Published in final edited form as:

Mol Biochem Parasitol. 2014 ; 194(0): 9–15. doi:10.1016/j.molbiopara.2014.03.006.

Genome-wide polymorphisms and development of a microarray platform to detect genetic variations in *Plasmodium yoelii*

Sethu C. Nair^{a,1}, Sittiporn Pattaradilokrat^{a,b,1}, Martine M. Zilversmit^{a,1}, Jennifer Dommer^c, Vijayaraj Nagarajan^c, Melissa T. Stephens^d, Wenming Xiao^e, John C. Tan^d, and Xin-zhuan Su^{a,*}

^aLaboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA ^bDepartment of Biology, Faculty of Science, Chulalongkorn University, Bangkok 10330, Thailand ^cBioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA ^dThe Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA ^eMetabolism Branch, National Cancer Institutes, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

The rodent malaria parasite *Plasmodium yoelii* is an important model for studying malaria immunity and pathogenesis. One approach for studying malaria disease phenotypes is genetic mapping, which requires typing a large number of genetic markers from multiple parasite strains and/or progeny from genetic crosses. Hundreds of microsatellite (MS) markers have been developed to genotype the *P. yoelii* genome; however, typing a large number of MS markers can be labor intensive, time consuming, and expensive. Thus, development of high-throughput genotyping tools such as DNA microarrays that enable rapid and accurate large-scale genotyping of the malaria parasite will be highly desirable. In this study, we sequenced the genomes of two *P. yoelii* strains (33X and N67) and obtained a large number of single nucleotide polymorphisms (SNPs). Based on the SNPs obtained, we designed sets of oligonucleotide probes to develop a microarray that could interrogate ~11,000 SNPs across the 14 chromosomes of the parasite in a single hybridization. Results from hybridizations of DNA samples of five *P. yoelii* strains or cloned lines (17XNL, YM, 33X, N67 and N67C) and two progeny from a genetic cross (N67 × 17XNL) to the microarray showed that the array had a high call rate (~97%) and accuracy (99.9%) in calling SNPs, providing a simple and reliable tool for typing the *P. yoelii* genome. Our data show that the *P. yoelii* genome is highly polymorphic, although isogenic pairs of parasites were also detected. Additionally, our results indicate that the 33X parasite is a progeny of 17XNL (or YM) and an unknown parasite. The highly accurate and reliable microarray developed in this

*Corresponding author. Tel.: +1 301 402 0876; fax: +1 301 402 2201. xsu@niaid.nih.gov (X.-z. Su).

¹These authors contributed equally to this work.

Competing interests

The authors declare that they have no competing interests.

study will greatly facilitate our ability to study the genetic basis of important traits and the disease it causes.

Keywords

Malaria; Rodent; High-throughput genotyping; Single nucleotide polymorphism; Genome sequencing

1. Introduction

Rodent malaria parasites can be easily maintained in laboratory animals and have been widely employed to investigate the complex biology and disease phenotypes of malarial infection and drug resistance [1–5]. The availability of large numbers of inbred mice allows study of the effects of parasite genetics on a disease phenotype by minimizing the influence of host genetic variation; on the other hand, infection of different inbred or knockout mice or progeny from crosses of different genetic backgrounds with a single parasite clone will help dissect the effects of host genetics on disease manifestation [6,7]. There are three subspecies of *Plasmodium yoelii* (*P. yoelii yoelii*, *P. yoelii nigeriensis*, and *P. yoelii killicki*) and a number of strains and clonal lines (all lines used in this study are cloned parasite lines) that exhibit wide variation in disease phenotypes, from mild disease to lethal infection [8,9]. For example, *P. y. yoelii* strain 17XNL (17XNL), the most common laboratory strain and the genome reference strain of *P. yoelii* [10], is a non-virulent parasite that causes a mild infection, whereas *P. y. yoelii* strain YM (YM) and *P. y. nigeriensis* N67C (N67C) parasites are highly virulent and generally produce lethal diseases [8,9,11]. Characterization of the genetic differences between lethal and non-lethal parasite lines may lead to discovery of molecular markers of severe malaria phenotypes. Indeed, recent mapping studies have linked parasite growth-related virulence and host cytokine/chemokine levels to various chromosomal loci [9,11,12]. However, these mapping studies largely relied on typing individual genetic markers such as microsatellites (MS) markers [9]. Although MS typing is flexible and reliable, genotyping of large numbers of progeny from genetic crosses using hundreds of MS markers can be labor intensive, time consuming, and expensive. The lack of genome-wide markers and genetic maps also impedes assembly of the parasite genome, leaving many unmapped contigs in the genome database and preventing thorough gene annotation and characterization. Thus, development of a genotyping tool that can screen a large number of single nucleotide polymorphisms (SNPs) and/or MS will be desirable for large-scale genetic studies.

Microarray-based hybridization has been widely employed to measure gene expression, to identify SNPs, and to genotype *Plasmodium* parasites [13]. In particular, a number of customized SNP-typing array platforms have been developed for genotyping the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax* [13–20]. These arrays can genotype large numbers of genetic markers simultaneously and have provided an unprecedented opportunity to investigate parasite population structure and recombination rate and to conduct genome-wide association studies (GWAS) to identify genes affecting important phenotypes [15,18,21–23]. Among the array platforms is the NimbleGen CGH array that uses a two-dye system to compare reference and test samples labeled with two

fluorescent dyes (i.e., Cy3 and Cy5). Nucleotide substitution(s) in target sequences can be detected due to reduction in hybridization efficiency to the microarray probe relative to a perfectly matched reference sample, creating a detectable signal in hybridization intensity. NimbleGen array formats also allow simultaneous hybridization of multiple samples (up to 12, 12 × 135 K probes) on a single slide, thus improving inter-array reproducibility and reducing cost. This microarray platform has been applied to detect copy number polymorphisms (CNPs) and SNPs in the *P. falciparum* parasite [16,23,24].

In the present study, we sequenced genomic DNA from *P. y. yoelii* 33X (33X) and *P. y. nigeriensis* N67 (N67), aligned the genome sequences with that of 17XNL in public databases [25], and identified hundreds of thousands of SNPs between the genomes. After applying various criteria on marker spacing and hybridization dynamics, we selected ~11,000 SNPs that were evenly distributed across the 14 chromosomes (~2800 were not mapped to chromosomes) and printed a microarray to detect these polymorphisms in *P. yoelii* parasites. We showed that the microarray could efficiently and accurately detect SNPs from DNA samples of different *P. yoelii* strains. This microarray platform will be useful for studying genome diversity and genetic recombination and for genetic mapping important traits of this malaria parasite.

2. Materials and methods

2.1. Preparation of malaria parasites

The cloned lines of rodent malaria parasites used in the present work were 17XNL, YM, 33X, N67 and N67C, and two progeny (F115#5 and F245#1) from a 17XNL × N67 cross [9]. The origin and maintenance of the parasites were previously described [9]. In brief, the parasite stocks were used to initiate blood-stage infections in female C57BL/6 inbred mice (*Mus musculus*), 6–8 weeks old at the time of primary infection. All experiments using the laboratory mice in this study were conducted in accordance with the NIH-approved animal study protocol LMVR-11E.

Blood-stage parasites were collected from tail blood of donor mice and suspended in heparinized phosphate buffered saline (pH 7.2) to prepare an inoculum containing 1×10^6 infected red blood cells. The inoculum was then injected intraperitoneally into a group of 2–5 mice to generate sufficient blood-stage parasites for genomic DNA preparation. Infected mice were housed in polypropylene cages with sawdust bedding, maintained at a constant temperature ($25 \pm 1^\circ\text{C}$), on 12:12 h (light:dark) cycle and supplied with sterile food and water according to the NIH guidelines. Blood-stage infections were monitored daily by microscopic examination of Giemsa-stained thin blood smears. Blood-stage parasites were harvested when parasitemia reached >10%. To remove host white blood cells, mouse whole blood was filtered through Plasmodipur filters (EuroProxima, Arnhem, The Netherlands) three times, and the parasites were released from infected red blood cells after saponin treatment and washed in DNase-free PBS. Parasite pellets were stored at -80°C prior to genomic DNA extraction.

2.2. Extraction of genomic DNA for illumina sequencing

Parasite pellets were resuspended in 400 µl of lysis solution (10 mM Tris-HCl, pH 8.0; 50 mM EDTA; 0.1% SDS; 1 mg ml⁻¹ proteinase K) and incubated at 42 °C overnight. After phenol/chloroform extraction, DNA was precipitated by addition of chilled absolute ethanol and 3 M sodium acetate pH 5.2 (Sigma Chemicals, St. Louis, MO). DNA pellets were re-suspended, RNase-treated, and adjusted to a final concentration of 100 ng/µl in TE buffer (10 mM Tris HCl and 0.01 mM EDTA, pH 8.0). Concentrations of the genomic DNA samples were measured using a Hoefer™ DyNA Quant 200 fluorometer, and absence of RNA contamination was confirmed using agarose gel electrophoresis.

Whole genome small insert libraries for Illumina sequencing were prepared as follows. Five micrograms of genomic DNA was sheared to ~450 bp using a Covaris S2 with settings: duty cycle 20%; intensity 4; cycles 200. The library was constructed according to the Illumina protocol in the Paired-End DNA Sample Prep Oligo Only Kit. All enzymes were purchased from New England Biolabs, with the exception of Platinum Pfx (from Invitrogen, CA). After paired-end adapter ligation, the sample was size selected on a 2% agarose gel stained with SYBRGold and visualized on a Dark Reader. To insure that the library was not over amplified, a test amplification was performed where aliquots of a PCR reaction were removed every 2 cycles from 4 to 16. These aliquots were evaluated on a 2% agarose gel and an optimal cycle number was selected for subsequent large-scale amplification. Amplification reactions were cleaned up using two rounds of Agencourt AMPure Beads. Libraries were quantitated by qPCR using Applied Biosystems Power SYBR Green PCR Master Mix and sequenced on an Illumina GA_{ii}x using ver 4 chemistry to obtain 100-bp paired end reads. Data was processed using RTA1.6.32.0 and GERALD 1.15.

2.3. Sequence reads alignment and data analysis

Reads from two lanes of the same strain were pooled and mapped to reference genome 17XNL (build 2002.09.10) using the Burrows-Wheeler aligner (BWA) software (v0.5.4) with default parameters [26]. Alignments from paired-end reads were utilized to generate SNP calls. We first removed redundant reads if they were mapped to a location with the same starting position. Any mismatches at a base from reads with Phred quality score less than 20 were ignored. The numbers of coverage and SNPs were counted based on chromosome location from the alignment. Initial SNPs were recorded if there were more than two indications of reads with alternative alleles (different from the reference 17XNL) from both forward and reverse strands and the ratio of reads with alternative allele to total coverage was greater than 75%. Additional criteria were applied to select for the final list of SNPs for printing on the microarray chip (see below).

2.4. Microarray probe selection and design

Several filtering steps were taken to select probe sets for array printing. First, to ensure that the SNP calls were reliable, we selected SNPs with at least 50 read coverage (53 to 5695 reads for each site) and 100% of the reads calling a SNP between 17XNL and N67 parasites. Second, we downloaded 100-bp sequences with the polymorphic sites at the middle, blasted the sequences against the available 17XNL genome sequences, and picked those that had a single match. Third, we picked only one SNP with the highest T_m if there were more than

two SNPs within a contig or DNA segment equal to or smaller than 2 kb. If multiple SNPs were present, we selected the one that also had a SNP in the 33X sequences and had the highest T_m . At the time of submitting our final probe sets for array printing, we also had access to the assembled YM genome sequence (<http://www.sanger.ac.uk/resources/downloads/protozoa/plasmodium-yoelii.html>). We blasted the probe sequences against the YM genome sequence and removed additional probes that had more than one hit. Considering that the *P. yoelii* genome was not completely assembled—and to provide potential SNPs in the “gap” regions—we included 1822 probe sequences that did not have a hit against the assembled YM sequence. We also repeated 428 probe set sequences (in triplicate) as within slide hybridization controls. Probe sequences of 29–41 bp long with similar melting temperature and hybridization dynamics (T_m of ~64–68 °C) and with polymorphic sites at the middle of the probes were selected using a custom script described previously (Table S1) [16]. Each polymorphic site was covered by eight probes; four probes on the plus strand had identical sequences except the polymorphic site, which had either G, A, T, or C in each. Another set of four probes on the minus strand had the same design.

Supplementary Table S1 related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.molbiopara.2014.03.006>.

2.5. Microarray hybridization

DNA samples for array hybridization were prepared similarly to above except only one Plasmodipur filter was used to remove host white cells. Labeling and hybridization were conducted essentially as described in the standard NimbleGen CGH procedures [16,27]. Briefly, 1 µg of genomic DNA was denatured at 98 °C for 10 min in the presence of 1 OD Cy3-labeled 65% AT random nonamer (TriLink Biotechnologies, San Diego, CA). The denatured sample was snap-chilled on ice and incubated with 50 units of Klenow fragment (New England Biolabs, Ipswich, MA) and dNTP mix (6 mM each in TE; Sigma Aldrich, St. Louis, MO) for 2 h at 37 °C. Reactions were terminated with 0.5 M EDTA and precipitated with 5 M NaCl in isopropanol. Labeled product was resuspended in water and 10 µg of labeled product dried down for resuspension in 44 µl of 2.54× Denhardt’s solution (Sigma–Aldrich), 45.67 µl 2× hybridization buffer, 18.27 µl hybridization component A, 1.88 µl alignment oligo, and 2.2 µl sample tracking control (Roche NimbleGen, Inc., Madison, WI). The sample was denatured at 95 °C for 5 min and allowed to hybridize on the array for 20 h at 42 °C in a NimbleGen hybridization system. Arrays were washed sequentially in wash buffer I (Roche NimbleGen, Inc.) for 2 min at room temperature (RT), wash buffer II for 1 min at RT, and wash buffer III for 15 s at RT, dried for 1 min in a microarray high-speed centrifuge (Arrayit Corp., Sunnyvale, CA), and were scanned at 2-µm resolution using a NimbleGen MS 200 microarray scanner (Roche NimbleGen, Inc.).

2.6. Base call after microarray hybridization

Scanned images were gridded and processed using NimbleGen v2.5 (Roche NimbleGen, Inc.) to extract signal data. SNP genotypes were called when the probe sets at each locus indicated complimentary SNP genotypes on the two strands. The accuracy of base calls was calculated as the percentage of base calls that matched the known sequences of N67 and 17XNL. Parental clones were hybridized in duplicate and compared to identify SNPs that

distinguished the two parental clones. Progeny clones were scored at each locus as inheriting one of the parental genotypes or as NA if a non-parental genotype was called.

3. Results and discussion

3.1. Genome sequencing and identification of SNPs

Development of a method that can be used to genotype large numbers of SNPs through a single hybridization will greatly facilitate genetic studies of the parasites. Unlike the human malaria parasites *P. falciparum* and *P. vivax* from which large numbers of SNPs have been identified, no systematic search for genome-wide SNPs has been conducted, and few SNPs between *P. yoelii* strains are known. To identify SNPs in the *P. yoelii* genome, we first sequenced the genomes of two *P. yoelii* parasites, 33X and N67, known to have genomes different from those of 17XNL and YM after typing with MS [9,28]. Genomic DNA from 33X and N67 were prepared and sent for commercial sequencing using an Illumina next-generation sequencer. A total of 88,205,714 reads from 33X and 56,011,125 reads from N67 parasites were obtained, providing ~202-fold and 367-fold coverage for 33X and N67, respectively (Table 1). Sequence reads from the parasites were aligned to the 17XNL draft genome sequences at the Sanger database (<http://www.sanger.ac.uk/resources/downloads/protozoa/plasmodium-yoelii.html>), covering over 95% of the available genome sequences. Approximately 49.4% of the reads from 33X and 76.8% from N67 were mapped to the 17XNL genome sequence. The sequence reads for N67 and 33X are available at the NCBI SRA database with accession numbers of SRR1037485 and SRR1187182, respectively.

Comparison of the 33X and N67 genome sequences with those of 17XNL identified 458,922 potential SNPs between parasites 17XNL/N67 and 74,906 SNPs between 17XNL/33X based on criteria of 10-read coverage and 75% reads calling for a SNP. If we assume that *P. yoelii* has a genome size of 23 Mb, there will be approximately one SNP per 50 bp DNA sequence between the 17XNL and N67 parasite genomes (or 2% differences) and one SNP per 300 bp between the 17XNL and 33X genomes (0.33% differences). There were also 58,058 SNPs shared by both 33X and N67 but different from the 17XNL parasite. The results show that N67 is very different from 17XNL and 33X; whereas the differences between 33X and 17XNL are relatively smaller, which is consistent with results from MS typing (Table S2) [9,28].

Supplementary Table S2 related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.molbiopara.2014.03.006>.

3.2. SNP selection and microarray probe set design

The first step in developing an array is to select a reliable set of SNPs. We therefore applied stringent criteria to ensure that the SNPs selected for array printing reflected true nucleotide substitutions and were not from sequencing errors or alignment mistakes. A total of 23,037 SNPs were initially selected from the 458,922 potential SNPs after applying the criteria described in Section 2. We designed all the probe sequences to be 29–41 bp long with melting temperature (T_m) centered at 66 °C (64–68 °C). After applying these criteria, we had final probe sets for 11,052 SNPs (Table S1). The probe set queried 9230 SNP loci matching a single site in the YM and 17XNL genome, 1822 loci with no hit (located at

current genome sequencing gaps), and 428 having triplicate probe sets. For each potential SNP, four probes were designed on each DNA strand, with each probe having one of the four potential nucleotides (G, A, T, or C). Therefore, each plex on the 12-plex microarray chip contained 95,264 experimental probes.

3.3. Evaluation of array call rate and accuracy

To evaluate the efficiency and accuracy of the microarray, we first hybridized genomic DNA from five strains of *P. yoelii* (17XNL, YM, N67, N67C, 33X) and two progeny from the N67 × 17XNL cross (F115#5 and F245#1) reported previously [9] to the array. The call rates for the hybridizations were generally 96–97% (Table 2). Except for the 1822 probes that could not be assigned to the assembled 14 chromosomes or contigs, allocation and distribution of the probes were proportional to chromosome sizes, ranging from 0.39 probe/kb on chromosome 2–0.47 probe/kb on chromosome 13 (average 0.46 probe/kb), suggesting relatively even distribution of the probes regardless of chromosome length (Table 2). No reliable probes could be designed at the subtelomeric regions due to repetitive sequences; in particular, a segment on chromosome 2 did not produce any signals from all the strains (Fig. 1). One potential explanation for the lack of probes on this end of chromosome 2 was deletion of this region, and there was no chromosomal sequence from this region for probe alignment. A deletion on one end of chromosome 2 of *P. falciparum* parasites is well recognized [29], although chromosome 2 of the two parasite species are not necessarily syntenic.

We also hybridized the same DNA samples from YM and N67 to the array twice and compared the SNP calls from the repeated hybridizations. Approximately 95% of the calls were the same from both hybridizations, and the majority of the no-match calls were “no calls” in one or both hybridizations; only on three occasions did the repeated hybridizations call different alleles in both YM-1/YM-2 and N67-1/N67-2 pairs of hybridizations (Table 3). These three differences between repeated hybridizations of the same parasites were found to be errors from one of the hybridizations after PCR amplification of the target sequences and sequencing verification (data not shown). If we exclude the no calls, the error rates are approximately 0.03%.

We also compared SNP calls between two parasite pairs: (i) YM and 17XNL and (ii) N67 and N67C. These two pairs of parasites were known to be closely related genetically, but differ in growth and virulent phenotypes [9,12]. Our results showed that hybridization patterns between the pair of 17XNL and YM and the pair of N67 and N67C were almost identical (Table 3 and Fig. 2). Similar to the repeated YM and N67 hybridizations, comparison of the calls from 17XNL/YM and N67/N67C pairs revealed matched call rates of 94–95%. Comparison of 17XNL with the calls from the two repeated YM hybridizations found only 56 and 53 probes calling different alleles, respectively. Similarly, only 21 and 22 probes called different nucleotides comparing the calls from N67C and the two hybridizations of N67 (Table 3). These results confirmed that YM/17XNL and N67/N67C are closely related parasites.

In the array, 428 SNPs had probe sets in triplicate. Comparison of the calls from these probes also showed highly agreeable results. Of the nine hybridizations (*five strains, two*

progeny, and *two* repeats of YM and N67), only one probe set at one SNP position called a different allele (Table 4). All these data showed that the array has a high call rate and accuracy.

3.4. Detection of genome-wide SNPs and genetic crossovers in progeny of a genetic cross

Comparison of hybridization results from DNA samples of N67, 33X, F115#5 and F245#1 with that of 17XNL detected large numbers of SNPs (Table 3). Comparison of 17XNL with those of 33X and N67 produced 4198 and 7740 SNPs, whereas progeny F115#5 and F245#1 had 2706 and 4640 SNPs, respectively. Interestingly, 33X appears to be a progeny of 17XNL (or YM) and an unknown parasite, because its chromosome 2 and segments of chromosomes 3, 5, 6, 7, 10, and 12 were the same as those of 17XNL or YM (Fig. 1).

Progeny F245#1 had more SNPs than F115#5 (comparing to reference 17XNL) because it inherited more chromosomes or chromosomal segments from N67 (Fig. 2). Hybridization of the two progeny DNA samples also revealed clear chromosome re-assortment and genetic recombination during fertilization and meiotic division. For the progeny F245#1, chromosomes 3, 5, 6, 7, 11, 12, and 14 were inherited from N67, chromosomes 1, 9, and 13 were from 17XNL, and chromosomes 2, 4, 8 and 10 were recombinant. For progeny F115#5, chromosomes 1, 3, 7, 10, and 12 were inherited from N67, chromosomes 4, 5, 13, and 14 were from 17XNL, and chromosomes 6, 8, 9, and 11 (double crossovers) were recombinant (Fig. 2). From the two progeny, we detected 16 crossovers. The genotypes across the 14 chromosomes identified by the array also matched perfectly with the results from the MS typing reported previously (Table S2) [9]. Because the 10,196 SNPs (minus the 2×428 repeated probes) could provide an average marker density of approximately one SNP per 2.2 kb (assuming a 23-Mb genome), genotyping using this array will allow mapping a gene to a relatively small chromosomal region if sufficient progeny from genetic crosses are obtained.

This work identified hundreds of thousands of SNPs from the *P. yoelii* genome and developed the first high-throughput, genome-wide method for efficiently typing the *P. yoelii* genome. In addition to confirming isogenic parasite strains (YM/17XNL and N67/N67C), we also showed several highly diverse *P. yoelii* genomes (17XNL, 33X, and N67), including up to ~2% differences in the nucleotide sequence between 17XNL and N67. Analysis and annotation of the genome sequences are ongoing to further characterize the parasite genomes.

The microarray we developed is highly reliable and has a call rate of 96–97%, which is equivalent to or better than other SNP arrays for malaria parasites [16,17,30,31]. Comparison of data from repeated hybridization of the same DNA samples also showed ~95% matched calls, with the majority of not-matched calls being no calls. In both YM and N67 repeated hybridizations, only three probe sets call a different allele in the repeated hybridizations. If we remove the no-calls, the correct call rate is approximately 99.97%, suggesting an accurate and reliable SNP calling method. Although the recent technical development and reduction in costs for next generation sequencing makes it attractive to perform genome-sequencing instead of microarray hybridization, we found that processing and accurately calling SNPs in the AT-rich malaria genomes is still challenging and time-

consuming. It can take months or even years to accurately verify many SNPs calls from genome sequencing, whereas results from our microarray calls can be available in the same day after signal scanning. Additionally, relatively small numbers of progeny from malaria genetic crosses (All the crosses have fewer than 50 progeny) will limit the numbers of recombination events; therefore, the SNPs in our array will provide sufficient resolution for detecting crossovers in the progeny of genetic crosses. Development of this microarray will greatly facilitate our efforts in mapping parasite determinants contributing to drug resistance, parasite development, and virulence. Hybridization of the additional progeny from genetic crosses will also allow development of a genetic map with high marker density that can help genome assembly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. We thank Brenda Rae Marshall, DPSS, NIAID, for editing. Because SCN, SP, MZ, JD, VN, WX, and X-zS are government employees and this is a government work, the work is in the public domain in the United States. Notwithstanding any other agreements, the NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. You can establish rights outside of the U.S. subject to a government use license.

References

1. Stephens R, Culleton RL, Lamb TJ. The contribution of *Plasmodium chabaudi* to our understanding of malaria. *Trends Parasitol.* 2012; 28:73–82. [PubMed: 22100995]
2. Carlton JM, Hayton K, Cravo PV, Walliker D. Of mice and malaria mutants: unravelling the genetics of drug resistance using rodent malaria models. *Trends Parasitol.* 2001; 17:236–42. [PubMed: 11323308]
3. Janse CJ, Kroeze H, van Wigcheren A, Mededovic S, Fonager J, Franke-Fayard B, et al. A genotype and phenotype database of genetically modified malaria-parasites. *Trends Parasitol.* 2011; 27:31–9. [PubMed: 20663715]
4. Li C, Seixas E, Langhorne J. Rodent malarias: the mouse as a model for understanding immune responses and pathology induced by the erythrocytic stages of the parasite. *Med Microbiol Immunol.* 2001; 189:115–26. [PubMed: 11388608]
5. Langhorne J, Quin SJ, Sanni LA. Mouse models of blood-stage malaria infections: immune responses and cytokines involved in protection and pathology. *Chem Immunol.* 2002; 80:204–28. [PubMed: 12058640]
6. Longley R, Smith C, Fortin A, Berghout J, McMorran B, Burgio G, et al. Host resistance to malaria: using mouse models to explore the host response. *Mamm Genome.* 2011; 22:32–42. [PubMed: 21116636]
7. Fortin A, Stevenson MM, Gros P. Complex genetic control of susceptibility to malaria in mice. *Genes Immun.* 2002; 3:177–86. [PubMed: 12058252]
8. Pattaradilokrat S, Cheesman SJ, Carter R. Congenicity and genetic polymorphism in cloned lines derived from a single isolate of a rodent malaria parasite. *Mol Biochem Parasitol.* 2008; 157:244–7. [PubMed: 18068827]
9. Li J, Pattaradilokrat S, Zhu F, Jiang H, Liu S, Hong L, et al. Linkage maps from multiple genetic crosses and loci linked to growth-related virulent phenotype in *Plasmodium yoelii*. *Proc Natl Acad Sci U S A.* 2011; 108:E374–82. [PubMed: 21690382]

10. Carlton JM, Carucci DJ. Rodent models of malaria in the genomics era. *Trends Parasitol.* 2002; 18:100–2. [PubMed: 11854077]
11. Pattaradilokrat, S.; Li, J.; Wu, J.; Qi, Y.; Eastman, RT.; Zilversmit, M., et al. Plasmodium genetic loci linked to host cytokine and chemokine responses. *Genes Immun.* 2014. <http://dx.doi.org/10.1038/gene.2013.74> [Epub ahead of print]
12. Pattaradilokrat S, Culleton RL, Cheesman SJ, Carter R. Gene encoding erythrocyte binding ligand linked to blood stage multiplication rate phenotype in *Plasmodium yoelii yoelii*. *Proc Natl Acad Sci U S A.* 2009; 106:7161–6. [PubMed: 19359470]
13. Pattaradilokrat, S.; Mu, J.; Awadalla, P.; Su, X-z. Genome diversity and applications in genetic studies of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. In: Carlton, J.; Deitsch, K.; Perkins, S., editors. *Comparative Genomics, Evolution and Molecular Biology*. Norwich, UK: Caister Academic Press; 2013. p. 59-90.
14. Kidgell C, Volkman SK, Daily J, Borevitz JO, Plouffe D, Zhou Y, et al. A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathog.* 2006; 2:e57. [PubMed: 16789840]
15. Mu J, Myers RA, Jiang H, Liu S, Ricklefs S, Waisberg M, et al. Plasmodium falciparum genome-wide scans for positive selection, recombination hot spots and resistance to antimalarial drugs. *Nat Genet.* 2010; 42:268–71. [PubMed: 20101240]
16. Tan JC, Miller BA, Tan A, Patel JJ, Cheeseman IH, Anderson TJ, et al. An optimized microarray platform for assaying genomic variation in *Plasmodium falciparum* field populations. *Genome Biol.* 2011; 12:R35. [PubMed: 21477297]
17. Neafsey DE, Schaffner SF, Volkman SK, Park D, Montgomery P, Milner DA Jr, et al. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* 2008; 9:R171. [PubMed: 19077304]
18. Van Tyne D, Park DJ, Schaffner SF, Neafsey DE, Angelino E, Cortese JF, et al. Identification and functional validation of the novel antimalarial resistance locus PF10 0355 in *Plasmodium falciparum*. *PLoS Genet.* 2011; 7:e1001383. [PubMed: 21533027]
19. Jiang H, Li N, Gopalan V, Zilversmit MM, Varma S, Nagarajan V, et al. High recombination rates and hotspots in a Plasmodium falciparum genetic cross. *Genome Biol.* 2011; 12:R33. [PubMed: 21463505]
20. Dharia NV, Bright AT, Westenberger SJ, Barnes SW, Batalov S, Kuhen K, et al. Whole-genome sequencing and microarray analysis of ex vivo *Plasmodium vivax* reveal selective pressure on putative drug resistance genes. *Proc Natl Acad Sci U S A.* 2010; 107:20045–50. [PubMed: 21037109]
21. Yuan J, Cheng KC, Johnson RL, Huang R, Pattaradilokrat S, Liu A, et al. Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science.* 2011; 333:724–9. [PubMed: 21817045]
22. Takala-Harrison S, Clark TG, Jacob CG, Cummings MP, Miotto O, Dondorp AM, et al. Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc Natl Acad Sci U S A.* 2013; 110:240–5. [PubMed: 23248304]
23. Cheeseman IH, Miller BA, Nair S, Nkhoma S, Tan A, Tan JC, et al. A major genome region underlying artemisinin resistance in malaria. *Science.* 2012; 336:79–82. [PubMed: 22491853]
24. Tan JC, Patel JJ, Tan A, Blain JC, Albert TJ, Lobo NF, et al. Optimizing comparative genomic hybridization probes for genotyping and SNP detection in *Plasmodium falciparum*. *Genomics.* 2009; 93:543–50. [PubMed: 19285129]
25. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Perteu M, Silva JC, et al. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature.* 2002; 419:512–9. [PubMed: 12368865]
26. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25:1754–60. [PubMed: 19451168]
27. Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, et al. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer.* 2005; 44:305–19. [PubMed: 16075461]

28. Li J, Zhang Y, Sullivan M, Hong L, Huang L, Lu F, et al. Typing *Plasmodium yoelii* microsatellites using a simple and affordable fluorescent labeling method. *Mol Biochem Parasitol.* 2007; 155:94–102. [PubMed: 17658627]
29. Biggs BA, Kemp DJ, Brown GV. Subtelomeric chromosome deletions in field isolates of *Plasmodium falciparum* and their relationship to loss of cytoadherence in vitro. *Proc Natl Acad Sci U S A.* 1989; 86:2428–32. [PubMed: 2648403]
30. Jiang H, Yi M, Mu J, Zhang L, Ivens A, Klimczak LJ, et al. Detection of genome wide polymorphisms in the AT rich *Plasmodium falciparum* genome using a high density microarray. *BMC Genomics.* 2008; 9:398. [PubMed: 18724869]
31. Dharia NV, Sidhu AB, Cassera MB, Westenberger SJ, Bopp SE, Eastman RT, et al. Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biol.* 2009; 10:R21. [PubMed: 19216790]

Fig. 1.

Genotypes of the *Plasmodium yoelii yoelii* 33X parasite from array hybridization.

Genotypes and inheritances of the parasite's 14 chromosomes are as marked. The segments with pure green color (■) suggest DNA inherited from 17XNL/YM, whereas the other regions are mixed with red and green markers (■ ■), suggesting 17XNL/YM and N67 genotypes, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 2.

Array hybridization signals from the 14 chromosomes of four *Plasmodium yoelii* strains and two progeny of a genetic cross. N67 and N67C, *Plasmodium yoelii nigeriensis* N67 and N67C; 17XNL and YM, *Plasmodium yoelii yoelii* 17XNL and YM; F115#5 and F245#1 are two progeny from the 17XNL × N67 cross. Alleles from N67 are labeled in red (■), and those from 17XNL are in green (■). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1

Summary of sequence reads and estimated coverages of *Plasmodium yoelii yoelii* 33X and *Plasmodium yoelii nigeriensis* N67.

Strain	33X	N67
Total reads	88,205,714	56,011,125
One read aligned	43,595,422	43,027,899
Both reads aligned	3,459,693	5,630,956
Total bp covered by mappable reads	5,101,995,608	9,260,362,154
Depth of coverage (fold)	202	367
% Genome coverage	97.02	95.03

Note: Both reads aligned, both reads from the paired ends were aligned; % Genome coverage is the percentage of the genome covered by at least one read.

Table 2

Estimates of array call rates from *Plasmodium yoelii* strains and two recombinant progeny of the N67 × 17XNL cross.

Chr ^d	No. probes ^b	Est. Chr size ^c	Probes/Kb	YM-1 ^d	YM-2 ^d	17XNL ^d	N67C ^e	N67-1 ^e	N67-2 ^e	33X ^d	F115#5 ^f	F245#1 ^f
1	287	650,457	0.44	275	272	254	282	279	277	275	279	276
2	327	833,371	0.39	316	318	297	313	313	311	318	316	315
3	334	759,305	0.44	324	324	306	325	321	321	328	324	325
4	390	914,108	0.43	378	382	381	386	376	376	380	381	382
5	485	1,063,341	0.46	469	455	472	476	466	459	466	473	462
6	480	1,062,600	0.45	464	465	469	470	460	458	465	471	463
7	394	946,893	0.42	372	374	378	384	379	375	382	382	382
8	698	1,589,685	0.44	683	682	685	679	672	672	681	689	679
9	830	1,812,809	0.46	800	802	806	819	808	803	802	813	809
10	799	1,790,134	0.45	772	779	785	780	783	774	778	784	784
11	858	1,921,449	0.45	836	836	839	843	839	827	837	841	838
12	855	1,913,402	0.45	830	828	837	846	835	833	835	842	842
13	1261	2,683,165	0.47	1219	1226	1239	1239	1228	1228	1234	1239	1226
14	1231	2,657,773	0.46	1199	1199	1203	1207	1189	1189	1203	1204	1200
Unassigned	1823			1689	1681	1718	1720	1697	1688	1687	1716	1710
Total	11,052			10,626	10,623	10,669	10,769	10,645	10,591	10,671	10,754	10,693
Call rate				0.96	0.96	0.97	0.97	0.96	0.96	0.97	0.97	0.97

^aChr, chromosome.

^bNo. probes, numbers of probes on each chromosome.

^cEst. Chr size, estimated chromosome sizes based on August 2012 release of the YM genome sequence.

^dYM-1, YM-2, 17XNL, and 33X are *Plasmodium yoelii yoelii*.

^eN67C, N67-1, and N67-2 are *Plasmodium yoelii nigeriensis*.

^fF115#5 and F245#1 are progeny of the N67 × 17XNL genetic cross described in Li *et al.* [9].

Table 3
Matched calls and SNPs detected from repeated hybridizations or parasites with similar genomes.

Parasite pair	Matched ^a	Not matched ^b	% Matched	NA calls ^c	SNPs
YM-1/YM-2	10,500	552	95.0	549	3
17XNL/YM-1	10,415	637	94.2	581	56
17XNL/YM-2	10,435	617	94.4	564	53
N67-1/N67-2	10,452	600	94.6	597	3
N67C/N67-1	10,528	524	95.3	502	22
N67C/N67-2	10,483	569	94.9	548	21
N67/17XNL	2638	8414	23.9	674	7740
33X/17XNL	6384	4668	57.7	393	4198
F115#5/17XNL	7856	3196	71.1	490	2706
F245#1/17XNL	5825	5227	47.3	587	4640

^a Matched, total probes excluding NA calls (no calls) and SNP.
^b Not matched, NA calls plus SNPs.
^c NA calls, one or both parasites had NA calls.

Table 4

Comparison of calls from 428 triplicate probes.

Parasite	All matched ^a	Not matched	% Matched	NA calls ^b	SNPs/errors ^c
YM-1	414	14	96.7	14	0
YM-2	415	13	96.7	13	0
N67-1	415	13	96.7	13	0
N67-2	408	20	95.3	20	0
N67C	421	7	98.4	7	0
17XNL	424	4	99.1	3	1
F115#5	419	9	97.7	9	0
F245#1	416	12	97.2	12	0
33X	408	20	95.3	20	0

^a All matched, probes where all three triplicates called the same SNP genotype or all indicated an "NA" genotype.

^b NA calls, one or two of the probe replicates had NA calls but all available genotypes match.

^c SNPs/errors, available genotypes did not match.