Finding drug resistance genes

Ian Cheeseman
Texas Biomedical Research Institute
KITP 2014

Acknowledgements

- Infectious Disease Genetics lab- TBRI
  - Tim Anderson
  - Shalini Nair
  - Standwell Nkhoma
  - Marina McDew-White
- Ferdig Lab- Notre Dame
  - Michael Ferdig
  - Becky Miller
  - John Tan
  - Asako Tan
- Shoklo Malaria Research Unit, Thailand
  - Francois Nosten
  - Aung Pyae Phyo
  - Rose McGready
  - Kanlaya Sriprawat
- Mahidol University/University of Oxford
  - Nick White
  - Arjen Dondorp
  - Liz Ashley
- Mahosot Hospital, Laos PDR
  - Paul Newton
  - Mayfong Mayxay

Some basic malariology…

- Causative agents are from the *Plasmodium* family of apicomplexan parasites
  - >70 species of *Plasmodium*
  - 6 infect humans (***P. falciparum****, P. vivax, P. knowlesi, P. ovale (wallikeri and curtisi), P. malariae*)

- Single cell eukaryotic pathogen, 23Mb haploid genome

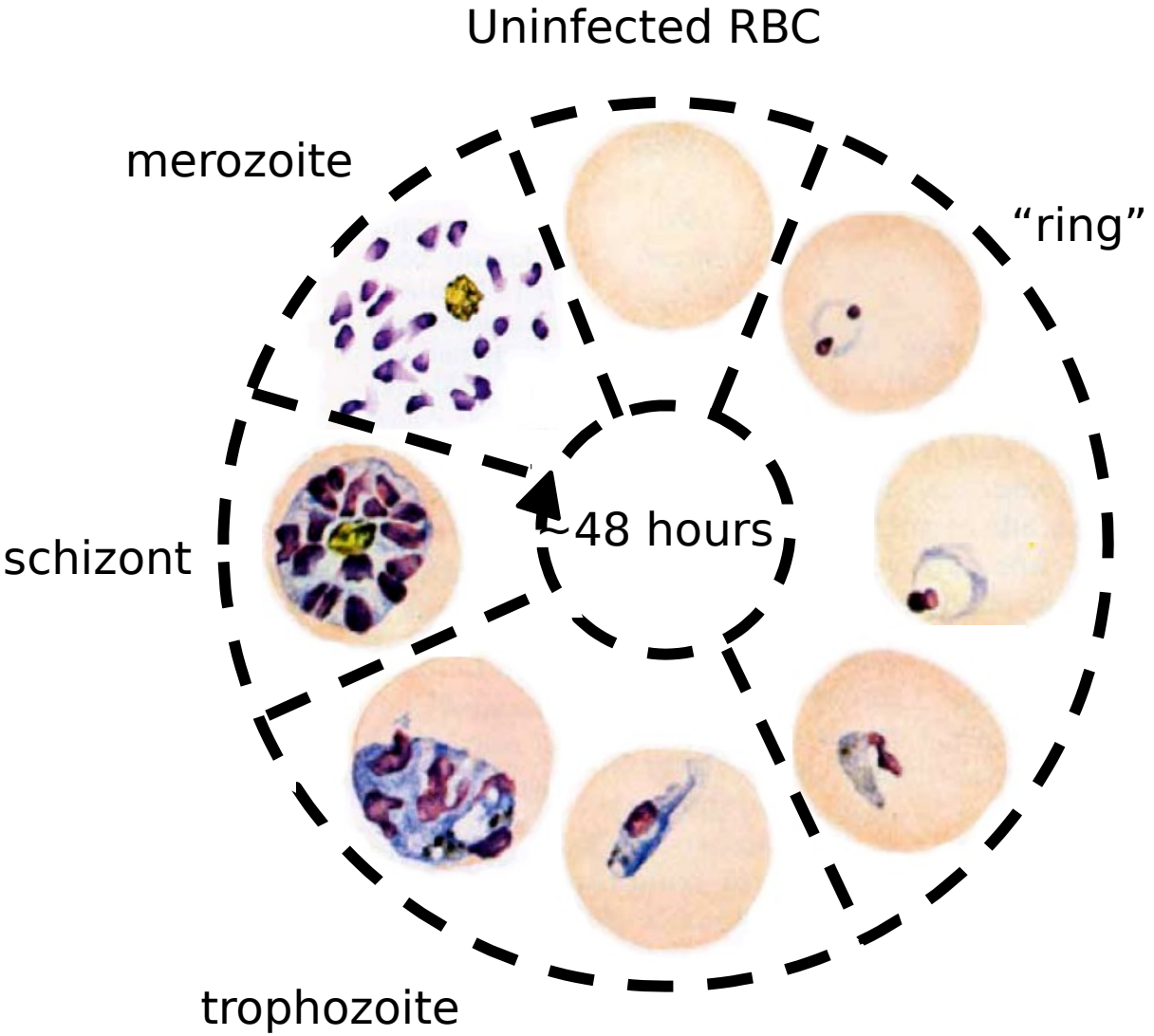
- Transmitted between people by female anopheline mosquitoes
  - Brief diploid phase where recombination occurs

- *P. falciparum* has a vast disease burden
  - ~700,000 deaths per year (90% of which in children <5 years old in sub-Saharan Africa)
  - ~500,000,000 infections per year

Life-cycle in the blood

- Complex life-cycle in 2 hosts, 2 very distinct stages in both (liver and blood stages in humans)

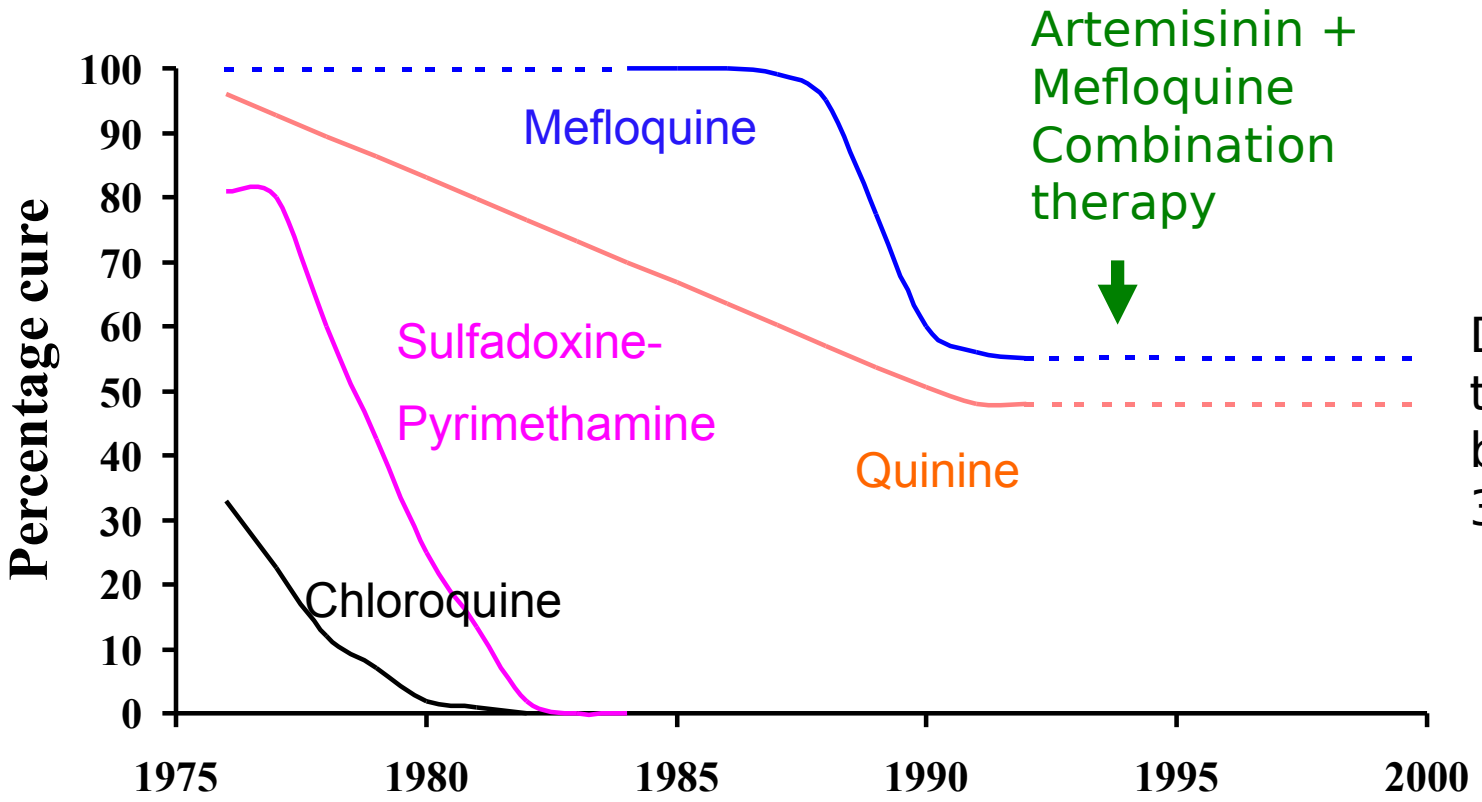
- Here is the human blood stage:
  - major pathogenic stage (causes all disease symptoms)
  - Major drug target (and major drug resistance target)



Drug resistance

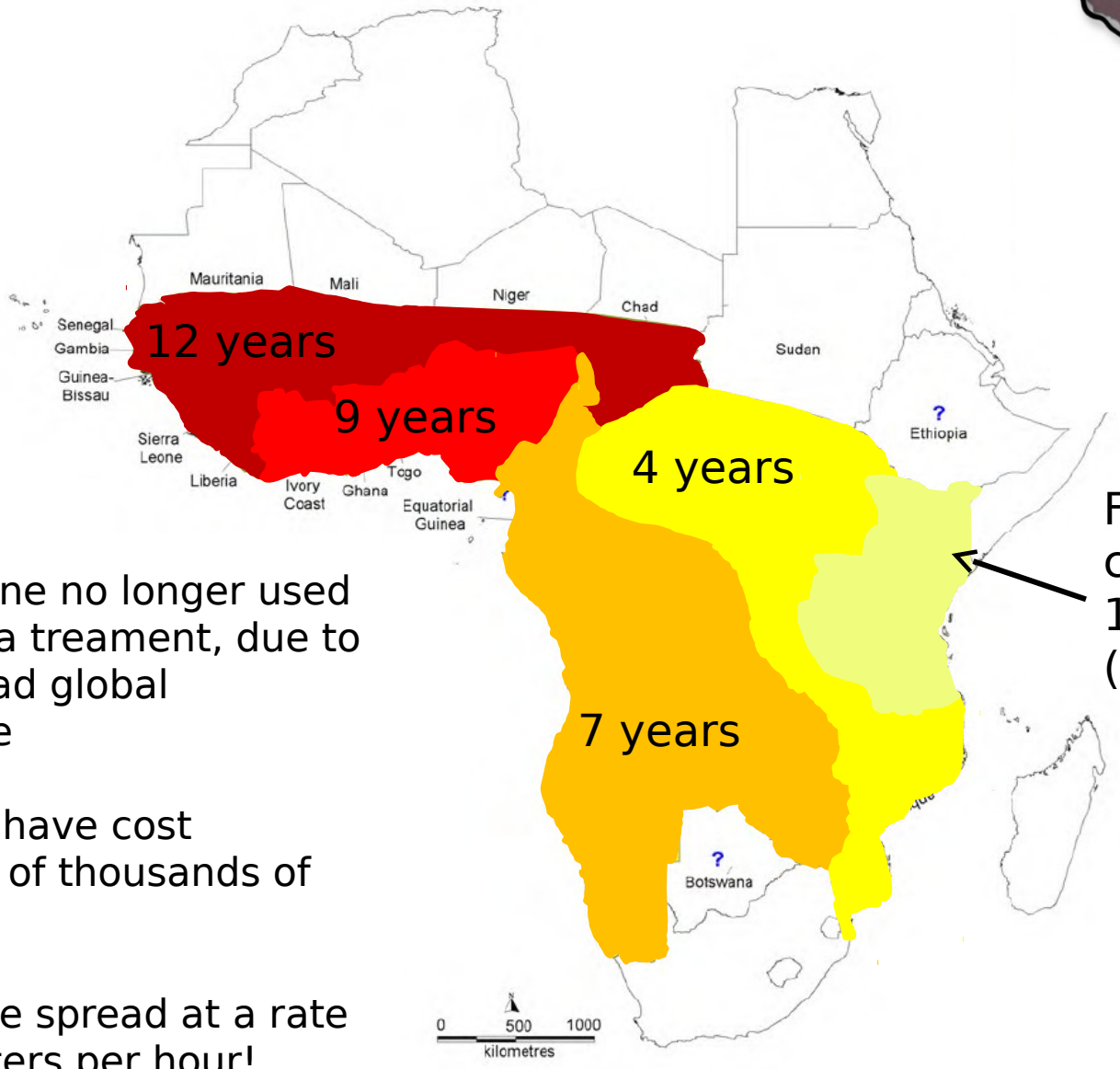
Resistance to all available anti-malarial drugs has occurred and rapidly spread

Chloroquine and Sulfadoxine-Pyrimethamine have both been withdrawn for malaria treatment in Africa



Drug resistance on the Thai-Burmese border over the last 35 years

Once resistance occurs a drug can be rapidly lost



First report of chloroquine resistance: 1978 Kenya/Tanzania (imported from SE Asia)

Chloroquine no longer used as malaria treament, due to widespread global resistance

This may have cost hundreds of thousands of lives!

Resistance spread at a rate of 60 meters per hour!

Artemisinin

A sesquiterpine lactone drug derived from the sweet wormwood plant (*Artemisia annua*)

Highly efficacious, clears infections in <24 hours, with a half-life of ~2 hours

Superior to previously used anti-malarials (30% decrease in mortality compared to quinine when used to treat severe malaria)

Few side effects and rapidly cleared from the bloodstream (~30 minutes)

Used alongside a long-lasting partner drug to slow the emergence of resistance/prevent parasite recrudescence

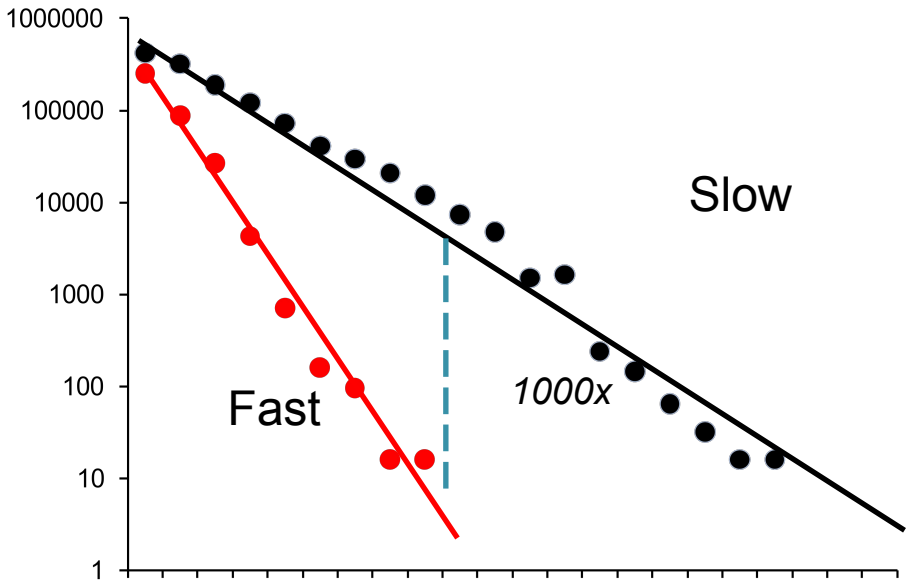
Artemisia annua

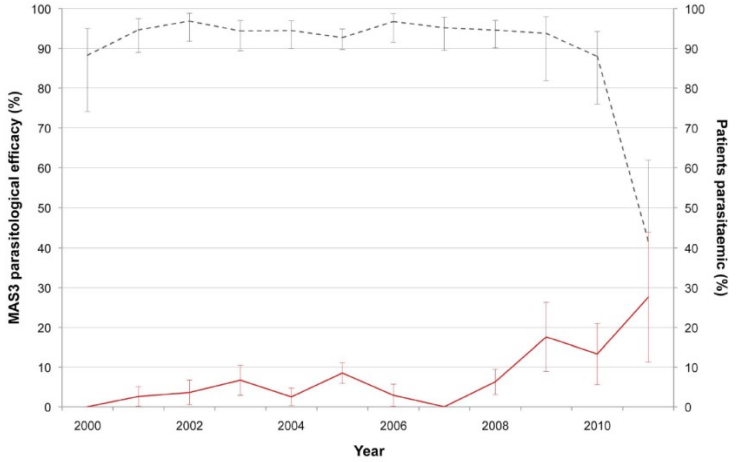
Could artemisinin resistance in malaria constitute a public health crisis?

- Parasites have acquired resistance to all other anti-malarial drugs and this can spread rapidly
- Loss of chloroquine/anti-folates caused a 2-3 fold increase in the mortality rate
- Over 1 million people treated with artemisinin annually (this figure is rising rapidly)

Artemisinin resistance

- Characterised by **clearance rate**
- No correlation with traditional *in vitro* IC50 measurements
- Is this resistance?
  - 1000 fold increase of clearance rate
  - Increase in treatment failure
  - Strong selection for slow clearance
  - Implications for control





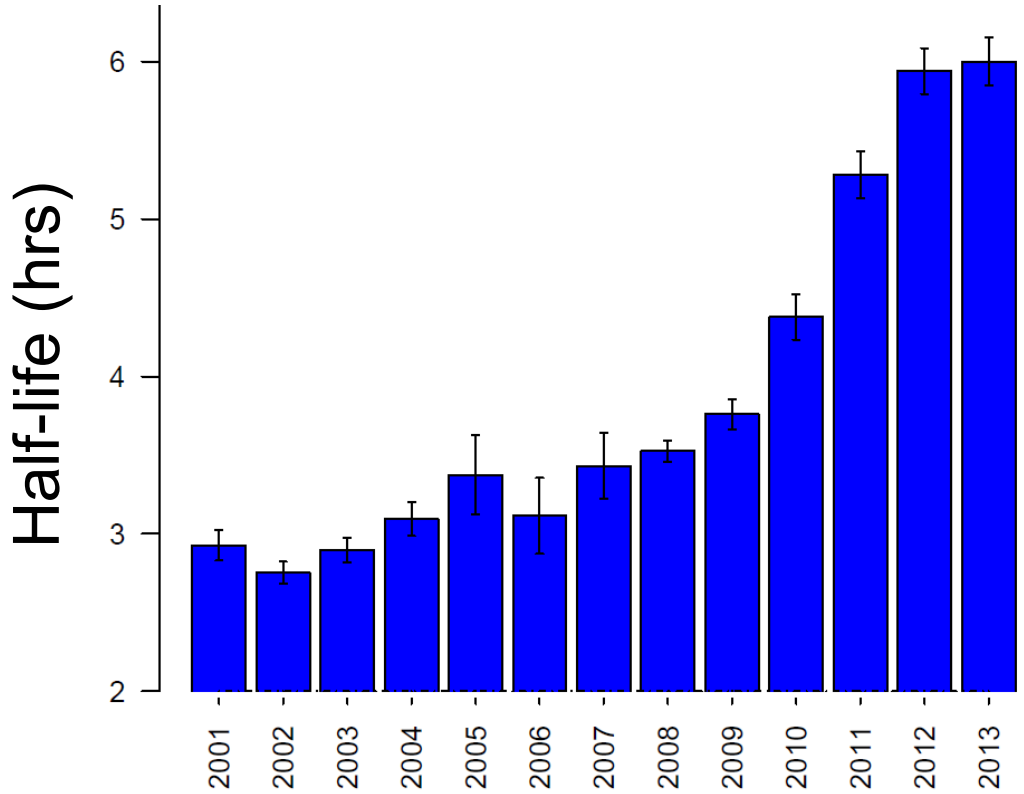
Carrara et al, PLoS Medicine 2013

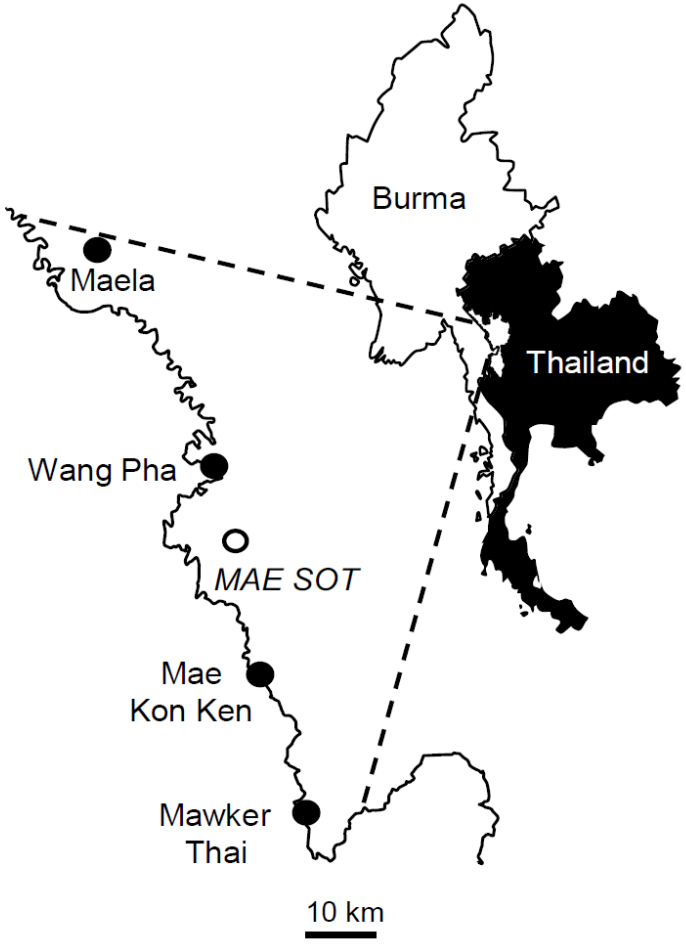
Measure parasite density every 6 hrs

Rapid spread of resistance

François Nosten
Shoklo Malaria Research Unit

- 4 clinics on the Thailand Burma border
- 6 hr parasite density measures in ~2500 people over 13 years
- Rapid changes in parasite clearance half-lives
- Finger prick blood samples collected



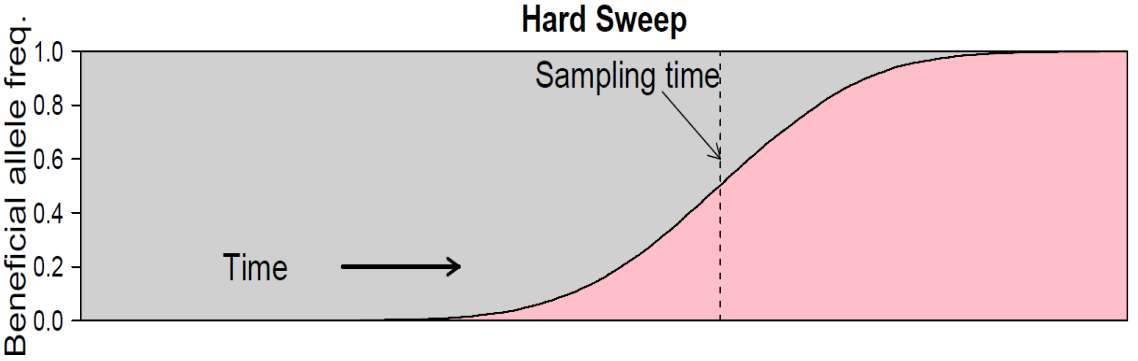


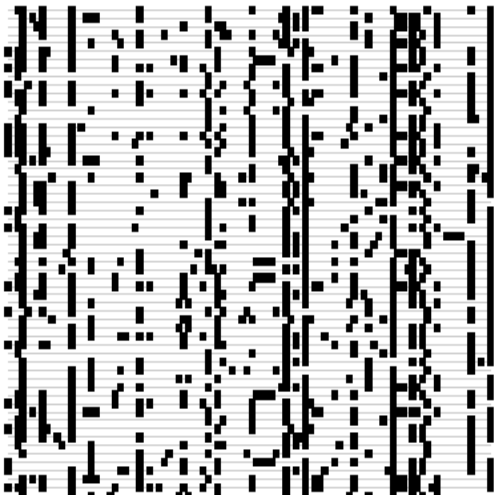
How do you find the mutations causing resistance?

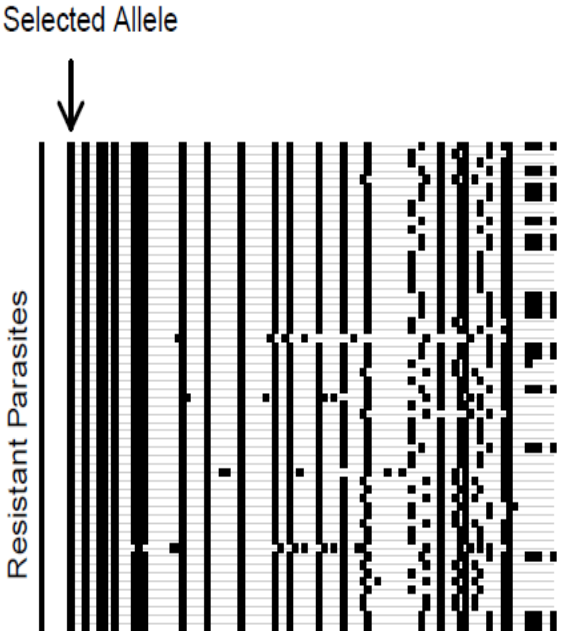
- Candidate genes
  - From other species where mechanism of resistance is known
  - From knowledge of the method of action of a drug
- QTL mapping
  - Generate cross between resistant and sensitive lines
  - Measure resistance in progeny
- Scan for signatures of strong recent selection
  - Sequence variants in tens/hundreds/thousands of individuals
- Genome-wide association study
  - Score genotypes and phenotypes in 100s-1,000s of individuals
  - Correlation between genotype and phenotype
- Experimental evolution
  - Place lab population under selective pressure
  - Sequence survivors
- ALL of these approaches have been used to find resistance genes in malaria parasites

What happens during the spread of drug resistance?

- Under a classic “hard” sweep model a beneficial mutation increases in frequency in a population
- The spread of this mutation generates low diversity, long haplotypes and allele frequency divergence
- It is possible to survey genomes from a population for these signatures to locate resistance genes

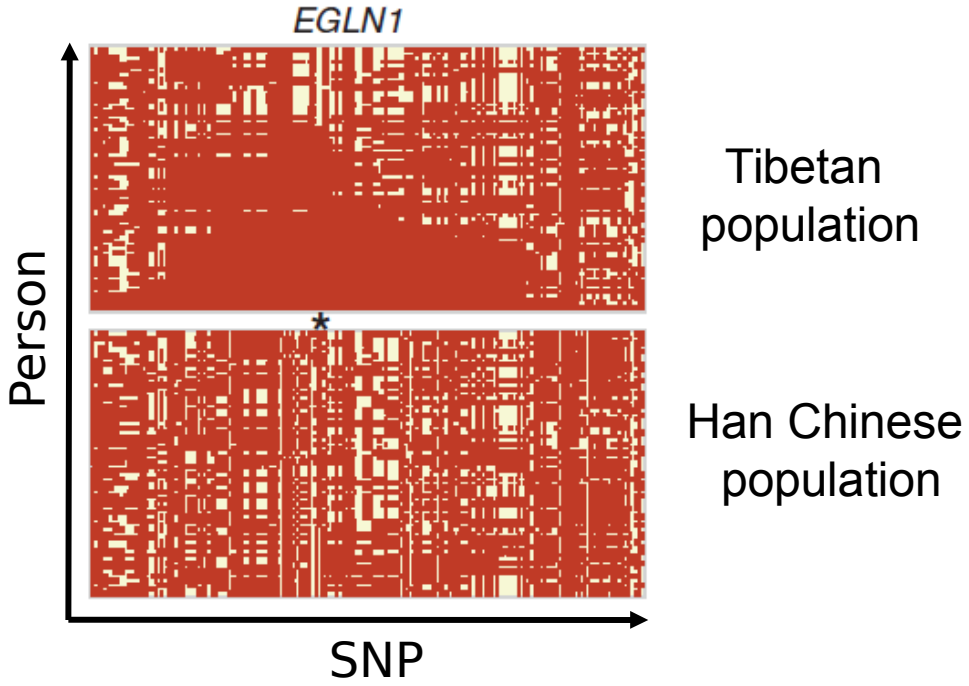






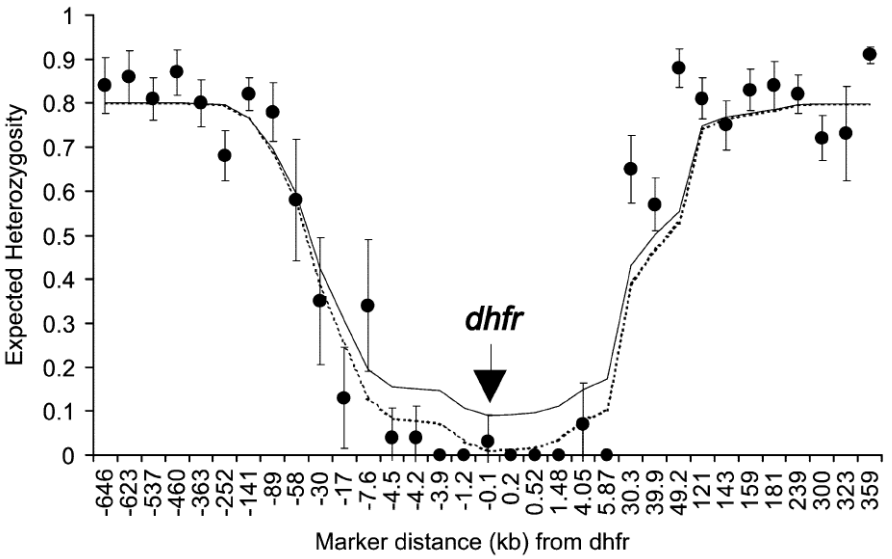
What does this look like in real populations?

Strong selection for high altitude environments has driven mutations of the EGLN1 gene (*) to high frequency in Tibetan populations but not Han Chinese and shows extreme XP-EHH.



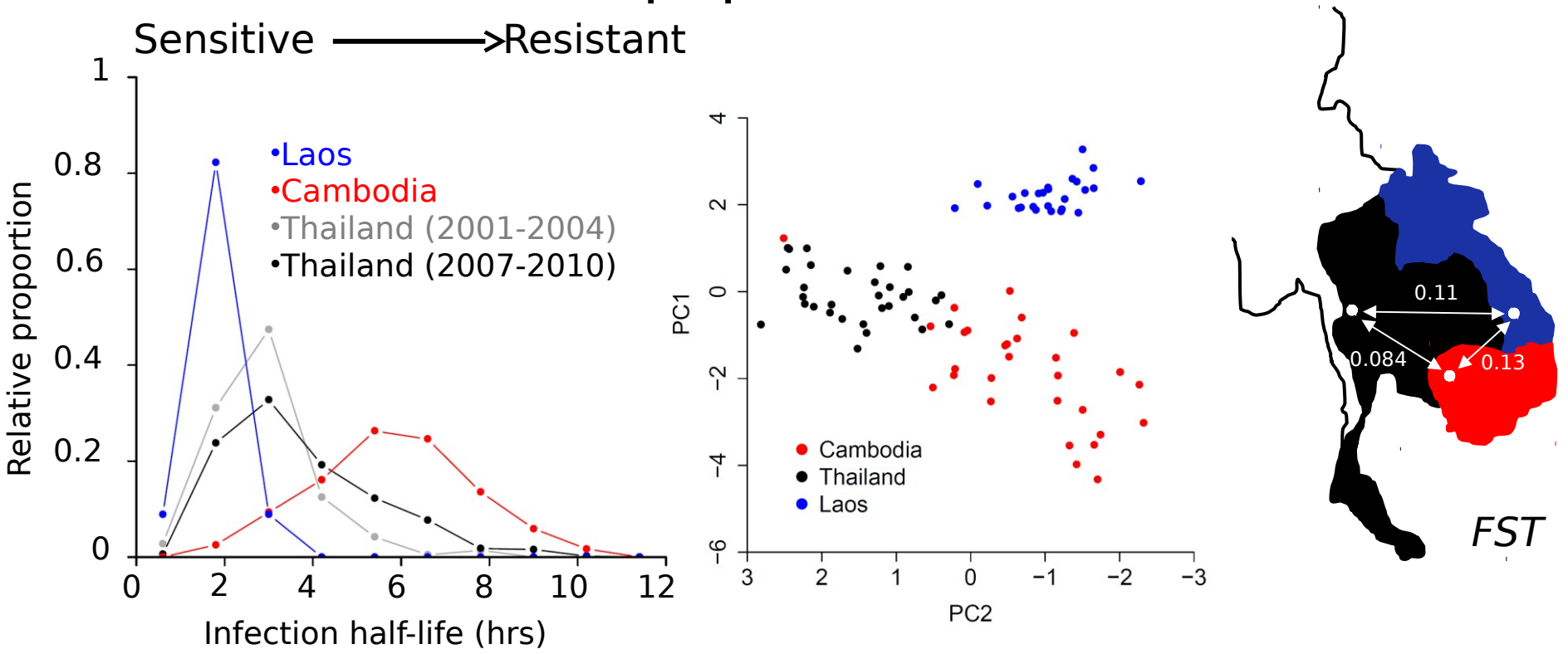
Simonson et al; Science, 2010

Strong selection for drug resistance conferred by alleles at the *dhfr* gene in *P. falciparum* has resulted in a selective sweep surrounding the gene



Nair et al; Molecular Biology and Evolution, 2003

Genetic and phenotypic variation between populations



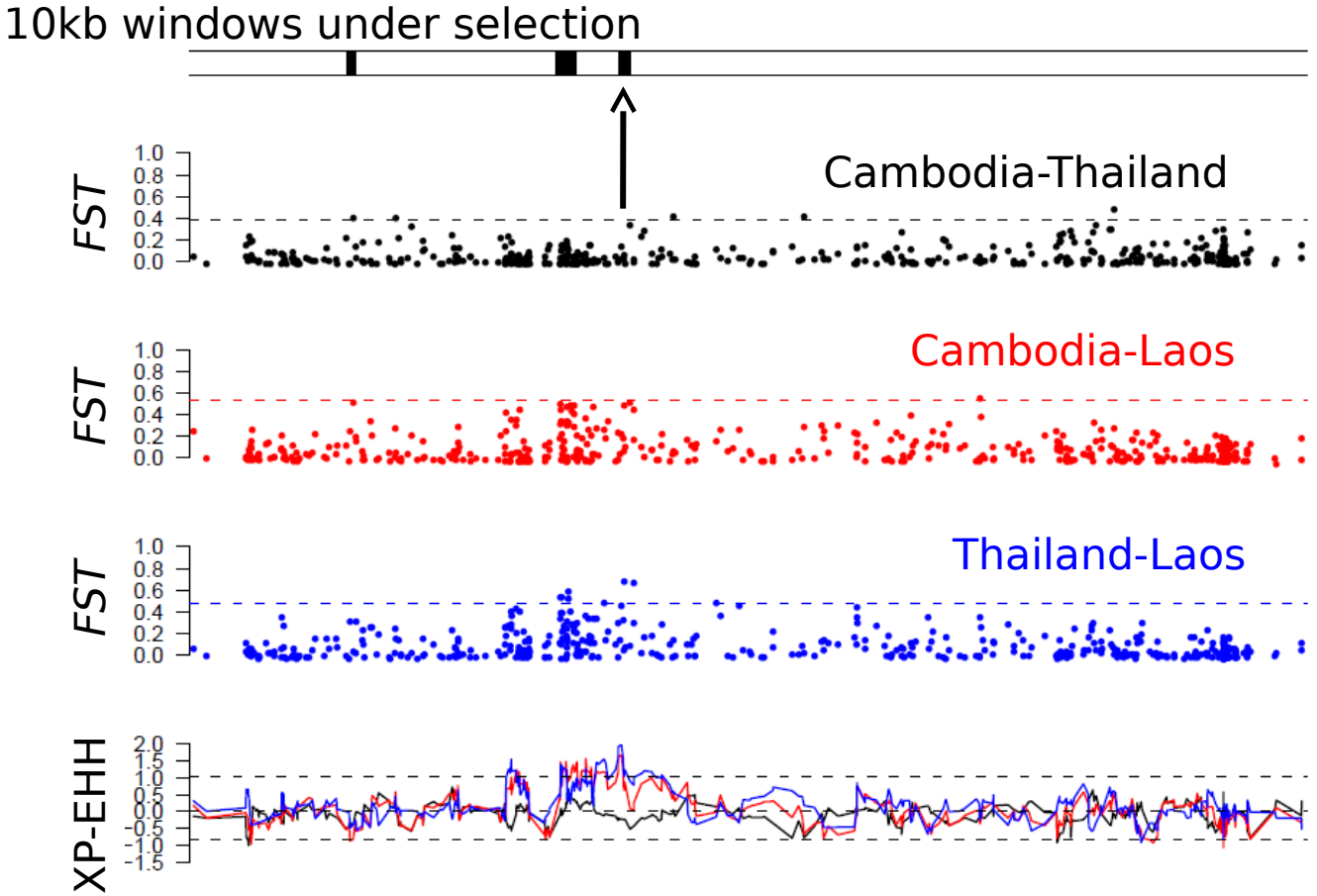
Extreme **phenotypic variation** between 3 SE Asian populations

Low **genetic differentiation** (though populations may be separated by *FST*/PCA analysis)

This provides a relatively low background level of differentiation against which selection can be identified

Cheeseman et al, Science 2012

Identifying known resistance genes using signatures of selection



Arrows shows the location of *dhps* an anti-folate resistance gene

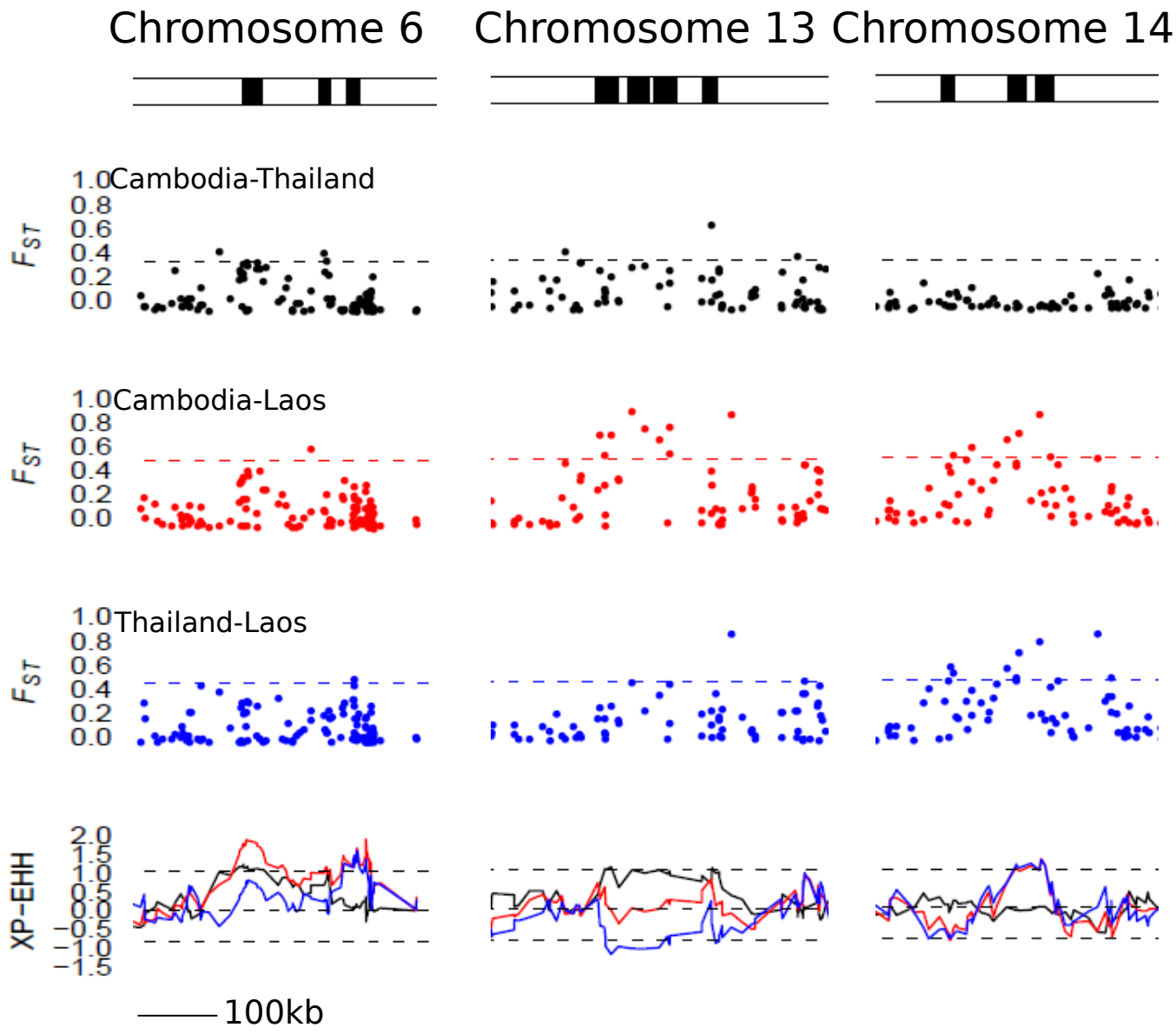
Cambodia and Thailand are anti-folate resistant

Laos is anti-folate sensitive

Three/five known drug resistance genes (*dhps*, *dhfr* and *pfcrt*) are directly implicated in our analysis.

Cheeseman et al, Science 2012

Identifying novel resistance genes using signatures of selection

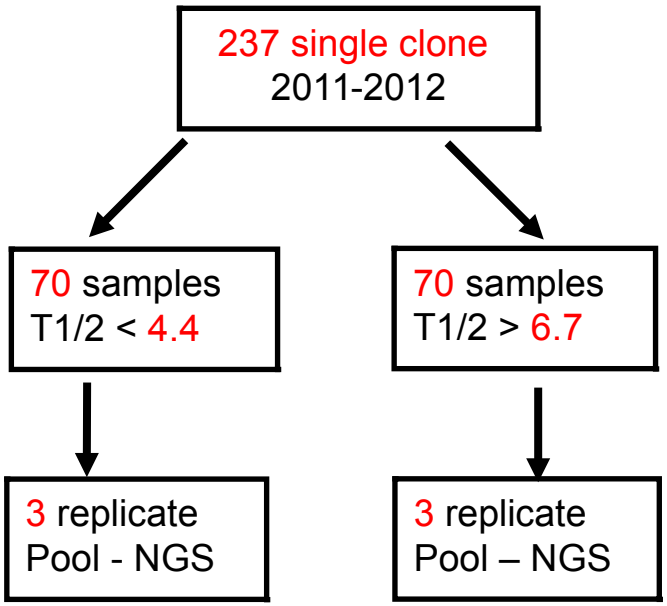


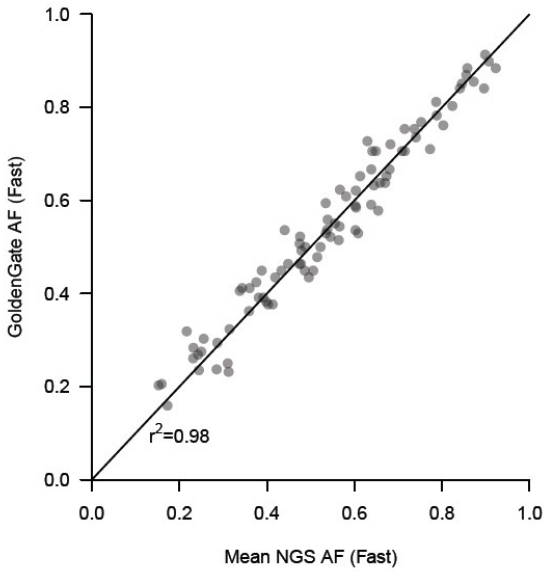
In total **33 loci** showed evidence of **recent positive selection** in at least one population comparison

here are some of the strongest candidates

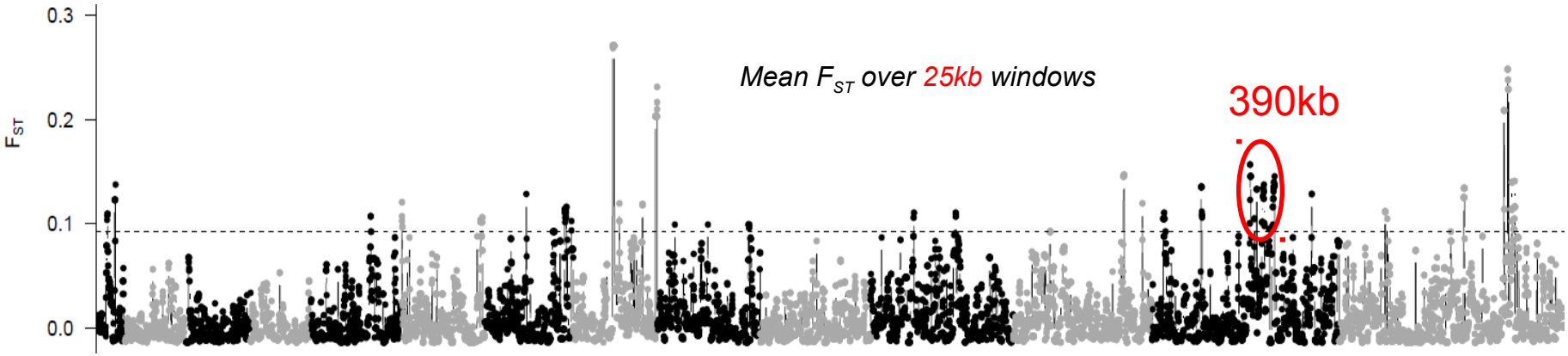
Cheeseman et al, Science 2012

Identifying additional targets of resistance





Allele frequency estimates are very accurate!



Pooled sequencing is powerful, accurate and cheap
Chr 13 region is replicated
Several other interesting regions

Cheeseman et al, in revision

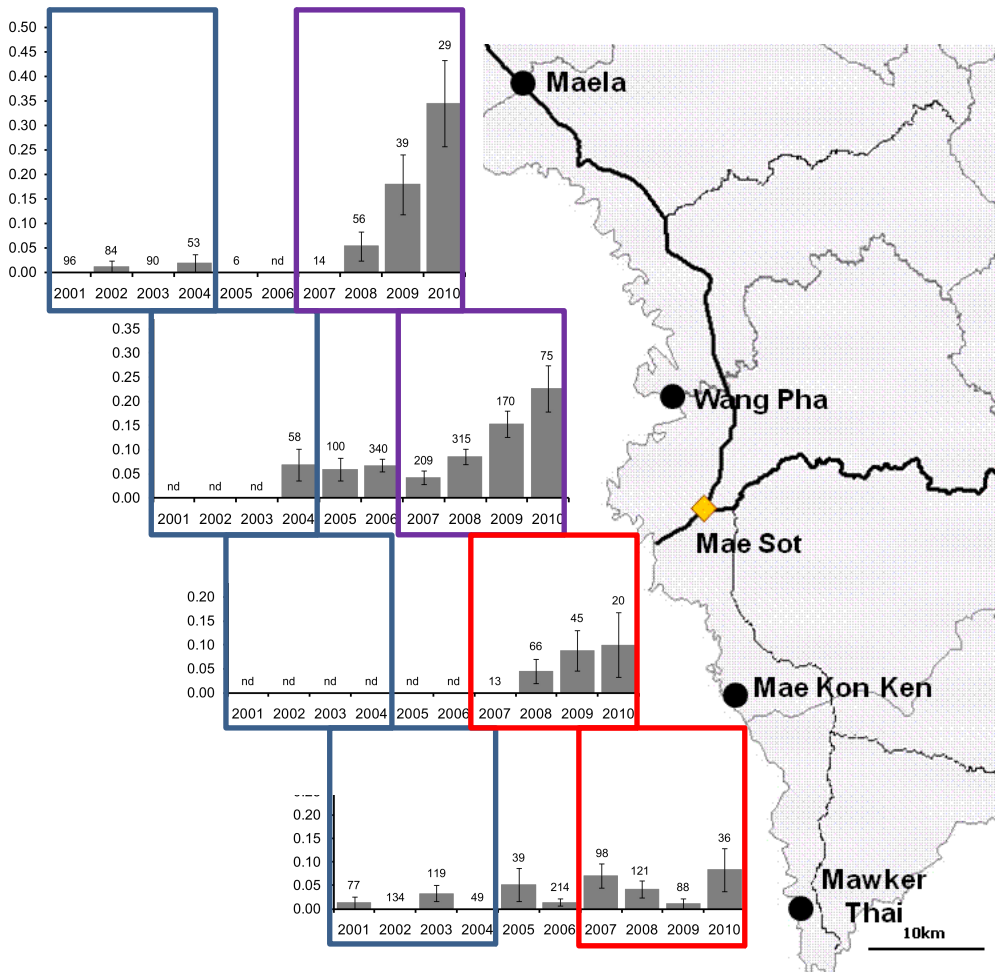
How close does this get you?

- Scans for signatures of strong recent selection may have arisen for many reasons, not just drug resistance
- Even for genome regions where selection is driven by resistance 10s of genes may be implicated
- To identify genes not only under selection but associated with resistance other methods are needed

GWAS for artemisinin resistance genes

- We can use genome-wide association studies to confirm regions of the genome under selection are due to artemisinin resistance
- We used a large cohort of 3202 patients from Western Thailand collected during the rise of resistance
- We genotyped 1,689 parasites at 96 high FST SNPs from within regions under strong selection and 96 neutral SNPs from throughout the genome
- We excluded mixed-clone infections and non-unique individuals, retaining 715 for association

Mapping artemisinin resistance in Western Thailand



Phyo, Nkhoma et al, The Lancet 2012

The last decade has seen a dramatic shift in the prevalence of artemisinin resistance

Clear patterns of spreading resistance are apparent in populations to the north of Mae Sot, though this is less clear to the south despite a general trend of increasing tolerance of the drug

For association mapping we have split the population into 3 groups:

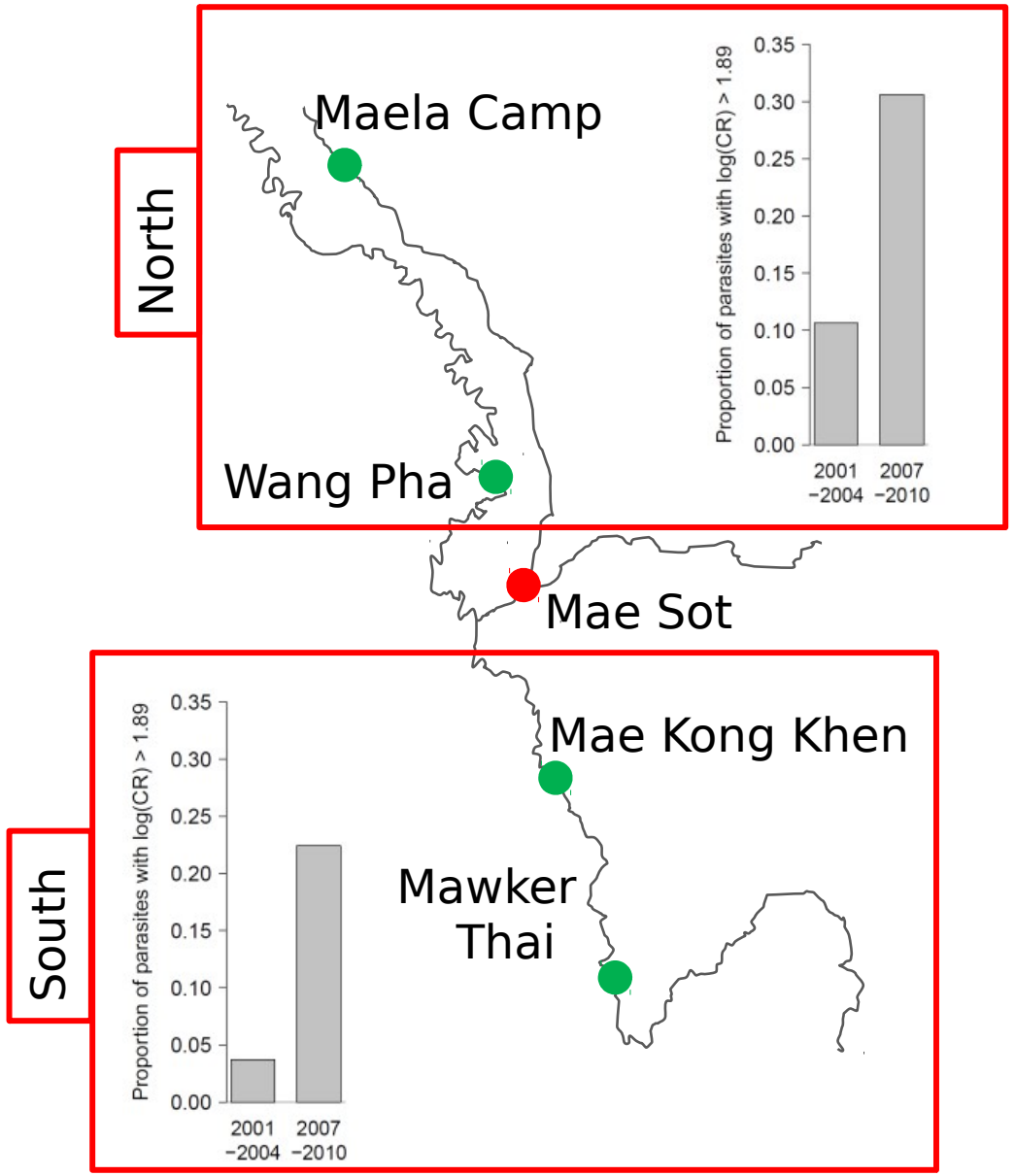
a) All parasites from 2001-2004 (no differences in phenotype detect between north and south)

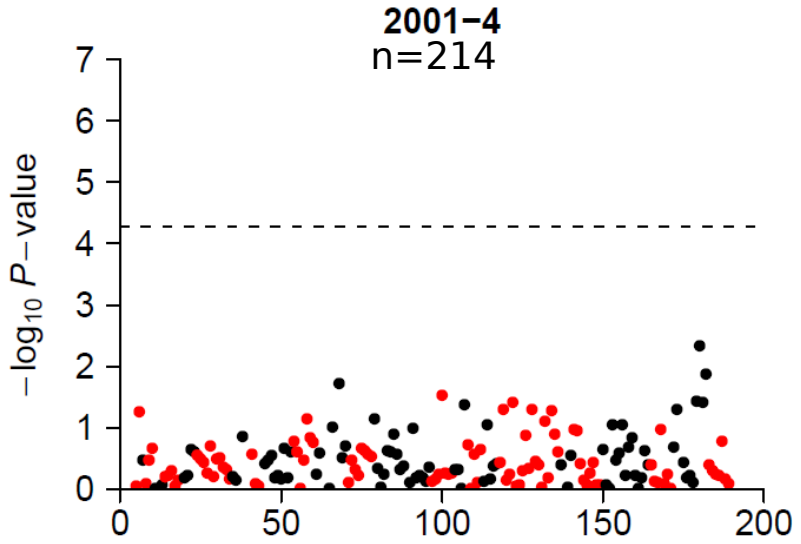
b) 2007-10 group of Maela and Wang Pha (north)

c) 2007-10 group of Mae Kon Ken and Mawker Thai (south)

b) and c) show significant differences in their distribution of phenotypes

Association mapping across selected loci

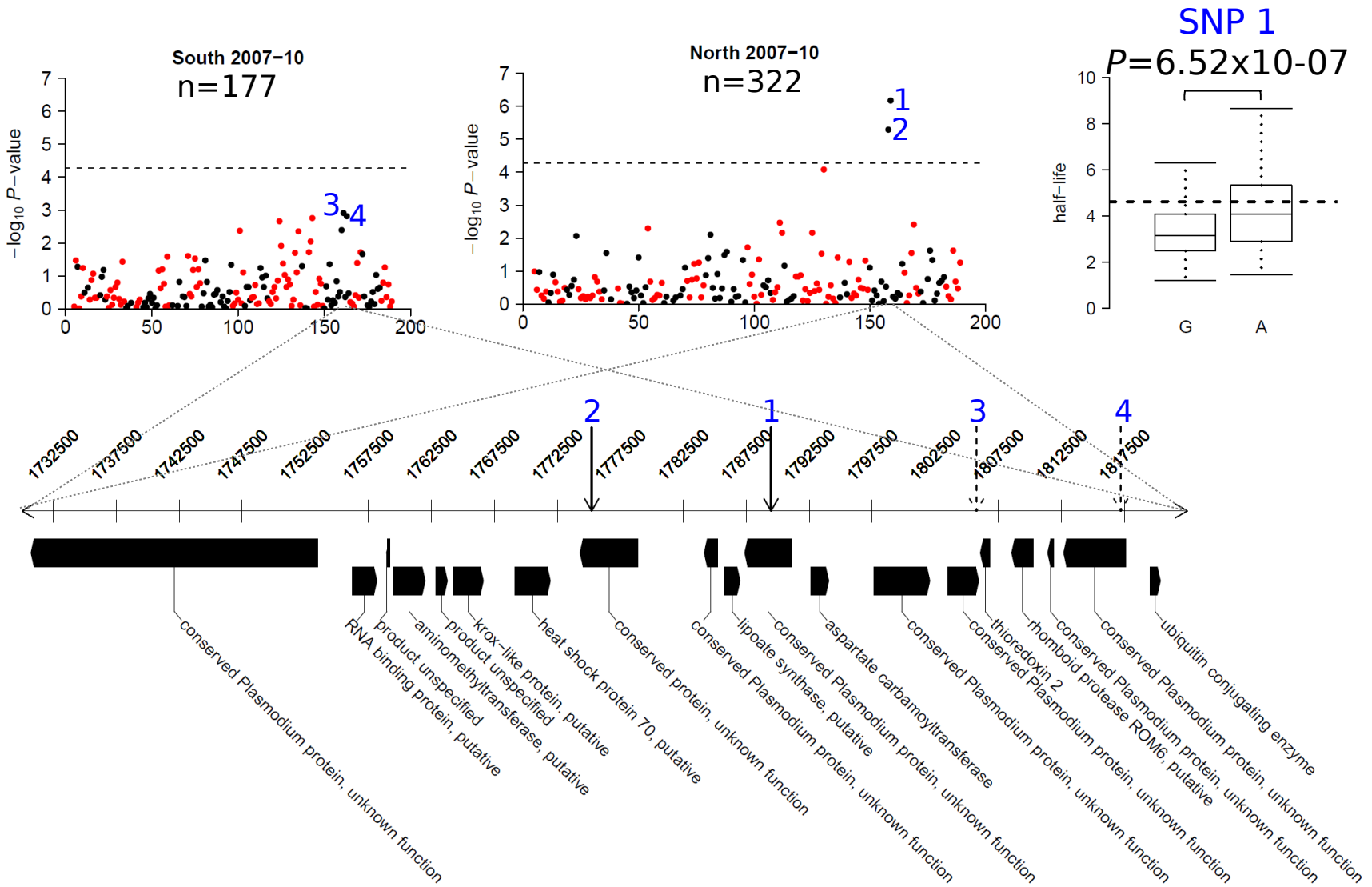




P-values from *t*-test for 96 SNPs from regions under selection **(Black** dots) and 96 neutral markers (**red** dots).

No significant association with artemisinin resistance in early timepoint- consistent with the absence of drug resistance alleles in the population

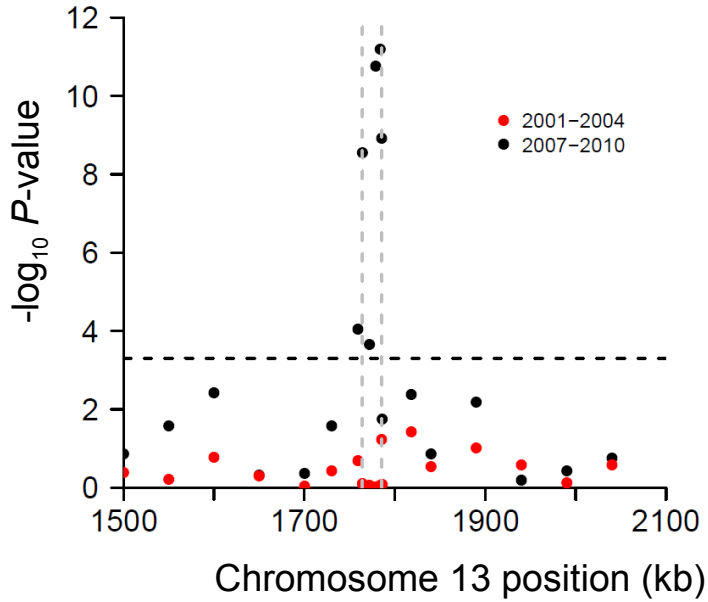
Association mapping across selected loci

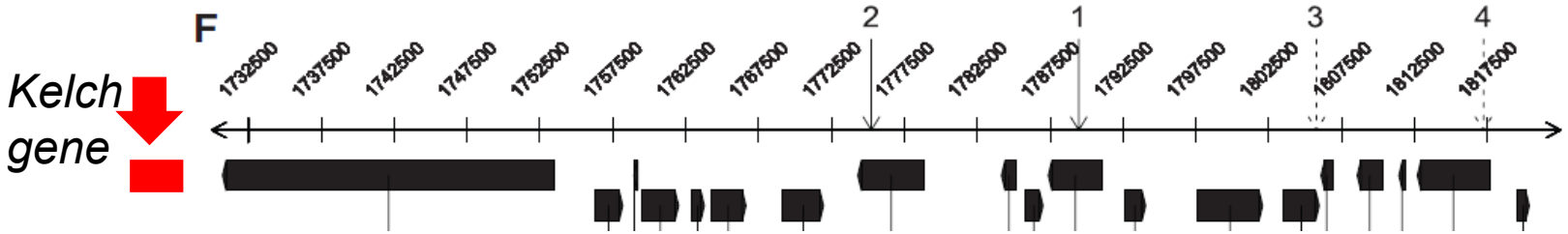


What gene(s) are involved?

A Major Genome Region Underlying Artemisinin Resistance in Malaria

Ian H. Cheeseman,[1] Becky A. Miller,[2] Shalini Nair,[1] Standwell Nkhoma,[1] Asako Tan,[2] John C. Tan,[2] Salma Al Saai,[1] Aung Pyae Phyo,[3] Carit Ler Moo,[3] Khin Maung Lwin,[3] Rose McGready,[3,4,5] Elizabeth Ashley,[3,4,5] Mallika Imwong,[4] Kasia Stepniewska,[4,5,7] Poravuth Yi,[8] Arjen M. Dondorp,[4,5] Mayfong Mayxay,[6] Paul N. Newton,[5,6] Nicholas J. White,[4,5] François Nosten,[3,4,5] Michael T. Ferdig,[2] Timothy J. C. Anderson[1]*





A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria

Frédéric Ariey[1,2]†, Benoit Witkowski[3], Chanaki Amaratunga[4], Johann Beghain[1,2]†, Anne-Claire Langlois[1,2], Nimol Khim[3], Saorin Kim[3], Valentine Duru[3], Christiane Bouchier[5], Laurence Ma[5], Pharath Lim[3,4,6], Rithea Leang[6], Socheat Duong[6], Sokunthea Sreng[6], Seila Suon[6], Char Meng Chuor[6], Denis Mey Bout[7], Sandie Ménard[8]†, William O. Rogers[9], Blaise Genton[10], Thierry Fandeur[1,3], Olivo Miotto[11,12,13], Pascal Ringwald[14], Jacques Le Bras[15], Antoine Berry[8]†, Jean-Christophe Barale[1,2]†, Rick M. Fairhurst[4]*, Françoise Benoit-Vical[16,17]*, Odile Mercereau-Puijalon[1,2]* & Didier Ménard[3]*

nature biotechnology

Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system

Mehdi Ghorbal[1–3], Molly Gorman[1,2], Cameron Ross Macpherson[1,2], Rafael Miyazawa Martins[1,2], Artur Scherf[1,2] & Jose-Juan Lopez-Rubio[1–3]

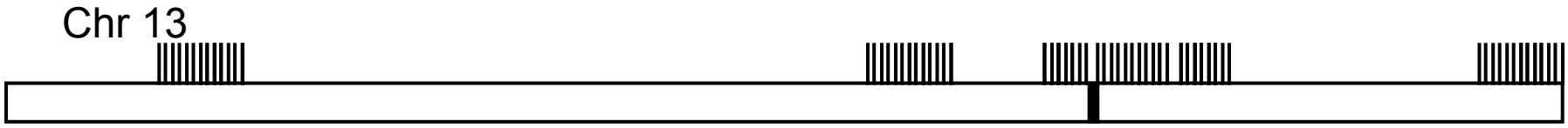
- We now know kelch is a major determinant of artemisinin resistance

- Our large longitudinal data will have captured the origin of this mutation and its spread through a population

- Does kelch fit our expectations of a drug resistance gene?



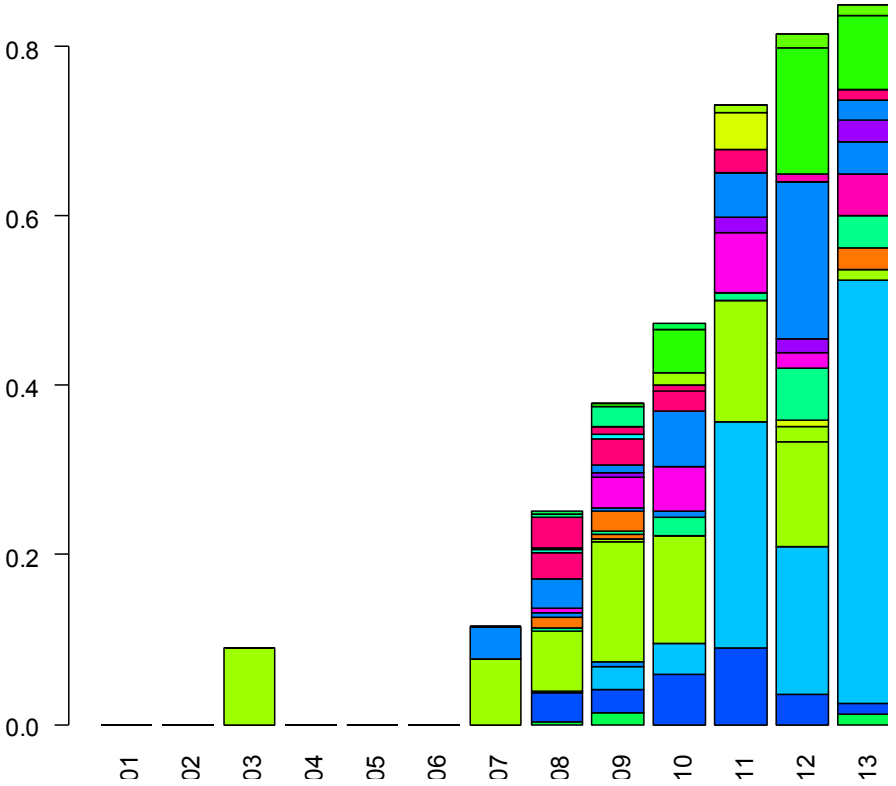

Study design

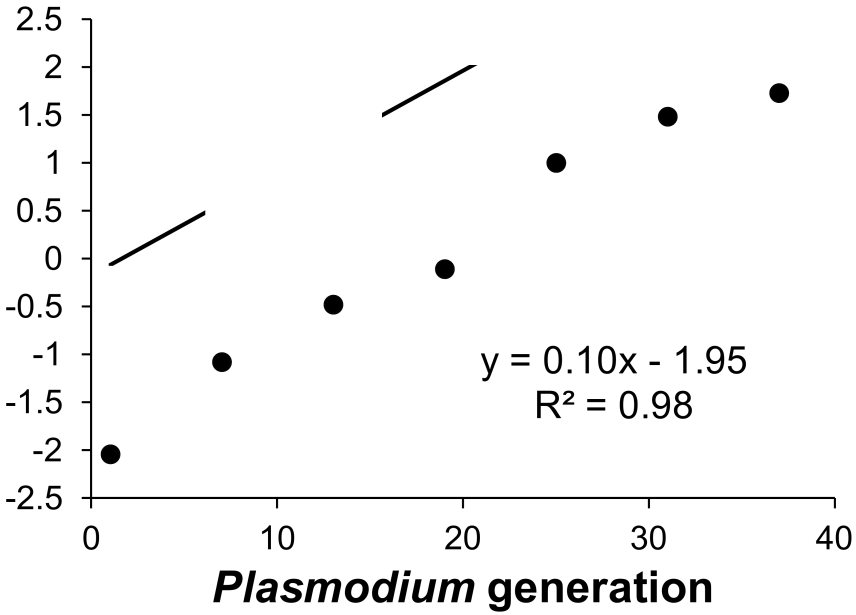


1138 infections with defined clearance rate (2001-13)

Sequence Kelch gene (726 amino acids)

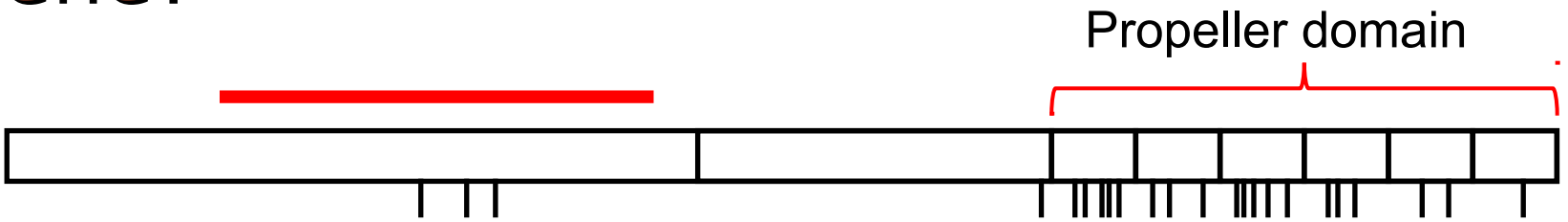
Golden Gate genotyping (74 flanking SNPs on chr 13)

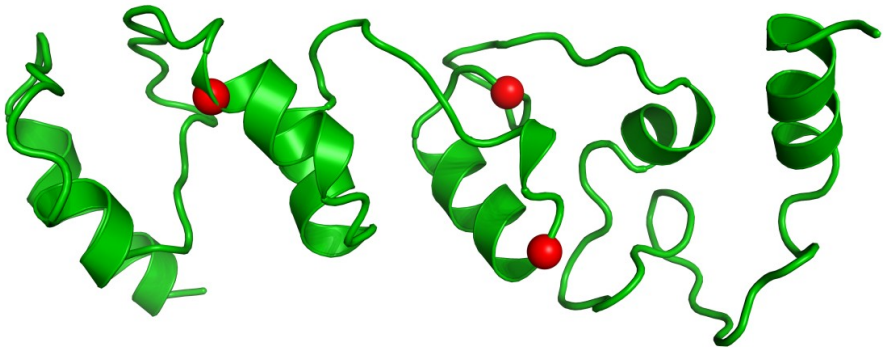


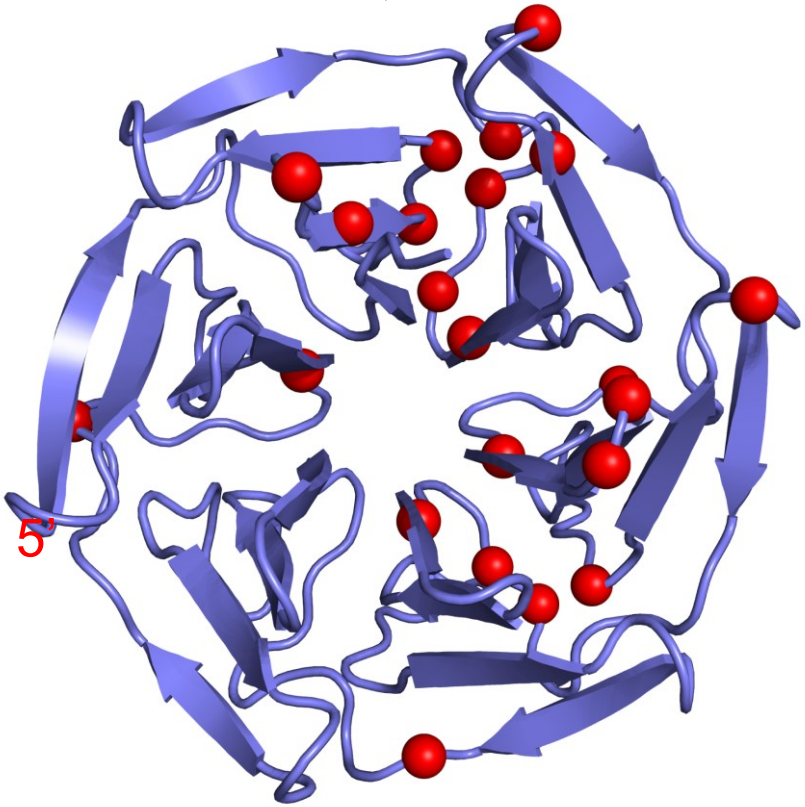


- Strong selection (s = 0.1)
- 27 derived mutations
- One mutation only per parasite genome

Where are the mutations in the kelch gene?





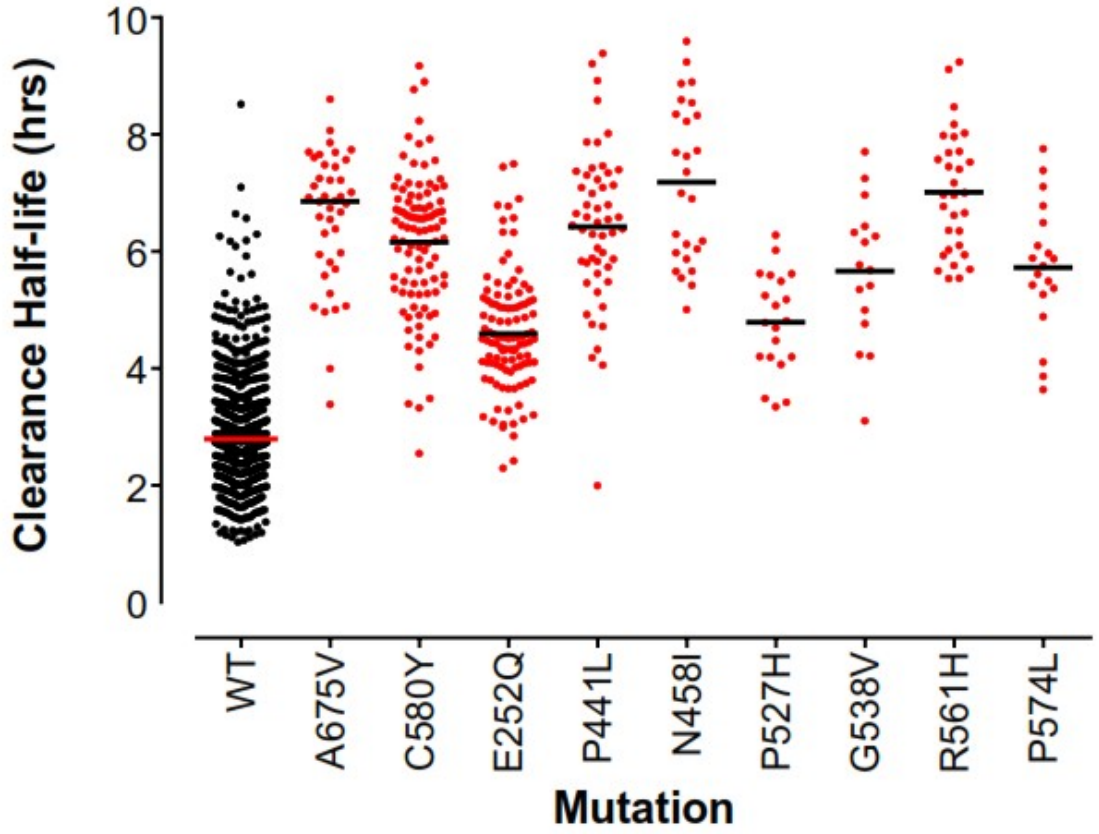


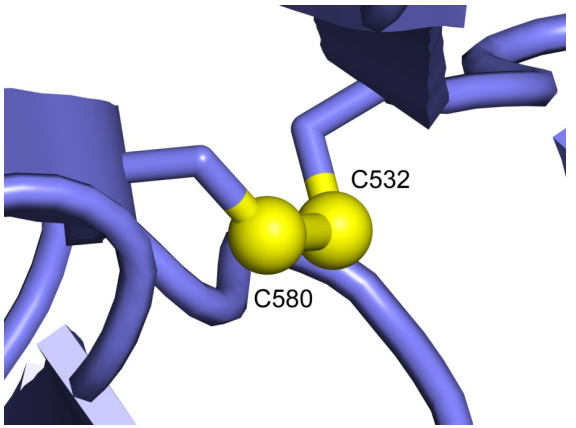
•Protein binding/oxidative stress

•Resistance SNPs in propeller domain and 5' region

•Loss-of-function (?) ➔ large mutational target (~838bp)

Impact on parasite clearance phenotype





- Resistance alleles differ in phenotype
- C580Y has moderate impact on T1/2, but spreading rapidly in SE Asia
- Disrupts a disulfide bond – impacts protein rigidity

What do we expect to see?

Population mutation parameter $\Theta = 2N_e\mu = 0.16$

$T_o = 1995$

N_e = ~10000 *(Joy et al. Science 2003)*

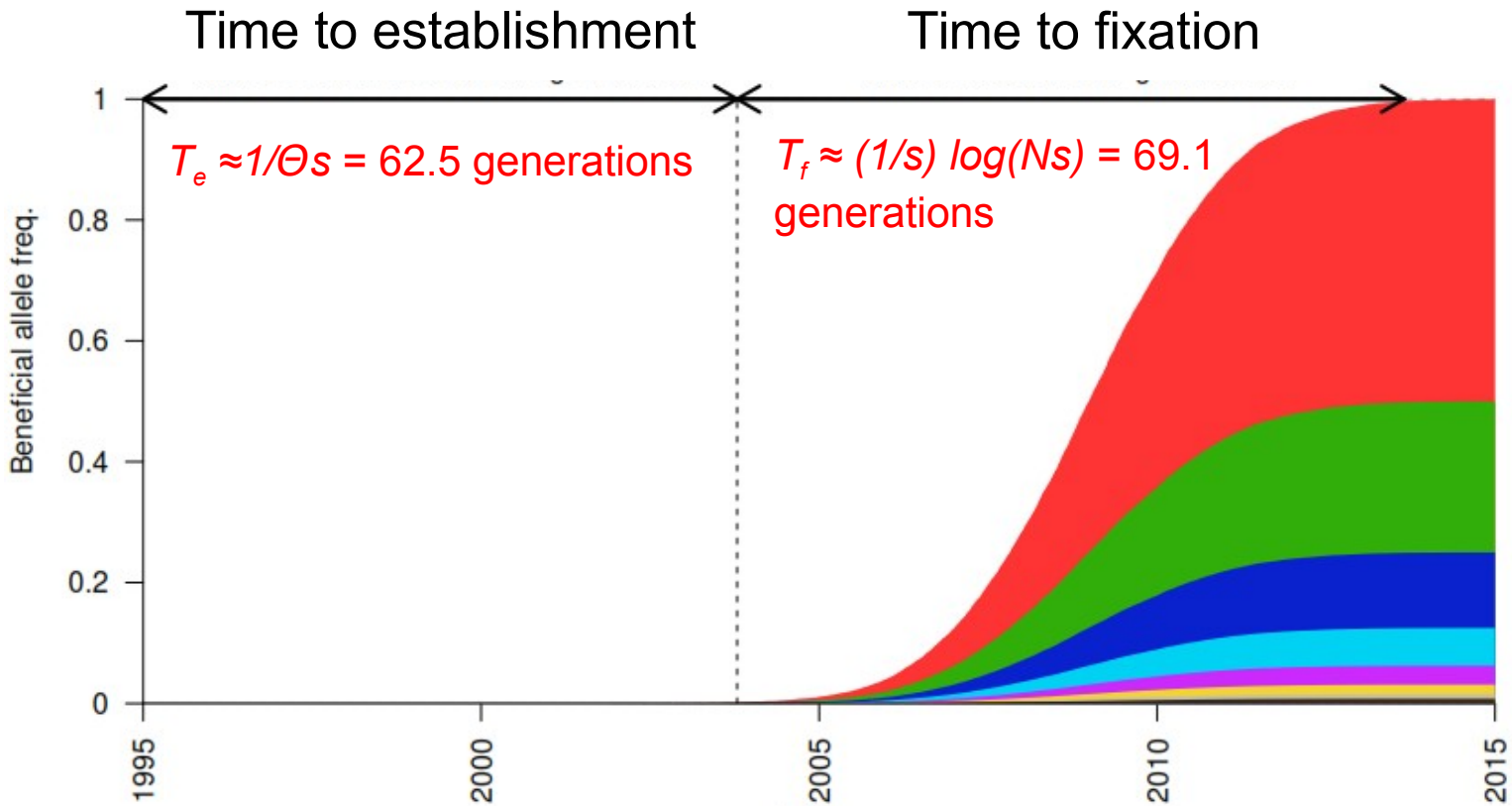
Generations per year = 7

$s = 0.1$

$\mu = 9.7 \times 10^{-9}$ *(Bopp et al. PloS Genetics 2013)*

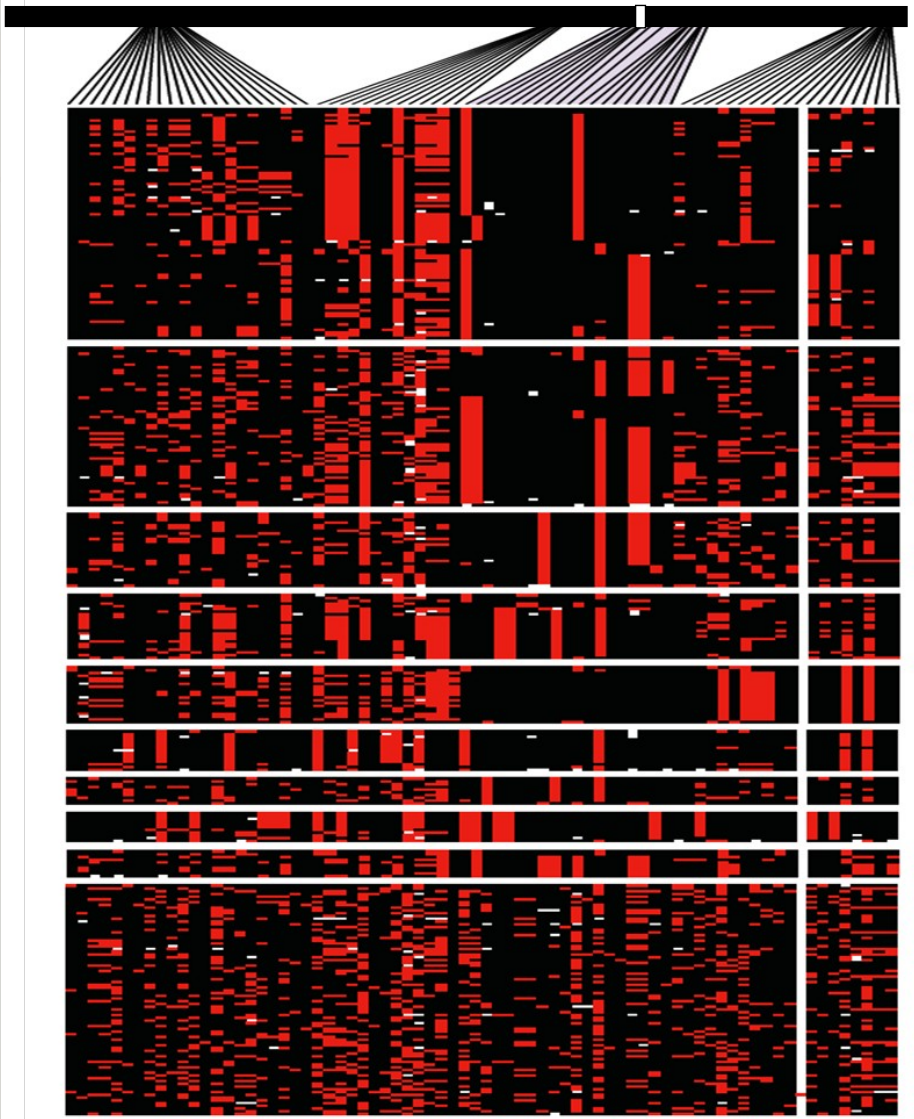
target size = 838bp

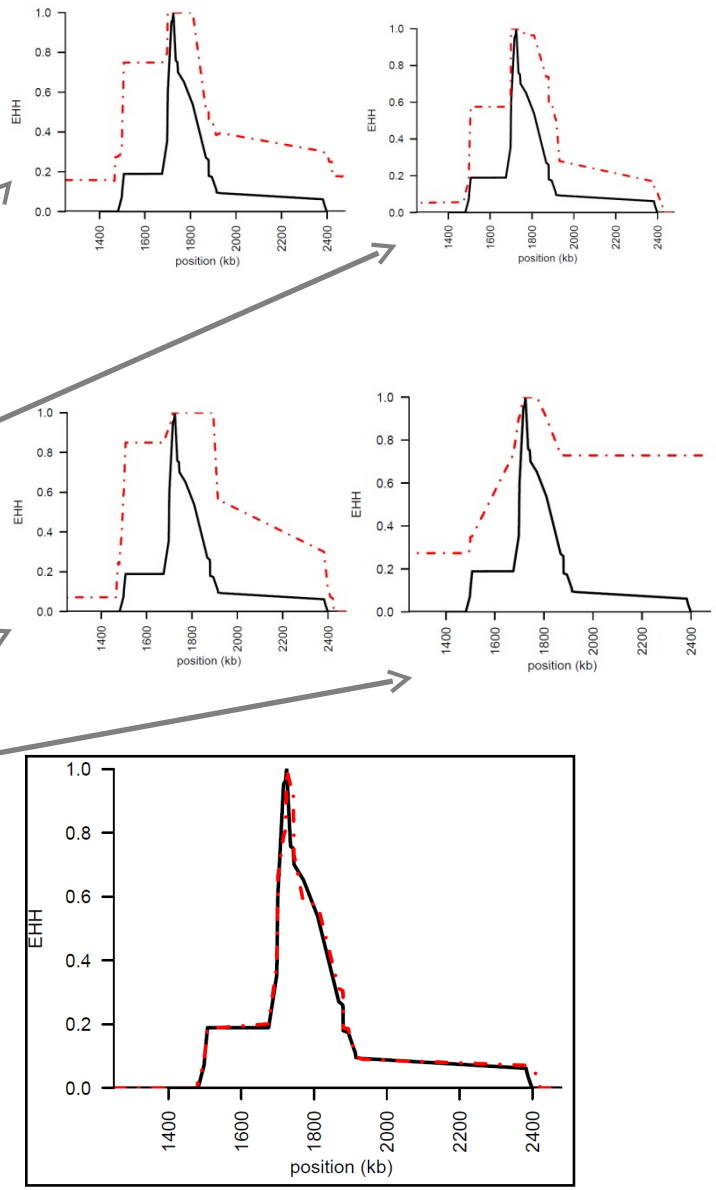
Soft sweeps likely when $\Theta \geq 0.1$ (Hermission and Pennings Genetics 2005)



Reasonable concordance between theory and observation

Impact on flanking LD



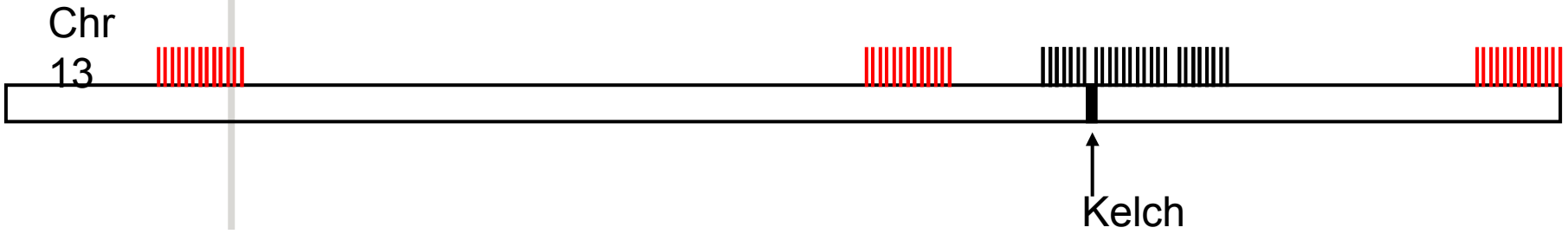


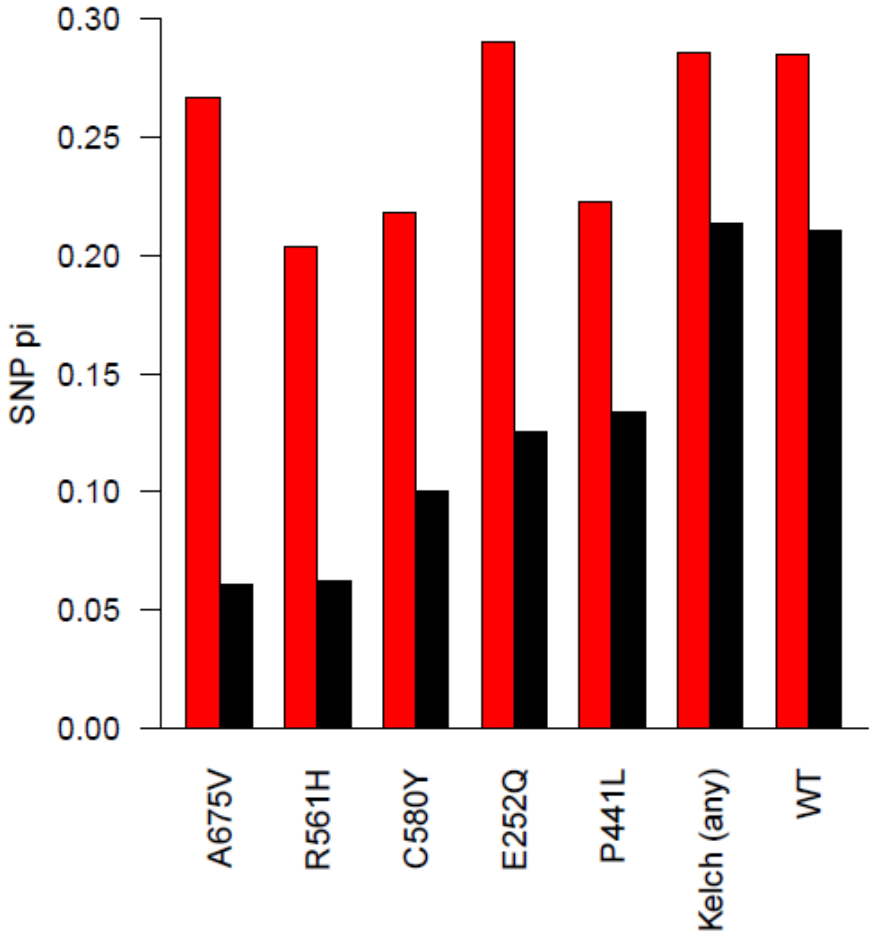
Distinctive haplotypes and elevated EHH around each resistance SNP

No difference between EHH for resistance vs WT alleles

➔ standard methods will be underpowered to detect this sweep

Impact on flanking diversity



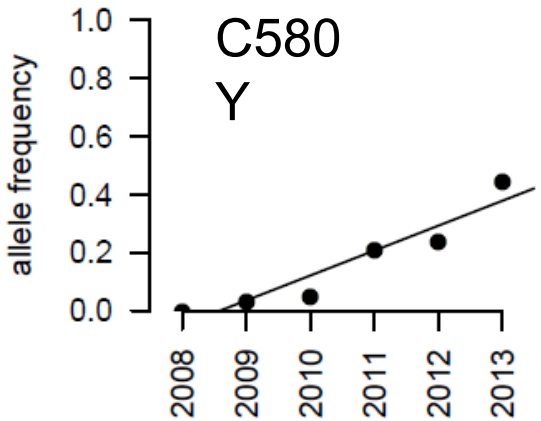


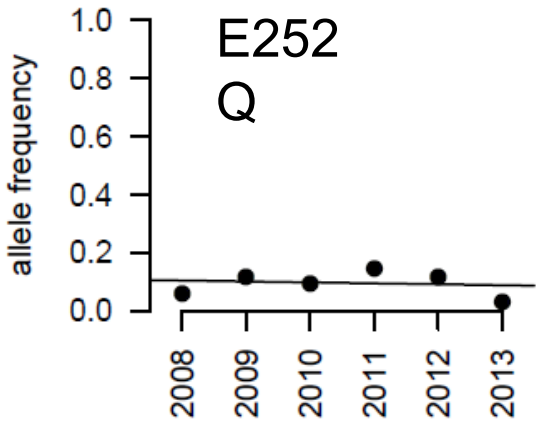
Individual resistance mutations have reduced diversity at flanking SNPs

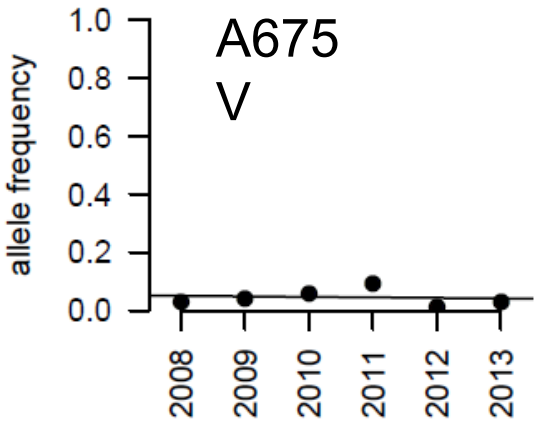
No reduction in diversity between wildtype and all mutant kelch alleles

Scans based on diversity will perform poorly

Dynamics of resistance alleles







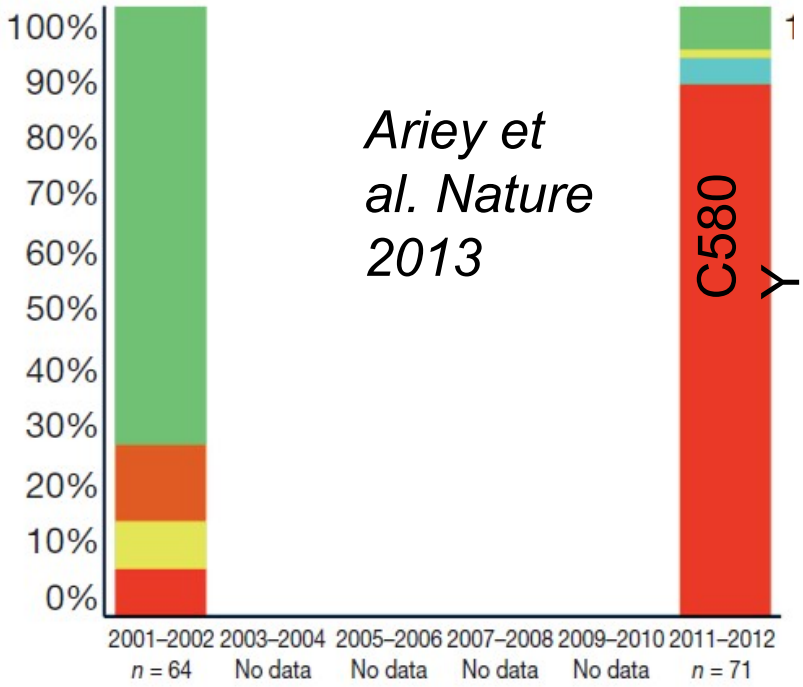
C580Y allele is increasing in frequency – now at 45% in Thailand

Other alleles at <20% frequency

In Cambodia, C580Y allele is at 90%

➔One allele replaces others

➔ Is soft sweep is becoming hard?



Ariey et al. Nature 2013

Associating kelch genotypes to clearance rate

- Typical association analysis takes a single marker at a time and associates it with a phenotype

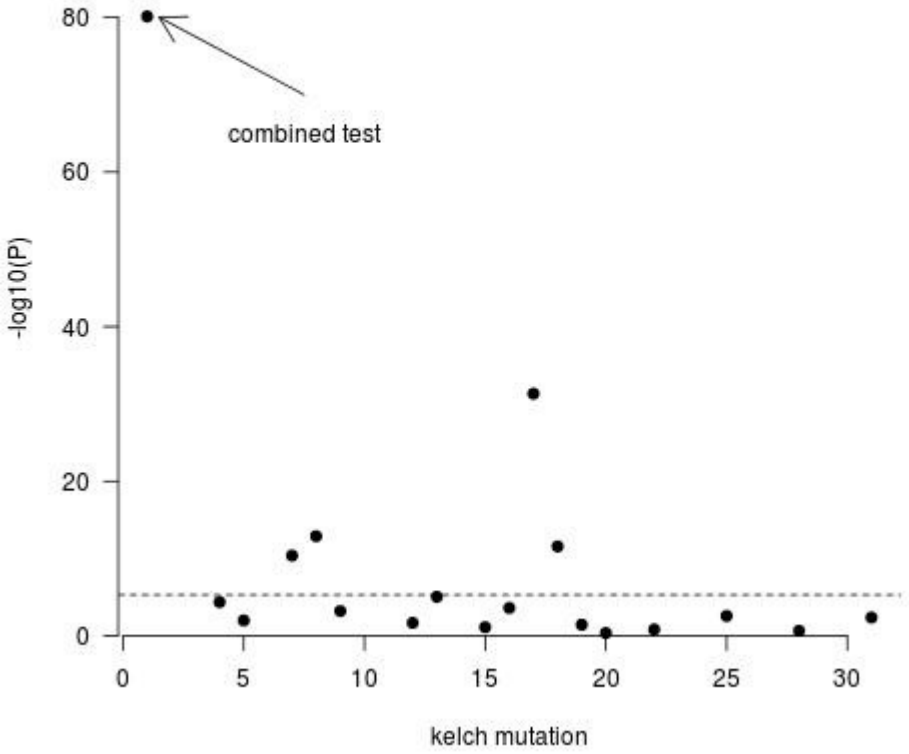
- When there are multiple, beneficial mutations in a single gene we may easily confound this analysis as each mutation explains a far smaller proportion of the phenotype than a single marker

- We may easily either miss an important effect or underestimate the importance a gene plays in resistance

Burden tests

- Burden tests are an emerging family of statistics which combine the effects of multiple markers in an association study

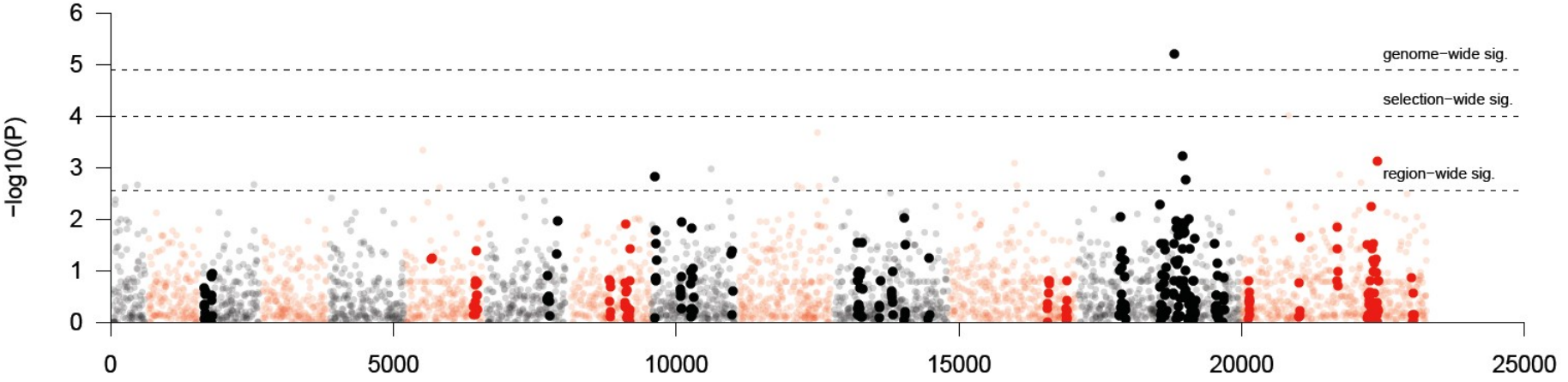
- If we compare single marker tests against these multi-marker tests we see a substantial increase in power



Genome-wide burden tests

- We performed whole genome sequencing on 38 isolates (22 slow clearing, 16 fast clearing) and applied SKAT-O tests to all non-synonymous SNPs for each gene

- Even with this small sample size we have enough power to detect kelch directly



Conclusions

- We can rapidly identify the genes which underlie the emergence of resistance
- Understanding the selective sweeps generated by the spread of resistance mutations can help us design better approaches to finding them