Order within chaos: Harnessing Plasmodium falciparum var gene extreme polymorphism for malaria epidemiology

Marc-Antoine Guery, Antoine Claessens

PLOS GENETICS

PERSPECTIVE

Order within chaos: Harnessing *Plasmodium falciparum var* gene extreme polymorphism for malaria epidemiology

Marc-Antoine Guery[1], **Antoine Claessens**[2]*

1 LPHI, Université de Montpellier, CNRS, Montpellier, France, **2** LPHI, MIVEGEC, Université de Montpellier, CNRS, Montpellier, France

* antoine.claessens@umontpellier.fr

Citation: Guery M-A, Claessens A (2021) Order within chaos: Harnessing *Plasmodium falciparum var* gene extreme polymorphism for malaria epidemiology. PLoS Genet 17(2): e1009344. https://doi.org/10.1371/journal.pgen.1009344

Editor: Carmen Buchrieser, Institut Pasteur, CNRS UMR 3525, FRANCE

Published: February 25, 2021

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Genotyping methods

The characterization of *Plasmodium falciparum* genetic diversity is key to understand evolutionary pressure, measure the impact of elimination campaigns, monitor drug resistance, etc. A common genotyping method is to PCR amplify microsatellite markers or polymorphic genes, such as the *msp* family, the size of the resulting amplicon(s) being variable between different isolates [1]. Amplicon ultra-deep sequencing now allows multiplexing samples and are particularly useful to determine the proportion of each strain in infections with multiple genotypes [2]. Alternatively, genotyping a few dozen single nucleotide polymorphisms (SNPs) generates a molecular barcode specific to each isolate. Whole genome sequencing (WGS) is the most comprehensive approach, but the cost remains high for large-scale studies.

Over the last decade, the Day lab has pioneered a PCR approach based on *P. falciparum*'s most polymorphic gene family: the *var* genes [3]. The roughly 60 *var* genes have a similar organization consisting in a succession of Duffy binding-like (DBL) and cysteine-rich interdomain region (CIDR) domains with the near ubiquitous presence of DBL*α* subtype (Fig 1A) [4]. Through this approach, the DBL*α* domain is amplified using degenerate primers that match two short conserved motifs on either side of the domain. The approximately 450 bp-long sequence in between is extremely variable, making the total number of unique DBL*α* domains worldwide virtually infinite [5]. This cost-effective method is used as a surveillance tool across a variety of epidemiological settings. Unlike microsatellites that are presumably neutral markers, the DBL*α* typing offers the added advantage of characterizing the parasite's most immunogenic surface protein family [6].

DBL*α* sequences differentiate *Plasmodium falciparum* populations

To test the suitability of DBL*α* sequences for population genetics analysis, Tonkin-Hill and colleagues processed 32,682 sequences from 1,248 *P. falciparum* isolates collected in 10 different countries located in South America, Africa, Asia, and Oceania [7]. To account for the frequent recombination events between *var* genes, they developed a novel computationally intensive method known as jumping hidden Markov Model (JHMM) which is able to infer the posterior probability that each location in an isolate's DBL*α* type amino acid sequence is most closely related to every other DBL*α* type [8,9]. They accumulated the probabilities between DBL*α* found in an isolate to provide an estimate of the expected proportion of relatedness between isolates. These proportions were then aggregated to provide estimates of an isolate's DBL*α* repertoire that most closely matched each country, hence attributing a geographical origin to each isolate (Fig 1B). Country-specific clustering was observed, even in Africa where *P. falciparum*

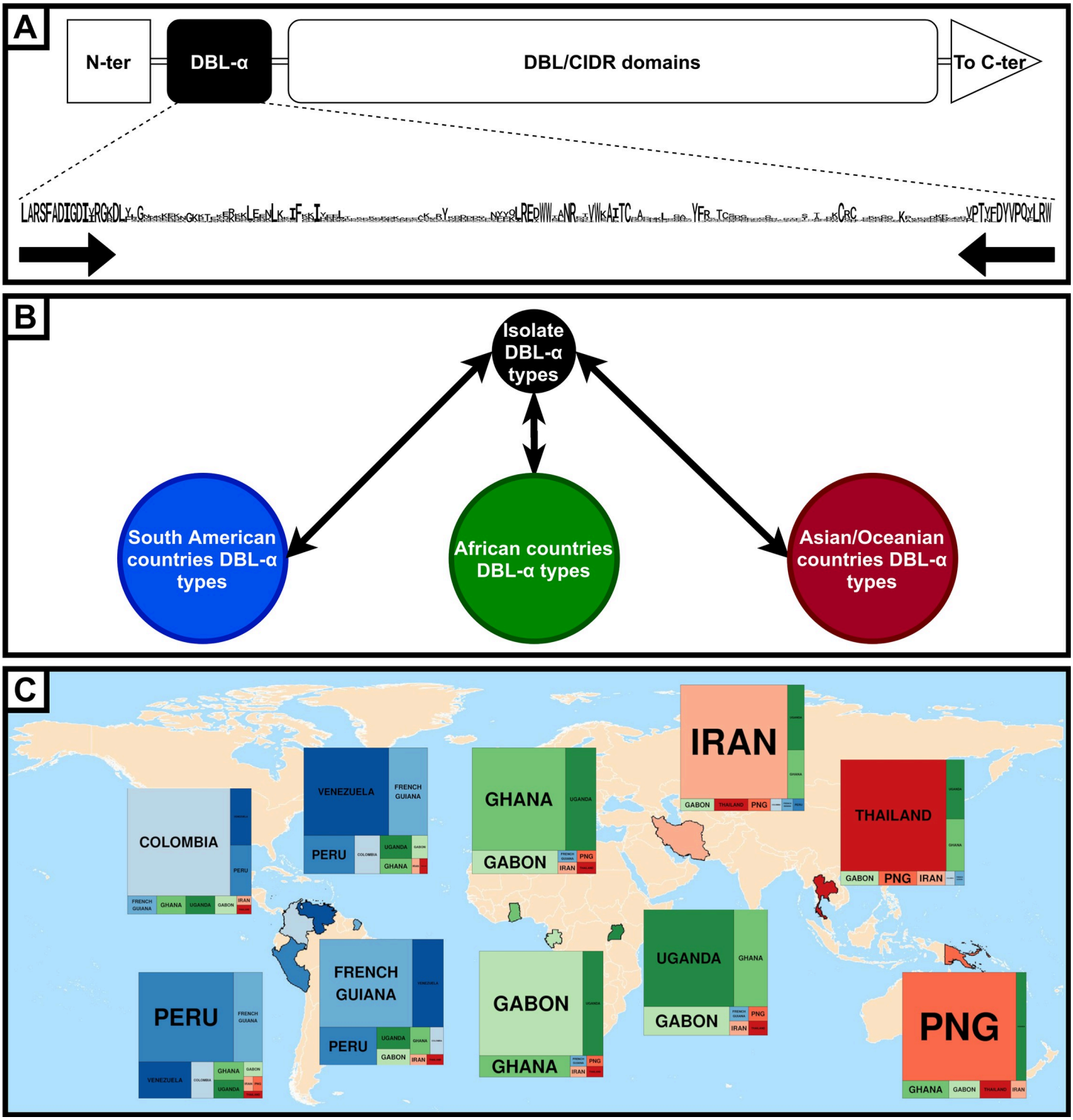

Fig 1. Relevance of using DBL*α* sequences for malaria epidemiology. (A) Schematic view of DBL*α* domain in PfEMP1, the protein encoded by *var* genes. The amino acid sequence of DBL*α* domain, with letter size proportional to the prevalence in the *P. falciparum* population, shows two conserved motifs on either side. Universal primers were designed to match the corresponding DNA sequence. (B) Each isolate's geographical origin is estimated by comparing DBL*α* types found in the isolate with DBL*α* types found in a specific country. (C) World map with each country from South America (blue), Africa (green), and Asia or Oceania (red) highlighted. Rectangle areas are proportional to the weight of each country in the overall estimated geographical origin of isolates, based on DBL*α* matching proportions. For example, isolates from Uganda share a relatively high number of DBL*α* with Ghana and Gabon, while PNG is more isolated. Each country is preferentially self-matching. Data from Tonkin-Hill and colleagues [7]. Made with Natural Earth.

https://doi.org/10.1371/journal.pgen.1009344.g001

genetic diversity is particularly high [10], indicating that the majority of DBLα types are country specific. Excluding intracountry comparisons, countries located in the same continent show greater matching proportions of DBLα sequences, especially for Africa and South America. Regarding intercontinent comparisons, DBLα types coming from South America show greater relatedness to DBLα types found in Africa compared to Asia or Oceania (Fig 1C).

These results are in line with conclusions drawn from previous works, indicating that *P. falciparum* was introduced from Africa to South America during the slave trade [11]. Also, the JHMM proportions suggest that the *var* populations in Asia/Oceania more closely resemble African populations than those seen in South America, consistent with the expansion of *P. falciparum* out of Africa toward Asia [12]. Finally, the DBLα method is also able to distinguish between *Laverania* species.

The exception to the rule: Conserved *var* genes

While most DBLα in the Tonkin-Hill dataset were detected only once, some appeared to be conserved at the global scale. The top 100 most prevalent DBLα, present in at least 50 isolates, were found to be distributed all over the countries sampled. This result was confirmed with a BLAST search against the NCBI and MalariaGEN-derived databases [13]. Although the in vitro recombination rate of group A *var* genes is lower than group B or C [9], the DBLα1 (specific to group A *var* genes) was proportionally represented in the Top 100. The overall mind-boggling genetic diversity of DBLα sequences make these specific "conserved" DBLα even more interesting.

What is driving this selection? Some DBLα could be in linkage disequilibrium with drug resistance alleles [14]. However, the authors showed that only 10% of conserved DBLα were associated with drug resistance markers, indicating that other selection forces are at play and remained to be discovered. Interestingly, they identified conserved DBLα on chromosome 6, in which long-range haplotypes have been reported in African *P. falciparum* genomes [15].

Plasmodium diversity: A matter of scale

In summary, Tonkin-Hill and colleagues produced the first global scale DBLα types analysis. Their JHMM approach is promising when dealing with relatedness between sets of sequences, even though its computational cost remains to be lowered. Comparing DBLα sequences led to a clustering similar to what was observed using WGS in Africa, for a fraction of the cost. To further investigate the ability of DBLα comparisons to track *P. falciparum* populations, more samples are needed, especially from Ethiopia and the Democratic Republic of The Congo, where divergent parasites have recently been identified [10]. Follow-up studies should also include countries where malaria is highly seasonal, to measure the impact of transmission (or lack thereof) on parasite population genetic diversity.

To conclude, Tonkin-Hill and colleagues provide compelling evidence for further DBLα-based genotyping studies, not just for "varologists" but most epidemiological studies.

References

1. Zhong D, Lo E, Wang X, Yewhalaw D, Zhou G, Atieli HE, et al. Multiplicity and molecular epidemiology of Plasmodium vivax and Plasmodium falciparum infections in East Africa. Malar J. 2018 May 2; 17 (1):185. https://doi.org/10.1186/s12936-018-2337-y PMID: 29720181
2. Gruenberg M, Lerch A, Beck H-P, Felger I. Amplicon deep sequencing improves Plasmodium falciparum genotyping in clinical trials of antimalarial drugs. Sci Rep. 2019 Nov 28; 9(1):17790. https://doi.org/10.1038/s41598-019-54203-0 PMID: 31780741

3. Taylor HM, Kyes SA, Harris D, Kriek N, Newbold CI. A study of var gene transcription in vitro using universal var gene primers. Mol Biochem Parasitol. 2000 Jan 5; 105(1):13–23. https://doi.org/10.1016/s0166-6851(99)00159-0 PMID: 10613695
4. Rask TS, Hansen DA, Theander TG, Pedersen AG, Lavstsen T. Plasmodium falciparum Erythrocyte Membrane Protein 1 Diversity in Seven Genomes–Divide and Conquer. PLOS Comput Biol. 2010 Sep 16; 6(9):e1000933. https://doi.org/10.1371/journal.pcbi.1000933 PMID: 20862303
5. Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, et al. Population Genomics of the Immune Evasion (var) Genes of Plasmodium falciparum. PLOS Pathog. 2007 Mar 16; 3(3):e34. https://doi.org/10.1371/journal.ppat.0030034 PMID: 17367208
6. Chan J-A, Howell KB, Reiling L, Ataide R, Mackintosh CL, Fowkes FJI, et al. Targets of antibodies against Plasmodium falciparum–infected erythrocytes in malaria immunity. J Clin Invest. 2012 Sep 4; 122(9):3227–38. https://doi.org/10.1172/JCI62182 PMID: 22850879
7. Tonkin-Hill G, Ruybal-Pesántez S, Tiedje KE, Rougeron V, Duffy MF, Zakeri S, et al. Evolutionary analyses of the major variant surface antigen-encoding genes reveal population structure of Plasmodium falciparum within and between continents. PLoS Genet. 2021. https://doi.org/10.1371/journal.pgen.1009269
8. Zilversmit MM, Chase EK, Chen DS, Awadalla P, Day KP, McVean G. Hypervariable antigen genes in malaria have ancient roots. BMC Evol Biol. 2013 May 31; 13(1):110. https://doi.org/10.1186/1471-2148-13-110 PMID: 23725540
9. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, et al. Generation of Antigenic Diversity in Plasmodium falciparum by Structured Rearrangement of Var Genes During Mitosis. Deitsch K, editor. PLoS Genet. 2014 Dec 18; 10(12):e1004812. https://doi.org/10.1371/journal.pgen.1004812 PMID: 25521112
10. Amambua-Ngwa A, Amenga-Etego L, Kamau E, Amato R, Ghansah A, Golassa L, et al. Major subpopulations of *Plasmodium falciparum* in sub-Saharan Africa. Science. 2019 Aug 23; 365(6455):813–6. https://doi.org/10.1126/science.aav5427 PMID: 31439796
11. Yalcindag E, Elguero E, Arnathau C, Durand P, Akiana J, Anderson TJ, et al. Multiple independent introductions of Plasmodium falciparum in South America. Proc Natl Acad Sci. 2012 Jan 10; 109(2):511–6. https://doi.org/10.1073/pnas.1119058109 PMID: 22203975
12. Tanabe K, Mita T, Jombart T, Eriksson A, Horibe S, Palacpac N, et al. Plasmodium falciparum Accompanied the Human Expansion out of Africa. Curr Biol. 2010 Jul 27; 20(14):1283–9. https://doi.org/10.1016/j.cub.2010.05.053 PMID: 20656209
13. Otto TD, Assefa SA, Böhme U, Sanders MJ, Kwiatkowski D, Berriman M, et al. Evolutionary analysis of the most polymorphic gene family in falciparum malaria. Wellcome Open Res. 2019 Dec 3; 4. https://doi.org/10.12688/wellcomeopenres.15590.1 PMID: 32055709
14. Fowler EV, Chavchich M, Chen N, Peters JM, Kyle DE, Gatton ML, et al. Physical Linkage to Drug Resistance Genes Results in Conservation of var Genes among West Pacific Plasmodium falciparum Isolates. J Infect Dis. 2006 Oct 1; 194(7):939–48. https://doi.org/10.1086/506619 PMID: 16960782
15. Amambua-Ngwa A, Danso B, Worwui A, Ceesay S, Davies N, Jeffries D, et al. Exceptionally long-range haplotypes in Plasmodium falciparum chromosome 6 maintained in an endemic African population. Malar J. 2016 Oct 21; 15(1):515. https://doi.org/10.1186/s12936-016-1560-7 PMID: 27769292