The utility of genomic data for *Plasmodium vivax* population surveillance

Rachel F. Daniels^{1,4}, Benjamin L. Rice², Noah M. Daniels³, Sarah K. Volkman^{4,5}, Daniel L. Hartl²

¹Harvard University, Cambridge, MA, USA, ²Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA, ³Computer Science and Artificial Intelligence Laboratory and Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA, ⁴Immunology and Infectious Diseases, Harvard School of Public Health, Cambridge, MA, USA, ⁵School of Nursing and Health Sciences, Simmons College, Boston, MA, USA

Genetic polymorphisms identified from genomic sequencing can be used to track changes in parasite populations through time. Such tracking is particularly informative when applying control strategies and evaluating their effectiveness. Using genomic approaches may also enable improved ability to categorise populations and to stratify them according to the likely effectiveness of intervention. Clinical applications of genomic approaches also allow relapses to be classified according to reinfection or recrudescence. These tools can be used not only to assess the effectiveness of malaria interventions but also to appraise the strategies for malaria elimination.

Keywords: *Plasmodium vivax*, Malaria, Eradication, Elimination, Genotyping

Introduction

Plasmodium vivax is an important emerging tropical disease

Of the billions of dollars that have been allocated to research, control and eradication of malaria, the vast majority (97%) has been devoted to *Plasmodium falciparum*.¹ The rationale for the disproportionate allocation is that *P. falciparum* malaria remains the primary cause of global malaria mortality, especially in children under the age of five. Relatively little attention has been given to malaria due to *P. vivax*, even though its global distribution is substantially wider than that of *P. falciparum* and it has a staggering economic and public health impact owing to the long-term morbidity of relapsed infections.

Plasmodium vivax malaria occurs in Africa, Asia, Central America and South America, and while it may not be as lethal as *P. falciparum* malaria, it is a serious, deadly disease. The burden of *P. vivax* can be seen using the disability-adjusted life-year (DALY) statistic, a summary index that quantifies mortality and disability. In Asia and South America, the DALY reaches 100–500 per 10 000 people for *P. vivax*,^{1,2} a level less than, but approaching, that of *P. falciparum* malaria in sub-Saharan Africa (1000 per 10,000

people). Moreover, this does not include massive indirect effects such as chronic anaemia or impaired growth. The numbers tell the tale: *P. vivax* malaria imposes important social and economic costs among the world's poorest countries, affecting 80–390 million people annually at a cost of 1.4–4.0 billion dollars.³

Control of *P. vivax* malaria is hampered by the evolution of drug resistance in the parasite, the evolution of insecticide resistance in the mosquito vectors, the absence of an effective vaccine and the inability to establish long term *in vitro* cultures of the parasite in the laboratory. Another major hindrance is the limited understanding of the geographical distribution and spatial epidemiology of *P. vivax* malaria. Recent data have reduced gaps in geographical knowledge,¹ which opens new opportunities for rational strategies of control and elimination. This is particularly important because autochthonous *P. vivax* epidemics are appearing in regions that have been malaria-free for decades,^{4,5} possibly related to global climate change.

Our understanding of the genetics of the *P. vivax* parasite — and especially in the case of drug resistance, the parasite's genetics themselves — seem to be changing too. Three sources of evidence support this hypothesis:

- (1) Several groups have reported *P. vivax* malaria infections in Duffy-negative individuals. The Duffy-negative blood group is due to a mutation

Correspondence to: Rachel Fath Daniels, Harvard University, Cambridge, MA, USA. Email: rdaniels@broadinstitute.org

previously thought to prohibit *P. vivax* invasion of red blood cells. The mutation gives evidence of being under strong positive selection in humans, which is commonly attributed to its protective effects.^{6–12} Like the sickle cell mutation, the protective Duffy-negative mutation is most common in African countries. The spread of *P. vivax* parasites that can infect Duffy-negative human hosts greatly increases the number of people at risk for infection.

- (2) Of particular concern are indications that *P. vivax* malaria infections are not as benign as previously believed. Reports of severe and even fatal *P. vivax* malaria with symptoms comparable to severe *P. falciparum* malaria are increasing, having been reported in Sudan, Pakistan, Papua New Guinea and India.¹³
- (3) Further, there are observations of increasing resistance to the few drugs known to be effective for the eradication of systemic and latent infective forms and these are of increased concern in light of the widespread reports of severe *vivax* disease.^{14,15}

This accumulating body of evidence overwhelmingly emphasises that *P. vivax* should be regarded as an important emerging tropical disease.

It also draws attention to existing gaps in our understanding. Chief among the remaining gaps in *P. vivax* biology is a narrow and inadequate knowledge of genetic differences among geographical populations and the epidemiological implications of these genetic differences. For example, two critical parameters for *P. vivax* elimination are latency time and relapse frequency.¹⁶ Geographic variation in these parameters is well documented, but it is still not known to what extent they are affected by genetic differences among parasites.¹⁷

Plasmodium vivax genotype tools

Easy-to-use genotyping tools are needed at this stage to track the geographical distribution as well as the spatial and temporal epidemiology of parasite genotypes. Such *P. vivax* genotyping tools should be inexpensive and applicable to both field and laboratory derived material.

The rationale for needing such tools to assay genetic variants is clear. On the one hand, such genotyping tools have accelerated progress in understanding the biology and epidemiology of *P. falciparum* malaria, and this has contributed to development of new strategies for control and eradication. On the other hand, the lessons learned from *P. falciparum* are not directly applicable to *P. vivax* owing to great differences in the parasites' evolutionary history, course and clinical features of infection, liability to relapse or recrudescence, and responsiveness to measures of control and eradication. In regions endemic for both major malaria pathogens, the reduction of *P. falciparum* malaria with no concomitant reduction of *P. vivax* malaria underscores how *P. vivax* is biologically and epidemiologically distinct from *P. falciparum*. The critical implication is that

approaches developed for control of *P. falciparum* may not be effective against *P. vivax*. Hence proposed control programmes modelled on those for *P. falciparum* but aimed specifically at *P. vivax* might very well be ineffective, unless there are established benchmarks against which the results of these strategies can be tested as they are implemented. Furthermore, strategies that target *P. falciparum* but not *P. vivax* may in fact be creating a niche for *P. vivax* to thrive and in this way compromise malaria elimination efforts.

Fine-grained mapping of genetic variation in global and local populations will yield baseline data on the spatial distribution and transmission of *P. vivax*, and observed changes in response to intensified intervention will help evaluate and enhance the effectiveness of strategies for control and elimination.

Initial studies that have used large-scale genotyping to look at patterns of transmission have often focussed on a specific parameter of malaria transmission, the complexity of infection (COI). Complex *P. vivax* infections, where multiple genotypes of *P. vivax* infect a single individual, are common, ranging, for example, from 19.3 to 100% in a review of infections in Southeast Asia.¹⁸ The potential utility of this parameter and genomic data in general, as well as challenges in their analysis, are discussed in this review, especially with respect to the four specific knowledge gaps detailed below.

We highlight four critical gaps in our current understanding of *P. vivax* populations: (1) the distribution and epidemiological relevance of genetic diversity at different geographic scales, (2) how *P. vivax* populations respond to control efforts at the genetic level, (3) the rates of relapse and reinfection and how these parameters may vary between populations, and (4) an understanding of the extent of variation in the specific genes involved in *P. vivax* infection and transmission.

We find that, although challenges remain, including understanding the persistence of complex infections and their effects on relapse, recent progress demonstrates the utility of genomic data in efforts to control this emerging pathogen (see Fig. 1).

Current Approaches

Using genetic markers to monitor P. vivax populations

The sequence of the *P. falciparum* reference genome was first assembled in 2002,²⁷ catalysing research efforts to understand parasite biology. However, *P. vivax* was not fully sequenced until 2008.²⁸ The unique and sometimes frustrating biology of *P. vivax*, including the inability to establish long-term culture *in vitro*, accounts for the extended delay and the limited number of isolates subjected to high-throughput sequencing. The parasite's continued recalcitrance to be cultured for more than several cycles *in vitro* has induced

(1) Characterising geographic variation and population surveillance:

Understanding patterns of genetic variation at different geographic scales and genetic differences between populations allows inferences of population size, connectivity between regions or structure between populations. While global genetic diversity has been found to be very high in comparison to *P. falciparum*,¹⁹ conflicting signatures have been found at the local scale and in near-elimination scenarios where *P. vivax* effective population size is thought to be small.²⁰

(2) Monitoring the response of *P. vivax* to control efforts

New tools are needed to track populations as programmes re-orient towards elimination. Genotyping data at multiple loci allows estimating the COI, a parameter reflecting the number of distinct parasite genotypes present and informative of population size. While this has been successfully applied in *P. falciparum* for surveillance of control efforts in Senegal,²¹ new tools recently available for *P. vivax* have yet to be widely applied.²²

(3) Distinguishing relapse from reinfection

Recurrent *P. vivax* infections can be the result of a new infection or activation of latent parasite clones persisting in the hypnozoite stage in the liver. Genotyping data allows one to distinguish between these possibilities,²³ and examine the patterns of genetic diversity among *P. vivax* clones present in a relapsed infection. Such parameters are critical to understanding *P. vivax* transmission dynamics²⁴ and have implications for drug resistance monitoring,²⁵ though challenges remain.

(4) Gaining a better understanding of the invasion biology of *P. vivax*

Genomic data can also be utilised to understand the genes that permit successful parasite invasion; there is also a need to monitor diversity and changes in invasion efficiency. Genetic data have been used to monitor loci associated with drug resistance²⁶ and the genes encoding antigens used by the parasite to invade the red blood cell.¹⁹ Increased reports of *P. vivax* infection of Duffy-negative individuals and putatively associated genomic changes in the parasite population are discussed.⁶

Figure 1 Applications of genomic data to the surveillance of *P. vivax* populations.

development of novel methods to enrich parasite content in patient-derived samples, including whole genome capture²⁹ and filtration of patient samples followed by short-term *ex vivo* culture.³⁰ These methods offer opportunity for increased numbers of parasite genomes to be sequenced, which will improve efforts to identify polymorphisms useful for population analysis.

Since publication of the reference *P. vivax* genome,²⁷ additional isolates from a few geographic origins have also been sequenced, revealing a genome much more genetically diverse than that of *P. falciparum*.^{19,31} The diversity includes antigenic variation, repeat-number variation in microsatellites, gene copy-number variation (CNV), and single-nucleotide polymorphisms (SNPs). In the long run, one would hope for a sufficient number of genomic sequences from geographically distinct populations to be available to identify sets of genetic variants

that are highly informative for population genomics at least on a continental scale, but preferably on the much smaller scale of countries or even subdivisions within.

Polymorphic genetic markers are useful to characterise *P. vivax* populations in several ways. First, genetic markers under selection, such as the parasite surface antigens CSP and the MSPs,^{32–34} can be used to determine relationships between immunity and transmission. These relationships are especially important in relation to malaria vaccine trials and implementation. Genetic markers associated with reduced drug sensitivity can also be used to infer the evolutionary history of selection events, detect emergence of new mutations, and track their spread across the globe. Surveillance of these markers allows control programmes to better tailor anti-malarial drug regimens and to respond to new resistance threats before they are introduced by migration

or emerge *de novo* in populations. However, because identity in kind (for example, identical *csp* or *msp* profiles) does not necessarily imply identity by descent,³⁵ alleles such as microsatellites or antigens classified by electrophoretic mobility are of limited use for identification of individual parasite lineages.³⁶

Selectively neutral markers can be used as measures of genetic diversity to describe parasite population structure. Putatively neutral markers include microsatellites, which in *P. vivax* have been used nearly exclusively for population analysis. A recent meta-analysis reviews many of these findings in an attempt to sort various reported microsatellite markers by informativeness for particular research questions.³⁷

However, other studies have cautioned against the universal interpretation of these markers as selectively neutral.³⁸ While informative for short-term studies within the same season or year, these markers also have the limitation of being unable to distinguish identity in kind from identity by descent, making them of limited use for long-term temporal studies such as those required to study the effects of vaccines and other control efforts.

More commonly used in studies of *P. falciparum*, putatively neutral single nucleotide polymorphisms (SNPs) offer the potential not only to describe population structure and responses to control efforts^{39–43} but can also offer the ability to identify individual parasites lineages and track them through space and time.^{21,44}

The limited number of SNP-based studies in *P. vivax* so far published have generally focussed on restricted genomic regions^{45–47}; however, a recent publication²² describes a *P. vivax* molecular barcode similar in concept to the useful *P. falciparum* barcode,⁴⁴ which offers opportunities for field-based studies of *P. vivax* population structure.

With sufficient data from a variety of populations, one can estimate genetic diversity between and among populations. Polymorphic SNP markers also reveal the extent of linkage disequilibrium (LD), or the nonrandom association of alleles of linked genes. Further, these variable markers provide useful tools with which to estimate the COI, a metric of the number of different genomes present in a patient sample. Amplicon sequencing approaches can also be used to interpret individual parasite types among polygenomic infections.²³

Overall, genetic markers help characterise individual populations and can reveal genetic relatedness between populations. The combined knowledge of population structure and parasite relatedness within geographical areas allows identification of parasite lineages and tracking of their movement within and beyond these regions. This knowledge will result in improved predictions of the likelihood of successful elimination,⁴⁸ and thus better prioritisation of limited resources.

Monitoring progress towards elimination

Understanding changes in population structure is of increasing importance as countries and regions move from control towards pre-elimination status. Under these conditions, the number of reported cases is considerably reduced and new surveillance strategies must be implemented. Rapid diagnostic tests (RDTs), the gold standard for malaria detection, require a minimum patient parasitaemia. When treatment programmes have sufficiently reduced the burden of malaria, the number of false-positive instances is expected to increase. Polymerase chain reaction (PCR) based methods of detection are an option with increased detection sensitivity, but thus far, implementation as part of control and elimination efforts has not been adequate to assess the true prevalence of the parasite in low-infection settings. In order to prepare for programme reorientations with increasing malaria control, it is essential to track progress towards elimination. This creates the need for alternative methods to monitor parasite population size as elimination is approached.

One such way to do this is to use the estimates of COI derived from assaying polymorphic markers across the parasite genome. In regions of high parasite transmission, it is common to find evidence for multiple parasite genomes within a single patient (high COI). In contrast, in areas with very low reported transmission, such as areas well into pre-elimination or experiencing imported or other small-scale epidemics, either a single or limited number of genomes are present (low COI). Therefore, COI from genomic data can be used as an indicator of *P. vivax* population status and aid in surveillance when RDT sensitivity and PCR-based detection are inadequate. However, a remaining challenge is that the rate and pattern of change in COI between these population states is unclear and thus far unreported. It is also at present unclear how the number of genomes within patients (monogenomic versus polygenomic infections and relative COI) affects *P. vivax* population structure.

In theory, as transmission decreases, the overall genetic diversity will also decrease as opportunities for meiotic crossover events between genetically distinct parasites within the mosquito midgut become more rare. The decrease in effective recombination will be coincident with a reduced complexity of infection. With fewer parasite types in the population, crossover events will often occur between identical or nearly identical individuals, increasing the level of inbreeding in the population.

However, several groups have reported continued transmission of polygenomic *P. falciparum* malaria infections even in low transmission areas, hinting that the mechanisms for parasite transmission may be complicated by the maintenance of mixed gamete

transmission within mosquitoes.^{49,50} A major differentiating biological feature between *P. vivax* and *P. falciparum* is relapse of *vivax* parasites, which can contribute to elevated COIs not associated with new infections. Indeed, a study of *P. vivax* in Sri Lanka reported a puzzling persistence of genetic diversity and a high proportion of complex infections even as prevalence dramatically declined due to control efforts.²⁰ The continued elevated COI could confound efforts to determine eradication status by COI alone, making efforts to overcome the current analysis challenges of even greater importance.

Relapse and reinfection

Another application of characterising within-host genetic diversity has been to analyse recurring infections in a patient to estimate the rates of relapse and reinfection. Relapse plays an important role in *P. vivax* transmission dynamics,²⁴ in initiating the transmission season in seasonal transmission settings (such as Southern Mexico⁵¹), contributes to high levels of genetic diversity in populations, and is responsible for a high percentage of *P. vivax* cases, for example more than 50% in a cohort in Cambodia.²³ It also confounds monitoring for treatment failures due to resistance, as the reappearance of blood stage infection can also be due to the emergence of previously latent parasites.²⁵ Relapse parameters, such as the proportion of infections that result in relapse, the period of time to first relapse, and the time between subsequent relapses, are therefore critical to understanding *P. vivax* epidemiology.

However, it is well documented that these parameters differ across the range of *P. vivax*. There is evidence that this geographical variation is in part due to historical adaptation to transmission in different climates,¹⁷ suggesting that some of the genetic variation distinguishing *P. vivax* populations underlies the variation in relapse phenotypes that has been observed. However, little work aimed at identifying the genetic basis of this variance or how high global levels of genetic diversity affect the parasite's adaptability has been performed to date.

Instead, much of the recent work has focussed on understanding the surprising amount of *P. vivax* genetic diversity that is observed in a relapsed infection. Three principal explanations for the high proportion of relapses that have complex infections containing heterologous or novel genotypes have been investigated with genomic or genotyping data.²³

The first possible explanation is that a long history of *P. vivax* exposure in endemic areas causes a reservoir of diverse hypnozoites to accumulate in the liver, from which a variable subset of clones activates upon relapse. The genotypes present in the relapsed infection therefore can be diverse and dissimilar from

the primary infection observed when study began. Such was the conclusion from genotyping a sample of 58 paired clinical isolates in Northwest Colombia⁵² as well as in multiple other studies. The second explanation, minority variant expansion, has been observed when using high-throughput DNA sequencing techniques to attain high sequencing coverage of a polymorphic amplicon.²³ Here, variants in a complex infection that are below the detection threshold of less sensitive methods such as microsatellite typing go unobserved in the primary infection, but are more abundant or predominant in later recurrences. Finally, recombination in the mosquito midgut prior to inoculation can result in an infection comprised of differing, but closely related relatives. Full genome sequencing indicated that the circulating genotypes in three relapses of a Sudanese patient who had travelled to a non-endemic country after primary infection likely originated from a single meiosis event.⁵³

While the three possible processes have all found support, confidently classifying recurrent infections as a relapse remains challenging. Major limitations are that accurately capturing the diversity of primary infections and an understanding of the distribution of genetic diversity in the *P. vivax* population are required to rule out reinfection. This again highlights the need for characterising the complexity of infection and for fine-grained mapping of genetic variation in local populations as discussed above.

Such an understanding of the levels of population diversity and how it is distributed would allow genome wide association studies to begin searching for the loci responsible for relapse phenotypes.⁴⁷ This is an intriguing opportunity for the application of genomics to provide much needed information.

Epidemiologically relevant *P. vivax* invasion biology

A fourth topic of current genomic research is aimed at improving our understanding of variation in the regions of the *P. vivax* genome that encode specific proteins relevant to infection and transmission. These include antigens involved in the binding and invasion of host cells,^{6,54} potential vaccine candidates,⁵⁵ and loci associated with drug resistance.²⁶

Initial comparisons of *P. vivax* genomes have found high levels of variation in many of the antigens involved in red blood cell invasion,¹⁹ and even suggestions of common copy number variation in antigen families such as *vir* and MSP3.⁵⁴ This indicates that different subpopulations of *P. vivax* could present very different antigenic sequences and even have different repertoires of antigens, with clear relevance for immunity and vaccine development.

Additionally, genome wide screens for genes with a pattern of polymorphism consistent with the action of natural selection have identified new loci that are of interest in understanding *P. vivax* invasion biology. As an example, a study by Escalante *et al.* (2014) identified 87 such genes, including a surface antigen in the MSP7 family.⁵⁶ These studies are limited, however, by both the small number of isolates with genomic sequence data available and the inherent difficulty in assembling and analysing very diverse regions of the genome. Especially the latter difficulty may require more targeted genomic approaches directly aimed at capturing the diversity in these specific regions.

An interesting example of this approach can be seen in a study of the *P. vivax* Duffy binding protein (DBP) from sequencing full genomes of field isolates from Madagascar.⁶ Reports that the Duffy-negative phenotype does not provide a barrier to *P. vivax* infection has raised concerns of an increased spread in Sub-Saharan Africa, the region already burdened with the highest *P. falciparum* morbidity.^{8,10} The reports also indicate that the well understood Duffy-mediated invasion pathway of *P. vivax* is an incomplete description of the parasite's invasion biology. Sequencing full genomes from Madagascar, a country where *P. vivax* infections of Duffy-negative individuals were previously found to be common,¹⁰ identified several genetic differences, including a duplication containing the gene encoding the parasite protein that binds to the host's Duffy antigen. The peak prevalence of the DBP duplication coincided with the area of Madagascar with the highest prevalence of *P. vivax* infections of Duffy-negative people. The duplication was also widely distributed in *vivax* patients in other regions of Africa and in Asia, though absent from Papua New Guinea, French Guiana and Turkey. Little sequence diversity was found in the DBP duplication, however, indicative of a recent spread.

The spread and variable distribution of the DBP duplication putatively associated with Duffy-independent invasion reaffirms the need to further understand spatial heterogeneity in *P. vivax* populations. Duffy-negative infection and the DBP duplication also demonstrate that genomic studies of *P. vivax* are useful in investigations aimed at relating that heterogeneity to clinically and epidemiologically relevant phenotypes.

Analysis Challenges

Estimating the number of clones in polygenomic infections

Like whole-genome sequencing, genotyping large numbers of samples presents challenges for data analysis. The biology of the parasite inherently

complicates both whole-genome and genotyping analysis. In *P. vivax*, a high proportion of polygenomic samples (those with more than one parasite genome present in a single sample) may arise from repeated cycles of recrudescence or high entomological inoculation rates in areas with high transmission.¹⁸ While multiple genotyping methods such as digital PCR⁵⁷ or deep sequencing of SNP markers and highly polymorphic amplicons can provide relative frequencies for sets of SNPs, high infection polyclonality make it difficult to assign individual identity to the minimum number of genomes present in a sample. However, computational methods applied to SNP profiles or barcodes may allow maximum likelihood estimates to be made of the number of distinct lineages in individual polygenomic infections and in turn that could be used to characterise parasite populations.

Deep sequencing of a single or limited number of highly polymorphic amplicons may offer one method for imputing the number of genomes present as a fraction of the total reads.⁵⁸

Another approach to estimating complexity of infection uses greedy algorithms, which at each step choose a local optimum estimate, combined with linear regression to infer the ratios of alleles and, consequently, the minimum ratio of genomes within a single sample. Using this estimator of allelic ratios, the next step is to infer the number and ratio of strains within a sample that best explains the data. One possible approach is matching pursuit,⁵⁹ an algorithm that identifies the minimal set of functions from a 'dictionary' to explain a given signal. In this instance, the 'dictionary' is the set of all possible genomes based on the SNPs interrogated (strains), and the signal is the set of allelic ratios over the entire set of SNPs. This approach relies on a termination threshold ϵ ; it does not report any strain that might be present in a ratio smaller than ϵ .

Finally, COI can be estimated from a set of SNP markers using COIL.⁶⁰ This approach uses a Bayesian approach to estimate the likelihood of distinct COI based on molecular barcode data.⁶⁰

Identification and application of markers of the status of these populations will enable a better understanding of the changing population genetics and ultimately lead to better strategies and more effective distribution of limited resources in the global effort to eradicate malaria.

Data Accessibility and Portability

In order for the research community to better compare disparate populations, it is essential to share not only results but also genotyping data. Building a shared source of data combined with simple and accessible

Table 1 Select new bioinformatics tools used in malaria genotyping

Programme	Function	Current location	Reference
(1) COIL	Bayesian/estimate/of/the/complexity/of/infection/from/SNP/markers	broadinstitute.org/infect/malaria/coil/	60
(2) LookSeq	Alignment/and/analysis/of/whole/genome/sequences	sanger.ac.uk/resources/software/lookseq/	64
(3) EuPathDB	Genome/sequence,/polymorphism,/and/gene/expression/database	eupathdb.org/	62
(4) Galaxy/ToolShed	A/rich/depository/of/free/population/genetic/analysis/tools	toolshed.g2.bx.psu.edu	63

analysis tools will provide researchers across the globe access to a standardised set of tools and language for making comparisons.

Several of these tools have already been developed (see Table 1). LookSeq, from the Sanger Institute, allows users to align and analyse whole genome sequences,⁶¹ while EupathDB offers access to a variety of study findings, including gene expression, whole genome sequencing, and genotyping data.⁶² Future software releases of EupathDB are expected to incorporate solutions such as Galaxy ToolShed⁶³ to offer a suite of population genetics analysis tools. Currently hosted at the Broad Institute, COIL may also be linked from a common analysis site.

Conclusion

Plasmodium vivax is a relatively neglected tropical disease as evidenced by the fact that it only receives about 3% of the expenditures on malaria worldwide. Yet *vivax* malaria is becoming increasingly important even as the campaign to eliminate *falciparum* malaria shows promising early results.

Chief among the remaining gaps in *P. vivax* biology is a narrow and inadequate knowledge of genetic differences within or between geographical populations and the epidemiological implications of these genetic differences. Although the lessons learned from *P. falciparum* can inform some of these processes, they are not always directly applicable to *P. vivax*, owing to great differences in the parasites' evolutionary history, course and clinical features of infection, liability to relapse or recrudescence, and responsiveness to measures of control and eradication. Foremost among these differences is the ability of *P. vivax* infections to relapse. Relapse can contribute to greater complexity of infection and higher levels of population diversity, but much remains to be known about its variability and distribution. Genotyping tools allow new investigation of relapse parameters, which when coupled to an increased understanding of the geographic distribution of genetic differences will allow exciting new avenues of research into the genetic basis of relevant *P. vivax* phenotypes such as Duffy-independent invasion and relapse phenotypes themselves.

What is needed at this stage of *P. vivax* research are genotyping tools to track geographical distribution and spatial/temporal epidemiology of parasite genotypes as well as concerted efforts to rapidly disseminate data and standardise analysis methodologies to develop a common language for comparative analysis across wide geographical settings. Application of these tools to larger data sets will allow us to leverage limited funds to best determine the most effective strategies for *P. vivax* control and elimination.

Disclaimer Statements

Contributors All authors contributed to the writing of this manuscript.

Funding Gates Foundation.

Conflicts of interest The authors have no conflict of interest.

Ethics approval Ethical approval was not required for this manuscript.

Acknowledgements

Sarah K. Volkman and Rachel F. Daniels supported by The Gates Foundation. Benjamin L. Rice and Daniel L. Hartl supported by NIH grant AI106734. Noah Manus Daniels supported by NIH grant GM108348.

References

- 1 Gething PW, Elyazar IR, Moyes CL, Smith DL, Battle KE, Guerra CA, Moore RG *et al.* A long neglected world malaria map: *Plasmodium vivax* endemicity in 2010. *PLoS Negl Trop Dis.* 2012;6:e1814.
- 2 World Health Organization. *World malaria report.* 246. Geneva: World Health Organization; 2011.
- 3 The other malaria. *The Economist.* 2011;:1–4. Available from: <<http://www.economist.com/node/18741412>>.
- 4 Klein TA, Kim H-C, Lee J-J, Rueda LM, Sattobongkot J, Moore RG, *et al.* *Reemergence, Persistence, and Surveillance of Vivax Malaria and Its Vectors in the Republic of Korea.* 2008.
- 5 Ioannidis A, Nicolaou C, Stoupi A, Kossyvakis A, Matsoukas P, Liakata MV, *et al.* First report of a phylogenetic analysis of an autochthonous *Plasmodium vivax* strain isolated from a malaria case in East Attica. Greece. *Malar J.* 2013;12:299.
- 6 Menard D, Chan ER, Benedet C, Ratsimbaoa A, Kim S, Chim P, *et al.* Whole genome sequencing of field isolates reveals a common duplication of the Duffy binding protein gene in Malagasy *Plasmodium vivax* strains. *PLoS Negl Trop Dis.* 2013;7:e2489.
- 7 Zimmerman PA, Ferreira MU, Howes RE, Mercereau-Puijalon O. Red blood cell polymorphism and susceptibility to *Plasmodium vivax*. *Adv Parasitol.* 2013;81:27–76.

- 8 Mendes C, Dias F, Figueiredo J, Mora VG, Cano J, de Sousa B, *et al*. Duffy negative antigen is no longer a barrier to *Plasmodium vivax* – molecular evidences from the African West Coast (Angola and Equatorial Guinea). *PLoS Negl Trop Dis*. 2011;5:e1192.
- 9 Ryan JR, Stoute JA, Amon J, Dunton RF, Mtalib R, Koros J. Evidence for transmission of *Plasmodium vivax* among a duffy antigen negative population in Western Kenya. *Am J Trop Med Hyg*. 2006;75:575–81.
- 10 Ménard D, Barnadas C, Bouchier C, Henry-Halldin C, Gray LR, Ratsimbao A. *Plasmodium vivax* clinical malaria is commonly observed in Duffy-negative Malagasy people. *Proc Natl Acad Sci U S A*. 2010;107:5967–71.
- 11 Wurtz N, Mint Lekweiry K, Bogreau H, Pradines B, Rogier C, Ould Mohamed Salem Boukhary A, *et al*. *Vivax* malaria in Mauritania includes infection of a Duffy-negative individual. *Malar J*. 2011;10:336–6.
- 12 Woldearegai TG, Kremsner PG, Kun JFJ, Mordmüller B. *Plasmodium vivax* malaria in Duffy-negative individuals from Ethiopia. *Trans R Soc Trop Med Hyg*. 2013;107:328–31.
- 13 Abdallah TM, Abdeen MT, Ahmed IS, Hamdan HZ, Magzoub M, Adam I. Severe *Plasmodium falciparum* and *Plasmodium vivax* malaria among adults at Kassala Hospital, eastern Sudan. *Malar J*. 2013;12:148–8.
- 14 Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M. Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Med*. 2008;5:e128–8.
- 15 Baird JK. Chloroquine resistance in *Plasmodium vivax*. *Antimicrob Agents Chemother*. 2004;48:4075–83.
- 16 Roy M, Bouma MJ, Ionides EL, Dhiman RC, Pascual M. The potential elimination of *Plasmodium vivax* malaria by relapse treatment: insights from a transmission model and surveillance data from NW India. *PLoS Negl Trop Dis*. 2013;7:e1979–9.
- 17 Battle KE, Karhunen MS, Bhatt S, Gething PW, Howes RE, Golding N, *et al*. Geographical variation in *Plasmodium vivax* relapse. *Malar J*. 2014;13:144.
- 18 Arnott A, Barry AE, Reeder JC. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malar J*. 2012;11:14–14.
- 19 Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat Genet*. 2012;44:1046–50.
- 20 Gunawardena S, Ferreira MU, Kapilana GMG, Wirth DF, Karunaweera ND. The Sri Lankan paradox: high genetic diversity in *Plasmodium vivax* populations despite decreasing levels of malaria transmission. *Parasitology*. 2014;141:880–90.
- 21 Daniels R, Chang HH, Sène PD, Park DC, Neafsey DE, Schaffner SF, *et al*. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS One*. 2013;8:e60780–0.
- 22 Baniecki ML, Faust AL, Schaffner SF, Park DJ, Galinsky K, Daniels RF, *et al*. Development of a single nucleotide polymorphism barcode to genotype *Plasmodium vivax* infections. *PLoS Negl Trop Dis*. 2015;9(3):e0003539.
- 23 Lin JT, Hathaway NJ, Saunders DL, Lon C, Balasubramanian S, Kharabara O, *et al*. Using amplicon deep sequencing to detect genetic signatures of *Plasmodium vivax* relapse. *J Infect*. 2015; pii:jiv142 [Epub ahead of print].
- 24 Lover AA, Zhao X, Gao Z, Coker RJ, Cook AR. The distribution of incubation and relapse times in experimental human infections with the malaria parasite *Plasmodium vivax*. *BMC Infect Dis*. 2014;14:539.
- 25 Price RN, von Seidlein L, Valecha N, Nosten F, Baird JK, White NJ. Global extent of chloroquine-resistant *Plasmodium vivax*: a systematic review and meta-analysis. *Lancet Infect Dis*. 2014;14:982–91.
- 26 Lin JT, Patel JC, Kharabara O, Sattabongkot J, Muth S, Ubalee R, *et al*. *Plasmodium vivax* isolates from Cambodia and Thailand show high genetic complexity and distinct patterns of *P. vivax* multidrug resistance gene 1 (*pvmr1*) polymorphisms. *Am J Trop Med Hyg*. 2013;88:1116–23.
- 27 Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 2002;419(6906):498–511.
- 28 Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, *et al*. Comparative genomics of the neglected human malaria parasite *Plasmodium vivax*. *Nature*. 2008;455:757–63.
- 29 Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics*. 2012;13:262.
- 30 Auburn S, Marfurt J, Maslen G, Campino S, Ruano Rubio V, Manske M, *et al*. Effective preparation of *Plasmodium vivax* field isolates for high-throughput whole genome sequencing. *PLoS One*. 2013;8:e53160–0.
- 31 Chan ER, Menard D, David PH, Ratsimbao A, Kim S, Chim P, *et al*. Whole genome sequencing of field isolates provides robust characterization of genetic diversity in *Plasmodium vivax*. *PLoS Negl Trop Dis*. 2012;6:e1811.
- 32 Henry-Halldin CN, Sepe D, Susapu M, McNamara DT, Bockarie M, King CL. High-throughput molecular diagnosis of circumsporozoite variants VK210 and VK247 detects complex *Plasmodium vivax* infections in malaria endemic populations in Papua New Guinea. *Infect Genet Evol*. 2011;11:391–8.
- 33 Koepfli C, Ross A, Kiniboro B, Smith TA, Zimmerman PA, Siba P, *et al*. Multiplicity and diversity of *Plasmodium vivax* infections in a highly endemic region in Papua New Guinea. *PLoS Negl Trop Dis*. 2011;5:e1424.
- 34 Putaporntip C, Miao J, Kuamsab N, Sattabongkot J, Sirichaisinthop J, Jongwutiwes S, *et al*. The *Plasmodium vivax* merozoite surface protein 3 β sequence reveals contrasting parasite populations in southern and northwestern Thailand. *PLoS Negl Trop Dis*. 2014;8:e3336.
- 35 Rice BL, Acosta MM, Pacheco MA, Escalante AA. Merozoite surface protein-3 alpha as a genetic marker for epidemiologic studies in *Plasmodium vivax*: a cautionary note. *Malar J*. 2013;12:288.
- 36 Brito CF, Ferreira MU. Molecular markers and genetic diversity of *Plasmodium vivax*. *Mem Inst Oswaldo Cruz*. 2011;106(Suppl 1):12–26.
- 37 Sutton PL. A call to arms: on refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. *Malar J*. 2013;12:447.
- 38 Rouse P, Mkulama MA, Thuma PE, Mharakurwa S. Distinction of *Plasmodium falciparum* recrudescence and re-infection by MSP2 genotyping: a caution about unstandardized classification criteria. *Malar J*. 2008;7:185.
- 39 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(2):273–85.
- 40 Tan JC, Tan A, Checkley L, Honsa CM, Ferdig MT. Variable numbers of tandem repeats in *Plasmodium falciparum* genes. *J Mol Evol*. 2010;71:268–78.
- 41 Campino S, Auburn S, Kivinen K, Zongo I, Ouedraogo JB, Mangano V, *et al*. Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. *PLoS One*. 2011;6:e20251–1.
- 42 Dharia NV, Sidhu AB, Cassera MB, Westenberger SJ, Bopp SE, Eastman RT, *et al*. Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol*. 2009;10:R21.
- 43 Jacob CG, Tan JC, Miller BA, Tan A, Takala-Harrison S, Ferdig MT, *et al*. A microarray platform and novel SNP calling algorithm to evaluate *Plasmodium falciparum* field samples of low DNA quantity. *BMC Genomics*. 2014;15:719–9.
- 44 Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar J*. 2008;7:223.
- 45 Taylor JE, Pacheco MA, Bacon DJ, Beg MA, Machado RL, Fairhurst RM, *et al*. The evolutionary history of *Plasmodium vivax* as inferred from mitochondrial genomes: parasite genetic diversity in the Americas. *Mol Biol Evol*. 2013;30:2050–64.
- 46 Feng X, Carlton JM, Joy DA, Mu J, Furuya T, Suh BB, *et al*. Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proc Natl Acad Sci U S A*. 2003;100:8502–7.
- 47 Orjuela-Sánchez P, Karunaweera ND, da Silva-Nunes M, da Silva NS, Scopel KK, Gonçalves RM, *et al*. Single-nucleotide polymorphism, linkage disequilibrium and geographic structure in the malaria parasite *Plasmodium vivax*: prospects for genome-wide association studies. *BMC Genet*. 2010;11:65–5.
- 48 Moonen B, Cohen JM, Snow RW, Slutsker L, Drakeley C, Smith DL, *et al*. Operational strategies to achieve and maintain malaria elimination. *Lancet*. 2010;376:1592–603.
- 49 Echeverry DF, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ. Long term persistence of clonal malaria parasite *Plasmodium falciparum* lineages in the Colombian Pacific region. *BMC Genet*. 2013;14:2.

- 50 Branch OH, Sutton PL, Barnes C, Castro JC, Hussin J, Awadalla P. *Plasmodium falciparum* genetic diversity maintained and amplified over 5 years of a low transmission endemic in the Peruvian Amazon. *Mol Biol Evol.* 2011;28:1973–86.
- 51 Gonzalez-Ceron L, Mu J, Santillán F, Joy D, Sandoval MA, Camas G. Molecular and epidemiological characterization of *Plasmodium vivax* recurrent infections in southern Mexico. *Parasit Vectors.* 2013;6:109.
- 52 Restrepo E, Imwong M, Rojas W, Carmona-Fonseca J, Maestre A. High genetic polymorphism of relapsing *P. vivax* isolates in northwest Colombia. *Acta Trop.* 2011;119:23–9.
- 53 Bright AT, Manary MJ, Tewhey R, Arango EM, Wang T, Schork NJ, *et al.* A high resolution case study of a patient with recurrent *Plasmodium vivax* infections shows that relapses were caused by meiotic siblings. *PLoS Negl Trop Dis.* 2014;8:e2882.
- 54 Rice BL, Acosta MM, Pacheco MA, Carlton JM, Barnwell JW, Escalante AA. The origin and diversification of the merozoite surface protein 3 (msp3) multi-gene family in *Plasmodium vivax* and related parasites. *Mol Phylogenet Evol.* 2014;78:172–84.
- 55 Parobek CM, Bailey JA, Hathaway NJ, Socheat D, Rogers WO, Juliano JJ. Differing patterns of selection and geospatial genetic diversity within two leading *Plasmodium vivax* candidate vaccine antigens. *PLoS Negl Trop Dis.* 2014;8:e2796.
- 56 Cornejo OE, Fisher D, Escalante AA. Genome-wide patterns of genetic polymorphism and signatures of selection in *Plasmodium vivax*. *Genome Biol Evol.* 2015;7:106–19.
- 57 Johnston WT, Mutalima N, Sun D, Emmanuel B, Bhatia K, Aka P. Relationship between *Plasmodium falciparum* malaria prevalence, genetic diversity and endemic Burkitt lymphoma in Malawi. *Sci Rep.* 2014;4:3741–1.
- 58 Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D. Use of massively parallel pyrosequencing to evaluate the diversity of and selection on *Plasmodium falciparum* csp T-cell epitopes in Lilongwe, Malawi. *J Infect Dis.* 2012;206:580–7.
- 59 Mallat SG, Zhifeng Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans Signal Process.* 1993;41:3397–415.
- 60 Galinsky K, Valim C, Salmier A, de Thoisy B, Musset L, Legrand E, *et al.* COIL: a methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data. *Malar J.* 2015;14(1):4.
- 61 Manske HM, Kwiatkowski DP. LookSeq: a browser-based viewer for deep sequencing data. *Genome Res.* 2009;19:2125–32.
- 62 Aurrecoechea C, Heiges M, Wang H, Wang Z, Fischer S, Rhodes P. ApiDB: integrated resources for the apicomplexan bioinformatics resource center. *Nucleic Acids Res.* 2007;35:D427–30.
- 63 Blankenberg D, Von Kuster G, Bouvier E, Baker D, Afgan E, Stoler N, *et al.* Dissemination of scientific software with Galaxy ToolShed. *Genome Biol.* 2014;15:403–3.
- 64 Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature.* 2012;487:375–9.