*MHap Analysis Report*: Drug Resistance Surveillance - Summary

02\_Plasmodium\_Genotype\_Population\_Analysis by Neafsey Lab

Table of Contents

## Introduction

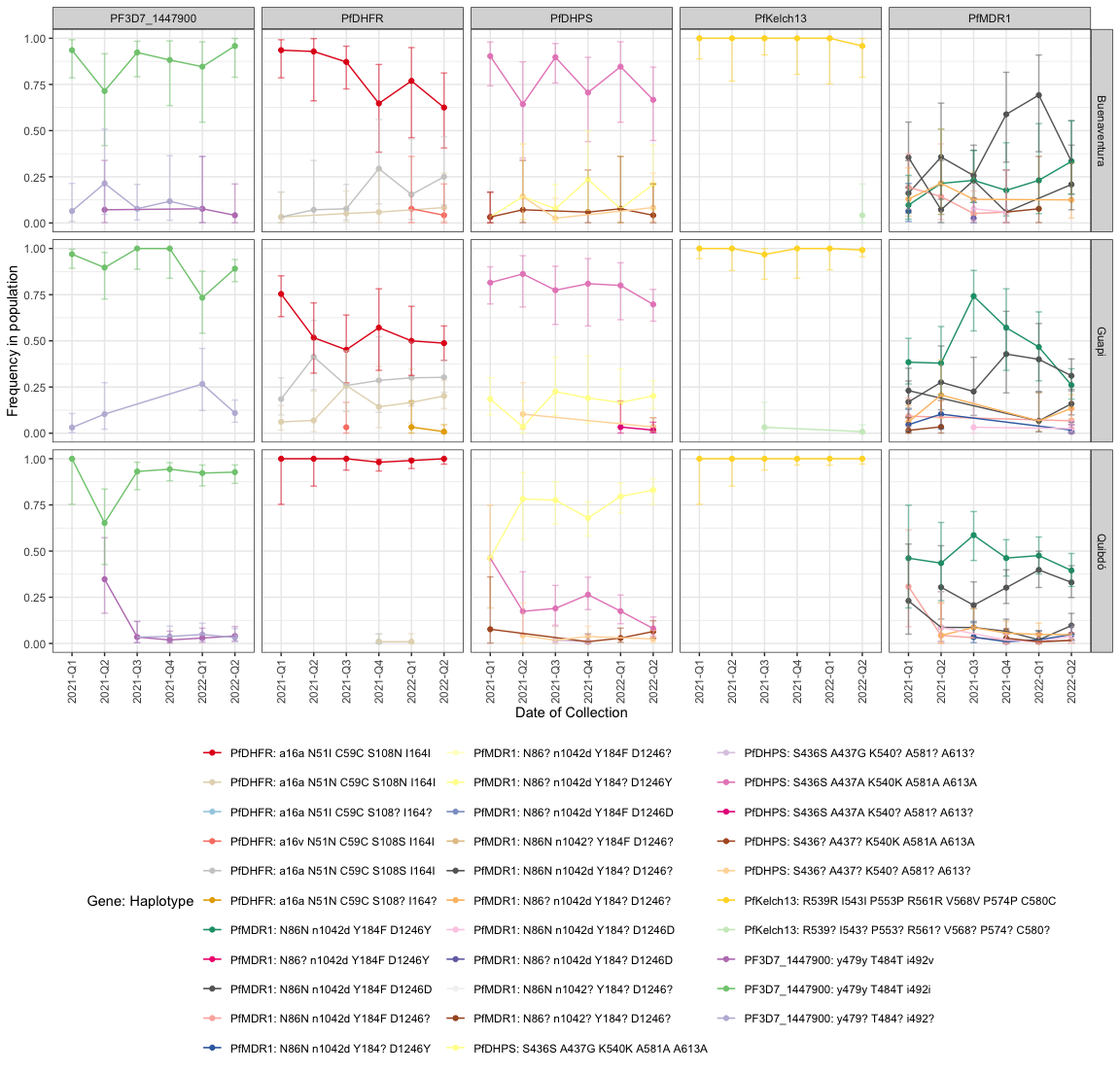
A cornerstone of genomic epidemiological surveillance is tracking alleles (gene variants) associated with resistance againsts antimalarials in Plasmodium parasites. Alleles arise when mutations happen in the sequence of nucleotides that compose a gene. This report relays change in frequency of gene variants in geographical and temporal contexts.

The combination of alleles in a sample is called the **genotype**. The genotypes here reported for each sample are combinations of the alleles present in the sample. This is because (1) patients may be infected with more than one parasite strain at the time of sampling and (2) amplicon sequencing does not infer the set of genes in each individual parasite. In this sense, this report reflects the frequency of alleles at the population level (see further explanations below).

drug\_resistant\_haplotypes\_plot$dna\_mutations %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(  
 buttons = c('csv', 'excel')))  
  
drug\_resistant\_haplotypes\_plot$aa\_mutations %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(dom = 'Blfrtip',  
 buttons = c('csv', 'excel')))  
  
drug\_resistant\_haplotypes\_plot$genotype\_phenotype\_table %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(dom = 'Blfrtip',  
 buttons = c('csv', 'excel')))  
  
drug\_resistant\_haplotypes\_plot$haplotype\_freq\_barplot$data %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(dom = 'Blfrtip',  
 buttons = c('csv', 'excel')))

## Haplotype Frequency

The haplotype is defined as the unique combination of alleles across loci (variant sites across different amplicons) in a gene. In monoclonal samples, one allele is observe at each locus; thus, genotype and haplotype are the same. In polyclonal samples, the complexity of infection (the number of strains present in the sample) equals the maximum number of alleles observed at any loci of a gene. Alleles are sorted in descending order of read depth. When different amplicons within a gene each show multiple alleles, and the number of alleles is different between these amplicons, the major allele on the locus with fewest alleles is used to impute the unknown/ambiguous haplotype(s) (haplotype with least read support). Finally, the total number of haplotypes counted (rather than the total number of infections) is used to calculate the haplotype frequency as haplotypes observed in monoclonal and polyclonal infections are given equal weight. It is for this reason that, as previously stated, this report reflects the frequency of alleles at the population level.



Haplotype frequencies for genes carrying mutations associated with resistance to antimalarials. For each cell, the vertical and horizontal axes show the **haplotype** frequency and the time of the year, respectively. Studied areas are rendered in alphabetic order in the vertical grid. The same applies for genes in the horizontal grid. Red lines indicate the presence of one or more resistance-associated mutations, while blue lines indicate antimalarial-sensitive alleles. The intensity of red indicates the number of resistance-associated mutations. Orange lines represents haplotypes carrying mutations with respect to the reference strain which have not been associated with resistance. Bars represent 95% Confidence Intervals.

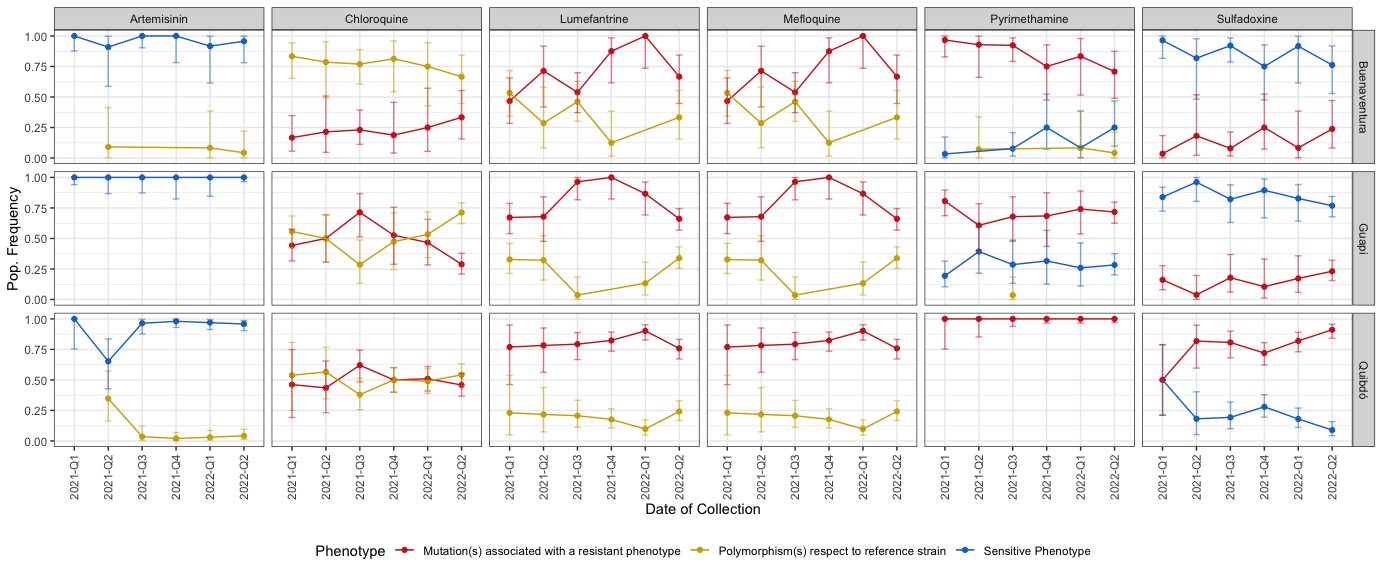
## DRUG PHENOTYPES

Resistance to a particular drug can be conferred by mutations in multiple genes. For this reason, the drug resistance profile of each sample is determined based on the joint analysis of all the genes that have resistance-associated mutations to a particular drug. For example, seven mutations have been associated with resistance to pyrimethamine. Six of these are found in the pfdhfr gene and one in the pfmdr2 gene. To define the possible phenotype of this drug, therefore, the workflow has analyzed the alleles present in each of these seven positions. This section also reports other mutations that occur in either of these two genes and whether amplification was unsuccessful for any of the amplicons.

The drug-resistance profiles are divided into five categories:

* Sensitive Phenotype: All amplicons amplified and no mutations were found with respect to the reference strain and the reference\_allele table.
* Polymorphism(s) with respect to the reference strain: At least one mutation with respect to the reference strain was found, but no resistance-associated mutations occur. Partial haplotypes are also included.
* Mutation(s) associated with a resistant phenotype: At least one resistance-associated mutations was found. Partial haplotypes are also included.
* Partial Haplotype: At least one amplicon has amplified and no mutations occur.
* Missing data: None of the amplicons amplified.

drug\_resistant\_haplotypes\_plot$drug\_phenotype\_table %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(dom = 'Blfrtip',  
 buttons = c('csv', 'excel')))  
  
drug\_resistant\_haplotypes\_plot$drug\_phenotype\_barplot$data %>%  
 DT::datatable(extensions = 'Buttons',  
 options = list(dom = 'Blfrtip',  
 buttons = c('csv', 'excel')))



## Line Plot

Frequency of resistant vs. sensitive phenotypes classified based on amino acid changes in drug resistance-associated genes. For each cell, the vertical and horizontal axes show the **phenotype** frequency and the time of the year, respectively. Studied areas are rendered in alphabetic order in the vertical grid. The same applies for genes in the horizontal grid. Red lines indicate the presence of one or more resistance-associated mutations, while blue lines indicate antimalarial-sensitive alleles. Yellow lines represents haplotypes carrying mutations with respect to the reference strain which have not been associated with resistance. Bars represent 95% Confidence Intervals.

## Frequency of resistance per geographical area.

if(!is.null(Longitude) & !is.null(Latitude)){  
 library(tmap)  
 tmap\_mode('view')  
 drug\_resistant\_haplotypes\_plot$i\_drug\_map   
}

## Glossary

Allele: Genotype: Locus (p. Loci): Haplotype: Complexity of infection: Confidence Interval: