the Fundacion Ramon Areces.

Author contributions: HI, JMR and CSW carried out the experiments; SI contributed to the data analysis; NFL contributed to the crossing the FUMOZ and FANG strains. We also thank Professor Maureen Coetzee for earlier work done on crossing the FUMOZ and FANG strains. We also thank Professor Janet Hemingway and Professor Hilary Ramsey for initial contribution in this work. JMR was supported by a fellowship from the Fundacion Ramon Areces.

This work was supported by a Wellcome Trust Research Career Development Fellowship to CSW. We thank Professor Janet Hemingway and Professor Hilary Ramsey for initial contribution in this work. JMR was supported by a fellowship from the Fundacion Ramon Areces. We thank Professor Maureen Coetzee for earlier work done on crossing the FUMOZ and FANG strains. We also thank Professor Janet Hemingway and Professor Hilary Ramsey for initial contribution in this work. JMR was supported by a fellowship from the Fundacion Ramon Areces.