The *Plasmodium berghei* RC strain is highly diverged and harbors putatively novel drug resistance variants

Warangkhana Songsungthong [1,2] , **Supasak Kulawonganunchai** [3] , **Alisa Wilantho** [3] , **Sissades Tongsima** [3] ,
Pongpisid Koonyosying [1] , **Chairat Uthaipibull** [1] , **Sumalee Kamchonwongpaisan** [Corresp., 1] , **Philip J Shaw** [Corresp. 1]

[1] Protein-Ligand and Molecular Biology Laboratory, Medical Molecular Biology Research Unit, National Center for Genetic Engineering and Biotechnology
(BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand

[2] Biomolecular Analysis and Application Laboratory, Biosensing Technology Research Unit, National Center for Genetic Engineering and Biotechnology
(BIOTEC), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand

[3] Biostatistics and Bioinformatics Laboratory, Genome Technology Research Unit, National Center for Genetic Engineering and Biotechnology (BIOTEC),
National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand

Corresponding Authors: Sumalee Kamchonwongpaisan, Philip J Shaw
Email address: sumaleek@biotec.or.th, philip@biotec.or.th

Background. The current first line drugs for treating uncomplicated malaria are artemisinin combination therapies. However, *Plasmodium falciparum* parasites resistant to artemisinin and partner drugs are spreading, which threatens malaria control efforts. Rodent malaria species are useful models for understanding antimalarial resistance, in particular genetic variants responsible for cross resistance to different compounds.

Methods. The *Plasmodium berghei* RC strain (*Pb*RC) is described as resistant to different antimalarials, including chloroquine and artemisinin. In an attempt to identify the genetic basis for the antimalarial resistance trait in *Pb*RC, its genome was sequenced and compared with five other previously sequenced *P. berghei* strains.

Results. We found that *Pb*RC is eight-fold less sensitive to the artemisinin derivative artesunate than the reference strain *Pb*ANKA. The genome of *Pb*RC is markedly different from other strains, and 6974 single nucleotide variants private to *Pb*RC were identified. Among these *Pb*RC private variants, non-synonymous changes were identified in genes known to modulate antimalarial sensitivity in rodent malaria species, including notably the ubiquitin carboxyl-terminal hydrolase 1 gene. However, no variants were found in some genes with strong evidence of association with artemisinin resistance in *P. falciparum* such as K13 propeller protein.

Discussion. The variants identified in *Pb*RC provide insight into *P. berghei* genome diversity and genetic factors that could modulate chloroquine and artemisinin resistance in *Plasmodium* spp.

1 **The *Plasmodium berghei* RC strain is highly diverged and harbors putatively novel drug**

2 **resistance variants**

3 Warangkhana Songsungthong[1, 2], Supasak Kulawonganunchai[3], Alisa Wilantho[3], Sissades

4 Tongsima[3], Pongpisid Koonyosying[1], Chairat Uthaipibull[1], Sumalee Kamchonwongpaisan[1] and

5 Philip J. Shaw[1]

7 [1]Protein-Ligand and Molecular Biology Laboratory, Medical Molecular Biology Research Unit,

8 National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and

9 Technology Development Agency (NSTDA), Pathum Thani, Thailand

10 [2]Biomolecular Analysis and Application Laboratory, Biosensing Technology Research Unit,

11 National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and

12 Technology Development Agency (NSTDA), Pathum Thani, Thailand

13 [3]Biostatistics and Bioinformatics Laboratory, Genome Technology Research Unit, National

14 Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and

15 Technology Development Agency (NSTDA), Pathum Thani, Thailand

17 Corresponding author:

18 Philip J. Shaw, National Center for Genetic Engineering and Biotechnology (BIOTEC),

19 National Science and Technology Development Agency (NSTDA), 113 Thailand Science Park,

20 Khlong Nueng, Khlong Luang, Pathum Thani 12120, Thailand

21 E-mail: philip@biotec.or.th

23   **Abstract**

24   **Background.**

25   The current first line drugs for treating uncomplicated malaria are artemisinin

26   combination therapies. However, *Plasmodium falciparum* parasites resistant to artemisinin and

27   partner drugs are spreading, which threatens malaria control efforts. Rodent malaria species are

28   useful models for understanding antimalarial resistance, in particular genetic variants responsible

29   for cross resistance to different compounds.

30   **Methods.**

31   The *Plasmodium berghei* RC strain (*Pb*RC) is described as resistant to different

32   antimalarials, including chloroquine and artemisinin. In an attempt to identify the genetic basis

33   for the antimalarial resistance trait in *Pb*RC, its genome was sequenced and compared with five

34   other previously sequenced *P. berghei* strains.

35   **Results.**

36   We found that *Pb*RC is eight-fold less sensitive to the artemisinin derivative artesunate

37   than the reference strain *Pb*ANKA. The genome of *Pb*RC is markedly different from other

38   strains, and 6974 single nucleotide variants private to *Pb*RC were identified. Among these *Pb*RC

39   private variants, non-synonymous changes were identified in genes known to modulate

40   antimalarial sensitivity in rodent malaria species, including notably the ubiquitin carboxyl-

41   terminal hydrolase 1 gene. However, no variants were found in some genes with strong evidence

42   of association with artemisinin resistance in *P. falciparum* such as K13 propeller protein.

43   **Discussion.**

44   The variants identified in *Pb*RC provide insight into *P. berghei* genome diversity and

45   genetic factors that could modulate chloroquine and artemisinin resistance in *Plasmodium* spp.

46

47

48

49

50   **Introduction**

51   The incidence of malaria is declining around the world, and efforts are being directed

52   towards elimination of this disease in many endemic areas (Tanner et al., 2015); (WHO Malaria

53   Policy Advisory Committee and Secretariat, 2015). However, malaria parasite resistance to first-

54   line artemisinin combination therapy is evolving in Southeast Asia (Woodrow & White, 2016),

55   including resistance to partner drugs such as piperaquine (Duru, Witkowski & Ménard, 2016);

56   (Amato et al., 2016); (Imwong et al., 2017). This is alarming given that parasites resistant to

57   chloroquine (CQ) (Wootton et al., 2002) and pyrimethamine (Nair et al., 2003) are highly

58   prevalent in this region. The specter of multi-drug resistant *Plasmodium falciparum* parasites

59   could undermine all recent advances in reducing the disease burden. Laboratory models of

60   antimalarial resistance are needed to develop new drugs effective against parasites resistant to

61   currently available antimalarials, and to understand the molecular mechanisms of resistance.

62   Rodent malaria parasites are widely used laboratory models for human malaria as they

63   can be studied *in vivo* in animal and mosquito hosts. Antimalarial-resistant parasites can be

64   selected by repeated dosing of infected animals, and stably resistant parasite clones can be

65   isolated after serial passage in animal hosts. Moreover, parasites cross-resistant to different drugs

66   have been obtained by this approach. CQ and artemisinin (ART) resistant *Plasmodium chabaudi*

67   rodent malaria parasites from laboratory selection were isolated in (Afonso et al., 2006).

68   Laboratory selected CQ and ART-resistant *Plasmodium yoelii* rodent malaria parasites display

69   impaired hemozoin production and elevated level of glutathione (GSH), suggestive of a common

70   mechanism of resistance against the two drugs (Witkowski et al., 2012). Laboratory selection of

71   antimalarial resistance has also been performed for the most virulent rodent malaria species

72   *Plasmodium berghei*. The *P. berghei* RC strain (*Pb*RC) was obtained by laboratory selection

73   with CQ, and is defective for production of hemozoin (Peters, 1964); (Peters, Fletcher &

74   Staeubli, 1965). This strain is also reported as resistant to other drugs, including ART (Pérez-

75   Rosado et al., 2002). Given that *Pb*RC also shows an elevated level of GSH (Vega-Rodríguez et

76   al., 2015), similar to cross-resistant *P. yoelii* (Witkowski et al., 2012), it may harbor resistance

77   mutations in the same genes as other drug-resistant rodent malaria parasites.

78   In this study, we found that *Pb*RC is resistant to artesunate, a water-soluble derivative of

79   ART. We sought the genetic factors responsible for the cross-drug resistant phenotype of *Pb*RC

80   by performing whole genome sequencing. The *Pb*RC genome is markedly different from other

81   characterized *P. berghei* strains, and we identified several variants unique to *Pb*RC in genes

82   which may modulate antimalarial resistance.

83

Materials & Methods

85   Six to ten-week old female BALB/c mice were purchased from the National Laboratory

86   Animal Center, Mahidol University, Thailand. The following parasites were obtained from the

87   Malaria Research and Reference Reagent Resource Center (MR4; http://www.beiresources.org),

88   a part of BEI Resources, NIAID, NIH: *Plasmodium berghei* RC, MRA-404, deposited by W

89   Peters and BL Robinson; *Plasmodium berghei* ANKA 507m6cl1, MRA-867 deposited by CJ

90   Janse and AP Waters. Artesunate was a gift from Dafra Pharma, Belgium. CF11 was purchased

91   from Whatman. Other reagents, unless otherwise noted, were purchased from Sigma Aldrich.

92

Four-day suppressive test for *in vivo* drug sensitivity

94   BALB/c mice were injected intravenously with 1 x 10^7 *P. berghei* infected red blood

95   cells. For *Pb*RC, five to six mice were used per group. For *Pb*ANKA, four to eight mice were

96   used per group. Oral doses of artesunate were given at 4, 24, 48 and 72 h post infection. Four

97   days post infection, parasitemia was determined by manually counting infected and uninfected

98   red blood cells in Giemsa-stained thin blood smears. Percent inhibition was calculated using the

99   following formula:

100   Percent inhibition = 100 – ((100 x parasitemia of each dose) / parasitemia of untreated control)

101   Data of percent inhibition at different doses of artesunate were fitted to the two-parameter

102   sigmoidal dose response equation using the drc package in R (Ritz & Streibig, 2005). This study

103   was carried out in accordance with the guidelines in the Guide for the Care and Use of

104   Laboratory Animals of the National Research Council, Thailand. All animal experiments were

105   performed with the approval of BIOTEC's Institutional Animal Care and Use Committee (Permit

106   number BT-Animal 02/2557). At the end of the experiments, mice were euthanized by CO_2

107   asphyxiation. All efforts were made to alleviate pain and suffering.

108

Sanger dideoxy sequencing of PCR products from selected regions

109

110 Selected genomic regions with candidate variants in the *P. berghei* RC strain were PCR-

111 amplified using high-fidelity Phusion DNA polymerase (New England Biolabs) using primers

112 listed in Supplemental Table S1. The resulting PCR products were sent for Sanger dideoxy

113 sequencing (1st BASE, Malaysia).

114

115 **Whole genome DNA sequencing**

116 Parasitized blood obtained from a single mouse infected with the *Pb*RC parasite was

117 passed through a CF11 (Whatman) column to remove white blood cells. Cells were harvested by

118 centrifugation and parasites were liberated from red cells by lysis of the red blood cell membrane

119 with 0.2% saponin. Parasites were washed twice with phosphate buffered saline (137 mM NaCl,

120 2.7 mM KCl, 4.3 mM Na$_2$HPO$_4$, 1.47 mM KH$_2$PO$_4$, pH 7.4), resuspended in lysis buffer (8.5

121 mM Tris-HCl, 342 mM NaCl, 2 mM Na$_2$-EDTA, 0.14 mg/ml proteinase K, 0.14% SDS, pH 8.2)

122 and incubated at 37°C overnight. Proteins were precipitated by the addition of NaCl to 1.5 M,

123 centrifuged, and discarded. Isopropanol was added to the supernatant to precipitate nucleic

124 acids. The nucleic acid pellet was dried and resuspended in TE buffer (10 mM Tris, 1 mM

125 EDTA pH 8.0). RNA was removed by digestion with RNAseA. Phenol chloroform extraction

126 and ethanol precipitation of DNA were performed. The integrity of genomic DNA was checked

127 by agarose gel electrophoresis. Genomic DNA was submitted to the Chulalongkorn Medical

128 Research Center, Bangkok, Thailand for genome sequencing. Genomic DNA was sheared by

129 sonication and DNA fragments < 1 kb were gel-purified. Sequencing libraries were constructed

130 from sheared genomic DNA using a TruSeq kit (Illumina). 2x 150 bp reads were obtained using

131 a MiSeq instrument (Illumina). The raw data are deposited in the NCBI Sequence Read Archive,

132 accession number PRJNA277169.

133

134 **Sequence Data Analysis**

135   Raw sequencing data were obtained in FASTQ format. Genome sequence data for strains

136   K173, NK65NY, NK65E, SP11_RLL, SP11_A (Otto et al., 2014) were downloaded from the

137   European Nucleotide Archive. Filtering of raw data to remove poor quality reads (average Q-

138   score<30) and removal of adapter sequences were performed using FASTQC (Andrews, 2010).

139   Preprocessed read data were aligned to the *P. berghei* ANKA version 3 reference genome

140   (downloaded from GeneDB; (Logan-Klumpler et al., 2012)) using BOWTIE version 2.2.2.6

141   software (Langmead et al., 2009) under the setting of length of seed substring 22 bases without

142   clipping. GATK software version 3.3.0 (McKenna et al., 2010) was used to identify potential

143   base-substitution single nucleotide variants (SNV) and small insertion/deletions (INDEL)

144   between all strains different from the reference strain by the HaplotypeCaller method combined

145   with the GenotypeGVCFs method, under the settings: --minReadPerAlignmentStart (10), --

146   min_base_quality_score (10), --sample_ploidy(2), --heterozygosity(0.001). SnpEff software

147   version 4.3g (Cingolani et al., 2012) was used to annotate the variants using the genome

148   annotation file downloaded from GeneDB. The .vcf file containing all raw variants called by

149   GATK is provided in Data S1. The raw variants were filtered to retain only high-confidence

150   SNVs for genome analyses. INDEL variants were not included, since many occur within simple

151   sequence repeats that could be prone to PCR and sequencing artifacts. Potential false positive

152   SNVs resulting from sequencing artifacts were removed by excluding variants with heterozygous

153   or missing genotypic calls in any strain. Potential false variants resulting from read alignment

154   error in repetitive regions were removed using the DustMasker program (Morgulis et al., 2006).

155   Variants present among multigene families listed in Additional File 4 of (Otto et al., 2014) were

156   also treated as potential false positives and were removed. After filtering, 8681 SNV markers

157   remained.

158   Principal Components Analysis (PCA) was performed using the ipPCA tool (Limpiti et

159   al., 2011), implemented in MATLAB version R2009b. The 8681 filtered SNV markers were

160   used as input for ipPCA. Tajima's D scores were calculated using PopGenome, a population

161   genomic analysis tool in R (Pfeifer et al., 2014). Scores were calculated in sliding non-

162   overlapping genomic windows of five consecutive variants (1730 windows in total), and

163   separately for annotated genes (1765 genes in total, Supplemental Table S2). Plots of Tajima's D

164   scores were made using the ggplot2 package in R (Wickham, 2009). Gene ontology analysis of

165   genes with Tajima D scores greater than 1 was performed using the Gene ontology web service

166 provided in the PlasmoDB website (Aurrecoechea et al., 2009). Terms were considered

167 significant using a Bonferroni-corrected p-value threshold of 0.05.

168 Copy number variants (CNVs) were identified from the whole-genome sequencing data

169 using the cn.MOPS tool implemented in R (Klambauer et al., 2012). The sequence data for all

170 six strains aligned to the reference genome were used as input for cn.MOPS, which was run

171 using the haploid genome setting and window size of 500 bp used for local modelling of read

172 counts. Mapped reads in genomic regions with putative variants were visualized using the

173 Integrative Genomics Viewer program (Thorvaldsdottir, Robinson & Mesirov, 2013).

174

175 **Results**

176 The *Pb*RC parasite is reported as CQ and ART resistant (Pérez-Rosado et al., 2002). We

177 tested the sensitivity of the *Pb*RC strain to artesunate, a water soluble derivative of ART. *Pb*RC

178 is approximately eight-fold less sensitive to artesunate compared with *Pb*ANKA (effective dose

179 for 50% inhibition of parasite (ED_{50}) of 2.8 (s.e. = 0.4) and 23.1 (s.e. = 7.3) mg/kg in *Pb*ANKA

180 and *Pb*RC, respectively; Fig. 1), which confirms the multi-drug resistant phenotype for this

181 strain.

182

183 ***Pb*RC is widely diverged from other *P. berghei* strains**

184 Illumina sequencing of *Pb*RC genomic DNA was performed. Sequence reads were

185 obtained from 17,801,263 clusters, and 80.5% of the preprocessed reads could be mapped to the

186 *Pb*ANKA v3 reference genome. Genomic sequence data of previously sequenced strains K173,

187 NK65NY, NK65E, SP11_RLL and SP11_A (Otto et al., 2014) were aligned to the reference

188 genome using the same analytical procedure. SNV and small INDEL variants were called from

189 the mapped reads for all strains. A total of 27,495 variant markers were identified among the six

190 strains (Data S1), of which 8681 SNV markers remained after applying stringent filtering

191 criteria. The numbers of variants identified in each strain are shown in Table 1.

192 The majority of filtered SNVs are private to each strain, and the *Pb*RC strain shows the

193 highest number of variants, indicating that it is markedly different from the other strains. This is

194 confirmed by Principal Components Analysis in which *Pb*RC is clearly separated from the other

195   strains in the first Principal Component, which captures most of the genotypic variance among

196   these strains (Fig. 2).

197

Genetic diversity across the *P. berghei* genome

199   A previous analysis of *P. berghei* genotypic diversity revealed a low SNV frequency

200   compared with *P. chabaudi*, another rodent malaria species (Otto et al., 2014). However, the

201   extreme divergence of *Pb*RC from other strains may provide further insight into *P. berghei*

202   intraspecific genetic diversity. The Tajima's D score was calculated for sliding genomic

203   windows along all chromosomes, and also for annotated genes. Tajima's D is a measure of the

204   observed versus expected genetic diversity, calculated as the difference between the mean

205   number of pairwise differences and the number segregating sites (Tajima, 1989). The mean

206   Tajima's D is negative for each chromosome, indicating a genome-wide pattern of negative

207   Tajima's D (Fig. 3A). However, there are several regions with positive Tajima's D values that

208   are spread throughout the genome (Fig. 3B). At the gene level, 122 genes were identified with a

209   Tajima's D score greater than one (Supplemental Table S2). No specific gene ontology terms or

210   biological processes are significantly enriched among these genes with high Tajima's D scores.

211   However, a few genes homologous to *Plasmodium* genes with known functions in host cell

212   invasion showed high Tajima's D scores, including PBANKA_132170 (berghepain 1),

213   PBANKA_041290 (circumsporozoite- and TRAP-related protein, *Pb*CTRP),

214   PBANKA_1115300 (glideosome-associated protein 40, *Pb*GAP40) and PBANKA_1137800

215   (glideosome-associated connector, *Pb*GAC).

216   Next, variants private to *Pb*RC located within genes were examined to identify putative

217   causal variants of the drug resistant trait in this strain. To our knowledge, the only prior evidence

218   of *Pb*RC genetic variants compared with the *Pb*ANKA reference is an Ile to Lys substitution at

219   residue 413 (I413K) of the gamma glutamylcysteine synthetase (*Pb*γ-*gcs*) gene (Pérez-Rosado et

220   al., 2002). We confirmed this mutation in *Pb*RC from whole genome and Sanger sequencing.

221   Comparison of *P. berghei* strain genomes identified this variant as private to *Pb*RC (Table 2). A

222   previous study reported a possible translocation of the multidrug resistance associated protein

223   (MRP) gene to chromosome 8 in *Pb*RC (Gonzalez-Pons et al., 2009). However, visualization of

224   *Pb*RC mapped reads showed contiguity in chromosome 14 in the vicinity of the MRP gene

225   reference location, and thus no evidence of translocation (Supplemental Figure S1).

226   We searched for candidate drug resistance variants in genes known to modulate CQ

227   and/or ART drug resistance in other *Plasmodium* spp. The variants in these genes private to

228   *Pb*RC are shown in Table 2. Mutations in the *P. chabaudi* ubiquitin carboxyl-terminal hydrolase

229   1 (*ubp1)* gene modulate CQ and ART resistance in laboratory-selected drug resistant parasites

230   (Hunt et al., 2007); (Henriques et al., 2013). Five non-synonymous variants private to *Pb*RC

231   were found in the homologous *Pbubp1* gene. Mutations in the chloroquine resistance transporter

232   (*crt*) gene modulate CQ resistance in *P. falciparum* (Martin & Kirk, 2004); (Ecker et al., 2012).

233   A non-synonymous V42F variant private to *Pb*RC was found in the *P. berghei* homologue

234   (*Pbcrt*). Mutation of the *P. falciparum* multidrug resistance associated protein 1 (*mdr1*) gene

235   modulates sensitivity to several antimalarials, including CQ and ART (Sanchez et al., 2010). A

236   non-synonymous V54A variant private to *Pb*RC was found in the homologous *Pbmdr1* gene.

237   Among other proteins implicated as modulators of ART sensitivity in *P. falciparum*, a non-

238   synonymous F320C variant private to *Pb*RC was found in the *Pbpi3k* gene homologous to

239   phosphatidylinositol-3-kinase (*Pf*PI3K) (Mbengue et al., 2015). *Pb*RC mutations were notably

240   absent from some genes strongly implicated as modulators of ART sensitivity, including the

241   homologues of μ subunit of adaptor protein 2 complex (AP2-μ; PBANKA_1433900), a gene

242   mutated in the *P. chabaudi* CQ- and ART-resistant AS-ART strain (Henriques et al., 2013) and

243   K13 Kelch propeller (*Pb*K13; PBANKA_1356700), mutations of which confer reduced ART

244   sensitivity in *P. falciparum* (Ariey et al., 2013); (Ghorbal et al., 2014).

245   Catabolism of hemoglobin is tied to the mechanisms of action of both CQ and ART. CQ

246   prevents crystallization of heme, a product of hemoglobin catabolism (Martin & Kirk, 2004),

247   whereas the antimalarial effect of ART is strongly dependent on catabolism of hemoglobin

248   (Klonis et al., 2011); (Xie et al., 2015). Given that the CQ and ART-resistant *Pb*RC is defective

249   for hemoglobin catabolism and hemozoin formation (Peters, 1964); (Peters, Fletcher & Staeubli,

250   1965), we investigated whether *Pb*RC harbored private variants in genes encoding proteins

251   known to function in catabolism of hemoglobin (Ponsuwanna et al., 2016); (Lin et al., 2015).

252   Among these genes, a non-synonymous variant private to *Pb*RC was found only in the dipeptidyl

253   aminopeptidase 3 gene (*Pbdpap3*) (Table 2).

254   In addition to genes previously associated with CQ and ART-resistance, *Pb*RC private
255   variants in other genes may be responsible for the traits associated with this strain. *Pb*RC does
256   not produce gametocytes (Peters, Fletcher & Staeubli, 1965), and so may harbor variants in
257   genes important for gametocytogenesis. *Pb*RC is also reported as slow growing compared with
258   the drug-sensitive N strain (Peters, 1964), and so may harbor variants in genes important for
259   growth. Furthermore, the ART resistance trait of *Pb*RC may involve mutations in several genes,
260   since ART resistance in *P. falciparum* is a heritable trait in which the expression patterns of
261   many genes are changed (Mok et al., 2015). *Pb*RC private variants were identified in genes
262   mutated in gametocyte non-producing lines derived from the reference strain *Pb*ANKA (Sinha et
263   al., 2014), genes with annotated function with respect to blood-stage growth (Bushell et al.,
264   2017) and *P. berghei* orthologues of genes with altered expression in ART-resistant *P.*
265   *falciparum* (Mok et al., 2015). The summaries of *Pb*RC private variants for each phenotypic
266   category are shown in Table 3, and all variants in each phenotypic category are shown in
267   Supplemental Table S3. Non-synonymous variants are more prevalent than synonymous for each
268   phenotypic category, including essential genes.

269   Finally, drug resistance in *Plasmodium* spp. is often modulated by copy number variants;
270   for instance, the *P. falciparum* gene *pfmdr1* is amplified in geographical regions where
271   antimalarial resistance is common (Nair et al., 2006). Copy number variants (CNVs) among *P.*
272   *berghei* strains were identified from whole-genome sequencing data (Supplemental Table S4).
273   All of the CNVs are present in chromosomal regions near chromosome ends. All of the genes
274   that overlap CNVs are members of multigene families. Amplified regions with copy number
275   greater than one are not private to any strain, including *Pb*RC. Large deletions (>1 kb) are
276   prevalent in all strains except *Pb*SP11_A, and *Pb*RC has the greatest number of large deletions.

277

278   **Discussion**
279   Our analysis of *P. berghei* genomes revealed that the majority of SNVs are private to
280   each strain, in which *Pb*RC is most diverged strain of all. Moreover, the genome-wide pattern of
281   negative Tajima's D score (Fig. 3) can be interpreted as evidence of an excess of rare variants
282   because of a recent population bottleneck. Other *Plasmodium* spp. also show overall negative
283   Tajima's D score (Parobek et al., 2016). The overall negative Tajima's D score (and other

284   measures (Rutledge et al., 2017); (Otto et al., 2016)), are consistent with recent population

285   bottlenecks in *Plasmodium* spp. Although Tajima's D score is negative overall for *P. berghei*,

286   we identified some genes with high Tajima's D scores. Some of these genes are known to

287   function in host cell invasion, including *Pb*CTRP, berghepain 1, *Pb*GAP40 and *Pb*GAC. The *P.*

288   *berghei* CTRP protein is important for mosquito midgut invasion by the ookinete (Dessens et

289   al., 1999). The berghepain 1 gene may also function in midgut invasion, as mutants of the *P.*

290   *falciparum* homologous gene (falcipain 1) are defective for oocyst production (Eksi et al., 2004).

291   The GAP40 and GAC genes may function as part of the glideosome complex, an actin- and

292   myosin-based machine conserved across Apicomplexa that powers parasite motility, migration,

293   host cell invasion and egress (Frénal et al., 2010); (Jacot et al., 2016). The high Tajima's D

294   scores for these genes suggests balancing selection could operate in *P. berghei* to favour alleles

295   that produce antigenically diverse proteins allowing invasive parasites to evade host immune

296   systems.

297   The large number of variants private to *Pb*RC makes it difficult to pinpoint causal

298   variants of its drug-resistant phenotype. Moreover, the N strain progenitor from which *Pb*RC

299   was derived by chloroquine selection was not available to us, and forward genetic mapping is not

300   possible since *Pb*RC does not produce gametocytes (Peters, Fletcher & Staeubli, 1965).

301   Therefore, we manually curated variants private to *Pb*RC to identify those in candidate genes

302   known to modulate drug sensitivity in other *Plasmodium* spp. The level of GSH modulates

303   chloroquine sensitivity in *P. falciparum*, which is controlled by the level of expression of γ-

304   glutamyl cysteine synthetase enzyme (Ginsburg & Golenser, 2003). *Pb*RC carries a I413K

305   variant in the *Pb*γ-*gcs* gene encoding this enzyme (Table 2). However, *P. berghei* parasites with

306   knockout of *Pb*γ-*gcs* do not show altered sensitivity to CQ or ART (Vega-Rodríguez et al.,

307   2015); (Songsungthong et al., 2016). Furthermore, the *Pb*γ-*gcs* I413K variant may have pre-

308   existed in the CQ-sensitive N strain progenitor (Pérez-Rosado et al., 2002), and so is not likely to

309   modulate drug sensitivity.

310   *Pb*RC harbors five non-synonymous variants in the *Pbubp1* gene. The homologous gene

311   is mutated in drug-resistant *P. chabaudi* (Hunt et al., 2007); (Henriques et al., 2013). The *ubp1*

312   mutated residues in drug-resistant *P. chabaudi* are located in a conserved putative ubiquitin

313   binding region of the protein, which may disrupt its function and lead to increased proteasomal

314   degradation of UBP-1 substrates (Hunt et al., 2007). However, none of the *Pbubp1* variants in

315   *Pb*RC correspond to the *ubp1* mutations in *P. chabaudi*, and except for N2279D, the *Pbubp1*

316   variant residues are not conserved among *Plasmodium* spp. Therefore, it is not known if the

317   *Pbubp1* variants in *Pb*RC affect the function of the protein. Disruption of ubiquitination

318   pathways has been implicated in CQ and ART resistance mechanisms, although such disruption

319   could occur by mutations in different genes.

320   Genome sequencing of the CQ- and ART-resistant clone AS-ART isolated in (Afonso et

321   al., 2006) revealed a non-synonymous mutation in the AP2-μ gene, which was not present in the

322   AS-15CQ progenitor strain (Henriques et al., 2013). Cross resistance to CQ and ART was thus

323   attributed to the presence of *ubp1* and AP2-μ mutations in AS-ART (Henriques et al., 2013).

324   AP2-μ gene mutation in the AS-ART parasite was proposed to change the balance of

325   endocytosis toward a clathrin-independent pathway (Henriques et al., 2013). In this scenario,

326   endocytosis of hemoglobin may be reduced with subsequent lower production of heme catabolite

327   and reduced CQ and ART efficacy. However, no variants in the *P. berghei* AP2-μ homologue

328   were found in *Pb*RC, suggesting that resistance pathways differ between AS-ART and *Pb*RC.

329   Mutations in the *P. falciparum* K13 gene are thought to lead to important changes in

330   ubiquitination patterns that affect sensitivity to ART (Dogovski et al., 2015); however, no *Pb*RC

331   variants were found in the homologous gene *Pb*K13. The activity of *Pf*PI3K modulates ART

332   sensitivity in *P. falciparum*, which is controlled by K13-mediated ubiquitination (Mbengue et al.,

333   2015). *Pb*RC harbors a F320C variant in the homologous *Pbpi3k* gene. The F320C variant is

334   located in an N terminal domain not present in homologous PI3K proteins with reported

335   structures (i.e. human and *Drosophila*); hence, it is difficult to predict if this mutation affects

336   *Pb*PI3K protein function.

337   Mutations in the *Pfcrt* gene modulate CQ sensitivity in *P. falciparum*, and a V42F variant

338   was found in the homologous gene (*Pbcrt*) of *Pb*RC (Table 2). The *Pbcrt* residue 42 is predicted

339   to be in the N-terminal cytosolic part of the protein before the first transmembrane domain;

340   however, the equivalent residue in *Pfcrt* is not reported as mutated among field isolates of CQ-

341   resistant *P. falciparum* (Ecker et al., 2012). It should be noted that other *crt* mutations, such as

342   the C101F variant in piperaquine-resistant *P. falciparum* parasites (Dhingra et al., 2017) can

343   cause defects in food vacuole morphology similar to *Pb*RC. It is difficult to predict the effect of

344   the V42F variant on *Pbcrt* function by comparison with *Pfcrt*, since allelic replacement of *Pbcrt*

345   with *Pfcrt* from CQ-resistant *P. falciparum* modulated CQ sensitivity only in sexual stages of *P.*

346   *berghei*, pointing to divergence of *crt* gene function between the two species (Ecker et al., 2011).

347   The degree of CQ sensitivity in *P. falciparum* with *Pfcrt* mutation is modulated by variants in

348   other genes such as *Pfmdr1*, as shown recently by genome editing studies (Veiga et al., 2016).

349   The *Pb*RC *Pbmdr1* V54A variant residue is not equivalent to any *Pfmdr1* residue shown to

350   modulate CQ or ART sensitivity (Veiga et al., 2016); hence, it is difficult to predict the effect of

351   the V54A variant on *Pbmdr1* function.

352   The *Pb*RC strain produces less hemozoin than other CQ-sensitive strains (Peters, 1964);

353   (Peters, Fletcher & Staeubli, 1965), which may be due to defective hemoglobin digestive

354   enzymes. A protein complex of hemoglobin digestive enzymes has been described in *P.*

355   *falciparum* (Chugh et al., 2013). This complex may have a simpler composition in *P. berghei*,

356   since this species possesses only one digestive falcipain-like enzyme (berghepain-2), one

357   plasmepsin (plasmepsin IV) and one hemoglobin digestive protein (HDP) (Ponsuwanna et al.,

358   2016). No variants were found among these genes in *Pb*RC, and of the downstream digestive

359   enzymes, only one variant in the *Pbdpap3* gene was found. This gene is non-essential in *P.*

360   *berghei* (Lin et al., 2015), and so mutation of this gene may play a minor role in modulating CQ

361   and ART sensitivity.

362   We extended the candidate gene approach towards association of *Pb*RC private variants

363   with other phenotypic traits (Table 3 and Supplemental Table S3). The inability of the *Pb*RC

364   strain to produce gametocytes is likely due to non-synonymous mutations in genes previously

365   shown to be mutated in gametocyte non-producing lines (Sinha et al., 2014). Most importantly,

366   among these genes the transcription factor AP2-G known to be essential for gametocytogenesis

367   harbors the missense mutations P86S, S159N and F546V. Using the same reasoning, the slow

368   growth of *Pb*RC is likely caused by the many non-synonymous mutations in genes which are

369   known to cause growth defects when knocked out (Bushell et al., 2017). The variants in genes

370   causing growth defects may also contribute to the drug-resistant trait, since nearly a third of

371   genes orthologous with those with altered expression in ART-resistant *P. falciparum* isolates

372   (Mok et al., 2015) are also essential (Bushell et al., 2017).

373   In addition to non-synonymous SNVs, CNVs could contribute towards some of the *Pb*RC

374   phenotypic traits. However, no CNVs were found in core genes that could modulate antimalarial

375   sensitivity or growth. CNVs were found only among sub-telomeric genomic regions

376   encompassing multigene families, which were reported previously to be variable among *P.*

377   *berghei* strains (Otto et al., 2014). Large deletions appear to be more prevalent in sub-telomeric

378   regions in *Pb*RC compared with other strains, although we are cautious in making any inferences

379   from these patterns owing to potential sequence alignment error in these genomic regions.

380

381   **Conclusions**

382   Analysis of genomic variants across six strains of *P. berghei* revealed an excess of rare

383   variants, consistent with a population bottleneck as reported for other *Plasmodium* spp. Several

384   variants were identified as private to *Pb*RC which could modulate drug sensitivity, although

385   direct testing of these variants using approaches such as genome editing is necessary to test

386   causality.

387

388   **Acknowledgements**

389   The following reagents were obtained through BEI Resources, NIAID, NIH: *Plasmodium*

390   *berghei*, Strain (ANKA) 507m6cl1, MRA-867, contributed by CJ Janse and AP Waters and

391   *Plasmodium berghei* RC, MRA-404, deposited by W Peters and BL Robinson. We thank Dafra

392   Pharma for artesunate, Dr. Thomas Otto for suggestions on genome analysis, Miss Wanrisa

393   Khamtawee for technical assistance, the reviewers for their suggestions, and the curators of

394   PlasmoDB and GeneDB,

395

396   **References**

397   Afonso A., Hunt P., Cheesman S., Alves AC., Cunha CV., do Rosário V., Cravo P. 2006.

398   Malaria parasites can develop stable resistance to artemisinin but lack mutations in

399   candidate genes atp6 (encoding the sarcoplasmic and endoplasmic reticulum Ca2+

400   ATPase), tctp, mdr1, and cg10. *Antimicrobial Agents and Chemotherapy* 50:480–489.

401   DOI: 10.1128/AAC.50.2.480-489.2006.

402   Amato R., Lim P., Miotto O., Amaratunga C., Dek D., Pearson RD., Almagro-Garcia J., Neal
403   AT., Sreng S., Suon S., Drury E., Jyothi D., Stalker J., Kwiatkowski DP., Fairhurst RM.
404   2016. Genetic markers associated with dihydroartemisinin–piperaquine failure in
405   *Plasmodium falciparum* malaria in Cambodia: a genotype–phenotype association study.
406   *The Lancet Infectious Diseases*. DOI: 10.1016/S1473-3099(16)30409-1.

407   Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available
408   online from: [http://www.bioinformatics.babraham.ac.uk/projects/fastqc] (accessed 12
409   June 2017)

410   Ariey F., Witkowski B., Amaratunga C., Beghain J., Langlois A-C., Khim N., Kim S., Duru V.,
411   Bouchier C., Ma L., Lim P., Leang R., Duong S., Sreng S., Suon S., Chuor CM., Bout
412   DM., Ménard S., Rogers WO., Genton B., Fandeur T., Miotto O., Ringwald P., Le Bras
413   J., Berry A., Barale J-C., Fairhurst RM., Benoit-Vical F., Mercereau-Puijalon O., Ménard
414   D. 2013. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria.
415   *Nature* 505:50–55. DOI: 10.1038/nature12876.

416   Aurrecoechea C., Brestelli J., Brunk BP., Dommer J., Fischer S., Gajria B., Gao X., Gingle A.,
417   Grant G., Harb OS., Heiges M., Innamorato F., Iodice J., Kissinger JC., Kraemer E., Li
418   W., Miller JA., Nayak V., Pennington C., Pinney DF., Roos DS., Ross C., Stoeckert CJ.,
419   Treatman C., Wang H. 2009. PlasmoDB: a functional genomic database for malaria
420   parasites. *Nucleic Acids Research* 37:D539-543. DOI: 10.1093/nar/gkn814.

421   Bushell E., Gomes AR., Sanderson T., Anar B., Girling G., Herd C., Metcalf T., Modrzynska K.,
422   Schwach F., Martin RE., Mather MW., McFadden GI., Parts L., Rutledge GG., Vaidya
423   AB., Wengelnik K., Rayner JC., Billker O. 2017. Functional Profiling of a *Plasmodium*
424   Genome Reveals an Abundance of Essential Genes. *Cell* 170:260–272.e8. DOI:
425   10.1016/j.cell.2017.06.030.

426   Chugh M., Sundararaman V., Kumar S., Reddy VS., Siddiqui WA., Stuart KD., Malhotra P.
427   2013. Protein complex directs hemoglobin-to-hemozoin formation in *Plasmodium*
428   *falciparum*. *Proceedings of the National Academy of Sciences of the United States of*
429   *America* 110:5392–5397. DOI: 10.1073/pnas.1218412110.

430   Cingolani P., Platts A., Wang LL., Coon M., Nguyen T., Wang L., Land SJ., Lu X., Ruden DM.
431   2012. A program for annotating and predicting the effects of single nucleotide

432   polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118;

433   iso-2; iso-3. *Fly* 6:80–92. DOI: 10.4161/fly.19695.

434 Dessens JT., Beetsma AL., Dimopoulos G., Wengelnik K., Crisanti A., Kafatos FC., Sinden RE.

435   1999. CTRP is essential for mosquito infection by malaria ookinetes. *The EMBO journal*

436   18:6221–6227. DOI: 10.1093/emboj/18.22.6221.

437 Dhingra SK., Redhi D., Combrinck JM., Yeo T., Okombo J., Henrich PP., Cowell AN., Gupta

438   P., Stegman ML., Hoke JM., Cooper RA., Winzeler E., Mok S., Egan TJ., Fidock DA.

439   2017. A Variant PfCRT Isoform Can Contribute to *Plasmodium falciparum* Resistance to

440   the First-Line Partner Drug Piperaquine. *mBio* 8:e00303-17. DOI: 10.1128/mBio.00303-

441   17.

442 Dogovski C., Xie SC., Burgio G., Bridgford J., Mok S., McCaw JM., Chotivanich K., Kenny S.,

443   Gnädig N., Straimer J., Bozdech Z., Fidock DA., Simpson JA., Dondorp AM., Foote S.,

444   Klonis N., Tilley L. 2015. Targeting the Cell Stress Response of *Plasmodium falciparum*

445   to Overcome Artemisinin Resistance. *PLOS Biology* 13:e1002132. DOI:

446   10.1371/journal.pbio.1002132.

447 Duru V., Witkowski B., Ménard D. 2016. *Plasmodium falciparum* Resistance to Artemisinin

448   Derivatives and Piperaquine: A Major Challenge for Malaria Elimination in Cambodia.

449   *The American Journal of Tropical Medicine and Hygiene* 95:1228–1238. DOI:

450   10.4269/ajtmh.16-0234.

451 Ecker A., Lakshmanan V., Sinnis P., Coppens I., Fidock DA. 2011. Evidence that mutant PfCRT

452   facilitates the transmission to mosquitoes of chloroquine-treated *Plasmodium*

453   gametocytes. *The Journal of Infectious Diseases* 203:228–236. DOI:

454   10.1093/infdis/jiq036.

455 Ecker A., Lehane AM., Clain J., Fidock DA. 2012. PfCRT and its role in antimalarial drug

456   resistance. *Trends in Parasitology* 28:504–514. DOI: 10.1016/j.pt.2012.08.002.

457 Eksi S., Czesny B., Greenbaum DC., Bogyo M., Williamson KC. 2004. Targeted disruption of

458   *Plasmodium falciparum* cysteine protease, falcipain 1, reduces oocyst production, not

459   erythrocytic stage growth: Targeted disruption of falcipain 1. *Molecular Microbiology*

460   53:243–250. DOI: 10.1111/j.1365-2958.2004.04108.x.

461   Frénal K., Polonais V., Marq J-B., Stratmann R., Limenitakis J., Soldati-Favre D. 2010.

462   Functional Dissection of the Apicomplexan Glideosome Molecular Architecture. *Cell*

463   *Host & Microbe* 8:343–357. DOI: 10.1016/j.chom.2010.09.002.

464   Ghorbal M., Gorman M., Macpherson CR., Martins RM., Scherf A., Lopez-Rubio J-J. 2014.

465   Genome editing in the human malaria parasite *Plasmodium falciparum* using the

466   CRISPR-Cas9 system. *Nature Biotechnology* 32:819–821. DOI: 10.1038/nbt.2925.

467   Ginsburg H., Golenser J. 2003. Glutathione is involved in the antimalarial action of chloroquine

468   and its modulation affects drug sensitivity of human and murine species of *Plasmodium*.

469   *Redox Report: Communications in Free Radical Research* 8:276–279. DOI:

470   10.1179/135100003225002907.

471   Gonzalez-Pons M., Szeto AC., Gonzalez-Mendez R., Serrano AE. 2009. Identification and

472   bioinformatic characterization of a multidrug resistance associated protein (ABCC) gene

473   in *Plasmodium berghei*. *Malaria Journal* 8:1. DOI: 10.1186/1475-2875-8-1.

474   Henriques G., Martinelli A., Rodrigues L., Modrzynska K., Fawcett R., Houston DR., Borges

475   ST., d'Alessandro U., Tinto H., Karema C., Hunt P., Cravo P. 2013. Artemisinin

476   resistance in rodent malaria--mutation in the AP2 adaptor μ-chain suggests involvement

477   of endocytosis and membrane protein trafficking. *Malaria Journal* 12:118. DOI:

478   10.1186/1475-2875-12-118.

479   Hunt P., Afonso A., Creasey A., Culleton R., Sidhu ABS., Logan J., Valderramos SG., McNae I.,

480   Cheesman S., Rosario V do., Carter R., Fidock DA., Cravo P. 2007. Gene encoding a

481   deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent

482   malaria parasites. *Molecular Microbiology* 65:27–40. DOI: 10.1111/j.1365-

483   2958.2007.05753.x.

484   Imwong M., Suwannasin K., Kunasol C., Sutawong K., Mayxay M., Rekol H., Smithuis FM.,

485   Hlaing TM., Tun KM., van der Pluijm RW., Tripura R., Miotto O., Menard D., Dhorda

486   M., Day NPJ., White NJ., Dondorp AM. 2017. The spread of artemisinin-resistant

487   *Plasmodium falciparum* in the Greater Mekong subregion: a molecular epidemiology

488   observational study. *The Lancet. Infectious Diseases* 17:491–497. DOI: 10.1016/S1473-

489   3099(17)30048-8.

490   Jacot D., Tosetti N., Pires I., Stock J., Graindorge A., Hung Y-F., Han H., Tewari R., Kursula I.,

491   Soldati-Favre D. 2016. An Apicomplexan Actin-Binding Protein Serves as a Connector

492   and Lipid Sensor to Coordinate Motility and Invasion. *Cell Host & Microbe* 20:731–743.

493   DOI: 10.1016/j.chom.2016.10.020.

494   Klambauer G., Schwarzbauer K., Mayr A., Clevert D-A., Mitterecker A., Bodenhofer U.,

495   Hochreiter S. 2012. cn.MOPS: mixture of Poissons for discovering copy number

496   variations in next-generation sequencing data with a low false discovery rate. *Nucleic*

497   *Acids Research* 40:e69–e69. DOI: 10.1093/nar/gks003.

498   Klonis N., Crespo-Ortiz MP., Bottova I., Abu-Bakar N., Kenny S., Rosenthal PJ., Tilley L. 2011.

499   Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and

500   digestion. *Proceedings of the National Academy of Sciences of the United States of*

501   *America* 108:11405–11410. DOI: 10.1073/pnas.1104063108.

502   Langmead B., Trapnell C., Pop M., Salzberg SL. 2009. Ultrafast and memory-efficient alignment

503   of short DNA sequences to the human genome. *Genome Biology* 10:R25. DOI:

504   10.1186/gb-2009-10-3-r25.

505   Limpiti T., Intarapanich A., Assawamakin A., Shaw PJ., Wangkumhang P., Piriyapongsa J.,

506   Ngamphiw C., Tongsima S. 2011. Study of large and highly stratified population datasets

507   by combining iterative pruning principal component analysis and STRUCTURE. *BMC*

508   *Bioinformatics* 12:255. DOI: 10.1186/1471-2105-12-255.

509   Lin J., Spaccapelo R., Schwarzer E., Sajid M., Annoura T., Deroost K., Ravelli RBG., Aime E.,

510   Capuccini B., Mommaas-Kienhuis AM., O'Toole T., Prins F., Franke-Fayard BMD.,

511   Ramesar J., Chevalley-Maurel S., Kroeze H., Koster AJ., Tanke HJ., Crisanti A.,

512   Langhorne J., Arese P., Van den Steen PE., Janse CJ., Khan SM. 2015. Replication of

513   *Plasmodium* in reticulocytes can occur without hemozoin formation, resulting in

514   chloroquine resistance. *The Journal of Experimental Medicine* 212:893–903. DOI:

515   10.1084/jem.20141731.

516   Logan-Klumpler FJ., De Silva N., Boehme U., Rogers MB., Velarde G., McQuillan JA., Carver

517   T., Aslett M., Olsen C., Subramanian S., Phan I., Farris C., Mitra S., Ramasamy G.,

518   Wang H., Tivey A., Jackson A., Houston R., Parkhill J., Holden M., Harb OS., Brunk

519   BP., Myler PJ., Roos D., Carrington M., Smith DF., Hertz-Fowler C., Berriman M. 2012.

520   GeneDB-an annotation database for pathogens. *Nucleic Acids Research* 40:D98–D108.

521   DOI: 10.1093/nar/gkr1032.

522   Martin RE., Kirk K. 2004. The malaria parasite's chloroquine resistance transporter is a member
523   of the drug/metabolite transporter superfamily. *Molecular Biology and Evolution*
524   21:1938–1949. DOI: 10.1093/molbev/msh205.

525   Mbengue A., Bhattacharjee S., Pandharkar T., Liu H., Estiu G., Stahelin RV., Rizk SS., Njimoh
526   DL., Ryan Y., Chotivanich K., Nguon C., Ghorbal M., Lopez-Rubio J-J., Pfrender M.,
527   Emrich S., Mohandas N., Dondorp AM., Wiest O., Haldar K. 2015. A molecular
528   mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature*
529   520:683–687. DOI: 10.1038/nature14412.

530   McKenna A., Hanna M., Banks E., Sivachenko A., Cibulskis K., Kernytsky A., Garimella K.,
531   Altshuler D., Gabriel S., Daly M., DePristo MA. 2010. The Genome Analysis Toolkit: a
532   MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
533   *Research* 20:1297–1303. DOI: 10.1101/gr.107524.110.

534   Mok S., Ashley EA., Ferreira PE., Zhu L., Lin Z., Yeo T., Chotivanich K., Imwong M.,
535   Pukrittayakamee S., Dhorda M., Nguon C., Lim P., Amaratunga C., Suon S., Hien TT.,
536   Htut Y., Faiz MA., Onyamboko MA., Mayxay M., Newton PN., Tripura R., Woodrow
537   CJ., Miotto O., Kwiatkowski DP., Nosten F., Day NPJ., Preiser PR., White NJ., Dondorp
538   AM., Fairhurst RM., Bozdech Z. 2015. Drug resistance. Population transcriptomics of
539   human malaria parasites reveals the mechanism of artemisinin resistance. *Science*
540   347:431–435. DOI: 10.1126/science.1260403.

541   Morgulis A., Gertz EM., Schäffer AA., Agarwala R. 2006. A fast and symmetric DUST
542   implementation to mask low-complexity DNA sequences. *Journal of Computational*
543   *Biology: A Journal of Computational Molecular Cell Biology* 13:1028–1040. DOI:
544   10.1089/cmb.2006.13.1028.

545   Nair S., Nash D., Sudimack D., Jaidee A., Barends M., Uhlemann A-C., Krishna S., Nosten F.,
546   Anderson TJC. 2006. Recurrent Gene Amplification and Soft Selective Sweeps during
547   Evolution of Multidrug Resistance in Malaria Parasites. *Molecular Biology and Evolution*
548   24:562–573. DOI: 10.1093/molbev/msl185.

549   Nair S., Williams JT., Brockman A., Paiphun L., Mayxay M., Newton PN., Guthmann J-P.,
550   Smithuis FM., Hien TT., White NJ., Nosten F., Anderson TJC. 2003. A selective sweep
551   driven by pyrimethamine treatment in southeast asian malaria parasites. *Molecular*
552   *Biology and Evolution* 20:1526–1536. DOI: 10.1093/molbev/msg162.

553 Otto TD., Böhme U., Jackson AP., Hunt M., Franke-Fayard B., Hoeijmakers WAM., Religa
554   AA., Robertson L., Sanders M., Ogun SA., Cunningham D., Erhart A., Billker O., Khan
555   SM., Stunnenberg HG., Langhorne J., Holder AA., Waters AP., Newbold CI., Pain A.,
556   Berriman M., Janse CJ. 2014. A comprehensive evaluation of rodent malaria parasite
557   genomes and gene expression. *BMC Biology* 12. DOI: 10.1186/s12915-014-0086-0.

558 Otto TD., Gilabert A., Crellen T., Böhme U., Arnathau C., Sanders M., Oyola S., Okauga AP.,
559   Boundenga L., Wuillaume E., Ngoubangoye B., Moukodoum N., Paupy C., Durand P.,
560   Rougeron V., Ollomo B., Renaud F., Newbold C., Berriman M., Prugnolle F. 2016.
561   Genomes of an entire *Plasmodium* subgenus reveal paths to virulent human malaria.
562   bioRxiv 095679; doi: https://doi.org/10.1101/095679.

563 Parobek CM., Lin JT., Saunders DL., Barnett EJ., Lon C., Lanteri CA., Balasubramanian S.,
564   Brazeau N., DeConti DK., Garba DL., Meshnick SR., Spring MD., Chuor CM., Bailey
565   JA., Juliano JJ. 2016. Selective sweep suggests transcriptional regulation may underlie
566   *Plasmodium vivax* resilience to malaria control measures in Cambodia. *Proceedings of*
567   *the National Academy of Sciences* 113:E8096–E8105. DOI: 10.1073/pnas.1608828113.

568 Pérez-Rosado J., Gervais GW., Ferrer-Rodríguez I., Peters W., Serrano AE. 2002. *Plasmodium*
569   *berghei*: analysis of the gamma-glutamylcysteine synthetase gene in drug-resistant lines.
570   *Experimental Parasitology* 101:175–182.

571 Peters W. 1964. Pigment formation and nuclear division in chloroquine-resistant malaria
572   parasites (*Plasmodium berghei*, Vincke and Lips, 1948). *Nature* 203:1290–1291.

573 Peters W., Fletcher KA., Staeubli W. 1965. Phagotrophy and pigment formation in a
574   chloroquine-resistant strain of *Plasmodium berghei* Vincke and Lips, 1948. *Annals of*
575   *Tropical Medicine and Parasitology* 59:126–134.

576 Pfeifer B., Wittelsbürger U., Ramos-Onsins SE., Lercher MJ. 2014. PopGenome: an efficient
577   Swiss army knife for population genomic analyses in R. *Molecular Biology and*
578   *Evolution* 31:1929–1936. DOI: 10.1093/molbev/msu136.

579 Ponsuwanna P., Kochakarn T., Bunditvorapoom D., Kümpornsin K., Otto TD., Ridenour C.,
580   Chotivanich K., Wilairat P., White NJ., Miotto O., Chookajorn T. 2016. Comparative
581   genome-wide analysis and evolutionary history of haemoglobin-processing and haem
582   detoxification enzymes in malarial parasites. *Malaria Journal* 15. DOI: 10.1186/s12936-
583   016-1097-9.

584 Ritz C., Streibig JC. 2005. Bioassay analysis using R. *Journal of Statistical Software* 12:1–22.

585 DOI: 10.18637/jss.v012.i05

586 Rutledge GG., Böhme U., Sanders M., Reid AJ., Cotton JA., Maiga-Ascofare O., Djimdé AA.,

587 Apinjoh TO., Amenga-Etego L., Manske M., Barnwell JW., Renaud F., Ollomo B.,

588 Prugnolle F., Anstey NM., Auburn S., Price RN., McCarthy JS., Kwiatkowski DP.,

589 Newbold CI., Berriman M., Otto TD. 2017. *Plasmodium malariae* and *P. ovale* genomes

590 provide insights into malaria parasite evolution. *Nature* 542:101–104. DOI:

591 10.1038/nature21038.

592 Sanchez CP., Dave A., Stein WD., Lanzer M. 2010. Transporters as mediators of drug resistance

593 in *Plasmodium falciparum*. *International Journal for Parasitology* 40:1109–1118. DOI:

594 10.1016/j.ijpara.2010.04.001.

595 Sinha A., Hughes KR., Modrzynska KK., Otto TD., Pfander C., Dickens NJ., Religa AA.,

596 Bushell E., Graham AL., Cameron R., Kafsack BFC., Williams AE., Llinas M., Berriman

597 M., Billker O., Waters AP. 2014. A cascade of DNA-binding proteins for sexual

598 commitment and development in *Plasmodium*. *Nature* 507:253–257. DOI:

599 10.1038/nature12970.

600 Songsungthong W., Koonyosying P., Uthaipibull C., Kamchonwongpaisan S. 2016. Inhibition of

601 Glutathione Biosynthesis Sensitizes *Plasmodium berghei* to Antifolates. *Antimicrobial*

602 *Agents and Chemotherapy* 60:3057–3064. DOI: 10.1128/AAC.01836-15.

603 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA

604 polymorphism. *Genetics* 123:585–595.

605 Tanner M., Greenwood B., Whitty CJM., Ansah EK., Price RN., Dondorp AM., von Seidlein L.,

606 Baird JK., Beeson JG., Fowkes FJI., Hemingway J., Marsh K., Osier F. 2015. Malaria

607 eradication and elimination: views on how to translate a vision into reality. *BMC*

608 *Medicine* 13. DOI: 10.1186/s12916-015-0384-6.

609 Thorvaldsdottir H., Robinson JT., Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-

610 performance genomics data visualization and exploration. Briefings in Bioinformatics

611 14:178–192. DOI: 10.1093/bib/bbs017.

612 Vega-Rodríguez J., Pastrana-Mena R., Crespo-Lladó KN., Ortiz JG., Ferrer-Rodríguez I.,

613 Serrano AE. 2015. Implications of Glutathione Levels in the *Plasmodium berghei*

614   Response to Chloroquine and Artemisinin. *PLOS ONE* 10:e0128212. DOI:

615   10.1371/journal.pone.0128212.

616 Veiga MI., Dhingra SK., Henrich PP., Straimer J., Gnädig N., Uhlemann A-C., Martin RE.,

617   Lehane AM., Fidock DA. 2016. Globally prevalent PfMDR1 mutations modulate

618   *Plasmodium falciparum* susceptibility to artemisinin-based combination therapies. *Nature*

619   *Communications* 7:11553. DOI: 10.1038/ncomms11553.

620 WHO Malaria Policy Advisory Committee and Secretariat 2015. Malaria Policy Advisory

621   Committee to the WHO: conclusions and recommendations of seventh biannual meeting

622   (March 2015). *Malaria Journal* 14:295. DOI: 10.1186/s12936-015-0787-z.

623 Wickham H. 2009. *Ggplot2: elegant graphics for data analysis*. New York: Springer.

624 Witkowski B., Lelièvre J., Nicolau-Travers M-L., Iriart X., Njomnang Soh P., Bousejra-Elgarah

625   F., Meunier B., Berry A., Benoit-Vical F. 2012. Evidence for the contribution of the

626   hemozoin synthesis pathway of the murine *Plasmodium yoelii* to the resistance to

627   artemisinin-related drugs. *PloS One* 7:e32620. DOI: 10.1371/journal.pone.0032620.

628 Woodrow CJ., White NJ. 2016. The clinical impact of artemisinin resistance in Southeast Asia

629   and the potential for future spread. *FEMS Microbiology Reviews*:fuw037. DOI:

630   10.1093/femsre/fuw037.

631 Wootton JC., Feng X., Ferdig MT., Cooper RA., Mu J., Baruch DI., Magill AJ., Su X. 2002.

632   Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature*

633   418:320–323. DOI: 10.1038/nature00813.

634 Xie SC., Dogovski C., Hanssen E., Chiu F., Yang T., Crespo MP., Stafford C., Batinovic S.,

635   Teguh S., Charman S., Klonis N., Tilley L. 2015. Haemoglobin degradation underpins

636   the sensitivity of early ring stage *Plasmodium falciparum* to artemisinins. *Journal of Cell*

637   *Science*. DOI: 10.1242/jcs.178830.

638

639

Summary of *P. berghei* genomic variants identified from whole genome sequencing.

The *Pb*RC strain was sequenced in this study; data from other strains were reported in (Otto et al., 2014).

	*Pb*RC	*Pb*K173	*Pb*NK65_E	*Pb*NK65_NY	*Pb*SP11_A	*Pb*SP11_RLL
Raw[a]	20399	5263	2034	2480	1561	5834
-private[b]	16774	1991	305	561	292	1693
-missing[c]	1295	1326	241	259	224	654
Filtered[d]	7726	1251	12	33	1	890
-private[b]	6974	504	1	17	1	311

3   [a]Single Nucleotide Variants (SNV) and small insertion/deletion (INDEL) markers called by the

4   GATK tool using default parameters.

5   [b]Markers with variant allele detected in only one strain.

6   [c]Markers with no genotype calls owing to insufficient mapped reads.

7   [d]SNV remaining after filtering to remove variants with heterozygous calls, located in repetitive

8   regions, or present in multigene families.

Non-synonymous variants private to *P. berghei* RC present in genes previously associated with chloroquine and/or artemisinin resistance in *Plasmodium* spp.

Single nucleotide variants (SNV) private to the *P. berghei* RC strain were identified by whole-genome sequencing comparing with the reference strain *Pb*ANKA v3 and other strains reported in (Otto et al., 2014).

Gene Description	Gene Symbol	Gene ID	Missense variants private to *Pb*RC	Confirmed by Sanger dideoxy sequencing
γ-glutamylcysteine synthetase	*Pbγ-gcs*	PBANKA_0819800	I413K[a]	YES
Ubiquitin carboxyl-terminal hydrolase 1	*Pbubp1*	PBANKA_0208800	R1561K	YES
			K1582E[a]	YES
			K2102E	YES
			N2279D	YES
			A2402V	YES
Chloroquine resistance transporter	*Pbcrt*	PBANKA_1219500	V42F	YES
Multidrug resistance gene	*Pbmdr1*	PBANKA_1237800	V54A	YES
Phosphatidyl-inositol-3-phosphate kinase	*Pbpi3k*	PBANKA_1114900	F320C	N.D.[b]
Dipeptidyl aminopeptidase 3	*Pbdpap3*	PBANKA_1002400	L3I	N.D.[b]

3 [a]Raw variants called by GATK, but removed by DustMasker filtering (variant located within a repetitive

4 region).

5 [b]Sanger dideoxy sequencing not done

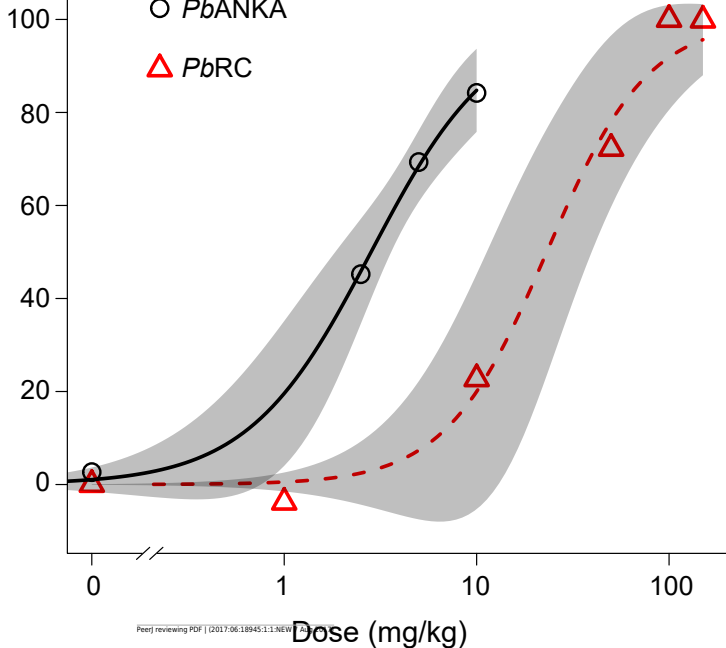
Summaries of *P. berghei* RC private variants located in genes associated with different phenotypes

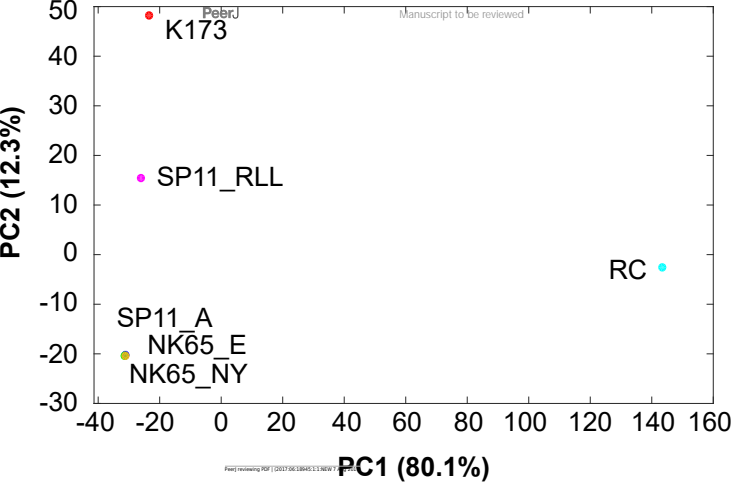
3538 single nucleotide variants (SNV) private to the *P. berghei* RC strain after filtering and located within annotated genes were identified by whole-genome sequencing comparing with the reference strain *Pb*ANKA v3 and other strains reported in (Otto et al., 2014) . The numbers of SNV in each functional category annotated by the SnpEff software version 3.6g are indicated in the columns. The numbers of variants in genes associated with different phenotypes are shown in each row. Note that some genes are associated with more than one phenotype.

	Missense variants	Synonymous variants	[a]Modifier variants	[b]Stop variants	No. genes with variants
All genes	2019	1458	55	6	1739
[c]Core	1980	1438	53	5	1705
[d]Gametocytogenesis	90	55	1	0	35
[e]Essential	563	435	8	2	484
[e]Slow	228	153	3	2	166
[e]Dispensable	380	280	11	4	342
[e]Fast	3	2	0	0	1
[f]ART-R	415	289	16	3	344

2
3 [a]Includes the following SnpEff functional categories: splice region variant & intron variant, non
4 coding transcript exon variant, splice region variant & stop retained variant, splice region variant
5 & synonymous variant, and modifier synonymous stop variant.
6 [b]Includes all stop codon variants predicted to alter the open reading frame.
7 [c]P. berghei genes orthologous across rodent and primate malaria Plasmodium spp. (Otto et al.,
8 2014).
9 [d]Genes mutated in gametocyte non-producing lines of P. berghei (Sinha et al., 2014).
10 [e]Growth phenotypes annotated from P. berghei gene knockout mutants in the PlasmoGEM
11 database (Bushell et al., 2017).
12 [f]P. berghei orthologues of P. falciparum genes with altered gene expression profiles in
13 artemisinin-resistant isolates (Mok et al., 2015).

Artesunate sensitivity of *P. berghei* RC (*Pb*RC) in infected mice.

BALB/c mice were infected intravenously with 1 x 10^7 infected red blood cells. Artesunate was given orally and parasitemia was determined four days post infection. Five to six mice per group were used for *Pb*RC and four to eight mice per group were used for *Pb*ANKA reference strain. Percent inhibition compared with untreated control (0%) was calculated for each dose of drug. Data were fitted to the two-parameter variable-slope sigmoidal dose-response equation. The average percent inhibitions for each dose are plotted, and the gray bars represent confidence regions calculated by the drc package.

Principal Components Analysis of *P. berghei strains.*

The genotypic data from 8681 Single Nucleotide Variants called from whole genome sequences were encoded as a matrix for Principal Components Analysis. The loadings from the first and second Principal Components for each strain are plotted.

Genome-wide distribution of Tajima's D scores in *P. berghei*.

Tajima's D score was calculated for sliding, non-overlapping genomic windows of five variants along all 14 chromosomes using the PopGenome tool. Scores were calculated for a total of 1730 windows from six strains compared with the reference strain *Pb*ANKA v3. The distributions of scores in each chromosome are shown as violin plots in part A. The scores in each window are plotted separately for each chromosome in part B.