Genetic Report Card User Guide

GRC V1.4 - 22 February 2024

Table of Contents

The information in this document apply to datasets released in the **GenRe-Mekong GRC V1.4** format.

Introduction

Genetic Report Card (GRC) datasets contain data resulting from the genetic analyses of patient blood samples, using the SpotMalaria platform. SpotMalaria provides data for a wide range of genetic markers and features of the malaria parasites in each sample, and predict their level of resistance to various antimalarials. The objective is to provide public health officials, researchers and other stakeholders with timely and comprehensive knowledge that can inform their policy decisions. Details of the project can be found in our paper[1]:

> Jacob CG, et al. (2021) Genetic surveillance in the Greater Mekong Subregion and South Asia to support malaria control and elimination. eLife 10:e62997 https://elifesciences.org/articles/62997

and the full methodology documentation is accessible from the project Resource Page at https://www.malariagen.net/resource/29.

The GRC data covers many aspects of malaria parasite genetics, and therefore its interpretation can be challenging. In the following sections we will provide a guide to interpreting the data, and explain the link between the genotyped mutations and the predicted resistance to antimalarial drugs.

The GRC Datasets

GRC Dataset Releases

GenRe-Mekong delivers GRC datasets as Microsoft Excel spreadsheets (.xlsx extension), containing a single worksheet, showing one row for each sample. The columns are grouped together in logical groups indicated by different colour headings, as shown in Table 1.

Colour heading	Corresponding columns
Gray	Sample metadata, e.g., location and date of sample collection
Pink	Private sample metadata (only for study partners)
Orange	Phenotype prediction
Turquoise	Sample profile, such as genetic barcodes, missingness, complexity of infection
Green	Drug-resistance loci
Blue	Amino acid alleles that are of interest
Lilac	Evidence from different genotyping methods
Red	Genetic barcode SNP

Table 1 –Colour headings for column groups in GRC datasets.

GenRe-Mekong has many collaborating studies, each with separate investigators (partners), who are considered the owners of the sample data that is generated. When we release a new update, each partner whose samples are updated receives a spreadsheet containing the full data for all processed samples submitted by their study.

From time to time, we also release aggregated public datasets, containing the data from samples contributed by all partners. The structure of these public datasets, and the meaning of the columns, will be the same as for the datasets returned to partners, except that study-specific information will be omitted from the public release. For example, the public release will not show the identifiers of the sites where the sample was collected, and the exact day of collection, but will still detail district-level geographical data and year of collection.

Predicted Drug Resistance Phenotypes (orange headers)

Columns with orange headers show the predicted level of antimalarial drug resistance for the parasites in each sample. The following drugs and combinations are covered:

Single Drugs:

- Artemisinin
- Piperaquine
- Mefloquine
- Chloroquine
- Pyrimethamine
- Sulfadoxine

Combination Therapies:

- Dihydroartemisinin-Piperaquine (column "DHA-PPQ")
- Artesunate-Mefloquine ("AS-MQ")
- Sulfadoxine-Pyrimethamine ("S-P")
- Sulfadoxine-Pyrimethamine for intermittent preventive treatment in pregnancy ("S-P-IPTp")

The level of resistance is colour-coded for easier interpretation of the data. It can take the following values:

- "**<NA>**" means that the sample was not tested for mutations that predict resistance to this antimalarial drug. This may be the case for samples that were processed before the tests were available.
- "**Missing**" means that the tests for mutations that predict resistance to this antimalarial drug were unsuccessful. Often this happens in samples with low parasite levels in the sample.
- "**Undetermined**" (yellow background) means that we were unable to predict the resistance to this antimalarial drug from the available genetic data.
- "**Resistant**" (red background) means that the sample is predicted to be resistant to this antimalarial drug or combination.
- "**Sensitive**" (green background) means that the sample is predicted to be sensitive to this antimalarial drug or combination.
- "**SensitiveWithMissingness**" (light green background) means that the sample is predicted to be sensitive to this antimalarial drug combination (only applicable for combination therapies). This value is given for antimalarial drug combinations where one drug is predicted to be sensitive, and one drug has no prediction (the prediction is missing). In such cases, we know that the parasite is sensitive- but unless predictions are available for both drugs, we could never identify parasites that are resistant. Therefore, samples with the value "**SensitiveWithMissingness**" should not be included in an estimation of the prevalence of resistance, as it would lead to a biased (unfair) prevalence outcome. For example: without testing for mefloquine resistance we cannot predict if a sample is resistant to AS-MQ. However, if the sample is predicted to be sensitive to artemisinin, we can predict that it is also sensitive to AS-MQ.

Drug Resistance Changes and Haplotypes

For each sample, we determine genotypes for several variations that are known to be associated to resistance to antimalarials from peer-reviewed scientific literature. Evidence for these associations is described in section “Resistance to Specific Antimalarials” of this document. These variations genotyped include: amino acid changes (mutations), haplotypes and gene amplifications:

- **Amino acid changes (mutations)** are caused by single nucleotide polymorphisms (SNPs) in a gene. These changes typically change the structure of the protein synthesized by the gene, and produce a change in sensitivity to the drug.
- **Combinations of multiple amino acid changes** in the same gene (**haplotypes**) are reported in genes where several amino acid changes are thought of being involved in drug resistance, so these changes are often collectively reported as one concatenated sequence.
- **Gene amplifications** occur when the parasite develops additional copies of a gene, which increases the expression level, with as a result more protein production, which subsequently modifies the response to the drug.

In the following sections, we detail the relevant columns for these groups.

Amino acid changes (blue headers)

We report 36 different amino acid changes for each sample, shown in Table A1 in the Appendix. These are amino acid alleles that are of interest; several of these are related to drug-resisatance and are components of haplotypes reported in fields of class "Resistance Locus". In each of these columns, we show the genotype at the target position as follows:

- One of the 20 amino acid 1-letter codes (e.g., **“L”** for leucine, etc.).
- Multiple letters, separated by commas (e.g., **“A,S”** for alanine+serine, etc.) if multiple alleles were detected (e.g., in the case of a multi-clonal infection).
- A dash (**“-”**) indicates a missing genotype, i.e., because the sample could not be genotyped at the relevant genome position.
- An asterisk (**“*”**) indicates that multiple alleles were detected (e.g., in the case of a multi-clonal infection) but there is no information as to which.
- **“<NA>”** indicates that the sample was not tested for this mutation, e.g. when a sample was processed before the test was available.

Details of how these genetic mutations relate to resistance to specific drugs can be found in section “Resistance to Specific Antimalarials” of this guide.

kelch13 mutations (column "Pfkelch13", green header)

Amino acid mutations in the *kelch13* gene (associated with artemisinin resistance) are reported in a different fashion. Several non-synonymous mutations (i.e. mutations that cause an amino acid change) in two domains of this gene have been associated with delayed clearance of artemisinin. Many *kelch13* mutations have been reported, although not all of them have been validated. Therefore, rather than genotype specific SNPs in this gene to report mutations, we scan the gene for any nonsynonymous mutation in the two domains of interest. These mutations are reported using the following notation:

- **"WT"** indicates that no nonsynonymous mutation is detected in the relevant domains.
- An amino acid mutation encoded using **"XnnnY"** notation, where X is the wild-type amino acid, Y is the mutated amino acid, and nnn is the amino acid position in the protein sequence. For example, **"C580Y"** indicates a C→Y mutation at position 580 in the *kelch13* gene.
- A comma-separated list of one or more of the above genotypes, in cases where more than one mutation was identified. This is likely to happen in multiclonal infections for example **"WT,C580Y"** would be reported for a sample that contains parasites with the C580Y mutation, as well as wild-type parasites.
- A dash (**"-"**) indicates a missing genotype, i.e., because a large portion of the *kelch13* gene could not be genotyped for this sample.
- **"<NA>"** indicates that the sample was not tested for *kelch13* mutations.

While many non-synonymous mutations confer artemisinin resistance, not all are confirmed. The WHO maintains a list of validates, suspected and potential mutations, which we use in predictions of artemisinin resistance phenotype. The current list is reported in Appendix Table A3.

Drug Resistance Haplotypes (green headers)

Haplotypes are generated by concatenating multiple amino acid from different positions in the same gene. The alleles used are also reported individually in the orange-header columns. For example, in the *pfdhfr* gene, we genotype amino acid positions 51, 59, 108 and 164, which are implicated in resistance to antifolates. The wild-type (WT) amino acids at these positions are N, C, S and I respectively we represent this haplotype as **"NCSI"**. If we report the haplotype **"IRNI"**, it means that the parasite is a triple mutant at positions 51 (I →N), 59 (C →R) and 108 (S →N).

Positions where multiple alleles are detected (e.g., in a multiclonal infection) are indicated with an asterisk (**"*"**); the alleles are listed in the relevant amino acid mutation column. A dash (**"-"**) indicates a missing genotype at a position. For example, a *pfdhfr* haplotype reported as **"I-N*"** has a missing genotype at position 59, and multiple alleles detected at position 108.

We have selected 5 haplotypes for their reported relevance to a number of antimalarials. These are listed in Table 2, and their relevance is detailed in section "Resistance to Specific Antimalarials".

Column	Antimalarial	Gene	Amino Acid Positions	Wild Type
PfCRT	chloroquine	*pfcrt*	72, 73, 74, 75, 76	CVMNK
PfDHFR	pyrimethamine	*pfdhfr*	51, 59, 108, 164	NCSI
PfDHPS	sulfadoxine	*pfdhps*	436, 437, 540, 581, 613	SAKAA
PfMDR1	chloroquine, amodiaquine, lumefantrine, mefloquine	*pfmdr1*	86, 184, 1246	NYD
pm23-Amp	piperaquine	*pfplasmepsin 2/3*	Amplication	WT
mdr1-Amp	mefloquine	*pfmdr1*	Amplication	WT
PGB (ART-R genetic background)	artemisinin	*pfarps10*	127, 128	VDDNIT
		ferredoxin	193	
		pfcrt	326, 356	
		pfmdr2	484	

Table 2 – List of reported haplotypes related to drug resistance.

Gene Amplification (green headers)

We report on two amplifications:

- Amplification of *pfmdr1* (column "**mdr1-Amp**") associated with resistance to mefloquine and possibly other drugs. Currently, detection of this amplification is performed by a quantitative polymerase chain reaction (qPCR) only.
- Amplification of *pfplasmepsin2 and pfplasmepsin3* (column "**pm23-Amp**") associated with resistance to piperaquine. Detection of this amplification can be performed by a quantitative polymerase chain reaction (qPCR) or by detecting a particular breakpoint sequence. Wherever possible we perform both tests and combine the results.

Amplification columns use the following encoding:

- **"WT"** indicates that no amplification was detected
- **"Amplified"** indicates that an amplification was detected (no copy numbers are specified)
- A dash (**"-"**) indicates that status of the amplification could not be determined.
- **"<NA>"** indicates that the sample was not tested for *the amplification*.

In addition, we also report in three other columns the results of the individual tests performed to identify the amplification:

- In column ("**pm23-break**") we report the results by detection of the *pfplasmepsin2/3* breakpoint sequence
- In column ("**pm23-qPCR**") we report the *pfplasmepsin2/3* results by qPCR
- In column ("**mdr1-qPCR**") we report the *pfmdr1* results by qPCR

The result in column "**pm23-Amp**" was obtained from columns "**pm23-break**" and "**pm23-qPCR**" as follows:

- If both columns report **"<NA >"**, then the result is **"<NA >"**
- Otherwise, if one column reports **"<NA>"**, then the result is the value in the other column
- Otherwise, if one column reports **"-"**, then the result is the value in the other column
- Otherwise, if the two columns agree, the result is the value of either column; if they disagree, the result is undefined (**"-"**)

Sample Profile (turquoise headers)

Genetic Barcode (column "GenBarcode")
Genetic barcodes are "summaries of the genome", used to perform a variety of analyses, such as:

- Compare samples and determine their similarity by the proportion of positions with identical alleles.
- Estimate the complexity of infection (COI) by the proportion of heterozygous calls.
- Identifying multiple samples with (almost) identical barcodes, suggesting an expanding strain.
- Reconstruct parasite origins, etc.

Genetic barcodes are formed by concatenated genotypes at 101 SNPs across the *P. falciparum* genome. These SNPs are biallelic, i.e., only two alleles are observed; they were chosen for their usefulness in analyses of relationships between parasites, and are not associated to drug resistance.

The full list of SNPs used is given in Appendix Table A3. The individual barcode component alleles are shown in 101 columns with red headers; the corresponding name of the column is given in the table.

SNPs within the barcode are represented by the observed nucleotide (A, T, C and G). If the genotype is missing (could not be detected), the symbol **"X"** is used; the symbol **"N"** indicated that both alleles were observed (heterozygous call).

Three additional columns (with purple headers) report genetic barcode statistics:

- Column "**GenBarcodeMissing**" reports the proportion of positions in the barcode that are missing (**"X"**). This can be used as a quality measure for the sample.
- Column "**GenBarcodeHet**" reports the proportion of non-missing positions in the barcode that are heterozygous (**"N"**). This can be used as an indicator of multiclonal samples.

Complexity of Infection (column "COI")

We use the sample barcodes to estimate the Complexity of Infection (COI) in each sample using the programs COIL[2] and The Real McCOIL[3]. The COI is expressed as an integer value, representing the estimated number of individual parasites within the infection from which the sample was taken (a value of 1 represents a monoclonal infection, 2 represents an infection with 2 clones, etc.). A dash (**"-"**) indicates that COI could not be estimated.

This COI estimate is an approximation that greatly simplifies the complexities of mixed infections. However, the prevalence of highly mixed infections is an useful indicator of transmission intensity, which can be used when comparing sites or seasons.

Species Co-Infections (column "Species")

We detect the presence of different *Plasmodium* species by testing the sequences of parasite mitochondria for species-specific alleles. We test for 5 different parasite species: *P. falciparum* (code: **"Pf"**), *P. vivax* (**"Pv"**), *P. knowlesi* (**"Pk"**), *P. malariae* (**"Pm"**), *P. ovale* (**"Po"**).

The value reported is a comma-separated list of the codes for all the species detected, e.g., **"Pf"** indicates that the sample only contains *P. falciparum* parasites, while **"Pf,Pv"** indicates that the sample contains both *P. falciparum* and *P. vivax* parasites. A dash (**"-"**) indicates that the species could not be determined.

Evidence (lilac headers)

Fields containing evidence from different genotyping methods, for variations that were tested in multiple ways.
For plasmepsin 2/3 amplications, we used two methods of detections by identifying a breakpoint sequence, or by qPCR.
Amplification of *mdr1* was detected by qPCR.
For specie detection, we used three detection methods: by amplicon sequencing, by qPCR and by presence of *Pf* barcode genotypes

Sample Metadata (gray headers)

The GRC provides the following metadata columns to describe the samples:

- Column "**SampleID**" contains a unique identifier (also known as "MalariaGEN identifier") for the sample. This is the primary reference for the sample.
- Column "**SeqNum**" assigns a sequential number to the sample.
- Column "**Source**" shows the project that contributed the sample. For GenRe-Mekong studies, the value is **"GenRe-Mekong"**.
- Column "**Process**" indicates the genotyping technology used to genotype the sample.
- Column "**Study**" provides the id of the study that contributed the sample.
- Column "**Year**" shows the year in which the sample was collected.
- Column "**Month**" shows the month in which the sample was collected.
- Column "**TimePoint**" indicates at what stage of the malaria infection the sample was collected.
    - Value is of the form **"DxxHyy"** where xx is the number of days and yy the number of hours after first presenting. For example, **"D00H00"** means a sample taken at admission time, and **"D01H00"** means a sample taken 24 hours after admission.

    - A value suffix **"_REC"** indicates a recurrence episode (e.g., a recrudescence).
    - A value of **"-"** indicates that the collection timepoint was not reported.
- Column "**Country**" shows the two-letter ISO 3166-1 code (https://www.iso.org/obp/ui/) of the country in which the sample was collected.
- Columns "**AdmDiv1**" and "**AdmDiv2**" contain the names of the first- and second-level administrative divisions of the location where the sample was collected.
    - These are generally defined by the country's governments.
    - The first and second level divisions are known by different terms in different countries. In several cases, they are known as "Province" and "District" respectively, but this varies, e.g., in Thailand they are called "Changwat" and "Amphoe".
- Columns "**AdmDiv1_GID**" and "**AdmDiv2_GID**" contain the identifiers of the first- and second-level administrative divisions of the location where the sample was collected, as defined by the GADM database (https://gadm.org/).

Private Sample Metadata (pink headers)

The GRC provides the following private metadata columns **only in study releases** (in public releases these columns are absent):

- Column "**SiteName**" is the name of the collection site as provided by the collecting study.
- Column "**Latitude**" and "**Longitude**" define where the collection site is located.
- Column "**CollectionDate**" shows the date when the sample was collected.
- Column "**CollectionID**" is the identifier assigned to the blood sample upon collection; it usually is provided by a barcoded sticker.
- Column "**ManifestID**" is the identifier of the batch in which the sample was submitted.
- Column "**DataRelease**" is the identifier of the processing release that generated the data for the sample.

Resistance to Specific Antimalarials

Artemisinin Drug Resistance

Non-synonymous *Kelch13* mutations
Several non-synonymous mutations in two domains (BTB/POZ and propeller) of the *kelch13* (PF3D7_1343700) gene have been associated with delayed clearance of artemisinin.[4-6] Many *kelch13* mutations have been reported around the world, but not all have been validated. We provide a list of associated mutations in Appendix Table A2.

Parasite genetic background (PGB) mutations
A study showed evidence of a genetic background of mutations that allowed for the emergence of *kelch13* mutations.[7] These mutations are:

- V127M and D128Y/H in the *pfarps10* (PF3D7_1460900) protein,
- D193Y in *ferredoxin* (*pffd*, PF3D7_1318100),
- N326S and I356T in *pfcrt* (PF3D7_0709000), and
- T484I in *pfmdr2* (PF3D7_1447900).

These are displayed in concatenated haplotype form, the reference allele (WT) being VDDNIT.

Chloroquine Drug Resistance Mutations

Chloroquine drug resistance is primarily mediated by mutations in the chloroquine resistance transporter (*pfcrt*, PF3D7_0709000).[8] An accessory mutation at position 86 in the multidrug resistance protein (*pfmdr1*, PF3D7_0523000) has been shown to accentuate this resistance phenotype in parasites.[9] The loci in *pfcrt* are represented as a 5-amino acid haplotype including positions 72-76, the wild-type (WT) haplotype being CVMNK.[10] The CVIET haplotype is the most widespread resistant haplotype in Asia and Africa, while SVMNT is common in resistant parasites in South America and Oceania.

The *pfmdr1* mutation at position 86 is the first in the 3 amino-acid haplotype reported for this gene. The WT variant is N, while the *pfmdr1* 86Y variant enhances resistance to chloroquine.

Mefloquine Drug Resistance Mutations

An amplification of the multidrug resistance 1 gene (*pfmdr1*, PF3D7_0523000) has been associated with parasite response to mefloquine, such that parasites with multiple copies of this gene are predicted to be resistant to this antimalarial.[11]

Also, there is limited evidence that variations at positions 86, 184, and 1246 increase susceptibility to mefloquine.[12,13]

Lumefantrine Drug Resistance Mutations

Mutations in multidrug resistance protein (*pfmdr1*, PF3D7_0523000) have been associated with parasite response to the drugs mefloquine and lumefantrine.[12] There is limited evidence that variations at positions 86, 184, and 1246 increase susceptibility to this antimalarial.

Amodiaquine Drug Resistance Mutations

Mutations in the multidrug resistance protein (*pfmdr1*, PF3D7_0523000) have been associated with parasite response to amodiaquine. There is limited evidence that mutations at positions 86 and 1246 can mediate response to this drug,[14] and these positions constitute the haplotype reported for this gene. In vitro experiments have shown that haplotypes containing the mutant 86Y have increased IC50's to chloroquine and amodiaquine.[15]

Piperaquine Drug Resistance

Gene amplification of a section of chromosome 14 involving the genes *plasmepsin 2* and *plasmepsin 3* has also been associated with increased resistance to piperaquine.[16] An assay designed to detect the breakpoint of a hybrid gene product can be used to detect this gene amplification, at least in Southeast Asia. In addition, a genome-wide association study (GWAS) has associated a SNP in a putative *exonuclease* gene (*pfexo*, PF3D7_1362500) with ex vivo piperaquine IC50 of parasite isolates from Cambodia.[16] This molecular marker is at position 415, and the 415G allele was significantly associated with increased tolerance of piperaquine with respect to the wild-type allele (415E).

Pyrimethamine and Sulfadoxine Drug Resistance Mutations

Pyrimethamine drug resistance is mediated by mutations in the bifunctional dihydrofolate reductase-thymidylate synthase (*pfdhfr*, PF3D7_0417200).[17] Four non- synonymous mutations at amino acid positions 51, 59, 108 and 164 have been established as important in drug resistance.[15] Resistant parasites are often characterized by the number of mutations they carry (single, double, triple, quadruple mutants), which is taken to be an indicator of the level of resistance to the drug.

Sulfadoxine drug resistance is mediated by mutations in the dihydropteroate synthetase (*pfdhps*, PF3D7_0810800). A variety of mutation combinations at positions 436, 437, 540, 581 and 613 are thought to confer resistance, and parasites with higher numbers of mutations often have higher levels of resistance.[18] Some mutations are geographically isolated, while others are seen globally.

pfdhps mutations are often seen in combination with pfdhfr mutations, since sulfadoxine and pyrimethamine are mostly used in combination (SP). One combination of particular importance is the quintuple mutant carrying mutations at positions 51, 59 and 108 in *pfdhfr*, and at positions 437 and 540 in *pfdhps* (IRNx + xGExx haplotypes). Such a combination of alleles is strongly predictive of SP treatment failure.[19]

Appendix

Table A1 – List of drug-resistance associated amino acid mutations genotyped by SpotMalaria

Num	Column	Gene	Amino Position	Wild-type allele
1	PfCRT:72	PfCRT	72	C
2	PfCRT:74		74	M
3	PfCRT:75		75	N
4	PfCRT:76		76	K
5	PfCRT:93		93	T
6	PfCRT:97		97	H
7	PfCRT:218		218	I
8	PfCRT:220		220	A
9	PfCRT:271		271	Q
10	PfCRT:326		326	N
11	PfCRT:333		333	T
12	PfCRT:353		353	G
13	PfCRT:356		356	I
14	PfCRT:371		371	R
15	PfDHFR:16	PfDHFR	16	A
16	PfDHFR:51		51	N
17	PfDHFR:59		59	C
18	PfDHFR:108		108	S
19	PfDHFR:164		164	I
20	PfDHFR:306		306	S
21	PfDHPS:436	PfDHPS	436	S
22	PfDHPS:437		437	A
23	PfDHPS:540		540	K
24	PfDHPS:581		581	A
25	PfDHPS:613		613	A
26	PfEXO:415	PfEXO	415	E
27	PfMDR1:86	PfMDR1	86	N
28	PfMDR1:184		184	Y
29	PfMDR1:1034		1034	S
30	PfMDR1:1042		1042	N
31	PfMDR1:1226		1226	F
32	PfMDR1:1246		1246	D
33	PfARPS10:127	PfARPS10	127	V
34	PfARPS10:128		128	D
35	PfFD:193	PfFerredoxin	193	D
36	PfMDR2:484	PfMDR2	484	T

Table A2 - Associated and validated *kelch13* resistance mutations. This classification, based on published studies, is provided by WHO. Several other less frequent variants were reported to be associated with in vivo or in vitro tests, or both: M476I; C469Y; A481V; S522C; N537I; N537D; G538V; M579I; D584V; H719N.

kelch13 mutation	Classification
E252Q	Not associated
P441L	Candidate
F446I	Candidate
G449A	Candidate
N458Y	Candidate
Y493H	Validated
R539T	Validated
I543T	Validated
P553L	Candidate
R561H	Validated
V568G	Candidate
P574L	Candidate
A578S	Not associated
C580Y	Validated
A675V	Candidate

Table A3 – List of the 101 barcode SNPs. The SNPs are presented in the order in which they are concatenated in the barcode. For each SNP, we show: the chromosome and position within the chromosome; the reference (3D7 strain) and non-reference (alternative) alleles; the ID and description of the gene containing the SNP; whether the SNP is synonymous or non-synonymous; the amino acid mutations caused by the SNP in that gene; and the coding strand for this gene (+ = sense, - = antisense).

Num	Chr	Pos	Ref	Gene Id	Gene Description	MutType	MutName	Strand
1	Pf3D7_02_v3	376222	A	PF3D7_0209000	6-cysteine protein (P230)	N	K1929E	+
2	Pf3D7_02_v3	470013	G	PF3D7_0211700	protein kinase, putative (TKL1)	N	G75E	+
3	Pf3D7_03_v3	656861	T	PF3D7_0316200	conserved Plasmodium protein, unknown function	S	129V	-
4	Pf3D7_04_v3	110442	C	PF3D7_0401900	acyl-CoA synthetase (ACS6)	N	G285E	-
5	Pf3D7_04_v3	881571	A	PF3D7_0419900	phosphatidylinositol 4-kinase, putative	S	1081R	+
6	Pf3D7_05_v3	350933	G	PF3D7_0508500	guanidine nucleotide exchange factor (RCC1)	S	1369N	-
7	Pf3D7_05_v3	369740	T	PF3D7_0508900	conserved Plasmodium protein, unknown function	S	907L	+
8	Pf3D7_06_v3	900278	G	PF3D7_0622100	conserved Plasmodium protein, unknown function	N	P696S	-
9	Pf3D7_07_v3	1044052	T	PF3D7_0724700	conserved Plasmodium protein, unknown function	S	686K	-
10	Pf3D7_08_v3	1314831	G	PF3D7_0830800	surface-associated interspersed gene 8.2 (SURFIN8.2)	S	1342K	+
11	Pf3D7_08_v3	413067	A	PF3D7_0808100	conserved Plasmodium protein, unknown function	S	1044V	+
12	Pf3D7_09_v3	900277	A	PF3D7_0922100	ubiquitin-like protein, putative	S	1534E	+
13	Pf3D7_11_v3	1018899	T	PF3D7_1126100	ThiF family protein, putative	S	1199L	-
14	Pf3D7_11_v3	1815412	C	PF3D7_1145800	conserved Plasmodium protein, unknown function	N	E765Q	-
15	Pf3D7_13_v3	1056452	T	PF3D7_1325400	conserved Plasmodium protein, unknown function	S	1234D	+
16	Pf3D7_13_v3	1466422	G	PF3D7_1335900	sporozoite surface protein 2+(TRAP)	N	N66K	-
17	Pf3D7_14_v3	137622	T	PF3D7_1403800	nuclear formin-like protein (MISFIT)	S	1179V	-
18	Pf3D7_14_v3	2164225	A	PF3D7_1452600	conserved Plasmodium protein, unknown function	S	2830S	+
19	Pf3D7_01_v3	145515	T	PF3D7_0103300	conserved Plasmodium protein, unknown function	S	294I	-
20	Pf3D7_03_v3	548178	C	PF3D7_0313400	conserved Plasmodium protein, unknown function	N	R2L	-
21	Pf3D7_04_v3	1102392	A	PF3D7_0424400	surface-associated interspersed gene 4.2, (SURFIN4.2)	N	E808D	+
22	Pf3D7_04_v3	139051	G	PF3D7_0402300	reticulocyte binding protein homologue 1+(RH1)	N	K438N	+
23	Pf3D7_04_v3	286542	G	PF3D7_0405300	sequestrin	N	H586N	-
24	Pf3D7_04_v3	529500	G	PF3D7_0411900	DNA polymerase alpha	S	1477Y	-
25	Pf3D7_05_v3	796714	A	PF3D7_0519300	cytochrome c+oxidase assembly protein, putative	S	396K	+
26	Pf3D7_07_v3	1256331	C	PF3D7_0729500	mRNA (N6-adenosine)-methyltransferase, putative	N	L321F	+
27	Pf3D7_07_v3	461139	G	PF3D7_0710100	conserved Plasmodium protein, unknown function	N	M361I	+
28	Pf3D7_07_v3	619957	G	PF3D7_0713600	mitochondrial ribosomal protein S5 precursor, putative	S	675R	+
29	Pf3D7_08_v3	417335	C	PF3D7_0808200	plasmepsin X	N	R244K	-
30	Pf3D7_09_v3	163977	C	PF3D7_0903500	conserved Plasmodium protein, unknown function	S	403D	+

Num	Chr	Pos	Ref	Gene Id	Gene Description	MutType	MutName	Strand
31	Pf3D7_10_v3	317581	A	PF3D7_1007900	eukaryotic translation initiation factor 3+subunit 7	S	311I	+
32	Pf3D7_10_v3	336274	A	PF3D7_1008100	conserved Plasmodium protein, unknown function	N	I1677V	+
33	Pf3D7_11_v3	1020397	C	PF3D7_1126100	ThiF family protein, putative	N	G700E	-
34	Pf3D7_11_v3	1294107	C	PF3D7_1133400	apical membrane antigen 1+(AMA1)	S	84A	+
35	Pf3D7_11_v3	1935227	T	PF3D7_1148700	Plasmodium exported protein (PHISTc), unknown function	N	R73S	-
36	Pf3D7_11_v3	477922	C	PF3D7_1112500	conserved Plasmodium protein, unknown function	N	H147Y	+
37	Pf3D7_12_v3	1663492	A	PF3D7_1239800	conserved Plasmodium protein, unknown function	S	1014E	+
38	Pf3D7_12_v3	2171901	T	PF3D7_1253100	Plasmodium exported protein (PHISTa), unknown function	N	V140D	+
39	Pf3D7_13_v3	1233218	T	PF3D7_1329100	myosin C+(MyoC)	N	N277S	-
40	Pf3D7_13_v3	1867630	G	PF3D7_1346400	conserved Plasmodium protein, unknown function	N	M4911I	+
41	Pf3D7_13_v3	2377887	C	PF3D7_1359600	conserved Plasmodium protein, unknown function	S	2002S	+
42	Pf3D7_14_v3	2355751	T	PF3D7_1457400	conserved Plasmodium protein, unknown function	N	H1589Q	+
43	Pf3D7_14_v3	3046108	C	PF3D7_1474400	conserved Plasmodium protein, unknown function	S	417V	+
44	Pf3D7_02_v3	529709	T	PF3D7_0212800	multidrug efflux pump, putative	N	F487L	+
45	Pf3D7_02_v3	714480	T	PF3D7_0217200	conserved Plasmodium protein, unknown function	N	D258G	-
46	Pf3D7_03_v3	155697	A	PF3D7_0302900	exportin 1, putative	S	150P	-
47	Pf3D7_04_v3	1037656	A	PF3D7_0422500	pre-mRNA-splicing helicase BRR2, putative (BRR2)	S	2776I	+
48	Pf3D7_04_v3	648101	G	PF3D7_0414200.1	calmodulin-like protein	S	51V	-
49	Pf3D7_05_v3	1204155	A	PF3D7_0529400.1	conserved Plasmodium protein, unknown function	S	1338I	+
50	Pf3D7_06_v3	1282691	G	PF3D7_0630600	conserved Plasmodium protein, unknown function	S	803K	+
51	Pf3D7_06_v3	1289212	A	PF3D7_0630800	conserved Plasmodium protein, unknown function	S	125T	+
52	Pf3D7_07_v3	1066698	G	PF3D7_0725100	conserved P. membrane protein, unknown function	N	G483S	+
53	Pf3D7_07_v3	1213486	G	PF3D7_0728200	actin-like protein, putative	N	S543N	+
54	Pf3D7_07_v3	704373	A	PF3D7_0716000	RNA binding protein, putative	S	389E	+
55	Pf3D7_08_v3	1313202	T	PF3D7_0830800	surface-associated interspersed gene 8.2 (SURFIN8.2)	S	799F	+
56	Pf3D7_08_v3	339406	A	PF3D7_0806300	ferlin like protein, putative	S	1283C	-
57	Pf3D7_08_v3	701557	T	PF3D7_0814500	conserved Plasmodium protein, unknown function	S	394G	+
58	Pf3D7_09_v3	452690	A	PF3D7_0910000	SET domain protein, putative (SET4)	S	1018I	-
59	Pf3D7_09_v3	599655	G	PF3D7_0914000	pseudouridylate synthase, putative	N	E654D	+
60	Pf3D7_10_v3	1383789	A	PF3D7_1034900	methionine-tRNA ligase, putative	N	N114H	+
61	Pf3D7_10_v3	1385894	C	PF3D7_1034900	methionine-tRNA ligase, putative	S	815P	+
62	Pf3D7_11_v3	1006911	A	PF3D7_1125700	kelch protein, putative	N	D124E	-
63	Pf3D7_11_v3	1295068	G	PF3D7_1133400	apical membrane antigen 1+(AMA1)	N	E405K	+
64	Pf3D7_11_v3	1802201	G	PF3D7_1145400	dynamin-like protein (DYN1)	S	450S	-
65	Pf3D7_12_v3	1667593	T	PF3D7_1239800	conserved Plasmodium protein, unknown function	S	2381N	+
66	Pf3D7_12_v3	1934745	G	PF3D7_1246500	conserved Plasmodium protein, unknown function	S	241L	-

Num	Chr	Pos	Ref	Gene Id	Gene Description	MutType	MutName	Strand
67	Pf3D7_12_v3	858501	C	PF3D7_1221400	membrane skeletal protein, putative (ALV3)	N	Q469K	+
68	Pf3D7_13_v3	1419519	T	PF3D7_1335100	merozoite surface protein 7+(MSP7)	N	Q208R	-
69	Pf3D7_13_v3	159086	A	PF3D7_1303100	methyltransferase-like protein, putative	S	21R	+
70	Pf3D7_13_v3	2161975	T	PF3D7_1354200	inositol-polyphosphate 5-phosphatase, putative	N	D252V	-
71	Pf3D7_13_v3	2573828	A	PF3D7_1364200	conserved Plasmodium protein, unknown function	N	I1153M	-
72	Pf3D7_13_v3	388365	A	PF3D7_1308400	conserved Plasmodium protein, unknown function	N	S1236R	-
73	Pf3D7_14_v3	2625887	C	PF3D7_1464700	ATP synthase (C/AC39) subunit, putative	N	M238I	-
74	Pf3D7_14_v3	3126219	C	PF3D7_1475900	conserved Plasmodium protein, unknown function	N	S628F	+
75	Pf3D7_14_v3	438592	A	PF3D7_1410900	conserved Plasmodium protein, unknown function	N	N348T	+
76	Pf3D7_01_v3	179347	A	PF3D7_0104100	conserved P. membrane protein, unknown function	S	311G	+
77	Pf3D7_01_v3	180554	G	PF3D7_0104100	conserved P. membrane protein, unknown function	N	D714N	+
78	Pf3D7_01_v3	283144	C	PF3D7_0106700	small ribosomal subunit assembling AARP2 protein	N	H664D	+
79	Pf3D7_01_v3	535211	C	PF3D7_0113800	DBL containing protein, unknown function	S	2521F	+
80	Pf3D7_02_v3	839620	T	PF3D7_0220800	cytoadherence linked asexual protein 2+(CLAG2)	S	260L	+
81	Pf3D7_04_v3	426436	A	PF3D7_0408900.1	peptidase, M22 family, putative	N	D560A	+
82	Pf3D7_04_v3	531138	G	PF3D7_0411900	DNA polymerase alpha	N	A992E	-
83	Pf3D7_04_v3	891732	A	PF3D7_0419900	phosphatidylinositol 4-kinase, putative	N	R4468S	+
84	Pf3D7_05_v3	172801	G	PF3D7_0504400	ATP-dependent helicase, putative	N	E218K	+
85	Pf3D7_06_v3	574938	A	PF3D7_0613800	transcription factor with AP2 domain(s) (ApiAP2)	N	I2934L	+
86	Pf3D7_07_v3	1308383	C	PF3D7_0730500	conserved Plasmodium protein, unknown function	N	G1945R	-
87	Pf3D7_07_v3	1358910	A	PF3D7_0731500	erythrocyte binding antigen-175 (EBA175)	N	K286E	+
88	Pf3D7_07_v3	1359218	A	PF3D7_0731500	erythrocyte binding antigen-175 (EBA175)	N	K388N	+
89	Pf3D7_07_v3	635985	T	PF3D7_0713900	conserved Plasmodium protein, unknown function	N	T598A	-
90	Pf3D7_08_v3	1056829	C	PF3D7_0824200	conserved Plasmodium protein, unknown function	N	L474I	+
91	Pf3D7_08_v3	150033	T	PF3D7_0802000	glutamate dehydrogenase, putative (GDHc)	S	1315I	+
92	Pf3D7_08_v3	399774	C	PF3D7_0807800	proteasome subunit alpha type 5, putative	S	421K	-
93	Pf3D7_09_v3	1379145	G	PF3D7_0935400	cytoadherence linked protein	N	R398Q	+
94	Pf3D7_10_v3	1386850	C	PF3D7_1035000	U2 snRNA/tRNA pseudouridine synthase, putative	S	927K	-
95	Pf3D7_11_v3	1935031	T	PF3D7_1148700	Plasmodium exported protein (PHISTc), unknown function	N	I139L	-
96	Pf3D7_11_v3	408668	T	PF3D7_1110200	pre-mRNA-processing factor 6, putative (PRPF6)	S	1058I	+
97	Pf3D7_11_v3	828596	T	PF3D7_1121800	petidase, M16 family	N	K240E	-
98	Pf3D7_12_v3	857245	A	PF3D7_1221400	membrane skeletal protein, putative (ALV3)	N	E50G	+
99	Pf3D7_14_v3	107014	G	PF3D7_1402900	conserved Plasmodium protein, unknown function	S	215K	+
100	Pf3D7_14_v3	1757603	A	PF3D7_1442900	guanine nucleotide exchange factor, putative	N	D1365G	+
101	Pf3D7_14_v3	2733656	C	PF3D7_1466800	conserved Plasmodium protein, unknown function	S	557C	+

References

1. Jacob CG, Thuy-Nhien N, Mayxay M, et al. Genetic surveillance in the Greater Mekong subregion and South Asia to support malaria control and elimination. *Elife* 2021; **10**.
2. Galinsky K, Valim C, Salmier A, et al. COIL: a methodology for evaluating malarial complexity of infection using likelihood from single nucleotide polymorphism data. *Malaria journal* 2015; **14**: 4.
3. Chang HH, Worby CJ, Yeka A, et al. THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. *PLoS Comput Biol* 2017; **13**(1): e1005348.
4. Ariey F, Witkowski B, Amaratunga C, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. *Nature* 2014; **505**(7481): 50-5.
5. Ashley EA, Dhorda M, Fairhurst RM, et al. Spread of artemisinin resistance in Plasmodium falciparum malaria. *N Engl J Med* 2014; **371**(5): 411-23.
6. MalariaGEN Plasmodium falciparum Community Project. Genomic epidemiology of artemisinin resistant malaria. *Elife* 2016; **5**.
7. Miotto O, Amato R, Ashley EA, et al. Genetic architecture of artemisinin-resistant Plasmodium falciparum. *Nat Genet* 2015; **47**(3): 226-34.
8. Wellems TE, Plowe CV. Chloroquine-resistant malaria. *The Journal of Infectious Diseases* 2001; **184**(6): 770-6.
9. Foote SJ, Kyle DE, Martin RK, et al. Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in Plasmodium falciparum. *Nature* 1990; **345**: 255-8.
10. Fidock DA, Nomura T, Talley AK, et al. Mutations in the P. falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. *Molecular Cell* 2000; **6**(4): 861-71.
11. Price RN, Uhlemann AC, Brockman A, et al. Mefloquine resistance in Plasmodium falciparum and increased pfmdr1 gene copy number. *Lancet* 2004; **364**(9432): 438-47.
12. Veiga MI, Dhingra SK, Henrich PP, et al. Globally prevalent PfMDR1 mutations modulate Plasmodium falciparum susceptibility to artemisinin-based combination therapies. *Nature Communications* 2016; **7**: 11553.
13. Malmberg M, Ferreira PE, Tarning J, et al. Plasmodium falciparum drug resistance phenotype as assessed by patient antimalarial drug levels and its association with pfmdr1 polymorphisms. *The Journal of Infectious Diseases* 2013; **207**(5): 842-7.
14. Venkatesan M, Gadalla NB, Stepniewska K, et al. Polymorphisms in Plasmodium falciparum chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for P. falciparum malaria after artemether-lumefantrine and artesunate-amodiaquine. *The American journal of tropical medicine and hygiene* 2014; **91**(4): 833-43.
15. Nsobya SL, Kiggundu M, Nanyunja S, Joloba M, Greenhouse B, Rosenthal PJ. In vitro sensitivities of Plasmodium falciparum to different antimalarial drugs in Uganda. *Antimicrobial agents and chemotherapy* 2010; **54**(3): 1200-6.
16. Amato R, Lim P, Miotto O, et al. Genetic markers associated with dihydroartemisinin-piperaquine failure in Plasmodium falciparum malaria in Cambodia: a genotype-phenotype association study. *Lancet Infect Dis* 2017; **17**(2): 164-73.
17. Peterson DS, Walliker D, Wellems TE. Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. *Proceedings of the National Academy of Sciences of the United States of America* 1988; **85**(23): 9114-8.
18. Gregson A, Plowe CV. Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev* 2005; **57**(1): 117-45.
19. Picot S, Olliaro P, de Monbrison F, Bienvenu AL, Price RN, Ringwald P. A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malaria journal* 2009; **8**: 89.