

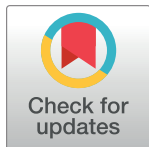
REVIEW

Turning the needle into the haystack: Culture-independent amplification of complex microbial genomes directly from their native environment

Olivia A. Pilling¹, Sesh A. Sundararaman², Dustin Brisson³, Daniel P. Beiting¹ ^{*}

1 Department of Pathobiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **2** Department of Pediatrics, Children's Hospital of Philadelphia, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **3** Department of Biology, School of Arts & Sciences, University of Pennsylvania, Pennsylvania, United States of America

* beiting@upenn.edu



Abstract

High-throughput sequencing (HTS) has revolutionized microbiology, but many microbes exist at low abundance in their natural environment and/or are difficult, if not impossible, to culture in the laboratory. This makes it challenging to use HTS to study the genomes of many important microbes and pathogens. In this review, we discuss the development and application of selective whole genome amplification (SWGA) to allow whole or partial genomes to be sequenced for low abundance microbes directly from complex biological samples. We highlight ways in which genomic data generated by SWGA have been used to elucidate the population dynamics of important human pathogens and monitor development of antimicrobial resistance and the emergence of potential outbreaks. We also describe the limitations of this method and propose some potential innovations that could be used to improve the quality of SWGA and lower the barriers to using this method across a wider range of infectious pathogens.

OPEN ACCESS

Citation: Pilling OA, Sundararaman SA, Brisson D, Beiting DP (2024) Turning the needle into the haystack: Culture-independent amplification of complex microbial genomes directly from their native environment. *PLoS Pathog* 20(9): e1012418. <https://doi.org/10.1371/journal.ppat.1012418>

Editor: Bjorn F. C. Kafsack, Joan and Sanford I Weill Medical College of Cornell University, UNITED STATES OF AMERICA

Published: September 12, 2024

Copyright: © 2024 Pilling et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded in part by a grant from the National Institute of Allergy and Infectious Diseases (5T32AI007532-24 to OAP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Pathogens and commensal microbes reside within extremely complex environments, including the mammalian gut, skin, urogenital tract, and upper airways, as well as environmental reservoirs, such as soil, water, and insect vectors. These niches require specialized adaptations, and our understanding of the nutrient requirements for microbes to survive in these habitats is extremely limited. As a result, many microbes are not readily amenable to in vitro culture in the laboratory and therefore are not experimentally tractable. One major development that opened the doors to culture-independent microbial genomics was the first use of “shotgun” metagenomic sequencing to reconstruct genomes and infer microbial function directly from high-throughput sequencing (HTS) data [1]. This work, along with computational methods developed to process the resulting complex data [2–6], represented a landmark breakthrough,

demonstrating that de novo assembly and functional annotation of whole bacterial genomes was possible for organisms that had never before been cultured in the laboratory. In the 20 years since this discovery, metagenomic sequencing and de novo assembly methods have yielded rich collections of metagenome-assembled genomes (MAGs) for bacteria, shedding light on novel microbial functions and evolution [7–11]. Unfortunately, several factors limit the more widespread use of metagenomics to study microbes that cause disease. Metagenomic data are often dominated by high levels of nonmicrobial “contaminating” sequences from the host or environment. For example, stool from healthy donors may contain as little as 10% host DNA, but in patients with gastrointestinal inflammation, this can increase to over 90% [12–14]. Similarly, saliva, nasal, buccal, and vaginal samples routinely comprise as much as 90% host DNA [14]. This issue, together with the relatively high cost associated with deep sequencing, often renders metagenomic methods inefficient for studying microbial genomes that are present at low abundance in complex samples. Recent innovations in computational and sequencing methods may ameliorate these issues, but even in the complete absence of contaminating host sequences, HTS of mixed communities often yields poor coverage of larger genomes, such as those from microbial eukaryotes including helminths, protozoa, and fungi. This makes assembly of whole or partial genomes of these microbes from metagenomic data problematic [15,16].

A hallmark of many infectious diseases is the chronic persistence of pathogens at low levels in host tissues, thereby avoiding immune-mediated clearance and maximizing the chance of transmission. Many parasites like *Trypanosoma cruzi* and *Toxoplasma gondii* establish persistent, often lifelong, infections that remain at low densities for decades. Similarly, *Mycobacterium tuberculosis* and *Mycobacterium bovis*, the cause of human and bovine tuberculosis, respectively, are extremely slow growing and develop latent infections that are difficult to directly detect [17,18]. Many chronic bacterial pathogens like *Borrelia burgdorferi*, the cause of Lyme disease, also persist at extremely low levels in human or animal hosts [19]. These examples underscore the challenges associated with studying the genomes of important human and animal pathogens.

In this review, we describe the development and use of selective whole genome amplification (SWGA) to extract whole or partial microbial genomes directly from their native animal and human hosts, even when the organism is present at extremely low abundance. SWGA uses short oligonucleotide primers designed to preferentially bind a target microbial genome over one or more contaminating “background” host genomes. These primers are used as a pool in a multiplexed isothermal reaction together with a high-fidelity, highly processive *Phi29* polymerase to preferentially amplify large segments (up to 70 to 100 kb) of a target genome [20,21]. The simplicity and low cost of SWGA makes it possible to implement in low- and middle-income countries (LMICs) that may have limited laboratory resources. This review covers nearly a decade of SWGA research and development spanning 3 bacterial species, eukaryotic microbes including the protozoa *Plasmodium* spp. and *Leishmania braziliensis*, and the helminth parasite, *Wuchereria bancrofti*. The broad use of SWGA across a diverse range of organisms highlights the impact of this method and the potential for approaches like SWGA to serve as a general toolkit for population genomics of pathogens that may otherwise be difficult to study in vitro.

The case for SWGA

To appreciate the need for, and utility of, SWGA, it is useful to first consider modern viral genomic surveillance (Fig 1). The small size of viral genomes, together with a high degree of sequence divergence between viruses and the human or animal host(s) they infect, means that

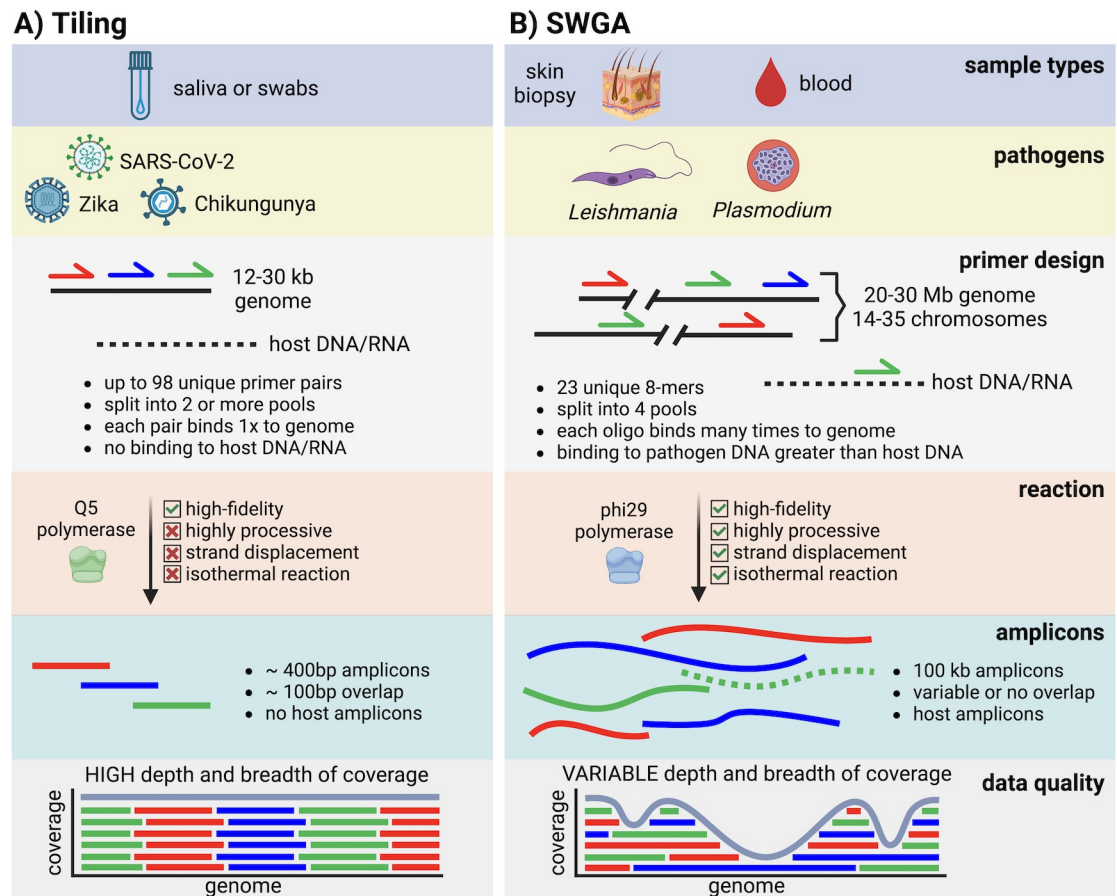


Fig 1. SWGA unlocks population genomics for nonviral pathogens. Schematic comparison of (A) tiling-based methods for viral genomic surveillance studies and (B) SWGA for bacterial and eukaryotic pathogens. Comparisons are made for sample type, pathogens targeted, primer design, PCR reactions, amplicons produced, and the quality of data produced by these 2 approaches. Created with [Biorender.com](https://www.biorender.com).

<https://doi.org/10.1371/journal.ppat.1012418.g001>

conventional PCR primers can readily be designed [22] that will produce PCR amplicons that tile (partially overlap) across the full length of the viral genome, with little or no off-target binding to host DNA or RNA present in the original sample (Fig 1A). Tiling primer sets are then combined into 2 or more pools (likely to avoid primer dimers) for PCR, and the resulting products are then purified and sequenced to generate high depth of coverage across the entire viral genome. This amplicon tiling approach proved instrumental in tracking the Zika and Chikungunya virus epidemics in South America [23], and more recently for tracking Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) genome evolution during the Coronavirus Disease 2019 (COVID-19) pandemic [24]. However, amplicon tiling requires a relatively large number of unique primer sequences for even the smallest viral genomes. For example, 35 primer pairs are needed to amplify the approximately 11 Kb genome of Zika [23], 44 primer pairs for the approximately 12 Kb genome of Chikungunya [25], and nearly 100 primer pairs for approximately 30 Kb genome of SARS-CoV-2 [24,26,27]. The genomes of bacterial and eukaryotic pathogens are orders of magnitude larger than viruses (Fig 1B), making tiling approaches impractical. Moreover, eukaryotic pathogens share a much higher degree of sequence similarity with the eukaryotic hosts they infect, resulting in primers that often exhibit off-target binding to host DNA. SWGA attempts to address these challenges by using

smaller pools of roughly 5 to 12, short (8 to 12 nt) oligos—rather than pathogen-specific primer pairs—which are designed to bind with higher frequency (rather than exclusively) to the target pathogen genome compared to the host background genome (Fig 1B). This distinguishes SWGA from traditional whole genome amplification (WGA), which also uses *Phi*29 polymerase but which employs random primers and therefore amplifies all genomes present in a complex sample. By combining SWGA oligos with a highly processive polymerase that exhibits strand displacement and proof-reading ability, long amplicons with low error rates can be generated for the genome of interest, thus allowing the amplification of very large pathogen genomes from complex samples that contain high levels of contaminating DNA.

SWGA—How it started, and how it's going

The first proof-of-concept use of SWGA on a complex sample involved the amplification of the bacterial symbiont, *Wolbachia pipientis*, directly from its *Drosophila* host (Fig 2). Using crude extracts prepared from infected *Drosophila*, this work showed that SWGA resulted in nearly 140-fold greater amplification of *W. pipientis* DNA as compared to *Drosophila* DNA. This degree of enrichment yielded approximately 90% coverage of the *W. pipientis* genome from a modest sequencing effort of just 0.6 to 2.2 billion base pairs, 10-fold less than would have been required from unenriched samples [21]. The authors of this study later produced the first program to design of SWGA primers given any target and background (off-target) genome [20], thus opening the doors to adapt this method more broadly. Since these foundational studies, SWGA has been adapted to target several other bacterial species, including the human pathogens *Coxiella burnetii* [28], *Neisseria meningitidis* [29], and *Treponema pallidum*

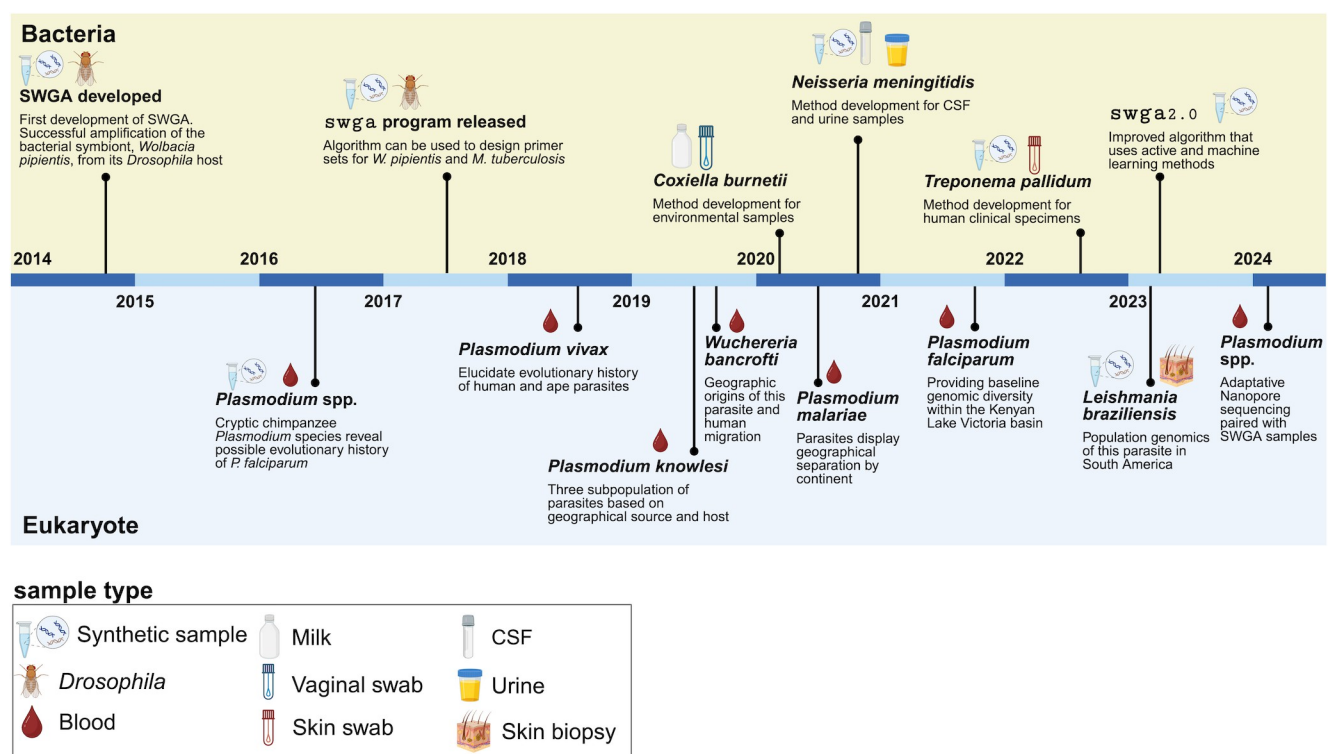


Fig 2. SWGA development over the past decade. Schematic showing major milestones in the development and application of SWGA across bacterial (top; yellow) and eukaryotic (bottom; blue) microbes in multiple sample types. Citations for milestones on top of timeline, from left to right are [20,21,28–30,35], and for bottom are [32,34,37,51,52,56,75,94]. Created with [Biorender.com](https://biorender.com).

<https://doi.org/10.1371/journal.ppat.1012418.g002>

[30] (Fig 2, top). *C. burnetii* is a Biosafety Level 3 (BSL3) pathogen and, as such, requires rigorous safety precautions and extensive training to handle pure cultures of this organism. SWGA enabled whole genome sequences to be generated for this pathogen directly from vaginal swabs, bypassing the need for culture [28]. Similarly, SWGA lowered the barrier to generate and study *N. meningitidis* genomes directly from urine and cerebrospinal fluid (CSF) [29]. Finally, Thurlow and colleagues developed SWGA for *T. pallidum*, the cause of syphilis, which has been difficult to study using genomics due to a paucity of in vitro culture systems [30]. The authors showed that high-quality *T. pallidum* genomes with $\geq 10\times$ coverage across up to 98% of the genome could be generated directly from swabs of syphilis skin lesions.

Although initially envisaged for bacterial organisms [21], SWGA has flourished as a tool to study eukaryotic pathogens (Fig 2, bottom)—most notably, *Plasmodium* species that cause malaria (Fig 3A). *Plasmodium falciparum* was viewed as particularly amenable for SWGA because of the AT-rich nature of its genome (approximately 80% compared to only approximately 59% in human [31]), thus making it easier to design primers that preferentially bind to *P. falciparum* DNA over the background human host DNA. Remarkably, these early studies

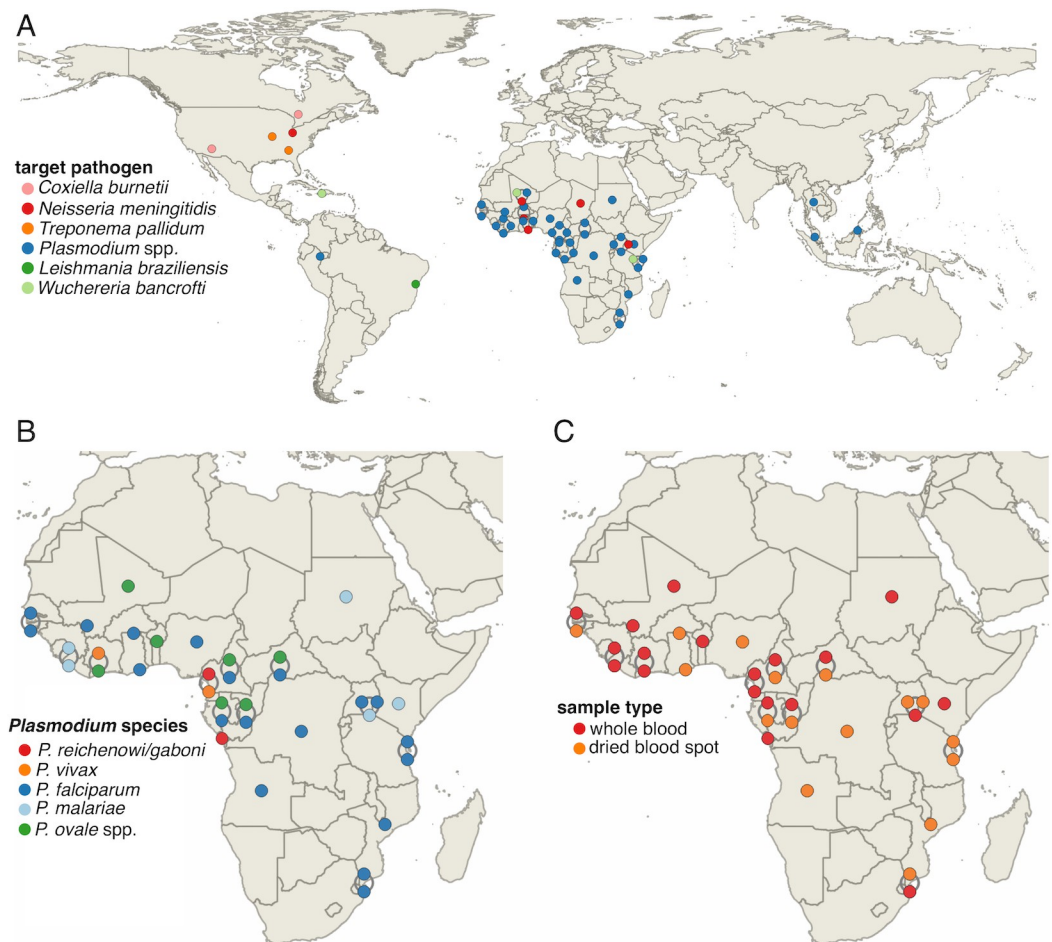


Fig 3. Geographic distribution of genome sequences derived from samples subjected to SWGA from the literature. (A) Geographic distribution colored by the pathogen targeted for study; (B) by *Plasmodium* species targeted; and (C) by sample type used for SWGA. Each point represents one or more samples from the same study for 21 or 14 published manuscripts (panels A and B/C, respectively). Data points are linked by curved lines if the points have the same GPS coordinates. Maps were created in QGIS software [95].

<https://doi.org/10.1371/journal.ppat.1012418.g003>

with *Plasmodium* were successful even on blood samples with parasitemia below the limit of detection by standard microscopic examination of blood smears [32]. SWGA thus enabled genomic analyses of chimpanzee and gorilla *Plasmodium* species from subclinical infections [32–34]. These species represent the closest ancestors of *P. falciparum* and *P. vivax* and revealed new insights into how these parasites evolved to infect humans. Spurred by this success, the *Plasmodium* research community has used SWGA extensively to study the population genomics of *Plasmodium* species from challenging samples with miniscule amounts of parasite DNA, including whole blood and dried blood spots (DBS) (Fig 2, bottom).

The expansion of SWGA into new diseases, host species, and tissues underscored a need for a more robust algorithm for primer selection. Solid tissue specimens, such as biopsies, contain far higher levels of contaminating host DNA than fluid specimens such as blood, CSF, and urine. In addition, pathogens that are similar in AT richness to their mammalian host pose a challenge to primer design. To address these challenges, Dwivedi-Yu and colleagues recently developed open-source software that uses active and machine learning methods to improve the speed and quality of SWGA primer design [35]. As a proof-of-concept for this improved algorithm, the authors successfully developed and tested SWGA primers for *Prevotella melani-nogenica*, an important pathogen in cystic fibrosis patients, but one that has similar GC content as humans and frequently eludes culture-based diagnosis and clinical epidemiology surveillance [36]. Pilling and colleagues used this improved primer design algorithm to develop SWGA primers that successfully amplified the genome of the protozoan parasite, *L. braziliensis*, directly from skin biopsies [37], generating $\geq 10\times$ coverage across more than 80% of the 32 Mb parasite genome. This marked the first time SWGA had been applied to solid tissues, where host DNA vastly outweighs parasite genetic material.

Population genomics powered by SWGA

Population genomics is a vital tool for revealing the evolutionary history of pathogens, shaping our view of the role people and animals play in transmission, and influencing the implementation of mitigation strategies to improve public health. SWGA has arguably advanced population genomic research for *Plasmodium* more so than for any other pathogen (Fig 3). Given the low genetic diversity of *Plasmodium* species, amplicon-based sequencing methods are of little use for population level analyses. While genome-wide single nucleotide polymorphism (SNP) studies can be helpful, they require a priori knowledge of what sites are polymorphic within a population of interest. Thus, whole genome sequencing (WGS)-based methods are ideal for analyzing *Plasmodium* population-level data.

Many techniques have been developed to enrich for *Plasmodium* DNA from clinical samples prior to sequencing; however, these often require significant resources and labor at the time of collection [38–44]. DBS (small amounts of blood blotted on filter paper) have been extensively used in malaria surveillance [45]. These samples are easy to collect and transport and do not require refrigeration [46]. Unfortunately, it is impossible to separate parasite DNA from contaminating host DNA in DBS, limiting their usefulness in *Plasmodium* population genomics. SWGA overcomes these limitations, enabling WGS from DBS, even from low parasitemia samples ($<0.03\%$ parasitemia, or approximately 1,200 parasites/ μl) [46,47]. Comparisons of SWGA-amplified DBS samples to venous blood assayed by traditional WGS have also showed high concordance of allele frequencies and SNP calls, indicating that genomes sequenced after SWGA are as accurate as direct sequencing and of sufficient quality for population genomic studies.

Frequent infections in malaria-endemic areas make it difficult to distinguish new infections from treatment failure due to either drug resistance or inadequate therapy. Guggisberg and

colleagues overcame this using SWGA, examining treatment failure in children-administered fosmidomycin-clindamycin for *P. falciparum* infections [48]. The authors found that initial and recurrent infections were genetically related, but there was no evidence of selection for known drug resistance markers. These results suggest that treatment failure was most likely due to inadequate therapeutic drug concentrations and not preexisting or de novo resistance mutations. More recently, Coonahan and colleagues used SWGA to evaluate the genomic landscape of *P. falciparum* infections in Mozambique, identifying high levels of resistance to sulfadoxine-pyrimethamine, no evidence of continued chloroquine resistance, and multiple genomic regions with signatures of positive selection [49]. While SWGA studies have focused largely on *P. falciparum* (Fig 3B) from whole blood or DBS (Fig 3C), the method has also proven useful in studies of other species, including exploring transmission dynamics of *P. vivax*, population structure of *P. malariae*, and even gene flow in the zoonotic parasite *P. knowlesi* [50–52].

The evolutionary history of *Plasmodium* species revealed by SWGA

Understanding the evolutionary history of pathogens can provide key insights into host–pathogen interactions and help predict the risk of future cross-species transmission. Early studies of *P. falciparum* and its chimpanzee relative *P. reichenowi* led to the hypothesis that *P. falciparum* had coevolved with humans over millions of years. In a landmark study, Liu and colleagues performed single genome amplification of *Plasmodium* DNA from the stool of wild-living chimpanzees, gorillas, and bonobos to show that the ancestor of *P. falciparum* was transmitted from gorillas into humans [53], resulting in one of the most devastating infectious diseases in human history. This finding raised key questions about the relationship of human and ape *Plasmodium* species and genetic traits that might predispose to cross-species transmission. However, the endangered status of African apes and lack of culture systems for ape parasites made comparative genomic studies challenging to carry out.

Given their protected status, the majority of blood samples from chimpanzees and gorillas in Africa are collected opportunistically (i.e., during routine health screens of apes in sanctuaries). These samples typically have extremely low levels of *Plasmodium* parasites (0.00081% to 0.14% *Plasmodium* DNA). To overcome this, Sundararaman and colleagues combined SWGA with methylation-dependent digestion (MDD) of host DNA [32]. MDD takes advantage of the differential methylation of *Plasmodium* and ape DNA to selectively degrade the latter. The addition of MDD to SWGA could improve enrichment in 2 ways. First, MDD directly degrades host DNA, thereby limiting off-target binding of SWGA primers. Second, off-target amplification of host DNA consumes deoxynucleotide triphosphates (dNTPs) and primers, which limits the total amount of DNA that can be added to a SWGA reaction while still achieving significant amplification. By decreasing off-target amplification, MDD could allow the use of more DNA and, therefore, more target *Plasmodium* genomes, per SWGA reaction. The addition of more target genomes improves the evenness of SWGA amplification, as well as the depth and breadth of coverage after sequencing.

Combining SWGA and MDD, Sundararaman and colleagues produced the first full-length genomes of *Plasmodium* species from naturally infected chimpanzees. Analyses of these sequences revealed that chimpanzee relatives of *P. falciparum* exhibit 10-fold higher within-species diversity than *P. falciparum*, suggesting that the jump from apes to humans may have occurred within the last 10,000 years [32,33]. Comparative genomics from SWGA-enriched samples also identified a horizontal transfer of 2 essential invasion genes, as well as the duplications and rapid diversification of the FIKK multigene family of protein kinases, which have

been implicated in remodeling infected erythrocytes [32]. These adaptations may have provided the necessary tools for *Plasmodium* to transmit and adapt to humans [54].

Studies of African apes also revealed close genetic relatives of *P. vivax* in both chimpanzees and gorillas. Loy and colleagues used SWGA to produce near full-length genomes of 2 *P. vivax*-like parasites from chimpanzees [34]. These data showed that, like *P. falciparum*, *P. vivax* strains circulating in chimpanzees and gorillas in Africa are 10 times more diverse than their human counterparts, suggesting an African origin for this important human pathogen and an extreme bottleneck in human *P. vivax* [34]. This contradicted the previous belief that *P. vivax* originated as an Asian malaria [55].

Challenges and opportunities for SWGA

SWGA fulfills an important need not currently met by traditional pathogen genomic approaches, such as amplicon tiling PCR for viral surveillance or WGS based on random hexamer priming. Despite the progression of SWGA from its initial development for bacteria to rapid adaptation for use in eukaryotic microbial genomics (Fig 2), the latter has largely been limited to *Plasmodium* in blood samples (Fig 3B and 3C), and only 2 neglected tropical diseases (filariasis [56] and leishmaniasis [37]) have been explored using this method. For SWGA to be more widely used across a diverse range of infectious diseases and sample types, there need to be improvements in method development. For example, we observed a relatively low success rate (approximately 25%) for SWGA on skin biopsies from patients harboring *L. braziliensis* infections [37], which can likely be attributed to the extremely low abundance of this pathogen together with the high levels of contaminating host DNA in solid tissue. This is supported by patient data showing that *L. braziliensis* burden is positively correlated with SWGA success (Fig 4A). Work on viruses has shown that when pathogen burden is low, primer dimer interactions increase resulting in amplicon “drop out,” which can be improved by primer redesign [27]. Moreover, for reasons that are not well understood, SWGA primer sets that yield a genome from one sample may not be successful for other samples, necessitating the use of multiple primer sets for each sample assayed [21,37]. The consequence of these shortcomings

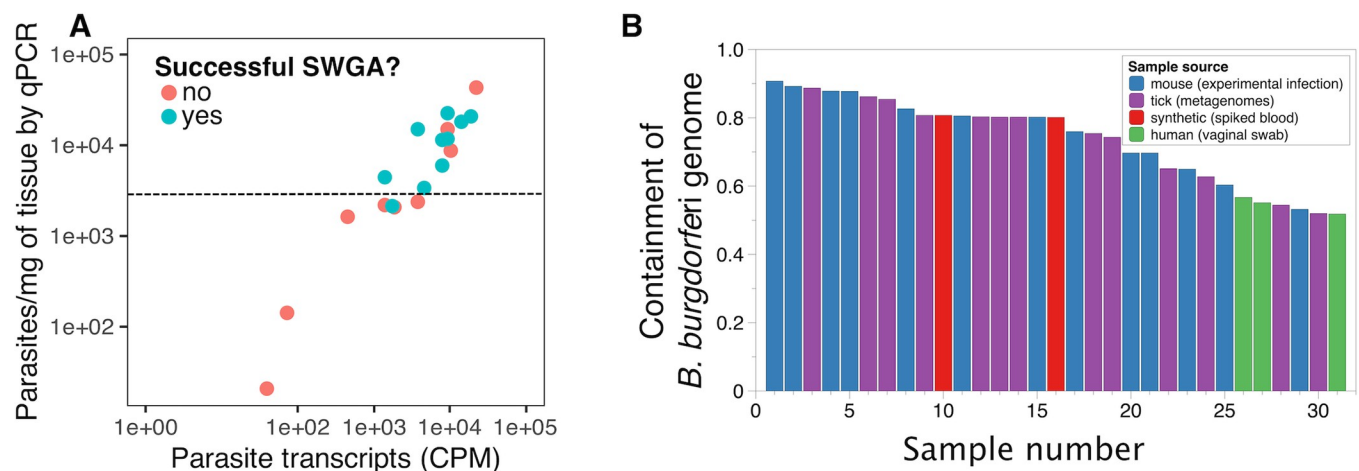


Fig 4. Overcoming limitations of low microbial load for SWGA. (A) Quantifying microbial load to prioritize samples for SWGA. Absolute quantification by qPCR (y-axis) of *L. braziliensis* from patient skin biopsies compared to relative quantification by RNA-seq (x-axis), from Pilling and colleagues [37]. Each point represents a single patient sample. Points are colored based on whether genomes were successfully generated by SWGA for each patient. Dotted line indicates a potential qPCR quantification cutoff that could be used for prioritizing samples for SWGA. Below this threshold, SWGA failed for 6/7 samples but succeeded for 9/12 above. (B) Results from searching approximately 500,000 metagenomes from SRA using the *B. burgdorferi* strain B31 reference genome with Sourmash Branchwater software [65]. Each bar represents a sample colored by source.

<https://doi.org/10.1371/journal.ppat.1012418.g004>

is that many SWGA reactions need to be carried out in parallel to successfully generate genomes for a microbe of interest. Fortunately, newer methods for HTS library preparation dramatically reduce the cost of moving large numbers of SWGA reactions forward for sequencing [57]. One area for improvement would be to integrate artificial intelligence into the SWGA primer design software with experimental data—such as SWGA reaction success as assessed by quantitative PCR (qPCR) or coverage data from sequencing—to identify the reasons why some sets work well while others do not. Similar improvements have been described for viral surveillance to adapt PCR primers and pools as viruses evolve and mutate during an outbreak [27,58,59]. Improved SWGA primer set design would accelerate adoption of this method for insect vectors, fungi, and the microbiome, all of which are areas where population genomics is critical but where SWGA has not yet been applied.

One major difference between SWGA and traditional PCR is the up-front computational cost involved in primer set design. SWGA primer design requires working with at least 2 genomes (target microbe and host background) and searching a very large potential sequence space for short oligos that are likely to match better with the target microbial genome than the host background genome—a process that is computationally intensive [20,35]. This represents a potential impediment to wider adoption of the method, particularly by researchers in lower- and middle-income countries (LMICs) working on neglected tropical diseases, where SWGA could make a major contribution to population genomic studies of pathogens. Therefore, a major area for growth for SWGA would be to develop a resource analogous to the ARTIC network for real-time molecular epidemiology and outbreak responses for viruses. Such a resource could provide online tools for SWGA primer design (e.g., similar to PrimalScheme for viruses [22]) and standardized protocols for carrying out SWGA and analyzing the resulting data. The biology of parasitic helminths and protozoa presents an interesting challenge for SWGA primer design. These microbes frequently have life cycles that involve multiple evolutionarily divergent host species. Our own data suggest that primers designed using one mammalian host as background (e.g., human) can work well when performing SWGA in another mammalian host (e.g., mouse), but not in more distantly related species such as an insect vector [37]. Therefore, an ideal primer design resource would allow users to generate SWGA primer sets for different hosts, effectively allowing population genomics to be carried out across the complete life cycle for a parasite species.

Given the issues with low success rate described above, identifying samples with the greatest potential to yield complete or nearly complete genomes by SWGA is critical. qPCR for microbe-specific marker genes is one simple way to prioritize samples based on amount of target gene present [37] (Fig 4A), but computational approaches may also help guide this prioritization process. Text matching algorithms, such as MinHash, have been adapted to search vast DNA sequence collections to identify samples that contain a microbial genome of interest [60–63], allowing rapid searching of nearly 1 million metagenomic samples publicly available through the Sequence Read Archive (SRA) [64,65] on a standard laptop computer [65]. Such large-scale in silico screens could be used to identify ideal samples, sample types, disease states, or experimental models that may yield the best results with SWGA. For example, searching all SRA metagenomes for *B. burgdorferi*, the cause of Lyme disease and a notoriously difficult pathogen to detect in vivo and culture in vitro, returns 31 samples from 3 public datasets (PRJNA208535, PRJNA723600, and PRJNA981116) that contain greater than 50% of the bacterial genome (inferred by shared k-mer content, or “containment”) (Fig 4B). Not surprisingly, among these samples are total DNA extracts from homogenates of whole *Ixodes scapularis* (deer tick) (Fig 4B, purple bars), as well as human blood spiked with *B. burgdorferi* DNA (Fig 4B, red bars) [66]. Nearly half of the samples identified in this screen were heart, skin, and joint tissue from experimentally infected mice (Fig 4B, blue bars) [66], while 3 samples were

from vaginal swabs collected from a single human individual in a microbiome study of bacterial vaginosis [67]. Interestingly, *B. burgdorferi* has previously been isolated from vaginal secretions [68] and can cause genital ulcers [69–71]. Collectively, these data underscore how in silico screens can identify sample types and disease states potentially amenable to SWGA. Regardless of how samples are selected for SWGA, the primers do not bind exclusively to the target genome, thus the resulting sequence data will still contain a substantial fraction of reads from the background genome(s). These contaminating reads represent wasted sequencing effort and increase experiment costs. Two recent publications combined SWGA with adaptive sequencing [72] on the Oxford Nanopore Technologies platform for use in genome assembly and genomic surveillance of *Plasmodium* spp. [73–75], enabling real-time selective sequencing of only the target microbial species for as little as \$25 per sample.

Other methods for targeted enrichment of pathogen genomes

While this review primarily focuses on SWGA, other methods have been developed for the targeted enrichment of pathogens from complex samples. For example, hybrid capture sequencing is a widely used method that employs custom biotinylated oligonucleotide probes, or “baits,” to capture the target genome allowing unbound contaminating DNA to be washed away, thus enriching for the genome of interest. Although initially developed for the capture of the human exome [76], hybrid capture (e.g., Agilent SureSelect system) has been successfully adapted for many of the pathogens described above [39,43,77–83]. In contrast to SWGA, however, hybrid capture requires expensive baits and high amounts of input DNA (>100 ng), thus limiting its widespread use in LMICs and in situations where the input sample may be limiting [84]. While hybrid capture is a positive selection method, negative selection approaches can also achieve target enrichment through depletion of host DNA. For example, in *P. falciparum* studies, leukocyte depletion has been used to remove nucleated host cells [38,40], but this method is time consuming, requires refrigeration, and cannot be used when blood volume is low [38,46,47,85]. Alternatively, differential DNA methylation patterns between humans and microbes can be exploited to selectively digest host DNA in complex samples [47,85–88]. This method may offer one appealing solution for *Leishmania*, particularly since *L. donovani* reportedly lacks C-5 DNA methylation [89], opening the doors to using methylation-sensitive restriction enzymes to preferentially degrade host DNA. Restriction enzymes paired with SWGA could improve SWGA success rate for *L. braziliensis*. However, multiple methylation-sensitive restriction enzymes have been tested in *Plasmodium* SWGA studies with varying rates of success. Finally, although several commercial kits have been developed for host depletion [90], SWGA required 9-fold less genomic DNA copies/μl compared to some of these kits to generate the same quality of genome sequence for *T. pallidum* [30].

Conclusions

Genome-wide studies of pathogens allow us to investigate population structure and transmission dynamics, link pathogen genotypes to pathogen virulence and persistence or host clinical traits, and monitor drug resistance genes in the population. SWGA complements amplicon tilting PCR for viral surveillance and WGS for generating microbial genomes from pure cultures. This method also extends the population genomics toolkit for pathogens by unlocking comparative genomics in less-than-ideal circumstances, including for pathogens that persist at extremely low abundance, are not experimentally tractable, or have reservoirs in endangered or protected host species such as gorillas and chimpanzees. In addition, eukaryotic pathogens not only have large and complex genomes [91], but they are also the cause of many neglected tropical diseases that lead to significant morbidity and mortality in LMICs. SWGA has and

will continue to unlock genomic surveillance for this important group of organisms and, when combined with mobile and adaptive sequencing technologies, will help to decentralize the process of tracking local infection dynamics [92,93].

Author Contributions

Writing – review & editing: Sesh A. Sundararaman, Dustin Brisson.

References

1. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 2004; 428:37–43. <https://doi.org/10.1038/nature02340> PMID: 14961025
2. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016; 102:3–11. <https://doi.org/10.1016/j.ymeth.2016.02.020> PMID: 27012178
3. Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol*. 2021; 6:3–6. <https://doi.org/10.1038/s41564-020-00834-3> PMID: 33349678
4. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015; 3:e1319. <https://doi.org/10.7717/peerj.1319> PMID: 26500826
5. Aineberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, et al. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014; 11:1144–1146. <https://doi.org/10.1038/nmeth.3103> PMID: 25218180
6. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
7. Youngblut ND, de la Cuesta-Zuluaga J, Reischer GH, Dauser S, Schuster N, Walzer C, et al. Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. *mSystems*. 2020; 5. <https://doi.org/10.1128/mSystems.01045-20> PMID: 33144315
8. Gurbich TA, Almeida A, Beracochea M, Burdett T, Burgin J, Cochrane G, et al. MGnify Genomes: A Resource for Biome-specific Microbial Genome Catalogues. *J Mol Biol*. 2023; 435:168016. <https://doi.org/10.1016/j.jmb.2023.168016> PMID: 36806692
9. Lesker TR, Durairaj AC, Gálvez EJC, Lagkouvardos I, Baines JF, Clavel T, et al. An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. *Cell Rep*. 2020; 30:2909–2922.e6. <https://doi.org/10.1016/j.celrep.2020.02.036> PMID: 32130896
10. Hisei P, Rudi K, Wilson RC, Hegge FT, Snipen L. HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome*. 2021; 9:165. <https://doi.org/10.1186/s40168-021-01114-w> PMID: 34330336
11. Wilkinson T, Korir D, Ogugo M, Stewart RD, Watson M, Paxton E, et al. 1200 high-quality metagenome-assembled genomes from the rumen of African cattle and their relevance in the context of sub-optimal feeding. *Genome Biol*. 2020; 21:229. <https://doi.org/10.1186/s13059-020-02144-7> PMID: 32883364
12. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric crohn's disease. *Cell Host Microbe*. 2015; 18:489–500. <https://doi.org/10.1016/j.chom.2015.09.008> PMID: 26468751
13. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012; 486:207–214. <https://doi.org/10.1038/nature11234> PMID: 22699609
14. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017; 550:61–66. <https://doi.org/10.1038/nature23889> PMID: 28953883
15. Lind AL, Pollard KS. Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing. *Microbiome*. 2021; 9:58. <https://doi.org/10.1186/s40168-021-01015-y> PMID: 33658077
16. Bazant W, Blevins AS, Crouch K, Beiting DP. Improved eukaryotic detection compatible with large-scale automated analysis of metagenomes. *bioRxiv [Preprint]*. 2022. <https://doi.org/10.1101/2022.03.09.483664>

17. US Preventive Services Task Force, Mangione CM, Barry MJ, Nicholson WK, Cabana M, Chelmow D, et al. Screening for latent tuberculosis infection in adults: US preventive services task force recommendation statement. *JAMA*. 2023; 329:1487–1494. <https://doi.org/10.1001/jama.2023.4899> PMID: 37129649
18. Bernitz N, Kerr TJ, Goosen WJ, Chileshe J, Higgitt RL, Roos EO, et al. Review of Diagnostic Tests for Detection of *Mycobacterium bovis* Infection in South African Wildlife. *Front Vet Sci*. 2021; 8:588697. <https://doi.org/10.3389/fvets.2021.588697> PMID: 33585615
19. Wormser GP, Liveris D, Hanincová K, Brisson D, Ludin S, Stracuzzi VJ, et al. Effect of *Borrelia burgdorferi* genotype on the sensitivity of C6 and 2-tier testing in North American patients with culture-confirmed Lyme disease. *Clin Infect Dis*. 2008; 47:910–914. <https://doi.org/10.1086/591529> PMID: 18724824
20. Clarke EL, Sundararaman SA, Seifert SN, Bushman FD, Hahn BH, Brisson D. swga: a primer design toolkit for selective whole genome amplification. *Bioinformatics*. 2017; 33:2071–2077. <https://doi.org/10.1093/bioinformatics/btx118> PMID: 28334194
21. Leichty AR, Brisson D. Selective whole genome amplification for resequencing target microbial species from complex natural samples. *Genetics*. 2014; 198:473–481. <https://doi.org/10.1534/genetics.114.165498> PMID: 25096321
22. PrimalScheme: primer panels for multiplex PCR. [cited 2024 Mar 25]. Available from: <https://primalscheme.com/>
23. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc*. 2017; 12:1261–1276. <https://doi.org/10.1038/nprot.2017.066> PMID: 28538739
24. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. *Cell*. 2020; 181:997–1003.e9. <https://doi.org/10.1016/j.cell.2020.04.023> PMID: 32359424
25. Diaz MH, Waller JL, Theodore MJ, Patel N, Wolff BJ, Benitez AJ, et al. Development and implementation of multiplex tagman array cards for specimen testing at child health and mortality prevention surveillance site laboratories. *Clin Infect Dis*. 2019; 69:S311–S321. <https://doi.org/10.1093/cid/ciz571> PMID: 31598666
26. Quick J. nCoV-2019 sequencing protocol v3 (LoCost) v3. 2020. <https://doi.org/10.17504/protocols.io.bp2l6n26rgqe/v3>
27. Itokawa K, Sekizuka T, Hashino M, Tanaka R, Kuroda M. Disentangling primer interactions improves SARS-CoV-2 genome sequencing by multiplex tiling PCR. *PLoS ONE*. 2020; 15:e0239403. <https://doi.org/10.1371/journal.pone.0239403> PMID: 32946527
28. Cocking JH, Deberg M, Schupp J, Sahl J, Wiggins K, Porty A, et al. Selective whole genome amplification and sequencing of *Coxiella burnetii* directly from environmental samples. *Genomics*. 2020; 112:1872–1878. <https://doi.org/10.1016/j.ygeno.2019.10.022> PMID: 31678592
29. Itsko M, Retchless AC, Joseph SJ, Norris Turner A, Bazan JA, Sadij AY, et al. Full Molecular Typing of *Neisseria meningitidis* Directly from Clinical Specimens for Outbreak Investigation. *J Clin Microbiol*. 2020; 58. <https://doi.org/10.1128/JCM.01780-20> PMID: 32938738
30. Thurlow CM, Joseph SJ, Ganova-Raeva L, Katz SS, Pereira L, Chen C, et al. Selective Whole-Genome Amplification as a Tool to Enrich Specimens with Low *Treponema pallidum* Genomic DNA Copies for Whole-Genome Sequencing. *mSphere*. 2022; 7:e0000922. <https://doi.org/10.1128/msphere.00009-22> PMID: 35491834
31. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409:860–921. <https://doi.org/10.1038/35057062> PMID: 11237011
32. Sundararaman SA, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, et al. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat Commun*. 2016; 7:11078. <https://doi.org/10.1038/ncomms11078> PMID: 27002652
33. Otto TD, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, et al. Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. *Nat Microbiol*. 2018; 3:687–697. <https://doi.org/10.1038/s41564-018-0162-2> PMID: 29784978
34. Loy DE, Plenderleith LJ, Sundararaman SA, Liu W, Gruszczyk J, Chen Y-J, et al. Evolutionary history of human *Plasmodium vivax* revealed by genome-wide analyses of related ape parasites. *Proc Natl Acad Sci U S A*. 2018; 115:E8450–E8459. <https://doi.org/10.1073/pnas.1810053115> PMID: 30127015
35. Dwivedi-Yu JA, Oppler ZJ, Mitchell MW, Song YS, Brisson D. A fast machine-learning-guided primer design pipeline for selective whole genome amplification. *PLoS Comput Biol*. 2023; 19:e1010137. <https://doi.org/10.1371/journal.pcbi.1010137> PMID: 37068103

36. Yu JA, Oppler ZJ, Mitchell MW, Song YS, Brisson D. A fast machine-learning-guided primer design pipeline for selective whole genome amplification. *bioRxiv* [Preprint]. 2022. <https://doi.org/10.1101/2022.04.27.489632>
37. Pilling OA, Reis-Cunha JL, Grace CA, Berry ASF, Mitchell MW, Yu JA, et al. Selective whole-genome amplification reveals population genetics of *Leishmania braziliensis* directly from patient skin biopsies. *PLoS Pathog.* 2023; 19:e1011230. <https://doi.org/10.1371/journal.ppat.1011230> PMID: 36940219
38. Auburn S, Campino S, Clark TG, Djimde AA, Zongo I, Pinches R, et al. An effective method to purify *Plasmodium falciparum* DNA directly from clinical blood samples for whole genome high-throughput sequencing. *PLoS ONE.* 2011; 6:e22213. <https://doi.org/10.1371/journal.pone.0022213> PMID: 21789235
39. Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU, et al. Whole genome sequencing analysis of *Plasmodium vivax* using whole genome capture. *BMC Genomics.* 2012; 13:262. <https://doi.org/10.1186/1471-2164-13-262> PMID: 22721170
40. Venkatesan M, Amaratunga C, Campino S, Auburn S, Koch O, Lim P, et al. Using CF11 cellulose columns to inexpensively and effectively remove human DNA from *Plasmodium falciparum*-infected whole blood samples. *Malar J.* 2012; 11:41. <https://doi.org/10.1186/1475-2875-11-41> PMID: 22321373
41. Auburn S, Marfurt J, Maslen G, Campino S, Ruano Rubio V, Manske M, et al. Effective preparation of *Plasmodium vivax* field isolates for high-throughput whole genome sequencing. *PLoS ONE.* 2013; 8:e53160. <https://doi.org/10.1371/journal.pone.0053160> PMID: 23308154
42. Nair S, Nkhoma SC, Serre D, Zimmerman PA, Gorena K, Daniel BJ, et al. Single-cell genomics for dissection of complex malaria infections. *Genome Res.* 2014; 24:1028–1038. <https://doi.org/10.1101/gr.168286.113> PMID: 24812326
43. Hupalo DN, Luo Z, Melnikov A, Sutton PL, Rogov P, Escalante A, et al. Population genomics studies identify signatures of global dispersal and drug resistance in *Plasmodium vivax*. *Nat Genet.* 2016; 48:953–958. <https://doi.org/10.1038/ng.3588> PMID: 27348298
44. Pearson RD, Amato R, Auburn S, Miotto O, Almagro-Garcia J, Amaratunga C, et al. Genomic analysis of local variation and recent evolution in *Plasmodium vivax*. *Nat Genet.* 2016; 48:959–964. <https://doi.org/10.1038/ng.3599> PMID: 27348299
45. Hsiang MS, Lin M, Dokomajilar C, Kemere J, Pilcher CD, Dorsey G, et al. PCR-based pooling of dried blood spots for detection of malaria parasites: optimization and application to a cohort of Ugandan children. *J Clin Microbiol.* 2010; 48:3539–3543. <https://doi.org/10.1128/JCM.00522-10> PMID: 20686079
46. Oyola SO, Ariani CV, Hamilton WL, Kekre M, Amenga-Etego LN, Ghansah A, et al. Whole genome sequencing of *Plasmodium falciparum* from dried blood spots using selective whole genome amplification. *Malar J.* 2016; 15:597. <https://doi.org/10.1186/s12936-016-1641-7> PMID: 27998271
47. Cowell AN, Loy DE, Sundararaman SA, Valdivia H, Fisch K, Lescano AG, et al. Selective Whole-Genome Amplification Is a Robust Method That Enables Scalable Whole-Genome Sequencing of *Plasmodium vivax* from Unprocessed Clinical Samples. *MBio.* 2017; 8. <https://doi.org/10.1128/mBio.02257-16> PMID: 28174312
48. Guggisberg AM, Sundararaman SA, Lanaspá M, Moraleda C, González R, Mayor A, et al. Whole-Genome Sequencing to Evaluate the Resistance Landscape Following Antimalarial Treatment Failure With Fosmidomycin-Clindamycin. *J Infect Dis.* 2016; 214:1085–1091. <https://doi.org/10.1093/infdis/jiw304> PMID: 27443612
49. Coonahan E, Gage H, Chen D, Noormahomed EV, Buene TP, Mendes de Sousa I, et al. Whole-genome surveillance identifies markers of *Plasmodium falciparum* drug resistance and novel genomic regions under selection in Mozambique. *MBio.* 2023; 14:e0176823. <https://doi.org/10.1128/mbio.01768-23> PMID: 37750720
50. Cowell AN, Valdivia HO, Bishop DK, Winzeler EA. Exploration of *Plasmodium vivax* transmission dynamics and recurrent infections in the Peruvian Amazon using whole genome sequencing. *Genome Med.* 2018; 10:52. <https://doi.org/10.1186/s13073-018-0563-0> PMID: 29973248
51. Ibrahim A, Diez Benavente E, Nolder D, Proux S, Higgins M, Muwanguzi J, et al. Selective whole genome amplification of *Plasmodium malariae* DNA from clinical samples reveals insights into population structure. *Sci Rep.* 2020; 10:10832. <https://doi.org/10.1038/s41598-020-67568-4> PMID: 32616738
52. Benavente ED, Gomes AR, De Silva JR, Grigg M, Walker H, Barber BE, et al. Whole genome sequencing of amplified *Plasmodium knowlesi* DNA from unprocessed blood reveals genetic exchange events between Malaysian Peninsular and Borneo subpopulations. *Sci Rep.* 2019; 9:9873. <https://doi.org/10.1038/s41598-019-46398-z> PMID: 31285495
53. Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, et al. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature.* 2010; 467:420–425. <https://doi.org/10.1038/nature09442> PMID: 20864995

54. Sargeant TJ, Marti M, Caler E, Carlton JM, Simpson K, Speed TP, et al. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* 2006; 7:R12. <https://doi.org/10.1186/gb-2006-7-2-r12> PMID: 16507167
55. Wiscovitch-Russo R, Narganes-Stordes Y, Cano RJ, Toranzos GA. Origin of the New World *Plasmodium vivax*: Facts and New Approaches. *Int Microbiol.* 2019; 22:337–342. <https://doi.org/10.1007/s10123-018-00053-1> PMID: 30810995
56. Small ST, Labbé F, Coulibaly YI, Nutman TB, King CL, Serre D, et al. Human Migration and the Spread of the Nematode Parasite *Wuchereria bancrofti*. *Mol Biol Evol.* 2019; 36:1931–1941. <https://doi.org/10.1093/molbev/msz116> PMID: 31077328
57. Gaio D, Anantanawat K, To J, Liu M, Monahan L, Darling AE. Hackflex: low-cost, high-throughput, Illumina Nextera Flex library construction. *Microb Genom.* 2022; 8. <https://doi.org/10.1099/mgen.0.000744> PMID: 35014949
58. Ulhuq FR, Barge M, Falconer K, Wild J, Fernandes G, Gallagher A, et al. Analysis of the ARTIC V4 and V4.1 SARS-CoV-2 primers and their impact on the detection of Omicron BA.1 and BA.2 lineage-defining mutations. *Microb Genom.* 2023; 9. <https://doi.org/10.1099/mgen.0.000991> PMID: 37083576
59. Lambisia AW, Mohammed KS, Makori TO, Ndwiga L, Mburu MW, Morobe JM, et al. Optimization of the SARS-CoV-2 ARTIC Network V4 Primers and Whole Genome Sequencing Protocol. *Front Med (Lausanne).* 2022; 9:836728. <https://doi.org/10.3389/fmed.2022.836728> PMID: 35252269
60. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016; 17:132. <https://doi.org/10.1186/s13059-016-0997-x> PMID: 27323842
61. Titus Brown C, Irber L. sourmash: a library for MinHash sketching of DNA. *J Open Source Softw.* 2016; 1. <https://doi.org/10.21105/joss.00027>
62. Lumian J, Sumner D, Grettenberger C, Jungblut AD, Irber L, Pierce-Ward NT, et al. Biogeographic Distribution of Five Antarctic Cyanobacteria Using Large-Scale k-mer Searching with sourmash branchwater. *bioRxiv [Preprint].* 2022. <https://doi.org/10.1101/2022.10.27.514113>
63. Viehweger A, Blumenscheit C, Lippmann N, Wyres KL, Brandt C, Hans JB, et al. Context-aware genomic surveillance reveals hidden transmission of a carbapenemase-producing *Klebsiella pneumoniae*. *Microb Genom.* 2021; 7. <https://doi.org/10.1099/mgen.0.000741> PMID: 34913861
64. Leinonen R, Sugawara H, Shumway M. International Nucleotide Sequence Database Collaboration. The sequence read archive. *Nucleic Acids Res.* 2011; 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019> PMID: 21062823
65. Irber L, Pierce-Ward NT, Brown CT. Sourmash Branchwater Enables Lightweight Petabyte-Scale Sequence Search. *bioRxiv [Preprint].* 2022. <https://doi.org/10.1101/2022.11.02.514947>
66. Jain K, Tagliafierro T, Marques A, Sanchez-Vicente S, Gokden A, Fallon B, et al. Development of a capture sequencing assay for enhanced detection and genotyping of tick-borne pathogens. *Sci Rep.* 2021; 11:12384. <https://doi.org/10.1038/s41598-021-91956-z> PMID: 34117323
67. Ravel J, Brotman RM, Gajer P, Ma B, Nandy M, Fadrosch DW, et al. Daily temporal dynamics of vaginal microbiota before, during and after episodes of bacterial vaginosis. *Microbiome.* 2013; 1:29. <https://doi.org/10.1186/2049-2618-1-29> PMID: 24451163
68. Middelveen MJ, Burke J, Sapi E, Bandoski C, Filush KR, Wang Y, et al. Culture and identification of *Borrelia spirochetes* in human vaginal and seminal secretions. [version 3; peer review: 2 approved, 2 not approved]. *F1000Res.* 2014; 3:309. <https://doi.org/10.12688/f1000research.5778.3> PMID: 28690828
69. Fesler MC, Middelveen MJ, Burke JM, Stricker RB. Erosive Vulvovaginitis Associated With *Borrelia burgdorferi* Infection. *J Investig Med High Impact Case Rep.* 2019; 7:2324709619842901. <https://doi.org/10.1177/2324709619842901> PMID: 31043089
70. Finch JJ, Wald J, Ferenczi K, Khalid S, Murphy M. Disseminated Lyme disease presenting with nonsexual acute genital ulcers. *JAMA Dermatol.* 2014; 150:1202–1204. <https://doi.org/10.1001/jamadermatol.2014.1072> PMID: 25162635
71. Middelveen MJ, Haggblad JS, Lewis J, Robichaud GA, Martinez RM, Shah JS, et al. Dermatological and genital manifestations of Lyme disease including morrellons disease. *Clin Cosmet Investig Dermatol.* 2021; 14:425–436. <https://doi.org/10.2147/CCID.S299526> PMID: 33986606
72. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* 2022; 23:11. <https://doi.org/10.1186/s13059-021-02582-x> PMID: 35067223
73. De Meulenaere K, Cuypers WL, Rosanas-Urgell A, Laukens K, Cuypers B. Selective whole-genome sequencing of *Plasmodium* parasites directly from blood samples by Nanopore adaptive sampling. *bioRxiv [Preprint].* 2022. <https://doi.org/10.1101/2022.11.29.518068>

74. de Cesare M, Mwenda M, Jeffreys AE, Chirwa J, Drakeley C, Schneider K, et al. Flexible and cost-effective genomic surveillance of *P. falciparum* malaria with targeted nanopore sequencing. *Nat Commun*. 2024; 15:1413. <https://doi.org/10.1038/s41467-024-45688-z> PMID: 38360754
75. Higgins M, Manko E, Ward D, Phelan JE, Nolder D, Sutherland CJ, et al. New reference genomes to distinguish the sympatric malaria parasites, *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. *Sci Rep*. 2024; 14:3843. <https://doi.org/10.1038/s41598-024-54382-5> PMID: 38360879
76. Chen R, Im H, Snyder M. Whole-Exome Enrichment with the Agilent SureSelect Human All Exon Platform. *Cold Spring Harb Protoc*. 2015; 2015:626–633. <https://doi.org/10.1101/pdb.prot083659> PMID: 25762417
77. Domagalska MA, Imamura H, Sanders M, Van den Broeck F, Bhattarai NR, Vanaerschot M, et al. Genomes of *Leishmania* parasites directly sequenced from patients with visceral leishmaniasis in the Indian subcontinent. *PLoS Negl Trop Dis*. 2019; 13:e0007900. <https://doi.org/10.1371/journal.pntd.0007900> PMID: 31830038
78. Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nat Microbiol*. 2016; 2:16190. <https://doi.org/10.1038/nmicrobiol.2016.190> PMID: 27748767
79. Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A, Herbig A, et al. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nat Microbiol*. 2016; 2:16245. <https://doi.org/10.1038/nmicrobiol.2016.245> PMID: 27918528
80. Chen W, Šmajš D, Hu Y, Ke W, Pospíšilová P, Hawley KL, et al. Analysis of *Treponema pallidum* Strains From China Using Improved Methods for Whole-Genome Sequencing From Primary Syphilis Chancres. *J Infect Dis*. 2021; 223:848–853. <https://doi.org/10.1093/infdis/jiaa449> PMID: 32710788
81. Clark SA, Doyle R, Lucidarme J, Borrow R, Breuer J. Targeted DNA enrichment and whole genome sequencing of *Neisseria meningitidis* directly from clinical specimens. *Int J Med Microbiol*. 2018; 308:256–262. <https://doi.org/10.1016/j.ijmm.2017.11.004> PMID: 29153620
82. Melnikov A, Galinsky K, Rogov P, Fennell T, Van Tyne D, Russ C, et al. Hybrid selection for sequencing pathogen genomes from clinical samples. *Genome Biol*. 2011; 12:R73. <https://doi.org/10.1186/gb-2011-12-8-r73> PMID: 21835008
83. Smith M, Campino S, Gu Y, Clark TG, Otto TD, Maslen G, et al. An In-Solution Hybridisation Method for the Isolation of Pathogen DNA from Human DNA-rich Clinical Samples for Analysis by NGS. *Open Genomics J*. 2012; 5. <https://doi.org/10.2174/1875693X01205010018> PMID: 24273626
84. Domagalska MA, Dujardin J-C. Next-Generation Molecular Surveillance of TriTryp Diseases. *Trends Parasitol*. 2020; 36:356–367. <https://doi.org/10.1016/j.pt.2020.01.008> PMID: 32191850
85. Joste V, Guillochon E, Clain J, Coppée R, Houzé S. Development and Optimization of a Selective Whole-Genome Amplification To Study *Plasmodium ovale* Spp. *Microbiol Spectr*. 2022; 10:e0072622. <https://doi.org/10.1128/spectrum.00726-22> PMID: 36098524
86. Teyssier NB, Chen A, Duarte EM, Sit R, Greenhouse B, Tessema SK. Optimization of whole-genome sequencing of *Plasmodium falciparum* from low-density dried blood spot samples. *Malar J*. 2021; 20:116. <https://doi.org/10.1186/s12936-021-03630-4> PMID: 33637093
87. Oyola SO, Gu Y, Manske M, Otto TD, O'Brien J, Alcock D, et al. Efficient depletion of host DNA contamination in malaria clinical sequencing. *J Clin Microbiol*. 2013; 51:745–751. <https://doi.org/10.1128/JCM.02507-12> PMID: 23224084
88. Shah Z, Adams M, Moser KA, Shrestha B, Stucke EM, Laufer MK, et al. Optimization of parasite DNA enrichment approaches to generate whole genome sequencing data for *Plasmodium falciparum* from low parasitaemia samples. *Malar J*. 2020; 19:135. <https://doi.org/10.1186/s12936-020-03195-8> PMID: 32228559
89. Cuypers B, Dumetz F, Meysman P, Laukens K, De Muylder G, Dujardin J-C, et al. The Absence of C-5 DNA Methylation in *Leishmania donovani* Allows DNA Enrichment from Complex Samples. *Microorganisms*. 2020; 8. <https://doi.org/10.3390/microorganisms8081252> PMID: 32824654
90. Heravi FS, Zakrzewski M, Vickery K, Hu H. Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J Microbiol Methods*. 2020; 170:105856. <https://doi.org/10.1016/j.mimet.2020.105856> PMID: 32007505
91. Alvarez-Jarreta J, Amos B, Aurrecoechea C, Bah S, Barba M, Barreto A, et al. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center in 2023. *Nucleic Acids Res*. 2024; 52:D808–D816. <https://doi.org/10.1093/nar/gkad1003> PMID: 37953350
92. Onwuamah CK, Kanteh A, Abimbola BS, Ahmed RA, Okoli CL, Shaibu JO, et al. SARS-CoV-2 sequencing collaboration in west Africa shows best practices. *Lancet Glob Health*. 2021; 9:e1499–e1500. [https://doi.org/10.1016/S2214-109X\(21\)00389-2](https://doi.org/10.1016/S2214-109X(21)00389-2) PMID: 34678187

93. Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022; 603:679–686. <https://doi.org/10.1038/s41586-022-04411-y> PMID: [35042229](https://pubmed.ncbi.nlm.nih.gov/35042229/)
94. Osborne A, Manko E, Takeda M, Kaneko A, Kagaya W, Chan C, et al. Characterizing the genomic variation and population dynamics of *Plasmodium falciparum* malaria parasites in and around Lake Victoria, Kenya. *Sci Rep*. 2021; 11:19809. <https://doi.org/10.1038/s41598-021-99192-1> PMID: [34615917](https://pubmed.ncbi.nlm.nih.gov/34615917/)
95. Welcome to the QGIS project! [cited 2024 Mar 27]. Available from: <https://www.qgis.org/en/site/>