Malaria Genomics in the Era of Eradication

Daniel E. Neafsey¹ and Sarah K. Volkman^{2,3,4}

¹Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142

²Department of Immunology and Infectious Disease, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115

³Infectious Disease Initiative, Broad Institute of MIT and Harvard, Cambridge Massachusetts 02142

⁴School of Nursing and Health Sciences, Simmons College, Boston, MA 02115

Correspondence: neafsey@broadinstitute.org; svolkman@hsph.harvard.edu

The first reference genome assembly for the *Plasmodium falciparum* malaria parasite was completed over a decade ago, and the impact of this and other genomic resources on malaria research has been significant. Genomic resources for other malaria parasites are being established, even as *P. falciparum* continues to be the focus of development of new genomic methods and applications. Here we review the impact and applications of genomic data on malaria research, and discuss future needs and directions as genomic data generation becomes less expensive and more decentralized. Specifically, we focus on how population genomic strategies can be utilized to advance the malaria eradication agenda.

The past decade has been marked by tremendous cost reductions for generating genomic data, coupled with considerable knowledge increases required for analysis and interpretation of these genomic data. Consequently, thousands of parasite and vector genomes have been sequenced since the original *Plasmodium falciparum* reference genome assembly in 2002 (Gardner et al. 2002). The maturation of next-generation sequencing (NGS) technology has reduced the cost of sequencing a malaria parasite genome from millions to tens of dollars. The challenge now is how to interpret these genomic data to inform important biological or operational questions, and to prospectively gen-

erate genomic data in a strategic manner. Some of the ways genomic approaches can guide or advance the eradication agenda include addressing questions such as: Are intervention approaches working? Where are new infections coming from? Are intervention approaches like drugs and vaccines inducing resistance in parasite populations or only working on a subset of the population? In this review, we examine the application of genomic data toward understanding and eradicating malaria, and outline opportunities for extracting even more value from such data through defined sampling strategies and collection of relevant metadata and phenotypes.

Editors: Dyann F. Wirth and Pedro L. Alonso

Additional Perspectives on Malaria available at www.perspectivesinmedicine.org

Copyright © 2017 Cold Spring Harbor Laboratory Press; all rights reserved; doi: 10.1101/cshperspect.a025544

Cite this article as *Cold Spring Harb Perspect Med* 2017;7:a025544

D.E. Neafsey and S.K. Volkman

AN ABUNDANCE OF GENOMIC DATA

To date, large genomic datasets have provided a framework for the forward-thinking work needed to apply malaria genomic information toward the goal of malaria eradication. Following the generation of reference genome assemblies, malaria investigators followed the path of other organisms in the postgenomic era and characterized genomic diversity through sequencing surveys (Jeffares et al. 2007; Mu et al. 2007; Volkman et al. 2007; Tan et al. 2011; Neafsey et al. 2012) and later genome-wide single nucleotide polymorphism (SNP) arrays (Neafsey et al. 2008, 2010; Mu et al. 2010; Tan et al. 2011; Van Tyne et al. 2011). Such work illuminated the recent demographic history of multiple parasite and vector species, defined sometimes complex gene flow boundaries, and revealed the impact of immune, drug, or insecticide selection on the genomes of malaria parasites and vectors. Subsequently, more comprehensive characterizations of genetic diversity have yielded thorough inventories of genomic diversity across many geographic regions (Amambua-Ngwa et al. 2012; Manske et al. 2012; Miotto et al. 2013, 2015; MalariaGEN *P. falciparum* Community Project 2016), refining our understanding of population boundaries and diversity differences between populations. Collectively, this deep genomic sequencing dataset, for example, represented in the open-access Pf3k collaboration and database (www.malariagen.net/projects/pf3k), facilitates informed development of markers for genotyping to analyze even larger sample collections. To ensure that sequencing data remain accessible and maximally useful to the malaria research community, it will be essential that prospective sequencing data generation efforts contribute sequencing data and a minimum increment of sample metadata (e.g., location and date of sample collection) to open repositories such as Pf3K and/or PlasmoDB in a timely manner with as few restrictions as possible. Strategies for how to accomplish this remain largely undetermined, but community-generated and -accessible databases are critical for leveraging genomic

information to advance malaria elimination efforts.

DEVELOPING THE TOOLKIT

Although whole genome sequencing (WGS) efforts have provided a critical foundation for genomic tool development, WGS data may not be necessary to address some key considerations for malaria eradication. In fact, a major current limitation to the use of genomics for malaria eradication efforts is the limited availability of samples of sufficient quality or amount for WGS that have been collected with important clinical, epidemiological, or biological metadata. Although the cost of genome sequencing has fallen substantially, it can still be technically difficult and costly to sequence clinical samples heavily contaminated with host DNA. Two strategies have been developed to address this challenge: processes that enrich the parasite DNA either at the time of collection or from the already extracted nucleic acid material and use of genotyping tools that do not require removal of host DNA. Approaches, including hybrid selection (Melnikov et al. 2011; Bright et al. 2012) and selective whole genome amplification (Leichty and Brisson 2014), enable efficient sequencing by enriching for parasite over host genetic material. Filtration methods at the time of collection have also been employed to reduce host material, with variable success (Venkatesan et al. 2012). However, polymerase chain reaction (PCR)-based approaches that either genotype small collections of SNPs or a limited number of highly polymorphic regions (amplicons) are inexpensive ways of extracting genomic data from samples that are limited in amount or contaminated with large amounts of host DNA.

Several groups have explored the use of so-called SNP “barcodes” to distinguish unique versus clonal parasite lineages and track changes in disease transmission over time (Campino et al. 2011; Daniels et al. 2013, 2008, 2015; Echeverry et al. 2013; Nkhoma et al. 2013). Other groups have used amplification of highly polymorphic regions of the parasite genome to create haplotypes, which is the basis of merozoite surface protein type 1 (MSP) typing strate-

gies used to distinguish parasite reinfection (Tanabe et al. 1998). This latter strategy is particularly useful for polygenomic infections, where more than one parasite genome contributes to human infection. Both approaches (SNP barcodes and MSP typing) allow one to estimate the complexity of infection (COI) level across patient populations, or the mean number of genetically distinct parasite lineages infecting a person, which has been used to distinguish between high and low transmission levels. Statistical tools like COIL (complexity of infection using likelihood) can estimate the COI level of a malaria sample using only knowledge of SNP minor allele frequencies (MAFs) and a sample's genotype (Galinsky et al. 2015). SNP-based barcodes have also proven useful for distinguishing malaria parasites hailing from different geographic regions (Preston et al. 2014; Baniecki et al. 2015), suggesting they could be helpful in identifying the source of imported infections in pre-elimination settings. Amplicon-based sequencing approaches have also been useful in other intervention contexts, such as characterizing polymorphism in vaccine candidates (Juliano et al. 2010; Bailey et al. 2012; Gandhi et al. 2012, 2014; Aragam et al. 2013; Neafsey et al. 2015; Mideo et al. 2016).

One consideration important for obtaining value from large collections of SNP genotyping data will be some means of ensuring that data and findings are portable between studies conducted by different investigators. Numerous groups have put forth proposed collections of nuclear genetic markers for genotyping, while other groups have pursued whole mitochondrial sequencing to make demographic inferences about populations. Comparing findings between studies that employed nuclear and mitochondrial markers (Sutton 2013; Koepfli et al. 2015) can be challenging because of the different modes of inheritance, genetic effective population sizes, and mutation rates of those genomes. Use of markers that are neutral or potentially under selection (e.g., immune or drug) adds another level of consideration about which markers are most informative and/or appropriate for a given question. Within the nuclear genome, SNPs and microsatellite markers exhibit

extremely different modes and rates of mutation, requiring very different analysis strategies for interpretation. Microsatellites themselves exhibit highly variable mutation rates, related to the length of the repeat unit and the number of copies of the repeat, with greater repeat counts generally leading to a higher mutation rate.

Studies that simultaneously explore the signals of multiple marker sets, for example, nuclear SNPs and microsatellite markers, will be necessary to understand the transferability of findings employing a single class of markers. As methods to call microsatellite repeat length variation from NGS data improve, there will be an opportunity to use WGS data from parasites to explore the degree of concordance within and among different classes of markers from the same samples. The development and refinement of new tools to analyze genotype data for malaria applications will help to organize the field around specific sets of markers. Ultimately, the most powerful way to ensure comparability of findings from different genotyping studies will be community adoption of a common set of genotypic markers. Even if there is motivation to include population-specific genotype markers, for example, because of local questions relating to drug resistance or selection, analysis of a common core of genotype markers will contribute to building a collection of data that can be critically compared across disease elimination settings in a meta-analytical framework. Observing the behavior of a common set of markers in a range of elimination settings with disparate outcomes will be key to furthering our understanding of effective uses of genotyping tools.

DEPLOYMENT OF A GENOMICS TOOLKIT FOR THE ERADICATION AGENDA

Now that we have extensive catalogs of genomic variation data for *P. falciparum*, with efforts underway to produce comparable compendia of genomic variants for *Plasmodium vivax* and other parasite species, how do we apply these resources as effectively as possible to the goal of malaria eradication? Below we outline several ways we believe genomic data can continue to

D.E. Neafsey and S.K. Volkman

Figure 1. Range of transmission intensity. A schematic representing signals detected from mosquito, human, and parasite populations that are anticipated as transmission intensity declines. As the transmission intensity decreases from high to low levels, changes in mosquito (entomological inoculation rate [EIR]); human complexity of infection [COI]), and parasite (genotyping) indicators are anticipated to change. With relatively high transmission (e.g., $EIR > 1$), we see high COI levels and a predominance of polygenomic infections as assessed through genotyping methods. As transmission decreases to more moderate (e.g., EIR from 0.1 to 1) or lower levels (e.g., EIR from 0.01 to 0.1), we detect decreases in COI and increases in the proportion of individuals harboring monogenomic infections. Eventually as transmission intensity is very low (e.g., EIR from 0 to 0.01) we detect evidence of COI = 1 and clonal parasite populations among monogenomic infections detected using genotyping methods.

be a cost-effective and invaluable resource for understanding and controlling malaria and providing critical information about elimination activities. We focus this discussion on ways genomic information can be used to monitor changes in transmission intensity, evaluate of the impact of intervention methods, and track the movement of parasites between human populations. Critical for the success of these applications is the collection of samples with epidemiological, clinical, or biological metadata. Here we focus on study design and use of epidemiological and clinical metadata to inform changes in transmission or evaluate the impact of an intervention. However, genomic approaches for biological discovery, as are generally applied to understanding of mechanisms

of drug resistance or drug action, are also an important application of genomics that will be considered.

GENOMICS FOR STUDYING TRANSMISSION

The use of genomics to monitor changes in transmission intensity and evaluate intervention impact is grounded in basic population genetic principles (Fig. 1). As transmission intensity decreases there is a reduction in outcrossing during the mosquito stages of the life-cycle. This predicts that, over time, COI will decline among human infections, and parasites will become increasingly genetically similar because of inbreeding and recent common ances-

try (Volkman et al. 2012). One can follow these signals over time to monitor increased parasite relatedness, as outward indicators of reduced transmission such as prevalence or incidence of malaria are detected using epidemiological or clinical measures. Such an approach has recently documented changing parasite population dynamics over time in Senegal, with modeling approaches employed to confirm sensitivity for detection of both transmission decrease and rebound patterns of malaria transmission (Daniels et al. 2015). There is tremendous value in samples collected over time, either across sequential transmission seasons, or in longitudinal or cohort design studies. To be useful for observing changes in transmission dynamics over time, longitudinal sample collections only need to be large enough to overcome a binomial sampling error within individual time points, and, often, signals indicating transmission changes can be detected by repeatedly sampling 100 samples or fewer. Thus, use of genomics has the potential to track transmission dynamics over time and monitor changes in these dynamics related to epidemiological and clinical variables.

Important for our understanding of transmission dynamics is determining the relative contributions of cotransmission and superinfection. Cotransmission occurs when multiple parasites enter the human host during a single mosquito-feeding event, whereas superinfection occurs when parasites enter the human host during multiple mosquito-feeding events. Cotransmission is expected to introduce parasites that are genetically more alike, perhaps coming from the same recombination events in the mosquito midgut. In contrast, depending on the genetic variation in the population, superinfection may introduce more highly diverse parasite types into the human host. Genomics, including single-cell genotyping and sequencing (Nkhoma et al. 2012; Nair et al. 2014), as well as population-based approaches, has the opportunity to help resolve distinct parasite types within either the human or mosquito host, and may be used to better estimate the contributions of these two different transmission patterns and their influence

Malaria Genomics in the Era of Eradication

under various transmission settings to onward infection.

INTERVENTION IMPACT ASSESSMENT

Related to the concept of monitoring parasite population structure is the use of genomics to evaluate the impact of interventions (Fig. 2). Two main strategies are employed, using either drugs or vaccines, to reduce the clinical burden of malaria. Genomics has the potential to assess how such interventions are working, and provide an early warning system to potential failure of these strategies as resistant parasite populations emerge through natural selection. For example, application of drug pressure selects for drug-resistant variants that increasingly contribute to the overall parasite population infecting humans. At a certain point, these drug-resistant variants are sufficiently prevalent in the population that drug responses are compromised and these agents become limited or ineffective clinically. Genomics has the power to not only to discover drug-resistant variants by detecting genomic variants that increase in frequency over time, but also to monitor the consequences of drug pressure, such as under mass drug administration (MDA) projects that are being rolled out as part of the malaria eradication efforts. Application of MDA can boost frequency of drug-resistant parasites in the population by simultaneously compressing the parasite population while selecting for drug-resistant variants, and potentially compromise drug efficacy for future MDA projects.

To circumvent the emergence of drug resistance, specific genomic loci can be identified (e.g., *dhodh* [Ross et al. 2014] and *pfcr1* [Lukens et al. 2014]) that exist in two alternate allelic states (Fig. 3). In one allelic state, the locus is susceptible to the first drug while resistant to a second drug. In the alternate allelic state, the locus is resistant to the first drug, but susceptible to the second drug. Thus, one compound selects for genetic variants that are susceptible to the second compound, and vice versa. Such approaches provide promise for use of novel drug combinations where one drug selects for para-

D.E. Neafsey and S.K. Volkman

Figure 2. Intervention effect on parasite population. This schematic represents the potential impact of different interventions (e.g., bednets, vaccines, or drugs) on allele frequencies in the parasite population. The different colored circles represent a specific locus of interest such as an allele or haplotype that may be subject to selection by a vaccine (green circle) or drug (red circle). As a vaccine (e.g., a monovalent protein subunit vaccine like RTS,S) is applied, there may be reduction in specific parasite types that harbor the target locus (e.g., the circumsporozoite [CS] locus that matches the type found in the vaccine), thus reducing parasite types with that specific locus from the parasite population. As a drug is applied, there may be selection for a specific drug resistance locus resulting in an increase in the frequency of parasite types that harbor that specific drug-resistant variant. For comparison, use of a bednet might reduce the overall parasite types (represented by fewer circles that are proportionally reduced), but may not specifically select for or against any particular locus or allele within parasites in that population.

sites that are then susceptible to a second drug. The combined use of such paired compounds could offset emergence of resistance altogether, or, used sequentially, one drug could drive the emergence of parasite populations susceptible to the second drug, and vice versa. Thus, alternating drug pressure using a paired set of compounds could maintain the effectiveness of the compounds over time. Genomics can be used to identify genetic loci that have these alternate allelic states and facilitate identification of paired drugs that could prevent emergence of drug-resistant parasites. Surveillance of these critical allelic states can then be used to identify critical inflection points where drug strategies need to shift to remain effective.

Several malaria vaccine candidates have been associated with allele-specific protective

responses (Takala et al. 2007, 2009; Takala and Plowe 2009; Thera et al. 2011; Neafsey et al. 2015). Allele-specific protective efficacy may compromise the overall efficacy of a vaccine, but can be characterized with PCR amplicon sequencing data. We recently demonstrated that the RTS,S/AS01 vaccine, which targets a highly polymorphic region of the circumsporozoite (CS) protein, exhibits significantly higher vaccine efficacy against infections matching the 3D7 vaccine strain compared with infections that do not match the vaccine strain (Neafsey et al. 2015). Thus, genomic information provides insight into vaccine effectiveness, and can perhaps contribute to improvements in vaccine efficacy for highly polymorphic vaccine targets, as well as helping predict the impact of a vaccine on a given target population.

Figure 3. Anticorrelated drug resistance—an evolutionary loop. This schematic represents a potential evolutionary loop whereby two different drugs (A in red and B in blue) act on the same locus with each drug selecting a different variant or haplotype than the alternate drug. For example, as drug A (red) is applied to the parasite population, the haplotype conferring resistance to A will increase in frequency in the population. If the alternate drug (B, blue) is applied to this population, the alternate haplotype conferring resistance to B will increase in the population. This creates a population with increased resistance to B, but increased susceptibility to the alternate drug A. Application of drug A will now reverse the dynamics of the haplotype frequency such that the variant that confers resistance to A will again increase in frequency in the population. This alternating use of paired drugs with anticorrelated resistance will thus create an evolutionary loop.

TRACKING PARASITES

Genomics also has the potential to track transmission patterns, including detecting the movement of parasites and identifying new sources of infection. Genotyping approaches such as barcode methods or amplicon-sequencing approaches can identify specific parasite types, much like creating a “fingerprint” that identifies each kind of parasite within an infection. Such methods provide an opportunity to track specific parasite types in space and time. Human migration patterns, identified either from demographic or cell phone data, can be used to predict sources and sinks of malaria parasite

migration. If such patterns are present, targeting interventions to the source site could consequently reduce infections in the sink site. Genomic tools can test this “sink-source” hypothesis, by fingerprinting parasites and following them in space and time.

As we are better able to detect and track parasites, we still need to develop more sensitive methods for parasite identification, especially as we move toward malaria elimination and parasite density within human infections decreases. Specifically, sensitive molecular genetic approaches can identify parasite reservoirs, possibly among large numbers of asymptomatic individuals, and define whether these contrib-

D.E. Neafsey and S.K. Volkman

ute to incident infection. One strategy is using case investigation or reactive case detection, where cases identified through passive case detection are then followed to determine whether other individuals in the household or neighborhood are also infected. Genomic tools have also been used for outbreak investigation (Obaldia et al. 2015) and may assist in identifying parasite types under “prevention of reintroduction” conditions in elimination settings. Genomic strategies not only provide a sensitive means of identifying who is infected, but also provide evidence for the spatial and temporal relationships among infected individuals. Such strategies can identify individuals who may not be treated as they are asymptomatic, yet may be a local source of new infection. Mapping of infected individuals may also identify pockets of infection that can then be investigated for mosquito breeding sites or other factors that contribute to maintaining these reservoirs of infection. For example, genomics can be coupled with strategies such as serology mapping in elimination phases to identify recent exposure history with serology, and then use molecular genetic approaches to determine who harbors infections and whether these contribute to new infections to map transmission networks and develop malaria risk maps.

Other useful information and strategies that might be obtained using genetic approaches include detection of gametocyte reservoirs, evaluating whether asymptomatic individuals harbor parasites that are infectious (i.e., determining the risk of seeding new infections), and understanding the intrahost dynamics between strains. As we move toward regional elimination, it is also critical to assess whether new cases are a result of rebound of autochthonous malaria, or whether it is a result of importation. In the context of “prevention of new infections” it will be important to be able to distinguish between local or external sources of these new infections to better define operational responses to distinct parasite sources. Some of the challenges we are likely to face as we move to elimination are the reduction of parasite types and increases in parasite relatedness. Thus, use of markers that differentiate between local and ex-

ternal populations will become increasingly challenging. Likely we will require use of a combination of markers including those that are rapidly evolving, such as microsatellites, and possibly the use of expression profiling such as can be obtained through RNAseq or Nanostring approaches.

FUTURE PERSPECTIVES FOR APPLICATION OF GENOMIC DATA

With a growing catalog of parasite population genomic information, we now need to develop strategies to apply these genomic data and approaches to promote malaria elimination. Although genomic sequencing provides a great deal of information about genetic variation within and between populations, key to application of these data to malaria eradication will involve population genetics strategies. Some of these approaches will involve use of changing parasite population genetics to monitor transmission dynamics and with modeling to inform best strategies and combinations of approaches to reduce the malaria burden. As drugs, vaccines, or other intervention measures impose selective pressures on the genomes of parasites and vectors, we can expect selected variants to measurably change in frequency within a few years, owing to the relatively short generation times. This method of detecting the targets of natural selection could identify the genomic basis of drug or insecticide resistance, as well as identify compensatory mutations that do not directly contribute to resistance but restore organismal fitness in the presence of resistance mutations.

To estimate changes in allele frequencies within a population over time, Figure 4 depicts the results of a binomial sampling-based simulation of changes in derived allele frequency (DAF) in a parasite population over time for neutral ($s = 0$; gray) and selected ($s = 0.01$; red) alleles, starting from a common DAF of 5% (assumes effective populations size (N_e) is constant and equal to 1000). Assuming that there are approximately 10 parasite generations per year, these simulations approximate the observable allele frequency changes in a clinical

Figure 4. Expected changes in derived allele frequency. A simulation of expected changes in derived allele frequency (DAF) over time in a population with 1000 individuals for alleles that are selectively neutral (gray lines) or that confer a 1% fitness advantage (red lines). Time is measured in generations.

sample set collected over the span of a decade, the interval of time during which artemisinin resistance became widespread in Southeast Asia (Dondorp et al. 2009). Both neutral and selected alleles oscillate in frequency because of random genetic drift, but a significant fraction of selected alleles increase markedly in frequency over this relatively short time interval, with a conservatively low fitness advantage (1%). This suggests that careful analysis of longitudinally collected genomic datasets, taking into account factors such as population structure and demography, could be highly powered to detect and monitor the evolutionary impact of disease intervention efforts in both parasites and vectors. To produce such datasets in a practical and cost-effective manner, technical innovations will be necessary to ensure that clinical samples can be efficiently sequenced from an easily collectable source material (e.g., dried blood spots on filter paper), using such methods as hybrid selection, host DNA depletion, and/or selective whole genome amplification.

Whether future malaria genomic datasets are produced from longitudinal studies or focused on individual transmission seasons, these data must be analyzed and made public as quickly as possible to be truly useful. It will be of little practical value to understand the nature

of a pre-elimination outbreak from a genomic perspective 2 years after the fact; as a community we need tools to render actionable information from applied sequencing studies very quickly. Modeling genomic data in a manner that is integrated with epidemiological, clinical, serological, and other data types will be essential to distill the most informative signals about when and how to act in the context of elimination activities. In the future, easy-to-use tools like simple cell phone applications that display converted genomic signals into decision-making outputs for program officers in National Malaria Control Programs within the Ministry of Health would be ideal.

Malaria genomics is ready to move from a descriptive field to a practical tool informing elimination efforts, but, to successfully undergo that transition, the field will need to produce data and analyses closely tied to epidemiological and public health needs. Furthermore, the field will need new metrics and accepted modes of recognizing important contributions. A manuscript published in a prestigious journal years after samples have been collected and analyzed is of value to the investigators for future grant applications, but release of a dataset accompanied by a more cursory but actionable analysis in a forum like BioRxiv (bioRxiv.org)

D.E. Neafsey and S.K. Volkman

may be more useful for rapid dissemination of new molecular drug-resistance markers. Community engagement about how to create, support, and recognize contributions required for this timely generation, sharing, and interpretation of genomic data to inform operational activities is paramount.

CONCLUDING REMARKS

Recent developments in the field of nanopore sequencing point to new directions that malaria genomics may follow in the era of eradication. The prospect of inexpensive, disposable sequencing hardware, paired with largely automated analysis for common applications, will be powerful forces in the decentralization of genomic data generation. Although there are many reasons to celebrate increased access to the means of genomic data generation by investigators in malaria-endemic countries, that access will be of low value unless it is accompanied by training in the analysis, interpretation, and use of such data for operational activities. Thus, genomic data must become a local, interpretable, and actionable resource for the malaria community. The malaria research community needs to work hard to ensure that scientists from malaria-endemic countries are empowered to generate and analyze their own data, as the era of exporting both samples and intellectual ownership for the execution of malaria genomics studies to nonendemic countries is rapidly drawing to a close.

To avoid unnecessary and counterproductive duplication of efforts in an era of decentralized data production, the field should identify “best practice” methodologies for the generation and analysis of data, and repositories that can efficiently absorb data produced by a large number of contributors, as well as develop guidelines for use of these data. The malaria genomics field is fortunate to already have several central data repositories, including PlasmoDB (<http://www.plasmodb.org>) and the Pf3K project (www.malariagen.net/projects/parasite/pf3k) for parasite data, as well as VectorBase (vectorbase.org) and the Ag1000g project (www.malariagen.net/projects/vector/

ag1000g) for vector data. It will be incumbent on the malaria research community, however, to ensure that the resources and tools provided by these repositories truly serve the goals of the malaria research community in the era of eradication. Care and attention to study design and sampling with connected metadata, including clinical, epidemiological, and biological information, will only enrich our ability to apply genomic strategies to increase our chances of reducing the malaria disease burden toward eradication.

ACKNOWLEDGMENTS

We thank Dyann F. Wirth, Daniel Hartl, and Bronwyn MacInnis for ongoing discussions about the ideas presented.

REFERENCES

- Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, Cheeseman IH, Newbold CI, Holder AA, Knuepfer E, et al. 2012. Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet* **8**: e1002992.
- Aragam NR, Thayer KM, Nge N, Hoffman I, Martinson F, Kamwendo D, Lin FC, Sutherland C, Bailey JA, Juliano JJ. 2013. Diversity of T cell epitopes in *Plasmodium falciparum* circumsporozoite protein likely due to protein–protein interactions. *PLoS ONE* **8**: e62427.
- Bailey JA, Mvalo T, Aragam N, Weiser M, Congdon S, Kamwendo D, Martinson F, Hoffman I, Meshnick SR, Juliano JJ. 2012. 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* **206**: 580–587.
- Baniecki ML, Faust AL, Schaffner SE, Park DJ, Galinsky K, Daniels RE, Hamilton E, Ferreira MU, Karunaweera ND, Serre D, et al. 2015. Development of a single nucleotide polymorphism barcode to genotype *Plasmodium vivax* infections. *PLoS Negl Trop Dis* **9**: e0003539.
- Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU, Schork NJ, Vinetz JM, Winzeler EA. 2012. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics* **13**: 262.
- Campino S, Auburn S, Kivinen K, Zongo I, Ouedraogo JB, Mangano V, Djimde A, Doumbo OK, Kiara SM, Nzila A, et al. 2011. Population genetic analysis of *Plasmodium falciparum* parasites using a customized Illumina GoldenGate genotyping assay. *PLoS ONE* **6**: e20251.
- Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, Rosen D, Angelino E, Sabeti PC, Wirth DF, et al. 2008. A general SNP-based molecular barcode for *Plas-*

- modium falciparum* identification and tracking. *Malaria J* 7: 223.
- Daniels R, Chang HH, Sene PD, Park DC, Neafsey DE, Schaffner SE, Hamilton EJ, Lukens AK, Van Tyne D, Mboup S, et al. 2013. Genetic surveillance detects both clonal and epidemic transmission of malaria following enhanced intervention in Senegal. *PLoS ONE* 8: e60780.
- Daniels RE, Schaffner SE, Wenger EA, Proctor JL, Chang HH, Wong W, Baro N, Ndiaye D, Fall FB, Ndiop M, et al. 2015. Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci* 112: 7067–7072.
- Dondorp AM, Nosten F, Yi P, Das D, Phylo AP, Tarning J, Lwin KM, Arie F, Hanpithakpong W, Lee SJ, et al. 2009. Artemisinin resistance in *Plasmodium falciparum* malaria. *N Engl J Med* 361: 455–467.
- Echeverry DE, Nair S, Osorio L, Menon S, Murillo C, Anderson TJ. 2013. Long term persistence of clonal malaria parasite *Plasmodium falciparum* lineages in the Colombian Pacific region. *BMC Genet* 14: 2.
- Galinsky K, Valim C, Salmier A, de Thoisy B, Musset L, Legrand E, Faust A, Baniecki ML, Ndiaye D, Daniels RE, et al. 2015. COIL: A methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data. *Malaria J* 14: 4.
- Gandhi K, Thera MA, Coulbaly D, Traore K, Guindo AB, Doumbo OK, Takala-Harrison S, Plowe CV. 2012. Next generation sequencing to detect variation in the *Plasmodium falciparum* circumsporozoite protein. *Am J Trop Med Hyg* 86: 775–781.
- Gandhi K, Thera MA, Coulbaly D, Traore K, Guindo AB, Ouattara A, Takala-Harrison S, Berry AA, Doumbo OK, Plowe CV. 2014. Variation in the circumsporozoite protein of *Plasmodium falciparum*: Vaccine development implications. *PLoS ONE* 9: e101783.
- Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, et al. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 419: 498–511.
- Jeffares DC, Pain A, Berry A, Cox AV, Stalker J, Ingle CE, Thomas A, Quail MA, Siebenthal K, Uhlemann AC, et al. 2007. Genome variation and evolution of the malaria parasite *Plasmodium falciparum*. *Nat Genet* 39: 120–125.
- Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Arie F, Wongsrichanalai C, Read A, Meshnick SR. 2010. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc Natl Acad Sci* 107: 20138–20143.
- Koepfli C, Rodrigues PT, Antao T, Orjuela-Sanchez P, Van den Eede P, Gamboa D, van Hong N, Bendezu J, Erhart A, Barnadas C, et al. 2015. *Plasmodium vivax* diversity and population structure across four continents. *PLoS Negl Trop Dis* 9: e0003872.
- Leichty AR, Brisson D. 2014. Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics* 198: 473–481.
- Lukens AK, Ross LS, Heidebrecht R, Javier Gamo F, Lafuente-Monasterio MJ, Booker ML, Hartl DL, Wiegand RC, Wirth DE. 2014. Harnessing evolutionary fitness in *Plasmodium falciparum* for drug discovery and suppressing resistance. *Proc Natl Acad Sci* 111: 799–804.
- MalariaGEN *Plasmodium falciparum* Community Project. 2016. Genomic epidemiology of artemisinin resistant malaria. *eLife* 5: e08714.
- Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, O'Brien J, Djimde A, Doumbo O, Zongo I, et al. 2012. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487: 375–379.
- Melnikov A, Galinsky K, Rogov P, Fennell T, Van Tyne D, Russ C, Daniels R, Barnes KG, Bochicchio J, Ndiaye D, et al. 2011. Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol* 12: R73.
- Mideo N, Bailey JA, Hathaway NJ, Ngasala B, Saunders DL, Lon C, Kharabara O, Jamnik A, Balasubramanian S, Bjorkman A, et al. 2016. A deep sequencing tool for partitioning clearance rates following antimalarial treatment in polyclonal infections. *Evol Med Public Health* 2016: 21–36.
- Miotto O, Almagro-Garcia J, Manske M, Macinnis B, Campino S, Rockett KA, Amaratunga C, Lim P, Suon S, Sreng S, et al. 2013. Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nat Genet* 45: 648–655.
- Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, et al. 2015. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. *Nat Genet* 47: 226–234.
- Mu J, Awadalla P, Duan J, McGee KM, Keebler J, Seydel K, McVean GA, Su XZ. 2007. Genome-wide variation and identification of vaccine targets in the *Plasmodium falciparum* genome. *Nat Genet* 39: 126–130.
- Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, Chotivanich K, Wilairatana P, Krudsood S, White NJ, et al. 2010. *Plasmodium falciparum* genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet* 42: 268–271.
- Nair S, Nkhoma SC, Serre D, Zimmerman PA, Goren A, Daniel BJ, Nosten F, Anderson TJ, Cheeseman IH. 2014. Single-cell genomics for dissection of complex malaria infections. *Genome Res* 24: 1028–1038.
- Neafsey DE, Schaffner SE, Volkman SK, Park D, Montgomery P, Milner DA Jr, Lukens A, Rosen D, Daniels R, Houde N, et al. 2008. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol* 9: R171.
- Neafsey DE, Lawniczak MK, Park DJ, Redmond SN, Coulbaly MB, Traore SE, Sagnon N, Costantini C, Johnson C, Wiegand RC, et al. 2010. SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. *Science* 330: 514–517.
- Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q, et al. 2012. The malaria parasite *Plasmodium vivax* exhibits greater genetic diversity than *Plasmodium falciparum*. *Nat Genet* 44: 1046–1050.
- Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, Lievens M, Abdulla S, Adjei S, Agbenyega T, et al. 2015. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. *N Engl J Med* 373: 2025–2037.
- Nkhoma SC, Nair S, Cheeseman IH, Rohr-Allegrini C, Singlam S, Nosten F, Anderson TJ. 2012. Close kinship

D.E. Neafsey and S.K. Volkman

- within multiple-genotype malaria parasite infections. *Proc Biol Sci* **279**: 2589–2598.
- Nkhoma SC, Nair S, Al-Saai S, Ashley E, McGready R, Phyo AB, Nosten F, Anderson TJ. 2013. Population genetic correlates of declining transmission in a human pathogen. *Mol Ecol* **22**: 273–285.
- Obaldia N 3rd, Baro NK, Calzada JE, Santamaria AM, Daniels R, Wong W, Chang HH, Hamilton EJ, Arevalo-Herrera M, Herrera S, et al. 2015. Clonal outbreak of *Plasmodium falciparum* infection in eastern Panama. *J Infect Dis* **211**: 1087–1096.
- Preston MD, Campino S, Assefa SA, Echeverry DE, Ocholla H, Amambua-Ngwa A, Stewart LB, Conway DJ, Borrmann S, Michon P, et al. 2014. A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. *Nat Commun* **5**: 4052.
- Ross LS, Gamo FJ, Lafuente-Monasterio MJ, Singh OM, Rowland P, Wiegand RC, Wirth DE. 2014. In vitro resistance selections for *Plasmodium falciparum* dihydroorotate dehydrogenase inhibitors give mutants with multiple point mutations in the drug-binding site and altered growth. *J Biol Chem* **289**: 17980–17995.
- Sutton PL. 2013. A call to arms: On refining *Plasmodium vivax* microsatellite marker panels for comparing global diversity. *Malaria J* **12**: 447.
- Takala SL, Plowe CV. 2009. Genetic diversity and malaria vaccine design, testing and efficacy: Preventing and overcoming “vaccine resistant malaria.” *Parasite Immunol* **31**: 560–573.
- Takala SL, Coulibaly D, Thera MA, Dicko A, Smith DL, Guindo AB, Kone AK, Traore K, Ouattara A, Djimde AA, et al. 2007. Dynamics of polymorphism in a malaria vaccine antigen at a vaccine-testing site in Mali. *PLoS Med* **4**: e93.
- Takala SL, Coulibaly D, Thera MA, Batchelor AH, Cummings MP, Escalante AA, Ouattara A, Traore K, Niangaly A, Djimde AA, et al. 2009. Extreme polymorphism in a vaccine antigen and risk of clinical malaria: Implications for vaccine development. *Sci Transl Med* **1**: 2ra5.
- Tan JC, Miller BA, Tan A, Patel JJ, Cheeseman IH, Anderson TJ, Manske M, Maslen G, Kwiatkowski DP, Ferdig MT. 2011. An optimized microarray platform for assaying genomic variation in *Plasmodium falciparum* field populations. *Genome Biol* **12**: R35.
- Tanabe K, Sakihama N, Kaneko O, Saito-Ito A, Kimura M. 1998. A PCR method for molecular epidemiology of *Plasmodium falciparum* Msp-1. *Tokai J Exp Clin Med* **23**: 375–381.
- Thera MA, Doumbo OK, Coulibaly D, Laurens MB, Ouattara A, Kone AK, Guindo AB, Traore K, Traore I, Kouriba B, et al. 2011. A field trial to assess a blood-stage malaria vaccine. *N Engl J Med* **365**: 1004–1013.
- Van Tyne D, Park DJ, Schaffner SE, Neafsey DE, Angelino E, Cortese JF, Barnes KG, Rosen DM, Lukens AK, Daniels RF, et al. 2011. Identification and functional validation of the novel antimalarial resistance locus PF10_0355 in *Plasmodium falciparum*. *PLoS Genet* **7**: e1001383.
- Venkatesan M, Amaratunga C, Campino S, Auburn S, Koch O, Lim P, Uk S, Socheat D, Kwiatkowski DP, Fairhurst RM, et al. 2012. Using CF11 cellulose columns to inexpensively and effectively remove human DNA from *Plasmodium falciparum*-infected whole blood samples. *Malaria J* **11**: 41.
- Volkman SK, Sabeti PC, DeCaprio D, Neafsey DE, Schaffner SE, Milner DA Jr, Daily JB, Sarr O, Ndiaye D, Ndir O, et al. 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat Genet* **39**: 113–119.
- Volkman SK, Neafsey DE, Schaffner SE, Park DJ, Wirth DE. 2012. Harnessing genomics and genome biology to understand malaria biology. *Nat Rev Genet* **13**: 315–328.

Cold Spring Harbor Perspectives in Medicine

Malaria Genomics in the Era of Eradication

Daniel E. Neafsey and Sarah K. Volkman

Cold Spring Harb Perspect Med 2017; doi: 10.1101/cshperspect.a025544 originally published online April 7, 2017

Subject Collection [Malaria: Biology in the Era of Eradication](#)

Modern Vector Control

Neil F. Lobo, Nicole L. Achee, John Greico, et al.

Anopheline Reproductive Biology: Impacts on Vectorial Capacity and Potential Avenues for Malaria Control

Sara N. Mitchell and Flaminia Catteruccia

Current and Future Prospects for Preventing Malaria Transmission via the Use of Insecticides

Hilary Ranson

Molecular Signaling Involved in Entry and Exit of Malaria Parasites from Host Erythrocytes

Shailja Singh and Chetan E. Chitnis

Vaccines to Accelerate Malaria Elimination and Eventual Eradication

Julie Healer, Alan F. Cowman, David C. Kaslow, et al.

Immune Responses in Malaria

Carole A. Long and Fidel Zavala

Antimalarial Drug Resistance: A Threat to Malaria Elimination

Didier Menard and Arjen Dondorp

Malaria during Pregnancy

Michal Fried and Patrick E. Duffy

Malaria Pathogenesis

Danny A. Milner, Jr.

Determinants of Malaria Transmission at the Population Level

Teun Bousema and Chris Drakeley

Host Cell Tropism and Adaptation of Blood-Stage Malaria Parasites: Challenges for Malaria Elimination

Caeul Lim, Selasi Dankwa, Aditya S. Paul, et al.

Malaria Transmission and Prospects for Malaria Eradication: The Role of the Environment

Marcia C. Castro

The Biology of *Plasmodium vivax*

John H. Adams and Ivo Mueller

Malaria Genomics in the Era of Eradication

Daniel E. Neafsey and Sarah K. Volkman

Malaria Epigenetics

Alfred Cortés and Kirk W. Deitsch

Malaria Parasite Liver Infection and Exoerythrocytic Biology

Ashley M. Vaughan and Stefan H.I. Kappe

For additional articles in this collection, see <http://perspectivesinmedicine.cshlp.org/cgi/collection/>