RESEARCH NOTE

Divergent pattern of genomic variation in *Plasmodium falciparum* and *P. vivax* [version 1; referees: 2 approved with reservations]

Preeti Goel, Gajinder Pal Singh

Molecular Medicine Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi, India

v1 **First published:** 25 Nov 2016, **5**:2763 (doi: 10.12688/f1000research.10255.1)
Latest published: 25 Nov 2016, **5**:2763 (doi: 10.12688/f1000research.10255.1)

Abstract

The two main species causing malaria in humans, *Plasmodium falciparum* and *P. vivax,* differ significantly from each other in their evolutionary response to common drugs, but the reasons for this are not clear. Here we utilized the recently available large-scale genome sequencing data from these parasites and compared the pattern of single nucleotide polymorphisms, which may be related to these differences. We found that there was a five-fold higher preference for AT nucleotides compared to GC nucleotides at synonymous single nucleotide polymorphism sites in *P. vivax*. The preference for AT nucleotides was also present at non-synonymous sites, which lead to amino acid changes favouring those with codons of higher AT content. The substitution bias was also present at low and moderately conserved amino acid positions, but not at highly conserved positions. No marked bias was found at synonymous and non-synonymous sites in *P. falciparum*. The difference in the substitution bias between *P. falciparum* and *P. vivax* found in the present study may possibly contribute to their divergent evolutionary response to similar drug pressures.

Open Peer Review

Referee Status: ? ?

	Invited Referees 1	2
version 1 published 25 Nov 2016	? report	? report

1 **Richard D. Pearson**, Wellcome Trust Sanger Institute UK, Wellcome Trust Centre for Human Genetics UK

2 **Charles Woodrow**, Mahidol University Thailand, **Ipsita Sinha**, Mahidol University Thailand

Discuss this article

Comments (0)

Corresponding author: Gajinder Pal Singh (gajinder.pal.singh@gmail.com)

How to cite this article: Goel P and Singh GP. **Divergent pattern of genomic variation in *Plasmodium falciparum* and *P. vivax* [version 1; referees: 2 approved with reservations]** *F1000Research* 2016, **5**:2763 (doi: 10.12688/f1000research.10255.1)

Grant information: The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297) and a Junior Research Fellowship from University Grants Commission (UGC) of India to P.G.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 25 Nov 2016, **5**:2763 (doi: 10.12688/f1000research.10255.1)

Introduction

Plasmodium falciparum and *P. vivax* are the two major species causing malaria in humans. These species differ greatly in their geographical distribution, mortality rates and resistance to anti-malarial drugs. *P. falciparum* is responsible for ~200 million malaria cases and ~440,000 deaths annually, of which ~90% occur in Africa, while *P. vivax* causes ~14 million malaria cases and 1400–15000 deaths annually, of which ~75% occur in South and South-East Asia[1]. Chloroquine was used as a frontline drug for both *P. falciparum* and *P. vivax*, but widespread resistance to chloroquine has only been observed in *P. falciparum*[2]. Therefore, chloroquine remains a frontline drug against *P. vivax* in most parts of the world, despite its usage for ~70 years[2]. The *crt* gene, which is involved in chloroquine resistance in *P. falciparum,* is not associated with chloroquine resistance in *P. vivax*[3], despite significant conservation of the protein in the two species, and the mechanism of chloroquine resistance in *P. vivax* remains unknown. *P. falciparum* strains have developed resistance to almost all currently used drugs, including artemisinin, the most effective anti-malarial drug, and dealing with drug-resistant *P. falciparum* is one of the main contemporary public health challenges[4,5]. While *P. vivax* has been exposed to artemisinin, due to its frequent co-infection with *P. falciparum*[6] and the usage of artemisinin against *P. vivax* in areas with chloroquine resistance[7], artemisinin resistance in *P. vivax* has not yet been observed[8–10]. Thus, the evolutionary response of *P. falciparum* and *P. vivax* against anti-malarial drugs appears to be different[11].

A large amount of genome sequencing data has recently been generated from thousands of *P. falciparum* and hundreds of *P. vivax* samples[12–14]. This provides an unprecedented opportunity to compare the evolutionary patterns in the two species. The present study analysed this genomic data, and found a large substitution bias in *P. vivax,* even at non-synonymous sites, leading to biased amino acid changes. This may be related to the differential evolutionary response to same anti-malarial drugs observed in the two parasites.

Methods

The single nucleotide polymorphism (SNP) data of *P. falciparum* and *P. vivax* was obtained from the MalariaGen community webpage (https://www.malariagen.net/data/p-falciparum-community-project-jan-2016-data-release; https://www.malariagen.net/data/p-vivax-genome-variation-may-2016-data-release). The SNP data for *P. falciparum* consists of filtered and high quality 939,687 exonic SNPs with 631,715 non-synonymous and 307,972 synonymous SNPs from 3,394 samples from 22 countries[13]. The SNP data for *P. vivax* consists of filtered and high quality 303,616 SNPs from 228 samples[14]. Of these there were 87,877 non-synonymous, 62,862 synonymous and 152,877 non-coding SNPs. Proteome sequences of *P. falciparum* 3D7, *P. berghei* ANKA, *P. chabaudi chabaudi*, *P. cynomolgi* B, *P. knowlesi* H, *P. reichenowi* CDC, *P. vivax* Sal1, *P. yoelii* 17X were downloaded from the PlasmoDB database (http://plasmodb.org/common/downloads/release-27/). Orthologous sequences were identified using best bidirectional hit algorithm[15] and aligned using ClustalO (http://www.clustal.org/omega/)[16]. The conservation score for *P. vivax* residues was calculated as the average substitution score using BLOSUM62 matrix across seven orthologs at non-gapped positions.

Statistical analysis

All statistical analyses were performed in R software v3.3.1 (https://www.r-project.org/). R commands `cor.test` was used for calculating the Spearman rank correlation coefficients.

Results

There is a large difference in the genomic AT content of the two *Plasmodium* species. *P. falciparum* has a genomic AT content of 81% compared to 58% for *P. vivax*, thus the two species have diverged in their AT content from their common ancestor[17]. It has been proposed that the common ancestor of the two species was AT rich[17] and *P. vivax* has increased its genomic GC content since its divergence from the common ancestor. We tested whether this is true during the recent evolution of *P. vivax* by analysing the SNP data. We found highly biased substitution patterns in *P. vivax,* such that SNPs that change GC to AT nucleotides were approximately three times more common than those that change AT to GC nucleotides (Figure 1). This bias was present at synonymous, non-synonymous and non-coding sites (Figure 1) and indicates a recent opposite substitution bias in *P. vivax* compared to the general increase in its genomic GC content since its divergence from the common ancestor of *P. falciparum* and *P. vivax*. The biased substitution pattern at non-synonymous sites in *P. vivax* was reflected in the pattern of amino acid changes at the polymorphic sites, such that amino acids with GC-rich codons are reduced in abundance, while amino acids with AT-rich codons are increased in abundance (Figure 2A).

We asked whether substitution bias in *P. vivax* might also influence amino acid changes at conserved positions. The substitution bias was present at amino acid positions that are low to moderately conserved, but not at highly conserved positions (Figure 1). As a control, there was no relationship between conservation and substitution bias at synonymous sites (Figure 1).

We next tested whether similar bias might be present in *P. falciparum.* There was no marked substitution bias at synonymous or non-synonymous sites in *P. falciparum* (Figure 1). Consequently, there was no bias in amino acid changes at polymorphic sites according to the GC content of its codons (Figure 2B).

Discussion

The present study finds a sharp recent reversal in the substitution bias in *P. vivax* favouring AT nucleotides compared to the general increase in its GC content since its divergence from the common ancestor of *P. vivax* and *P. falciparum*. This substitution bias has a consequence for the pattern of amino acid changes even at moderately conserved, and thus functionally important, sites (Figure 1). No such bias was observed in *P. falciparum* (Figure 1). The large difference in the substitution bias between *P. vivax* and *P. falciparum* may lead to different evolutionary solutions to similar drug pressure. It has been proposed that differences in the life cycle of the two *Plasmodium* species, specifically the early onset of gametocyte stage in *P. vivax,* which allows transmission before the malaria symptoms and drug treatment, may impede the spread of drug resistance in *P. vivax*[11]. It is also possible that the strength of negative selection might be different between the two species. The ratio of non-synonymous to synonymous polymorphisms (N/S) is much higher in *P. falciparum* compared to

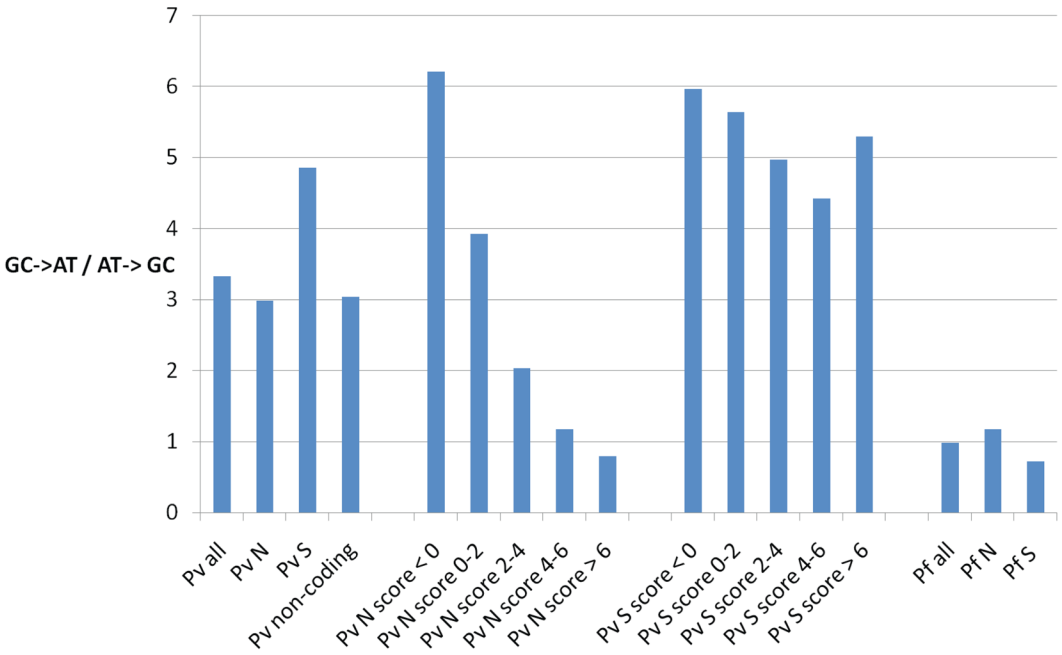

Figure 1. Substitution bias favoring AT nucleotides in *Plasmodium vivax*. There are three times as many single nucleotide polymorphisms in *P. vivax* (Pv) that change GC to AT nucleotides compared to those that change AT to GC nucleotides. This bias becomes higher at synonymous sites (S). At non-synonymous sites (N) the bias becomes lower with a higher conservation score. The conservation score for each amino acid at non-synonymous sites was calculated as the average BLOSUM62 substitution score across seven *Plasmodium* orthologs at non-gapped positions. No such bias was observed in *P. falciparum* (Pf).

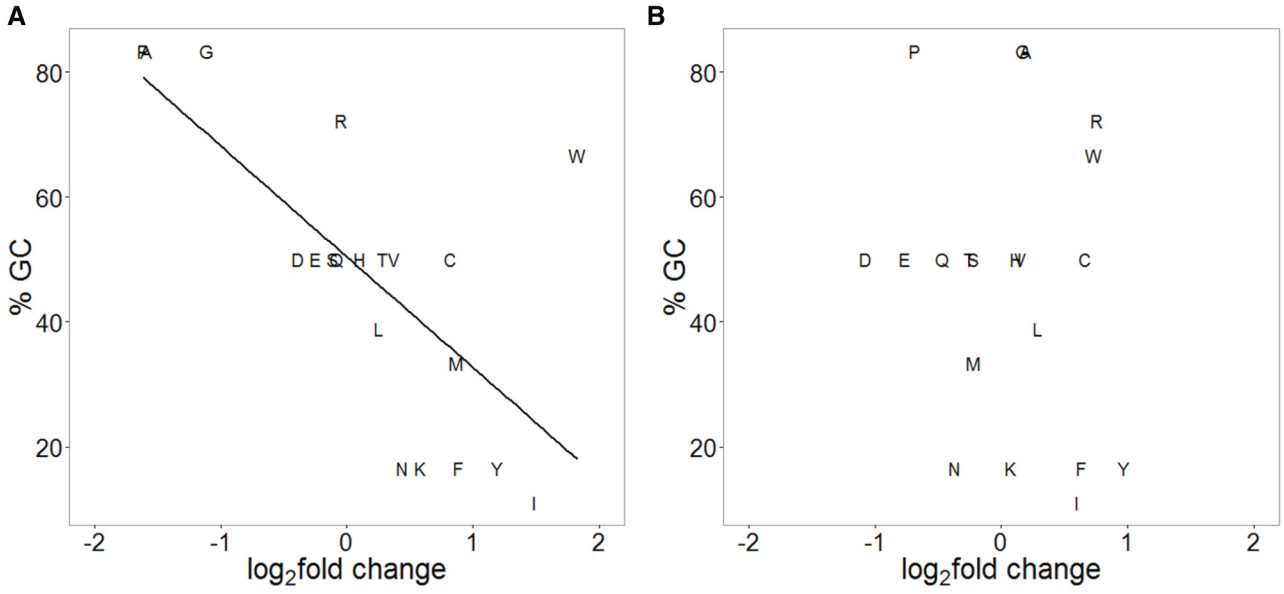

Figure 2. Substitution bias favoring AT nucleotides in *Plasmodium vivax* leads to biased amino acids changes favoring those with AT-rich codons. (**A**) *P. vivax*. The average % GC content of codons for different amino acids is plotted on the Y-axis and the $\log_2$ fold change in the amino acid abundance at non-synonymous sites is plotted on the X-axis. A significant negative correlation is observed (Spearman correlation coefficient 0.69, p=0.0008). (**B**) *P. falciparum*. No correlation was observed (Spearman correlation coefficient -0.09, p=0.7).

P. vivax (2.1[13] and 1.4[14], respectively). We found that the difference in N/S was also present when considering amino acid sites conserved across *Plasmodium* species (0.64 and 0.37 for *P. falciparum* and *P. vivax,* respectively). A higher tolerance for non-synonymous changes at conserved amino acid positions in *P. falciparum* suggests that fitness reducing mutations might have a higher likelihood to be established in *P. falciparum* compared to *P. vivax*. Since drug resistance evolution often entails fitness cost[18], it might be easier to acquire fitness reducing drug resistance mutations in *P. falciparum* compared to *P. vivax*. It is likely that a combination of these and other factors might contribute towards differences in the drug resistance evolution in the two species.

It has been proposed that the common ancestor of *P. falciparum* and *P. vivax* was AT-rich and *P. vivax* has subsequently been evolving towards higher GC content[17]. Here we find a recent reversal in the substitution bias in *P. vivax,* where it is now evolving towards increasing AT content. Interestingly, we observed a lower substitution bias in the non-coding regions in *P. vivax* compared to synonymous sites (Figure 1), which might suggest a higher functional constraint in the non-coding regions compared to synonymous sites. This observation may be utilized to identify non-coding regions in *P. vivax* genomes that are under higher functional constraint as more genomics data becomes available in the future.

Data availability

This publication uses data from the MalariaGEN *Plasmodium falciparum* Community Project, as described in 'Genomic epidemiology of artemisinin resistant malaria', eLife, 2016 (DOI: 10.7554/eLife.08714)[13], and the MalariaGEN *P. vivax* Genome Variation project, as described by Pearson *et al*. in Nature Genetics, 2016 (DOI: 10.1038/ng.3599)[14]. This data is also available from the MalariaGEN website (https://www.malariagen.net/data/p-falciparum-community-project-jan-2016-data-release; https://www.malariagen.net/data/p-vivax-genome-variation-may-2016-data-release).

Author contributions

G.P.S. conceived and designed the study. G.P.S. and P.G. performed the research. G.P.S. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

The work is supported by an Early Career Fellowship to G.P.S. by the Wellcome Trust/DBT India Alliance (IA/E/15/1/502297) and a Junior Research Fellowship from University Grants Commission (UGC) of India to P.G.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

Genome sequencing was performed by the Wellcome Trust Sanger Institute and the Community Project is coordinated by the MalariaGEN Resource Centre with funding from the Wellcome Trust.

References

1. World Health Organization: **The World Malaria Report 2015.** Reference Source
2. Baird JK: **Chloroquine resistance in *Plasmodium vivax*.** *Antimicrob Agents Chemother.* 2004; **48**(11): 4075–4083. PubMed Abstract | Publisher Full Text | Free Full Text
3. Nomura T, Carlton JM, Baird JK, *et al.*: **Evidence for different mechanisms of chloroquine resistance in 2 *Plasmodium* species that cause human malaria.** *J Infect Dis.* 2001; **183**(11): 1653–1661. PubMed Abstract | Publisher Full Text
4. Woodrow CJ, White NJ: **The clinical impact of artemisinin resistance in Southeast Asia and the potential for future spread.** *FEMS Microbiol Rev.* 2016; pii: fuw037. PubMed Abstract | Publisher Full Text
5. Lover AA, Gosling R, Feachem R, *et al.*: **Eliminate now: seven critical actions required to accelerate elimination of *Plasmodium falciparum* malaria in the Greater Mekong Subregion.** *Malar J.* 2016; **15**(1): 518. PubMed Abstract | Publisher Full Text | Free Full Text
6. Mayxay M, Pukrittayakamee S, Newton PN, *et al.*: **Mixed-species malaria infections in humans.** *Trends Parasitol.* 2004; **20**(5): 233–240. PubMed Abstract | Publisher Full Text
7. World Health Organization: **Global report on antimalarial drug efficacy and drug resistance: 2000–2010.** Geneva, World Health Organization (WHO). 2010. Reference Source
8. Wang M, Siddiqui FA, Fan Q, *et al.*: **Limited genetic diversity in the PvK12 Kelch protein in *Plasmodium vivax* isolates from Southeast Asia.** *Malar J.* 2016; **15**(1): 537. PubMed Abstract | Publisher Full Text | Free Full Text
9. Deng S, Ruan Y, Bai Y, *et al.*: **Genetic diversity of the *Pvk12* gene in *Plasmodium vivax* from the China-Myanmar border area.** *Malar J.* 2016; **15**(1): 528. PubMed Abstract | Publisher Full Text | Free Full Text
10. Brazeau NF, Hathaway N, Parobek CM, *et al.*: **Longitudinal Pooled Deep Sequencing of the *Plasmodium vivax K12* Kelch Gene in Cambodia Reveals a Lack of Selection by Artemisinin.** *Am J Trop Med Hyg.* 2016; pii: 16-0566. PubMed Abstract | Publisher Full Text
11. Schneider KA, Escalante AA: **Fitness components and natural selection: why are there different patterns on the emergence of drug resistance in *Plasmodium falciparum* and *Plasmodium vivax*?** *Malar J.* 2013; **12**: 15. PubMed Abstract | Publisher Full Text | Free Full Text
12. Hupalo DN, Luo Z, Melnikov A, *et al.*: **Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*.** *Nat Genet.* 2016; **48**(8): 953–958. PubMed Abstract | Publisher Full Text
13. MalariaGEN Plasmodium falciparum Community Project: **Genomic epidemiology of artemisinin resistant malaria.** *eLife.* 2016; **5**: pii: e08714. PubMed Abstract | Publisher Full Text | Free Full Text

14. Pearson RD, Amato R, Auburn S, *et al.*: **Genomic analysis of local variation and recent evolution in *Plasmodium vivax*.** *Nat Genet.* 2016; **48**(8): 959–964.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Wolf YI, Koonin EV: **A tight link between orthologs and bidirectional best hits in bacterial and archaeal genomes.** *Genome Biol Evol.* 2012; **4**(12): 1286–1294.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Sievers F, Wilm A, Dineen D, *et al.*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol.* 2011; **7**(1): 539.
PubMed Abstract | Publisher Full Text | Free Full Text

17. Nikbakht H, Xia X, Hickey DA: **The evolution of genomic GC content undergoes a rapid reversal within the genus *Plasmodium*.** *Genome.* 2014; **57**(9): 507–511.
PubMed Abstract | Publisher Full Text

18. Rosenthal PJ: **The interplay between drug resistance and fitness in malaria parasites.** *Mol Microbiol.* 2013; **89**(6): 1025–1038.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Referee Status: ? ?

Version 1

Referee Report 16 January 2017

doi:10.5256/f1000research.11044.r18806

? **Charles Woodrow**, **Ipsita Sinha**
Mahidol-Oxford Tropical Medicine Research Unit (MORU), Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

This is a secondary analysis of sequencing data produced by the MalariaGen community project. The authors describe a strong imbalance in SNPs at synonymous positions and less conserved non-synonymous positions, with GC to AT several times more common than AT to GC substitutions. If true, the core finding is of considerable interest.

As it stands, there is a substantial methodological issue relating to the assumption that the reference sequence is the ancestral one. For the result to be taken seriously by the wider community this central issue will needs to be addressed by the authors and then tested in future work by others. The authors attempt to discuss evolutionary significance of the finding but these sections are highly speculative. The authors also omit acknowledgment of possible weaknesses / limitations in their work.

Major issues:

1. Ancestral vs. derived alleles

   No attention is given to the critical issue of which are ancestral and derived alleles for each SNP and until this issue is attended to, the findings should be treated with caution. The paper does not explicitly state how the direction of each mutation was determined, but presumably they used the Sall reference sequence available at the MalariaGen website as the ancestral and the alternative allele as the derived allele. Sall is of course not an ancestral sequence but simply the complete sequence of a single isolate and hence the direction of some SNPs is bound to be misclassified.

   The authors need to consider and describe the extent to which this affects their overall results and conclusion. Then, all possible approaches to addressing and/or resolving the issue need to be explored – the onus is on the authors to subject their finding to the highest level of scrutiny and not simply assume it is correct because it falls out that way on first analysis. The paper describing the database (Pearson et al.) discusses the use of *P. cynomolgi* as outlier sequence and the definition of an ancestral sequence for around 30% of SNPs – the authors could focus on this subset of SNPs. Other approaches might include looking at SNPs of low frequency and / or private to one of the Asian populations since these are highly likely to be derived – however in theory this also increases the possibility of including SNPs that do not actually exist (false positives), a risk that needs to be made explicit. It might also be relevant to consider the subset of SNPs at Sall only (those with 100% frequency in the samples) – here the reference becomes the consensus

sequence of the entire sample dataset and Sall becomes the derived sequence.

2. Units and interpretation

   There is room for confusion in the way the term 'bias' is used. *P. falciparum* has a highly AT-biased genome but the AT to CG vs. CG to AT ratio is around 1 indicating that the *P. falciparum* SNPs are also heavily biased towards AT (put another way the system is at equilibrium). In contrast *P. vivax* has a relatively unbiased genome but SNPs that appear imbalanced towards AT. The SNP bias in the *P. vivax* population might therefore be roughly the same as for *P. falciparum* (so it is inappropriate to say that the patterns are divergent).

   It would be interesting to calculate the underlying SNP rate for A to G (and G to A, C to T and T to C) where the denominator is the starting nucleotides. In other words, what proportion of A nucleotides mutate to G, and vice-versa? One approach would be to focus on 4-fold synonymous sites, calculate the underlying nucleotide content of these 4-fold synonymous sites across the genome, and then determine the mutation rates for each of the 12 possible mutation directions using the underlying nucleotide content as denominator. This would also test whether the AT-bias in SNPs (if it exists) applies to transversions (currently ignored).

3. Secondary use of data

   The set of SNPs described in Pearson et al. have clearly undergone extensive filtering and quality control and the possibility that a substantial number are incorrect (or have been missed out) seems remote. Furthermore, the imbalance seems to disappear in non-synonymous positions – evidence against a substantial number of artefactual SNPs. Nevertheless any bias in terms of false positive or false negative SNPs could in theory generate the imbalance in SNPs observed and the authors, who did not generate the data, cannot simply assume that the SNPs are 100% accurate – since they cannot see the underlying sequence reads which produced them. At the very least the authors need to acknowledge this potential weakness and discuss explicitly their reasons for being confident in terms of the SNP calls.

Minor issues:

The authors suggest that *P. falciparum* and *P. vivax* have evolved in distinct ways to antimalarials but the mutations associated with antifolate resistance in the two organisms are highly analogous.

The authors attempt to link their (possible) finding on SNP bias to wider areas of malaria biology (lifecycle, drug resistance etc..) in the Introduction and Discussion. Currently these sections are highly speculative and not rooted in the extensive literature on this subject that has already been generated for other organisms. The starting point should be that codon bias is of no functional impact. If there are known examples where codon bias has a functional effect on an organism these need to be brought in – are there any examples where codon bias is thought to influence amino acid constitution (hydrophilicity / secondary structure) or generate repeat sequences (a particular feature of *P. falciparum*)? Differing codon biases might instead be a secondary consequence of the organism's biology given the link between population size / genetic drift and fixation of new mutations. The authors need to place their finding in the context of wider literature.

When discussing possible changes over time there are statements which lack basis. The authors suggest that the bias in *P. vivax* SNPs has changed recently; this is only one possible explanation for the possible

findings. All the authors can say is that there is an imbalance in SNPs compared to the underlying genome.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Referee Report 20 December 2016

doi:10.5256/f1000research.11044.r18003

Richard D. Pearson[1,2]
[1] Wellcome Trust Sanger Institute, Hinxton, UK
[2] MRC Centre for Genomics and Global Health, Wellcome Trust Centre for Human Genetics, Oxford, UK

Overview

This article reports a difference in the types of single nucleotide polymorphisms observed between the malaria parasites P. falciparum and P. vivax. Specifically, a higher rate of G/C to A/T substitutions is seen in P. vivax than in P. falciparum. The analysis is based on data produced by the Malaria Genomic Epidemiology Network (MalariaGEN, https://www.malariagen.net/).

The substitution bias reported in P. vivax is clearly present. This can be observed by using the MalariaGEN Data app at https://www.malariagen.net/apps/pvgv (select Variants tab, click Define query, click + and set Mutation Type Equals S, click OK, click Create plot then Bar graph, Group by: Reference Allele, Secondary group: Alternate Allele, click Create plot). It is clear that C->T and G->A mutations are much more common than other types of substitution. Using this app we can also see that the result continues to hold if we focus only on high frequency mutations (e.g. MAF All >= 0.25), though the difference is less dramatic for such variants. It can also be observed that the lack of such substitution bias is present in P. falciparum by using the MalariaGEN Data app for the Pf3k project at https://www.malariagen.net/apps/pf3k (select Variants 3.1 - Samtools Mpileup tab, click Define query, click + and set Is SNP Equals Yes, click + and set Is Coding Equals Yes, click OK, click Create plot then Bar graph, Group by: Reference Allele, Secondary group: Alternate Allele, click Create plot).

The substitution bias in P. vivax and lack of such bias in P. falciparum observed by the authors is an interesting result that warrants further explanation.

Major points

- The first and last sentence of the abstract concern responses to drug pressures. This leaves the impression that the manuscript will describe an important result regarding drug pressure. However, there is no evidence provided in the manuscript to support such a result. The hypothesis that the difference in genomic variation seen is due to drug pressure appears to be pure speculation. Unless the authors can provide any evidence that the patterns of genomic variation seen can attributed to drug pressure, such speculation should be removed from the abstract, and if mentioned at all, should only be in the discussion and made clear that this is an unsubstantiated

hypothesis.

- The title of the manuscript describes a "divergent" pattern of genomic variation in the two species. The word divergent suggests that the pattern of genomic variation might be growing apart over time between the two species in some way. However, if the main results holds, and other types of mutation are GC content-neutral, then over time it should be expected that the GC content of P. vivax will decrease, which in this sense would result in the P. vivax genome becoming more like that of P. falciparum, so if anything this might be described as a convergent rather than a divergent pattern. A more accurate description might be that there is a "different" pattern of genomic variation in the two species.

- The authors describe "similar drug pressure" in P. falciparum and P. vivax, but I think it is well established that the drug pressures are quite different in the two species, so such phrases should be removed.

- The authors describe a "recent reversal in the substitution bias in P. vivax". In order to make such a claim, the authors would first need to show that there has, until recent times, been an opposite substitution bias in P. vivax, but no such evidence is given, only evidence from others that the GC content of P. vivax might have increased since the common ancestor of P. vivax and P. falciparum, which is not the same as showing an historic substitution bias.

Specific points

- The final sentence of the abstract mentions "similar drug pressures" in P. falciparum and P. vivax. However, I think it is well established that the drug pressures are quite different in the two species, for example due to difference in gametocytogenesis and longevity of gametocytes as discussed in ref 11 by Schneider and Escalante. As stated above, speculation regarding drug pressure should be removed from the abstract.

- In the first paragraph of the Introduction, it is not stated that there is resistance to chloroquine in P. vivax in some parts of the world. A reader not aware of this might be confused to read the phrase "the mechanism of chloroquine resistance in P. vivax remains unknown", because the previous sentences give no indication that there is any such resistance. It might be helpful to the reader to mention that there is resistance to chloroquine in P. vivax in some parts of the world.

- The final sentence of the Introduction states that the substitution bias observed may be related to differential evolutionary response to the same anti-malarial drugs. No evidence is given in the remainder of the manuscript to say why this particular hypothesis should be preferred to any other potential hypothesis we might want to imagine. Also, it seems highly unlikely that the substitution bias observed might be related to the differential evolutionary response to the same anti-malarial drugs, as to date only a small number of genomic loci have been shown to be associated with drug resistance, whereas the pattern of substitution bias observed is (presumably) genome-wide.

- In the first paragraph of the results, the authors use the phrase "We tested whether this is true during the recent evolution of P. vivax". It is not entirely clear what "this" refers to here but presumably the authors are saying they have tested the hypothesis that P. vivax has increased its GC content during recent evolution. However, this is not what they have tested. Firstly, the authors have not apparently considered which is the ancestral and which is the derived allele. The SNPs used are differences from the 3D7 reference genome. The authors appear to have assumed that

the substitutions have been from the reference to the alternative allele, but it might be the case that for many of these SNPs, the alternative allele is ancestral and the 3D7 allele is derived. Secondly, it is possible that the GC content could be increasing even if substitutions are biased to decrease GC content, for example if other types of mutation such as short indels, larger structural variants and/or gene conversion result in an increase in GC content. The authors should either make it very clear that their results do not give any definitive conclusions as to the change in GC content over recent evolution, or else should perform a more thorough analysis taking into account the ancestral status of alleles and analysing non-SNP types of genomic variation.

- The first sentence of the discussion states that the present study finds a sharp recent reversal in the substitution bias in P. vivax. In order to make such a claim, the authors would first need to show that there has, until recent times, been an opposite substitution bias in P. vivax. The authors appear to be making the assumption that
a) the conclusion from reference 17 holds (i.e. that the GC content of P. vivax has increased since the common ancestor of P. vivax and P. falciparum) and that
b) this increase in GC content was due to substitutions.
However, even if a) is true, there is no evidence to support b), and it could be that the GC content has increased due to other types of mutation such as short indels, larger structural variants and/or gene conversion. The authors should remove any reference to reversal in substitution bias unless they can provide evidence of such a reversal.

- The fourth sentence of the discussion again mentions "similar drug pressures", but as above I think it is well established that the drug pressures are quite different between the species. This sentence should be removed.

- The final sentence of the first paragraph of the discussion is somewhat unclear, for example what exactly is meant by "these" factors? Also, is it really "likely" that these might contribute towards differences in drug resistance or is this simply "possible".

- The final paragraph again refers to a recent reversal in the substitution bias in P. vivax. Unless the authors can demonstrate that this is true, this sentence should be removed.

- It was not clear to me how the observation of lower substitution bias in non-coding regions compared to synonymous sites could be utilized to identify non-coding regions that are under higher functional constraint (final sentence). This should either be explained, or the sentence dropped.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.