


RESEARCH

Open Access



Plasmodium falciparum genetic variation of *var2csa* in the Democratic Republic of the Congo

Robert Verity^{1*} , Nicholas J. Hathaway^{2,3}, Andreea Waltmann⁴, Stephanie M. Doctor⁵, Oliver J. Watson¹, Jaymin C. Patel⁵, Kashamuka Mwandagilirwa⁶, Antoinette K. Tshetu⁷, Jeffrey A. Bailey^{2,3}, Azra C. Ghani¹, Jonathan J. Juliano^{5,8,9} and Steven R. Meshnick⁵

Abstract

Background: The Democratic Republic of the Congo (DRC) bears a high burden of malaria, which is exacerbated in pregnant women. The VAR2CSA protein plays a crucial role in pregnancy-associated malaria (PAM), and hence quantifying diversity at the *var2csa* locus in the DRC is important in understanding the basic epidemiology of PAM, and in developing a robust vaccine against PAM.

Methods: Samples were taken from the 2013–14 Demographic and Health Survey conducted in the DRC, focusing on children under 5 years of age. A short subregion of the *var2csa* gene was sequenced in 115 spatial clusters, giving country-wide estimates of sequence polymorphism and spatial population structure.

Results: Results indicate that *var2csa* is highly polymorphic, and that diversity is being maintained through balancing selection, however, there is no clear signal of phylogenetic or geographic structure to this diversity. Linear modelling demonstrates that the number of *var2csa* variants in a cluster correlates directly with cluster prevalence, but not with other epidemiological factors such as urbanicity.

Conclusions: Results suggest that the DRC fits within the global pattern of high *var2csa* diversity and little genetic differentiation between regions. A broad multivalent VAR2CSA vaccine candidate could benefit from targeting stable regions and common variants to address the substantial genetic diversity.

Background

Pregnancy-associated malaria (PAM) is a major public health concern. In areas of stable malaria transmission in sub-Saharan Africa approximately one in four pregnant women have evidence of malaria infection at time of delivery [1, 2]. PAM is detrimental to both mother and child, placing the mother at increased risk of severe anaemia whilst increasing the chance of adverse birth outcomes, including stillbirth, preterm birth and low birth weight (LBW) [1–3].

The adverse effects of PAM are mediated by the sequestration of infected erythrocytes in the placental microvasculature through binding of VAR2CSA—a large and genetically diverse parasite protein expressed during pregnancy—to human chondroitin sulfate A (CSA) [4, 5]. Naturally occurring anti-VAR2CSA antibodies provide partial protection against future episodes of PAM, such that primigravid women are most susceptible and risk of severe infection and LBW decreases in subsequent pregnancies [4, 6, 7]. Vaccines against VAR2CSA are currently undergoing initial trials [8–10].

Extensive effort has gone into characterizing the particular sub-region of the VAR2CSA protein responsible for binding CSA and inducing a protective immune response. These efforts have led to the identification of ID1-DBL2X; a 1.6 kb segment encoding the minimal

*Correspondence: r.verity@imperial.ac.uk

¹ Medical Research Council Centre for Outbreak Analysis & Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

Full list of author information is available at the end of the article

binding epitope [11]. ID1-DBL2X has been shown to raise antibodies that abrogate the adhesion of infected erythrocytes to CSA with the same efficacy and specificity as the full-length extra-cellular protein, while maintaining high cross-reactivity to multiple parasite lines [12]. Both leading PAM vaccine candidates, PlacMalVac and PRIMALVAC, use recombinant proteins that target overlapping constructs of this region [9, 10].

The ability of any such vaccine to have a sustained impact on malaria will depend on the extent of antigenic variation in the vaccinated population, which in turn depends on the level of sequence polymorphism in the *var2csa* gene. Studies into global diversity at the *var2csa* locus have identified extremely high sequence polymorphism, with evidence that diversity is being maintained by balancing selection [13–16]. Furthermore, this high level of diversity occurs against the backdrop of a major dimorphic split in the N-terminal segment of the VAR2CSA protein, leading to multiple sequence clusters, each of which has been found to associate with a different level of parasitaemia [16] and a different risk of poor birth outcomes [17]. Sequence polymorphism at the *var2csa* locus is, therefore, both functionally relevant in vaccine design and clinically relevant in understanding the basic epidemiology of PAM.

The Democratic Republic of Congo (DRC) bears one of the highest malaria burdens in sub-Saharan Africa, with over 1 million *Plasmodium falciparum* affected pregnancies each year [18]. Transmission is stable throughout the country, and prevalence is estimated at 34.1% on average by PCR [19]. The majority of studies into the genetics of *P. falciparum* in DRC have focussed on issues of drug resistance (see Additional file 1). Studies into *dhps* mutations, which confer resistance to sulfadoxine, have found distinct geographic clustering of the most highly resistant haplotypes in the east of the country [20, 21]. In contrast, countrywide studies into neutral genetic variation [22] and variation in the *pfama1* gene [23] have found little signal of population structure or isolation by distance even over large geographic scales. To date, no study has explored the geographic and genetic structure of *var2csa* in DRC, and so it is unknown what challenges may lie ahead in terms of vaccine design.

This study focused on quantifying genetic variation at the *var2csa* locus in samples obtained from the 2013–14 Demographic and Health Survey (DHS); a large, cross-sectional study separated into spatial clusters spanning the DRC. Central aims of this study were: (1) to explore *var2csa* diversity in these samples in the context of global diversity at this locus, (2) to quantify the level of spatial structure in the DRC, and (3) to determine what epidemiological factors (if any) are predictors of observed levels of diversity at this locus. These questions will be

important for any future interventions aimed at reducing PAM in the DRC.

Methods

Sample collection

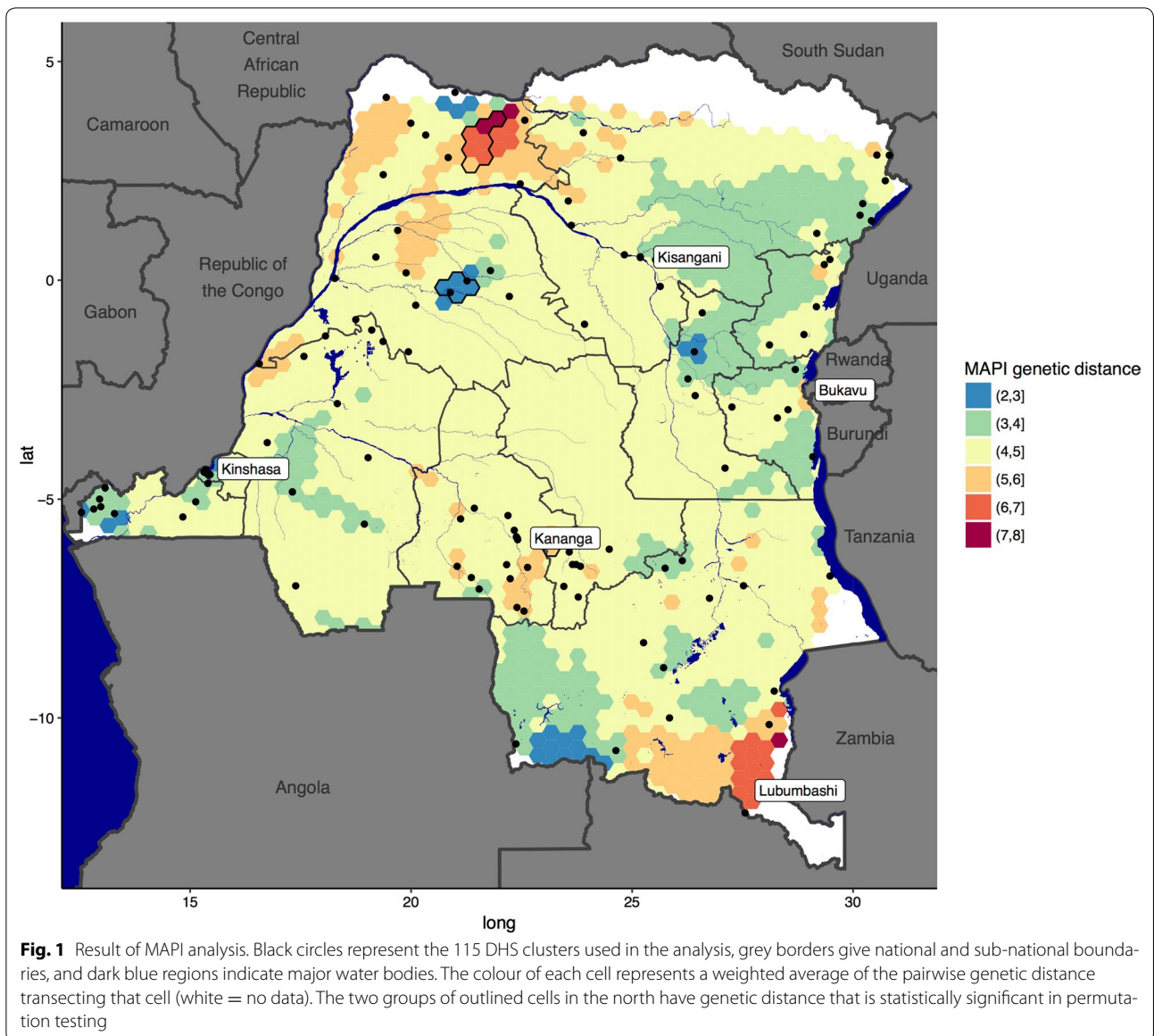
Samples from children under 5 years of age were collected as part of the 2013–2014 DHS study [19]. Heel- or finger-prick blood from each child was used to prepare smears for identification of malaria parasites via light microscopy, and to prepare dried blood spots (DBS) using previously described methods [24]. DBS were transported to the University of North Carolina, where genomic DNA (gDNA) was extracted using previously described methods [25]. Each gDNA sample was tested in duplicate in a duplex quantitative real-time PCR assay targeting the *P. falciparum* lactate dehydrogenase gene (*pfldh*) [26] and, as a control, the human β -tubulin gene [27]. The full DHS study was conducted in 540 clusters at the neighbourhood or village level, and the current study was performed on a subset of these clusters. A total of 115 clusters were selected that contained at least three parasitaemic children (i.e. smear and *pfldh*-positive), and that were spatially representative of the whole country (Fig. 1). For each of the 115 clusters, all the gDNA samples were pooled in equimolar volumes to ensure equal contribution to the pool of each infected child.

DNA amplification and sequencing

A 400 bp region of the binding domain of VAR2CSA was amplified by primers listed in Table 1. Samples from the 115 clusters were pooled and PCR amplified in technical duplicates. The forward primer included replicate specific molecular identifiers (MIDs) which allowed amplicon pooling prior to library preparation using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, United States). The Illumina primers included Illumina's dual barcoding system. This approach enabled multiplexing of pooled samples and libraries into one 2×300 sequencing run on the Illumina MySeq platform at the University of North Carolina High Throughput Sequencing Facility.

Haplotype determination

The paired-end sequences were stitched by FLASH [28], and haplotype determination was performed using the program SeekDeep [29] using a quality cut-off of $Q30 > 0.75$. In brief, sequence reads were first demultiplexed according to MIDs into amplicon-specific data, resulting in two independent amplicon reads sets for each of the 115 pools, after which MIDs were trimmed and clustered. Only predicted haplotypes that appeared in both replicates for each pool and occurred at a frequency of $\geq 0.5\%$ were utilized to determine the most likely

**Table 1 Primers used**

Forward primer	Reverse primer
ATCATGGTGAACACGAACA	GTACCCGCTTACGGTTTCG

haplotypes within each pool. SeekDeep also removed any haplotypes that may have resulted from chimerization during PCR. This pipeline resulted in the identification of 583 unique haplotypes throughout the 115 clusters.

Phylogenetic analysis

The similarity between sequences in this study and previously identified *var2csa* variants was established by

conducting a BLAST search of all 583 sequences individually against the NCBI database. 81% of top BLAST hits came from just two published collections, which were downloaded and incorporated, bringing an additional 84 sequences from Benin [16] and 36 sequences from Senegal [30, 31]. The corresponding region of the 3D7 genome was also downloaded from PlasmoDB [32], leading to an expanded data set of 704 sequences. These additional sequences were used to improve sequence alignment, and to root DRC sequences relative to wider African variation.

Initial attempts at multiple sequence alignment failed due to difficulties in aligning several highly variable regions containing large numbers of insertions and

deletions, and so an iterative approach to alignment was used. First, all pairwise alignments were produced using the *msa* package in the R programming language [33, 34]. Sequences were then grouped into 11 groups using agglomerative hierarchical clustering, such that sequences with a high proportion of matched bases clustered together. Multiple alignment was then carried out on each group independently using the ClustalW algorithm in MEGA7 [35], and any obvious alignment errors were removed by hand. Aligned groups were then combined using profile alignment in ClustalX [36]. Regions with large numbers of insertions or deletions were identified by counting the proportion of gaps at each locus, and subsequent analyses were restricted to four contiguous regions with zero gaps to ensure that results were not influenced by ambiguous alignment in a few highly variable regions (Fig. 2).

A neighbour-joining tree was constructed in MEGA7 using default substitution parameters, and bootstrapped with 100 replicates. Focusing on the 583 sequences from DRC hereafter, Tajima's D statistic was calculated [37] and a Z-test for selection was carried out based on relative rates of synonymous and non-synonymous variation in MEGA7 [38].

Spatial and epidemiological analysis

Sequence variants were grouped into the 115 DHS clusters from which they derived. The number of clusters occupied by a variant was correlated against the top BLAST hit percentage for that variant, to establish whether variants that were common in DRC also tended to be similar to previously identified sequences. Heterozygosity within and between clusters was calculated, and nearness to fixation was measured via G_{ST} [39]. Isolation by distance was explored by regressing Nei's genetic

distance [40] against geographic distance, and more subtle signals of spatial patterning—for example barriers to gene flow—were explored using the program MAPI [41]. A matrix of pairwise genetic distances between DHS clusters was analysed using MAPI to produce a smoothed surface indicating spatial connectivity between regions, coupled with a permutation test to identify regions of significantly high or low connectivity. MAPI was run using default parameters ($\beta = 0.25$, eccentricity = 0.975, error circle = 0.1), and statistical significance was assessed using 1000 permutations at a threshold of $\alpha = 0.05$.

Epidemiological analyses focused on finding covariates that explained a significant proportion of the observed genetic variation through generalized linear modelling (GLM). Predictor variables included prevalence of a cluster (defined as the proportion of PCR-positive individuals), sample size, and median cycle threshold (C_t) value; C_t values represent the number of PCR cycles needed before a signal is detected in the DNA amplification stage, and therefore C_t is roughly inversely proportional to the amount of nucleic acid in the sample. Results of the DHS questionnaire for each cluster were also downloaded from the DHS website [42], and both “province” and “urbanicity” of the cluster (categorized as “large city”, “small town” and “countryside”) were included as potential predictors. The response variable was the allelic richness, defined as the number of unique haplotypes in a given cluster. A range of GLMs were considered, starting with a full model containing all the predictors listed above and all second-order interaction terms (5 + 25 terms in total). Backwards stepwise selection was then used to identify the model with the lowest Akaike information criterion (AIC). This process was repeated for both Poisson and negative binomial error structures. A separate analysis explored the relationship between C_t

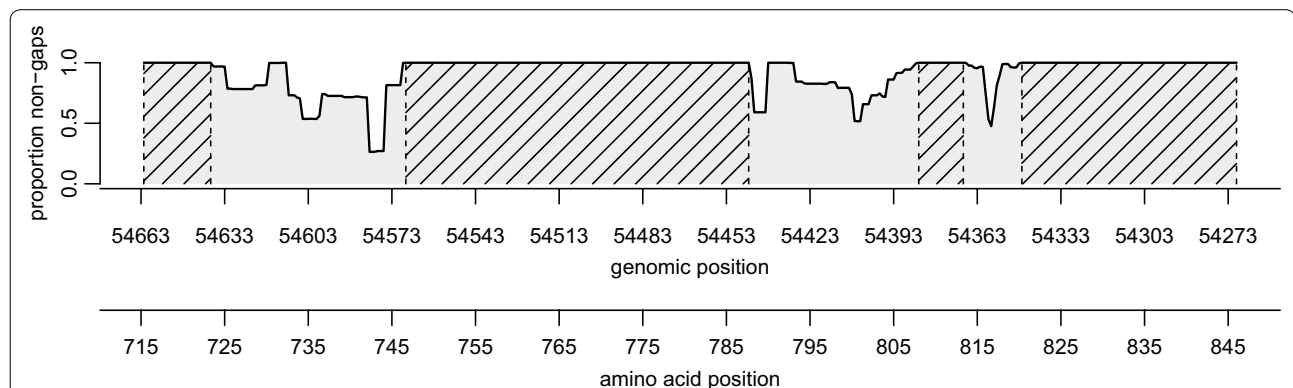


Fig. 2 Schematic of *var2c5a* sequence variation. The proportion of non-gap sequences at each locus (i.e. one minus the proportion of gaps) is plotted for all nucleotide positions relative to *Pf3D7* chromosome 12 and corresponding amino-acid positions. The four shaded regions have zero gaps, and were used in all analyses after the sequence alignment step. Our sequences mapped to 3D7 positions 54,271:54,663 and amino acid positions 715:845

value and prevalence using both a simple linear model and a mixed-effects model with cluster modelled as a random effect. Mixed effects models do not generate p values for individual variables, and so 95% confidence intervals for fitted values were obtained by simulation using the *confint()* function in the R package *lme4* [43].

Results

Phylogenetic analysis

Sequences ranged from 365 to 506 bp (median 413), and had an elevated average GC concentration of 36% compared with the genome-wide average of around 19% [44]. BLAST results revealed that 39 of the 583 sequences (7%) were perfect matches to previously identified *var2csa* accessions in the NCBI database, and the remaining 544 sequences had an average similarity of 87% to their top BLAST hit. After combining our sequences with collections from Benin and Senegal downloaded from NCBI, and carrying out multiple sequence alignment as detailed above, three highly variable regions were clearly visible amid four regions with zero gaps (Fig. 2). Subsequent analyses focused exclusively on these zero-gap regions, trading off the additional information contained in the pattern of insertions and deletions between sequences with increased reliability of any detected polymorphisms. The neighbour-joining tree revealed very little phylogenetic signal (see Additional file 2), and corresponding bootstrap values revealed little confidence in the tree, with only 11% of clades having > 80% confidence. Tajima's D was estimated at 1.95, indicating that the allele frequencies of variants are more evenly distributed than expected (fewer variants at high frequency). The Z-test for selection revealed an excess of synonymous mutations ($dN-dS = -1.445$), although this was non-significant at the 5% threshold ($p = 0.151$).

Spatial and epidemiological analyses

After grouping sequence variants into the 115 DHS clusters from which they originated, there was a median 7 variants per cluster (range 1–28). Variants that were present in many clusters also tended to have high BLAST scores (see Additional file 3), and this correlation was significant ($\rho = 0.176$, $p < 0.001$). The most common variant in the DRC was a 100% match to a variant previously identified from Senegal [31]. Heterozygosity was high both within DHS clusters (mean = 0.62, $sd = 0.21$) and between clusters (mean = 0.81, $sd = 0.07$), and pairwise G_{ST} was also high (mean = 0.25, $sd = 0.13$). However, regression of genetic distance against geographic distance revealed no simple signal of isolation by distance (see Additional file 4), and the relationship was non-significant ($p = 0.217$). Analysis in MAPI also showed relatively uniform connectivity between clusters spanning

the DRC, with no major barriers to gene flow and just one small area of significantly high genetic distance (i.e. low connectivity) in the Nord-Ubangi province, and one area of significantly low genetic distance (i.e. high connectivity) in the Tshuapa province (Fig. 1).

In the analysis of allelic richness, GLMs with negative binomial error structures consistently yielded better model fits (lower AIC values) compared with the Poisson model, suggesting that allelic richness is over-dispersed within clusters. The best-fitting GLM included just 3 out of a potential 30 predictors, including linear and quadratic terms for prevalence and a linear term for sample size (Table 2). Neither median C_t value per cluster nor the demographic parameters “province” and “urbanicity” were identified as important predictors of observed genetic variation. Overall the best-fitting GLM was a good fit to the data (Fig. 3, see also Additional file 5). In the separate analysis of C_t values, the mixed effects model gave a better fit (AIC = 4346) than the simple linear model (AIC = 4389), and C_t value was found to have a negative relationship with prevalence [coefficient = -2.603 , CI ($-4.359, -0.846$)], suggesting

Table 2 Parameters of best-fitting GLM (Negative-binomial model: AIC = 634.8, compared with Poisson model: AIC = 662.1)

Predictor	Estimate	Std. error	Z value	p value
Intercept	-3.682	2.643	-1.393	0.164
Prevalence	27.063	8.905	3.039	0.002**
Sample size	0.232	0.093	2.485	0.013*
Prevalence ²	-18.229	8.694	-2.097	0.036*

Significance codes: ** < 0.01, * < 0.1

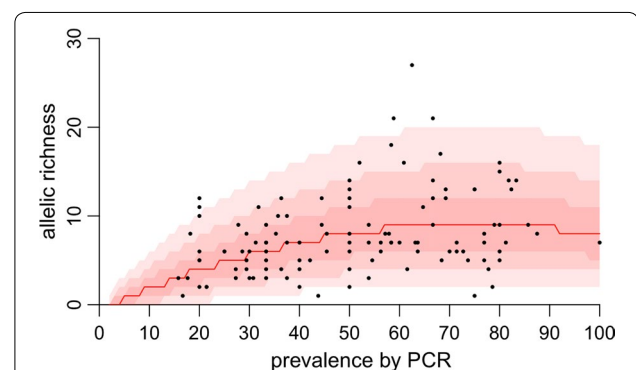


Fig. 3 Observed and model-predicted relationship between prevalence and allelic richness. Black circles represent data for each of the 115 clusters. Red shaded regions indicate the 95, 80 and 50% predictive intervals of the best-fitting model, and the red line represents the median prediction. The best fitting model also includes a sample size term, and so these predictions were generated assuming a sample size of 14 (the median sample size in the data)

that higher quantities of DNA tend to occur in samples obtained from high prevalence clusters.

Discussion

In recent years, a great deal of effort has gone into studying the VAR2CSA protein, both in terms of its immune profile and its impact on clinical outcomes. However, while good data exists on levels of *var2csa* sequence polymorphism in several countries spanning four continents [15], to date no study has explored variation at this locus in samples obtained from DRC. More generally, understanding of the basic epidemiology and population genetics of this large and diverse country has been limited historically by a lack of good quality data. This study aimed to address this issue by quantifying *var2csa* variation in DRC and relating it back to basic epidemiological questions.

The results of this study demonstrate that *var2csa* is highly diverse within DRC, with 583 sequence variants identified among 812 children. High Tajima's D and an excess of synonymous mutations suggest that this high diversity is being maintained by long-term balancing selection, as would be expected of a gene involved in antigenic variation, and as found in previous VAR2CSA studies [13–16]. However, there was no clear signal of phylogenetic structure to this diversity. Low bootstrap values in the neighbour joining tree indicate that different loci provide contradictory information about the position of sequences in the phylogeny, suggesting that recombination is acting to intertwine the phylogenetic branches and break down patterns across loci (see Additional file 2). In the spatial analysis, there was high heterozygosity within and between clusters but no signal of increasing genetic distance with geographic distance or other barriers to gene flow. Together these results indicate that the combined effects of rapid diversification through balancing selection, gene flow between spatial clusters, and recombination within the gene are acting to break down any clear signal of geographic population structure in this particular sub-region of *var2csa*. This contrasts with the result of Doritchamou et al. [16], who found clear signal of population structure in the N-terminal subregion just a few hundred bp upstream of the target region used here, in samples taken from Benin. However, when the Doritchamou et al. [16] sequences are truncated to the same target region used here, there is no clear signal of population structure (results not shown). Therefore, it appears that the extent of *var2csa* phylogenetic signal varies greatly even within narrow sub-regions of this highly diverse gene. It is not clear from this study

what impact this diversity has on acquired immunity, however the excess of synonymous mutations suggests some role in diversification of the VAR2CSA antigen.

Results of general linear modelling demonstrate that *var2csa* diversity is directly related to epidemiological factors. The best-fitting model found that the allelic richness of a cluster tends to increase with prevalence, eventually plateauing out at high prevalence. Crucially, this analysis takes sampling effort into account—that is, prevalence is an important predictor of allelic richness even after accounting for the absolute number of infected children in the analysis. This result is in line with a wider body of evidence showing that genetic diversity tends to be higher in areas of intense transmission, perhaps due to increased opportunity for recombination [45–47]. In terms of vaccine design, this may indicate that a vaccine would be more likely to succeed in areas of low transmission where populations tend to be more clonal. The finding that common alleles in DRC also tend to be represented in other African countries (see Additional file 3) is also encouraging for vaccine development, as it indicates that some variants are common over large geographic scales, despite the general trend of strong local diversification.

One strength of this study is the use of samples derived from the DHS, which is both cross-sectional and nationally representative. This ensures that asymptomatic and low-density (i.e. sub-microscopic) infections are captured in the analysis, which may include a different subset of strains to those found in clinically ill individuals [48]. A caveat is that samples were extracted from children under 5 years of age, and not from infected pregnant women, and so the pool of strains sampled here may not perfectly reflect those implicated in clinical disease. The use of pooled samples also makes it possible to explore a wide geographic area, and to answer questions about spatial connectivity (Fig. 1). However, the use of pooled samples also limits analysis to making observations at the cluster level, and cannot reliably determine factors such as the complexity of infection of each individual within a cluster.

The results of this study highlight the extreme evolutionary pressures acting at the *var2csa* locus to promote antigenic variability in the DRC. This variability creates a substantial hurdle in the development of VAR2CSA-based vaccine, which is extenuated by the highly-connected population structure found within the DRC. A broad multivalent VAR2CSA vaccine candidate could thus benefit from targeting stable regions and common variants to address the substantial genetic diversity [49].

Additional files

Additional file 1. Details of systematic literature search in PubMed and Web of Science, including full search terms and details of all studies identified.

Additional file 2. Neighbour-joining tree of samples from DRC, Benin and Senegal, produced in Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>). Edges are coloured from red to blue according to their bootstrap percentage (black edges are terminal and so have no bootstrap value). Dotted lines leading away from the tree are coloured to indicate the origin of the sample.

Additional file 3. The number of clusters in which a variant was found plotted against percentage identity of top BLAST hit for that variant. Multiple variants have the same combination of BLAST percentage and cluster representation, so size of circles indicates the number of variants for a given combination.

Additional file 4. Regression of Nei's (1972) genetic distance against great circle distance. The fitted relationship is non-significant ($\rho = -0.04$, $p = 0.217$).

Additional file 5. Best-fitting GLM compared to data. Black circles represent the median model prediction for each of the 115 clusters, and vertical bars represent the 95% predictive interval. Predictions are presented relative to the observed allelic richness, meaning the model is a good fit wherever the interval crosses the dashed zero line.

Authors' contributions

SRM conceptualized the study. KM and AKT provided field support in obtaining samples. SMD, NJH, AW and JCP carried out lab work and bioinformatics. RV and OJW performed phylogenetic and statistical analyses. JAB, JJJ, ACG and SRM provided supervision and coordination. All authors contributed to the final manuscript. All authors read and approved the final manuscript.

Author details

¹ Medical Research Council Centre for Outbreak Analysis & Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, UK. ² Program in Bioinformatics and Integrative Biology, University of Massachusetts, Worcester, MA, USA. ³ Division of Transfusion Medicine, Department of Medicine, University of Massachusetts, Worcester, MA, USA. ⁴ Institute for Global Health and Infectious Diseases, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, USA. ⁵ Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, USA. ⁶ Kinshasa School of Public Health, Hôpital Général Provincial de Référence de Kinshasa, Kinshasa, Democratic Republic of Congo. ⁷ Community Health, Kinshasa School of Public Health, School of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of Congo. ⁸ Division of Infectious Diseases, University of North Carolina at Chapel Hill, 130 Mason Farm Road, Chapel Hill 27599, USA. ⁹ Curriculum in Genetics and Microbiology, University of North Carolina at Chapel Hill, 321 South Columbia Street, Chapel Hill, NC 27516, USA.

Acknowledgements

We would like to thank Olivia Anderson and Joris L. Likwela for their help and support with this project. Also, sincere thanks to Sujata Balasubramanian for technical advice, and Jonathan Parr for DRC and DHS data knowledge. Finally, we are grateful to the many thousands of researchers and participants in the DHS study, without whom this work would not be possible.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All sequences are publicly available at NCBI GenBank, Accession Numbers MG029761–MG030343.

Ethics approval and consent to participate

All survey respondents provided verbal informed consent for the collection of blood spots. Consent procedures, survey administration, and blood sample collection were approved by the ICF Institutional Review Board (IRB) and the Ethics Committee of the Kinshasa School of Public Health. De-identified testing for malaria parasites was approved by the Institutional Review Board of the University of North Carolina.

Funding

RV is funded by a Skills Development Fellowship (MR/N01507X/1): this award is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. JJJ received support from National Institute of Health (NIH) (R01AI121558, R21AI121465). SRM received support from NIH (5R01AI107949, R56 AI106129 01A1). OJW is a Wellcome Trust funded Ph.D. Student (109312/Z/15/Z). JAB and NJH received funding through NIH Grants (5 R01 AI099473).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 22 November 2017 Accepted: 17 January 2018

Published online: 24 January 2018

References

- Steketee RW, Nahlen BL, Parise ME, Menendez C. The burden of malaria in pregnancy in malaria-endemic areas. *Am J Trop Med Hyg*. 2001;64:28–35.
- Desai M, ter Kuile FO, Nosten F, McGready R, Asamo K, Brabin B, et al. Epidemiology and burden of malaria in pregnancy. *Lancet Infect Dis*. 2007;7:93–104.
- Van Geertruyden JP, Thomas F, Erhart A, D'Alessandro U. The contribution of malaria in pregnancy to perinatal mortality. *Am J Trop Med Hyg*. 2004;71(2_suppl):35–40.
- Salanti A, Dahlbäck M, Turner L, Nielsen MA, Barfod L, Magistrado P, et al. Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J Exp Med*. 2004;200:1197–203.
- Pereira MA, Clausen TM, Pehrson C, Mao Y, Resende M, Daugaard M, et al. Placental sequestration of *Plasmodium falciparum* malaria parasites is mediated by the interaction between VAR2CSA and chondroitin sulfate A on syndecan-1. *PLoS Pathog*. 2016;12:e1005831.
- Brabin BJ. An analysis of malaria in pregnancy in Africa. *Bull World Health Organ*. 1983;61:1005–16.
- Eisele TP, Larsen DA, Anglewicz PA, Keating J, Yukich J, Bennett A, et al. Malaria prevention in pregnancy, birthweight, and neonatal mortality: a meta-analysis of 32 national cross-sectional datasets in Africa. *Lancet Infect Dis*. 2012;12:942–9.
- Chène A, Houard S, Nielsen MA, Hundt S, D'Alessio F, Sirima SB, et al. Clinical development of placental malaria vaccines and immunoassays harmonization: a workshop report. *Malar J*. 2016;15:476.
- European Vaccine Initiative. <http://www.euvaccine.eu/portfolio/project-index/placmalvac>. Accessed 8 Jan 2018.
- European Vaccine Initiative. <http://www.euvaccine.eu/portfolio/project-index/primalvac>. Accessed 8 Jan 2018.
- Clausen TM, Christoffersen S, Dahlbäck M, Langkilde AE, Jensen KE, Resende M, et al. Structural and functional insight into how the *Plasmodium falciparum* VAR2CSA protein mediates binding to chondroitin sulfate A in placental malaria. *J Biol Chem*. 2012;287:23332–45.
- Bordbar B, Tuikue-Ndam N, Bigey P, Doritchamou J, Scherman D, Deloron P. Identification of Id1-DBL2X of VAR2CSA as a key domain inducing highly inhibitory and cross-reactive antibodies. *Vaccine*. 2012;30:1343–8.
- Trimnell AR, Kraemer SM, Mukherjee S, Phippard DJ, Janes JH, Flamoe E, et al. Global genetic diversity and evolution of var genes associated with placental and severe childhood malaria. *Mol Biochem Parasitol*. 2006;148:169–80.

14. Bockhorst J, Lu F, Janes JH, Keebler J, Gamain B, Awadalla P, et al. Structural polymorphism and diversifying selection on the pregnancy malaria vaccine candidate VAR2CSA. *Mol Biochem Parasitol.* 2007;155:103–12.
15. Bordbar B, Ndam NT, Renard E, Jafari-Guemouri S, Tavul L, Jennison C, et al. Genetic diversity of VAR2CSA ID1-DBL2Xb in worldwide *Plasmodium falciparum* populations: impact on vaccine design for placental malaria. *Infect Genet Evol.* 2014;25:81–92.
16. Doritchamou J, Sabbagh A, Jespersen JS, Renard E, Salanti A, Nielsen MA, et al. Identification of a major dimorphic region in the functionally critical N-Terminal ID1 domain of VAR2CSA. *PLoS ONE.* 2015;10:e0137695.
17. Patel JC, Hathaway NJ, Parobek CM, Thwai KL, Madanitsa M, Khairallah C, et al. Increased risk of low birth weight in women with placental malaria associated with *P. falciparum* VAR2CSA clade. *Sci Rep.* 2017;7:7768.
18. Taylor SM, Van Eijk AM, Hand CC, Mwandagaliwa K, Messina JP, Tshefu AK, et al. Quantification of the burden and consequences of pregnancy-associated malaria in the Democratic Republic of the Congo. *J Infect Dis.* 2011;204:1762–71.
19. Meshnick SR, et al. Democratic Republic of the Congo 2013–14—Supplemental Malaria Report. Demographic and Health Survey (DRC-DHS II). <https://dhsprogram.com/pubs/pdf/FR300/FR300.Mal.pdf>. Accessed 8 Jan 2018.
20. Taylor SM, Antonia AL, Parobek CM, Juliano JJ, Janko M, Emch M, et al. *Plasmodium falciparum* sulfadoxine resistance is geographically and genetically clustered within the DR Congo. *Sci Rep.* 2013;3:1165.
21. Taylor SM, Antonia AL, Harrington WE, Goheen MM, Mwapasa V, Chaluluka E, et al. Independent lineages of highly sulfadoxine-resistant *Plasmodium falciparum* haplotypes, eastern Africa. *Emerg Infect Dis.* 2014;20:1140.
22. Carrel M, Patel J, Taylor SM, Janko M, Mwandagaliwa MK, Tshefu AK, et al. The geography of malaria genetics in the Democratic Republic of Congo: a complex and fragmented landscape. *Soc Sci Med.* 2015;133:233–41.
23. Miller RH, Hathaway NJ, Kharabora O, Mwandagaliwa K, Tshefu A, Meshnick SR, et al. A deep sequencing approach to estimate *Plasmodium falciparum* complexity of infection (COI) and explore apical membrane antigen 1 diversity. *Malar J.* 2017;16:490.
24. Doctor SM, Liu Y, Whitesell A, Thwai KL, Taylor SM, Janko M, et al. Malaria surveillance in the Democratic Republic of the Congo: comparison of microscopy, PCR, and rapid diagnostic test. *Diagn Microbiol Infect Dis.* 2016;85:16–8.
25. Plowe CV, Djimde A, Bouare M, Doumbo O, Wellems TE. Pyrimethamine and proguanil resistance-conferring mutations in *Plasmodium falciparum* dihydrofolate reductase: polymerase chain reaction methods for surveillance in Africa. *Am J Trop Med Hyg.* 1995;52:565–8.
26. Taylor SM, Juliano JJ, Trotman PA, Griffin JB, Landis SH, Kitsa P, et al. High-throughput pooling and real-time PCR-based strategy for malaria detection. *J Clin Microbiol.* 2010;48:512–9.
27. Beshir KB, Hallett RL, Eziefula AC, Bailey R, Watson J, Wright SG, et al. Measuring the efficacy of anti-malarial drugs in vivo: quantitative PCR measurement of parasite clearance. *Malar J.* 2010;9:312.
28. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27:2957–63.
29. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res.* 2017. <https://doi.org/10.1093/nar/gkx1201> (Epub ahead of print).
30. Ndam NT, Bischoff E, Proux C, Lavstsen T, Salanti A, Guitard J, et al. *Plasmodium falciparum* transcriptome analysis reveals pregnancy malaria associated gene expression. *PLoS ONE.* 2008;3:e1855.
31. Sander AF, Salanti A, Lavstsen T, Nielsen MA, Magistrado P, Lusingu J, et al. Multiple var2csa-type PfEMP1 genes located at different chromosomal loci occur in many *Plasmodium falciparum* isolates. *PLoS ONE.* 2009;4:e6667.
32. Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, et al. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* 2003;31:212–5.
33. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. *Bioinformatics.* 2015;31:3997–9.
34. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2011. <http://www.R-project.org/>.
35. Kumar S, Stecher G, Tamura K. MEGA7: molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
36. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
37. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123:585–95.
38. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986;3:418–26.
39. Nei M. Definition and estimation of fixation indices. *Evolution.* 1986;40:643–5.
40. Nei M. Genetic distance between populations. *Am Nat.* 1972;106:283–92.
41. Piry S, Chapuis MP, Gauffre B, Papaix J, Cruaud A, Berthier K. Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure. *Methods Ecol Evol.* 2016;7:1463–75.
42. DHS Program. <http://dhsprogram.com/data/available-datasets.cfm>. Accessed 10 Mar 2017.
43. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823.* 2014.
44. Hamilton WL, Claessens A, Otto TD, Kekre M, Fairhurst RM, Rayner JC, et al. Extreme mutation bias and high AT content in *Plasmodium falciparum*. *Nucleic Acids Res.* 2016;45:1889–901.
45. Conway DJ, Roper C, Oduola AM, Arnot DE, Kreamsner PG, Grobusch MP, et al. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc Natl Acad Sci USA.* 1999;96:4506–11.
46. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol.* 2000;17:1467–82.
47. Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AP, et al. Population genetic correlates of declining transmission in a human pathogen. *Mol Ecol.* 2013;22:273–85.
48. Daubersies P, Sallenave-Sales S, Magne S, Trape JF, Contamin H, Fandeur T, et al. Rapid turnover of *Plasmodium falciparum* populations in asymptomatic individuals living in a high transmission area. *Am J Trop Med Hyg.* 1996;54:18–26.
49. Gratepanche S, Gamain B, Smith JD, Robinson BA, Saul A, Miller LH. Induction of crossreactive antibodies against the *Plasmodium falciparum* variant protein. *Proc Natl Acad Sci USA.* 2003;100:13007–12.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

