

Markers to phenotype mapping - SOP

v1.2 - February 2018

This document describes how we assign phenotypes to individual samples given their genotypes at relevant loci. It doesn't aim at being definitive and exhaustive, but choices made have to be sensible, clear and transparent. Feedback from the community will inform future versions of this document.

Data

The procedures described here are particularly focused on genotypes calls and copy number alleles as described in the Pf6 manuscript (in preparation). They can obviously be applied to other data sources, where relevant, without compromising the logic. The most relevant alternatives are:

- genotypes calls from Agena and, in future, amplicon sequencing;
- copy number estimated by qPCR;
- genotypes calls from Olivo's "GRC BAM genotyper", where a BAM file is available but not the corresponding VCF.

Heterozygous and missing genotypes are not always explicitly handled at the moment, in particular when combinations of SNPs are required. For that reason, the order these rule are applied matters.

Drugs

For the time being, we only consider the following six major drugs:

1. Chloroquine
2. Artemisinin
3. Piperaquine
4. Mefloquine
5. Pyrimethamine
6. Sulfadoxine

We additionally consider the following drugs, but we will **not explicitly report** on them yet:

7. Sulfadoxine-Pyrimethamine (SP)
8. Proguanil
9. Atovaquone

We are **NOT** considering:

1. Explicit combination therapies, other than SP
2. Lumefantrine
3. Amodiaquine
4. Quinine

Chloroquine

Locus utilized: PF3D7_0709000 (*crt*), codon 76

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|-------------------|-------------------------------------|----------------|----------------------|----------------------|
| 76-Het | <code>crt_76[K]</code> contains “,” | Het | Resistant/het | Resistant |
| 76-Missing | <code>crt_76[K]</code> == “-” | Missing | Undetermined/missing | Undetermined |
| K76 | <code>crt_76[K]</code> == “K” | WT | Sensitive | Sensitive |
| 76T | <code>crt_76[K]</code> == “T” | Mutant | Resistant [1] | Resistant |
| 76-nonT | otherwise | Unknown mutant | Undetermined/unknown | Undetermined |

References:

1. [https://doi.org/10.1016/S0140-6736\(01\)06040-8](https://doi.org/10.1016/S0140-6736(01)06040-8)

Notes:

- There are evidences linking SNPs in the gene *mdr1* with some level of chloroquine resistance. However their impact on clinical phenotypes and the interactions with *crt* mutations is not fully understood hence the decision to not use that in the phenotype determination.
- Work in progress from Rutledge et al. shows the effect of other mutations on the level or resistance (super-resistance lineages). This information could be incorporated in future versions.

Artemisinin

Locus utilized: PF3D7_1343700 (*kelch13*), codons 349-726

Mapping:

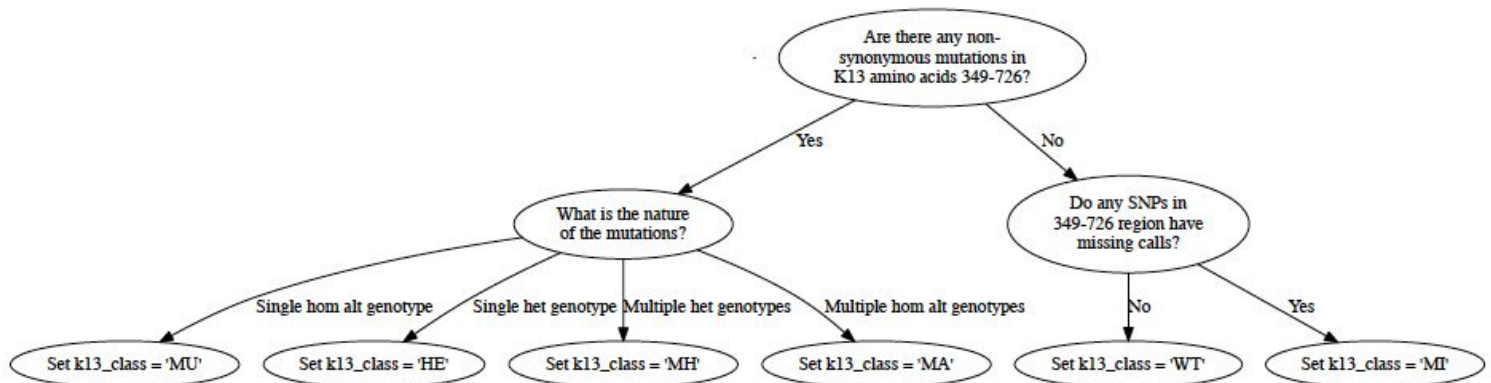
| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|--|--|--------------------------|----------------------|----------------------|
| Missing | k13_class == "MI" | Missing | Undetermined/missing | Undetermined |
| None | k13_class == "WT" | Wild-type | Sensitive | Sensitive |
| 458Y, 493H, 539T, 543T, 561H, 580Y | k13_class == "MU" and k13_allele == {"N458Y", "Y493H", "R539T", "I543T", "R561H", "C580Y"} | Mutant - validated | Resistant [1] | Resistant |
| 441L, 446I, 449A, 538V, 553L, 568G, 574L, 675V | k13_class == "MU" and k13_allele == {"P441L", "F446I", "G449A", "G538V", "P553L", "V568G", "P574L", "A675V"} | Mutant - candidate | Resistant/candidate | Resistant |
| 476I, 469Y, 469F, 476I, 479I, 481V, 515K, 522C, 527L, 537I, 537D, 538V, 575K, 579I, 584V, 667T, 673I, 719N | k13_class == "MU" and k13_allele == {"M476I", "C469Y", "C469F", "M476I", "K479I", "A481V", "R515K", "S522C", "P527L", "N537I", "N537D", "G538V", "R575K", "M579I", "D584V", "P667T", "F673I", "H719N"} | Mutant - associated | Resistant/associated | Resistant |
| 458Y, 493H, 539T, 543T, 561H, 580Y, 441L, 446I, 449A, 538V, 553L, 568G, 574L, 675V, 476I, 469Y, 469F, 476I, 479I, 481V, 515K, 522C, 527L, 537I, 537D, 538V, 575K, 579I, 584V, 667T, 673I, 719N | k13_class == {"HE", "MH"} and k13_allele contains {"n458y", "y493h", "r539t", "i543t", "r561h", "c580y", "p441l", "f446i", "g449a", "g538v", "p553l", "v568g", "p574l", "a675v", "m476i", "c469y", "c469f", "m476i", "k479i", "a481v", "r515k", "s522c", "p527l", "n537i", "n537d", "g538v", "r575k", "m579i", "d584v", "p667t", "f673i", "h719n"} | Mutant - heterozygous | Resistant/het | Resistant |
| 252Q, 578S | k13_class == {"MU", "HE"} and k13_allele contains {"E252Q", "A578S"} | Mutant - not associated | Sensitive/mutant [1] | Sensitive |
| otherwise | otherwise | Mutant - not in WHO list | Undetermined/unknown | Undetermined |

References:

1. <http://apps.who.int/iris/bitstream/10665/255213/1/WHO-HTM-GMP-2017.9-eng.pdf?ua=1>

Notes:

- We can assume “possibly resistant” to be resistant for the time being. They should represent a minority of cases anyway.
- A sample has k13_class == “MI” if and only if there is at least one pass variant in the VCF that has a missing genotype (./.) and there are no variants leading to non-synonymous mutations.
- The algorithm looks like:



Piperaquine

Loci utilized: amplification of both PF3D7_1408000 (*plasmepsin 2*) and PF3D7_1408100 (*plasmepsin 3*)

Mapping:

| Change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|-----------------|---------------|----------------|----------------------|----------------------|
| Missing | dup_pm2 == -1 | Missing | Undetermined/missing | Undetermined |
| Single copy | dup_pm2 == 0 | Wild-type | Sensitive | Sensitive |
| Multiple copies | dup_pm2 == 1 | Mutant | Resistant [1,2] | Resistant |

References:

1. [http://dx.doi.org/10.1016/S1473-3099\(16\)30409-1](http://dx.doi.org/10.1016/S1473-3099(16)30409-1)
2. [https://doi.org/10.1016/S1473-3099\(16\)30415-7](https://doi.org/10.1016/S1473-3099(16)30415-7)

Notes:

- Where neither sequence nor qPCR data are available, we could use either the exo-E415G (PF3D7_1362500; PF3D7_13_v3:2504560) or the PCR/Agena breakpoint assay, but caveats need to be made explicit.

Mefloquine

Locus utilized: amplification of PF3D7_0523000 (*mdr1*)

Mapping:

| Change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|-----------------|----------------|----------------|----------------------|----------------------|
| Missing | dup_mdr1 == -1 | Missing | Undetermined/missing | Undetermined |
| Single copy | dup_mdr1 == 0 | Wild-type | Sensitive | Sensitive |
| Multiple copies | dup_mdr1 == 1 | Mutant | Resistant [1] | Resistant |

References:

1. [https://doi.org/10.1016/S0140-6736\(04\)16767-6](https://doi.org/10.1016/S0140-6736(04)16767-6)

Notes:

Pyrimethamine

Locus utilized: PF3D7_0417200 (*dhfr*), codon 108

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|-------------------|--------------------------|----------------|----------------------|----------------------|
| 108-Het | dhfr_108[S] contains “ ” | Het | Resistant/het | Resistant |
| 108-Missing | dhfr_108[S] == “-” | Missing | Undetermined/missing | Undetermined |
| S108 | dhfr_108[S] == “S” | WT | Sensitive | Sensitive |
| 108N | dhfr_108[S] == “N” | Mutant | Resistant [1] | Resistant |
| 108-nonN | otherwise | Unknown mutant | Undetermined/unknown | Undetermined |

References:

1. PMID: 3057499

Notes:

- Greater number of mutations in DHFR leads to greater drug resistance, but we are not considering that for the moment since it appears that 108N is necessary and sufficient.

Sulfadoxine

Locus utilized: PF3D7_0810800 (*dhps*), codons 436, 437, 540, 581, 613

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|----------------------------------|--|--------------------------------|-----------------|----------------------|
| S436 + A437 + K540 + A581 + A613 | dhps_436[S] == "S" and dhps_437[G] == "A" and dhps_540[K] == "K" and dhps_581[A] == "A" and dhps_613[A] == "A" | WT | Sensitive | Sensitive |
| 436F + 613S/T | dhps_436[S] == "F" and dhps_613[A] == {"S", "T"} | Mutant | Resistant [1] | Resistant |
| 581G | dhps_581[A] == "G" | Mutant | Resistant [1] | Resistant |
| 437G + 540E | dhps_437[G] == "G" and dhps_540[K] == "E" | Mutant | Resistant [2,3] | Resistant |
| 436F + 613S/T | (dhps_436[S] == "F" and dhps_613[A] contains ",") or (dhps_436[S] contains "," and dhps_613[A] == {"S", "T"}) | Mutant - het | Resistant/het | Resistant |
| 581G | dhps_581[A] contains "," | Mutant - het | Resistant/het | Resistant |
| 437G + 540E | (dhps_437[G] == "G" and dhps_540[K] contains "," and) or (dhps_437[G] contains "," and dhps_540[K] == "E") | Mutant - het | Resistant/het | Resistant |
| otherwise | otherwise | Missing or unknown combination | Undetermined | Undetermined |

References:

1. PMID: 7925353
2. PMID: 9395372
3. DOI 10.1093/emboj/17.14.3807

Notes:

- Heterozygous and missing genotypes are not explicitly handled at the moment.

Sulfadoxine-Pyrimethamine

Locus utilized: PF3D7_0417200 (*dhfr*), codon 51, 59, 108 and PF3D7_0810800 (*dhps*), codons 437, 540

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|---|--|--------------------------------|------------------|----------------------|
| dhfr: N51 + C59 + S108 dhps: A437 + K540 | dhfr_51[N] == "N" and dhfr_59[C] == "C" and dhfr_108[S] == "S" and dhps_437[G] == "A" and dhps_540[K] == "K" | WT | Sensitive | N/A |
| dhfr: 51I + 59R + 108N dhps: 437G | dhfr_51[N] == "I" and dhfr_59[C] == "R" and dhfr_108[S] == "N" and dhps_437[G] == "G" | Mutant | Resistant [1, 2] | N/A |
| dhfr: 59R + 108N dhps: 437G + 540E | dhfr_59[C] == "R" and dhfr_108[S] == "N" and dhps_437[G] == "G" and dhps_540[K] == "E" | Mutant | Resistant [3, 4] | N/A |
| het | Only one het in any of the mutant combinations above | Mutant - het | Resistant/het | N/A |
| otherwise | otherwise | Missing or unknown combination | Undetermined | N/A |

References:

- doi: 10.1128/AAC.44.4.991-996.2000
- DOI: [10.1080/00034980120103234](https://doi.org/10.1080/00034980120103234)
- DOI: <https://doi.org/10.4269/ajtmh.1996.55.209>
- PMID: 11280065

Notes:

- These two drugs are almost invariably used in combination. As a consequence, clinical studies tend to analyse the failure of the therapy rather than the relative effect of Sul and Pyr resistance. We might consider reporting this phenotype instead of the two distinct resistance.
- For the reason stated above, it's possible that a sample will result to be S-P resistant but not Sul or Pyr resistant, and vice versa. These cases will need to be handled manually if we present Sul, Pyr, and S-P as three different phenotypes.

Proguanil

Locus utilized: PF3D7_0417200 (*dhfr*), codon 16 and 108

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|------------------------------|--|---------------------|----------------------|----------------------|
| 16-Missing or 108-Missing | dhfr_16[A] == "-" or dhfr_108[S] == "-" | Missing | Undetermined/missing | N/A |
| A16 + S108 | dhfr_16[A] == "A" and dhfr_108[S] == "S" | WT | Sensitive | N/A |
| A16 + 108N/T or 16V + 108N/S | (dhfr_16[A] == "A" and dhfr_108[S] == {"N", "T"}) or (dhfr_16[A] == "V" and dhfr_108[S] == {"S", "N"}) | Mutant - sensitive | Sensitive/mutant | N/A |
| 16V + 108T | dhfr_16[A] == "V" and dhfr_108[S] == "T" | Double mutant | Resistant [1] | N/A |
| 16V + 108T | dhfr_16[A] contains "V" and dhfr_108[S] contains "T" and exactly one contains "," | Double mutant - het | Resistant/het | N/A |
| A16 + (S108 or 108N) | (dhfr_16[A] == "A" or dhfr_108[S] == {"N", "S"}) and exactly one contains "," | Mutant - het | Sensitive/het | N/A |
| 108-het | dhfr_108[S] contains "N" and "S" | Mutant - het | Sensitive/het | N/A |
| 16V + 108T | dhfr_16[A] contains "V" and dhfr_108[S] contains "T" and both contains "," | Double mutant - het | Undetermined/het | N/A |
| otherwise | otherwise | Unknown combination | Undetermined/unknown | N/A |

References:

1. PMID: 2183221

Notes:

- We only see three samples (plus one failing QC) that are proguanil-resistant. We can mention that in text but it's not necessary to report on the frequency of these mutations at this stage.

Atovaquone

Locus utilized: mal_mito_3 (*cytB*), codon 268

Mapping:

| Amino acid change | Pf6 coding | Interpretation | Phenotype | MalariaGEN analytics |
|-------------------|------------|----------------|---------------|----------------------|
| Y268 | - | Wild-type | Sensitive | N/A |
| 268S/N/C | - | Mutant | Resistant [1] | N/A |

References:

1. PMID: 10898682

Notes:

- We don't observe any mutant in Pf6. Complication derives from the fact that mutant parasites can be present in the infections at extremely low frequency and rapidly selected upon atovaquone exposure (Siegel et al, in preparation). However, we have analysed the read counts for all samples and we haven't found any evidence in this data set supporting low frequency mutants.