



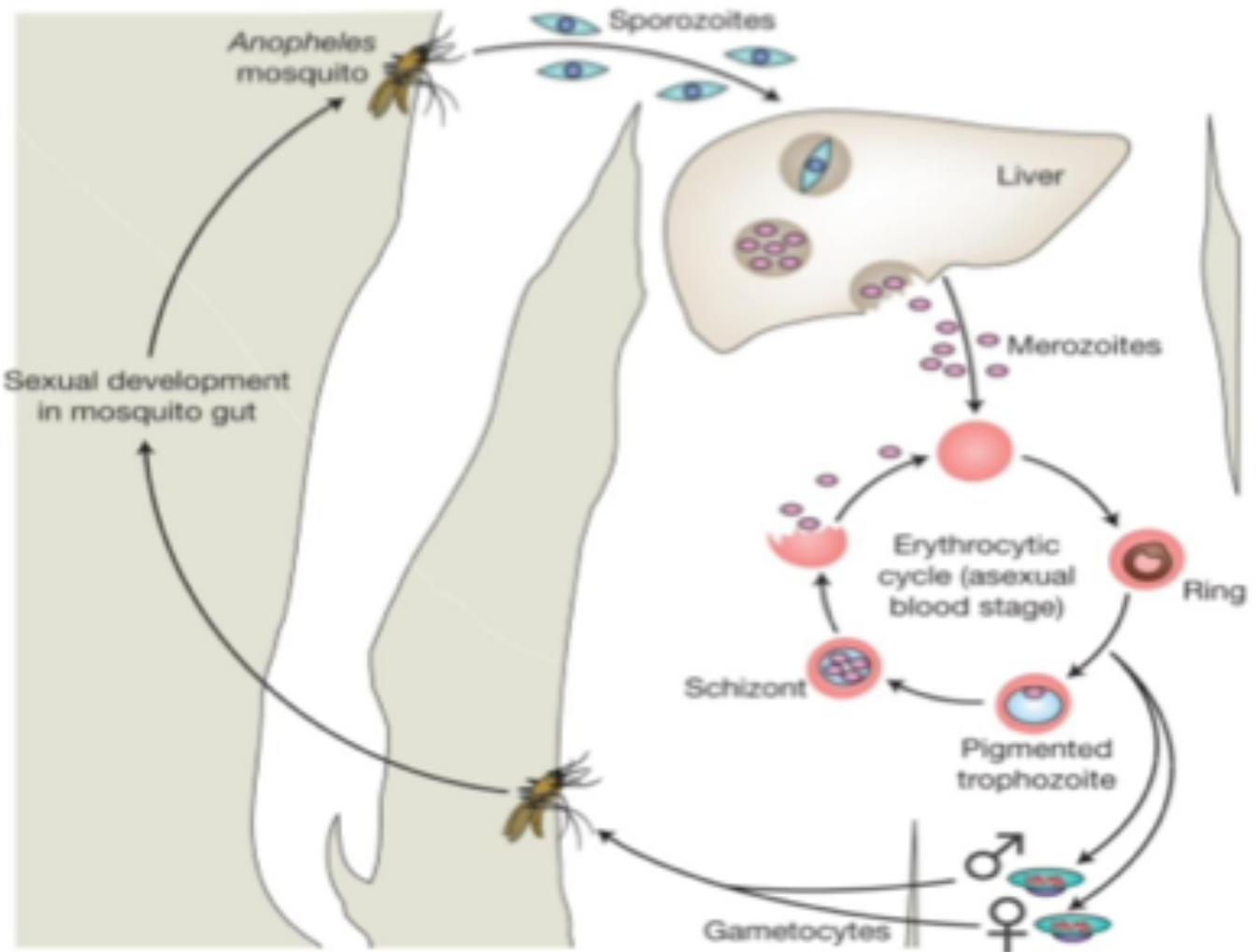
Estimating the micro-indel mutation rate in *Plasmodium falciparum* using genomes from mutation accumulation experiments

Aakanksha Singh
S3 Internship deference
25th February 2021
Supervisor- Antoine Claessens

i

Introduction

- Malaria is a life-threatening disease caused by parasites of the *Plasmodium* genus.
- P. falciparum genome
 - Haploid
 - 23 Mb genome consists of 14 chromosomes
 - 5,300 genes
 - 80% AT rich in coding regions and 90% AT rich in noncoding regions.



Life cycle of P. falciparum

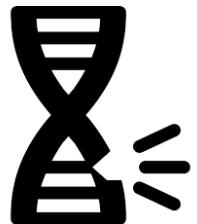


Hypothesis

- Despite Africa bearing the largest disease load, South East Asia has been the hotspot of the evolution of drug resistance in malaria.
- At the parasite genomic level, *P. falciparum* genomes from Africa and Asia are clearly distinct, with thousands of SNPs being specific to a continent (Manske et al., 2012)

Hypothesis

“South East Asian parasites display a faster mutation rate than the African *P. falciparum* parasites, in turn promoting the emergence and spread of drug resistance. ”



Mutations

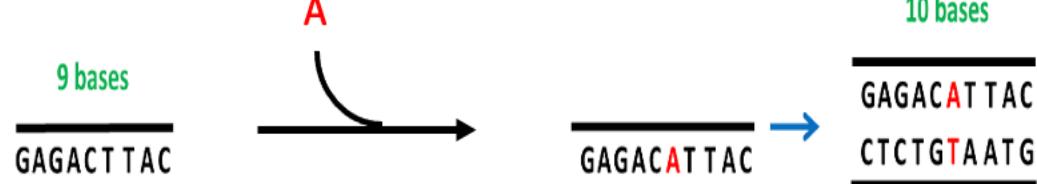
- A mutation is the change in our DNA sequence.
- Types of Mutations
 - Base pair substitutions (SNPs)
 - Insertions and Deletions
 - Copy Number Variants
 - Micro-indels

'Micro-indels are defined as insertions or deletions of typically 1-20 bp found in the microsatellite regions.'

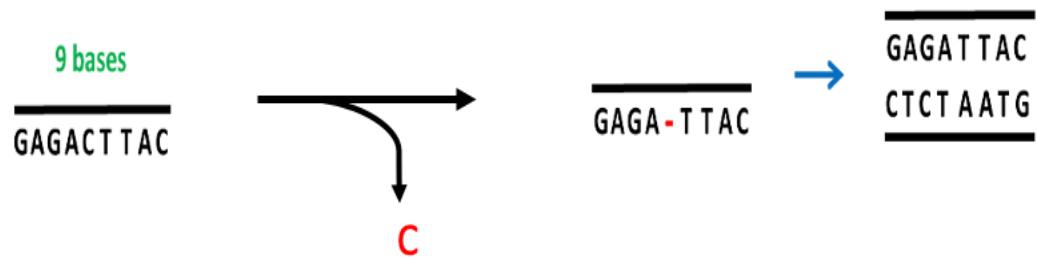
Substitutions



Insertions



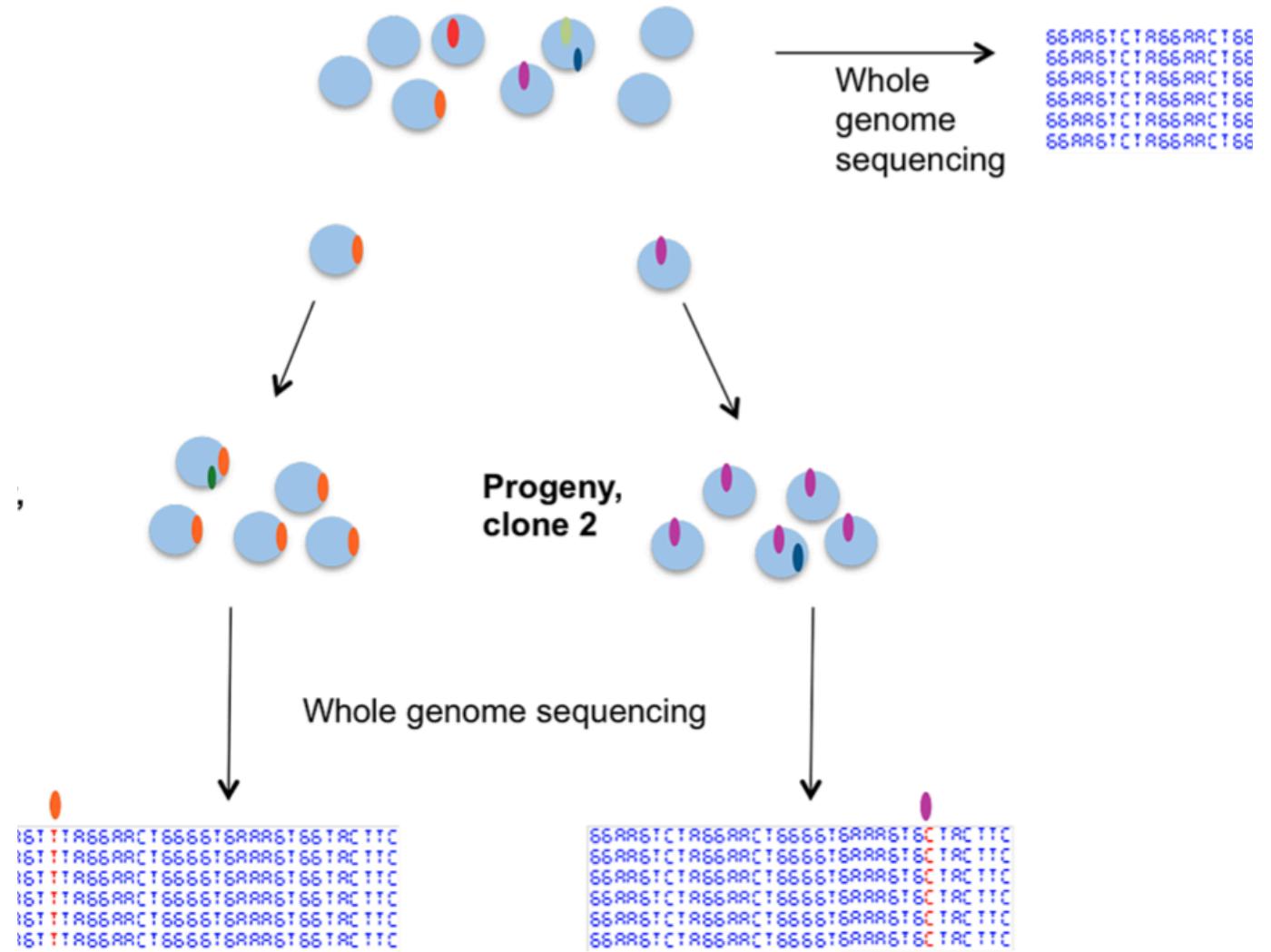
Deletions





De novo mutations

- Each individual *P. falciparum* parasite accumulates some de novo mutations after each round of mitosis.
- De novo mutations are the mutations that are not inherited from the parents.

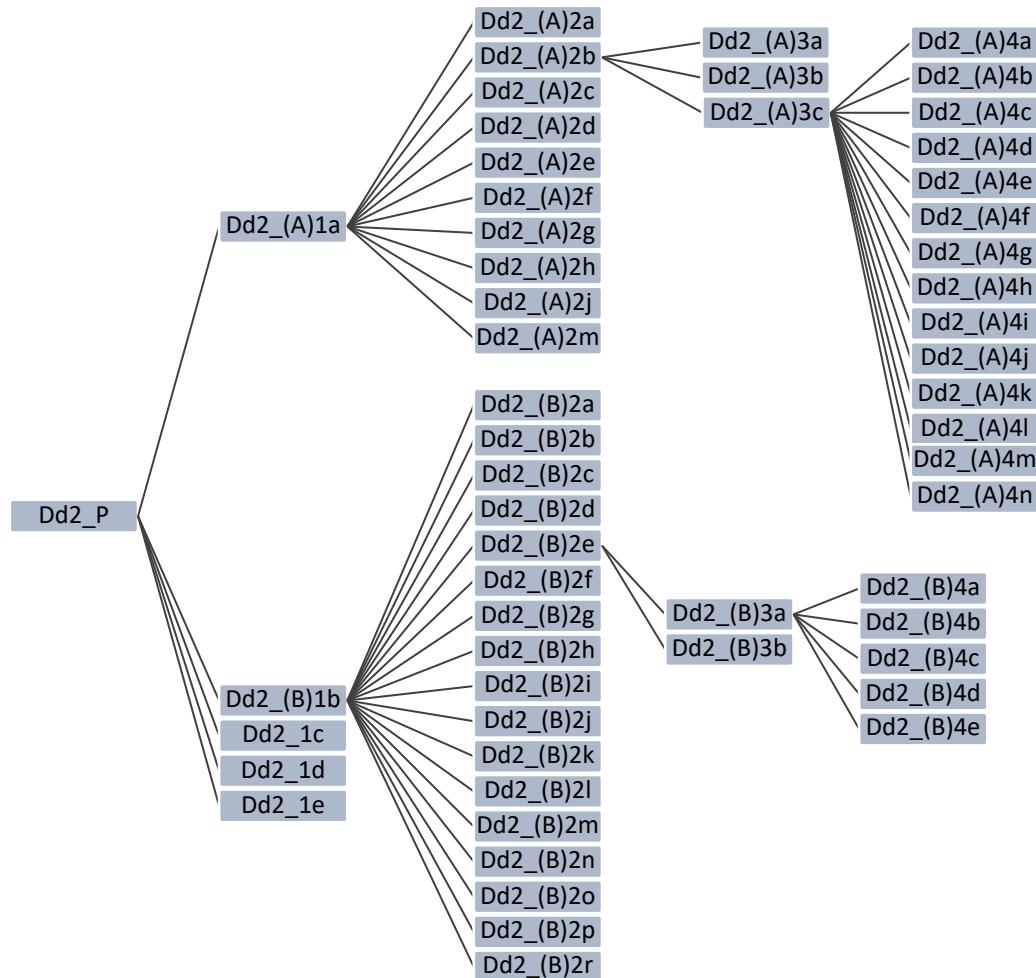


What has already been done?

- Claessens et al.(2014) cultured six distinct *P. falciparum* strains (3D7, W2, Dd2, HB3, KH-01, KH-02) and obtained specific clone trees for each of these strains.
- SNP mutation rate already calculated by Hamilton et al.(2016)

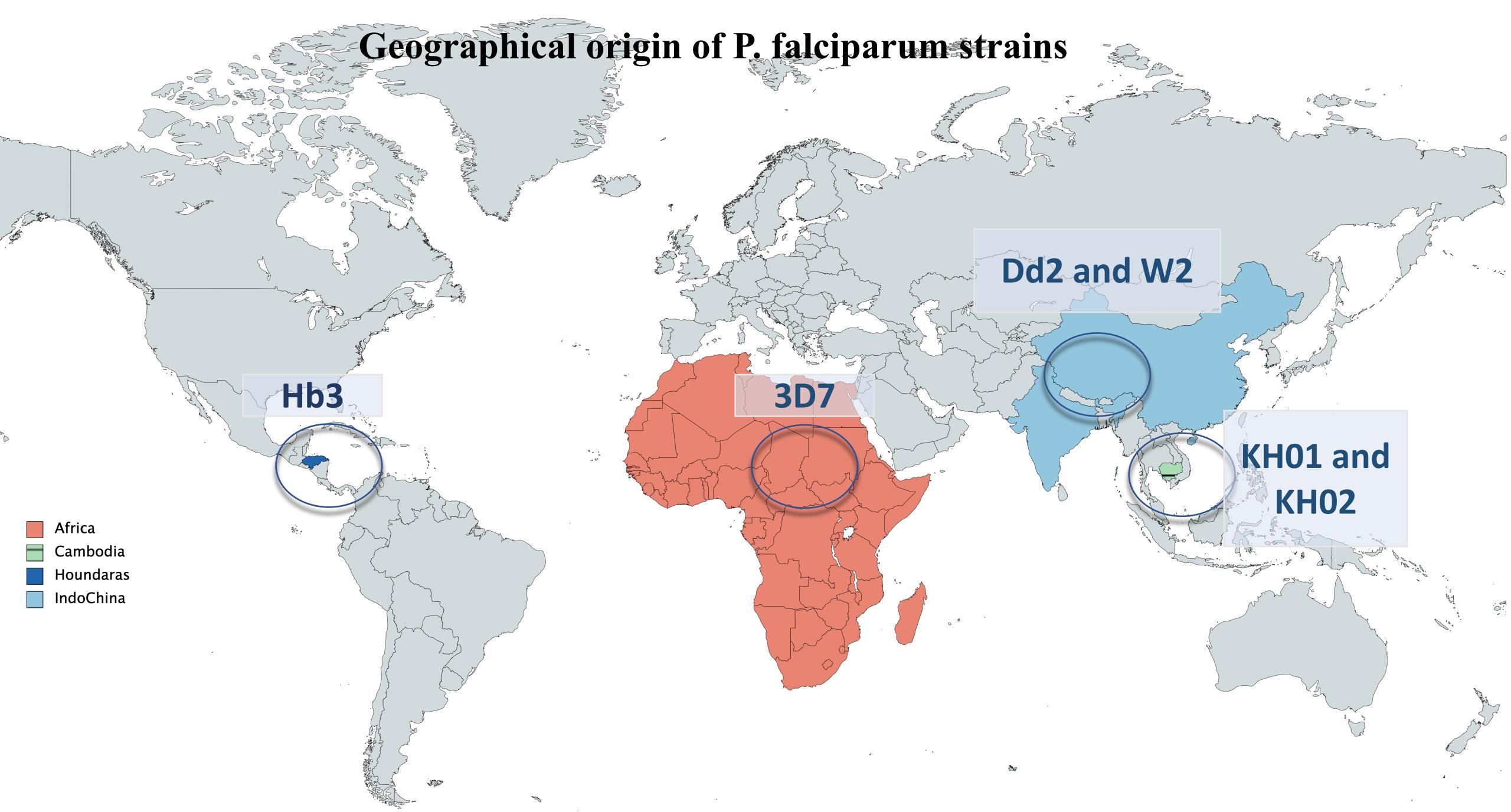
Geographical origin	Strain name	No. of Clonal populations sequenced
Africa	3D7	33
Indochina	W2	19
Indochina	Dd2	56
Honduras	HB3	83
Cambodia	KH-01	60
Cambodia	KH-02	27
	Total	278

Each rectangle represents a genome sequenced from a clonal population



Dd2 Clone Tree (Source- Antoine Claessens)

Geographical origin of *P. falciparum* strains





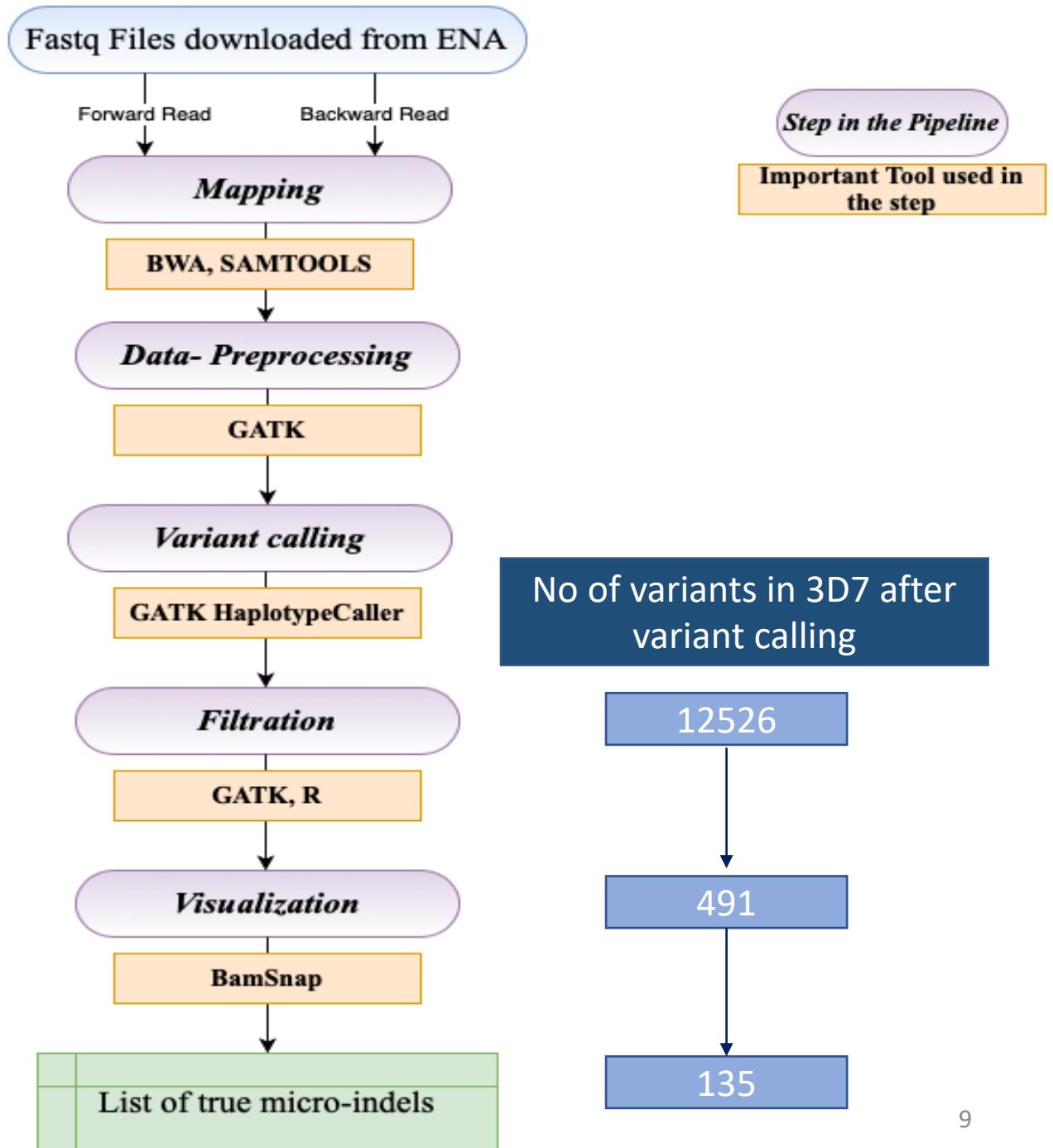
Objective

- Is the micro-indel mutation rate similar across *P. falciparum* strains?
- Could different mutation rates of *P. falciparum* strains from South-East Asia and Africa explain the higher predisposition of South-East Asian strains to evolve drug resistance?



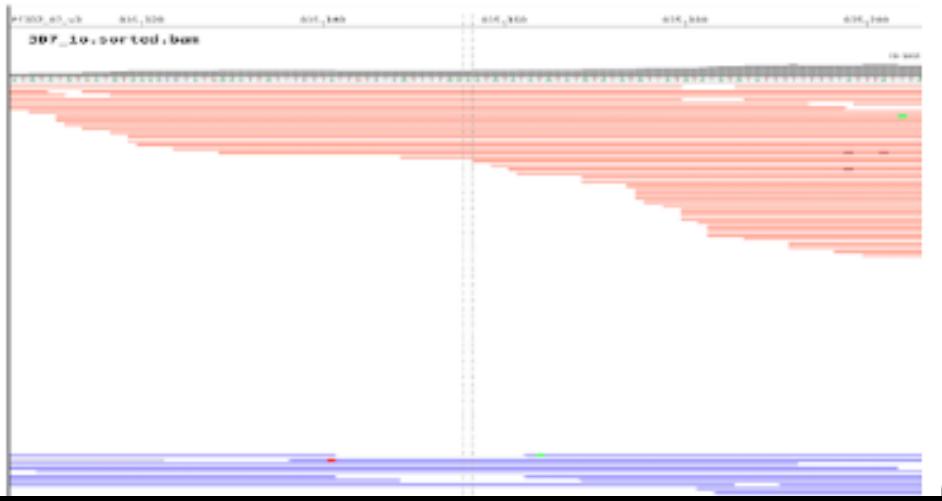
Variant Discovery Pipeline

Pipeline was validated by finding the same number of SNPs (100%) and 135(82 %) indels out of a total of 164 indels as were found by Hamilton et al. (2016) for 3D7.

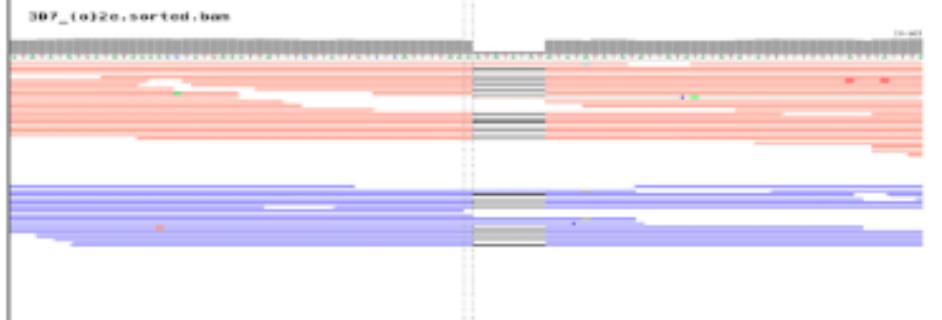


De novo mutation

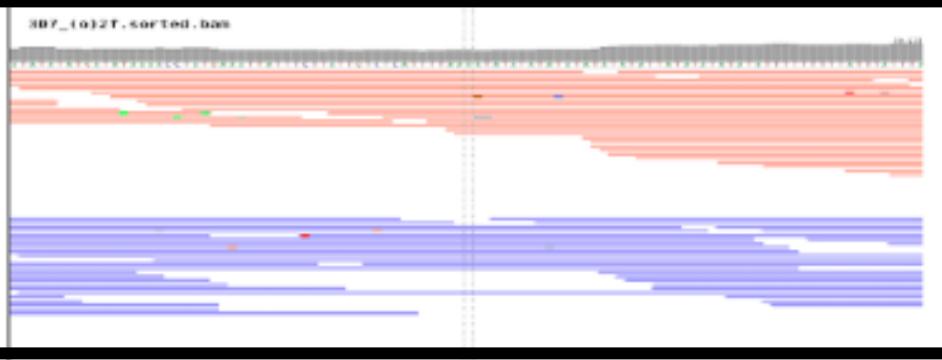
Parent



Sample

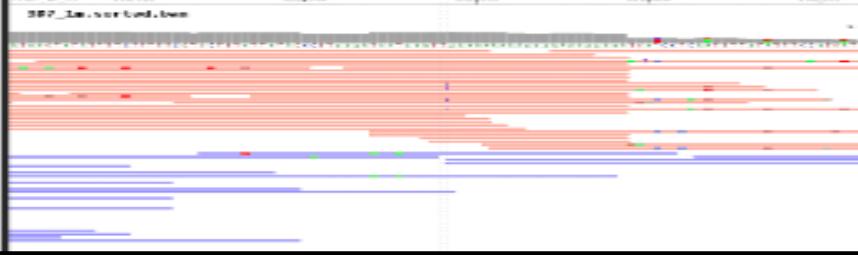


Sibling

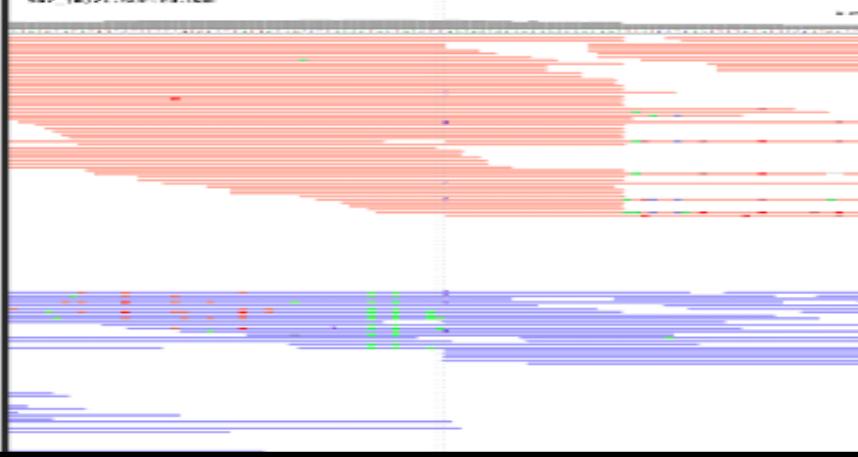


Non de novo mutation

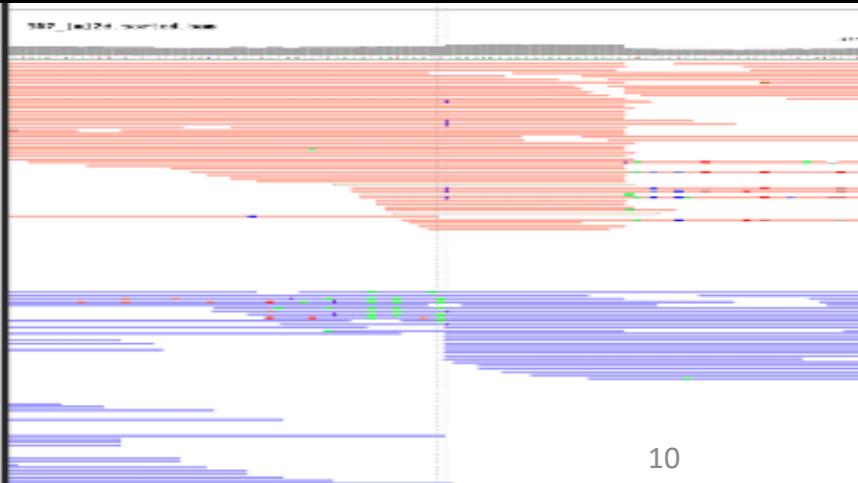
Parent



Sample



Sibling





Micro-Indel mutation rate

- Mutation rate is calculated according to the formula

$$\mu = \frac{\{\sum i / \sum c\}}{L * G}$$

μ = mutation rate,

Σi = number of micro-indels across the clonal generation

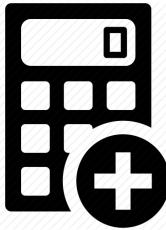
Σc = sum of all the clones for that dilution generation

L = is the total number of life cycles between the respective clonal ‘generations’

G = Genome size of *P. falciparum*.

Strain name	No. of subclones analysed (excluding 1st gen)	Days in culture	Total no. of indels identified	Indels/ELC/nt
3D7	13	203	24	5.510E-02
Dd2	47	298	56	4.42E-02
W2	8	40	10	4.25E-02
Hb3b	14	92	10	3.19E-02
KH-02	23	49	24	4.25E-02

Micro-indel mutation rate

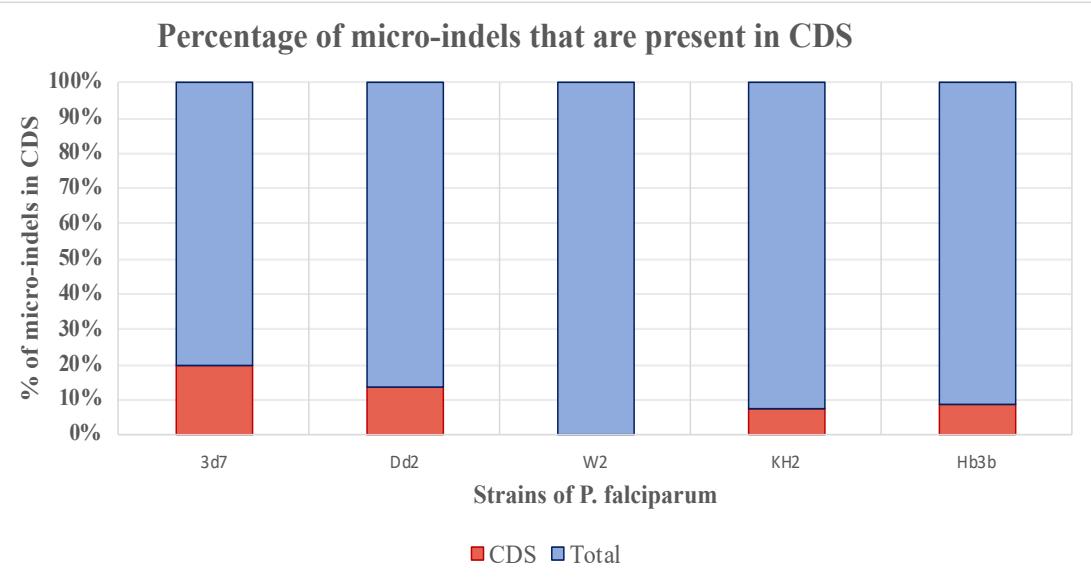


Do these micro-indels change the reading frame?

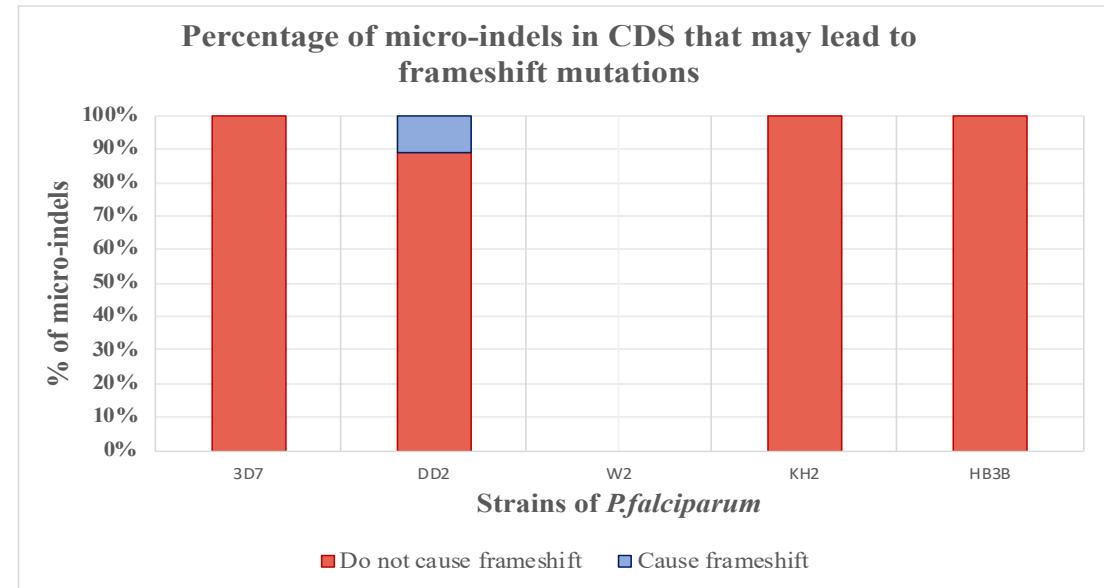
- Micro-indels that are multiples of three preserve the reading frame.

CDS – Coding Sequences

A.



B.

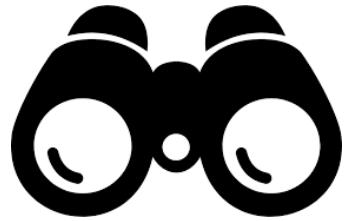


Percentage of micro-indels in CDS (A) and the percentage of micro-indels in CDS whose nucleotide lengths are divisible (red) or indivisible (blue) by three. (B)₁₂



Discussion

- Variant Discovery pipeline
 - Continuously evolving tools.
 - Problems with Haploid genome of *P. falciparum*.
- Majority of the micro-indels founds in all strains
 - Occur outside the coding sequences
 - the ones found in the coding sequences do not cause frameshift
- Other hypotheses as to why drug resistance is first detected in South East Asia
 - High drug pressure in Asia due to socio-economic factors
 - Multiplicity of Infection



Conclusion

The hypothesis is rejected, the micro-indel mutation rate is relatively constant across the *P. falciparum* strains.

Uncovering the mutations that lead the parasite to adapt to the changes in its environment would bring us a step closer to eliminating the disease that still kills nearly half a million people each year.

Acknowledgement

- Antoine Claessens
- Marc-Antoine Guery
- Renaud Vitalis
- Yannis Michalakis
- Sandrine Maurice



Bonus

