White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/152559/

Version: Published Version

Article:
Bah, S.Y., Morang'a, C.M., Kengne-Ouafo, J.A. et al. (2018) Highlights on the application of genomics and bioinformatics in the fight against infectious diseases : challenges and opportunities in Africa. Frontiers in Genetics, 9. 575.

https://doi.org/10.3389/fgene.2018.00575

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

frontiers in Genetics

REVIEW
published: 27 November 2018
doi: 10.3389/fgene.2018.00575

Highlights on the Application of Genomics and Bioinformatics in the Fight Against Infectious Diseases: Challenges and Opportunities in Africa

Saikou Y. Bah[1,2], Collins Misita Morang'a[1†], Jonas A. Kengne-Ouafo[1†], Lucas Amenga–Etego[1] and Gordon A. Awandare[1]**

[1] West African Centre for Cell Biology of Infectious Pathogens, University of Ghana, Accra, Ghana, [2] Vaccine and Immunity Theme, MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Banjul, Gambia

OPEN ACCESS

Edited by:
Ancha Baranova, George Mason University, United States

Reviewed by:
David John Studholme, University of Exeter, United Kingdom
Sandeep Kumar Dhanda, La Jolla Institute for Allergy and Immunology (LJI), United States

***Correspondence:**
Saikou Y. Bah
sbah@ug.edu.gh;
sabah@mrc.gm
Gordon A. Awandare
gawandare@ug.edu.gh

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics

Received: *18 June 2018*
Accepted: *08 November 2018*
Published: *27 November 2018*

Citation:
Bah SY, Morang'a CM, Kengne-Ouafo JA, Amenga–Etego L and Awandare GA (2018) Highlights on the Application of Genomics and Bioinformatics in the Fight Against Infectious Diseases: Challenges and Opportunities in Africa. Front. Genet. 9:575. doi: 10.3389/fgene.2018.00575

Genomics and bioinformatics are increasingly contributing to our understanding of infectious diseases caused by bacterial pathogens such as *Mycobacterium tuberculosis* and parasites such as *Plasmodium falciparum*. This ranges from investigations of disease outbreaks and pathogenesis, host and pathogen genomic variation, and host immune evasion mechanisms to identification of potential diagnostic markers and vaccine targets. High throughput genomics data generated from pathogens and animal models can be combined with host genomics and patients' health records to give advice on treatment options as well as potential drug and vaccine interactions. However, despite accounting for the highest burden of infectious diseases, Africa has the lowest research output on infectious disease genomics. Here we review the contributions of genomics and bioinformatics to the management of infectious diseases of serious public health concern in Africa including tuberculosis (TB), dengue fever, malaria and filariasis. Furthermore, we discuss how genomics and bioinformatics can be applied to identify drug and vaccine targets. We conclude by identifying challenges to genomics research in Africa and highlighting how these can be overcome where possible.

Keywords: bioinformatics, genomics, infectious diseases, antimicrobial resistant, diagnosis

INTRODUCTION: OMICS AND BIOINFORMATICS IN INFECTIOUS DISEASES

Genomics and bioinformatics have contributed immensely to our understanding of infectious diseases: from disease pathogenesis, mechanisms and the spread of antimicrobial resistance, to host immune responses. Herein, we review some of the major contributions of genomics and bioinformatics in infectious disease research using examples of three diseases that account for large proportions of morbidity and mortality as well as a neglected tropical disease. Specifically, we review *M. tuberculosis*, which causes TB, a disease responsible for approximately two million deaths globally per year. Dengue virus (DENV) causes Dengue fever, which is a re-emerging mosquito borne viral disease, responsible for more than 350 million cases annually (WHO, 2017; World Health Organization Western Pacific Region, 2018). *Plasmodium falciparum* causes malaria,

a parasitic disease that accounts for the highest morbidity and mortality in Sub-Saharan Africa, especially in children under five and pregnant women (WHO, 2018b), and Filariasis, which is a neglected tropical disease. **Figure 1** shows a circular wheel of genomics/bioinformatics as can be applied in infectious diseases as discussed herein, ranging from understanding host and pathogen genome biology to genome-wide association studies (GWAS) as well as the identification of drug targets and drug resistance surveillance to patient management. This encompasses molecular techniques, bioinformatics and clinical applications (**Figure 1**). We also highlight the application of genomics and bioinformatics to the identification of vaccine targets and drug discovery. We conclude by highlighting some challenges of conducting bioinformatics research in resource-limited countries in sub-Saharan Africa.

OMICS OF TUBERCULOSIS PATHOGENS AND HOST RESPONSES

Tuberculosis caused by members of the *M. tuberculosis* complex is a leading cause of death, with about 9 million cases and two million deaths per year globally (WHO, 2018a). The mycobacterial genome was first sequenced in 1998 and many more *M. tuberculosis* genomes have since been sequenced (Cole et al., 1998; Guerra-Assunção et al., 2015; Yun et al., 2016). These genomes provide great avenues for the genomic characterization, development of improved diagnostic tools, drug susceptibility testing, and molecular epidemiology of circulating mycobacterial strains. Host-pathogen genomics and transcriptomics have over the past decade enhanced our understanding of human-mycobacterium interactions and in the identification of potential diagnostic and prognostic markers (Anderson et al., 2014; Maertzdorf et al., 2015).

An understanding of the *M. tuberculosis* genome biology is invaluable in the control of TB. The *M. tuberculosis* genome is GC rich and consists of about 4000 genes and, unlike other bacteria, a large proportion of its genome encodes proteins and enzymes involved in lipogenesis and lipolysis (Cole et al., 1998), reflecting its thick lipid cell wall. TB control is hampered by antimycobacterial resistance, multidrug resistance (MDR) and, recently, extensively drug resistant (XDR) mycobacterial strains (Leisching et al., 2016). Genomics analysis has immensely contributed to the identification of drug resistance-conferring mutations and surveillance (Köser et al., 2013). Whole genome analyses have demonstrated that mycobacterial drug resistance is largely attributed to single nucleotide polymorphisms (SNPs); for example, rifampicin (RIF) resistance arises from mutations in the *rpoB* gene and mutations in the *katG* and *inhA* lead to isoniazid resistance (da Silva et al., 2011). Newly characterized genetic mutations in *M. tuberculosis* genomes have also been shown to play key roles in the emergence of antimycobacterial drug resistance (Sun et al., 2012). Analyses of 161 drug resistant *M. tuberculosis* genomes identified 72 genes, 28 intergenic regions and 21 SNPs with strong and consistent associations with drug resistance (Zhang et al., 2013). Genomic analysis has also identified lineage mutation rate differences and predicted the emergence of antimycobacterial resistance (Ford et al., 2013). A retrospective analysis of thousands of *M. tuberculosis* genomes collected from African and European patients identified 120 resistance-determining mutations for first and second line antimycobacterial drugs, which could be valuable in developing new assays for drug susceptibility testing (Walker et al., 2015). Furthermore, genomics through the use of GWAS has been used to identify novel mutations associated with resistance to cycloserine, ethionamide, and para-aminosalicylic acid, suggesting the involvement of efflux pump in the emergence of resistance (Coll et al., 2018). A number of genomics-based tools have been developed to detect drug resistance including Mykrobe Predictor, PhyResSE, and TB-Profiler, which are easy to use by researchers with no bioinformatics expertise and can predict drug resistance within minutes after obtaining sequences (Bradley et al., 2015; Coll et al., 2015; Feuerriegel et al., 2015). Mykrobe Predictor has a sensitivity and specificity of 82.6 and 98.5%, respectively (Bradley et al., 2015). TB-Profiler was developed using a mutation library consisting of 1,325 mutations in different genes associated with drug resistance in 15 anti-tuberculosis drugs and had more than 75% sensitivity as well as more than 90% specificity for all drugs tested (Coll et al., 2015). A recent study evaluating the performance of these tools showed that their sensitivity ranges from 74 to 80% along with a specificity of more than 95% (van Beek et al., 2018). However, there is still a need for optimization of analysis pipelines to make them applicable in field settings where the disease burden is usually the highest.

Genomics analysis has also been used to determine the evolutionary history and spread of mycobacterial strains such as the Beijing strain, demonstrating its spread from the Far East (Merker et al., 2015). An investigation of *M. tuberculosis* transmission dynamics is important in monitoring outbreak; Mehaffy et al. (2014) demonstrated that whole genome analysis can be used to monitor infections to decipher transmission dynamics. Furthermore, genomics has also been applied to decipher transmission dynamics of *M. tuberculosis* in Vietnam, suggesting that SNPs in ESX-5 type VII secreted protein EsxW could potentially contribute to enhancing transmission (Holt et al., 2018). Furthermore, genomics has been applied to investigate TB outbreaks, genotyping of the outbreak associated lineages, and their evolution during the outbreak (Jamieson et al., 2014; Stucki et al., 2015). Indeed, analysis tools have been developed for the prediction of *M. tuberculosis* spoligotypes from raw sequence reads, and in combination with other analysis tools also determine antibiotic resistance as well as transmission dynamics (Coll et al., 2012; Bradley et al., 2015). Some genomics methods can also be employed to identify mixed infections as well as infections with a single strain and have recently been applied to clinical isolates from Malawi (Sobkowiak et al., 2018).

Genome-wide association study (GWAS) has also been used to identify candidate gene variants associated with susceptibility to active tuberculosis. GWAS analyses in African patients from Ghana, Gambia, Uganda and Tanzania identified TB disease-associated SNPs located on three chromosomal loci: 18q11, 11p13, and 5q33 (Thye et al., 2010, 2012; Sobota et al., 2016). Similarly, GWAS studies have also been done in Europe

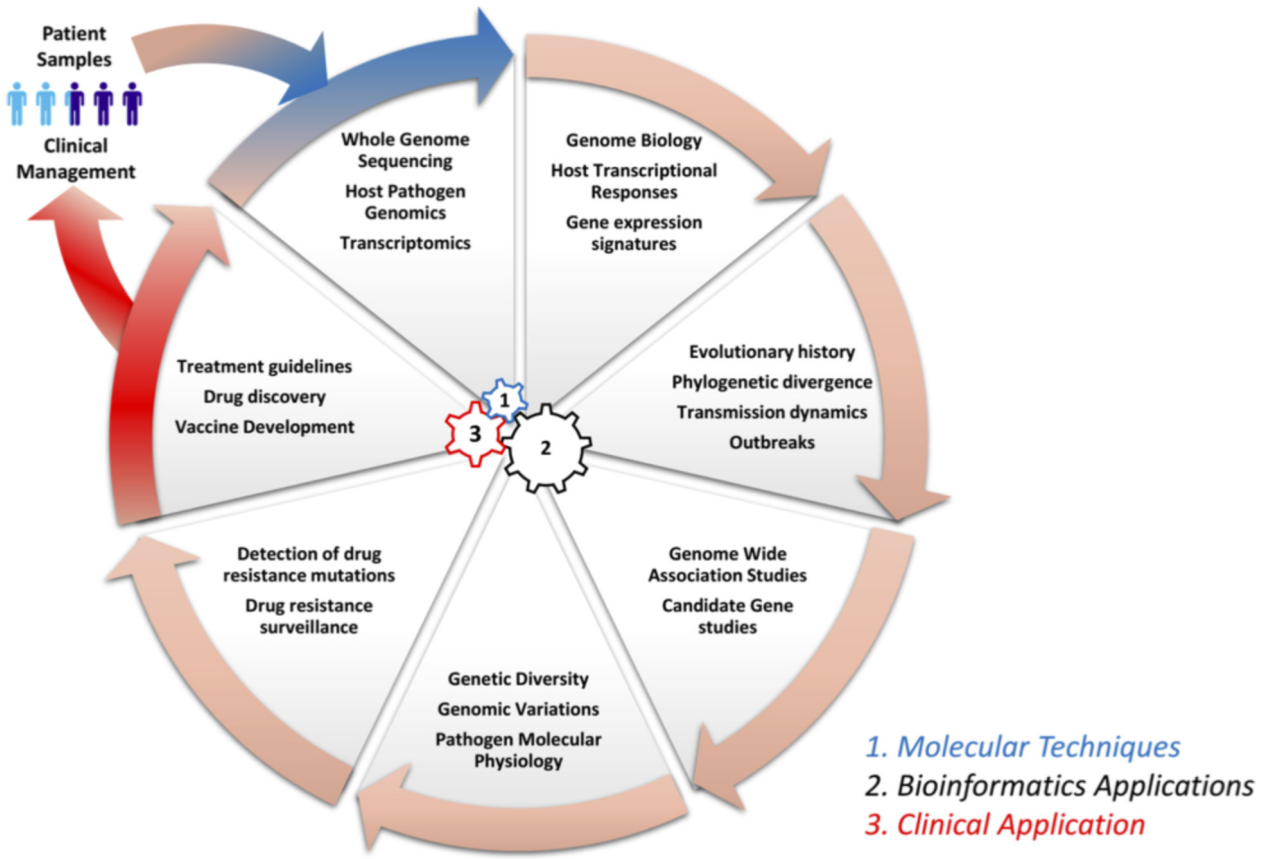

FIGURE 1 | Circle of infection disease genomics/bioinformatics. This shows a representation of genomic laboratory technologies, bioinformatics analysis, and potential applications. This includes: (1) molecular techniques such as whole genome sequencing by methods like Illumina to generate sequence reads, which are needed for the (2) bioinformatics application to study host systemic responses, pathogen genomics, and transmission dynamics. Further, bioinformatics can be applied to determine genetic diversity, investigation of drug resistance mechanisms and surveillance, and the identification of vaccine targets in systems vaccinology. Finally (3), all this information can be integrated to define treatment guidelines and patient management.

identifying SNPs in the *ASAP1* gene on chromosome 8q24 and in a genomic region in which class II human leucocyte antigen (HLA II) is encoded (Curtis et al., 2015; Sveinbjornsson et al., 2016). Recently, a GWAS study in a Han Chinese population also found SNPs in mitofusin-2 (*MFN2*), regulator of G protein signaling 12 (*RGS12*) and HLA II beta chain to be associated with active TB (Qi et al., 2017). This highlights that host genetics play significant roles in susceptibility to active TB and may explain why some individuals remain latently infected while some develop active TB despite having similar exposure levels. Furthermore, based on host genetic variants, GWAS analysis could be applied to identify latently infected individuals who are at a high risk of developing active TB for preventative interventions. Once validated, identified SNPs can be used to develop point of care diagnostics to identify high risk people for mass preventative treatment.

Host transcriptomics are increasingly being used to understand systemic responses to infections and to identify diagnostic and prognostic markers. Mistry et al. (2007) were among the first to use microarray technology to study host systemic response to TB, identifying a nine gene-signature with potential for TB diagnosis. Jacobsen et al. (2007) applied microarray analysis to investigate the host pathway biology and potential diagnostic biomarkers. Analyzing peripheral blood mononuclear cells (PBMCs), they found a monocyte-derived gene expression signature identifying CD64, lactoferrin and Ras-Associated GTPase-33A as potential diagnostic biomarkers, which were further validated in another independent study population in South Africa (Maertzdorf et al., 2011). Applying gene set enrichment analysis to microarray gene expression identified metabolic pathways such as insulin metabolism, immune cell differentiation and inflammation in TB (Lesho et al., 2011). A neutrophil-driven interferon signature consisting both type I and type II interferon during TB was also identified using microarray analysis (Berry et al., 2010). The type I interferon pathway was also observed by Ottenhoff et al. (2012) identifying IL15RA, UBE2L6, and GBP4 as the main molecules involved. A 393-transcript signature for active TB and an 86-transcript signature with a potential for distinguishing TB from other inflammatory diseases were also identified (Berry et al., 2010). In addition, a biosignature consisting of 27 transcript signatures to distinguish active from latent TB and 44 transcript signatures to

distinguish active TB from other diseases were recently identified (Kaforou et al., 2013). Microarrays have also been used to demonstrate that host transcriptional responses to *M. africanum* and *M. tuberculosis* differ following treatment (Tientcheu et al., 2015), which could be important in the management of patients infected with the different mycobacterial strains. Furthermore, host gene expression has also been used to monitor treatment responses and predict treatment outcome, which will be valuable in testing new drug regimens and new antimycobacterial drugs (Thompson et al., 2017). These studies prove the potential of host genomics in providing a better understanding of disease pathophysiology, prognosis and host pathway biology in response to an infectious agent.

In addition, arrays have also been applied to childhood TB, to identify signatures for active tuberculosis and a signature that distinguishes active tuberculosis from other diseases in sub-Saharan Africa (Anderson et al., 2014). Similarly, a 9-gene signature was also identified in Warao Amerindian children, further highlighting the potential of using host biomarkers for TB diagnosis (Verhagen et al., 2013). Host transcriptional analysis is moving from array-based technologies to RNA sequencing and has been applied to 16 gene signatures that identified people with a high risk of developing TB 2 years before diagnosis in sub-Saharan Africa (Zak et al., 2016). However, it is noteworthy that identified biosignatures have a variable number of genes, from about 10 to more than 100, and there is very little overlap between some signatures. It will be valuable to conduct a meta-analysis of available datasets to increase statistical power and identify high confidence signatures across studies regardless of circulating pathogens and local environmental factors. In doing such analysis, confounders due to technologies, age and circulating endemic pathogens can be accounted for to give a strong as well as diagnostic and prognostic signature. These studies highlight the potential application of genomics and bioinformatics to interrogate host response for the diagnosis and prognosis of TB, which will contribute immensely to curbing TB morbidity and mortality.

DENGUE VIRUS RESEARCH IN THE ERA OF BIOINFORMATICS

Dengue virus (DENV) is a pathogenic single-stranded RNA virus that belongs to the flavivirus genus, which comprises other known pathogenic viruses such as West Nile, yellow fever, Japanese encephalitis, St. Louis encephalitis, tick-borne encephalitis, Omsk hemorrhagic fever and Zika virus (Gould and Solomon, 2008). The re-emergence, evolution, diversity and geographic distribution of flaviviruses make them interesting pathogens (Moureau et al., 2015). Phylogenetic analysis of divergence times suggests that flaviviruses originated from a common ancestor (100,000 years ago) and later split into mosquito and tick borne flaviviruses about 40,000 years ago (Holbrook, 2017). Approximately 40% of the world population is at risk of DENV infection with more than 350 million cases reported annually.

Illumina SNPs genotyping and SNPs identified through whole genome analysis have been used in case-control GWAS statistical analysis to identify SNPs that predispose or confer protection against DENV infection (de Carvalho et al., 2017). The DENV shock syndrome (DSS) has been shown in a GWAS analysis of SNPs in a cohort of 2008 pediatric cases to have a strong association ($P < 0.5 \times 10^{-8}$) with the human major histocompatibility complex (*MHC*) (rs3132468) on chromosome 6 and phospholipase C (rs3740360 and rs3765524) on chromosome 10 (Khor et al., 2011). Dang et al. replicated the study in 917 Thai children with DSS and confirmed that alleles rs3132468 [MHC I chain related protein A (*MICB*)] and rs3765524 [phospholipase C epsilon 1 (*PLCEI*)] predispose Southeast Asians to DSS (Dang et al., 2014). In contrast, Whitehorn et al. (2013) genotyped 3,961 confirmed cases and 5,968 controls and found that rs3132468 *MICB* and rs3740360 alleles *PLCEI* were associated with less severe phenotypes of DENV infection in both infants and adults. This implies that the effect of these SNPs could be population-specific. Other candidate genes include dendritic cell-specific intracellular adhesion molecule (*ICAM*)*-3* grabbing non-integrin (*DC-SIGN*), C-Type Lectin Domain Containing 5A (*CLEC5A*), immunoglobulin gamma constant fragment receptor (*FCGRIIA*), Toll-Like receptors (*TLRs*), Tumor necrosis Factor (*TNF*), Interferons (*IFNs*), 2′-5′-oligoadenylate synthase (*OASs*), Janus Kinase (*JAK*), Stimulator of Interferon Genes (*STING*), cytokines, chemokines, *ICAM-1* and tryptase 1 proteases (de Carvalho et al., 2017).

Whole genome sequencing (WGS) and phylogenetic methods have been used to investigate DENV outbreaks. Faria et al. (2017) analyzing 92 viral genomes from DENV patients during the 2012 outbreak in Rio de Janeiro, found that at least two thirds of infections went unnoticed and their analysis highlighted the scale of the epidemic spread of DENV after the outbreak. Ahn et al. (2015) investigated the genetic variations in 8,826 nucleotide sequences of whole-genome DENV virus, and demonstrated that there was a distinctive genetic pattern between the four DENV subtypes across different regions (American, Oceanian, Asian, and Africa).

Analyses of envelope encoding nucleotide sequences from India have shown a shift from DENV subtype III to subtype IV, suggesting some level of positive selection (Manakkadan et al., 2013). These phylodynamic methods, which indicate evolutionary process or patterns of genetic diversity of the DENV virus, have also been reconciled with the virus epidemiology so as to decrease the variation between the two methods that are mainly used to study the population dynamics or viral behaviors (Pybus et al., 2012; Rasmussen et al., 2014). Due to the importance of genomics and bioinformatics in viral research, a range of tools has been developed to analyze viral genomes and make inferences (Stamatakis, 2014; Brody et al., 2017).

The use of RNA folding, structural predictions and functional studies has shown that genetic variation of the DENV occurs in nature due to high rates of recombination and error-prone RNA polymerases. A deleterious DENV genome was first shown by Aaskov et al. (2006) whereby a stop codon in the envelope coding region resulted in a defective DENV. Li et al. (2011)

also discovered defective interfering viral particles by analyzing short fragments of DENV, suggesting that they may be part of a broader disease attenuating process mediated by the deleterious virus and the defective interfering particles are important in viral replication, thereby enhancing the overall transmission capability of DENV (Li and Aaskov, 2014). Structural RNA predictions have implicated other elements in modulating replication of the virus, such as the downstream cyclization sequence (Friebe et al., 2012), cis-acting elements occurring in the capsid coding region (de Borba et al., 2015), and elements in the promoter Stem Loop A (SLA) and non-structural protein 5 (NS5) regions (Gebhard et al., 2011).

Understanding intra- and inter-host genetic diversity was previously mired with experimental and analytical methods that did not fully account for errors in viral amplifications. Thai et al. (2012) used various statistical approaches to correct for the artefactual mutations resulting from PCR amplifications and sanger sequencing, and showed that the genetic diversity index (Pi) of the DENV was low, ranging from 0 to 0.0013. This suggested sequence conservation, but they were able to show mixed infections and phylogenetically distinct DENV lineages present within the same host. Furthermore, genome-wide scans for patterns of intra-host diversity in DENV identified variants between genes suggesting significant differences in intra-host diversity of the virus in the Nicaraguan population (Parameswaran et al., 2012). Functional annotation of the variants showed the impact of viral mutations on protein function, which strongly suggested purifying selection across transmission events.

Deep sequencing, RNA structural analysis and fitness evaluation have been used to determine processes that DENV employs for host specialization (mosquito or human) using RNA elements in the 3′-UTR (Villordo et al., 2015). A host adaptable stem loop structure was found to be duplicated, which DENV uses to accumulate mutations that are beneficial in one host and deleterious in another host, but the duplication confers a robust mechanism during host switching (Villordo et al., 2015). Recently, Waman et al. (2016) used population genetics methods to compute the genotype diversity and evolution of 990 DENV genomes, and revealed that the DENV-2 population is subdivided into 15 lineages. Their study also indicated the presence of intra-genotype diversity and that the population structure of DENV-2 is spatiotemporal, shaped by episodic positive selection and viral recombination (Waman et al., 2016). The application of genomics and bioinformatics in the study of DENV shows the complexity of the virus biology, which can be exploited in target identification for drug discovery and vaccine development (Guy et al., 2016; Low et al., 2017).

PROGRESS IN MALARIA GENOMICS

Malaria incidence and mortality rates decreased by 21 and 29%, respectively, between 2010 and 2015 (WHO, 2018b). The genetic landscape of *P. falciparum,* the main cause of malaria, is increasingly being unraveled by using deep sequencing to identify polymorphisms and structural and copy number variations, which are fundamental for parasite evolution (Kwiatkowski, 2015). Sequencing consortia such as the MalariaGEN improve our understanding of genomics of both the Anopheles vector and the plasmodium species[1]. A recent study on genotyping accuracy using deep sequencing of *Plasmodium* parental generations and their progenies revealed that polymorphism frequencies can be used as markers of high recombination rates (Miles et al., 2016), which is an important contributor to enhancing immune evasion and drug resistance. Using whole genome deep sequencing and micro-array analysis, a study observed 18 deletions on regions encoding multigene families that are associated with immune evasion (Bopp et al., 2013). The authors showed the presence of chromosomal crossovers in six of the deletions and were able to estimate mutation rates of *P. falciparum* (Bopp et al., 2013).

Bioinformatics has contributed to our understanding of resistant mechanisms to previous drugs such as chloroquine and the emerging resistance to artemisinin-based combination therapies (ACT). Robinson et al. deployed next generation sequencing to investigate multi-clonality, population genetics and drug-resistant genotypes (Robinson et al., 2011). More recently, WGS was used to discover that mutations in the Kelch propeller domain (K-13) are associated with ACT resistance in Cambodia (Ariey et al., 2014; Straimer et al., 2015). Profiling of the drug resistance genes [*P. falciparum* chloroquine resistance transporter (*pfcrt)*, *P. falciparum* multidrug resistance (*pfmdr1*), *P. falciparum* dihydrofolate reductase (*dhfr*) and *P. falciparum* dihydropteroate synthetase (*dhps*), and *P. falciparum* Kelch protein 13 (*pfk13*)] was done using Illumina next generation sequencing and demonstrated that the resistance-associated K-13 variants were largely absent in Africa (MalariaGEN Plasmodium falciparum Community Project, 2016; Nag et al., 2017).

Furthermore, bioinformatics tools have been used to demonstrate multi-locus linkage disequilibrium and local diversity, recent selection through integrated haplotype scores, regional gene flow and allele frequency differentiations (Duffy et al., 2017). Intra-host diversity can now be statistically characterized using the F*ws* metrics because sequencing platforms are able to generate read count data. Auburn et al. characterized within host diversity in 64 samples from West Africa, capturing a multiplicity of infections, number of clone ratios, clonal variation and within-host diversity (Auburn et al., 2012). Bioinformatics analysis of deep sequencing revealed large-scale genetic variations in *P. falciparum* (86158 SNPs), and genome wide allelic frequencies, population structure, linkage disequilibrium and intra-host diversity (Manske et al., 2012). The genetic diversity of *P. falciparum* is dependent on directional and balancing selection, whereby drug pressure and host immunity are the major selective agents, respectively (Mobegi et al., 2014; Duffy et al., 2015).

Genomics has been used to discover novel malaria resistance loci in humans, which provide 33% protection from severe malaria (Malaria Genomic Epidemiology Network, 2015). In Ghana, GWAS identified two unknown genetic loci associated with severe malaria: 1q32 within the ATPase Plasma Membrane Ca^{2+} Transporting 4 (*ATP2B4*) gene and the 16q22.2 linked

[1]https://www.malariagen.net/

to a tight junction protein known as *MARVELD3* (Timmann et al., 2012). Most recently, GWAS was used in a longitudinal surveillance to detect K-13 signatures, which led to the identification of a Kelch variant that is suggested to be a potential modulator of artemisinin resistance (Cerqueira et al., 2017).

The *Plasmodium* pathophysiology is increasingly being explored using transcriptomics and proteomics. Bioinformatics and statistical models have been used to describe the genome-wide translational dynamics of *P. falciparum*, showing that parasite transcription and translation are tightly coupled presenting a broad and high resolution of parasite gene expression profiles (Caro et al., 2014). ChIP-Seq and RNA sequencing have been used for polysome profiling to understand the regulation of *Plasmodium* gene expression in humans. Bunnik et al. (2013) observed a delay in peak polysomal transcript abundance for several genes as compared to the mRNA fraction, which they reported to be alternative polysomal mRNA splicing events of non-coding transcripts.

DNA microarray technologies had been used to describe the gene expression patterns of *P. falciparum* during the intra-erythrocytic stage (Bozdech et al., 2003), gametocyte (Young et al., 2005), sporozoite (Siau et al., 2008), liver stage (Tarun et al., 2008), and even between three different strains (Llinás et al., 2006). Recently, microarrays have been used to characterize parasite transcriptomes during cerebral and asymptomatic malaria, which revealed some differentially expressed genes encoding proteins involved in protein trafficking, Maurer's cleft proteins, transcriptional factor proteins and several hypothetical proteins (Almelli et al., 2014). RNA sequencing has also been used to describe *P. falciparum* expression profiles at different time points and has found novel gene transcripts, alternative splicing events and predicted untranslated regions of some genes providing further information on the parasite biology (Otto et al., 2010). Yamagishi et al. (2014) simultaneously analyzed the human host and the parasite transcriptomes using RNA sequencing, and showed that several human and parasite genes such as Toll-like receptor 2 and TIR domain-containing adapter molecule 2 (*TICAM2*) correlated with clinical symptoms. RNA sequencing has also been employed to study the transcriptome of *P. vivax,* which revealed a hotspot of *vir* genes on chromosome 2, new gene transcripts and the presence of species-specific genes (Zhu et al., 2016). It would be valuable to compare this data with similar data from other related *Plasmodium* species to identify species-specific transcriptomes. Analyzing the transcriptome of Chloroquine sensitive and resistant parasites identified 89 upregulated genes and 227 downregulated genes that were associated with resistance (Antony et al., 2016). These differentially expressed genes are involved in immune evasion mechanisms, pathogenesis, and various host-parasite interactions and could be targeted for drug and vaccine development.

Currently, single-cell RNA sequencing is revolutionizing the study of cell-to-cell heterogeneity. For example, the use of this method led to the discovery of novel variations in the expression of specific gene families that are involved in host-parasite interactions among asexual populations (Reid et al., 2018). Altogether, these studies demonstrate the profound impact of malaria parasite transcriptomics and genomics on our understanding of the parasite (Lee et al., 2017), and identify possible candidate targets for drugs, vaccines and diagnostics (Ludin et al., 2012; Hoo et al., 2016).

GENOMICS RESEARCH IN FILARIASIS

Filariasis is a neglected chronic disease caused by tissue-dwelling nematodes (filariae) with onchocerciasis and lymphatic filariasis (LF), causing significant health concerns with a disease burden approaching 86 million cumulatively (WHO/Department of Control of Neglected Tropical Diseases, 2016). Onchocerciasis is caused by *Onchocerca volvulus* while LF is caused by three different parasites, namely *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori* (Taylor et al., 2010). Elimination of filariasis is challenging because of the unavailability of sensitive diagnostic tools, lack of appropriate treatments and inadequate control measures in resource limited countries.

The *W. bancrofti* and *O. volvulus* genomes have been sequenced, providing opportunities for further genomic analyses (Desjardins et al., 2013; Cotton et al., 2016). Bioinformatics revealed the presence of gene coding for host immune system regulators such as human-like autoantigens as well as serine and cysteine protease inhibitors (Molehin et al., 2012; Cotton et al., 2016).

Molecular studies coupled with computational analyses have demonstrated an association between human host factors and filariasis clinical manifestations. LF infections have been shown to cluster in some families using pedigree studies (Cuenco et al., 2004; Chesnais et al., 2016). These studies show that genetic factors are involved in the regulation of LF infections and affect both the presence and intensity of microfilariae. However, a GWAS would be more comprehensive to demonstrate this genetic susceptibility to LF as has been the case for a tropical lymphedema (Podoconiosis) of non-filarial origin (Tekola Ayele et al., 2012). It is worth mentioning that lymphedema, or elephantiasis, is one of the main features of LF and normally occurs as a result of a compromised lymphatic system (Addiss, 2010). As opposed to LF, which is infectious, Podoconiosis is a non-communicable disease caused by soil particles such as aluminum and silica predominant in volcanic regions (Price, 1976; Davey et al., 2007). A comparative genomics-based study of LF would help to better understand these clinical manifestations.

Most of the pathological features of LF are associated with human-immunogenetics (Taylor, 2003; Junpee et al., 2010), which has been investigated using genomics and bioinformatics. Gene candidate-based genomics studies carried out in Thailand revealed that polymorphisms in the *TLR-2* gene (−196 to −173 deletion, +597 T > C and +1350 T > C) have a strong linkage disequilibrium and were associated with increased risk of asymptomatic LF (Junpee et al., 2010). In a functional study, individuals with the −196 to −173 deletion were found to have significantly low transcription levels compared to those with the wild-type gene (Junpee et al., 2010). Further analyses showed strong association of a mutation (M196A) in human tumor necrosis factors (TNF) receptor-II with hydrocele development,

while the A288S mutation of endothelin-1 (ET-1) correlated with low ET-1 1 plasma levels and elephantiasis (Panda et al., 2011).

Population genetics is very important for assessing and understanding the epidemiology and transmission dynamics of filarial diseases (Small et al., 2016; Doyle et al., 2017). Population genomics of *O. volvulus* samples collected from different geographical zones – West Africa (WA), Uganda and Ecuador – demonstrated some level of population structure between WA and other populations (Choi et al., 2016). Furthermore, phylogenetic signals indicative of gene flow and genetic admixture between WA forest and savanna populations were identified. These signals could serve as markers to delineate forest from savanna populations and/or sort out admixed populations (Choi et al., 2016). A study using both nuclear and mitochondrial sequences identified regions in the *W. bancrofti* genome that exhibited an arrangement which was consistent with both balancing and directional selection (Small et al., 2016).

The control of filariasis in general is difficult due to the complex parasite life cycle. In an attempt to demystify the complex life cycle of the parasite, RNA sequencing has been used to investigate gene expression profiles of different developmental stages of *Brugia malayi* (Choi et al., 2011). Transcriptomics analyses revealed stage-specific gene expression correlating with stage-specific pathway activation. Upregulated proteins included cathepsin L and Z-like cysteine proteases that were previously demonstrated to be essential for larva molting in *O. volvulus* (Lustigman et al., 2004) and cuticle and eggshell remodeling in filarial nematodes in general (Guiliano et al., 2004). Another study using a filarial microarray chip composed of 18,104 gene probes revealed that gene expression in *B. malayi* infective larvae (L3s) depends on environmental factors (Li et al., 2009). The gene expression patterns in irradiated L3s, laboratory-adapted L3s and those collected from mosquitoes were found to be different. Gene Ontology analyses showed that upregulated genes in laboratory-adapted and mosquito-derived L3s were mostly involved in growth and invasion, whereas those in irradiated L3s were enriched with immunogenic proteins and proteins involved in radiation repair (Li et al., 2009). Such high throughput genomics analysis is important for understanding the biology/development, invasion, and immune evasion mechanisms of the parasite and could help improve disease control measures (Choi et al., 2011).

Mass drug treatment with Ivermectin (IVM) or Mectizan® and Albendazole is the main strategy for filariasis control in Africa and has been going on for decades (Amazigo, 2008). However, cases of drug resistance have been reported and genomic methods are increasingly being used to investigate mechanisms of resistance. Genotyping and sequencing studies have shown an association between SNPs in some *O. volvulus* genes (P-glycoprotein-like protein, β-tubulin) and the development of resistance (Nana-Djeunga et al., 2012; Osei-Atweneboana et al., 2012). P-glycoprotein was recently demonstrated to be associated with resistance to IVM in a horse filarial species (cyathostomins) with transcript levels measured by RNA-Seq and confirmed by RT q-PCR found to be significantly higher in the resistant compared to sensitive worm population (Peachey et al., 2017). Moreover, GWAS demonstrated that reduced sensitivity of *O. volvulus* to IVM is accounted for by genetic drift and soft selective sweeps. Pooled next generation sequencing of *O. volvulus* worms collected from Ghana and Cameroon repeatedly treated with IVM and phenotypically characterized into poor responder (PR) and good responder (GR) parasites identified genetic variants that considerably delineate GR and PR parasites. One of these variants (SNP, OM1b_7179218) was common in both Cameroon and Ghana worm populations, whereas the others were country-specific (Nana-Djeunga et al., 2014; Doyle et al., 2017). These variants were found to be grouped in quantitative trait loci (QTLs) in which published genes associated with IVM resistance were scarcely found. Gene Ontology[2] analysis revealed that genes found in those QTLs regions were linked to pathways involved in neurotransmission, development, and stress responses (Harris et al., 2004; Doyle et al., 2017). The involvement of neurotransmission is a promising finding here because one of the main targets of IVM is a ligand-gated channel at neuromuscular junctions (Cully et al., 1994).

The molecular mechanism of Ivermectin is not clearly understood and has been investigated using bioinformatics approaches. RNA-Seq analyses of ivermectin-challenged *B. malayi* adult female worms revealed that genes involved in cell division (meiosis) and oxidative phosphorylation were drastically downregulated as early as 24 h post-exposure (Ballesteros et al., 2016). A similar study in which the worms were instead challenged with flubendazole (FLBZ), a potential macrofilaricide, demonstrated the effect of FLBZ on embryogenesis and cuticle integrity (O'Neill et al., 2016a). Expression of cuticle-related genes and those involved in mitosis or meiosis were notably affected by the treatment. These studies further elucidate the drug-induced inhibition of embryogenesis and microfilarial release from the female worm uterus during larval development as previously demonstrated (O'Neill et al., 2015, 2016b). Knowledge of this mechanism could help in drug repurposing whereby drugs known to have a similar mode of action or mechanism, but are used for the treatment of other parasitic diseases, could be tested for their efficacy on filarial parasites.

APPLICATION OF OMICS TO VACCINE TARGET IDENTIFICATION AND DRUG DISCOVERY

The availability of whole genome sequences of both the host and pathogens in different databases such as GenBank[3] (Benson et al., 2004), EuPathDB ([4]formerly ApiDB), WormBase[5], Virus Pathogen Database and Analysis Resource (ViPR) has led to tremendous advances in the search for new drug and vaccine targets (Yan et al., 2015; Xia, 2017). This enables high throughput *in silico* screening for the identification of vaccine and drug targets, thus focusing expensive laboratory screening on selected high affinity targets. Though not yet fully implemented in Africa,

[2]http://geneontology.org/

[3]http://www.ncbi.nlm.nih.gov

[4]http://EuPathDB.org

[5]http://www.wormbase.org

omics technologies and bioinformatics analyses have aided significantly in the generation of new knowledge toward drug and vaccine target discovery (Yan et al., 2015; Xia, 2017). Genomic, transcriptomic and proteomic analyses of pathogens such as *filariasis* parasites have identified new potential biomarkers that can be invaluable in diagnostics, vaccine and drug development (Armstrong et al., 2016; Bennuru et al., 2017). Kumar et al. (2007), using genome wide *C. elegans* RNA-interference data as proxy, identified a set of 3,059 essential genes in the *B. malayi* genome, from which 589 were characterized as potential drug targets. The prioritization algorithm helps in the prediction of the efficacy, selectivity and tractability of each target.

Phylogenomic analyses across *Plasmodium* spp. and comparative genomic studies in humans have led to the identification of new drug targets in *P. falciparum*. Identification of essential genes (targets) responsive to specific inhibitors led to the discovery of 40 potential drug targets, which includes known ones such as calcium dependent protein kinase and previously unknown ones such as phosphoisomerase and carboxylase (Ludin et al., 2012). Comparing the transcriptomes of six *Plasmodium* spp. during blood stage infection revealed about 800 genes that have similar expression patterns across species, among which 240 were demonstrated to be druggable by online drug target prioritization databases (Hoo et al., 2016). Similarly, genomic and transcriptomic analyses have been carried out with other pathogens with encouraging results in fungi (Kaltdorf et al., 2016), bacteria (Turab Naqvi et al., 2017), and viruses (Dapat and Oshitani, 2016).

In vaccine target identification, pathogen genomes are being scanned in a bid to identify genes encoding proteins or molecules with vaccine candidate properties such as low antigenic variation, polymorphism, and immunogenicity (Masignani et al., 2002; De Groot et al., 2008). Despite the success of whole-organism vaccines such as those for polio, whole-organism vaccines for pathogens such as *Plasmodium spp.*, *Mycobacterium spp.* and HIV remain a challenge (Doolan et al., 2014; Proietti and Doolan, 2015). Genomics offers a potential way around this challenge through the discovery of immunogenic antigens using whole-genome scans (Doolan et al., 2014; Proietti and Doolan, 2015). Here, omics techniques and bioinformatics tools are used to determine genes or proteins that are involved in the virulence of the pathogen and pathogenesis of the disease by comparing, for example, attenuated and pathogenic disease agents. Algorithms can be used to predict T cell epitopes or regions with high affinity within HLA molecules in translated peptides found in databases (Grubaugh et al., 2013; Davies et al., 2015) in order to inform the choice of the right antigens for vaccine design. Omics technologies have been reviewed in the context of vaccine target identification by He (2012).

Most of the tools used for epitope identification rely on statistics and machine learning. Some of them include servers to predict MHC-binding, peptides namely RANKPEP (Reche et al., 2004), which uses Position Specific Scoring Matrices (PSSMs), and nHLAPred[6] (Bhasin and Raghava, 2007), based on Artificial Neural Networks (ANNs) and quantitative matrices among others. Some severs are specific for B-cell epitope prediction, such as Bcepred[7] (Saha and Raghava, 2004), ABCpred[8] (Saha and Raghava, 2006), and BepiPred[9] (Jespersen et al., 2017). These tools work based on the physicochemical properties and location of the peptides. They function alongside epitope-containing databases such as Swiss-Prot, SYFPEITHI, and IEDB (Fleri et al., 2017). The list of tools, methods and databases mentioned here is not exhaustive, however, they have been extensively reviewed elsewhere (Soria-Guerra et al., 2015).

Nowadays, due to advances in the fields of computer sciences, genomics, proteomics, bioinformatics and management of patients' health records, etc., there seems to be a paradigm shift from generalized medicine to personalized therapy (Sorber et al., 2017). For example, many drugs are metabolized by cytochrome P450 enzymes with drug action depending on the expressed gene variant (BlueCross and BlueShield Association, 2004; Daly et al., 2006). Moreover, malaria patients with glucose-6-phosphate (G6p) deficiency have been reported with severe complications such as cardiotoxicity and acute hemolytic anemia following treatment with quinidine gluconate (Damhoff et al., 2014). These complications have been described as a consequence of inherited (X-linked trait) mutations in the *g6p* gene (Luzzatto and Seneca, 2014). These mutations do not cause the complete loss of the G6P enzyme but instead affect its stability and level in red blood cells (Luzatto et al., 2001). In the same line rifampicin, which is the drug of choice for TB treatment, is transported after administration by a human anion transporter encoded by the *SLCO1B1* gene. Studies have shown that mutations in the *SLCO1B1* gene, namely rs11045819 and rs4149032, are associated with decreased RIF plasma levels in South-African populations (Weiner et al., 2010; Chigutsa et al., 2011; Gengiah et al., 2014). However, this finding could not be replicated in Malawian and South Indian populations, implying that this could be population-specific (Ramesh et al., 2016; Sloan et al., 2017). These show, in a nutshell, the implication of genomics and bioinformatics in drug discovery and precision therapy (Hamburg and Collins, 2010; Rabbani et al., 2016).

CHALLENGES AND OPPORTUNITIES IN CONDUCTING OMICS AND BIOINFORMATICS STUDIES IN AFRICA

Bioinformatics is increasingly becoming an important cornerstone in contemporary research on infectious diseases (Mulder et al., 2017), where Africa has the highest morbidity and mortality but less genomics research output compared to other regions of the world (Fatumo et al., 2014; Karikari, 2015). This slow pace of genomics research output is due to several challenges in omics and bioinformatics research facilities in Africa; three of the major ones are briefly discussed.

[6]http://www.imtech.res.in/raghava/nhlapred/

[7]http://www.imtech.res.in/raghava/bcepred/

[8]http://www.imtech.res.in/raghava/abcpred/

[9]http://www.cbs.dtu.dk/services/BepiPred/

Inadequate Infrastructure

Bioinformatics and genomics analysis require powerful computers and a reliable source of electricity for large data storage and high throughput analyses (H3Africa Consortium et al., 2014). With the exception of some South African universities, most sub-Saharan African universities lack high performance computing facilities (Karikari et al., 2015; Mulder et al., 2016). There is also a limitation of high-speed internet for sharing data and accessing bioinformatics databases and repositories (Fatumo et al., 2014; Karikari, 2015). This hinders the application of cloud-based web services which could have circumvented the need for local high-performance computing facilities (Navale and Bourne, 2018). Furthermore, few research institutions in Africa have sequencing facilities and therefore resort to sequencing abroad through collaborations. Such collaborations often result in a loss of ownership of the data and resulting publications usually have the external collaborators as lead and correspondence authors. Notable efforts being made to bridge this infrastructural gap include the installation of high-performance computers (HPCs) at The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) at the University of Science Technique and Technologies of Bamako, Mali, the West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), University of Ghana and the Medical Research Council Unit, The Gambia at the London School of Hygiene and Tropical Medicine, to support storage and high throughput analyses of genomic data. These HPC facilities are complemented by NGS sequencing facilities at WACCBIP and MRC in addition to some institutions in East Africa such the International Livestock Research Institute (ILRI-Kenya). This infrastructural development, and pressure from initiatives such as Human Heredity and Health in Africa (H3Africa), will hopefully serve as a springboard for Africa to increase her involvement in the study design, sample collection, analysis and ownership of data rather than just collecting samples for international collaborators.

Lack of Training Opportunities and Well-Structured Bioinformatics Courses

Until the recent introduction of bioinformatics training courses by H3ABioNet, there were limited bioinformatics training courses in Africa. Such training programs were mostly short courses organized by local bioinformaticians with support from experts in the field across Africa and other external collaborators (Gurwitz et al., 2017). Very few African universities have structured bioinformatics courses, most of these universities are South African, while some are North African and few are in sub-Saharan Africa (Bishop et al., 2015). The DELGEME, through funding from the Wellcome Trust, is also providing funding for Master of Science courses in bioinformatics, which are mostly done in South Africa. The other form of bioinformatics training is through local capacity building, which institutions organize for staff with support usually through North-South collaborations and transfer of expertise. However, the downside of short courses is that there is no mentorship beyond the course, which hinders consolidation of the knowledge gained. In addition to these, some organizations working predominantly on crop production, such as the International Institute of Tropical Agriculture Bioscience Center[10] and Consultative Group on International Agricultural Research institute[11], offer short bioinformatics training opportunities to African scholars. Sometimes some students from Africa get training from European universities, but the challenge is that most of the trainees do not come back to join local institutions because of poor infrastructures. Furthermore, there is a disconnect between biologists and other scientific disciplines such as computer science, statistics and mathematics in most African universities. This affects multidisciplinary research, which is crucial in modern-day infectious disease research. Ultimately, the lack of well-structured bioinformatics curricula hampers the development and maintenance of highly needed experts in the field in Africa, since they often move to Europe and North America for better career prospects.

Limited Research Funding

A major challenge to research on the African continent is the lack of funding for biomedical research. Current research is mainly funded from international donors, with limited or no funding from national governments and African regional bodies such as the African Union (Hamburg and Collins, 2010; Karikari, 2015). However, a few countries such as South Africa, through the South Africa's National Research Foundation and Medical Research Council, do provide funding for genomics research projects (Karikari et al., 2015). Until the initiation of H3Africa, through funding from the National Institute of Health (United States) and the Wellcome Trust (United Kingdom), there was limited to no funding for genomics and bioinformatics in Africa (Adoga et al., 2014; Mulder et al., 2017).

CONCLUSION AND PERSPECTIVE

Herein we highlight how genomics and bioinformatics has contributed to our understanding of infectious diseases of significant health concern, ranging from bacterial and viral to parasitic infections, as well as their applications to drug and vaccine target identification. This ranges from understanding pathogenesis, host systemic responses and host-pathogen interactions to identification of prognostic and diagnostic markers. However, in Africa, despite the high morbidity and mortality due to infectious diseases, there is limited expertise in the field of bioinformatics and hence limited bioinformatics research output in terms of publications. Thus, there is a need to strengthen training and capacity building in bioinformatics in Africa to improve infectious disease genomics and host-pathogen genomics on the continent. This can be achieved through the establishment of well-structured courses, mentorship for junior

[10] http://bioscience.iita.org/index.php/en/services/bioinformatics

[11] https://www.cgiar.org/

and trainee bioinformaticians and better career prospects to maintain trained bioinformaticians on the continent.

AUTHOR CONTRIBUTIONS

All authors listed contributed substantially to the intellectual, writing and editing, and approved the manuscript for publication.

FUNDING

All authors were supported by a DELTAS Africa grant (DEL-15-007: GA). 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 is supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency), with funding from the Wellcome Trust (107755/Z/15/Z: GA) and the United Kingdom 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 United Kingdom Government.

ACKNOWLEDGMENTS

We are grateful to WACCBIP for providing us with the funding and conducive environment to do quality research.

REFERENCES

Aaskov, J., Buzacott, K., Thu, H. M., Lowry, K., and Holmes, E. C. (2006). Long-term transmission of defective RNA viruses in humans and aedes mosquitoes. *Science* 311, 236–238.

Addiss, D. G. (2010). Global elimination of lymphatic filariasis: addressing the public health problem. *PLoS Negl. Trop. Dis.* 4:e741. doi: 10.1371/journal.pntd.0000741

Adoga, M. P., Fatumo, S. A., and Agwale, S. M. (2014). H3Africa: a tipping point for a revolution in bioinformatics, genomics and health research in Africa. *Source Code Biol. Med. BioMed Central* 9:10. doi: 10.1186/1751-0473-9-10

Ahn, I., Jang, J.-H., Han, Y., and Tung, T. Q. (2015). "Phylogenetic analysis of dengue viruses using bioinformatics techniques," in *Paper Presented at the Bioscience and Medical Research 2015*, Cambridge, MA, 12–18.

Almelli, T., Nuel, G., Bischoff, E., Aubouy, A., Elati, M., Wang, C. W., et al. (2014). Differences in gene transcriptomic pattern of *Plasmodium falciparum* in children with cerebral malaria and asymptomatic carriers. *PLoS One* 9:e114401. doi: 10.1371/journal.pone.0114401

Amazigo, U. (2008). The african programme for onchocerciasis control (APOC). *Ann. Trop. Med. Parasitol.* 102(Suppl. 1), 19–22.

Anderson, S. T., Kaforou, M., Brent, A. J., Wright, V. J., Banwell, C. M., Chagaluka, G., et al. (2014). Diagnosis of childhood tuberculosis and host RNA expression in Africa. *N. Engl. J. Med.* 370, 1712–1723. doi: 10.1056/NEJMoa1303657

Antony, H. A., Pathak, V., Parija, S. C., Ghosh, K., and Bhattacherjee, A. (2016). Transcriptomic analysis of chloroquine-sensitive and chloroquine-resistant strains of *Plasmodium falciparum*: toward malaria diagnostics and therapeutics for global health. *OMICS* 20, 424–432. doi: 10.1089/omi.2016.0058

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.-C., Khim, N., et al. (2014). A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 505, 50–55. doi: 10.1038/nature12876

Armstrong, S. D., Xia, D., Bah, G. S., Krishna, R., Ngangyung, H. F., LaCourse, E. J., et al. (2016). Stage-specific proteomes from *Onchocerca ochengi*, sister species of the human river blindness parasite, uncover adaptations to a nodular lifestyle. *Mol. Cell. Proteom.* 15, 2554–2575. doi: 10.1074/mcp.M115.055640

Auburn, S., Campino, S., Miotto, O., Djimde, A. A., Zongo, I., Manske, M., et al. (2012). Characterization of within-host *Plasmodium falciparum* diversity using next-generation sequence data. *PLoS One* 7:e32891. doi: 10.1371/journal.pone.0032891

Ballesteros, C., Tritten, L., O'Neill, M., Burkman, E., Zaky, W. I., Xia, J., et al. (2016). The effects of ivermectin on *Brugia malayi* females in vitro: a transcriptomic approach. *PLoS Negl. Trop. Dis*. 10:e0004929. doi: 10.1371/journal.pntd.0004929

Bennuru, S., O'Connell, E. M., Drame, P. M., and Nutman, T. B. (2017). Mining filarial genomes for diagnostic and therapeutic targets. *Trends Parasitol*. 34, 80–90. doi: 10.1016/j.pt.2017.09.003

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2004). GenBank. *Nucleic Acids Res.* 33, D34–D38.

Berry, M. P. R., Graham, C. M., McNab, F. W., Xu, Z., Bloch, S. A. A., Oni, T., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466, 973–977. doi: 10.1038/nature09247

Bhasin, M., and Raghava, G. P. S. (2007). A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. *J. Biosci.* 32, 31–42.

Bishop, Ö. T., Adebiyi, E. F., Alzohairy, A. M., Everett, D., Ghedira, K., Ghouila, A., et al. (2015). Bioinformatics education-perspectives and challenges out of Africa. *Brief. Bioinform*. 16, 355–364. doi: 10.1093/bib/bbu022

BlueCross and BlueShield Association (2004). Special report: genotyping for cytochrome P450 polymorphisms to determine drug-metabolizer status. *Technol. Eval. Cent. Assess. Program Exec. Summ*. 19, 1–2.

Bopp, S. E. R., Manary, M. J., Bright, A. T., Johnston, G. L., Dharia, N. V., Luna, F. L., et al. (2013). Mitotic evolution of *Plasmodium falciparum* shows a stable core genome but recombination in antigen families. *PLoS Genet.* 9:e1003293. doi: 10.1371/journal.pgen.1003293

Bozdech, Z., Llinás, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* 1:e5. doi: 10.1371/journal.pbio.0000005

Bradley, P., Gordon, N. C., Walker, T. M., Dunn, L., Heys, S., Huang, B., et al. (2015). Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. *Nat. Commun*. 6:10063. doi: 10.1038/ncomms10063

Brody, T., Yavatkar, A. S., Park, D. S., Kuzin, A., Ross, J., and Odenwald, W. F. (2017). Flavivirus and filovirus evoprinters: new alignment tools for the comparative analysis of viral evolution. *PLoS Negl. Trop. Dis.* 11:e0005673. doi: 10.1371/journal.pntd.0005673

Bunnik, E. M., Chung, D.-W. D., Hamilton, M., Ponts, N., Saraf, A., Prudhomme, J., et al. (2013). Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome Biol.* 14:R128. doi: 10.1186/gb-2013-14-11-r128

Caro, F., Ahyong, V., Betegon, M., and DeRisi, J. L. (2014). Genome-wide regulatory dynamics of translation in the *Plasmodium falciparum* asexual blood stages. *eLife* 3:e04106. doi: 10.7554/eLife.04106

Cerqueira, G. C., Cheeseman, I. H., Schaffner, S. F., Nair, S., McDew-White, M., Phyo, A. P., et al. (2017). Longitudinal genomic surveillance of *Plasmodium falciparum* malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. *Genome Biol.* 18:78. doi: 10.1186/s13059-017-1204-4

Chesnais, C. B., Sabbagh, A., Pion, S. D., Missamou, F., Garcia, A., and Boussinesq, M. (2016). Familial aggregation and heritability of *Wuchereria bancrofti* infection. *J. Infect. Dis.* 214, 587–594. doi: 10.1093/infdis/jiw212

Chigutsa, E., Visser, M. E., Swart, E. C., Denti, P., Pushpakom, S., Egan, D., et al. (2011). The SLCO1B1 rs4149032 polymorphism is highly prevalent in South Africans and is associated with reduced rifampin concentrations: dosing

implications. *Antimicrob. Agents Chemother.* 55, 4122–4127. doi: 10.1128/AAC.01833-10

Choi, Y.-J., Ghedin, E., Berriman, M., McQuillan, J., Holroyd, N., Mayhew, G. F., et al. (2011). A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, *Brugia malayi*. *PLoS Negl. Trop. Dis.* 5:e1409. doi: 10.1371/journal.pntd.0001409

Choi, Y.-J., Tyagi, R., McNulty, S. N., Rosa, B. A., Ozersky, P., Martin, J., et al. (2016). Genomic diversity in *Onchocerca volvulus* and its *Wolbachia* endosymbiont. *Nat. Microbiol.* 2:16207. doi: 10.1038/nmicrobiol.2016.207

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.

Coll, F., Mallard, K., Preston, M. D., Bentley, S., Parkhill, J., McNerney, R., et al. (2012). SpolPred: rapid and accurate prediction of *Mycobacterium tuberculosis* spoligotypes from short genomic sequences. *Bioinformatics* 28, 2991–2993. doi: 10.1093/bioinformatics/bts544

Coll, F., McNerney, R., Preston, M. D., Guerra-Assunção, J. A., Warry, A., Hill-Cawthorne, G., et al. (2015). Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med.* 7:51. doi: 10.1186/s13073-015-0164-0

Coll, F., Phelan, J., Hill-Cawthorne, G. A., Nair, M. B., Mallard, K., Ali, S., et al. (2018). Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* 50, 307–316. doi: 10.1038/s41588-017-0029-0

Cotton, J. A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., et al. (2016). The genome of *Onchocerca volvulus*, agent of river blindness. *Nat. Microbiol.* 2:16216. doi: 10.1038/nmicrobiol.2016.216

Cuenco, K. T., Halloran, M. E., Louis-Charles, J., and Lammie, P. J. (2004). A family study of lymphedema of the leg in a lymphatic filariasis-endemic area. *Am. J. Trop. Med. Hyg.* 70, 180–184.

Cully, D. F., Vassilatis, D. K., Liu, K. K., Paress, P. S., Van der Ploeg, L. H. T., Schaeffer, J. M., et al. (1994). Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* 371, 707–711.

Curtis, J., Luo, Y., Zenner, H. L., Cuchet-Lourenço, D., Wu, C., Lo, K., et al. (2015). Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat. Genet.* 47, 523–527. doi: 10.1038/ng.3248

da Silva, P. E. A., Palomino, J. C., Almeida Da Silva, P. E. A., and Palomino, J. C. (2011). Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J. Antimicrob. Chemother.* 66, 1417–1430. doi: 10.1093/jac/dkr173

Daly, A. K., King, B. P., and Leathart, J. B. S. (2006). Genotyping for cytochrome P450 polymorphisms. *Methods Mol. Biol.* 320, 193–207.

Damhoff, H. N., Kuhn, R. J., and Stadler, L. P. (2014). Severe malaria complicated by G6PD deficiency in a pediatric tanzanian immigrant. *J. Pediatr. Pharmacol. Ther.* 19, 325–334. doi: 10.5863/1551-6776-19.4.325

Dang, T. N., Naka, I., Sa-Ngasang, A., Anantapreecha, S., Chanama, S., Wichukchinda, N., et al. (2014). A replication study confirms the association of GWAS-identified SNPs at MICB and PLCE1in Thai patients with dengue shock syndrome. *BMC Med. Genet.* 15:58. doi: 10.1186/1471-2350-15-58

Dapat, C., and Oshitani, H. (2016). Novel insights into human respiratory syncytial virus-host factor interactions through integrated proteomics and transcriptomics analysis. *Expert Rev. Anti Infect. Ther.* 14, 285–297. doi: 10.1586/14787210.2016.1141676

Davey, G., Tekola, F., and Newport, M. J. (2007). Podoconiosis: non-infectious geochemical elephantiasis. *Trans. R. Soc. Trop. Med. Hyg.* 101, 1175–1180. doi: 10.1016/j.trstmh.2007.08.013

Davies, D. H., Duffy, P., Bodmer, J.-L., Felgner, P. L., and Doolan, D. L. (2015). Large screen approaches to identify novel malaria vaccine candidates. *Vaccine* 33, 7496–7505. doi: 10.1016/j.vaccine.2015.09.059

de Borba, L., Villordo, S. M., Iglesias, N. G., Filomatori, C. V., Gebhard, L. G., and Gamarnik, A. V. (2015). Overlapping local and long-range RNA-RNA interactions modulate dengue virus genome cyclization and replication. *J. Virol.* 89, 3430–3437. doi: 10.1128/JVI.02677-14

de Carvalho, C. X., Cardoso, C. C., de Souza Kehdy, F., Pacheco, A. G., and Moraes, M. O. (2017). Host genetics and dengue fever. *Infect. Genet. Evol.* 56, 99–110. doi: 10.1016/j.meegid.2017.11.009

De Groot, A. S., McMurry, J., and Moise, L. (2008). Prediction of immunogenicity: in silico paradigms, ex vivo and in vivo correlates. *Curr. Opin. Pharmacol.* 8, 620–626. doi: 10.1016/j.coph.2008.08.002

Desjardins, C. A., Cerqueira, G. C., Goldberg, J. M., Dunning Hotopp, J. C., Haas, B. J., Zucker, J., et al. (2013). Genomics of Loa loa, a Wolbachia-free filarial parasite of humans. *Nat. Genet.* 45, 495–500. doi: 10.1038/ng.2585

Doolan, D. L., Apte, S. H., and Proietti, C. (2014). Genome-based vaccine design: the promise for malaria and other infectious diseases. *Int. J. Parasitol.* 44, 901–913. doi: 10.1016/j.ijpara.2014.07.010

Doyle, S. R., Bourguinat, C., Nana-Djeunga, H. C., Kengne-Ouafo, J. A., Pion, S. D. S., Bopda, J., et al. (2017). Genome-wide analysis of ivermectin response by *Onchocerca volvulus* reveals that genetic drift and soft selective sweeps contribute to loss of drug sensitivity. *PLoS Negl. Trop. Dis.* 11:e0005816. doi: 10.1371/journal.pntd.0005816

Duffy, C. W., Assefa, S. A., Abugri, J., Amoako, N., Owusu-Agyei, S., Anyorigiya, T., et al. (2015). Comparison of genomic signatures of selection on *Plasmodium falciparum* between different regions of a country with high malaria endemicity. *BMC Genomics* 16:527. doi: 10.1186/s12864-015-1746-3

Duffy, C. W., Ba, H., Assefa, S., Ahouidi, A. D., Deh, Y. B., Tandia, A., et al. (2017). Population genetic structure and adaptation of malaria parasites on the edge of endemic distribution. *Mol. Ecol.* 26, 2880–2894. doi: 10.1111/mec.14066

Faria, N. R., da Costa, A. C., Lourenço, J., Loureiro, P., Lopes, M. E., Ribeiro, R., et al. (2017). Genomic and epidemiological characterisation of a dengue virus outbreak among blood donors in Brazil. *Sci. Rep.* 7:15216. doi: 10.1038/s41598-017-15152-8

Fatumo, S. A., Adoga, M. P., Ojo, O. O., Oluwagbemi, O., Adeoye, T., Ewejobi, I., et al. (2014). Computational biology and bioinformatics in nigeria. *PLoS Comput. Biol.* 10:e1003516. doi: 10.1371/journal.pcbi.1003516

Feuerriegel, S., Schleusener, V., Beckert, P., Kohl, T. A., Miotto, P., Cirillo, D. M., et al. (2015). PhyResSE: a web tool delineating *Mycobacterium tuberculosis* antibiotic resistance and lineage from whole-genome sequencing data. *J. Clin. Microbiol.* 53, 1908–1914. doi: 10.1128/JCM.00025-15

Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., et al. (2017). The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. *Front. Immunol.* 8:278. doi: 10.3389/fimmu.2017.00278

Ford, C. B., Shah, R. R., Maeda, M. K., Gagneux, S., Murray, M. B., Cohen, T., et al. (2013). *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* 45, 784–790. doi: 10.1038/ng.2656

Friebe, P., Peña, J., Pohl, M. O. F., and Harris, E. (2012). Composition of the sequence downstream of the dengue virus 5′ cyclization sequence (dCS) affects viral RNA replication. *Virology* 422, 346–356. doi: 10.1016/j.virol.2011.10.025

Gebhard, L. G., Filomatori, C. V., and Gamarnik, A. V. (2011). Functional RNA elements in the dengue virus genome. *Viruses* 3, 1739–1756. doi: 10.3390/v3091739

Gengiah, T., Botha, J., Soowamber, D., and Naidoo, K., Abdool Karim, S. S. (2014). Low rifampicin concentrations in tuberculosis patients with HIV infection. *J. Infect. Dev. Ctries.* 8, 987–993. doi: 10.3855/jidc.4696

Gould, E., and Solomon, T. (2008). Pathogenic flaviviruses. *Lancet* 371, 500–509.

Grubaugh, D., Flechtner, J. B., and Higgins, D. E. (2013). Proteins as T cell antigens: methods for high-throughput identification. *Vaccine* 31, 3805–3810. doi: 10.1016/j.vaccine.2013.06.046

Guerra-Assunção, J., Crampin, A., Houben, R., Mzembe, T., Mallard, K., Coll, F., et al. (2015). Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife* 4:e05166. doi: 10.7554/eLife.05166

Guiliano, D. B., Hong, X., McKerrow, J. H., Blaxter, M. L., Oksov, Y., Liu, J., et al. (2004). A gene family of cathepsin L-like proteases of filarial nematodes are associated with larval molting and cuticle and eggshell remodeling. *Mol. Biochem. Parasitol.* 136, 227–242.

Gurwitz, K. T., Aron, S., Panji, S., Maslamoney, S., Fernandes, P. L., Judge, D. P., et al. (2017). Designing a course model for distance-based online bioinformatics training in Africa: the H3ABioNet experience. *PLoS Comput. Biol.* 13:e1005715. doi: 10.1371/journal.pcbi.1005715

Guy, B., Lang, J., Saville, M., and Jackson, N. (2016). Vaccination against dengue: challenges and current developments. *Annu. Rev. Med*. 67, 387–404. doi: 10.1146/annurev-med-091014-090848

H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V. M., Adebamowo, C., et al. (2014). Research capacity. Enabling the genomic revolution in Africa. *Science* 344, 1346–1348.

Hamburg, M. A., and Collins, F. S. (2010). The path to personalized medicine. *N. Engl. J. Med*. 363, 301–304.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 32, D258–D261.

He, Y. (2012). Omics-based systems vaccinology for vaccine target identification. *Drug Dev. Res*. 73, 559–568.

Holbrook, M. R. (2017). Historical perspectives on flavivirus research. *Viruses* 9, 1–19.

Holt, K. E., McAdam, P., Thai, P. V. K., Thuong, N. T. T., Ha, D. T. M., Lan, N. N., et al. (2018). Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat. Genet*. 50, 849–856. doi: 10.1038/s41588-018-0117-9

Hoo, R., Zhu, L., Amaladoss, A., Mok, S., Natalang, O., Lapp, S. A., et al. (2016). Integrated analysis of the *Plasmodium* species transcriptome. *EBioMedicine* 7, 255–266. doi: 10.1016/j.ebiom.2016.04.011

Jacobsen, M., Repsilber, D., Gutschmidt, A., Neher, A., Feldmann, K., Mollenkopf, H. J., et al. (2007). Candidate biomarkers for discrimination between infection and disease caused by *Mycobacterium tuberculosis*. *J. Mol. Med.* 85, 613–621.

Jamieson, F. B., Teatero, S., Guthrie, J. L., Neemuchwala, A., Fittipaldi, N., and Mehaffy, C. (2014). Whole-genome sequencing of the *Mycobacterium tuberculosis* manila Sublineage results in less clustering and better resolution than Mycobacterial interspersed repetitive-unit-variable-number tandem-repeat (MIRU-VNTR) typing and Spoligotyping. *J. Clin. Microbiol*. 52, 3795–3798. doi: 10.1128/JCM.01726-14

Jespersen, M. C., Peters, B., Nielsen, M., and Marcatili, P. (2017). BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.* 45, W24–W29. doi: 10.1093/nar/gkx346

Junpee, A., Tencomnao, T., Sanprasert, V., and Nuchprayoon, S. (2010). Association between Toll-like receptor 2 (TLR2) polymorphisms and asymptomatic bancroftian filariasis. *Parasitol. Res.* 107, 807–816. doi: 10.1007/s00436-010-1932-9

Kaforou, M., Wright, V. J., Oni, T., French, N., Anderson, S. T., Bangani, N., et al. (2013). Detection of tuberculosis in HIV-infected and -uninfected African adults using whole blood RNA expression signatures: a case-control study. *PLoS Med.* 10:e1001538. doi: 10.1371/journal.pmed.1001538

Kaltdorf, M., Srivastava, M., Gupta, S. K., Liang, C., Binder, J., Dietl, A.-M., et al. (2016). Systematic identification of anti-fungal drug targets by a metabolic network approach. *Front. Mol. Biosci.* 3:22. doi: 10.3389/fmolb.2016.00022

Karikari, T. K. (2015). Bioinformatics in africa: the rise of ghana? *PLoS Comput. Biol.* 11:e1004308. doi: 10.1371/journal.pcbi.1004308

Karikari, T. K., Quansah, E., and Mohamed, W. M. Y. (2015). Developing expertise in bioinformatics for biomedical research in Africa. *Appl. Transl. Genomics* 6, 31–34.

Khor, C. C., Chau, T. N., Pang, J., Davila, S., Long, H. T., Ong, R. T., et al. (2011). Genome-wide association study identifies susceptibility loci for dengue shock syndrome at MICB and PLCE1. *Nat. Genet*. 43, 1139–1141. doi: 10.1038/ng.960

Köser, C. U., Bryant, J. M., Becq, J., Török, M. E., Ellington, M. J., Marti-Renom, M. A., et al. (2013). Whole-genome sequencing for rapid susceptibility testing of *M. tuberculosis*. *N. Engl. J. Med*. 369, 290–292.

Kumar, S., Chaudhary, K., Foster, J. M., Novelli, J. F., Zhang, Y., Wang, S., et al. (2007). Mining predicted essential genes of *Brugia malayi* for nematode drug targets. *PLoS One* 2:e1189. doi: 10.1371/journal.pone.0001189

Kwiatkowski, D. (2015). Malaria genomics: tracking a diverse and evolving parasite population. *Int. Health* 7, 82–84. doi: 10.1093/inthealth/ihv007

Lee, H. J., Walther, M., Georgiadou, A., Nwakanma, D., Stewart, L. B., Levin, M., et al. (2017). Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. *Sci. Transl. Med*. 10:eaar3619. doi: 10.1126/scitranslmed.aar3619

Leisching, G., Pietersen, R.-D., Mpongoshe, V., van Heerden, C., van Helden, P., Wiid, I., et al. (2016). The host response to a clinical MDR mycobacterial strain cultured in a detergent-free environment: a global transcriptomics approach. *PLoS One* 11:e0153079. doi: 10.1371/journal.pone.0153079

Lesho, E., Forestiero, F. J., Hirata, M. H., Hirata, R. D., Cecon, L., Melo, F. F., et al. (2011). Transcriptional responses of host peripheral blood cells to tuberculosis infection. *Tuberculosis* 91, 390–399. doi: 10.1016/j.tube.2011.07.002

Li, B.-W., Rush, A. C., Mitreva, M., Yin, Y., Spiro, D., Ghedin, E., et al. (2009). Transcriptomes and pathways associated with infectivity, survival and immunogenicity in *Brugia malayi* L3. *BMC Genomics* 10:267. doi: 10.1186/1471-2164-10-267

Li, D., and Aaskov, J. (2014). Sub-genomic RNA of defective interfering (D.I.) dengue viral particles is replicated in the same manner as full length genomes. *Virology* 468–470, 248–255. doi: 10.1016/j.virol.2014.08.013

Li, D., Lott, W. B., Lowry, K., Jones, A., Thu, H. M., and Aaskov, J. (2011). Defective interfering viral particles in acute dengue infections. *PLoS One* 6:e19447. doi: 10.1371/journal.pone.0019447

Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T., and DeRisi, J. L. (2006). Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Res.* 34, 1166–1173.

Low, J. G. H., Ooi, E. E., and Vasudevan, S. G. (2017). Current status of dengue therapeutics research and development. *J. Infect. Dis.* 215(Suppl. 2), S96–S102. doi: 10.1093/infdis/jiw423

Ludin, P., Woodcroft, B., Ralph, S. A., and Mäser, P. (2012). In silico prediction of antimalarial drug target candidates. *Int. J. Parasitol.* 2(Suppl. C), 191–199. doi: 10.1016/j.ijpddr.2012.07.002

Lustigman, S., Zhang, J., Liu, J., Oksov, Y., and Hashmi, S. (2004). RNA interference targeting cathepsin L and Z-like cysteine proteases of *Onchocerca volvulus* confirmed their essential function during L3 molting. *Mol. Biochem. Parasitol.* 138, 165–170.

Luzatto, L., Mehta, A., and Vulliamy, T. (2001). "Glucose 6-phosphate dehydrogenase deficiency," in *The Metabolic and Molecular Bases of Inherited Disease,* 8th Edn, eds C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (New York, NY: McGraw-Hill), 4517–4553.

Luzzatto, L., and Seneca, E. (2014). G6PD deficiency: a classic example of pharmacogenetics with on-going clinical implications. *Br. J. Haematol.* 164, 469–480. doi: 10.1111/bjh.12665

Maertzdorf, J., McEwen, G., Weiner, J. III, Tian, S., Lader, E., Schriek, U., et al. (2015). Concise gene signature for point-of-care classification of tuberculosis. *EMBO Mol. Med*. 8, 86–95. doi: 10.15252/emmm.201505790

Maertzdorf, J., Ota, M., Repsilber, D., Mollenkopf, H. J., Weiner, J., Hill, P. C., et al. (2011). Functional correlations of pathogenesis-driven gene expression signatures in tuberculosis. *PLoS One* 6:e26938. doi: 10.1371/journal.pone.0026938

Malaria Genomic Epidemiology Network (2015). A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* 526, 253–257. doi: 10.1038/nature15390

MalariaGEN Plasmodium falciparum Community Project (2016). Genomic epidemiology of artemisinin resistant malaria. *eLife* 5:e08714. doi: 10.7554/eLife.08714

Manakkadan, A., Joseph, I., Prasanna, R. R., Kunju, R. I., Kailas, L., and Sreekumar, E. (2013). Lineage shift in Indian strains of Dengue virus serotype-3 (Genotype III), evidenced by detection of lineage IV strains in clinical cases from Kerala. *Virol. J.* 10:37. doi: 10.1186/1743-422X-10-37

Manske, M., Miotto, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., et al. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487, 375–379. doi: 10.1038/nature11174

Masignani, V., Rappuoli, R., and Pizza, M. (2002). Reverse vaccinology: a genome-based approach for vaccine development. *Expert Opin. Biol. Ther.* 2, 895–905.

Mehaffy, C., Guthrie, J. L., Alexander, D. C., Stuart, R., Rea, E., and Jamieson, F. B. (2014). Marked microevolution of a unique *Mycobacterium tuberculosis* strain in 17 years of ongoing transmission in a high risk population. *PLoS One* 9:e112928. doi: 10.1371/journal.pone.0112928

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., et al. (2015). Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat. Genet*. 47, 242–249. doi: 10.1038/ng.3195

Miles, A., Iqbal, Z., Vauterin, P., Pearson, R., Campino, S., Theron, M., et al. (2016). Indels, structural variation, and recombination drive genomic diversity in *Plasmodium falciparum*. *Genome Res.* 26, 1288–1299. doi: 10.1101/gr.203711.115

Mistry, R., Cliff, J. M., Clayton, C. L., Beyers, N., Mohamed, Y. S., Wilson, P. A., et al. (2007). Gene-expression patterns in whole blood identify subjects at risk for recurrent tuberculosis. *J. Infect. Dis.* 195, 357–365.

Mobegi, V. A., Duffy, C. W., Amambua-Ngwa, A., Loua, K. M., Laman, E., Nwakanma, D. C., et al. (2014). genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in west african populations of differing infection endemicity. *Mol. Biol. Evol.* 31, 1490–1499. doi: 10.1093/molbev/msu106

Molehin, A. J., Gobert, G. N., and McManus, D. P. (2012). Serine protease inhibitors of parasitic helminths. *Parasitology* 139, 681–695. doi: 10.1017/S0031182011002435

Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N. L., Khasnatinov, M., et al. (2015). New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences. *PLoS One* 10:e0117849. doi: 10.1371/journal.pone.0117849

Mulder, N. J., Adebiyi, E., Adebiyi, M., Adeyemi, S., Ahmed, A., Ahmed, R., et al. (2017). Development of bioinformatics infrastructure for genomics research. *Glob. Heart* 12, 91–98. doi: 10.1016/j.gheart.2017.01.005

Mulder, N. J., Christoffels, A., de Oliveira, T., Gamieldien, J., Hazelhurst, S., Joubert, F., et al. (2016). The development of computational biology in south africa: successes achieved and lessons learnt. *PLoS Comput. Biol.* 12:e1004395. doi: 10.1371/journal.pcbi.1004395

Nag, S., Dalgaard, M. D., Kofoed, P.-E., Ursing, J., Crespo, M., Andersen, L. O., et al. (2017). High throughput resistance profiling of *Plasmodium falciparum* infections based on custom dual indexing and Illumina next generation sequencing-technology. *Sci. Rep.* 7:2398. doi: 10.1038/s41598-017-02724-x

Nana-Djeunga, H., Bourguinat, C., Pion, S. D., Kamgno, J., Gardon, J., and Njiokou, F. (2012). Single nucleotide polymorphisms in beta-tubulin selected in *Onchocerca volvulus* following repeated ivermectin treatment: possible indication of resistance selection. *Mol. Biochem. Parasitol.* 185, 10–18. doi: 10.1016/j.molbiopara.2012.05.005

Nana-Djeunga, H. C., Bourguinat, C., Pion, S. D., Bopda, J., Kengne-Ouafo, J. A., and Njiokou, F. (2014). Reproductive status of *Onchocerca volvulus* after ivermectin treatment in an ivermectin-naïve and a frequently treated population from cameroon. *PLoS Negl. Trop. Dis.* 8:e2824. doi: 10.1371/journal.pntd.0002824

Navale, V., and Bourne, P. E. (2018). Cloud computing applications for biomedical science: a perspective. *PLoS Comput. Biol.* 14:e1006144. doi: 10.1371/journal.pcbi.1006144

O'Neill, M., Ballesteros, C., Tritten, L., Burkman, E., Zaky, W. I., Xia, J., et al. (2016a). Profiling the macrofilaricidal effects of flubendazole on adult female *Brugia malayi* using RNAseq. *Int. J. Parasitol. Drugs Drug Resist.* 6, 288–296. doi: 10.1016/j.ijpddr.2016.09.005

O'Neill, M., Mansour, A., DiCosty, U., Geary, J., Dzimianski, M., McCall, S. D., et al. (2016b). An in vitro/in vivo model to analyze the effects of flubendazole exposure on adult female *Brugia malayi*. *PLoS Negl. Trop. Dis.* 10:e0004698. doi: 10.1371/journal.pntd.0004698

O'Neill, M., Geary, J. F., Agnew, D. W., Mackenzie, C. D., and Geary, T. G. (2015). In vitro flubendazole-induced damage to vital tissues in adult females of the filarial nematode *Brugia malayi*. *Int. J. Parasitol. Drugs Drug Resist.* 5, 135–140. doi: 10.1016/j.ijpddr.2015.06.002

Osei-Atweneboana, M. Y., Boakye, D. A., Awadzi, K., Gyapong, J. O., and Prichard, R. K. (2012). Genotypic analysis of beta-tubulin in *Onchocerca volvulus* from communities and individuals showing poor parasitological response to ivermectin treatment. *Int. J. Parasitol. Drugs Drug Resist.* 2, 20–28. doi: 10.1016/j.ijpddr.2012.01.005

Ottenhoff, T. H. M., Dass, R. H., Yang, N., Zhang, M. M., Wong, H. E. E., Sahiratmadja, E., et al. (2012). Genome-wide expression profiling identifies type 1 interferon response pathways in active tuberculosis. *PLoS One* 7:e45839. doi: 10.1371/journal.pone.0045839

Otto, T. D., Wilinski, D., Assefa, S., Keane, T. M., Sarry, L. R., Böhme, U., et al. (2010). New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Mol. Microbiol.* 76, 12–24. doi: 10.1111/j.1365-2958.2009.07026.x

Panda, A. K., Sahoo, P. K., Kerketta, A. S., Kar, S. K., Ravindran, B., and Satapathy, A. K. (2011). Human lymphatic filariasis: genetic polymorphism of endothelin-1 and tumor necrosis factor receptor II correlates with development of chronic disease. *J. Infect. Dis.* 204, 315–322. doi: 10.1093/infdis/jir258

Parameswaran, P., Charlebois, P., Tellez, Y., Nunez, A., Ryan, E. M., Malboeuf, C. M., et al. (2012). Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. *J. Virol.* 86, 8546–8558. doi: 10.1128/JVI.00736-12

Peachey, L. E., Pinchbeck, G. L., Matthews, J. B., Burden, F. A., Lespine, A., von Samson-Himmelstjerna, G., et al. (2017). P-glycoproteins play a role in ivermectin resistance in cyathostomins. *Int. J. Parasitol. Drugs Drug Resist.* 7, 388–398. doi: 10.1016/j.ijpddr.2017.10.006

Price, E. W. (1976). The association of endemic elephantiasis of the lower legs in East Africa with soil derived from volcanic rocks. *Trans. R. Soc. Trop. Med. Hyg.* 70, 288–295.

Proietti, C., and Doolan, D. L. (2015). The case for a rational genome-based vaccine against malaria. *Front. Microbiol.* 5:741. doi: 10.3389/fmicb.2014.00741

Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W., et al. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15066–15071. doi: 10.1073/pnas.1206598109

Qi, H., Zhang, Y.-B., Sun, L., Chen, C., Xu, B., Xu, F., et al. (2017). Discovery of susceptibility loci associated with tuberculosis in Han Chinese. *Hum. Mol. Genet.* 26, 4752–4763. doi: 10.1093/hmg/ddx365

Rabbani, B., Nakaoka, H., Akhondzadeh, S., Tekin, M., and Mahdieh, N. (2016). Next generation sequencing: implications in personalized medicine and pharmacogenomics. *Mol. Biosyst.* 12, 1818–1830.

Ramesh, K., Hemanth Kumar, A. K., Kannan, T., Vijayalakshmi, R., Sudha, V., Manohar Nesakumar, S., et al. (2016). SLCO1B1 gene polymorphisms do not influence plasma rifampicin concentrations in a South Indian population. *Int. J. Tuberc. Lung Dis.* 20, 1231–1235. doi: 10.5588/ijtld.15.1007

Rasmussen, D. A., Boni, M. F., and Koelle, K. (2014). Reconciling phylodynamics with epidemiology: the case of dengue virus in southern vietnam. *Mol. Biol. Evol.* 31, 258–271. doi: 10.1093/molbev/mst203

Reche, P., Glutting, J.-P., Zhang, H., and Reinherz, E. (2004). Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* 56, 405–419.

Reid, A. J., Talman, A. M., Bennett, H. M., Gomes, A. R., Sanders, M. J., Illingworth, C. J. R., et al. (2018). Single-cell RNA-seq reveals hidden transcriptional variation in malaria parasites. *Elife* 7:e33105. doi: 10.7554/eLife.33105

Robinson, T., Campino, S. G., Auburn, S., Assefa, S. A., Polley, S. D., Manske, M., et al. (2011). Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. *PLoS One* 6:e23204. doi: 10.1371/journal.pone.0023204

Saha, S., and Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 65, 40–48.

Saha, S., and Raghava, G. P. S. (2004). "BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties," in *Proceedings of the 3rd International Conference Artificial Immune Systems, ICARIS 2004*, eds G. Nicosia, V. Cutello, P. J. Bentley, and J. Timis (Berlin: Springer).

Siau, A., Silvie, O., Franetich, J. F., Yalaoui, S., Marinach, C., Hannoun, L., et al. (2008). Temperature shift and host cell contact up-regulate sporozoite expression of *Plasmodium falciparum* genes involved in hepatocyte infection. *PLoS Pathog.* 4:e1000121. doi: 10.1371/journal.ppat.1000121

Sloan, D. J., McCallum, A. D., Schipani, A., Egan, D., Mwandumba, H. C., Ward, S. A., et al. (2017). Genetic determinants of the pharmacokinetic variability of rifampin in malawian adults with pulmonary tuberculosis. *Antimicrob. Agents Chemother.* 61:e00210-17. doi: 10.1128/AAC.00210-17

Small, S. T., Reimer, L. J., Tisch, D. J., King, C. L., Christensen, B. M., Siba, P. M., et al. (2016). Population genomics of the filarial nematode parasite *Wuchereria*

bancrofti from mosquitoes. *Mol. Ecol.* 25, 1465–1477. doi: 10.1111/mec.13574

Sobkowiak, B., Glynn, J. R., Houben, R. M. G. J., Mallard, K., Phelan, J. E., Guerra-Assunção, J. A., et al. (2018). Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genomics* 19:613. doi: 10.1186/s12864-018-4988-z

Sobota, R. S., Stein, C. M., Kodaman, N., Scheinfeldt, L. B., Maro, I., Wieland-Alter, W., et al. (2016). A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. *Am. J. Hum. Genet*. 98, 514–524. doi: 10.1016/j.ajhg.2016.01.015

Sorber, L., Zwaenepoel, K., Deschoolmeester, V., Van Schil, P. E. Y., Van Meerbeeck, J., Lardon, F., et al. (2017). Circulating cell-free nucleic acids and platelets as a liquid biopsy in the provision of personalized therapy for lung cancer patients. *Lung Cancer* 107, 100–107. doi: 10.1016/j.lungcan.2016.04.026

Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., and Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: implications on vaccine development. *J. Biomed. Inform.* 53, 405–414. doi: 10.1016/j.jbi.2014.11.003

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Straimer, J., Gnädig, N. F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A. P., et al. (2015). Drug resistance. K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science* 347, 428–431. doi: 10.1126/science.1260867

Stucki, D., Ballif, M., Bodmer, T., Coscolla, M., Maurer, A. M., Droz, S., et al. (2015). Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. *J. Infect. Dis*. 211, 1306–1316. doi: 10.1093/infdis/jiu601

Sun, G., Luo, T., Yang, C., Dong, X., Li, J., Zhu, Y., et al. (2012). Dynamic population changes in *mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J. Infect. Dis*. 206, 1724–1733. doi: 10.1093/infdis/jis601

Sveinbjornsson, G., Gudbjartsson, D. F., Halldorsson, B. V., Kristinsson, K. G., Gottfredsson, M., Barrett, J. C., et al. (2016). HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. *Nat. Genet*. 48, 318–322. doi: 10.1038/ng.3498

Tarun, A. S., Peng, X., Dumpit, R. F., Ogata, Y., Silva-Rivera, H., Camargo, N., et al. (2008). A combined transcriptome and proteome survey of malaria parasite liver stages. *Proc. Natl. Acad. Sci. U.S.A.* 105, 305–310. doi: 10.1073/pnas.0710780104

Taylor, M. J. (2003). Wolbachia in the inflammatory pathogenesis of human filariasis. *Ann. N. Y. Acad. Sci.* 990, 444–449.

Taylor, M. J., Hoerauf, A., and Bockarie, M. (2010). Lymphatic filariasis and onchocerciasis. *Lancet* 376, 1175–1185.

Tekola Ayele, F., Adeyemo, A., Finan, C., Hailu, E., Sinnott, P., Burlinson, N. D., et al. (2012). HLA class II locus and susceptibility to podoconiosis. *N. Engl. J. Med*. 366, 1200–1208. doi: 10.1056/NEJMoa1108448

Thai, K. T. D., Henn, M. R., Zody, M. C., Tricou, V., Nguyet, N. M., Charlebois, P., et al. (2012). High-resolution analysis of intrahost genetic diversity in dengue virus serotype 1 infection identifies mixed infections. *J. Virol.* 86, 835–843. doi: 10.1128/JVI.05985-11

Thompson, E. G., Du, Y., Malherbe, S. T., Shankar, S., Braun, J., Valvo, J., et al. (2017). Host blood RNA signatures predict the outcome of tuberculosis treatment. *Tuberculosis* 107, 48–58. doi: 10.1016/j.tube.2017.08.004

Thye, T., Owusu-Dabo, E., Vannberg, F. O., van Crevel, R., Curtis, J., Sahiratmadja, E., et al. (2012). Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet*. 44, 257–259. doi: 10.1038/ng.1080

Thye, T., Vannberg, F. O., Wong, S. H., Owusu-Dabo, E., Osei, I., Gyapong, J., et al. (2010). Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet*. 42, 739–741. doi: 10.1038/ng.639

Tientcheu, L. D., Maertzdorf, J., Weiner, J., Adetifa, I. M., Mollenkopf, H.-J., Sutherland, J. S., et al. (2015). Differential transcriptomic and metabolic profiles of *M. africanum*- and *M. tuberculosis*-infected patients after, but not before, drug treatment. *Genes Immun.* 16, 347–355. doi: 10.1038/gene.2015.21

Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., et al. (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* 489, 443–446. doi: 10.1038/nature11334

Turab Naqvi, A. A., Rahman, S., Rubi., Zeya, F., Kumar, K., Choudhary, H., et al. (2017). Genome analysis of Chlamydia trachomatis for functional characterization of hypothetical proteins to discover novel drug targets. *Int. J. Biol. Macromol.* 96, 234–240. doi: 10.1016/j.ijbiomac.2016.12.045

van Beek, J., Haanperä, M., Smit, P. W., Mentula, S., and Soini, H. (2018). Evaluation of whole genome sequencing and software tools for drug susceptibility testing of *Mycobacterium tuberculosis*. *Clin. Microbiol. Infecti*. doi: 10.1016/j.cmi.2018.03.041 [Epub ahead of print].

Verhagen, L. M., Zomer, A., Maes, M., Villalba, J. A., del Nogal, B., Eleveld, M., et al. (2013). A predictive signature gene set for discriminating active from latent tuberculosis in Warao Amerindian children. *BMC Genomics* 14:74. doi: 10.1186/1471-2164-14-74

Villordo, S. M., Filomatori, C. V., Sánchez-Vargas, I., Blair, C. D., and Gamarnik, A. V. (2015). Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog.*11:e1004604. doi: 10.1371/journal.ppat.1004604

Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Del Ojo Elias, C., Bradley, P., et al. (2015). Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect. Dis*. 15, 1193–1202.

Waman, V. P., Kolekar, P., Ramtirthkar, M. R., Kale, M. M., and Kulkarni-Kale, U. (2016). Analysis of genotype diversity and evolution of Dengue virus serotype 2 using complete genomes. *PeerJ* 4:e2326. doi: 10.7717/peerj.2326

Weiner, M., Peloquin, C., Burman, W., Luo, C.-C., Engle, M., Prihoda, T. J., et al. (2010). Effects of tuberculosis, race, and human gene SLCO1B1 polymorphisms on rifampin concentrations. *Antimicrob. Agents Chemother.* 54, 4192–4200. doi: 10.1128/AAC.00353-10

Whitehorn, J., Chau, T. N. B., Nguyet, N. M., Kien, D. T. H., Quyen, N. T. H., Trung, D. T., et al. (2013). Genetic variants of MICB and PLCE1 and associations with non-severe dengue. *PLoS One* 8:e59067. doi: 10.1371/journal.pone.0059067

WHO (2017). *WHO Dengue and Severe Dengue*. Geneva: World Health Organization.

WHO (2018a). *Global Tuberculosis Report 2018*. Geneva: World Health Organization.

WHO (2018b). *World Malaria Day 2018: Ready to Beat Malaria*. Geneva: World Health Organization, 1–3.

WHO/Department of Control of Neglected Tropical Diseases (2016). Global programme to eliminate lymphatic filariasis: progress report, 2015. *Wkly. Epidemiol. Rec.* 39, 441–460.

World Health Organization Western Pacific Region (2018). Dengue Situation Update Number 500: Update on the Dengue situation in the Western Pacific Region (Northern Hemisphere)., (500), pp. 1–5*.

Xia, X. (2017). Bioinformatics and drug discovery. *Curr. Top. Med. Chem.* 17, 1709–1726.

Yamagishi, J., Natori, A., Tolba, M. E. M., Mongan, A. E., Sugimoto, C., Katayama, T., et al. (2014). Interactive transcriptome analysis of malaria patients and infecting *Plasmodium falciparum*. *Genome Res.* 24, 1433–1444. doi: 10.1101/gr.158980.113

Yan, S.-K., Liu, R.-H., Jin, H.-Z., Liu, X.-R., Ye, J., Shan, L., et al. (2015). 'Omics' in pharmaceutical research: overview, applications, challenges, and future perspectives. *Chin. J. Nat. Med.* 13, 3–21.

Young, J. A., Fivelman, Q. L., Blair, P. L., de la Vega, P., Le Roch, K. G., Zhou, Y., et al. (2005). The *Plasmodium falciparum* sexual development transcriptome: a microarray analysis using ontology-based pattern identification. *Mol. Biochem. Parasitol.* 143, 67–79.

Yun, M.-R., Han, S. J., Yoo, W. G., Kwon, T., Lee, S., Lee, J. S., et al. (2016). Draft genome sequence of *Mycobacterium tuberculosis* KT-0204, isolated in South Korea. *Genome Announc.* 4:e01519-15. doi: 10.1128/genomeA.01519-15

Zak, D. E., Penn-Nicholson, A., Scriba, T. J., Thompson, E., Suliman, S., Amon, L. M., et al. (2016). A blood RNA signature for tuberculosis disease risk: a prospective cohort study. *Lancet* 387, 2312–2322. doi: 10.1016/S0140-6736(15)01316-1

Zhang, H., Li, D., Zhao, L., Fleming, J., Lin, N., Wang, T., et al. (2013). Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat. Genet.* 45, 1255–1260. doi: 10.1038/ng.2735

Zhu, L., Mok, S., Imwong, M., Jaidee, A., Russell, B., Nosten, F., et al. (2016). New insights into the *Plasmodium vivax* transcriptome using RNA-Seq. *Sci. Rep.* 6:20498. doi: 10.1038/srep20498

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Bah, Morang'a, Kengne-Ouafo, Amenga–Etego and Awandare. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.