INTRODUCTION

Whereas the first reference genome sequences of species of *Plasmodium*1–3 and *Anopheles*4 were generated by Sanger sequencing methods, advances in how genetic material is sequenced using so-called “deep sequencing” technology mean that sequence data can now be generated faster, more cheaply, and at much higher volume than before.2 The malaria community has embraced this revolution, and studies using deep sequencing methods have recently appeared that 1) probe the population genetics of *Plasmodium falciparum* on local6 and global7 scales, 2) enable comparative genomic studies between closely related species, for example, *Plasmodium vivax* and its sister taxon *Plasmodium cynomolgi* in the monkey malaria clade,9,9 and 3) generate additional reference genomes for 16 species of *Anopheles*.9 With the development of novel host deoxyribonucleic acid (DNA) depletion methods such as “hybrid selection,”10 we now stand poised to move genomics from the bench to the field.10

The rapid accumulation of new genome sequences of malaria parasites collected worldwide has also been accompanied by an increasing interest in understanding the patterns and mechanisms of genetic variation at the population level. “Population genomics” provides the tools to examine, for example, how new antigenic variants or drug-resistant strains emerge and spread in field parasite populations. A variety of DNA barcoding technologies are now available to sample subsets of genetic variants, such as single-nucleotide polymorphisms (SNPs), identified from whole genome sequencing (WGS). These genetic variants can collectively portray evolutionary changes in a population over time. In particular, changes in population structure can indicate changes in transmission, an important variable in malaria epidemiology.

*Address correspondence to Jane M. Carlton, Center for Genomics and Systems Biology, New York University, 12 Waverly Place, New York, NY 10003. E-mail: jane.carlton@nyu.edu*
ways that ICEMR groups are utilizing independent, high MAF SNPs to address changes in a parasite population over time, as a possible proxy for changing transmission dynamics. (SNP genotyping can be carried out using several different platforms that vary in the chemistries used, and each with its own advantages and disadvantages, but these are not discussed here.)

High MAF SNPs have been identified by screening available genome sequence data from multiple geographic settings to detect neutral, unlinked SNPs that are highly variable among the global parasite population. One set of these markers is the *P. falciparum* molecular barcode tool that surveys 24 unlinked SNPs (“24 SNP barcode”). This barcode has been used to demonstrate changes in parasite population structure coincident with a decline of malaria transmission in the west Africa ICEMR. The basic principle is that under high-transmission settings, there are a relatively large number of parasite genotypes that are being transmitted by mosquitoes to the human population, since cross-fertilization generates novel genotypes in the mosquito. As control interventions are deployed that successfully reduce transmission, the number of parasite genotypes in the population will be reduced, increasing the chance of self-fertilization. Eventually, as transmission decreases, the parasites within infected individuals will become genetically more similar. In the most extreme case, where only self-fertilization occurs, there will be evidence of clonal parasite genotypes being transmitted. Thus, determining the relatedness of parasites within a population can provide important data about malaria transmission, as well as enable evaluation of the success of interventions that reduce malaria transmission over time. In west Africa, the tool has been applied to show emergence of highly related parasites—including evidence of clonal parasite population structures—with increased deployment of malaria-reducing interventions.

Several other ICEMR groups have also been deploying genotyping methods, including Malawi, southern Africa, southwest Pacific, and Latin America. For example, the southern Africa ICEMR (Zambia and Zimbabwe) has used the 24 SNP barcode to interrogate the impact of intervention strategies, detecting a decrease in parasite population signals corresponding with decreases in transmission (Sungano Mharakurwa and others, unpublished). In the Latin American ICEMR, this tool too has been used for outbreak investigation to show the clonal expansion of parasite populations in Panama.

Several groups have been working on the development of a similar tool for *P. vivax*, distilling SNPs from *P. vivax* WGS data. One of these tools is based on markers that have a high MAF among a global parasite population and contains approximately 40 informative markers that can distinguish populations based on continental-level geography. A second approach has been to use SNP markers that separate parasites in close geographic proximity based on differences in allelic frequencies (Alyssa Barry and others, southwest Pacific ICEMR, unpublished). Although these tools are still under development, they promise to provide useful genome-wide markers for asking important questions about *P. vivax* population structure.

**Microsatellite genotyping.** Despite the continuous development of novel molecular genotyping methods and high-throughput platforms, microsatellites (tandem repeats of motifs of two to six nucleotides) remain among the most popular and informative markers in population genetics. Several ICEMRs are using them in their studies, and although more details can be found in the accompanying Molecular Epidemiology supplement, we briefly review their use here as well.

Microsatellite abundance seems to correlate positively with genome adenine–thymine (AT) content, which is extremely high in *P. falciparum* (the average AT content reaches 95% in repetitive domains) but low in chromosome-internal regions of *P. vivax*. As a consequence, only ~160 short repetitive sequences, many of them with features of classical microsatellites, have been characterized across the *P. vivax* genome. The extensive variation found in these genetic markers arises mainly from strand-slippage events during DNA replication. Observed microsatellite mutation rates (10⁻³ to 10⁻⁴ per locus per generation) result from the interplay between strand-slippage events and mismatch repair, which counteracts DNA slippage during replication. Thus, the high mutation rates of microsatellites allow their use whenever there is a need to understand population genetic patterns that emerged relatively recently and locally. On the other hand, SNPs and partial or whole genome sequence data are more appropriate for comparisons across the worldwide distribution of malaria parasites, since these patterns are the result of long-term processes (e.g., global patterns of population structure). The combined use of microsatellite markers with SNPs is a powerful population genomics approach that has been used in genome-wide association studies (GWAS), for example, to identify loci linked to artemisinin resistance, or multiple origins of drug-resistant haplotypes. Several ICEMRs are too using a combination of both markers for their projects, for example, the India ICEMR’s planned *P. vivax* and *P. falciparum* GWAS studies use microsatellites to identify unrelated samples without extreme population structuring that are suitable for WGS.

**SEQUENCING APPROACHES TO EXAMINING MALARIA INFECTIONS**

Sanger sequencing (DNA sequencing based on incorporation of chain-terminating dideoxynucleotides during in vitro replication of a DNA molecule) has long been the workhorse of molecular biology efforts in malaria research. Although the first reference genome sequences of species of *Plasmodium* and *Anopheles* were generated by Sanger sequencing, deep sequencing is now almost exclusively used for whole genome studies, including probing the population genetics of *P. falciparum* on local and global scales, and comparative genomic studies between closely related species, for example, *P. vivax* and its sister taxon *P. cynomolgi* in the monkey malaria clade. Indeed, the use of deep sequencing is now encroaching on areas where Sanger sequencing used to rule, as exemplified below.

An example of deep sequencing methods providing extremely high resolution and sensitivity over Sanger sequencing is the use of the technology to examine within-host diversity of *Plasmodium* infections by amplicon sequencing. Briefly, this method uses targeted amplification of a polymorphic region in the *Plasmodium* genome followed by next-generation sequencing (NGS) so that hundreds or thousands of sequence reads are obtained that can be used to estimate the relative abundance of different clones in a single infection. Amplicon sequencing has been used to assess the genetic diversity in *P. falciparum* clinical samples, for example, using 454 pyrosequencing (454 Life Sciences, Branford, CT) to amplify the circumsporozoite (CS)
gene locus, researchers could successfully identify 57 unique parasite haplotypes in 100 patient samples. Another study involving comparison of 454 amplicon sequencing and Sanger sequencing data for the CS gene revealed that the former could detect more variation than the latter and resolved the genetic diversity in complex infections much more sensitively.

Within the ICEMRs, the India group is using “selection differential” amplicon sequencing to identify the change in frequency of clones in *P. vivax* and *P. falciparum* samples from its three field sites in Chennai, Raurkela, and Nadiad. Preliminary experiments have used a *Plasmodium chabaudi* rodent malaria model, previously used to study aspects of mixed genotype infections, such as the dynamics of multiple infection, within-host competition in genetically diverse infections, and competitive release of drug resistance on drug treatment in mixed infections. *Plasmodium chabaudi* has several genetically distinct laboratory clones that can be distinguished based on sequence variation at the polymorphic antigen *Pcmsp1*. In a proof-of-principle experiment, artificial mixtures of DNA from several *P. chabaudi* clones grown in mice were generated, and the *Pcmsp1* locus amplified and sequenced to high coverage on an Ion Torrent sequencer (Life Technologies Corporation, Carlsbad, CA). Sequencing reads could be aligned uniquely to the *Pcmsp1* sequence of each laboratory clone, and were used to quantify the relative frequency of each clone in the artificial mixtures. Technical and biological sequencing replicates were found to be highly reproducible, and independent estimates of the relative abundance of different clones obtained using quantitative polymerase chain reaction (qPCR) confirmed the accuracy of the method (Figure 1).

The use of amplicon sequencing is well suited to interrogate field samples, as it requires much less DNA than WGS, and can be obtained from filter paper blood spots. However, the use of NGS is not without challenges. For example, as sequence reads are prone to error, it can be difficult to accurately quantify the genetic diversity in a heterogeneous sample. A few computational methods are available for performing global haplotype reconstruction by clustering deep sequencing reads based on sequence variation in the presence or absence of a reference genome (see References 29, 30). The India ICEMR is also developing statistical methods to provide estimates of the power of the approach in detection of low-frequency variants and estimation of population structure, while considering the various sources of errors.

**WGS PROJECTS**

Human, *Anopheles*, and *Plasmodium* genome-scale studies can present many challenges, from the retrieval of sufficient high-quality biological material, to storage and manipulation of large sequence data files, and data analysis and visualization. In particular, low parasitemia and human DNA “contamination” in clinical samples can complicate WGS of *Plasmodium* field isolates. Most current protocols for *Plasmodium* clinical sample processing include CF11 column filtration to remove human leukocytes (the source of contaminating host DNA), or for *P. vivax*, 48 hour schizont maturation ex vivo to increase parasite DNA. The Amazonia ICEMR has found a simple and efficient alternative for removing contaminating host DNA from relatively small volumes (10–50 mL) of venous blood, by adapting commercially available leukocyte depletion filters (Fresenius Kabi BioR Plus or Max) that are commonly used in clinical hemotherapy. The simple filtering procedure described in Figure 2 produces 88% red blood cell recovery and ~94% of the sequence reads mapping to the *P. vivax* Salvador 1 reference genome. In contrast, the India ICEMR is using “hybrid selection” to enrich parasite material from *P. falciparum* and *P. vivax* samples used for whole genome studies.

WGS of *Plasmodium* genomes is complicated by extreme AT-rich regions, particularly in the case of *P. falciparum* (80.6% AT), although the subtelomeric ends of *P. vivax* chromosomes are also highly AT biased. This results in low sequencing coverage of such regions, making genome assembly and variant calling challenging, although in *P. vivax* both of these processes are somewhat easier. Another difficulty in sequencing *Plasmodium* clinical samples is posed by polyclonality, that is, mixed genotype infections. The presence of

**FIGURE 1.** Quantification of multiclonal *Plasmodium chabaudi* mixtures using Ion Torrent amplicon sequencing and quantitative polymerase chain reaction (qPCR). Stacked plots representing proportion of *P. chabaudi* clones ER (black), BC (gray), and AT (white) obtained for two sets of parasite mixtures, shown in panels (A) and (B), containing low concentrations of the ER and AT clones, respectively. qPCR replicates are represented as qPCR_1 and qPCR_2, whereas sequencing replicates are grouped as Run1 and Run2 (biological replicates) and Lane1 and Lane2 (technical replicates).
multiple genotypically distinct parasites in a single patient sample can result in an overestimation of the level of genetic diversity in a parasite genome, and potentially confound subsequent population genetic analysis. It is important to understand and consider these issues when analyzing WGS of *Plasmodium* field isolates. A summary of the WGS projects being undertaken as part of the ICEMR network is shown in Table 1 and described below.

**Mosquito WGS. Anopheles gambiae 1000 Genomes Project.** The east Africa ICEMR has contributed specimens to the *An. gambiae* 1000 Genomes Project, an international consortium of northern and southern partners that uses Illumina-based WGS to develop a detailed understanding of genetic variation in wild-caught populations of the major malaria vectors within the *An. gambiae* complex. The initiative is part of the MalariaGen Network (http://www.malariagen.net/projects/vector/ag1000g) and sequencing is performed at the Wellcome Trust Sanger Institute. Samples have been obtained from across the species range of *An. gambiae* and *Anopheles coluzzii*, with some collections also including *Anopheles arabiensis*. The east Africa ICEMR has contributed specimens from Tororo, eastern Uganda, a region with a high entomological inoculation rate and extensive insecticide resistance. Data from 103 female *An. gambiae* mosquitoes are currently available (http://www.malariagen.net/data/ag1000g-phase1-preview), and the data

---

**Table 1**

| Species          | ICEMR      | Country                      | Sequencing approach | Number of genomes | Main research questions                                      |
|------------------|------------|------------------------------|---------------------|-------------------|-------------------------------------------------------------|
| *Anopheles darlingi* | Amazonia   | Brazil                       | ddRAD               | NK                | Population genetics with regard to deforestation           |
| *Anopheles gambiae*  | East Africa| Uganda                       | WGS                 | 115               | Population genetics                                         |
| *Anopheles arabiensis* | East Africa| Uganda                       | WGS                 | 82                | Interspecies introgression                                  |
| *Plasmodium falciparum* | Southeast Asia| China-Myanmar              | WGS                 | > 150             | Drug resistance                                             |
| *P. falciparum*      | India      | India                        | WGS                 | 30                | Genetic determinants of cerebral malaria                    |
| *Plasmodium vivax*   | Latin America, southeast Asia, Brazil, Colombia, Papua New Guinea, Peru, India, China, Mexico, Thailand | WGS                 | ~180              | Global genetic diversity map                                |
| *P. vivax*           | India      | India                        | WGS                 | > 30              | Genetic diversity map, GWAS of traits                       |
| *P. vivax*           | Amazonia   | Brazil                       | WGS                 | 9                 | Sympatric population                                        |
| *P. vivax*           | Amazonia   | Peru                         | WGS                 | NK                | Methods for enriching parasite-derived DNA in field samples |

ddRAD = double-digest restriction associated DNA; DNA = deoxyribonucleic acid; GWAS = genome-wide association studies; ICEMR = International Centers of Excellence for Malaria Research; NK = not known; WGS = whole genome sequencing.
may be explored using a purpose-built, highly interactive web browser (http://www.malariagen.net/apps/ag1000g/phase1-preview/static/main.html). In later phases of the project, WGS data from sympatric An. arabiensis will be included to investigate the phenomenon of interspecific introgression. Data from a 1,536 SNP Goldengate microarray have demonstrated that there is extensive contemporary gene flow between the two species in this region, and it will be determined if there is any evidence for the transfer of adaptively advantageous alleles as has been observed between An. gambiae and An. coluzzii. In due course, it is likely that other ICEMR partners will join the consortium as the geographic and taxonomic range of the sequencing effort is extended.

Population genomics of Anopheles darlingi in Brazil. Other groups have used partial genome sequencing to sample the diversity of Anopheles. For example, to characterize the genetic diversity of An. darlingi populations collected in different rural settlements in Acre State in the Amazon Forest in Brazil, researchers in the Amazonian ICEMR have used ddRAD-Seq to generate a catalog of ∼150,000 genome-wide 100-bp tag sequences from 90 individual mosquitoes. ddRAD-Seq (restriction-site associated DNA sequencing) uses a pair of restriction enzymes to generate fragments of a genome that are then sequenced to produce genetic markers for population genomic surveys, and is a “reduced representation” method that enables sampling of large, complex genomes such as Anopheles without sequencing the whole genome. Population genetic analysis of different populations with this catalog of sequences is beginning to reveal that 1) genetic diversity of An. darlingi populations appears to be inversely proportional to deforestation; and 2) populations separated by 100 km, indistinguishable in terms of microsatellite diversity, can be differentiated by the SNP polymorphisms identified in the ddRAD-Seq tags.

**P. falciparum WGS.** WGS of drug-resistant *P. falciparum* from patients at the China–Myanmar border. Genome-based approaches to determine the molecular bases of antimalarial drug resistance are a powerful tool. With the recent detection of artemisinin-resistant parasites in multiple countries of the Greater Mekong subregion of southeast Asia, the southeast Asia ICEMR is focusing its drug resistance monitoring and research efforts on parasite populations along the China-Myanmar and Thailand-Myanmar borders, which have unique antimalarial drug use histories. Using parasites procured from clinical cases of malaria from these regions, they are using GWAS to identify genes or genomic regions associated with altered sensitivities to artemisinin and artemisinin combination therapy (ACT) partner drugs. With an aim of obtaining complete genome sequences from ~150 single parasite isolates, this study will provide an unprecedented opportunity to perform population genomic studies and GWAS in the deciphering of artemisinin and ACT drug resistance mechanisms.

WGS of *P. falciparum* from cerebral malaria patients in India. The identification of parasite genetic determinants of cerebral malaria (CM), the most severe complication of *P. falciparum* infection, has been unsuccessful so far. This has been mainly attributed to the highly diverse parasite population in hyper-endemic areas where the analyses were carried out, and to the variation in host susceptibility among patients, which is likely to play an important role in the progression of the disease. In addition, the clinical diagnosis of CM is not straightforward, exemplified by a study conducted in Malawi that revealed ~25% of pediatric patients with World Health Organization-defined CM had a non-malaria cause of death. This misdiagnosis may have represented an important bias in the selection of “false” CM isolates of *P. falciparum* in previous genetic studies. The India ICEMR is evaluating the genetic basis of malaria disease phenotype using WGS of clinical isolates collected from adult patients as part of an ongoing study of CM pathology by magnetic resonance imaging (MRI; see accompanying Pathogenicity article). Since falciparum malaria is complex and results in a broad spectrum of disease in adults, the project focuses on comparing the genomes of *P. falciparum* isolates from CM and uncomplicated malaria (UM) patients, at the opposite ends of the disease severity spectrum. These patient categories are carefully defined as part of the MRI study, and the occurrence of CM is assessed by combining the World Health Organization criteria with a systematic ophthalmic examination for the presence of malaria-associated retinopathies, a clinical feature associated with the severity of malaria, mortality, and duration of coma. Collection of isolates during the same malaria season in a mesoendemic transmission area in combination with population genetics will reduce confounding factors of population structure and reduce the sample selection biases to a minimum. *Plasmodium falciparum* DNA is being extracted directly from venous blood samples and hybrid selection used for parasite DNA enrichment, as described above. WGS on the Illumina HiSeq2500 platform (Illumina Inc., San Diego, CA) will yield up to ~130× coverage of the *P. falciparum* genome, and data processing will include quality control, sequence alignment, and variant discovery based on the *P. falciparum* 3D7 reference sequence.

**P. vivax WGS.** Although *P. falciparum* can be cultured in vitro, so that laboratory-adapted clones can be generated from patient infections and phenotyped for important traits such as drug resistance, no such luxury is available to the *P. vivax* researcher. Thus, genomics is a key methodology that can provide an enormous amount of information about *P. vivax* for relatively little input. Below we describe several of the ICEMR efforts in this area.

Cross-ICEMR global genetic diversity map. One of the largest cross-ICEMR efforts is being undertaken by the India ICEMR as part of a collaboration with the Broad Institute sequencing center. Approximately 180 clinical *P. vivax* isolates collected from vivax-endemic regions under ICEMR purview have been sequenced at New York University (NYU) and the Broad (Table 2) as part of a project to develop a global genetic diversity map of the species (see https://olive.broad institute.org/projects/plasmodium_vivax_hybrid_selection). Each isolate has been verified by species-specific PCR, as well as genotyped at one or several loci to determine if infections were multiclonal, and many isolates are also linked with metadata. As this project progresses, analysis will turn to bioinformatic methods to assess mixed genotype infections that will overlap with the molecular data to establish a firm understanding of each isolate’s history of infection. Additional goals are to use the variant calls and sequence data to generate phylogenetic histories and assess the impact of natural selection on this global population of *P. vivax*.

Whole genome studies of local populations. Although the study described above will provide a global overview of diversity and population genomics, several ICEMRs are undertaking
TABLE 2
Geographical location and number of Plasmodium vivax isolates from ICEMR sites sequenced for generation of a global genetic diversity map

| Country          | ICEMR       | No. of isolates | Comments                  |
|------------------|-------------|-----------------|---------------------------|
| Brazil           | Amazonia    | 20              | Sequenced at Broad        |
| China            | Southeast Asia | 8               | Sequenced at Broad        |
| Colombia         | Latin America | 31             | Sequenced at Broad        |
| India            | India       | 9               | Sequenced at NYU          |
| Mexico           | NA          | 20              | Sequenced at Broad        |
| Peru             | Amazonia    | 47              | Sequenced at Broad        |
| Papua New Guinea | Pacific     | 23              | Sequenced at Broad        |
| Thailand         | NA          | 20              | Sequenced at Broad        |
| Nicaragua, Panama, Thailand, Vietnam | NA | 4 | Monkey-adapted strains from MR4 |

Total: 11 countries 5 ICEMRs 182 isolates –

Broad – Broad Institute of Harvard and MIT (Massachusetts Institute of Technology); MR4 – Malaria Research and Reference Reagent Resource; NA – not applicable; NYU – New York University.

more localized whole genome analyses. For example, researchers in the India ICEMR are planning WGS of tens of P. vivax isolates in their three field sites for a genetic diversity map of the species in India, as well as identifying phenotypes of interest such as severe vivax infections that could be interrogated in a GWAS study. In the Amazonia ICEMR, researchers are analyzing sympatric P. vivax from the ICEMR field site of Remansinho (Acre, Brazil) where the species is responsible for about 85% of malaria cases. In Remansinho area, P. vivax transmission has declined substantially over the last 3 years and P. falciparum has not been found since March 2011. The genomic study of sympatric isolates from an area of low malaria endemicity is particularly interesting in this case due to three factors. First, since these isolates are on the whole more closely related, they are suitable for revealing recent selective pressure events. Second, sympatric isolates are less likely to present population substructure, whose genetic effects (e.g., linkage disequilibrium or particular allele frequency patterns) could be confused with the signature of natural selection. Third, as shown for P. falciparum isolates from Amazonian Peru, sympatric parasite samples are likely to comprise one or both parental genotypes and the recombinant progeny of natural genetic crosses, allowing for a preliminary evaluation of the relative roles of mutations and recombination to generate genome-wide diversity in local parasites. With that in mind, the team has recently sequenced the complete nuclear genomes of nine P. vivax genomes collected in a well-defined area (radius, 100 km) around Remansinho, in the western Amazon Basin of Brazil, close to the border with Bolivia (Table 1).

Comparative genomics of P. vivax. The origin of P. vivax as a species comprises a series of complex evolutionary events. The lineage leading to P. vivax is part of a diverse monophyletic group of nonhuman primate malaria parasites that radiated in southeast Asia, and among them, the macaque parasite P. cynomolgi seems to be the closest known species to P. vivax in that clade. This species is being used in the ICEMRs as part of comparative genomic studies to better understand the polymorphism in P. vivax. The emerging patterns indicate that there is extraordinary polymorphism in P. vivax multigene families that emerged after the P. cynomolgi–P. vivax split. These studies are allowing us to carefully explore fissures in Plasmodium genomes such as the discovery that the msp-3 gene family is not homologous between P. vivax (and related species) and P. falciparum, and to describes how natural selection acts on the P. vivax genome. The adaptive value of such patterns is a matter that will be explored in the context of P. vivax population genomics investigations, and these studies are further stressing that discoveries in P. falciparum do not immediately translate into P. vivax.

IMPLEMENTING GENOTYPING AND GENOMICS TECHNOLOGY IN A FIELD SETTING

Three general challenges must be met to successfully deploy and use genotyping or genomics technology in field-based settings. First, the purchase of durable and practical equipment is key, and this decision is based on a number of factors including cost, equipment versatility (e.g., ability to perform different types of assays), and flexibility of the technology for other purposes (e.g., the Luminex platform can perform immunology-based assays). Once the major use has been ascertained, other considerations including the cost per assay, throughput capabilities, and the sensitivity of the assay in terms of limits of detection, must be considered. Second, there is the issue of purchasing reagents and sustainability of the technology in-country. Reagent procurement can be facilitated in one of two ways: either with collaborators assisting the purchasing, or with the endemic country scientist implementing their own procurement system. These requirements often come with an additional cost burden since shipping such reagents directly can be extremely expensive; there may be a cold chain requirement, and access to reagents may be sporadic. Machine maintenance is another issue, since repairs and preventative maintenance may also be limited. One model that has been successful for several ICEMRs is the utilization of collaborations for procurement and shipping of reagents and supplies and even for swapping out machines for repair. Leveraging these types of relationships also usually has additional cost benefits for purchase of reagents. The third challenge is training in the transfer of the technology, study design, data analysis, and data interpretation, and here the ICEMRs have been highly productive because of their “network” of interactions. For example, workshops have been held in concert with the annual ICEMR meeting (e.g., the Workshop on Population Genetics in Lima, Peru held in August 2014, and the Workshop Using Protein Array Data in Guilin, China held in August 2013), or for those in close proximity there have been joint ICEMR workshops (e.g., the Zambia January 2013 genotyping workshop held jointly between the southern and west African ICEMRs). In addition, the eukaryotic pathogen database EuPathDB has focused on including ICEMR members in its workshops held close to ICEMR sites. These include workshops in Bogota, Colombia (July 2013) and Singapore (February 2014). EuPathDB provides an online compendium of its workshops material (e.g., see here http://workshop.euopathdb.org/singapore/2014/ for the 2014 Singapore workshop material). Thus, the ICEMR
network has promoted training opportunities that move malaria-endemic country scientists toward independence and sustainability of these skills, so that they can better address malaria control and elimination issues.

**Developing small-scale field-based genotyping capability.** Several ICEMRs have invested in training in generation and analysis of genotyping data—including SNPs, copy number variants, and microsatellites—used to understand changes in population structure of both parasite and mosquito. Many genotyping technologies use PCR-based methods to amplify a region surrounding the variant locus, and differ by the marker type and the technology platform used to assay for it. Examples of technologies that are being used by ICEMR sites in the field include: 1) PCR amplification followed by sequencing or restriction endonuclease digestion,54–56 2) TaqMan genotyping,12,57–59 3) high resolution melting,60,61 and (3) ligase detection reaction-fluorescent microsphere assays (e.g., the Luminex platform62). Although the chemistry underpinning these technologies varies, and consequently each has its own advantages and disadvantages, basically all of these and other technologies may be used to identify genetic variants.

**Developing next-generation sequencing expertise in malaria-endemic countries.** Next generation sequencing has revolutionized the field of biomedical research over the past decade, and while the leading sequencing platforms were initially tailored toward large-scale applications, recent technical improvements have led to the development of modestly priced benchtop instruments with fast turnover rates aimed at small-scale laboratories. Three of the most widely used benchtop instruments are the Roche 454 GS Junior (454 Life Sciences, Branford, CT), Illumina MiSeq (Illumina Inc., San Diego, CA), and the Ion Torrent Personal Genome Machine (PGM) (Life Technologies Corporation, Carlsbad, CA). A performance comparison of these sequencing platforms based on technical specifications, data quality, and throughput, as well as a review of their potential applications, setup, and running costs has been described in detail elsewhere.63,64 Although there are advantages and disadvantages associated with each platform, depending on the type of applications

![Figure 3](image-url)

**FIGURE 3.** Screen shots from PlasmoDB illustrating methods to access single-nucleotide polymorphism (SNP) data from International Centers of Excellence for Malaria Research (ICEMR) projects. (A) PlasmoDB home page showing where SNP data can be accessed (red rectangles). (B) Map of the geographic distribution of sequenced *Plasmodium* isolates from ICEMR locations. (C) Sequenced isolates can be compared in PlasmoDB using the search “Identify SNPs based on differences between groups of isolates.” (D) Metadata characteristics such as geographical location can be leveraged to identify SNPs that differentiate isolates from Peru and Colombia. (E) Results are returned in a table containing the location of the SNP and various SNP statistics. (F) The sequence alignment around any SNP can be visualized in the PlasmoDB genome browser.
being considered, they are closely matched in terms of utility and ease of workflow. Therefore, the choice of platform often depends on other factors, such as existing infrastructure, available finances, and personal experience of users.

A view from India. The India ICEMR chose to purchase the Ion Torrent PGM platform (with the fastest throughput, shortest run time, and least expense) for amplicon sequencing and P. vivax whole genome re-sequencing of Indian samples in Delhi, with a sister machine at NYU. Thus protocols can be tested at NYU and directly transferred to the machine in Delhi. The instrument comes with the Torrent Suite software and browser, a web-based interface for planning, monitoring, viewing, and processing results from a sequencing run, including sequence alignment, coverage analysis, and variant detection, which can be used to create a custom analysis workflow. The ease of use and minimum information technology expertise required to operate the data analysis tools on the PGM machine make it an attractive option for endemic country scientists without advanced training in bioinformatics, and without access to a high-performance compute cluster. Currently, the PGM in Delhi has been run successfully several times using clinical samples, and will be running continuously in 2015.

A view from Brazil. High-throughput sequencing facilities are becoming more common in Brazil, and becoming more efficient as they accrue experience. Nevertheless, due to local bureaucratic, tariff/tax, and market issues, genome sequencing can be significantly more expensive than in other countries, and weeks or even months can pass before reagents are available. For that reason, many researchers use the services of foreign companies or facilities for high-throughput sequencing—which solves some of the problems, but still suffers from bureaucracy and is therefore impractical for larger studies. Challenges related to the analysis of high-throughput data are also present in Brazil, for example, incipient computational infrastructure for bioinformatics at universities, and a relative scarcity of students, technicians, and researchers possessing the required quantitative and computational skills to deal with large-scale biological data.

Nonetheless, members of the Amazonia ICEMR have used the Ion Torrent PGM and Ion Proton platforms in Brazil and analyzed the resulting sequence at the National Laboratory for Scientific Computing in Petrópolis, Rio de Janeiro. Reads of ~200 bases in length, and sequencing depths of 50- to 80-fold coverage are being achieved, and SNP analysis is being performed using GATK, and assembly using Newbler (Roche 454), for comparison with the P. vivax Salvador I reference genome and with other less complete P. vivax genomes. Future analysis of the nine Brazilian sympatric genomes will include investigations of genome structure and evolution, recombination, population structure, selective pressure, and phylogenetics.

**FIGURE 4.** Screen shots from PlasmoDB depicting a search for genes containing at least 20 non-synonymous SNPs based on data from ICEMR isolates. (A) The search for genes based on SNP characteristics allows filtering of isolates based on metadata, such as geographic location, and defining the type of SNPs of interest. (B) Results are returned in a table that shows which genes contain the SNPs of interest.
GENOME DATABASES

The eukaryotic pathogen database EuPathDB (http:// eupathdb.org) is an NIH-funded bioinformatic resource center. The goal of this free and widely available online resource is to integrate a variety of large-scale datasets from eukaryotic pathogens and their hosts, and facilitate data interrogation in an easily accessible system. Data types already available in these include genome sequence and annotation, proteomic data (including quantitative), transcriptomic data (RNAseq, microarray), epigenomic data (ChIP-Chip and ChIP-seq), population data based on high-throughput sequencing, and phylogenetic data. The ability to effectively integrate population data necessitates the collection of useful metadata from the data provider that is interoperable between different providers, allowing questions to be asked not only about a specific dataset but also across datasets from diverse groups or projects. EuPathDB has invested a significant effort in ensuring data integration by mapping metadata to already established ontologies such as the Ontology of Biomedical Investigation (see accompanying manuscript describing data management of ICEMR projects).

PlasmoDB (http://plasmodb.org) and HostDB (http://hostdb.org) are components of EuPathDB that focus on integrating data from Plasmodium species and their host organisms. Importantly, the ability of these databases to accommodate high-throughput sequencing data from field studies along with their associated metadata allows sophisticated filtering and searching tools, making both databases suited for cross-ICEMR data interrogation. High-throughput sequencing data from ICEMR projects has been integrated into PlasmoDB and can be searched using gene-specific searches that identify genes with user-defined SNPs, or searches that identify SNPs regardless of their presence in genes (Figure 3). Genes may also be searched based on SNP characteristics identified from ICEMR sequencing data (Figure 4).

Received January 20, 2015. Accepted for publication April 13, 2015.

Acknowledgments: We thank the members of the ICEMR teams who could not be given authorship due to space constraints.

Financial support: This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, as part of the International Centers of Excellence for Malaria Research.

Disclaimer: The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This manuscript bears the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

REFERENCES

1. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuloi SV, Merino EF, Amedo P, Cheng Q, Coulson RMR, Crabbs BS, del Portillo HA, Essien K, Feldblyum TV, Fernandes-Bebbnera C, Gilson PR, Gueye AH, Guo X, Kooij TWA, Korszynska M, Meyer EVS, Nene V, Paulsen I, White O, Ralph SA, Ren O, Sargeant TJ, Salberg SL, Stockeert C, Sullivan SA, Yamamoto MM, Hoffman SL, Wirtman JR, Gardner MJ, Galinski MR, Barnwell JW, Fraser-Liggitt CM. 2008. Comparative genomics of the neglected human malaria parasite Plasmodium vivax. Nature 455: 757–763.

2. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salberg SL, Craig A, Kyes S, Chan MS, Nene V, Shalloom SJ, Suth B, Peterson J, Angiuloi S, Pertea M, Allen J, Selenutk J, Haft D, Mather MV, Vaidya AB, Martin DM, Fairall AB, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LMM, Subramanian GM, Mungall C, Venter JC, Caruucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrett B. 2002. Genome sequence of the Plasmodium falciparum malaria parasite. Nature 419: 511–517.

3. Pain A, Böhme U, Berry AE, Mungall K, Finn RD, Jackson AP, Mourier T, Mistry J, Pasini EM, Adlem E, Aslett M, Balasubramanian S, Borgwardt K, Brooks K, Carret C, Chersevach I, Chillingworth T, Galinski MR, Hall N, Harper D, Harris D, Hauser H, Ivens A, Janssen CS, Keane T, Larke N, Lapp S, Marti M, Motile S, Meyer IM, Ormond D, Peters N, Sanders M, Sanders S, Sergeant TJ, Simmonds M, Smith F, Squares R, Thurstong S, Walker D, White B, Zuidewijk E, Churcher C, Quail MA, Cowman AF, Turner CM, Rajandream MA, Kocken CH, Thomas AW, Newbold CI, Barrett B, Berriman M. 2008. The genome of the simian and human malaria parasite Plasmodium knowlesi. Nature 455: 799–803.

4. Holt RA, Subramaniam GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft C, Abril JF, Anthonv A, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Bossen D, Brown-Stead M, Cai S, Center A, Chaturvedi K, Christofides GK, Chrystall MA, Clamp M, Cravchik A, Curwen V, Daniel A, Delcher A, Dew I, Evans CA, Flanagan M, Grundsbacher-Freimoser A, Fredli E, Gu Z, Guan P, Guigo R, Hildenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsov A, Letunic I, Levitsky A, Liang Y, Lin J, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller M, Mobarry C, Mongin E, Murphy SD, O Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D,
Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brown PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Holland SL, 2002. The genome sequence of the malaria mosquito Anopheles gambiae. Science 298: 129–149.

5. Schuster SC, 2008. Next-generation sequencing transforms today’s biology. Nat Methods 5: 16–18.

6. Chung HH, Park DJ, Galinski KJ, Schaffner SF, Ndiaye D, Ndour O, Mboup S, Wiegand RC, Volkman SK, Sabeti PC, Wirth DF, Neafsey DE, Hartl DL, 2012. Genome sequencing of Plasmodium falciparum malaria parasites from Senegal reveals the demographic history of the population. Mol Biol Evol 29: 3427–3439.

7. Mañake M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, Djimde A, Doumbo O, Zongo I, Ouedraogo JB, Michon P, Mueller I, Siba P, Nzila A, Borrmann S, Kiara SM, Marsh K, Jiang H, Su XZ, Amarantungu C, Fairhurst R, Soechat D, Nosten F, Imwong M, White NJ, Sanders M, Anastasi E, Alcock D, Drury E, Oyola S, Quail MA, Turner DJ, Ruano-Rubio V, Jyothi D, Amengual L, Hubbell C, Jeffreys A, Rowlands K, Sutherland C, Roper C, Mangano V, Modiano D, Tan JC, Ferdig MT, Ambumba-Ngwa A, Conway DJ, Takala-Harrison S, Plowe CV, Rayner JC, Rockett KA, Clark TG, Newbold CI, Bertriman R, MacInnis B, Kwiatkowski DP, 2012. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature 487: 375–379.

8. Tachibana S, Sullivan AA, Kawai S, Nakamura S, Kim HR, Goto N, Arisue N, Palacpac NM, Homma H, Yagi M, Tougan T, Takata Y, Kaneko O, Mita T, Kita K, Yasutomi Y, Sutton PL, Shakhbatyan R, Horii T, Yasunaga T, Barnwell JW, Escalante AA, Carlton JM, Tanabe K, 2012. Plasmodium cynomolgi genome sequences provide insight into the evolution of Plasmodium vivax and the monkey malaria clade. Nat Genet 44: 1051–1055.

9. Neafsey DE, Christopides GK, Collins FH, Emrich SJ, Fontaine MC, Gelbart W, Hahn MW, Howell PL, Kafatos FC, Lawson D, Muskvitch MA, Waterhouse RM, Williams LJ, Besansky NJ, 2013. The evolution of the Anopheles 16 genomes project. G3 (Bethesda) 3: 1191–1194.

10. Melnikov A, Galinsky K, Rogov P, Fennell T, Mylonakis E, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brown PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Holland SL, 2002. The genome sequence of the malaria mosquito Anopheles gambiae. Science 298: 129–149.

11. Cheeseman IH, Miller BA, Nair S, Nkhoma S, Tan A, Tan JC, Schaffner SF, Rogers WH, Fauquier L, Lepai O, Lepothevin C, Malausa T, Revarel S, Salin F, Petit RJ, 2011. Current trends in microsatellite genotyping. Mol Ecol Resour 11: 591–611.

12. Schlotterer C, 2004. The evolution of molecular markers—just a matter of fashion? Nat Rev Genet 5: 63–69.

13. Chenet SM, Schneider KA, Villegas L, Escalante AA, 2012. Local population structure of Plasmodium: impact on malaria control and clinical resistance. Malar J 11: 412.

14. Haas BJ, Payseur BA, 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106: 158–171.

15. Vinayak S, Alam MT, Sem R, Shah NK, Susanti AI, Lim P, Muth S, Maguire JD, Rogers WH, Fauquier L, Barnwell JW, Escalante AA, Wongsrichanalai C, Arify E, Meshnick SR, Udhayakumar V, 2010. Multiple genetic backgrounds of the amplified Plasmodium falciparum multidrug resistance (pmdr1) gene and selective sweep of 184F mutation in Cambodia. J Infect Dis 201: 1551–1560.

16. McCollum AM, Poe AC, Hames M, Huber C, Zhou Z, Shi YP, Ouma P, Vulule J, Boland P, Slutsker L, Barnwell JW, Udhayakumar V, Escalante AA, 2006. Antifolate resistance in Plasmodium falciparum: multiple origins and identification of novel dhfr alleles. J Infect Dis 194: 189–197.

17. Mideo N, Kennedy DA, Carlton JM, Bailey JA, Juliano JJ, Read AF, 2013. Ahead of the curve: next generation estimators of drug resistance in malaria infections. Trends Parasitol 29: 321–329.

18. Bailey JA, Mvolo T, Aragam N, Deng C, Kamwendo D, Martinson F, Hoffman I, Meshnick SR, Juliano JJ, 2012. Use of massively parallel pyrosequencing to evaluate the diversity of and selection on Plasmodium falciparum csp T-cell epitopes in Lisongwe, Malawi. J Infect Dis 206: 580–587.

19. Gandhi K, Thera MA, Coulibaly D, Traoré K, Guindo AB, Doumbo OK, Takala-Harrison S, Plowe CV, 2012. Next generation sequencing to detect variation in the Plasmodium falciparum circumsporozoite protein. Am J Trop Med Hyg 86: 775–781.

20. de Roode JC, Culleton R, Cheesman SJ, Carter R, Read AF, 2004. Host heterogeneity is a determinant of competitive exclusion or coexistence in genetically diverse malaria infections. Proc Biol Sci 271: 1073–1080.

21. Bell AS, Roode JC, Sim D, Read AF, 2006. Within-host competition in genetically diverse malaria infections: parasite virulence and competitive success. Evolution 60: 1358–1371.

22. Hujiben S, Nelson WA, Wargo AR, Sim DG, Drew DR, Read AF, 2010. Chemotherapy, within-host ecology and the fitness of drug-resistant malaria parasites. Evolution 64: 2952–2968.

23. Cheesman SJ, de Roode JC, Read AF, Carter R, 2003. Real-time quantitative PCR for analysis of genetically mixed infections of malaria parasites: technique validation and applications. Mol Biochem Parasitol 131: 83–91.

24. Galinsky K, Valim C, Salmier A, dos Thaís B, Musset L, Legrand E, Faust A, Baniecki ML, Ndiaye D, Daniels RF, Hartl DL, Sabeti PC, Wirth DF, Volkman SK, Neafsey DE, 2015. Development of a field-deployable genomic barcode for Plasmodium vivax detection and global surveillance. PLoS NTD 9: e0003539.

25. Guichoux E, Lagache L, Wagner S, Chaumeil P, Leger P, Lepai O, Lepothevin C, Malausa T, Revarel S, Salin F, Petit RJ, 2011. Current trends in microsatellite genotyping. Mol Ecol Resour 11: 591–611.

26. Bright AT, Tewhey R, Abeles S, Chuquiyauri R, Llanos-Cuentas A, Ferreira MU, Schork NJ, Vinetz JM, Winzeler EA, 2012.
Whole genome sequencing analysis of Plasmodium vivax using whole genome capture. BMC Genomics 13: 262.

33. Neafsey DE, Galinsky K, Jiang RH, Young L, Sykes SM, Saif S, Gujja S, Goldberg JM, Young S, Zeng Q, Chapman SB, Dash AP, Anvikar AR, Sutton PL, Birren BW, Escalante AA, Barnwell JW, 2011. The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. Nat Genet 44: 1046–1050.

34. Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, Kamya MR, Staedke SG, Donnelly MJ, Drakeley C, Greenhouse B, Dorsey O, Lindsay SW, 2014. Estimating the annual entomological inoculation rate for Plasmodium falciparum transmitted by Anopheles gambiae s.s. using three sampling methods in three sites in Uganda. Malar J 13: 111.

35. Mawejje HD, Wilding CS, Rippon EJ, Hughes A, Weetman D, Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, Kamya MR, Staedke SG, Donnelly MJ, Wilding CS, 2013. Insecticide resistance monitoring of field-collected Anopheles gambiae s.l. populations from Jinja, eastern Uganda, identifies high levels of pyrethroid resistance. Med Vet Entomol 27: 276–283.

36. Wilding CS, Weetman D, Rippon EJ, Steen K, Mawejje HD, Barsukov I, Donnelly MJ, 2015. Parallel evolution or purifying selection, not introgression, explains similarity in the pyrethroid detoxification linked GSTE4 of Anopheles gambiae and An. arabiensis. Mol Genet Genomics 290: 201–215.

37. Wean MD, Steen K, Rippon EJ, Mawejje HD, Jordan NJ, Donnelly MJ, Wilding CS, 2014. Contemporary gene flow between wild An. gambiae s.s. and An. arabiensis. Parasit Vectors 7: 345.

38. Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG, Webster M, Antao T, MacInnis B, Kwiatkowski D, Donnelly MJ, 2014. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. Nat Commun 5: 4248.

39. Luca F, Hudson RR, Witosky DB, Di Rienzo A, 2011. A reduced representation approach to population genetic analyses and applications to human evolution. Genome Res 21: 1087–1098.

40. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis PA, Mullikin JC, 2012. The malaria parasite Plasmodium falciparum exhibits greater genetic diversity than Plasmodium falciparum. Proc Natl Acad Sci USA 109: 331–336.

41. Nocell H, Se Y, Schaecker K, Smith BL, Socheat D, Fukuda MM, Artemisinin Resistance in Cambodia 1 Study C, 2008. Evidence of artemisinin-resistant malaria in western Cambodia. N Engl J Med 359: 2619–2626.

42. Dondorp AM, Nosten F, Pi Y, Das D, Phyo AP, Tarning J, Lwin NG, Lewallen S, Liomba NG, Molyneux ME, 2004. Differences in Plasmodium falciparum genotypes in Malawi. Plasmodium falciparum genotypes in Malawi. Nature 431: 50–55.

43. Muhamed P, Chaijaoenkul W, Conguong W, Na-Bangchang K, 2011. SYBR Green I and TaqMan quantitative real-time polymerase chain reaction methods for the determination of amplification of Plasmodium falciparum multidrug resistance 1 gene (pmdr1). J Parasitol 97: 939–942.

44. Milner DA Jr, Vareta J, Valim C, Daniels RF, 2001. A molecular marker for chloroquine-resistant falciparum malaria. N Engl J Med 344: 257–263.

45. Arfne J, Witkowski B, Amonatunga C, Beghain J, Langlois AC, 2012. A microsatellite associated with anti-malarial drug resistance. Drug Dev Res 73: 373–380.

46. Cornejo OE, Fisher D, Escalante AA, 2015. Genome-wide patterns of genetic polymorphism and signatures of selection in Plasmodium vivax. Genome Biol Evol 7: 106–117.

47. Plowe CV, Dijimde A, Bouare M, Doumbou O, Wellems TE, 1995. Pyrrimethamine and proguanil resistance-conferring mutations in Plasmodium falciparum dihydrofolate reductase: polymerase chain reaction methods for surveillance in Africa. Am J Trop Med Hyg 52: 565–568.

48. Dijimde A, Doumbou OK, Cortese JF, Kiyengao K, Doumbou S, Diourte Y, Courby D, Dioko A, Xu Z, Nomura T, Fidock DA, Wellems TE, Plowe CV, 2001. A molecular marker for chloroquine-resistant falciparum malaria. Nature 404: 257–263.

49. Arfne J, Witkowski B, Amonatunga C, Beghain J, Langlois AC, 2012. A microsatellite associated with anti-malarial drug resistance. Drug Dev Res 73: 373–380.

50. Arfne J, Witkowski B, Amonatunga C, Beghain J, Langlois AC, 2012. A microsatellite associated with anti-malarial drug resistance. Drug Dev Res 73: 373–380.
2009. Rapid detection of point mutations in *Plasmodium falciparum* genes associated with antimalarial drugs resistance by using High-Resolution Melting analysis. *J Microbiol Methods* 78: 165–170.

62. McNamara DT, Kasehagen LJ, Grimberg BT, Cole-Tobian J, Collins WE, Zimmerman PA, 2006. Diagnosing infection levels of four human malaria parasite species by a polymerase chain reaction/ligase detection reaction fluorescent microsphere-based assay. *Am J Trop Med Hyg* 74: 413–421.

63. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M, 2012. Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012: 251364.

64. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ, 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30: 434–439.

65. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y, 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13: 341.

66. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20: 1297–1303.

67. Aurrecoechea C, Barreto A, Brestelli J, Brunk BP, Cade S, Doherty R, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Hu S, Iodice J, Kissing JC, Kraemer ET, Li W, Pinney DF, Pitts B, Roos DS, Srinivasamoorthy G, Stoeckert CJ Jr, Wang H, Warrenfeltz S, 2013. EuPathDB: the eukaryotic pathogen database. *Nucleic Acids Res* 41: D684–D691.