PlasmoView: A Web-based Resource to Visualise Global *Plasmodium falciparum* Genomic Variation

Mark D. Preston,[1] Samuel A. Assefa,[1] Harold Ocholla,[2,3] Colin J. Sutherland,[1] Steffen Borrmann,[4,5] Alexis Nzila,[6,7] Pascal Michon,[8] Tran Tinh Hien,[9] Teun Bousema,[1] Christopher J. Drakeley,[1] Issaka Zongo,[10] Jean-Bosco Ouédraogo,[11] Abdoulaye A. Djimde,[12,13] Ogobara K. Doumbo,[14] Francois Nosten,[15,16,17] Rick M. Fairhurst,[18] David J. Conway,[1] Cally Roper,[1] and Taane G. Clark[1]

[1]Department of Pathogen Molecular Biology, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom; [2]Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre 3, Malawi and [3]Liverpool School of Tropical Medicine, Liverpool L3 5QA, United Kingdom; [4]KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya and [5]Department of Infectious Diseases, Heidelberg University School of Medicine, Heidelberg 69120, Germany; [6]KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya and [7]King Fahd University of Petroleum and Minerals, PO Box 468, Dhahran 31262, Kingdom of Saudi Arabia; [8]Papua New Guinea Institute of Medical Research, PO Box 483, Madang, Papua New Guinea; [9]Oxford University Clinical Research Unit, Wellcome Trust Major Overseas Program, Hospital for Tropical Diseases, District 5, Ho Chi Minh City, Vietnam; [10]Institut de Recherche en Sciences de la Sant, Bobo–Dioulasso, Burkina Faso; [11]Institut de Recherche en Sciences de la Sant, BP 545, Bobo-Dioulasso 01, Burkina Faso; [12]Malaria Research and Training Centre, Faculty of Medicine, Pharmacy and Dentistry, University of Bamako, Bamako, Mali; and [13]Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; [14]Malaria Research and Training Centre, Faculty of Medicine, Pharmacy and Dentistry, University of Bamako, Bamako, Mali; [15]Mahidol-Oxford Tropical Medicine Research Unit, Bangkok 10400, Thailand; [16]Centre for Tropical Medicine, University of Oxford, Oxford OX3 7LJ, United Kingdom and [17]Shoklo Malaria Research Unit, Mae Sot TAK 63110, Thailand and [18]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Malaria is a global public health challenge, with drug resistance a major barrier to disease control and elimination. To meet the urgent need for better treatments and vaccines, a deeper knowledge of *Plasmodium* biology and malaria epidemiology is required. An improved understanding of the genomic variation of malaria parasites, especially the most virulent *Plasmodium falciparum* (*Pf*) species, has the potential to yield new insights in these areas. High-throughput sequencing and genotyping is generating large amounts of genomic data across multiple parasite populations. The resulting ability to identify informative variants, particularly single-nucleotide polymorphisms (SNPs), will lead to the discovery of intra- and inter-population differences and thus enable the development of genetic barcodes for diagnostic assays and clinical studies. Knowledge of genetic variability underlying drug resistance and other differential phenotypes will also facilitate the identification of novel mutations and contribute to surveillance and stratified medicine applications. The *PlasmoView* interactive web-browsing tool enables the research community to visualise genomic variation and annotation (eg, biological function) in a geographic setting. The first release contains over 600 000 high-quality SNPs in 631 *Pf* isolates from laboratory strains and four malaria-endemic regions (West Africa, East Africa, Southeast Asia and Oceania).

Keywords. *Plasmodium falciparum*; malaria; genomics; drug resistance; vaccine targets; visualization.

Malaria parasites cause disease in approximately 650 million people and *Plasmodium falciparum* (*Pf*) in

particular kills up to 1 million people each year [1]. Anti-malarial drug resistance is a major public health problem that hinders disease control and elimination efforts [2]. *Pf* parasites from almost all malaria-endemic countries show modest levels of drug resistance, especially to chloroquine [3]. Recent evidence indicates that *Pf* parasites in Cambodia and Thailand are developing resistance to artemisinin – currently the most effective anti-malaria intervention [4–7]. An improved understanding of *Pf* genetics has provided new insights into the molecular mechanisms of drug resistance [8–10] and may ultimately lead to new treatments and reduce the global disease burden [11].

High-throughput sequencing technologies and large-scale genotyping chips are generating genome-wide *Pf* data on an unprecedented scale. This means that it is now possible to densely map genomic variation (eg, single-nucleotide polymorphisms, SNPs) and assess global diversity. Knowledge of the diversity of variants across populations will enable the biological interrogation of novel mutations and identification of candidate vaccines. Other potential applications include SNP-based assays to barcode parasites over time and space for epidemiological, diagnostic and clinical studies. However there are several roadblocks to the translation of genomic variation into useful public health tools. These include difficulties in obtaining robust phenotypic data (eg, drug resistance and mosquito transmission) on samples and the lack of web-based informatics tools to access robust genomic data, which are needed to launch further experiments and translational activities.

Whilst the *plasmodDB* [12] and *genedb* [13] web-browsers provide genomic annotation and support the investigation of SNPs in individual parasite strains, there is a need for additional information (eg, allele frequencies and population statistics) across the growing collection of publicly available sequences and genotypes from clinical *Pf* isolates. We are making available to the malaria community the *PlasmoView* web-browser (http://pathogenseq.lshtm.ac.uk/plasmoview) to facilitate the investigation of genome-wide polymorphisms in parasites from different malaria-endemic regions.

RESULTS

SNP Data

We aligned raw sequence and genotyping data (to the 3D7 reference genome, version 3, 80.9% AT, 23 Mb, 14 chromosomes) from 971 *Pf* samples from four regions: West Africa (WAF, Burkina Faso, Gambia, Ghana, Mali and Senegal), East Africa (EAF, Kenya and Malawi), Southeast Asia (SEA, Cambodia, Thailand and Vietnam) and Oceania (OCE, Papua New Guinea, Supplementary Table S1). After quality control procedures, 631 samples (8 laboratory/reference, 367 WAF, 88 EAF, 151 SEA and 17 OCE) were retained. The alignments and high coverage (>10-fold) enabled the identification of 593 579 (2.6% of all nucleotides) high-quality SNPs across the nuclear genome (51% exonic, 33% non-synonymous, Supplementary Figure 1A). These SNPs include the 86 158 SNPs identified in 227 samples [14]. Seventy-three per cent (434 290) of SNPs are rare (minor allele frequency, MAF < 1%) and 4% (29 036) are common (MAF > 10%, see Supplementary Figure 1B).

The linkage disequilibrium between non-rare markers reveals decay with physical genetic distance (Supplementary Figure 1C). This decay is higher in the African regions (WAF and EAF) than in SEA or OCE. As expected, a principal component analysis reveals that the SNP data differentiate the

samples by geographic region (Supplementary Figure 1D) and population (not shown). To identify the polymorphisms driving the population differentiation, we applied a SNP-wise F_{ST} approach where values between 0 and 1, higher values imply more differentiation [15]. The application across regions identifies 467 SNPs with $F_{ST} > 0.1$ indicating a high level of regional differentiation. As expected, the F_{ST} values of intra-region comparisons are substantially lower (% SNPs with $F_{ST} > 0.1$: WAF 0.2%, EAF 1.6%, SEA 6.6% and OCE n/a) than inter-region comparisons (% SNPs with $F_{ST} > 0.1$: 27.9%) in common SNPs (MAF > 15%).

PlasmoView

The web-based *PlasmoView* tool presents genome-wide variation and geographical information on *Pf*. The implementation contains a first release of over 370 million data points (approximately 600 k SNPs in over 600 samples described above). *PlasmoView* provides real-time visualisation and summary statistics and is a timely tool for the high-level interrogation of large genomic datasets. *Pf* data are presented in two views: (a) the matrix view provides a colour-coded indication of mutation locations ordered by genetic and geographical location; and (b) the map view shows a global view of SNP prevalence by country. The matrix view includes graphical and real-time textual SNP-by-SNP genomic annotation (reference alleles, subtelomeric regions, gene regions, amino acid changes and genomic uniqueness) as well as MAF and F_{ST} graphs plotted for the selected data. Both matrix and map views are interactive enabling researchers to easily navigate the data via the menu at the top of the application or using mouse buttons. The *PlasmoView* launch page includes a facility to search by common gene name (eg, *RH5* or *CRT*), previous id (*PFD1145c* or *MAL7P1.27*) and the latest gene nomenclature (*PF3D7_0424100* or *PF3D7_0709000*). The tool is scalable for the significant increase in data that is anticipated over the next few years. In the next sections we demonstrate the utility of *PlasmoView* using specific loci in the *Pf* genome.

Drug resistance in *Pf* is commonly due to SNPs in genes involved in the biological pathways that antimalarials target. The matrix view in *PlasmoView* is ideally suited to display all observed mutations on a gene-by-gene basis. Figure 1 shows the mutations for each sample in *PfCRT* (*PF3D7_0709000*), a gene on chromosome 7 associated with chloroquine resistance [16]. Sixty-one SNPs (18 non-synonymous) have been detected in this gene, 5 of which are known to be involved in drug resistance. The high regional differentiation is measured by the SNP-wise F_{ST} (maximum 0.63, see blue histogram track in Figure 1). In *PlasmoView* the geographic spread of any mutation is further visualised in the map view, which displays SNP frequencies by geographical location. For example, Figure 2 shows the frequency of mutations observed across samples from each country at chromosome 7, position 403 615 (the
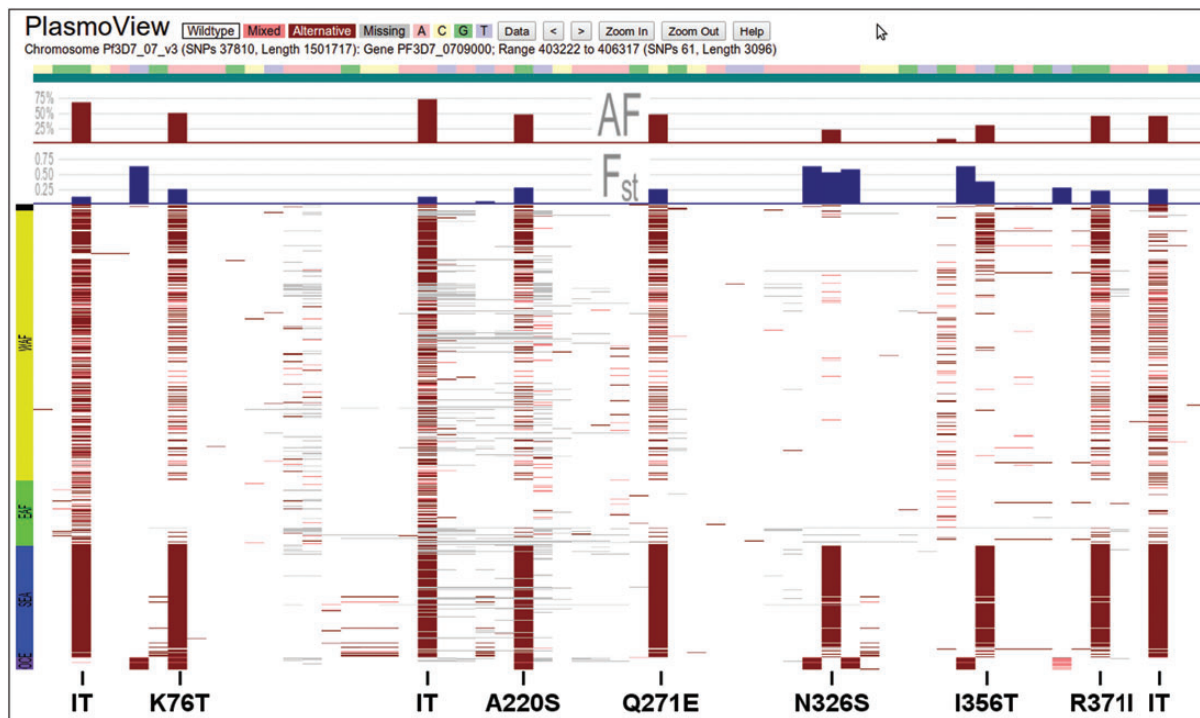
Figure 1. Global mutations in the *Pf* chloroquine resistance transporter gene (*PfCRT*). The 61 SNPs (18 non-synonymous) identified in *PfCRT* (Chromosome 7, *PF3D7_0709000*) are shown. Some markers (eg, Q271E) show evidence that they have gone to fixation in SEA (blue on the left axis) and are common throughout the rest of the world (WAF: green; EAF: yellow; OCE: purple on the left axis). Six known drug resistance markers are shown, as well as 3 common SNPs located in intronic regions (IT).
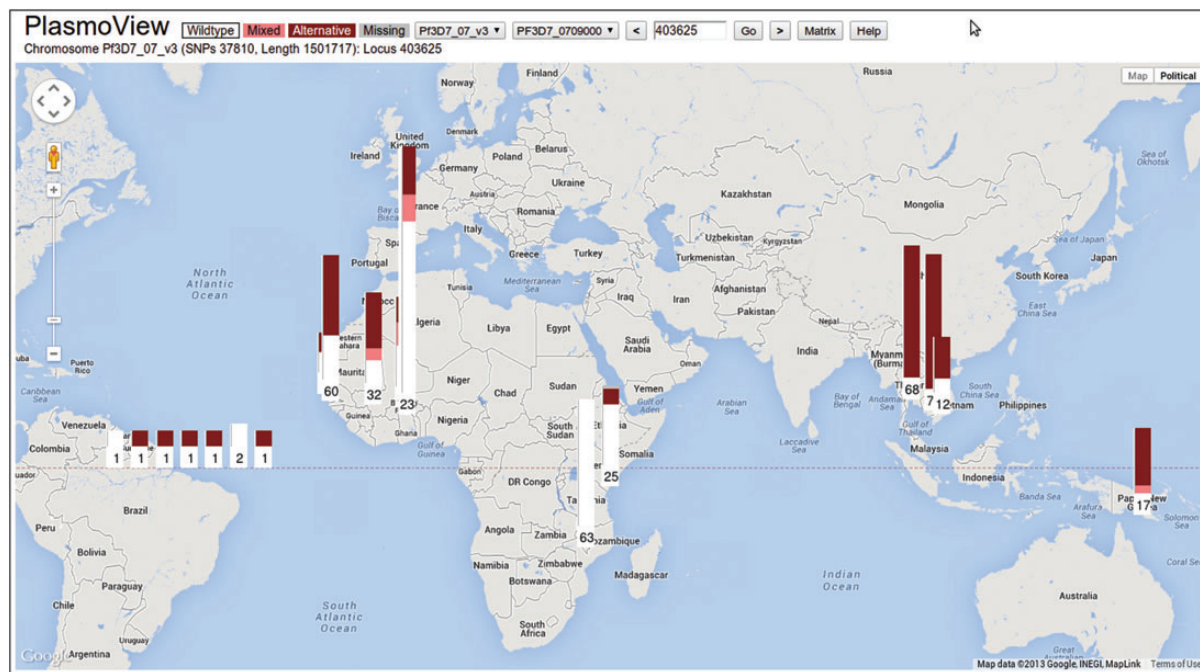
Figure 2. Global distribution of the resistance-conferring *PfCRT* mutation K76T. The global prevalence of the K76T mutation (Chromosome 7, position 403 615) can be seen with the mutation fixed or close to fixed in most countries. Parasites may have reverted back to the wildtype allele, due to the complete withdrawal of chloroquine from some countries (eg, in East Africa, yellow on the left axis). The LAB *Pf* strains (laboratory-adapted and imported *Pf* strains) are located over the South Atlantic. Information on each bar chart is available by holding the cursor over it.

PfCRT K76T amino acid mutation, [17, 18]). *PlasmoView* enables the easy identification of mutation sites within each gene, estimate their allele frequencies and investigate their geographical context, via the matrix view, the AF and F_{ST} graphs and the map view, respectively. The matrix views for three other genes (*PfDHFR*, *PfDHPS* and *PfMDR1*) involved in drug resistance are presented in Supplementary Figures 2A–C and also provided as part of the set of pre-loaded examples of interest for the web-browser.

Whilst a number of SNPs within drug resistance loci have been well characterised, other more recent and important regions may require investigation. For example, a major region of chromosome 13 (1.73 Mb to 1.82 Mb) in the *Pf* genome was identified to be undergoing a selective sweep due to artemisinin resistance in Southeast Asia [8]. It is possible to visualise the SNP variations in this region for all 631 global parasite isolates and laboratory-adapted parasite strains. The SNP density is very high in this region (1355 SNPs in 50 kb, 1.77 Mb to 1.82 Mb, Figure 3), so we use the SNP-by-SNP information to direct further investigation. There are interesting SNPs progressing to fixation (high MAF and low F_{ST}) and those showing large geographical differences (high F_{ST}). For example, the SNP at position 1 793 121 in *Pf3D7_1344700* (Supplementary Figures 2D and E, MAF 17%, F_{ST} 0.38) has been identified as having the highest association with parasite clearance rate [[8], denoted SNP 1]. *PlasmoView* shows 46 SNPs in this region whose F_{ST} is greater than 0.1 including 14 non-synonymous SNPs and one synonymous in *PF3D7_1343800* (22 785 bp, 591 SNPs) and four non-synonymous SNPs and one synonymous SNP in *PF3D7_1344300* (4768 bp, 132 SNPs).

The lack of an effective licensed vaccine remains one of the most significant gaps in the arsenal to control and eliminate *Pf* malaria [19]. The *Pf* Reticulocyte Binding Protein Homologue 5 (encoded by the *PfRH5* gene, *PF3D7_0424100*) is considered a promising candidate antigen, as it seems to be essential for the invasion of multiple laboratory-adapted *Pf* lines and clinical *Pf* isolates into red blood cells [19]. High genetic variation in target genes can pose difficulties in the design of a vaccine that covers the full range of diversity across populations. The diversity observed for *PfRH5* across 227 clinical Pf isolates was low, with only 5 non-synonymous SNPs (out of 12) reaching a frequency greater than 10% in at least 1 population [19]. Using *PlasmoView* to analyze more than double the number of samples and 5 additional populations, we identify 22 non-synonymous SNPs (out of 28) and confirm the locus has low diversity (Figure 4) with only 5 SNPs exhibiting a frequency greater than 10%.

Detecting balancing selection is one method to identify signatures of acquired immunity and therefore potential targets for vaccines. As immunity to the commonest alleles rises in malaria-endemic areas, parasites expressing rarer alleles experience a selective advantage. This process maintains a balance of alleles in the population, with neither the common alleles moving to fixation nor the rare alleles moving to extinction. When multiple alleles are maintained within populations and none of them achieves fixation, balancing selection forces are believed to be present. In *PlasmoView*, loci under balancing selection are readily be visualised as SNPs with intermediate MAF (10% to 40%) and low F_{ST} (indicating little population differentiation). These signatures are shown in vaccine targets previously identified using methods to detect balancing selection, including the *MSP3.8* (merozoite surface protein 3.8, *PF3D7_1036300*, Supplementary Figure 2F) and *AMA1* (apical membrane antigen 1, *PF3D7_1133400*) genes [11, 20, 21]. The malaria vaccine FMP2.1/AS02$_A$ is a recombinant protein (FMP2.1) based on *AMA1* and has been tested in clinical trials [22], see Figure 5.

The *Pf* mitochondrial genome (*Pf_M76611*, 6 kb, 3 genes (*Cox3, Cox1, CytB*), GC content 31.6%) is uniparentally inherited and does not undergo recombination. Therefore mitochondrial DNA (*mt*) sequence divergence and variation data have been used to study the evolutionary history and migration of *Pf* [23–25], particularly gene flow out of Africa. Analyses of the global patterns of *mt* sequence variation have revealed geographic differentiation [24–26]. *PlasmoView* shows 85 SNPs, including the three SNPs with MAF greater than 5% and F_{ST} in excess of 0.05 (*mt772, mt1692 and mt1776*, see Figure 6), which have previously been utilised in diversity studies [23, 24, 26]. We do not observe a mutation at amino acid position 268 in *CytB* (cytochrome B, *cob, MAL_MITO_3:4293*) that had been correlated to *atovaquone* resistance following treatment failure [25].

DISCUSSION

The continued public health burden of malaria and emergence of drug resistance requires the development of new treatments, vaccines and control measures. These efforts are likely to benefit from a deeper understanding of *Pf* biology and malaria epidemiology, in part deriving from analysis of parasite genetic variation. New genomic approaches, using massively parallelisable sequencing and genotyping technologies, are generating vast amounts of *Pf* genetic data. However, to fully achieve their scientific potential, including the initiation of further experiments and translational activities, web-based informatics tools are needed to help researchers to access this genomic information. The *PlasmoView* web-browser condenses and summarises SNP information derived from these technologies into a shared and interpretable visual form. The first release contains nearly 600 k SNPs in over 600 samples that have gone through the same high-quality control procedures as previous studies [14, 27, 28], with additional validation using clonal samples that underwent both sequencing and genotyping. Genome annotation and measures of quality control (eg, uniqueness) incorporated
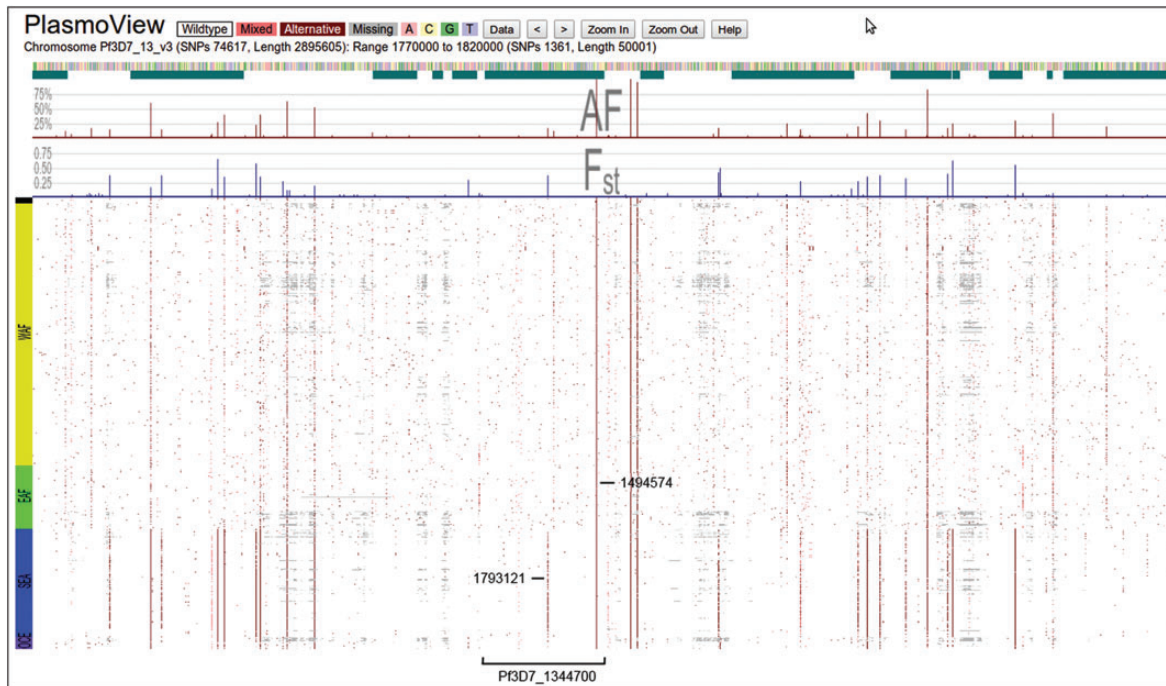
Figure 3. *Pf* chromosome 13, 1770 kb to 1820 kb. Cheeseman et al [8] described this region as being potentially important to the evolution of artemisinin-resistant *Pf* parasites in Southeast Asia. Of the 1355 SNPs in this 50-kb region, more than 20 exhibit high regional specificity (predominantly in SEA with high F_{ST} (>0.1) and low MAF (<30%)) and at least 10 may have spread or are currently spreading to other regions (high MAF and low F_{ST}). Gene *Pf3D7_1344700* (see Supplementary Figure 2*D*) contains an example of both: a SNP with high regional specificity (position 1 793 121, labelled as SNP 1 in [8], Supplementary Figure 2*E*) and one moving towards global fixation (position 1 794 574).
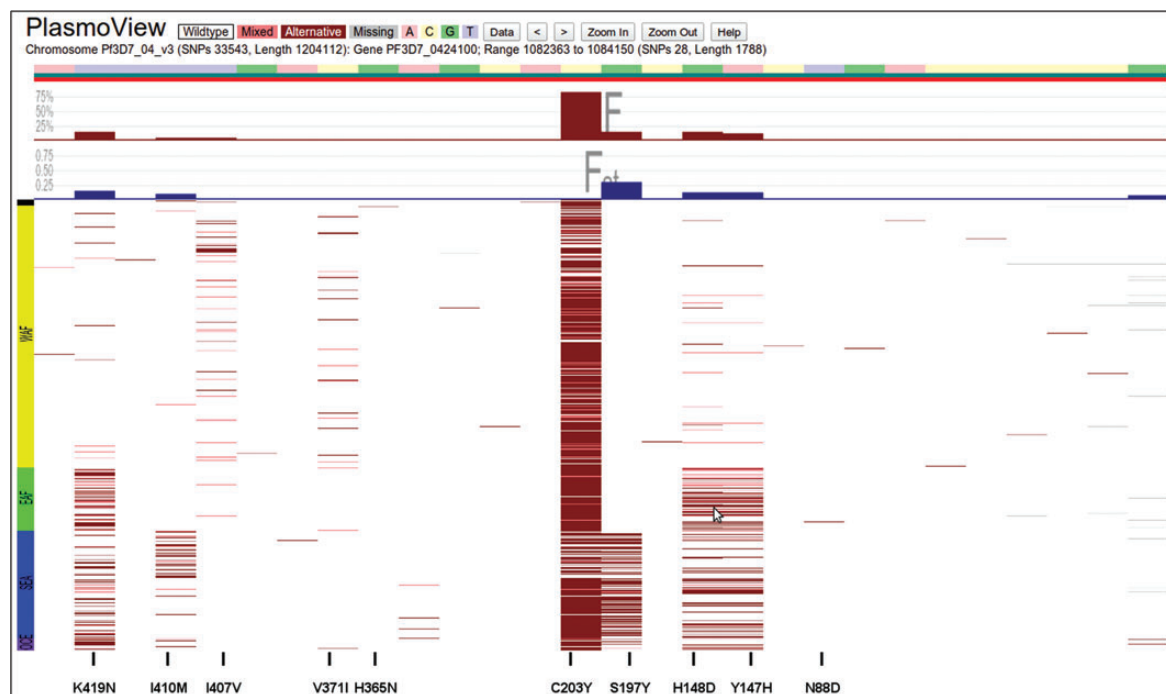
Figure 4. *PfRH5* vaccine candidate in chromosome 4. The *Pf* Reticulocyte Binding Protein Homologue 5 (*PfRH5*, *PF3D7_0424100*) gene is considered a promising vaccine candidate, as it seems to be essential for blood-stage parasite invasion of red blood cells [20]. The locus contains 28 SNPs, including 5 non-synonymous mutations with MAF >10% (I410M, S197Y, H148D, Y147H [19] and K419N [20]).
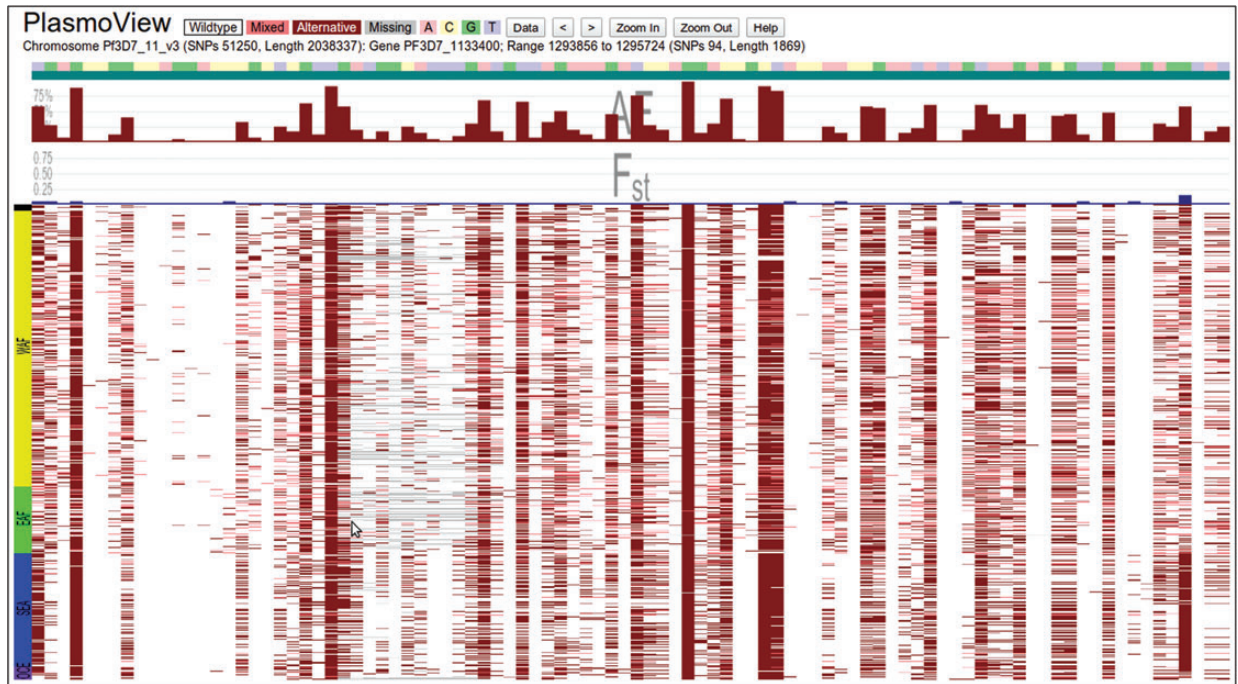
Figure 5. *AMA1* vaccine candidate in chromosome 11. The apical membrane antigen 1 gene (*AMA1*, Chromosome 11, *PF3D7_1133400*) has long been recognised as a vaccine candidate and is currently being evaluated in clinical trials [22]. The high number of SNPs with intermediate MAF and low F_{ST} indicate that this locus is under balancing selection.
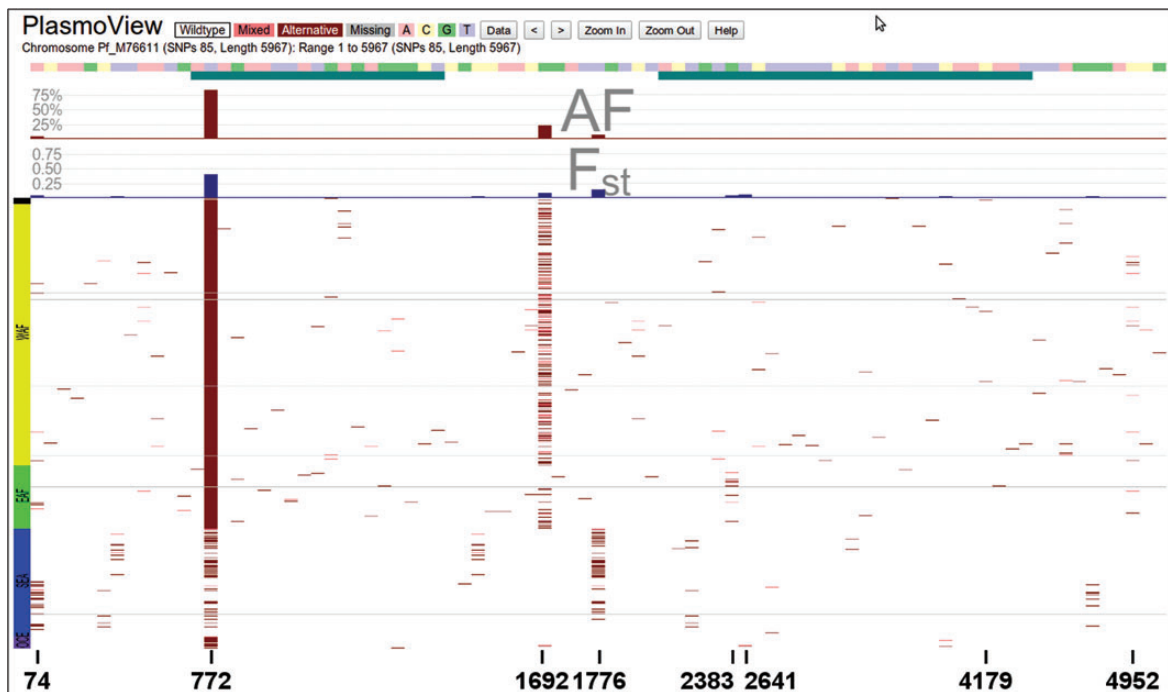
Figure 6. Mitochondrial genome. The *Pf* mitochondrion is a small, uniparentally inherited organelle and its DNA (6 kb) is used for investigating *Pf* evolution and migration [23, 24, 26]. Six SNPs exhibit population differentiation ($F_{ST} > 0.05$; *mt74, mt772, mt1692, mt1776, mt2383* and *mt2641*) including three common alleles (MAF > 5%; *mt772, mt1692* and *mt1776*) used to support an African origin for the species [24, 26].

into the tool allow a further level of assessment of variant quality. The ultimate validation is likely to come from independent studies using whole-genome sequencing technologies. When these data are placed into the public domain with appropriate meta-data (eg, location of samples) then these raw sequence data will be processed using our proven pipeline and included in *PlasmoView*.

The case studies above demonstrate the utility of *PlasmoView* to support informed decision-making processes, whether they take place in the clinic, laboratory or public policy arena. The genomic variation identified in drug resistance (eg, *PfCRT*, *PfDHFR* and *PfMDR1*) and vaccine candidate (eg, *PfRH5*, *MSP3.8* and *AMA1*) genes will help to define the potential repertoire of polymorphisms for follow-up experiments. Similarly, the identification of polymorphisms that drive genetic differentiation at the continental, regional and village level (eg, in drug resistance and *surfin* genes) will facilitate the barcoding of parasites for use in surveillance applications. *PlasmoView* has the functionality to display the variation with annotations, geographical distribution and frequency data and population differentiation metrics in real time.

Whilst *PlasmoView* is presented as a tool to visualise and summarise *Pf* variation, it will be extended to other Plasmodium species as the data becomes available. The next phase of the work is to characterise and display variation other than SNPs, leveraging off the high-coverage, paired-end nature of sequence data [29] and results from genotyping approaches (eg, comparative genomic hybridization [30]). Functionality will expand to include additional population genetic statistics (eg, *Tajima's D* [31]).

In summary, *PlasmoView* is a powerful, scalable tool for the interactive geographic visualisation of *Pf* mutations. Using a high-quality set of polymorphisms, this study shows that *PlasmoView* is useful in confirming existing results, identifying potential avenues for further research and presenting complex genetic data to a broad audience.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature **2005**; 434:214–7.
2. World Health Organization. Global malaria control and elimination: report of a meeting on containment of artemisinin tolerance. Geneva WHO Report **2008**.
3. Payne D. Spread of chloroquine resistance in Plasmodium falciparum. Parasitol Today **1987**; 3:241–6.
4. Amaratunga C, Sreng S, Suon S, et al. Artemisinin-resistant Plasmodium falciparum in Pursat province, western Cambodia: a parasite clearance rate study. Lancet Infect Dis **2012**; 12:851–8.
5. Dondorp AM, Nosten F, Yi P, et al. Artemisinin Resistance in Plasmodium falciparum Malaria. N Engl J Med **2009**; 361:455–67.
6. Noedl H, Se Y, Schaecher K, Smith BL, Socheat D, Fukuda MM. Evidence of Artemisinin-Resistant Malaria in Western Cambodia. N Engl J Med **2008**; 359:2619–20.
7. Phyo AP, Nkhoma S, Stepniewska K, et al. Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. Lancet **2012**; 379:1960–6.
8. Cheeseman IH, Miller BA, Nair S, et al. A Major Genome Region Underlying Artemisinin Resistance in Malaria. Science **2012**; 336:79–82.
9. Roper C, Pearce R, Nair S, Sharp B, Nosten F, Anderson T. Intercontinental Spread of Pyrimethamine-Resistant Malaria. Science **2004**; 305:1124–4.
10. Takala-Harrison S, Clark TG, Jacob CG, et al. Genetic loci associated with delayed clearance of Plasmodium falciparum following artemisinin treatment in Southeast Asia. Proc Natl Acad Sci **2012**; 110:240–5.
11. Amambua-Ngwa A, Tetteh KKA, Manske M, et al. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS Genet **2012**; 8:e1002992.
12. PlasmoDB. The Plasmodium genome resource. Available at: http://plasmodb.org/plasmo/. Accessed 15 January 2013.
13. Homepage - GeneDB. Available at: http://www.genedb.org/Homepage. Accessed 15 January 2013.
14. Manske M, Miotto O, Campino S, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature **2012**; 487:375–9.
15. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat Rev Genet **2009**; 10:639–50.
16. Wootton JC, Feng X, Ferdig MT, et al. Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature **2002**; 418: 320–3.
17. Fidock DA, Nomura T, Talley AK, et al. Mutations in the P. falciparum Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance. Mol Cell **2000**; 6:861–71.
18. Sidhu ABS, Verdier-Pinard D, Fidock DA. Chloroquine Resistance in Plasmodium falciparum Malaria Parasites Conferred by pfcrt Mutations. Science **2002**; 298:210–3.
19. Bustamante LY, Bartholdson SJ, Crosnier C, et al. A full-length recombinant Plasmodium falciparum PfRH5 protein induces inhibitory antibodies that are effective across common PfRH5 genetic variants. Vaccine **2013**; 31:373–9.
20. Borrmann S, Mwai L, Abdi A, et al. Genome-wide screen identifies new candidate genes associated with artemisinin susceptibility in Plasmodium falciparum in Kenya. Nat Sci Reports **2013**; 3.
21. Ochola H, Mipando M, Jensen AT, et al. Whole-genome sequencing of Malawi Plasmodium falciparum paediatric isolates reveal selective pressure on putative drug and vaccine genes and distrinct selection differences from West Africa and Asia. (in preparation).
22. Thera MA, Doumbo OK, Coulibaly D, et al. Safety and Immunogenicity of an AMA-1 Malaria Vaccine in Malian Adults: Results of a Phase 1 Randomized Controlled Trial. PLoS One **2008**; 3:e1465.

23. Conway DJ, Fanello C, Lloyd JM, et al. Origin of Plasmodium falciparum malaria is traced by mitochondrial DNA. Mol Biochem Parasitol **2000**; 111:163–71.

24. Joy DA, Feng X, Mu J, et al. Early Origin and Recent Expansion of Plasmodium falciparum. Science **2003**; 300:318–21.

25. Musset L, Le Bras J, Clain J. Parallel evolution of adaptive mutations in Plasmodium falciparum mitochondrial DNA during atovaquone-proguanil treatment. Mol Biol Evol **2007**; 24:1582–5.

26. Tanabe K, Jombart T, Horibe S, et al. Plasmodium falciparum mitochondrial genetic diversity exhibits isolation-by-distance patterns supporting a sub-Saharan African origin. Mitochondrion **2013**; 13:630–6.

27. Miotto O, Almagro-Garcia J, Manske M, et al. Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet **2013**; 45:648–55.

28. Robinson T, Campino SG, Auburn S, et al. Drug-Resistant Genotypes and Multi-Clonality in Plasmodium falciparum Analysed by Direct Genome Sequencing from Peripheral Blood of Malaria Patients. Dimopoulos G, editor. PLoS One **2011**; 6:e23204.

29. Sepúlveda N, Campino SG, Assefa SA, Sutherland CJ, Pain A, Clark TG. A Poisson hierarchical modelling approach to detecting copy number variation in sequence coverage data. BMC Genomics **2013**; 14:1–12.

30. Almagro-Garcia J, Manske M, Carret C, et al. SnoopCGH: software for visualizing comparative genomic hybridization data. Bioinformatics **2009**; 25:2732–3.

31. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **1989**; 123:585–95.

32. Home | www.malariagen.net. Available at: http:// www.malariagen.net/. Accessed 15 January 2013.

33. Malaria programme - Wellcome Trust Sanger Institute. Available at: http://www.sanger.ac.uk/research/projects/malariaprogramme-kwiatkowski/. Accessed 15 January 2013.

34. SMALT - Wellcome Trust Sanger Institute. Available at: http://www.sanger.ac.uk/resources/software/smalt/. Accessed 15 January 2013.

35. SAMtools. Available at: http://samtools.sourceforge.net/. Accessed 15 January 2013.

36. VCFtools. Available at: http://vcftools.sourceforge.net/. Accessed 15 January 2013.

37. Preston MD, Clark TG. GGV: An open resource for visualising Global Genomic Variation.

38. Auburn S, Campino S, Miotto O, et al. Characterization of Within-Host Plasmodium falciparum Diversity Using Next-Generation Sequence Data. PLoS One **2012**; 7:e32891.

39. Preston MD, Manske M, Horner N, et al. VarB: a variation browsing and analysis tool for variants derived from next-generation sequencing data. Bioinformatics **2012**; 28:2983–5.

40. Park DJ, Lukens AK, Neafsey DE, et al. Sequence-based association and selection scans identify drug resistance loci in the Plasmodium falciparum malaria parasite. Proc Natl Acad Sci **2012**; 109:13052–7.