Genetic surveillance in the Greater Mekong Subregion to support malaria control and elimination

Imputation of genotypes for markers of drug resistance

V1.0, February 2020

Change History

Version	Date	Notes
1.0	28 February 2020	Details SpotMalaria V2.1 and previous versions. Released with Jacob CG *et al.* (2020)

Introduction

This document defines a set of rules used by the GenRe-Mekong project to infer missing genotypes in drug resistance markers, based on genotypes at other sites that are significantly associated, to improve GenRe-Mekong's power to predict resistant phenotypes. The imputation rules presented here are specifically focussed on haplotypes in genes *crt*, *dhfr* and *dhps*.

Statistical associations were identified by analyzing whole-genome sequencing (WGS) data from over 7,000 global parasite samples in the MalariaGEN Plasmodium falciparum Community Project V6.0 (Pf6, see https://www.malariagen.net/resource/26). To test the predictive association of an allele with another allele at a different position, we created confusion matrices and translated highly significant associations into imputation rules. Each association chosen for a rule was supported by $p < 0.01$ in a Fisher's exact test, unless otherwise stated. No association with $p > 0.05$ was used for rules.

Haplotype counts were estimated in different geographical regions, as follows: WAF = West Africa; CAF = Central Africa; EAF = East Africa; SAS = South Asia; WSEA = West South-East Asia; ESEA = East South-East Asia; OCE = Oceania; SAM = South America. In cases where allele associations were only observed in specific geographical regions, we defined geographically-restricted imputation rules.

Rules are applied as follows:

- Each rule has a *match pattern* (a specific set of alleles at one or more positions) that triggers the rule if it is fully matched by the sample haplotype call; and an *output operation* that specifies

alleles to be assigned at specific positions if the rule is triggered, and if these positions are missing in the sample haplotype call.

- All rules are executed on every sample, in the sequence given in this document. Universal rules are applied before region-specific rules. Triggering a matched rule does not terminate the rule sequence.
- The match pattern is tested against the original sample haplotype call, ignoring any alleles imputed by previously executed rules
- The output operation only imputes missing genotypes, and has no effect on positions that have a valid genotype, including those that have been assigned a genotype by a previously executed rule.

In the first part of this document, we document the imputation rules gene by gene. In the second section, we detail the evidence for the allele associations underpinning each rule.

Prediction Rules

PF3D7_0709000 (pfcrt)

Haplotypes

Plasmodium falciparum chloroquine transporter (*PfCRT*), encoded by *pfcrt* gene, is a transmembrane protein located at digestive vacuole membrane. It is implicated in resistance to chloroquine, amongst others. The *core haplotype* of interest for consists of the amino acid residues 72-76. The K76T mutation is considered the most important marker of chloroquine resistance.

Codon	Wild-type Allele	Alternative Alleles
72	C	S, Y
73	V	-
74	M	I
75	N	D, E
76	K	T

The following table shows the distribution of *PfCRT* core haplotypes in different geographical subcontinents in Pf6:

Haplotype	SAM	EAF	CAF	WAF	SAS	ESEA	WSEA	OCE	Total
CVIDT	0	0	0	0	0	237	0	0	237
CVIET	2	160	314	975	157	1493	1285	0	4386
CVMET	61	0	0	0	0	0	0	0	61
CVMNK	1	711	165	1275	8	43	9	2	2214
CVMNT	21	0	0	0	0	0	0	0	21
SVMNT	12	0	0	0	0	0	0	198	210
YVIET	0	0	0	0	0	4	0	0	4
Total	97	871	479	2250	165	1777	1294	200	7133

Universal Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
All	`S----`	`S-MNT`	`crt_72 == S`	`crt_74 <- M; crt_75 <- N; crt_76 <- T`
All	`--I--`	`--I-T`	`crt_74 == I`	`crt_76 <- T`
All	`---D-`	`C-IDT`	`crt_75 == D`	`crt_72 <- C; crt_74 <- I; crt_76 <- T`
All	`---E-`	`---ET`	`crt_75 == E`	`crt_76 <- T`
All	`---N-`	`--MN-`	`crt_75 == N`	`crt_74 <- M`
All	`----K`	`C-MNK`	`crt_76 == K`	`crt_72 <- C; crt_74 <- M; crt_75 <- N`

Geographical Region	Match	Imputed	Test	Operation
All	`-----`	`-V---`	*<none>*	`crt_73 <- V (invariant)`

Geography-Specific Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
WAF, CAF, EAF, SAS	`--I--`	`C-IET`	`crt_74 == I`	`crt_72 <- C; crt_75 <-E; crt_76 <- T`
WSEA, ESEA	`---E-`	`--IET`	`crt_75 == E`	`crt_74 <- I; crt_76 <- T`
WAF, CAF, EAF, SAS	`---E-`	`C-IET`	`crt_75 == E`	`crt_72 <- C; crt_74 <- I; crt_76 <- T`
WAF, CAF, EAF, SAS, WSEA, ESEA	`---N-`	`C-MNK`	`crt_75 == N`	`crt_72 <- C; crt_74 <- M; crt_76 <- K`
OCE	`--M--`	`--MN-`	`crt_74 == M`	`crt_75 <- N`
WAF, CAF, EAF, SAS, WSEA, ESEA	`--M--`	`C-MNK`	`crt_74 == M`	`crt_72 <- C; crt_75 <- N; crt_76 <- K`
WSEA, ESEA	`----T`	`--I-T`	`crt_76 == T`	`crt_74 <- I`
WAF, CAF, EAF, SAS	`----T`	`C-IET`	`crt_76 == T`	`crt_72 <- C; crt_74 <- I; crt_75 <- E`

PF3D7_0417200 (pfdhfr)

Haplotypes

In Plasmodium falciparum, the *pfdhfr* gene encodes an enzyme, dihydrofolate reductase, which is required for tetrahydrofolate synthesis. Pyrimethamine, an antimalarial drug, interferes with the regeneration of tetrahydrofolic acid from dihydrofolate by competitively inhibiting *pfdhfr*. Parasites with *pfdhfr* mutations may be resistant to pyrimethamine. The core haplotype of interest for consists of the amino acid residues 51, 59, 108, 164. The K76T mutation is considered the most important marker of chloroquine resistance. The *pfdhfr* S108N mutation is an important predictor for pyrimethamine resistance in Plasmodium falciparum, while S108T mutation has been associated with resistance to chlorcycloguanil drugs.

Codon	Wild-type	Mutation
51	**N**	**I**
59	**C**	**R**
108	**S**	**N,T**
164	**I**	**L**

The following table shows the distribution of *pfdhfr* haplotypes in different geographical subcontinents in Pf6, indicating that occurrences of 51I and 59R single mutations (ICSI, NRSI) and double mutation of 51I/59R (IRSI) are rare. In addition, the 164L mutation rarely occurs in Africa, but it is more common in Southeast Asia.

Haplotype	SAM	EAF	CAF	WAF	SAS	ESEA	WSEA	OCE	TOTAL
NCTI	1	0	0	0	0	2	0	0	**3**
ICNL	4	1	0	0	0	0	0	0	**5**
NCNI	43	0	2	6	0	1	1	1	**54**
NRNL	0	0	0	0	15	3	81	0	**99**
ICNI	24	39	56	16	1	0	0	0	**136**
NCSI *(wild-type)*	21	13	7	315	3	8	0	1	**368**
NRNI	0	30	3	112	53	95	72	211	**576**
IRNL	0	3	0	1	29	788	933	0	**1754**
IRNI	1	780	434	1598	31	835	140	0	**3819**
TOTAL	**94**	**866**	**502**	**2048**	**132**	**1732**	**1227**	**213**	**6814**

Universal Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
Any	I---	I-N-	dhfr_51 == I	dhfr_108 <- N
Any	-R--	-RN-	dhfr_59 == R	dhfr_108 <- N
Any	---L	--NL	dhfr_164 == L	dhfr_108 <- N
Any	--S-	NCSI	dhfr_108 == S	dhfr_51 <- N; dhfr_59 <- C; dhfr_164 <- I
Any	--T-	NCTI	dhfr_108 == T	dhfr_51 <- N; dhfr_59 <- C; dhfr_164 <- I

Geography-Specific Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
WSEA, ESEA	-C--	NC--	dhfr_59 == C	dhfr_51 <- N
WSEA, ESEA	I---	IRN-	dhfr_51 == I	dhfr_59 <- R; dhfr_108 <- N
SAS, WSEA, ESEA	-C--	-C-I	dhfr_59 == C	dhfr_164 <- I
SAS, WSEA, ESEA	---L	-RNL	dhfr_164 == L	dhfr_59 <- R; dhfr_108 <- N

PF3D7_0810800 (pfdhps)

Haplotypes

The Plasmodium falciparum dihyropteroate synthase (*pfdhps*) gene encodes an important protein in the folate biosynthesis pathway of the parasite. Polymorphisms of *pfdhps* have been associated with reduced susceptibility to antifolate treatment, e.g. sulfadoxine and sulfadoxine/pyrimethamine.

Codon	Wild-type	Mutation
436	S	A, C, F, H, Y
437	A	G
540	K	E, I, K, N, Y
581	A	G
613	A	S, T

The following table shows the distribution of PfDHPS haplotypes for different geographical regions in Pf6:

Haplotype	SAM	EAF	CAF	WAF	SAS	ESEA	WSEA	OCE	TOTAL
AAKAA	0	13	9	384	0	16	0	0	422
AAKAS	0	0	0	2	0	0	0	0	2
AAKGS	0	0	0	1	0	0	0	0	1
AGEAA	0	1	0	0	71	342	283	0	697
AGEAT	0	0	0	0	0	2	0	0	2
AGEGA	0	0	0	0	0	1	0	0	1
AGIAA	0	0	0	0	0	0	1	0	**1**
AGKAA	0	0	24	367	0	135	0	0	**526**
AGKAS	0	0	1	63	0	0	0	0	**64**
AGKAT	0	0	0	0	0	3	0	0	**3**
AGKGS	0	0	4	43	0	0	0	0	**47**
AGNAA	0	0	0	0	0	0	18	0	**18**
CAKAA	0	0	0	5	0	1	0	3	**9**
FAKAA	0	0	0	0	0	3	0	0	**3**
FAKAS	0	1	0	8	0	1	0	0	**10**
FGEAS	0	0	0	0	0	37	0	0	**37**
FGEAT	0	0	0	0	0	2	4	0	**6**
FGKAS	1	2	0	0	0	1	0	0	**4**
FGKAT	0	3	0	0	0	0	0	0	**3**
HAKAA	0	1	0	0	0	0	0	0	**1**
HGEAA	0	7	0	0	0	0	0	8	**15**
SAKAA (wild-type)	45	57	14	120	13	143	3	79	**474**
SGEAA	0	634	25	12	15	85	21	96	**888**
SGEGA	4	68	9	1	13	49	752	0	**896**
SGKAA	42	12	398	755	3	113	2	11	**1336**
SGKAS	0	0	0	2	0	0	0	0	**2**
SGKGA	5	4	0	1	5	18	22	0	**55**
SGNGA	0	0	0	0	0	633	72	0	**705**
SGYAA	0	0	0	0	0	2	0	0	**2**
YAKAS	0	0	0	7	0	0	0	0	**7**
Total	**97**	**803**	**484**	**1771**	**120**	**1587**	**1178**	**197**	**6237**

Universal Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
Any	F----	F--A-	dhps_436 == F	dhps_581 <- A
Any	H----	H--A-	dhps_436 == H	dhps_581 <- A
Any	C----	CAKA-	dhps_436 == C	dhps_437 <- A; dhps_540 <- K; dhps_581 <- A
Any	Y----	YAKA-	dhps_436 == Y	dhps_437 <- A; dhps_540 <- K; dhps_581 <- A
Any	-A---	-AKA-	dhps_437 == A	dhps_540 <- K; dhps_581 <- A
Any	--N--	-GN-A	dhps_540 == N	dhps_437 <- G; dhps_613 <- A
Any	--E--	-GE--	dhps_540 == E	dhps_437 <- G
Any	---G-	-G-G-	dhps_581 == G	dhps_437 <- G

Geography-Specific Imputation Rules

Geographical Region	Match	Imputed	Test	Operation
EAF, CAF, WAF	F----	F-KA-	dhps_436 == F	dhps_540 <- K; dhps_581 <- A
EAF, CAF, WAF	--E--	-GE-A	dhps_540 == E	dhps_437 <- G; dhps_613 <- A
EAF, CAF, WAF	----S	--K-S	dhps_613 == S	dhps_540 <- K
EAF, CAF, WAF	----T	--K-T	dhps_613 == T	dhps_540 <- K
SAS, WSEA, ESEA	S----	S---A	dhps_436 == S	dhps_613 <- A
SAS, WSEA, ESEA	---G-	-G-GA	dhps_581 == G	dhps_437 <- G; dhps_613 <- A
SAS, WSEA, ESEA	----S	F--AS	dhps_613 == S	dhps_436 <- F; dhps_581 <- A
SAS, WSEA, ESEA	----T	---AT	dhps_613 == T	dhps_581 <- A

Statistical Evidence for Imputation Rules

PF3D7_0709000 (pfcrt)

Codons 72 and 74

There is a significant association between PfCRT codons 72 and 74 in parasites from all regions:

All regions	C72	72S	72Y	Total
74I	4838	0	4	4842
M74	2362	211	0	2573
Total	7200	211	4	7415

suggesting the following:

Rule: 72S predicts M74

Also:

- 72S was found only in South America and Oceania, and
- 72Y mutation only occurred in Southeast Asia

suggesting additional rules:

Rule: M74 predicts C72 except in SAM and OCE

Rule: 74I predicts C72 except in SAM, OCE, WSEA and ESEA

Codons 72 and 75

There is a significant association between PfCRT codons 72 and 75 in the parasites from all regions:

All regions	C72	72S	72Y	Total
75D	237	0	0	237
75E	4662	0	4	4666
N75	2301	211	0	2512
Total	7200	211	4	7415

suggesting the following:

Rule: 75D predicts C72

Rule: 72S predicts N75

Also:

- 72S was found only in South America and Oceania, and
- 72Y mutation only occurred in Southeast Asia

suggesting additional rules:

Rule: N75 predicts C72 except in SAM and OCE

Rule: 75E predicts C72 except in SAM, OCE, WSEA and ESEA

Codons 72 and 76

There is significant association between PfCRT codons 72 and 76 in the parasites from all regions:

All regions	C72	72S	72Y	Total
K76	2280	0	0	2280
76T	4920	211	4	5135
Total	7200	211	4	7415

suggesting the following:

Rule: K76 predicts C72
Rule: 72S predicts 76T

Also:

- 72S was found only in South America and Oceania, and
- 72Y mutation only occurred in Southeast Asia

suggesting additional rules:

Rule: 76T predicts C72 **except in SAM, OCE, WSEA and ESEA**

Codons 74 and 75

There is a significant association between PfCRT codons 74 and 75 in the parasites from all regions:

All regions	74I	M74	Total
75D	237	0	237
75E	4605	61	4666
N75	0	2512	2512
Total	4842	2573	7415

suggesting the following:

Rule: 75D predicts 74I
Rule: N75 predicts M74

When disregarding the parasites from *South America* (SAM):

All except South America	74I	M74	Total
75D	237	0	237
75E	4388	0	4388
N75	0	2411	2411
Total	4625	2411	7036

suggesting the following:

Rule: 75E predicts 74I **except in SAM**
Rule: M74 predicts N75 **except in SAM**

Also:

- 75D was found only in Southeast Asia

suggesting the following:

Rule: 74I predicts 75E	**except in SAM, WSEA and ESEA**

Codons 74 and 76

There is a significant association between PfCRT codons 74 and 76 in the parasites from all regions:

All regions	74I	M74	Total
K76	0	2280	2280
76T	4842	293	5135
Total	4842	2573	7415

suggesting the following:

Rule: K76 predicts M74	(Fisher's exact test: P < 0.01)
Rule: 74I predicts 76T	(Fisher's exact test: P < 0.01)

When disregarding the parasites from *South America* and *Oceania* (SAM and OCE):

All regions except SAM, OCE	74I	M74	Total
K76	0	2211	2211
76T	4625	0	4625
Total	4625	2211	6836

suggesting the following:

Rule: M74 predicts K76	**except in SAM, OCE**
Rule: 76T predicts 74I	**except in SAM, OCE**

Codons 75 and 76

There is a significant association between PfCRT codons 75 and 76 in the parasites from all regions:

All regions	75D	75E	N75	Total
K76	0	0	2280	2280
76T	237	4666	232	5135
Total	237	4666	2512	7415

suggesting the following:

Rule: K76 predicts N75
Rule: 75D predicts 76T
Rule: 75E predicts 76T

Also:

Rule: N75 predicts K76	**except in SAM, OCE**

Association between K76T and other *crt* markers for chloroquine resistance

K76T mutation has been used as a marker for chloroquine resistance; however, single *pfcrt* K76T mutation is not sufficient to cause altered chloroquine susceptibility. The data from MalariaGEN Plasmodium falciparum Community Project V6.0 was assessed for the association between K76T mutation and other markers for chloroquine resistance, including residues 75, 220 and 326. We can infer that K76T is a predictor for these additional chloroquine-resistant markers.

	K76	76T	Total
N75+A220+N326	2056	0	2056
N75+220S+N326	3	0	3
75D+220S+N326	0	210	210
75E+A220+N326	0	2	2
75E+220S+N326	0	1658	1658
75E+220S+326S	0	2060	2060
75N+220S+326D	0	222	222
Total	2059	4152	6211

PF3D7_0417200 (pfdhfr)

To impute missing codons in *pfdhfr*, the associations between amino acids at position 51, 59, 108 and 164 were assessed to generate the imputation rules using the data from Pf6. These generally support evidence from studies of the evolutionary process of *pfdhfr* mutation, proposing that 164L mutation arose after C59R+S108N and N51I+C59R+S108N mutations [1, 2].

Codons 51 and 108

There is a significant association between *pfdhfr* codons 51 and 108 in the parasites from all regions:

All regions	N51	51I	Total
108N	746	6283	**7029**
S108	383	0	**383**
108T	4	0	**4**
Total	**1133**	**6283**	**7416**

suggesting the following:

Rule: 51I predicts 108N
Rule: S108 predicts N51
Rule: 108T predicts N51

Codons 59 and 108

There is a significant association between *pfdhfr* codons 59 and 108 in the parasites from all regions:

All regions	C59	59R	Total
108N	204	6953	**7157**
S108	378	0	**378**
108T	3	0	**3**
Total	**585**	**6953**	**7538**

suggesting the following:

Rule: 59R predicts 108N
Rule: S108 predicts C59
Rule: 108T predicts C59

Codons 108 and 164

There is a significant association between *pfdhfr* codons 108 and 164 in the parasites from all regions:

All regions	I164	164L	Total
108N	5270	1911	7181
S108	391	0	391
108T	4	0	4
Total	5665	1911	7576

suggesting the following:

Rule: 164L predicts 108N
Rule: S108 predicts I164
Rule: 108T predicts I164

Codons 51 and 59

There is a significant association between *pfdhfr* codons 108 and 164 but insufficient for an imputation rule that covers the parasites from all regions:

All regions	51I	N51	Total
C59	148	455	**603**
59R	6031	696	**6727**
Total	**6179**	**1151**	**7330**

However, when considering only parasites in *Southeast Asia* (ESEA and WSEA):

Southeast Asia	51I	N51	Total
C59	0	12	**12**
59R	2840	261	**3101**
Total	**2840**	**273**	**3113**

suggesting the following:

Rule: C59 predicts N51 in WSEA, ESEA
Rule: 51I predicts 59R in WSEA, ESEA

Codons 59 and 164

There is a significant association between *pfdhfr* codons 59 and 164 but insufficient for an imputation rule that covers the parasites from all regions:

All regions	C59	59R	Total
I164	612	4890	**5502**
164L	5	1899	**1904**
Total	**617**	**6789**	**7406**

However, when considering only parasites in *Asia* (SAS, ESEA and WSEA):

Asia	C59	59R	Total
I164	16	1288	**1304**
164L	0	1895	**1895**
Total	**16**	**3183**	**3199**

suggesting the following:

Rule: C59 predicts I164 in SAS, WSEA, ESEA
Rule: 164L predicts 59R in SAS, WSEA, ESEA

References

1. Sirawaraporn W, Sathitkul T, Sirawaraporn R, Yuthavong Y, Santi DV. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. Proc Natl Acad Sci U S A. 1997;94(4):1124-9.
2. Lozovsky ER, Chookajorn T, Brown KM, Imwong M, Shaw PJ, Kamchonwongpaisan S, et al. Stepwise acquisition of pyrimethamine resistance in the malaria parasite. Proc Natl Acad Sci U S A. 2009;106(29):12025-30.

PF3D7_0810800 (pfdhps)

Codons 436 and 581

There is a significant association between *pfdhps* codons 436 and 581 in the parasites from all regions:

All regions	A581	581G	Total
436A	1954	58	**2012**
436C	9	0	**9**
436F	63	0	**63**
436H	17	0	**17**
S436	3114	1716	**4830**
436Y	9	0	**9**
Total	**5166**	**1774**	**6940**

suggesting the following:

Rule: 436F predicts A581
Rule: 436H predicts A581

Fisher's exact test: $p < 0.05$

Codons 437 and 581

There is a significant association between *pfdhps* codons 437 and 581 in the parasites from all regions:

All regions	A581	581G	Total
A437	1032	1	**1033**
437G	4370	1780	**6150**
Total	**5402**	**1781**	**7183**

suggesting the following:

Rule: A437 predicts A581
Rule: 581G predicts 437G

Codons 613 and 581

There is no association with $p < 0.05$ between *pfdhps* codons 613 and 581 in the parasites from all regions:

All regions	A581	581G	Total
A613	5682	1746	**7428**
613S	147	57	**204**
613T	14	0	**14**
Total	**5843**	**1803**	**7646**

However, when considering only parasites in Asia (SAS, ESEA and WSEA):

All regions	A581	581G	Total
A613	1473	1646	**3119**
613S	40	0	**40**
613T	11	0	**11**
Total	**1524**	**1646**	**3170**

suggesting the following:

Rule: 613S predicts A581 **in SAS, WSEA, ESEA**
Rule: 613T predicts A581 **in SAS, WSEA, ESEA**
Rule: 581G predicts A613 **in SAS, WSEA, ESEA**

Codons 540 and 613

There is a significant association between 540N mutation and A613 in parasites from all regions:

All regions	A613	613S	613T	Total
540E	2739	40	8	**2787**
540I	1	0	0	**1**
K540	3809	169	6	**3984**
540N	751	0	0	**751**
540Y	2	0	0	**2**
Total	**7302**	**209**	**14**	**7525**

suggesting the following:

Rule: 540N predicts A613

When only considering *African* parasites data (subcontinent codes: WAF, CAF and EAF):

Africa	A613	613S	613T	Total
540E	835	0	0	**835**
K540	2789	140	3	**2932**
Total	**3624**	**140**	**3**	**3767**

suggesting the following:

Rule: 540E predicts A613 in WAF, CAF and EAF
Rule: 613S predicts K540 in WAF, CAF and EAF
Rule: 613T predicts K540 in WAF, CAF and EAF

Codons 436 and 613

The data showed significant association only when considering Asian parasites (SAS, ESEA, WSEA):

Asia	A613	613S	613T	Total
436A	920	0	5	**925**
436C	1	0	0	**1**
436F	3	42	6	**51**
S436	2114	0	0	**2114**
Total	**3038**	**42**	**11**	**3091**

suggesting the following:

Rule: S436 predicts A613 in SAS, WSEA, ESEA
Rule: 613S predicts 436F in SAS, WSEA, ESEA

Codons 436 and 540

There are significant associations between K540 and 436C mutation and between K540 and 436Y mutation in the parasites from all regions:

All regions	540E	540I	K540	540N	540Y	Total
436A	717	1	1270	18	0	**2006**
436C	0	0	9	0	0	**9**
436F	45	0	20	0	0	**65**
436H	16	0	1	0	0	**17**
S436	1864	0	2155	720	2	**4741**
436Y	0	0	9	0	0	**9**
Total	**2642**	**1**	**3464**	**738**	**2**	**6847**

suggesting the following:

Rule: 436C predicts K540
Rule: 436Y predicts K540

When only considering *African* parasites data (subcontinent codes: WAF, CAF and EAF):

Africa	540E	K540	Total
436A	1	1033	**1034**
436C	0	5	**5**
436F	0	14	**14**
436H	7	1	**8**
S436	802	1410	**2212**
436Y	0	9	**9**
Total	**810**	**2472**	**3282**

suggesting the following:

Rule: 436F predicts K540 in WAF, CAF and EAF

Codons 437 and 540

There is a significant association between codons 437 and 540 in parasites from all regions:

All regions	540E	540I	K540	540N	540Y	Total
A437	0	0	1024	0	0	**1024**
437G	2758	1	2686	742	2	**6189**
Total	**2758**	**1**	**3710**	**742**	**2**	**7213**

suggesting the following:

Rule: 540E predicts 437G
Rule: 540N predicts 437G
Rule: A437 predicts K540

Codons 436 and 437

There is a significant association between codons 436 and 437 in parasites from all regions:

All regions	A437	437G	Total
436A	454	1495	**1949**
436C	9	0	**9**
436F	13	53	**66**
436H	1	16	**17**
S436	491	4344	**4835**
436Y	9	0	**9**
Total	**977**	**5908**	**6885**

suggesting the following:

Rule: 436C predicts A437
Rule: 436Y predicts A437