Extensive genetic diversity in *Plasmodium vivax* from Sudan and its genetic relationships with other geographical isolates

Musab M Ali. Albsheer^{1,2}, Eyoab Iyasu Gebremeskel³, Daniel Kepple⁴, Eugenia Lo⁴, Virginie Rougeron⁵, Muntaser E. Ibrahim¹, Muzamil M. Abdel Hamid^{1†}

¹ Department of Parasitology and Medical Entomology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan

² Faculty of Medical Laboratory Sciences, Sinnar University, Sudan

³ Department of Biology, Mai Nefhi College of Science, Mai-Nefhi, Eritrea

⁴ Department of Biological Sciences, the University of North Carolina at Charlotte, USA

⁵ Laboratoire MIVEGEC (Université de Montpellier-CNRS-IRD), CREES, 34394 Montpellier, France

†Correspondence: Muzamil M. Abdel Hamid, University of Khartoum, Sudan (mahdi@iend.org)

Running title: Population history of *Plasmodium vivax* isolates from Sudan

Abstract

Plasmodium vivax malaria is a neglected tropical disease in Africa due to low occurrence rates and lack of accurate diagnosis. Recently, there has been a dramatic increase in *P. vivax* cases in East Africa and reportedly spreading to western countries. This study investigated the geographical origin and genetic diversity of *P. vivax* in Sudan by 14 microsatellite markers. A total of 113 clinical *P. vivax* samples were collected from two districts, New Halfa and Khartoum in Sudan. In addition, data from 841 geographical samples retrieved from the database for global genetic analysis were included in the analysis to further the genetic relationships among the *P. vivax* isolates at regional and worldwide scales. On a regional scale, we observed 91 unique and 8 shared haplotypes amongst the Sudan samples. Such a high genetic diversity compared to other geographical isolates lends support to hypothesis that *P. vivax* was originated from Africa. On a global scale, as already demonstrated, we observed distinct genetic clustering of *P. vivax* isolates from Africa, South America, and Asia (including Papua New Guinea and Solomon Island) with limited admixture in all three clusters. The principal component analysis and phylogenetic tree showed similar clustering patterns and highlighted the contribution of the African isolates to the genetic variation observed globally. The East African *P. vivax* showed similarity with some of the Asian isolates suggesting potential recent introductions. Our results show extensive genetic diversity co-occurring with significant multi-locus linkage disequilibrium, demonstrating the effectiveness of using microsatellite markers to implement effective control measures.

Background

Plasmodium vivax malaria is a neglected tropical disease, and is regarded as a benign and self-limiting infection despite the increasing evidence that the overall burden, economic impact, and the severity of disease caused by *P. vivax* is on the rise [1,2]. Malaria caused by *P. vivax* is posing a significant re-emerging disease threat in many African, South American, and Asian countries caused by both resistance to available therapies [3–7] and the parasite's unique ability to form hypnozoites, a mechanism that enables *P. vivax* to hibernate and relapse in the liver [8–10]. Moreover, the current knowledge on genetic diversity and population structure of *P. vivax* lags behind that of *P. falciparum* due to lack of appropriate genetic markers for the *P. vivax* genome [11] and its inability to be cultured for long-term in the laboratory except in non-human primates [12, 13] further complicating control efforts. The complete sequencing of the *P. vivax* genome in 2008 provided a valuable resource and added a stimulus to the much-needed study of this neglected parasite [11] and subsequent studies [13] revealed greater genetic diversity of *P. vivax* in comparison to *P. falciparum*.

Differences in *Anopheles* mosquito dynamics provide *P. vivax* with a broader temperature tolerance and ability to be transmitted in climates not tolerated by *P. falciparum* [12, 15–18]. Studies conducted from different geographical regions showed distinctive phenotypic features of the parasite populations, which might be attributed to the presence of barriers to gene flow or strong regional selections [18]. Therefore, analyzing the genetic make-up of the isolates not only would be useful in identifying the geographic origins of these parasites [20–22], but also facilitate meaningful parasite-specific control and preventive measures to be implemented.

To understand the geographical origin and genetic diversity of isolates, microsatellite markers (MS) provide an effective means to determine population structure and for high-

throughput genotyping due to their higher polymorphic nature [23–28]. The polymorphic nature of MS loci also permits increased detection of multiclonal infections [28], which can be useful when describing the history of endemicity and the stability of transmission of malaria parasites within a specific global region [29], and can describe infection dynamics *P. vivax* across time [31, 32]. A recent study conducted in Sudan [2] showed a rise in *P. vivax* prevalence with concerning proportions and spreading to new areas including east and central African countries [16], posing new challenges to malaria treatment and control in Sudan. The overall prevalence of *P. vivax* among the malaria cases was 26.6%; of the suspected cases, the proportion of positive results for *P. vivax* was 13.3%, showing significant variations between the states and residence status [2], reflecting the stability of the transmission.

In line with the growing number of *P. vivax* cases, this study analyzed the genetic diversity and population structure of *P. vivax* isolates from Sudan using 14 highly polymorphic tri and tetranucleotide microsatellite markers [30]. An additional 841 global samples were included in the analysis to further understand the genetic origin and diversity of the isolates at a regional and global level. Microsatellite analyses resulted into high number of unique haplotypes and clustering of the isolates into geographical origins. Genetic analysis of *P. vivax* using microsatellites is robust to address questions on the geographical origin and spread of the parasites and contribute to the current malaria control and elimination effort in Sudan.

Materials and Methods

Sample collection and ethics statement

P. vivax microsatellite typing was conducted on 113 field isolates collected from two districts in Sudan: 21 from Halfa (2013) and 92 from Khartoum (2013 and 2015). A field isolate is defined in this study as a sample of parasites derived from a single-infected patient collected on a single occasion (19). All the patients were febrile with 37.5°C or above and were diagnosed as having *P. vivax* malaria by microscopic examination of Giemsa-stained thick and thin blood smears. Finger-prick blood samples, for molecular analysis, were collected on Whatman 3 mm filter paper (Whatman International, Ltd., Maidstone, England), dried, and kept in a dark place. Filter paper samples were coded to keep the anonymity of the patients and kept at Parasitology and Medical Entomology laboratory at the Institute of Endemic Diseases, the University of Khartoum for further analysis. Samples were obtained and analyzed after approval from the Ethics Committee of the Institute of Endemic Diseases, University of Khartoum (Reference number: 9/2016).

Microsatellite typing of *P. vivax*

DNA from blood samples was extracted using DNeasy Blood and Tissue kit (Qiagen, France) according to manufacturer's recommendations, and eluted in 100 µl of elution buffer per 200 µl of whole blood or per filter plot. For each DNA sample, whole genome amplification was carried out using an Illustra-GenomiPhi™ V₂ DNA Amplification Kit (GE Healthcare, Uppsala, Sweden) following the manufacturer's protocols. The product (DNA + Sample buffer) was denatured by heating at 95°C for 3 min and then kept on ice. 9µL of Reaction Buffer (Illustra™ GenomiPhi™ V2 DNA Amplification Kit - GE Healthcare, Uppsala, Sweden) and 1µL of enzyme Phi 29 was added and incubated at 30°C for 2h for genome amplification. The amplification was stopped by placing the samples at 65°C for 10 min, before being stored at -20°C. The product DNA was used as a template for the PCR-based

amplification of 14 polymorphic microsatellite markers distributed among 10 of the 14 chromosomes of *P. vivax*. The PCR protocol was adapted from Karunawera et al. (30) as follows: a 20 µl reaction mix was made of 0.3 µl of each primer (10 µM), with the forward being labelled with a fluorochrome, 0.1 µl template DNA, 0.5 µl dNTP mix (10 pmol/µl), 0.7 µl of MgCl₂ (50mM), 2.5 of 10X buffer, 0.2 µl Taq polymerase (5 UI/µl, Invitrogen). The fourteen highly polymorphic microsatellite markers used in the study for typing consist of tri-nucleotide tandem repeat motifs (MS1, MS6, MS10, MS8, MS9, MS20, MS16, MS2, MS5, MS12, MS15, MS3, MS4, and MS7) and tetra-nucleotide repeat array (MS2) (30). Amplifications were carried out in a thermal cycler: 40 cycles of denaturation at 94°C for the 30s, the annealing temperature of each locus for 1 min at 58°C and extension at 72°C for 30s, followed by a final extension at 72°C for 7 min. The reaction products were visualized on a 1.5% agarose gel stained with EZ-vision. Fluorescence-labeled PCR products were sized on Applied Biosystems ABI3500XL (GenSeq platform, Montpellier), with a Genescan 500 LIZ internal size standard.

Data analyses

In addition to the 113 *P. vivax* samples from Sudan, 841 global samples were retrieved from the database for global genetic analysis (<http://plasmodb.org> Accessed November 12, 2008) to analyze the ancestry of the isolates and to have a regional and global perspective of the parasites studied following the protocols followed in [31]. To maintain consistency, only 11 loci of the 17 populations were considered for analysis.

As blood-stage malaria parasites are haploid, the presence of one or more additional alleles at a particular locus was interpreted as a co-infection with two or more genetically distinct clones (multiple clone infections) in the same isolate [20, 34, 35] . Therefore in co-

infection cases, the single or predominant allele at each locus was considered for computing allele frequencies and to define haplotypes [23–28, 36]. The genetic diversity of the parasite populations from Sudan was determined by calculating the heterozygosity (H_E) at each locus in each separate population. Virtual heterozygosity is the average probability that a pair of alleles randomly obtained from the population is different and was defined as $H_E = [n / (n - 1)] [1 - \sum p_i^2]$, with n is the number of isolates analyzed and p_i the frequency of the i th allele in the population. Virtual heterozygosity ranges between 0 and 1, with values close to 1 reflecting high genetic diversity levels in a population.

Data were also analyzed using LIAN software version 3.5 [35] to calculate a standardized index of association (I^S_A) to test for evidence of overall multilocus linkage disequilibrium. This test compares the variance (V_D) of the number of alleles shared between all pairs of haplotypes observed in the population with the variance expected under random association of alleles (V_E) and is defined as $I^S_A = (V_D/V_E - 1)/(r - 1)$, with r being the number of loci analysed [36]. V_E is derived from 10,000 simulated data sets in which alleles were randomly reshuffled among haplotypes. Significant linkage disequilibrium is detected if V_D is > 95% of the values derived from the reshuffled data sets [37]. Moreover, GenAlEx program version 6.5 [36,37] was used to estimate allele richness and heterozygosity.

The Bayesian clustering model to assign isolates to K populations according to allele frequencies at each locus was applied using STRUCTURE 2.3.1 software [40] to test whether microsatellite haplotypes clustered according to the geographic origins of the isolates. The program was run 5 times at each of 4 different K values (K 2–6) with a burn-in period of 50,000 iterations followed by 1×10^5 iterations. The admixture model was used in all analyses. Isolates with predominant ancestry (> 70%) were considered as members of that

particular population [19]. PAST software [41] was used to determine the genetic variation among the populations. Phylogenetic tree of the 17 populations from this study and global samples was constructed using GenoDive v. 3.0 [42]. Moreover AMOVA analysis implemented in GENALEX program [38] was conducted to determine the distribution of genetic variation within and among populations included in the study.

RESULTS

Microsatellite variation in *P. vivax* isolates from Sudan

All parasite isolates from New Halfa and Khartoum were highly polymorphic and showed 68 unique and 12 shared haplotypes (Supplementary Table 2), with the number of alleles per locus (N_a) ranging from 5 to 13 in New Halfa and 6 to 21 in Khartoum (Table 1). The number of alleles observed in isolates from Khartoum is more than the number of alleles observed in isolates from New Halfa (Table 1). The most diverse alleles observed in isolates from New Halfa (0.895) and Khartoum (0.870) is locus MS15 (Table 1). High number of private alleles shared among both populations (Supplementary Figure 1). The virtual heterozygosity (H_E) values (Table 2) for the isolates from both New Halfa and Khartoum pooled together ranged from 0.70275 to 0.8931 (mean 0.80293 ± 0.06557). Significant multilocus linkage disequilibrium (Table 3) was found in the isolates ($n = 113$) from Sudan ($I_S^A = 0.1486$, $P < 0.001$). The association was high (Table 3) when isolates from Khartoum ($I_S^A = 0.1967$, $P < 0.01$) were analysed separately.

Geographic structure of *P. vivax*

The clustering patterns obtained with $K = 3$ (Figure 1) showed clear structure of *P. vivax* isolates from Africa, South America, and Asia (including Papua New Guinea and Solomon

Island) with much admixture in all three clusters. Cluster 1 (Red) included *P. vivax* isolates from Peru (90.8%), Brazil (95.4%), Central America (50.3%). Cluster 2 (Green) included *P. vivax* isolates from New Halfa (84.5%), Khartoum (81.5%), other parts of Sudan (89.9%), and of other African countries (77.3%), including Madagascar (77.4%). Cluster 3 (Blue) included *P. vivax* isolates from Cambodia (54.4%), Vietnam (76.4%), Indonesia (51.2%), Papua New Guinea low land (92.5%), Papua New Guinea high land (89.4%), and Solomon Island (79.9%). Isolates from the rest three Asian countries Armenia and India were assigned to cluster 2 with 87.5% and 74.4% proportion of the membership, respectively, whereas isolates from Uzbekistan were predominantly present in cluster 1 with 77.5% proportion of the membership. The proportion of the membership of each pre-defined population in each of the three clusters was higher than 75% except for Central America and Indonesia. When the number of clusters was set to 4 or 5 (Figure 1), the separation between countries was limited except for the isolates from Khartoum that formed a separate cluster from isolates from New Halfa and other isolates from Africa. The New Halfa and other isolates from Africa clustered with those from Peru and Uzbekistan

Principal component analysis (PCA) (Figure 2) generated using GenoDive v. 3.0 (40) showed clear clustering of all the populations analyzed into their geographical location to corroborate population structure results observed at K=3 level (Figure 1B). In general, PCA, NJ tree and clustering analyses showed similar structuring patterns. The phylogenetic tree (Figure 3) constructed using PAST further highlights the importance of isolates from Africa and Sudan to the genetic variation observed globally. AMOVA (Table 4) analysis of the 17 populations showed that most of the variation lies within the population (89%) and 11% variation observed among the populations.

Discussion

P. vivax cases in Sudan rose five times in six years [43]. The rise and re-emergence of *P. vivax* infectious could be related to the parasite's ability to develop alternative mechanisms to invade human erythrocytes other than the Duffy antigen [44], which also is associated with severe cases once more highlighting of undertaking more studies of the parasite which once was thought as benign and neglected. Continentally, the Duffy negative phenotype ranges from 60% in the eastern countries to 98% in the western countries [45] suggesting a prolonged period of host-parasite co-evolution [46]. Genomic analyses have identified that the evolution of Duffy negative infections occurred independently in a mixed Duffy population [45]. Prior studies have also shown cases of *P. vivax* transmitted from eastern African refugees to natives in Germany and Switzerland [51, 52].

Genetic markers like microsatellites (MS) are selectively neutral and still powerful tools for investigating the population structure parasites because of their polymorphic nature [50]. MS is a popular alternative to polymorphic antigenic genes in studying malaria parasites due to their purported neutral ubiquity throughout genomes and utility for describing the evolutionary history of global populations [51]. Furthermore, the relatively unconstrained polymorphic nature of MS loci permits increased detection of multi-clonal infections [28] which can be useful when describing the history of endemicity and the stability of transmission within a specific global region [29]. In line with this, we tried to characterize the genetic diversity of *P. vivax* populations from blood samples collected in New Halfa and Khartoum in Sudan, by employing MS genotyping and analysis. Overall, in the two populations studied high microsatellite polymorphisms, characterized by high heterozygosity [50] was recorded. This high heterozygosity observed is extended to all individual loci analyzed with mean H_E of 0.80293 (SD=0.06557). The observed high level of

genetic diversity in low to meso-endemic setting countries like Sudan is not uncommon as it corroborates with other studies [20, 34, 56–59], lending support to the notion that Africa is the origin of *P. vivax* [60, 61] though it is disputed [59] .

The use of 11 highly polymorphic microsatellite markers identified population structure of *P. vivax* isolates as Asian and Latin American isolates and the isolates from Africa clustering with SE Asian isolates. With the addition of a third population ($K = 3$), further differentiation between global isolates isolates from Brazil sitting in one cluster and the isolates from the Solomon Islands and Papua New Guinea assuming a separate cluster. When the number of clusters was set to 4 or 5, the separation between countries was limited [23] except for the isolates from Khartoum then assuming a separate cluster from isolates from New Halfa and other isolates from Africa, which the latter clustered with isolates from Peru and Uzbekistan highlighting the importance of designing studies at sub-regional or sub-population level in this part of Africa as was seen in human population studies [58]. The genetic association observed between the East African and Asian isolates may imply a recent re-introduction via human migration of *P. vivax* from Asia to East Africa. With all the limitation of using microsatellites in studying population genetics of parasites [52], the population structure generated using these microsatellites show promise in identifying origin [19] and providing useful information regarding the genetic structure of parasites [61, 62].

Conclusions

Microsatellite typing remains an important tool to study *P. vivax* population structure on a global scale, and differences in diversity reflect transmission intensity and isolation of parasite populations as was reported elsewhere [23]. Microsatellite-based analysis of *P.*

vivax parasites from Sudan showed extensive genetic diversity co-occurring with significant multilocus linkage disequilibrium. These parasite populations were seen to cluster according to their geographic locations and ancestry. As inferred elsewhere the predictive power of ancestry with this microsatellite data as a model is promising and could be useful in identifying the origin of *P. vivax* malaria cases and enabling meaningful control and preventive measures to be implemented. Moreover, dissecting the parasite population into subpopulations and sub-regions especially in this part of the world would give a meaningful interpretation in the origin and structuring of *P. vivax* parasites in the future.

Acknowledgements

We are greatly indebted to the patients for participating in the study by donating blood and the staffs and technicians from the Institute of Endemic Diseases, University of Khartoum for assisting sample collection; the communities and hospitals for their support and willingness to participate in this research. Authors would like also to thank IRD, CNRS-INEE, and the ANR Tremplin EVAD 2017 for generation of the data.

Funding

This research received was funded by Third World Academy of Sciences (Research Grant Agreement (RGA) No. 19-225 RG/BIO/AF/AC G - FR3240310127 The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM. Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* [Internet]. 2007 Dec [cited 2017 May 17];77(6 Suppl):79–87. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18165478>
2. Elgoraish AG, Eldin S, Elzaki G, Tagelsir R, Ibrahim A. Epidemiology and distribution of *Plasmodium vivax* malaria in Sudan Epidemiology and distribution of *Plasmodium vivax* malaria in Sudan. *Trans R Soc Trop Med Hyg*. 2019;00(June):1–8.
3. Leclerc MC, Menegon M, Cligny A, Noyer JL, Mammadov S, Aliyev N, et al. Genetic diversity of *Plasmodium vivax* isolates from Azerbaijan. *Malar J*. 2004;10:1–10.
4. Severini C, Menegon M, Luca M Di, Abdullaev I, Majoria G, ARazakov S, et al. Risk of *Plasmodium vivax* malaria reintroduction in Uzbekistan: Genetic characterization of parasites and status of potential malaria vectors in the Surkhandarya region. *Trans R Soc Trop Med Hyg*. 2004;98(10):585–92.
5. Bahk YY, Lee H, Na B, Kim J, Jin K, Hong YS, et al. Epidemiological Characteristics of Re-emerging Vivax Malaria in the Republic of Korea (1993-2017). *Korean J Parasitol*. 2018;56(6):531–43.
6. Faulde MK, Hoffmann R, Fazilat KM, Hoerauf A. Malaria Reemergence in Northern Afghanistan. *Emerg Infect Dis*. 2007;13(9):1402–1404.
7. Baird JK. Resurgent Malaria at the Millennium : Control Strategies in Crisis. *Drugs*. 2000;59(4):719–43.
8. Joshi H, Prajapati SK, Verma A, Kang’a S, Carlton JM. *Plasmodium vivax* in India. *Trends Parasitol*. 2008;24(5):228–35.

- 309 9. Campo B, Vandal O, Wesche DL, Burrows JN. Killing the hypnozoite – drug discovery
310 approaches to prevent relapse in *Plasmodium vivax*. *Pathog Glob Health*.
311 2015;109(3):107–22.
- 312 10. Markus MB. Malaria : Origin of the Term “Hypnozoite.” *J Hist Biol*. 2011;44(4):781–6.
- 313 11. Carlton J, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, et al. Comparative
314 genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*.
315 2008;455(October 9):757–63.
- 316 12. Roobsoong W, Tharinjaroen CS, Rachaphaew N, Chobson P, Schofield L, Cui L, et al.
317 Improvement of culture conditions for long-term in vitro culture of *Plasmodium vivax*.
318 *Malar J*. 2015;14(1):1–11.
- 319 13. Neafsey DE, Galinsky K, Jiang RHY, Young L, Sykes SM, Saif S, et al. The malaria
320 parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium*
321 *falciparum*. *Nat Genet*. 2012;44(9):1046–50.
- 322 14. Bi Y, Yu W, Hu W, Lin H, Guo Y, Zhou XN, et al. Impact of climate variability on
323 *Plasmodium vivax* and *Plasmodium falciparum* malaria in Yunnan Province, China.
324 *Parasites and Vectors*. 2013;6(1):1–12.
- 325 15. Battle KE, Gething PW, Elyazar IRF, Moyes CL, Sinka ME, Howes RE, et al. The Global
326 Public Health Significance of *Plasmodium vivax*. Vol. 80, *Advances in Parasitology*.
327 2012. p. 1–111.
- 328 16. Gething PW, Elyazar IRF, Moyes CL, Smith DL, Battle KE, Guerra CA, et al. A Long
329 Neglected World Malaria Map: *Plasmodium vivax* Endemicity in 2010. Carlton JM,
330 editor. *PLoS Negl Trop Dis* [Internet]. 2012 Sep 6 [cited 2017 May 22];6(9):e1814.
331 Available from: <http://dx.plos.org/10.1371/journal.pntd.0001814>
- 332 17. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, et al. Global

- epidemiology of *Plasmodium vivax*. *Am J Trop Med Hyg*. 2016;95(Suppl 6):15–34.
18. Benavente ED, Ward Z, Chan W, Mohareb FR, Sutherland CJ, Roper C, et al. Genomic variation in *Plasmodium vivax* malaria reveals regions under selective pressure. *PLoS One* [Internet]. 2017;12(5):1–15. Available from: <https://doi.org/10.1371/journal.pone.0177134>
19. Gunawardena S, Karunaweera ND, Ferreira MU, Phone-Kyaw M, Pollack RJ, Alifrangis M, et al. Geographic structure of *Plasmodium vivax*: Microsatellite analysis of parasite populations from Sri Lanka, Myanmar, and Ethiopia. *Am J Trop Med Hyg*. 2010;82(2):235–42.
20. Arnott A, Barry AE, Reeder JC. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malar J* [Internet]. 2012 [cited 2017 May 22];11(1):14. Available from: <http://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-11-14>
21. Arnott A, Mueller I, Ramsland PA, Siba PM, Reeder JC, Barry AE. Global Population Structure of the Genes Encoding the Malaria Vaccine Candidate, *Plasmodium vivax* Apical Membrane Antigen 1 (PvAMA1). *PLoS Negl Trop Dis*. 2013;7(10).
22. Sutton PL. A call to arms : on refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. *Malar J*. 2013;12:447.
23. Koepfli C, Rodrigues PT, Antao T, Orjuela-Sánchez P, Van den Eede P, Gamboa D, et al. *Plasmodium vivax* Diversity and Population Structure across Four Continents. Vinetz JM, editor. *PLoS Negl Trop Dis* [Internet]. 2015 Jun 30 [cited 2019 Nov 1];9(6):e0003872. Available from: <https://dx.plos.org/10.1371/journal.pntd.0003872>
24. Kittichai V, Koepfli C, Nguitragool W, Sattabongkot J, Cui L. Substantial population structure of *Plasmodium vivax* in Thailand facilitates identification of the sources of

- residual transmission. PLoS Negl Trop Dis. 2017;11(10):1–14.
25. Congpuong K, Ubalee R. Population genetics of Plasmodium vivax in four high malaria endemic areas in Thailand. Korean J Parasitol. 2017;55(5):465–72.
26. Barry AE, Waltmann A, Koepfli C, Barnadas C, Mueller I. Uncovering the transmission dynamics of Plasmodium vivax using population genetics. Pathog Glob Health. 2015;109(3):142–52.
27. Iwagami M, Fukumoto M, Hwang SY, Kim SH, Kho WG, Kano S. Population structure and transmission dynamics of Plasmodium vivax in the Republic of Korea based on microsatellite DNA analysis. Vol. 6, PLoS Neglected Tropical Diseases. 2012.
28. Havryliuk T, Ferreira MU. A closer look at multiple-clone Plasmodium vivax infections : detection methods , prevalence and consequences. 2009;104(February):67–73.
29. Branch OH, Sutton PL, Barnes C, Castro JC, Hussin J, Awadalla P, et al. Plasmodium falciparum genetic diversity maintained and amplified over 5 years of a low transmission endemic in the peruvian amazon. Mol Biol Evol. 2011;28(7):1973–86.
30. Karunaweera ND, Ferreira MU, Hartl DL, Wirth DF. Fourteen polymorphic microsatellite DNA markers for the human malaria parasite Plasmodium vivax. Mol Ecol Notes. 2007;7:172–5.
31. Aurrecoechea C, Brestelli J, Brunk BP, Dommer J, Fischer S, Gajria B, et al. PlasmoDB: A functional genomic database for malaria parasites. In: Nucleic Acids Research. 2009. p. 539–43.
32. Imwong M, Nair S, Pukrittayakamee S. Contrasting genetic structure in Plasmodium vivax populations from Asia and South America. Int J Parasitol. 2007;37(8–9):1013–22.

33. Mathema VB, Nakeesathit S, Pagornrat W, Smithuis F, White NJ, Dondorp AM, et al. Polymorphic markers for identification of parasite population in Plasmodium malariae. Malar J [Internet]. 2020;19(1):1–12. Available from: <https://doi.org/10.1186/s12936-020-3122-2>
34. Karunaweera ND, Ferreira MU, Munasinghe A, Barnwell JW, Collins WE, King CL, et al. Extensive microsatellite diversity in the human malaria parasite Plasmodium vivax. Vol. 410, Gene. 2008. p. 105–12.
35. Haubold B, Hudson RR. Lian 3.0: Detecting Linkage disequilibrium in multilocus data. Bioinformatics [Internet]. 2000;16(9):847–8. Available from: <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/16/9/847>
36. Hudson RR. Analytical results concerning linkage disequilibrium in models with genetic transformation and recombination. J Evol Biol 1994;7: 535 – 548.
37. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis . Bioinformatics 2010;16 : 847 – 848.
38. Peakall R, Smouse PE. GenAEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6:288–95.
39. Peakall R, Smouse PE. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. Bioinformatics. 2012;28(19):2537–9.
40. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. Genetics [Internet]. 2000;155(2):945–59. Available from: <http://www.genetics.org/cgi/content/abstract/155/2/945>
41. Hammer Ø, Harper D, Ryan P. PAST: Paleontological Statistics Software Package for Education and data analysis. Vol. 4, Palaeontol Electron. 2001. p. 1–9.

42. Meirmans PG. genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. Mol Ecol Resour. 2020;(September 2019):1–6.
43. Elgoraish AG, Elzaki SEG, Ahmed RT, Ahmed AI, Fadlalmula HA, Abdalgader Mohamed S, et al. Epidemiology and distribution of Plasmodium vivax malaria in Sudan. Trans R Soc Trop Med Hyg [Internet]. 2019 Sep 6 [cited 2019 Nov 1];113(9):517–24. Available from: <https://academic.oup.com/trstmh/article/113/9/517/5510701>
44. Suliman M, Hamad BM, Albasheer MMA, Elhadi M, Mustafa MA, Elobied M, et al. Molecular Evidence of High Proportion of Plasmodium vivax Malaria Infection in White Nile Area in Sudan. J Parasitol Res. 2016;1–4.
45. Howes RE, Patil AP, Piel FB, Nyangiri OA, Kabaria CW, Gething PW, et al. The global distribution of the Duffy blood group. Nat Commun [Internet]. 2011 Apr [cited 2017 May 22];2(1):266. Available from: <http://www.nature.com/doifinder/10.1038/ncomms1265>
46. Mu J. Host Switch Leads to Emergence of Plasmodium vivax Malaria in Humans. Mol Biol Evol [Internet]. 2005 Apr 27 [cited 2017 May 22];22(8):1686–93. Available from: <https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msi160>
47. Sibley CH. A Solid Beginning to Understanding Plasmodium vivax in Africa. J Infect Dis. 2019;220(11):1716–8.
48. Roggeline L, Tappe D, Noack B, Addo MM, Tannich E, Rothe C. Sharp increase of imported Plasmodium vivax malaria seen in migrants from Eritrea in Hamburg, Germany. Malar J. 2016;15(1):4–8.
49. Eperon G, Durieux-Paillard S, Mauris A, Chappuis F, Gysin N. Malaria cases in Switzerland from 2005 to 2015 and recent rise of imported Plasmodium vivax malaria. Swiss Med Wkly. 2017;147(October):w14510.

- 429 50. Havryliuk T, Ferreira MU. A closer look at multiple-clone *Plasmodium vivax* infections:
430 Detection methods, prevalence and consequences. Vol. 104, *Memorias do Instituto*
431 *Oswaldo Cruz*. 2009. p. 67–73.
- 432 51. Sutton PL. A call to arms: On refining *Plasmodium vivax* microsatellite marker panels
433 for comparing global diversity. *Malar J*. 2013;12(1):1–12.
- 434 52. Ellegren H. Microsatellites: Simple sequences with complex evolution. *Nat Rev Genet*.
435 2004;5(6):435–45.
- 436 53. Abdelraheem MH, Bansal D, Idris MA, Mukhtar MM, Hamid MMA, Imam ZS, et al.
437 Genetic diversity and transmissibility of imported *Plasmodium vivax* in Qatar and
438 three countries of origin. *Sci Rep [Internet]*. 2018;8(1):1–9. Available from:
439 <http://dx.doi.org/10.1038/s41598-018-27229-z>
- 440 54. Imwong M, Sudimack D, Pukrittayakamee S, Osorio L, Carlton JM, Day NPJ, et al.
441 Microsatellite Variation, Repeat Array Length, and Population History of *Plasmodium*
442 *vivax*. *Mol Biol Evol [Internet]*. 2006 Feb 22 [cited 2019 Sep 25];23(5):1016–8.
443 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16507919>
- 444 55. Kim JR, Imwong M, Nandy A, Chotivanich K, Nontprasert A, Tonomsing N, et al.
445 Genetic diversity of *Plasmodium vivax* in Kolkata, India. *Malar J*. 2006;5:1–10.
- 446 56. Moon S-U, Lee H-W, Kim J-Y, Na B-K, Cho S-H, Lin K, et al. High frequency of genetic
447 diversity of *Plasmodium vivax* field isolates in Myanmar. *Acta Trop [Internet]*. 2009
448 Jan [cited 2019 Sep 25];109(1):30–6. Available from:
449 <https://linkinghub.elsevier.com/retrieve/pii/S0001706X08002350>
- 450 57. Loy DE, Liu W, Li Y, Learn GH, Plenderleith LJ, Sundararaman SA, et al. Out of Africa:
451 origins and evolution of the human malaria parasites *Plasmodium falciparum* and
452 *Plasmodium vivax*. *Int J Parasitol [Internet]*. 2017;47(2–3):87–97. Available from:

<http://dx.doi.org/10.1016/j.ijpara.2016.05.008>

58. Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA, et al. African origin of the malaria parasite *Plasmodium vivax*. *Nat Commun*. 2014;5.
59. Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, et al. A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite. *Proc Natl Acad Sci [Internet]*. 2005 Feb 8 [cited 2017 May 21];102(6):1980–5. Available from: <http://www.pnas.org/cgi/doi/10.1073/pnas.0409652102>
60. Iacovacci G, Atanasio ED, Marini O, Coppa A, Sellitto D, Trombetta B, et al. Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. *Forensic Sci Int Genet*. 2017;27:123–31.
61. Severini C, Menegon M, Gradoni L, Majori G. Use of the *Plasmodium vivax* merozoite surface protein 1 gene sequence analysis in the investigation of an introduced malaria case in Italy. *Acta Trop [Internet]*. 2002 Nov [cited 2017 May 24];84(2):151–7. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0001706X02001869>
62. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010 May;10(3):564–7.

Tables

Table 1. Isolates from New Halfa (N) and Khartoum (N) were analyzed separately using the GENALEX program to determine allelic diversity in the populations at an individual locus. The table shows sample size (N), number of alleles (Na), number of effective alleles (Ne), information index (I), Diversity (h) and unbiased diversity (uh) for every 11 individual loci.

Table 2. All isolates from Sudan (N) were analysed using GENALEX program to determine the number of alleles (Na), the number of effective alleles (Ne), information index (I), Diversity (h), unbiased diversity (uh) and virtual heterozygosity (H_E) for each locus.

Table 3. The genetic diversity of the parasite isolates from New Halfa, Khartoum and all samples from Sudan (all samples pooled) were determined by calculating the virtual heterozygosity (H_E) using LIAN software.

Table 4. AMOVA results showing genetic variation within and among populations implemented in GENALEX program. The P value is 0.001.

Figures

Figure 1. Population structure of 17 populations analysed at the K=2, 3, 4, 5 and 6. Cluster 1 (Red) includes *P. vivax* isolates from Peru, Brazil, and Central America. Cluster 2 (Green) includes *P. vivax* isolates from New Halfa, Khartoum, other parts of Sudan and Africa, Madagascar and isolates from Peru. Cluster 3 (Blue) includes *P. vivax* isolates from Cambodia,

Vietnam, Indonesia, Papua New Guinea low land, Papua New Guinea high land and Solomon Island.

Figure 2. Principal component analysis on Sudanese and global *P. vivax* populations constructed using PAST software based on the genetic distance data generated from GenAlex software. Red includes isolates from New Halfa, Khartoum, Madagascar and isolates from another part of Sudan and Africa; Green includes isolates from Asia (Armenia, India, Cambodia, Vietnam, Indonesia, and Uzbek); Yellow includes isolates from Papua New Guinea and the Solomon Islands; and Blue includes isolates from Brazil, Central America, and Peru.

Figure 3. Phylogenetic tree generated for Sudanese and global *P. vivax* populations constructed using GenoDive v. 3.0 [43] based on the genetic distance data generated from GenAlex software. AR: Armenia, BR: Brazil, CA: Central Americas, HLF: New Halfa, Id and In: Indonesia/India, KH: Khartoum, Ma: Madagascar, Pe: Peru, PH: Papua New Guinea highlands, PL: Papua New Guinea Lowlands, Su: Other Sites in Sudan, Uz: Uzmbek, and Vi: Vietnam.

Supplementary Files

Supplementary Figure 1. Estimate of allele richness and private alleles in both populations analysed. Data generated using GenAlEx program.

Supplementary Table 1. Haplotypes generated using Arlequin program [63] with 68 unique haplotypes and 12 shared haplotypes in all the isolates from this study.

$K=2$

$K=3$

$K=4$

$K=5$

$K=6$

$K=2$

$K=3$

$K=4$

$K=5$

$K=6$