

REVIEW

Open Access



Applying next-generation sequencing to track *falciparum* malaria in sub-Saharan Africa

Sofonias K. Tessema^{1*} , Jaishree Raman², Craig W. Duffy³, Deus S. Ishengoma⁴, Alfred Amambua-Ngwa⁵ and Bryan Greenhouse^{1,6}

Abstract

Next-generation sequencing (NGS) technologies are increasingly being used to address a diverse range of biological and epidemiological questions. The current understanding of malaria transmission dynamics and parasite movement mainly relies on the analyses of epidemiologic data, e.g. case counts and self-reported travel history data. However, travel history data are often not routinely collected or are incomplete, lacking the necessary level of accuracy. Although genetic data from routinely collected field samples provides an unprecedented opportunity to track the spread of malaria parasites, it remains an underutilized resource for surveillance due to lack of local awareness and capacity, limited access to sensitive laboratory methods and associated computational tools and difficulty in interpreting genetic epidemiology data. In this review, the potential roles of NGS in better understanding of transmission patterns, accurately tracking parasite movement and addressing the emerging challenges of imported malaria in low transmission settings of sub-Saharan Africa are discussed. Furthermore, this review highlights the insights gained from malaria genomic research and challenges associated with integrating malaria genomics into existing surveillance tools to inform control and elimination strategies.

Keywords: Next-generation sequencing, Malaria genomics, Molecular epidemiology, Tracking parasites

Background

In 2017, there were approximately 198 million cases of malaria and 400,000 malaria-related deaths in sub-Saharan Africa (SSA) [1]. This burden is highly heterogeneous; 10 countries contributed 70% of the total reported cases [1], while others made substantial progress towards realizing elimination as an attainable goal [2]. This heterogeneous nature of malaria distribution in SSA has resulted in countries with very low transmission sharing borders with higher transmission countries, a scenario clearly evidenced in southern Africa [3]. On a finer scale, this phenomenon extends to neighboring regions within the same country, and has resulted in some SSA countries

implementing subnational targeted malaria elimination strategies [4, 5].

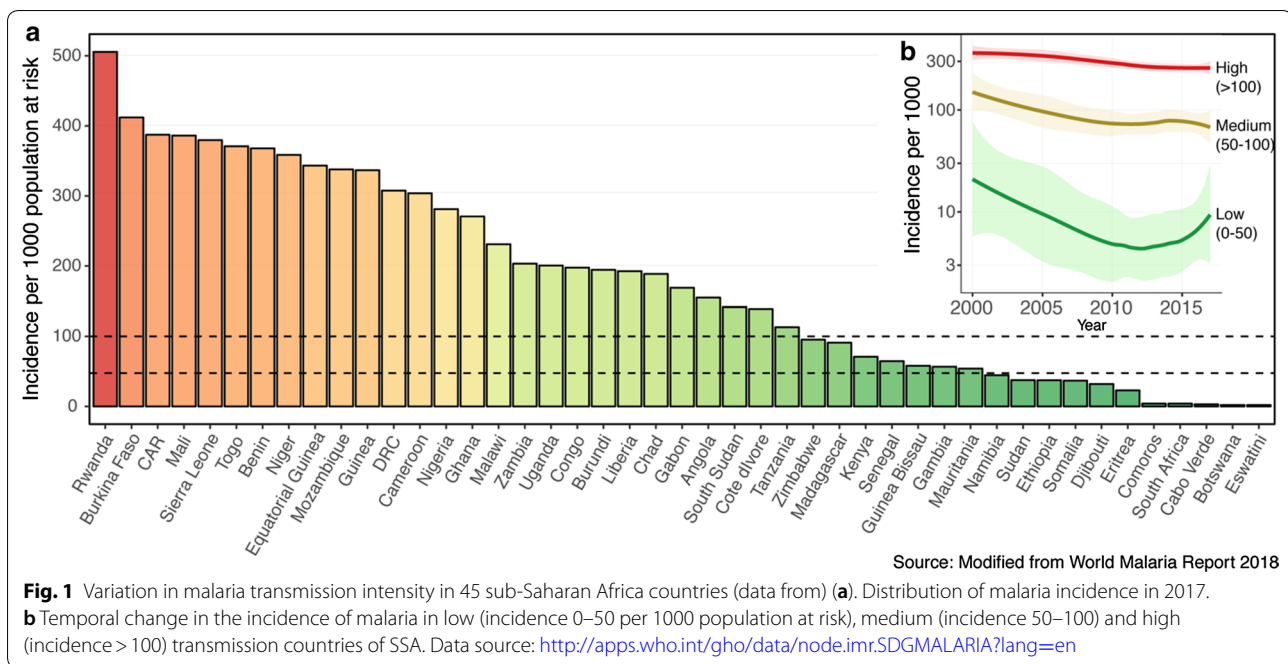
Spread of malaria from higher to lower transmission areas is not a new challenge, but has become a more pressing issue given the recent progress in malaria control and increasing connectivity between SSA countries driven by increased population movements [6]. In areas of low malaria transmission, an apparent “outbreak” of cases could merely arise from imported cases and could present a barrier to achieving malaria elimination [7] (Fig. 1). Imported malaria remains difficult to identify and address using the surveillance tools and limited resources currently available to most National Malaria Control Programmes (NMCPs). While existing surveillance measures (e.g. parasite rate, case incidence and reported travel histories) are fundamental to answering some of these questions, they are often limited by consistency and accuracy, particularly in areas with highly mobile and migrant populations.

*Correspondence: sofoniaskifle@gmail.com

¹ EPPICenter Program, Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

Full list of author information is available at the end of the article





Recent advances in malaria genomics, when combined with epidemiologic data, offer the potential to improve our ability to track the spread of malaria [8–12]. Several studies have highlighted the value of malaria genomics in epidemiological and public health settings (reviewed by [11–13]). Most of these studies have relied on traditional genotyping methods to understand transmission dynamics [14, 15]; to determine connectivity between populations [16–18]; to classify local and imported infections [19] and to track the spread of drug resistance as discussed by Ishengoma et al. in this technical series [20]. The increasing accessibility of high throughput next-generation sequencing (NGS) now provides an opportunity to refine existing tools and develop high resolution approaches [17, 21–27]. Despite this potential, molecular epidemiology approaches are not yet fully operationalized to inform intervention strategies in SSA. In this review, how malaria genomics can improve the understanding of malaria transmission dynamics by enabling the tracking of parasite flow between populations and its promise to address the challenges of identifying imported infections are discussed. This review highlights the major challenges in study design, gaps in laboratory and analytical methods and translating genomic data into calibrated epidemiologic parameters. Furthermore, issues surrounding logistics, local capacity, communication between researchers and NMCPs and the need to establish a framework for actionable genomic epidemiology studies in SSA are highlighted.

Genomic epidemiology of malaria: from transmission to translation

Estimating population level transmission from genomic data

Parasite genetic data from polymorphic loci (e.g. microsatellites and highly diverse genes); single nucleotide polymorphisms (SNP) and whole genome sequences (WGS) have been used to characterize transmission dynamics [11, 13, 14]. It is generally believed that as malaria transmission declines, complexity of infection and genetic diversity decreases, leading to spatial fragmentation of parasite populations [11, 28–31]. This association can ideally be leveraged to assess the level of ongoing local transmission and evaluate the impact of control or elimination interventions [32]. In this context, genomic data would augment traditional surveillance methods, and may be particularly useful in settings where surveillance infrastructure is limited or where traditional metrics are insensitive to relevant changes in transmission. Within-host and population level genetic indices have both been used to measure transmission intensity.

Measuring within-host genetic indices

Within-host parasite diversity can arise from sequential bites from mosquitoes that are infected with genetically distinct strains (superinfection) or more commonly from a single mosquito bite containing multiple strains, some of which may be closely related (co-transmission) [33]. Traditionally, within-host parasite diversity is measured by characterizing a few polymorphic loci to

determine the minimum number of genetically distinct parasite strains present (i.e. multiplicity of infections, MOI) (reviewed by Zhong et al. [33]), but these methods have limited inter-laboratory reproducibility and comparability among studies. SNP panels and analytical tools have been developed to mitigate the limitations of size-polymorphic markers [15, 34–37]. However, SNPs begin to lose resolution in highly complex infections where most loci with reasonable diversity will result in heterozygous calls. Advances in NGS technologies have made it possible to develop highly sensitive targeted deep sequencing approaches [22, 38–43]. These data can provide information not only the number of strains, but also the genetic diversity [e.g. the within-host diversity index (*F_{ws}*)] as well as relatedness and genetic structure of parasites in an infection [22, 43–46]. Multi-locus deep sequencing data can provide information on the number and genetic relatedness of strains in an infection, which can be utilized to evaluate spatial differences and temporal changes in transmission. A recent exciting development is the ability to sequence single infected red blood cells, providing high-resolution data, including the ability to understand drivers of genetic diversity within naturally occurring infections and to unambiguously phase genetic data from mixed infections [33, 47]. However, these methods are not yet easily scalable to the levels required for most epidemiologic work, requiring additional development.

In malaria-endemic regions, multiple infections are common, and within-host diversity indices broadly correlate with endemicity in a non-linear fashion [14, 28–32, 44, 48–50]. Reduced within-host diversity has been associated with increased ITN use [51] and temporal changes in transmission [50, 52, 53], indicating that these metrics may be reasonable indicators of changes in transmission intensity. However, these relationships may be more difficult to interpret in areas where infections are not at a steady-state, and in particular may not reflect local transmission intensity, even amongst locally acquired infections, in areas with high rates of importation [54]. Although NGS technologies are advancing quantification of within-host diversity, they have also introduced new challenges. For example, the depth of sequencing can dramatically affect within-host diversity indices and comparability of findings between studies. Therefore, study design and the sequencing workflow including sample preparation, choice of sequencing platform, sequencing depth and sequence processing pipelines would benefit from standardization across molecular epidemiology studies. Similarly, the sensitivity of computational methods to identify genetic variants and associated errors are crucial for the successful establishment of NGS-based inference of

transmission and the possibility to integrate these tools for routine surveillance.

Measuring population level genetic indices

While population level genetic diversity is complex, there are a range of genetic measures, including heterozygosity (or homozygosity), effective population size, linkage disequilibrium, proportions of genetically related infections and others [11, 12]. These indices have been traditionally generated using microsatellites and SNP panels. However, NGS offer the potential to provide data at a greater breadth to evaluate the genetic relatedness of parasites within and between infections. Several methods (such as DEploidIBD [44], DEploid [55], hmmIBD [56] and isoRelate [57]) are now available to identify genomic regions that are shared between isolates thereby detecting whether pairs of infections within and between populations are genetically related or not.

Several studies have compared the spatial differences in these indices and evaluated changes over time. Declining transmission has been associated with increase in linkage disequilibrium, reduction in parasite genetic diversity and effective population size likely as a result of population bottlenecks and fragmentation of parasite populations [11, 12]. However, the relationships between transmission intensity and changes in genomic indices are unlikely to be linear and uniform across settings. A recent study in the Kingdom of Eswatini [54], a country with a yearly malaria incidence of <1 malaria case per 1000 population at risk, revealed a high level of parasite diversity. The authors found that this seemingly paradoxical finding was consistent with transmission driven by a high proportion of imported infections with a few, short chains of local transmission, a finding supported by the NMCP's classification of most cases as imported based on travel history data. Findings such as this highlight the challenge of applying broad genetic relationships to scenarios where transmission is dynamic—with dramatic changes in transmission occurring over short spatial and/or temporal scales. Such areas may represent the rule rather than the exception in low and very low transmission settings of SSA where effective interventions may increase heterogeneity in transmission [16]. Therefore, additional work is needed to determine sampling frames, relevant meta-data and appropriate statistical frameworks to obtain calibrated transmission metrics from genomic data such that these data can complement existing tools and improve the scale and resolution of population level inference of transmission.

Inferring connectivity from parasite genetic data

Genetic data are often used to estimate connectivity between populations and track the spread of infectious

diseases. In malaria, the genetic relatedness of infections can be compared at different scales to determine spatial patterns of transmission [8, 16, 22, 58, 59]. Infections with the same or highly similar genomes linked by epidemiologic data, are likely to represent infections that are related by transmission. The genetic resolution required to distinguish the relatedness of infections over a given number of transmission events depends on the underlying transmission intensity; the spatial scale of the study; sampling frameworks and parasite density (e.g. symptomatic and/or asymptomatic samples) and the ability of the assay to detect genetically distinct strains in polyclonal and low-density infections.

Over the years, several analytical methods to infer population structure have been developed [60–62]. However, genetic relatedness-based approaches have recently been shown to provide a greater resolution to identify geographic regions that are linked by transmission [16, 58, 59]. These approaches have the potential to define spatial patterns in low to moderate transmission areas. However, genetic relatedness metrics alone are not enough to accurately estimate importation; analytical tools to translate genetic connectivity to demographic connectivity (e.g. the number of parasites imported from one location to the other) are required. This is a major problem in most areas of SSA where human connectivity and parasite flow occurs between genetically intermixed populations. However, the availability of reference WGS data and resultant ability to inform geographically informative target selection frameworks targets, would improve the resolution of genetic data to accurately assign parasites to their geographic origin.

Beyond population level estimates of connectivity, parasite genomics offers enormous potential to determine person-to-person transmission chains [63]. These approaches are routine for viral and bacterial diseases [64, 65] but are not yet readily applicable for sexually recombining pathogens such as malaria. Tracking of infections using this approach is based on the genetic similarity between the ensemble of parasites present in each of a pair of infections representing a potential transmission event. Therefore, the accuracy of transmission chains is dependent on the genetic resolution (i.e. the informativeness of genetic targets and the sensitivity of genotyping methods); the proportion of infections sampled; the available epidemiologic meta-data; the analytical methods and the context within which these methods are applied. In SSA, where imported infections are often highly complex and diverse [16, 66], genotyping methods must be able to detect minority strains from low density infections by targeting regions of the genome that exhibit genetic diversity sufficient to provide individual strain identification in mixed infections.

Classification of local and imported infections using parasite genetic data

Malaria importation poses a complex challenge in low transmission countries of SSA [7]. Despite the success of malaria control over the past decade, recent reports show an increase in the number of clinical cases in pre-elimination countries, which is mainly attributed to importation from neighboring higher transmission areas [1, 3] (Fig. 1b). These challenges are more pronounced in southern Africa, where malaria eliminating countries are being challenged by cross-border importation from higher transmission countries [3, 67, 68]. For example, more than 50% of malaria cases in Eswatini and South Africa were classified as imported based on reported travel history of patients [3, 67, 68]. Progress towards elimination requires novel approaches with improved accuracy to classify imported malaria and measure its impact on local transmission. Molecular epidemiology provides an as yet untapped opportunity to address these challenges.

NGS can provide high-throughput genomic data from routinely collected field samples (e.g. dried blood spots and used RDTs), however even the best genetic data alone cannot be expected to accurately reconstruct transmission chains. Availability of a toolbox of appropriate analytical methods that combine parasite genomics and epidemiological data [8], would be very useful for reconstruction of transmission chains and derivation of other transmission metrics. Furthermore, this approach can provide information on the number of secondary cases arising from each imported infection, allowing programmes to directly measure receptivity and determine what if any interventions are required to prevent re-establishment of transmission in an eliminating area. Therefore, a multidisciplinary team of modellers, molecular epidemiologists, programme officers and other stakeholders need to work together to fully harness the potential of malaria genomics to address the current and future challenges of imported malaria in very low transmission settings of SSA.

Challenges and priorities for the translation of malaria genomic epidemiology

Genomic epidemiology has grown rapidly over the past decade, facilitated by decreasing costs, increased efficiency of NGS technologies as well as improved infrastructure for data management and analysis. These improved technologies have enabled researchers to explore how malaria genomics can help improve the understanding of transmission dynamics and flow of parasites. However, translation of findings to inform control and elimination and its applications in SSA has lagged

behind other regions owing to a number of factors, some of which are illustrated in Fig. 2 and discussed below.

Challenges of integrating with routine surveillance systems

Integration of genomic data into malaria surveillance is a relatively recent phenomenon, largely facilitated by the rapid increase in accessibility to these technologies. While this makes the field very exciting scientifically, there is limited evidence on the utility of these methods to provide timely and actionable data. There are even fewer data available to guide standardization or best practices on the types of sampling frames to consider, genomic data to generate, and analytical plans to employ in order to answer epidemiologically relevant questions. Although these types of evidence are rapidly being generated, it is understandably difficult for research groups and other stakeholders interested in evaluating or deploying these technologies to obtain buy in from the relevant field partners. Some NMCPs may not consider genomic epidemiology as an integral part of their

suite of surveillance tools or even one worth investigating through operational research. For under-resourced NMCPs where achieving adequate coverage of basic control interventions is a major challenge, genomics may be viewed as a costly luxury that they can ill afford. What is often not appreciated by programmes and donors, and in truth is currently difficult to formally evaluate, is the cost of deploying (or not deploying) expensive interventions based on substandard surveillance data. Therefore, there is very urgent need for discussions and operationally relevant studies that highlight the value of genomic epidemiology. In addition, it is essential that standardized frameworks are created which enable the integration of genomic epidemiology into existing surveillance systems in low and very low transmission areas of SSA. The potential applications of genomic epidemiology depend on the timely, efficient collection of samples, as well as access to sample information and associated meta-data, which should be leveraged through existing surveillance programs such as health facility-based surveillance, malaria indicator surveys (MIS), therapeutic efficacy

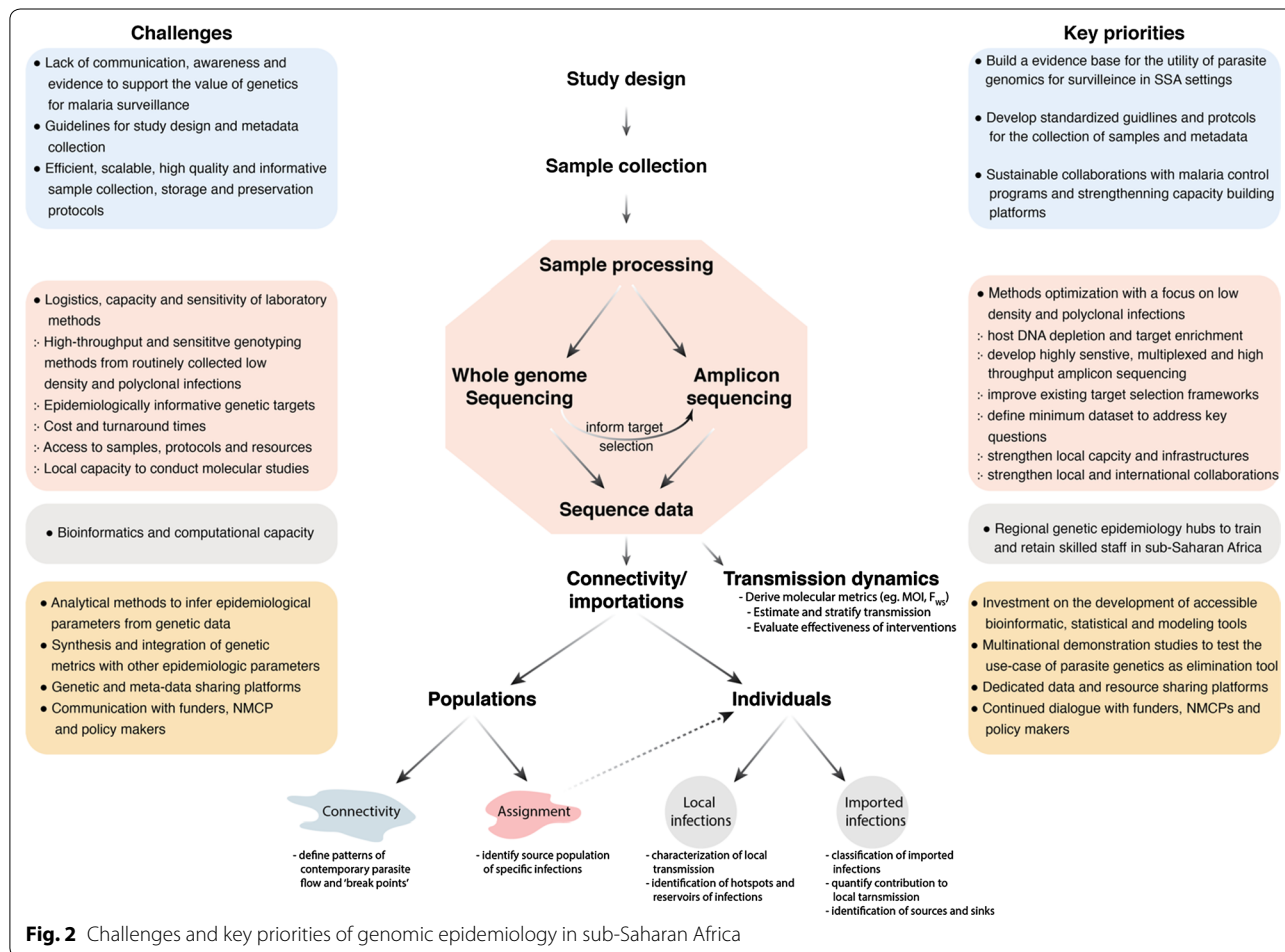


Fig. 2 Challenges and key priorities of genomic epidemiology in sub-Saharan Africa

studies (TES) and others. These approaches provide cost-effective and scalable platforms, which help establish efficient and equitable collaboration with NMCPs while addressing programmatic questions without requiring much if any costs for additional sampling.

Challenges with laboratory methods and data analyses

Two major challenges associated with laboratory-based analyses in SSA are the sensitivity of the technology required to generate high quality data from low-density infections and the detection of relevant strains in highly polyclonal infections. The ability to obtain epidemiologically useful information from the genomes of mixed infections is an issue in most of SSA, where polyclonal infections are often the rule rather than the exception even in relatively low transmission areas [16, 66, 69]. Advances in NGS have enabled the whole genome and targeted sequencing of malaria from routinely collected field samples [70], but no single method is universally applicable for all contexts. WGS can provide the near complete genome of the parasite, while targeted NGS can generate data at specific loci of interest. Although both approaches rely on a more or less similar laboratory workflow, the required quantity and quality of DNA; cost, laboratory processing, data storage, and computational time differ significantly. Unlike WGS, targeted sequencing provides greater depth of sequencing allowing for improved detection of minority strains in an infection and more robust comparisons between samples, since amplification of targeted loci is deeper and more consistent. However, identification of informative genomic regions and the capability to effectively multiplex these targets are challenges that need to be overcome to fully realize the utility of this method. Which and how many targets provide sufficient information to track the flow of parasites or determine their origin from different regions of SSA is currently unknown. Furthermore, it is appealing to develop combined panels to efficiently answer multiple questions, such as transmission epidemiology and prevalence of molecular markers of drug resistance. Therefore, optimization of sensitive laboratory methods and the development of well calibrated and accessible data analyses tools need to be prioritized and addressed.

Challenges with local infrastructure and capacity

Despite the potential of NGS technologies, there are many challenging and costly bridges to cross in order to realize its full potential and application in SSA. The lack of local capacity in many SSA countries is the result of many complex issues, including limited infrastructure, the shortage of African expertise in malaria genomics and the lack of adequate support by African governments and international donors. There are efforts to overcome

some of these limitations (e.g. The Malaria Genomic Epidemiology Network (MalariaGEN), The Plasmodium Diversity Network Africa (PDNA), MRCG-LSHTM genomics and high-performance computer centers and others), however a sustainable approach will, in part, depend on African scientists advocating for governments and international funders to build local facilities and establish sustainable collaborative platforms to train and attract skilled staff necessary to lead malaria genomics in SSA. One model would be to advocate for funding of regional genomics centers, which can provide services to geopolitical blocs. These would help to galvanize adoption and ensure sustainability, since local scientists and control program staff would be directly invested in the process and as such be able to better direct the outcome.

Challenges with communication of findings

Genomic surveillance, coupled with high quality meta-data, can provide valuable input for the development of policies for malaria elimination and eventual eradication. Critical to the success of any control/elimination programme is the involvement of and ownership by the respective local governments. For example, WHO-recommended TES have allowed strong partnership between NMCPs and researchers to plan, implement and report findings [71]. These partnerships also extend to the dissemination of findings to policy makers and development of policy recommendations, particularly when changes of first-line anti-malarials are required. Experience gained from such initiatives need to be formalized and utilized to develop efficient and sustainable partnerships between researchers and NMCPs for a broader range of molecular epidemiology studies. In the context of wide adoption and utilization of genomic data to support malaria elimination strategies, NMCPs should be capacitated on the added value and how genomic data can be integrated to inform routine activities of the programmes. The programmes also need to be supported and sensitized to appreciate how genetic data can potentially support and influence the process of changing different malaria control/elimination policies. Whenever possible, the NMCPs should play a central role in the planning and implementation of genomic studies in order to ensure that the research address questions of relevance and priority in the local context. Ensuring adoption of malaria control or elimination policies guided by research findings must be a joint effort of NMCP, researchers and other key stakeholders. Finally, the parasite genetic surveillance findings must be disseminated to key stakeholders and policy makers in language that they clearly understand. Making the data interpretable and accessible, e.g. by using maps and interactive visualization tools, must be a priority for genomics surveillance.

Data sharing and accessibility

The cost of sequencing has dropped significantly, resulting in the rapid generation of genomic data by researchers from developing and developed countries. An important benefit of NGS data is that it can be shared and may be directly compared, even if the sequences were generated by different laboratories. Data sharing is essential for enabling and promoting malaria genomic research in a way that will maximize the outcomes from molecular epidemiology studies and provide resources for the development and validation of laboratory and analytical tools. There are encouraging efforts in data sharing by the malaria community (e.g. the pioneering efforts demonstrated by MalariaGEN and PlasmoDB). However, consistent deposition of sequence data from various scientists into standardized repositories is not yet coordinated at the scale seen in other fields. It is challenging to formulate efficient solutions for data standardization, formatting, archiving and access in the absence of a dedicated repository for malaria genomic data. In the context of SSA, limited access to high-speed internet and computational capacity presents a barrier for African researchers to access and utilize publicly available genomic data. In the future, platforms such as PlasmoDB or any newly designed dedicated data sharing platform must adequately address the challenges and bottlenecks in malaria genomic data sharing (e.g. ethical, administrative, logistic and data management issues) and their accessibility by SSA researchers.

Conclusions and future perspectives

NGS technologies hold enormous potential for scientists and malaria control program personnel in SSA to improve their understanding of malaria transmission dynamics, including the ability to tracking the spread of parasites between populations or from one individual to the other. Ultimately, with sufficient reference genomes from multiple sites, it may be possible to accurately assign infections to their geographic origin. This information would help to address the formidable challenges of accurately evaluating imported malaria in low and very low transmission settings. It is clear that sensitive laboratory methods, appropriate analytical and dissemination tools, and accessible data sharing platforms need to be prioritized and developed to materialize the promises of malaria genomics, especially in SSA. Strategies for sample collection that leverage existing surveillance systems and, when appropriate, standardization of study designs will help to facilitate sustainable partnerships between researchers, NMCPs, and other stakeholders. African researchers need to play a leading role in advocating for international donors and African

governments to build local capacity, possibly through regional genomic epidemiology hubs that can spearhead the generation of data to inform malaria elimination and eradication policies on the continent.

Authors' contributions

SKT and BG conceived the idea and outlined the draft. SKT, JR, CWD, DSI, AAN and BG wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Funding

BG is a Chan Zuckerberg Biohub investigator. DSI and AN are partly supported by the DELTAS Africa Initiative [Grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [Grant 107740/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government. We would like to thank Jonathan Juliano for reviewing the draft manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ EPPICenter Program, Division of HIV, Infectious Diseases, and Global Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ² Centre for Emerging Zoonotic and Parasitic Diseases, National Institute for Communicable Disease, Sandringham, Gauteng, South Africa. ³ Department of Infection Biology, University of Liverpool, Liverpool, UK. ⁴ National Institute for Medical Research, Tanga Research Centre, Tanga, Tanzania. ⁵ MRC Gambia Unit, The Gambia at LSHTM, Fajara, The Gambia. ⁶ Chan Zuckerberg Biohub, San Francisco, CA, USA.

Received: 28 April 2019 Accepted: 11 July 2019

Published online: 03 September 2019

References

1. WHO. World Malaria Report 2018. Geneva: World Health Organization; 2018.
2. Global Elimination Forum. Malaria elimination: report from the inaugural global forum of countries with potential to eliminate malaria by 2020. *Wkly Epidemiol Rec.* 2017;92:578–86.
3. SADC. Southern African Development Community Malaria Report; 2017.
4. Muchena G, Dube B, Chikodzore R, Pasipamire J, Murugasampillay S, Mberikunashe J. A review of progress towards sub-national malaria elimination in Matabeleland South Province, Zimbabwe (2011–2015): a qualitative study. *Malar J.* 2018;17:146.
5. Balawanth R, Ba I, Qwabe B, Gast L, Maharaj R, Raman J, et al. Assessing Kwa-Zulu-Natal's progress towards malaria elimination and its readiness for sub-national verification. *Malar J.* 2019;18:108.
6. UNCTAD. Economic Development in Africa Report 2018: Migration for Structural Transformation. New York: United Nations Publications; 2018.
7. Sturrock HJW, Roberts KW, Wegbreit J, Ohrt C, Gosling RD. Tackling imported malaria: an elimination endgame. *Am J Trop Med Hyg.* 2015;93:139–44.

8. Wesolowski A, Taylor AR, Chang H-H, Verity R, Tessema S, Bailey JA, et al. Mapping malaria by combining parasite genomic and epidemiologic data. *BMC Med*. 2018;16:190.
9. Greenhouse B, Smith DL. Malaria genotyping for epidemiologic surveillance. *Proc Natl Acad Sci USA*. 2015;112:6782–3.
10. Kwiatkowski D. Malaria genomics: tracking a diverse and evolving parasite population. *Int Health*. 2015;7:82–4.
11. Auburn S, Barry AE. Dissecting malaria biology and epidemiology using population genetics and genomics. *Int J Parasitol*. 2017;47:77–85.
12. Escalante AA, Ferreira MU, Vinetz JM, Volkman SK, Cui L, Gamboa D, et al. Malaria molecular epidemiology: lessons from the International Centers of Excellence for Malaria Research Network. *Am J Trop Med Hyg*. 2015;93:79–86.
13. Koepfli C, Mueller I. Malaria epidemiology at the clone level. *Trends Parasitol*. 2017;33:974–85.
14. Anderson TJ, Haubold B, Williams JT, Estrada-Franco JG, Richardson L, Mollinedo R, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. *Mol Biol Evol*. 2000;17:1467–82.
15. Daniels R, Volkman SK, Milner DA, Mahesh N, Neafsey DE, Park DJ, et al. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar J*. 2008;7:223.
16. Tessema S, Wesolowski A, Chen A, Murphy M, Wilhelm J, Mupiri A-R, et al. Using parasite genetic and human mobility data to infer local and cross-border malaria connectivity in Southern Africa. *eLife*. 2019;8:e43510.
17. Mobegi VA, Duffy CW, Amambua-Ngwa A, Loua KM, Laman E, Nwakanma DC, et al. Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Mol Biol Evol*. 2014;31:1490–9.
18. Salgueiro P, Vicente JL, Figueiredo RC, Pinto J. Genetic diversity and population structure of *Plasmodium falciparum* over space and time in an African archipelago. *Infect Genet Evol*. 2016;43:252–60.
19. Patel JC, Taylor SM, Juliao PC, Parobek CM, Janko M, Gonzalez LD, et al. Genetic evidence of importation of drug-resistant *Plasmodium falciparum* to Guatemala from the Democratic Republic of the Congo. *Emerg Infect Dis*. 2014;20:932–40.
20. Ishengoma D, Saidi Q, Sibley C, Roper C, Alifrangis M. Deployment and utilization of next-generation sequencing of *Plasmodium falciparum* to guide anti-malarial drug policy decisions in sub-Saharan Africa: opportunities and challenges. *Malar J*. 2019 (in press).
21. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, Milner DA, et al. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol*. 2008;9:R171.
22. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature*. 2012;487:375–9.
23. MalariaGEN. Plasmodium Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife*. 2016;5:e08714.
24. Duffy CW, Assefa SA, Abugri J, Amoako N, Owusu-Agyei S, Anyorigiya T, et al. Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics*. 2015;16:527.
25. Juliano JJ, Parobek CM, Brazeau NF, Ngasala B, Randrianarivelojosia M, Lon C, et al. Pooled amplicon deep sequencing of candidate *Plasmodium falciparum* transmission-blocking vaccine antigens. *Am J Trop Med Hyg*. 2016;94:143–6.
26. Quan J, Langelier C, Kuchta A, Batson J, Teyssier N, Lyden A, et al. FLASH: A next-generation CRISPR diagnostic for multiplexed detection of antimicrobial resistance sequences. *bioRxiv*. 2018. <http://biorxiv.org/lookup/doi/10.1101/426338>. Accessed 2 Apr 2019.
27. Aydemir O, Janko M, Hathaway NJ, Verity R, Mwandagalirwa MK, Tshetu AK, et al. Drug-resistance and population structure of *Plasmodium falciparum* across the Democratic Republic of Congo using high-throughput molecular inversion probes. *J Infect Dis*. 2018;218:946–55.
28. Bei AK, Niang M, Deme AB, Daniels RF, Sarr FD, Sokhna C, et al. Dramatic changes in Malaria Population genetic complexity in Dielmo and Ndiop, Senegal, revealed using genomic surveillance. *J Infect Dis*. 2018;217:622–7.
29. Redmond SN, MacInnis BM, Bopp S, Bei AK, Ndiaye D, Hartl DL, et al. De novo mutations resolve disease transmission pathways in clonal malaria. *Mol Biol Evol*. 2018;35:1678–89.
30. Waltmann A, Koepfli C, Tessier N, Karl S, Fola A, Darcy AW, et al. Increasingly inbred and fragmented populations of *Plasmodium vivax* associated with the eastward decline in malaria transmission across the Southwest Pacific. *PLOS Negl Trop Dis*. 2018;12:e0006146.
31. Fola AA, Harrison GLA, Hazairin MH, Barnadas C, Hetzel MW, Iga J, et al. Higher complexity of infection and genetic diversity of *Plasmodium vivax* than *Plasmodium falciparum* across all malaria transmission zones of Papua New Guinea. *Am J Trop Med Hyg*. 2017;96:630–41.
32. Volkman SK, Neafsey DE, Schaffner SF, Park DJ, Wirth DF. Harnessing genomics and genome biology to understand malaria biology. *Nat Rev Genet*. 2012;13:315–28.
33. Nkhoma SC, Trevino SG, Gorena KM, Nair S, Khoswe S, Jett C, et al. Resolving within-host malaria parasite diversity using single-cell sequencing. *bioRxiv*. 2018. <https://doi.org/10.1101/391268>.
34. Zhong D, Koepfli C, Cui L, Yan G. Molecular approaches to determine the multiplicity of Plasmodium infections. *Malar J*. 2018;17:172.
35. Omedo I, Mogeni P, Bousema T, Rockett K, Amambua-Ngwa A, Oyier I, et al. Micro-epidemiological structuring of *Plasmodium falciparum* parasite populations in regions with varying transmission intensities in Africa. *Wellcome Open Res*. 2017;2:10.
36. Chang H-H, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, et al. THE REAL McCOIL: a method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput Biol*. 2017;13:e1005348.
37. 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;14:4.
38. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Ariey F, et al. Exposing malaria in-host diversity and estimating population diversity by capture-recapture using massively parallel pyrosequencing. *Proc Natl Acad Sci USA*. 2010;107:20138–43.
39. Levitt B, Obala A, Langdon S, Corcoran D, O'Meara WP, Taylor SM. Overlap extension barcoding for the next generation sequencing and genotyping of *Plasmodium falciparum* in individual patients in Western Kenya. *Sci Rep*. 2017;7:41108.
40. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. *BMC Genomics*. 2017;18:864.
41. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. *Nucleic Acids Res*. 2018;46:e21.
42. Miller RH, Hathaway NJ, Kharabara O, Mwandagalirwa K, Tshetu A, Meshnick SR, et al. A deep sequencing approach to estimate *Plasmodium falciparum* complexity of infection (COI) and explore apical membrane antigen 1 diversity. *Malar J*. 2017;16:490.
43. Auburn S, Campino S, Miotto O, Djimde AA, Zongo I, Manske M, et al. Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS ONE*. 2012;7:e32891.
44. Zhu SJ, Hendry JA, Almagro-Garcia J, Pearson RD, Amato R, Miles A, et al. The origins and relatedness structure of mixed infections vary with local prevalence of *P. falciparum* malaria. *bioRxiv*. 2018;387266.
45. Early AM, Lievens M, MacInnis BL, Ockenhouse CF, Volkman SK, Adjei S, et al. Host-mediated selection impacts the diversity of *Plasmodium falciparum* antigens within infections. *Nat Commun*. 2018;9:1381.
46. Taylor AR, Jacob PE, Neafsey DE, Buckee CO. Estimating relatedness between malaria parasites. *bioRxiv*. 2019;575985.
47. Trevino SG, Nkhoma SC, Nair S, Daniel BJ, Moncada K, Khoswe S, et al. High-resolution single-cell sequencing of malaria parasites. *Genome Biol Evol*. 2017;9:3373–83.
48. Duffy CW, Amambua-Ngwa A, Ahouidi AD, Diakite M, Awandare GA, Ba H, et al. Multi-population genomic analysis of malaria parasites indicates local selection and differentiation at the *gdf1* locus regulating sexual development. *Sci Rep*. 2018;8:15763.
49. Mobegi VA, Loua KM, Ahouidi AD, Satoguina J, Nwakanma DC, Amambua-Ngwa A, et al. Population genetic structure of *Plasmodium falciparum* across a region of diverse endemicity in West Africa. *Malar J*. 2012;11:223.

50. 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;22:273–85.
51. Mueller I, Schoepflin S, Smith TA, Benton KL, Bretscher MT, Lin E, et al. Force of infection is key to understanding the epidemiology of *Plasmodium falciparum* malaria in Papua New Guinean children. *Proc Natl Acad Sci USA*. 2012;109:10030–5.
52. Amambua-Ngwa A, Jeffries D, Amato R, Worwui A, Karim M, Ceesay S, et al. Consistent signatures of selection from genomic analysis of pairs of temporal and spatial *Plasmodium falciparum* populations from The Gambia. *Sci Rep*. 2018;8:9687.
53. Daniels R, Schaffner SF, Wenger EA, Proctor JL, Chang H-H, Wong W, et al. Modeling malaria genomics reveals transmission decline and rebound in Senegal. *Proc Natl Acad Sci USA*. 2015;112:7067–72.
54. Roh ME, Tessema SK, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, et al. High genetic diversity of *Plasmodium falciparum* in the low transmission setting of the Kingdom of Eswatini. *bioRxiv*. 2019;522896.
55. Zhu SJ, Almagro-Garcia J, McVean G. Deconvolution of multiple infections in *Plasmodium falciparum* from high throughput sequencing data. *Bioinformatics*. 2018;34:9–15.
56. Schaffner SF, Taylor AR, Wong W, Wirth DF, Neafsey DE. hmmlBD: software to infer pairwise identity by descent between haploid genotypes. *Malar J*. 2018;17:196.
57. Henden L, Lee S, Mueller I, Barry A, Bahlo M. Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS Genet*. 2018;14:e1007279.
58. Taylor AR, Schaffner SF, Cerqueira GC, Nkhoma SC, Anderson TJC, Sriprawat K, et al. Quantifying connectivity between local *Plasmodium falciparum* malaria parasite populations using identity by descent. *PLoS Genet*. 2017;13:e1007065.
59. Chang H-H, Wesolowski A, Sinha I, Jacob CG, Mahmud A, Uddin D, et al. Mapping imported malaria in Bangladesh using parasite genetic and human mobility data. *eLife*. 2019;8:e43481.
60. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
61. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. *PLoS Genet*. 2012;8:e1002453.
62. Verity R, Nichols RA. Estimating the number of subpopulations (K) in structured populations. *Genetics*. 2016;203:1827–39.
63. Archie EA, Luikart G, Ezenwa VO. Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends Ecol Evol*. 2009;24:21–30.
64. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*. 2017;544:309–15.
65. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature*. 2017;546:401–5.
66. Roh ME, Tessema SK, Murphy M, Nhlabathi N, Mkhonta N, Vilakati S, et al. High genetic diversity of in the low transmission setting of the Kingdom of Eswatini. *bioRxiv*. 2019;522896.
67. Tejedor-Garavito N, Dlamini N, Pindolia D, Soble A, Ruktanonchai NW, Alegana V, et al. Travel patterns and demographic characteristics of malaria cases in Swaziland, 2010–2014. *Malar J*. 2017;16:359.
68. Raman J, Morris N, Frean J, Brooke B, Blumberg L, Kruger P, et al. Reviewing South Africa's malaria elimination strategy (2012–2018): progress, challenges and priorities. *Malar J*. 2016;15:438.
69. Pringle JC, Tessema S, Wesolowski A, Chen A, Murphy M, Carpi G, et al. Genetic evidence of focal *Plasmodium falciparum* transmission in a pre-elimination setting in Southern Province, Zambia. *J Infect Dis*. 2018;219:1254–63.
70. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, et al. Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar J*. 2016;15:597.
71. WHO. Methods for surveillance of antimalarial drug efficacy. Geneva: World Health Organization; 2009.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

