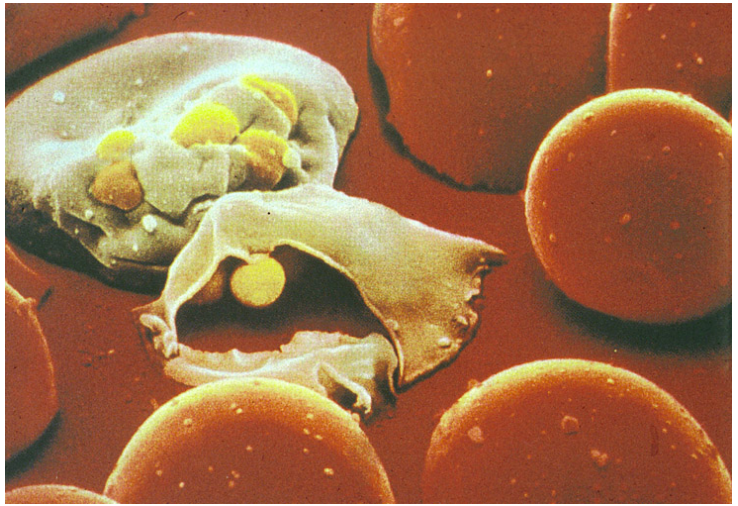
HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH

BROAD INSTITUTE

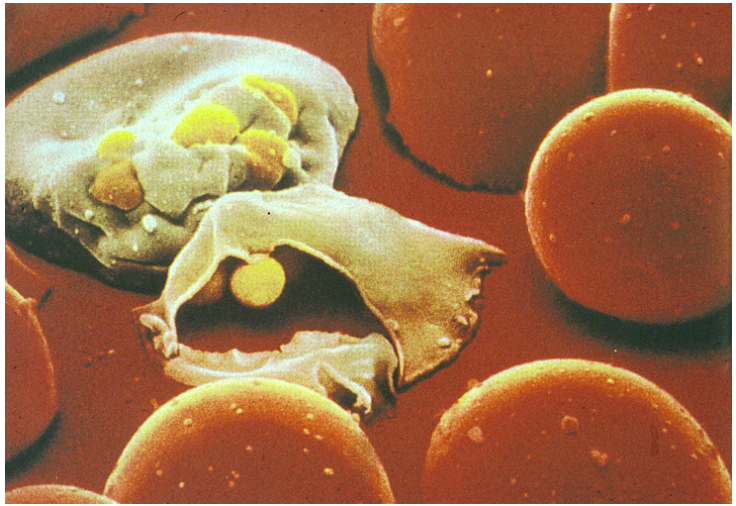
How Genetics Helps Us Understand Malaria

Malaria: From Innovation to Eradication
Pre-Meeting Workshop



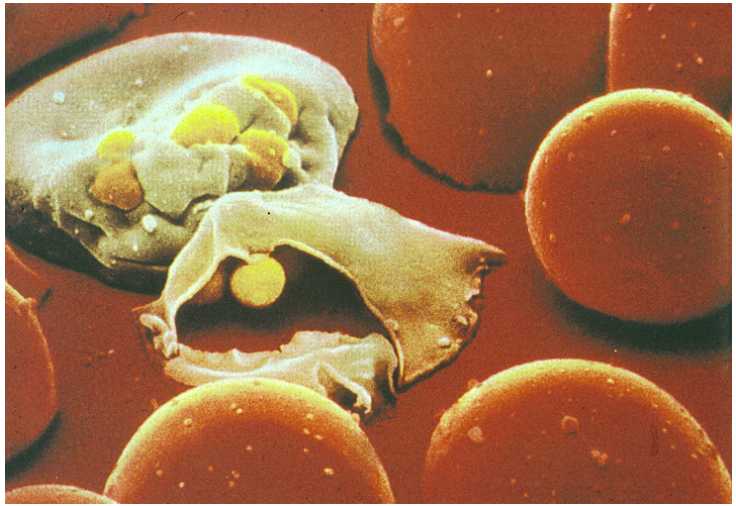
Malaria Innovation to Eradication: Genetic Concepts

- Introduce and provide context for genetics related to topics to be discussed at the meeting.
- Use of genetics to assess transmission reduction and identify sources or reservoirs of incident infection.
- Use of genetic data in the context of epidemiological modeling of transmission declines and rebounds.
- Genetic signatures of intervention use and impact assessment.
- Tracking and surveillance using genetics for enhanced resolution.

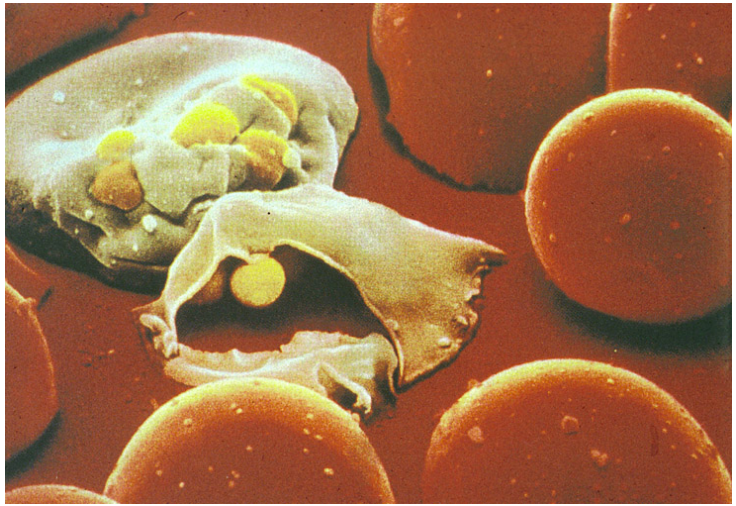


Malaria Innovation to Eradication: Genetic Concepts

1. Genetics toolkit and general considerations.
2. Use of genetics to identify drug resistance loci—genome wide association studies and chemogenomic strategies.
3. Using genetics and epidemiological modeling to detect transmission declines and rebounds.
4. Assessment of intervention impact and expected signatures of intervention impact.
5. Tracking parasites to detect reservoirs, sources of incident infection, and map transmission networks.



Concept 1: Genetic toolkit and general considerations



Genetics Glossary

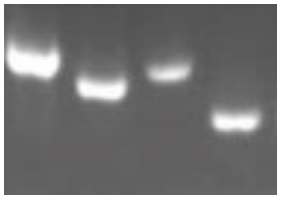
- Genotyping—detection of genetic variants
- Microsatellite marker—repetitive DNA sequence made up of repeats 2 – 8 nucleotides in length.
- Single nucleotide polymorphism (SNP)—changes in nucleotide sequence
- Haplotype—set of linked variants
- Monogenomic—one parasite genome
- Polygenomic—more than one parasite genome
- Genome Wide Association Study (GWAS)—identification of genetic variants associated with a specific phenotype using whole genome information
- Directional selection—selection favoring one allele over another that can rise in frequency
- Diversifying selection—selection for multiple alleles maintained in the population at frequencies longer than expected by genetic drift.

Questions to Ask Before Starting

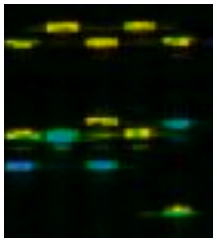
- What is the basic biological question and hypothesis? Are genetic approaches appropriate for testing this hypothesis?
- Given the basic biological question, what is the best strategy or approach?
- What is known about parasite population structure?
- What are the best markers for this question?
- What type of samples and sampling scheme should be used?
- How many samples are needed?
- What type of analysis will be used given the sampling scheme, the sample number, the data type, and the question of interest?
- Take Home Lessons:
  - One Size Does NOT Fit All
  - Do Your Homework

Genetic Markers

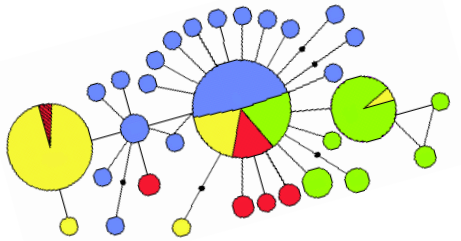
- Highly polymorphic (usually antigen) markers



- Microsatellite markers

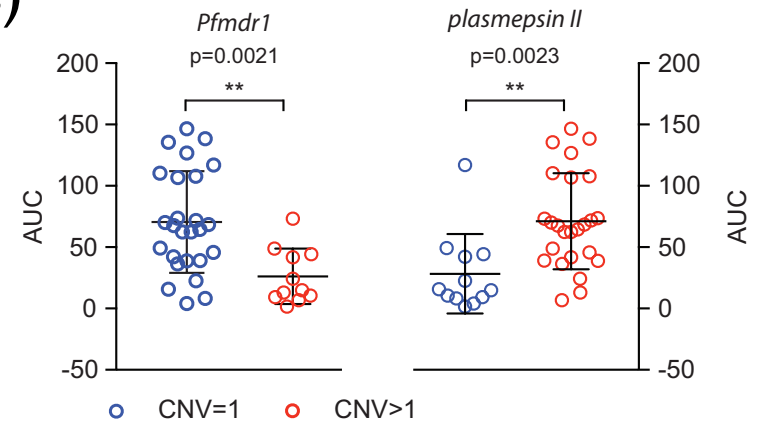


- Organelle makers—mitochondria and apicoplast



- SNP markers

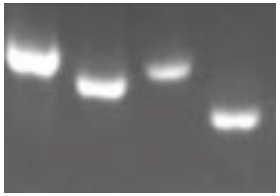
- Copy number variants (CNVs)

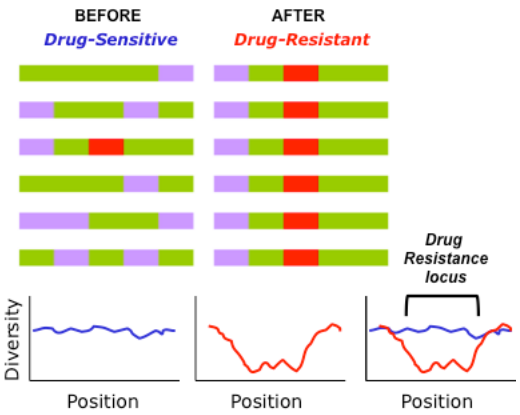


- Sequencing

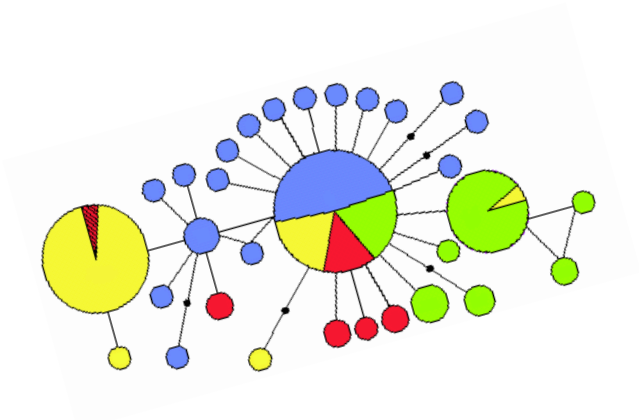
Genetic Markers

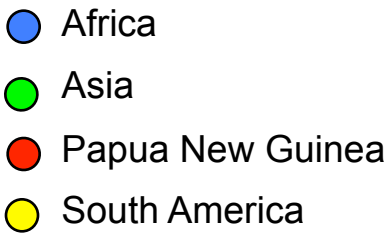
- Highly polymorphic (usually antigen) markers
  - *msp1, msp2, glurp*
  - monogenomic vs. polygenomic
  - reinfection vs. recrudescence

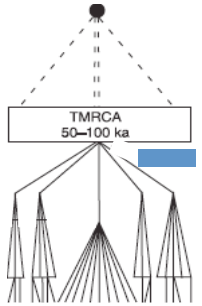




- Microsatellite markers
  - high mutation rate
  - highly related parasites can be tracked
  - useful for tracking region of genome—drug resistance loci
  - hard to distinguish between identity by state or identity by descent
- Organelle makers—mitochondria and apicoplast
  - Good for tracking population history because no recombination

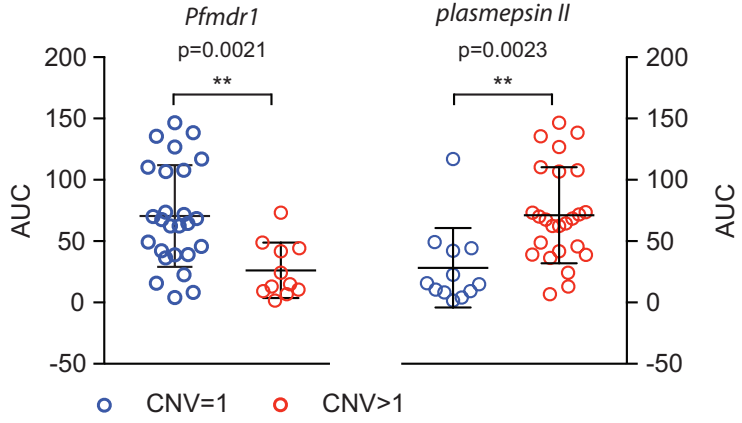




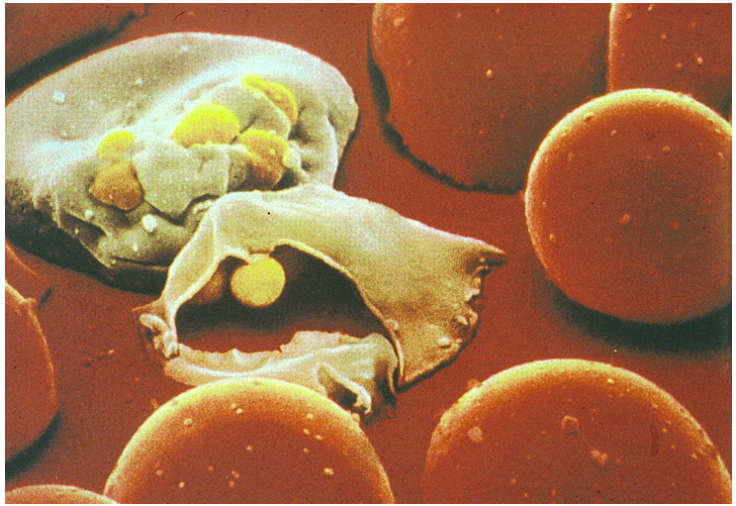


Genetic Markers

- SNP markers
  - Monogenomic vs. polygenomic
  - Distinguish identity by descent
- Copy number variants (CNVs)
  - Found associated with key drug resistance phenotypes
  - May require drug pressure to maintain CNVs
- Sequencing
  - Simultaneous sampling of multiple variants
  - Highly polymorphic low-complexity sequence regions (e.g., vars) are challenging.
  - New computational approaches being developed to infer CNVs, and to disentangle polygenomic infections.



Concept 2: Drug Resistance Loci Discovery: GWAS & Chemogenomics

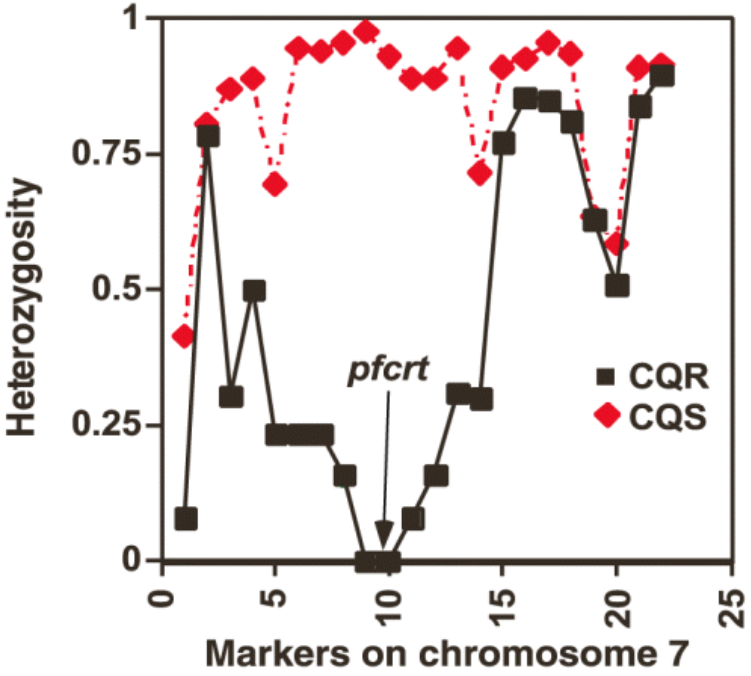


Genomic Approaches to Identify Drug Resistant Loci

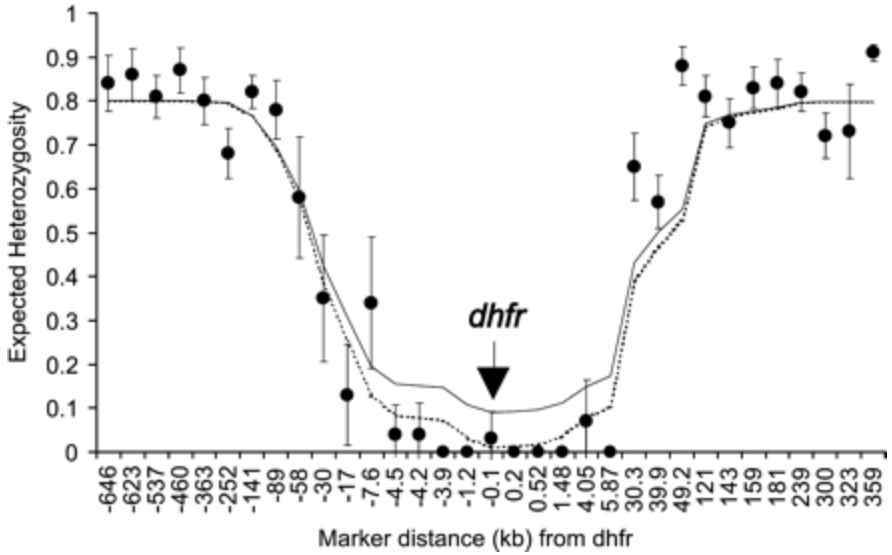
- *Association*
  - Which genetic variants are associated with a specific phenotype?
  - Utilize panel of isolates with either drug sensitive or resistant phenotypes, along with genome-wide genotyping or sequencing data.
- *Selection*
  - Drug pressure selects for parasites that harbor drug resistant variants.
  - Scan genome for "selective sweeps" to find regions with low genetic diversity among drug resistant parasites, as compared to drug sensitive ones.
- *Linkage*
  - Use highly related parasites to find regions of the genome that are inherited with a specific phenotype.
  - Apply to progeny of a genetic cross between drug resistant and drug sensitive parents.

Drug Resistance Discovery: Selective Sweep

- Compare genetic variation between drug sensitive and resistant parasites across a specific genomic region.
- Look for regions where there is reduced diversity—selective sweep—among drug resistant parasites.
- Identification of selective sweep for drug resistance loci including *pfcrt* and *dhfr*.



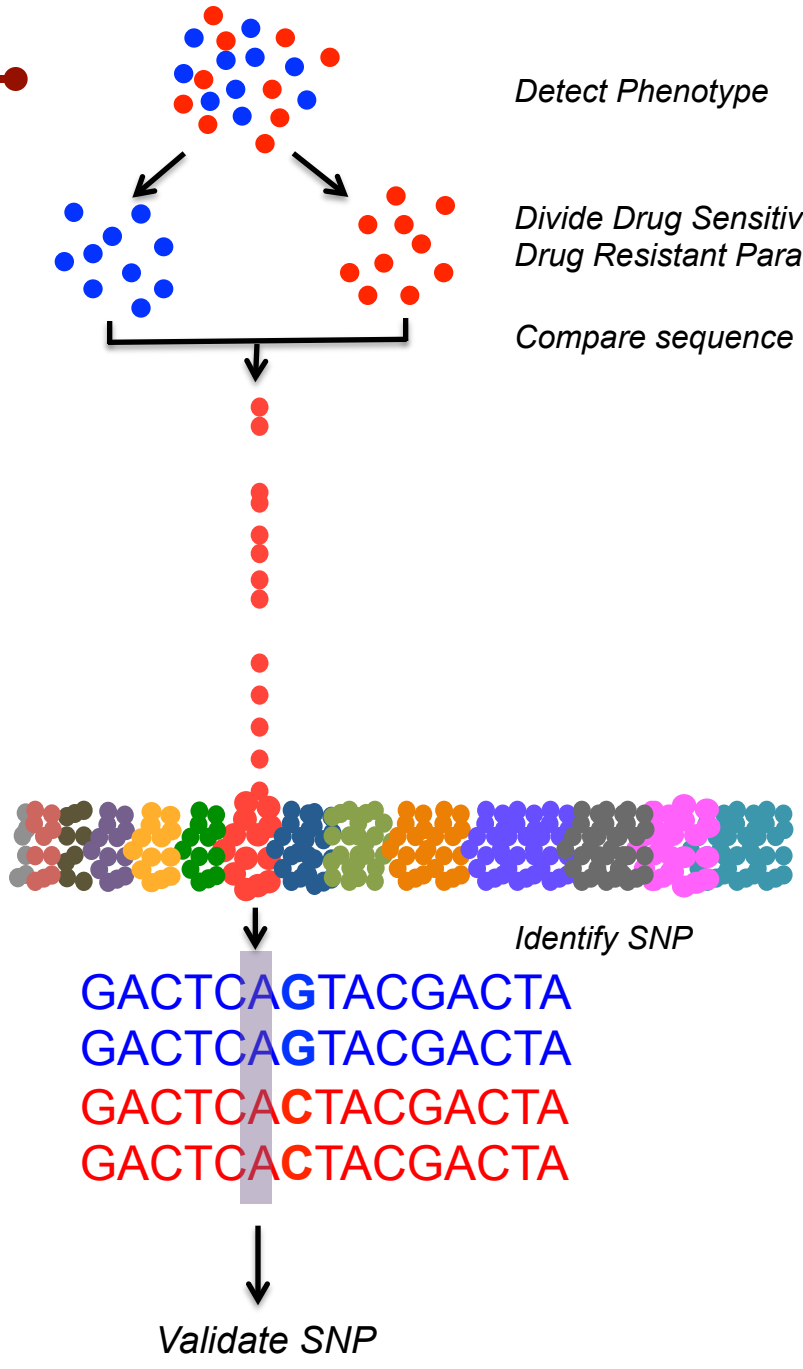
Su & Wootton, MolMicro (2004)



Nair et al, MBE (2003)

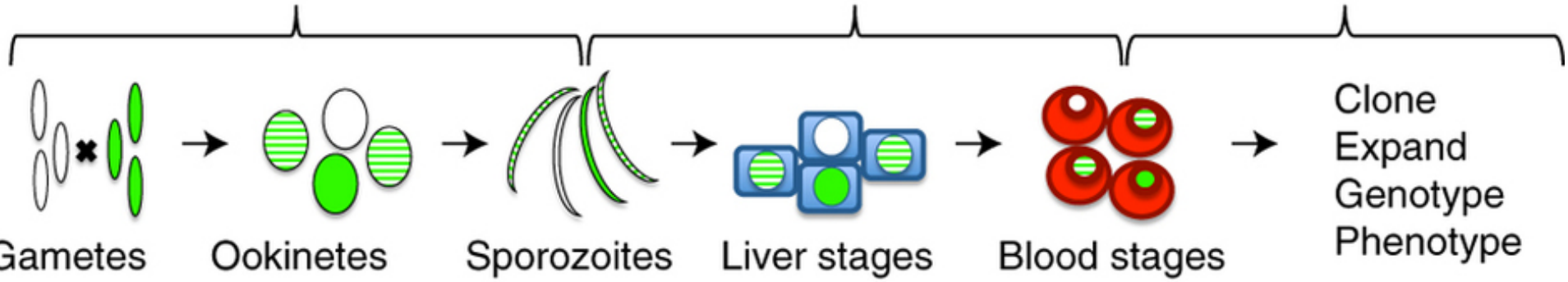
Drug Resistance Discovery: GWAS

- Sequence drug sensitive and drug resistant parasites
- Compare sequence data
- Identify SNPs associated with drug resistant phenotype but absent from drug sensitive parasites
- Carry out validation of SNP using gene-editing strategies.
- Used to identify *pfcrt*, *pfkelch13*, and other genetic loci

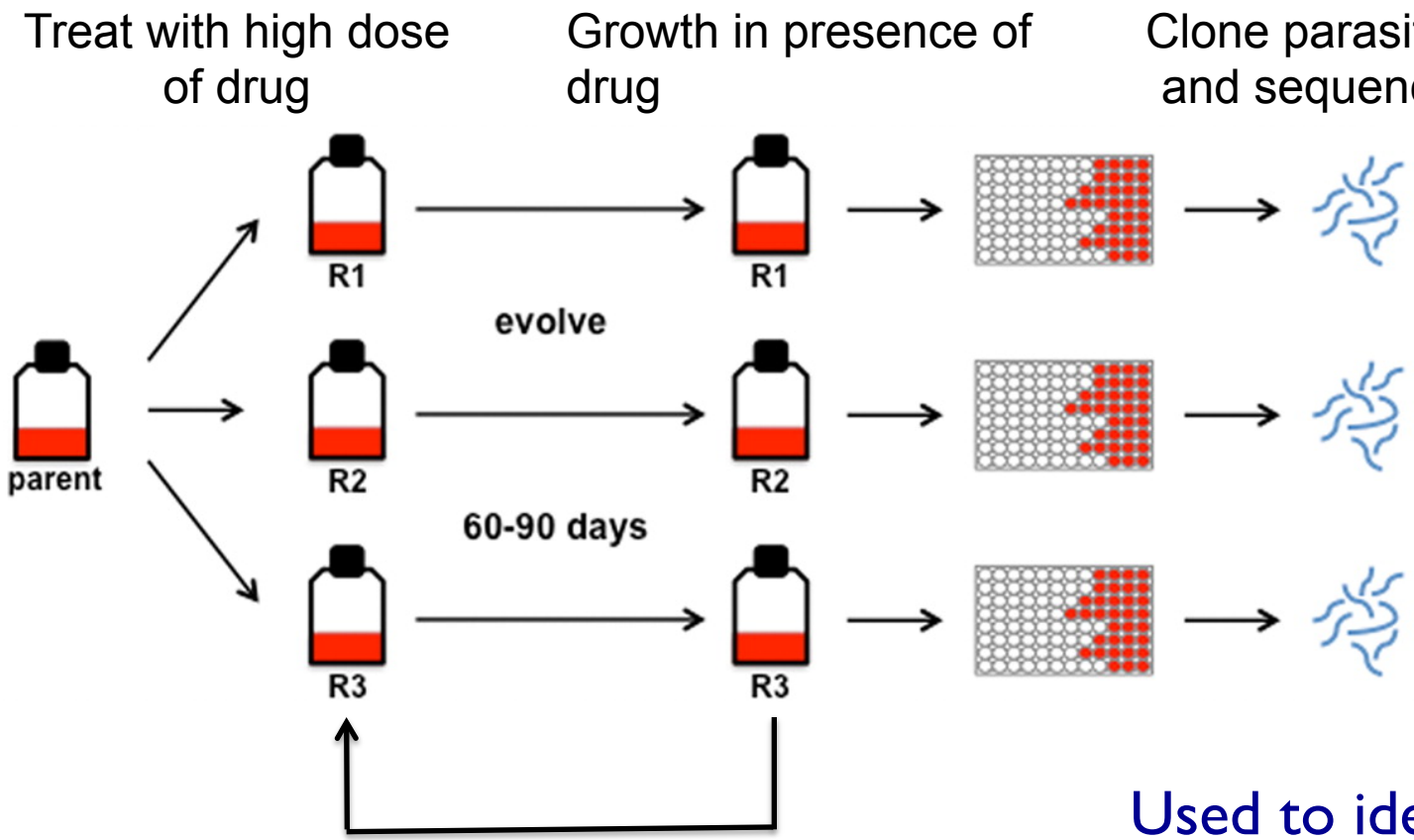


Drug Resistance Discovery: Genetic Cross

- Carry out genetic cross in animals between drug resistant and drug sensitive parasites.
- Isolate progeny and sequence.
- Identify genomic regions that are linked with drug resistance.
- Used to identify *pfcrt*.

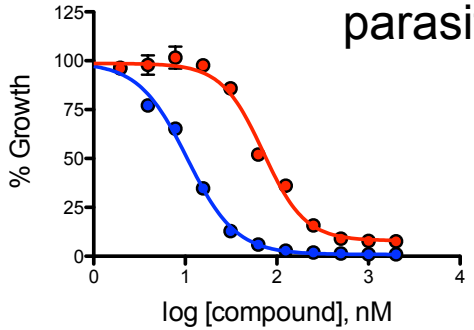


Drug Resistance Discovery: Chemogenomics



Used to identify *pfkelch13*

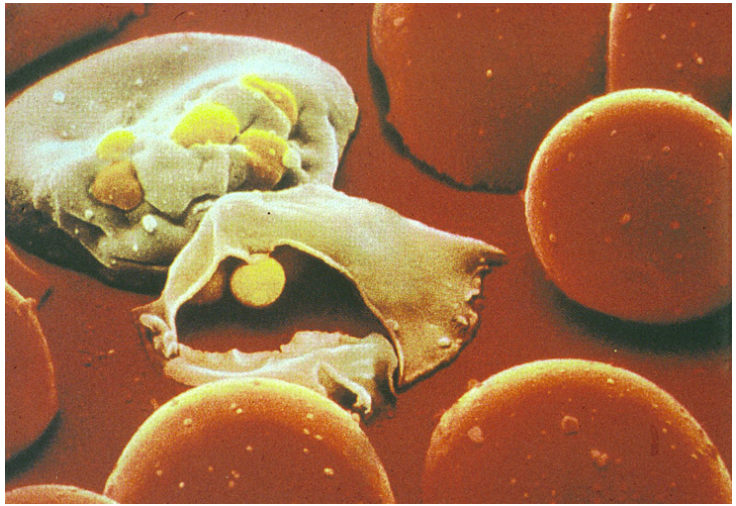
Remove drug, allow parasites to recover



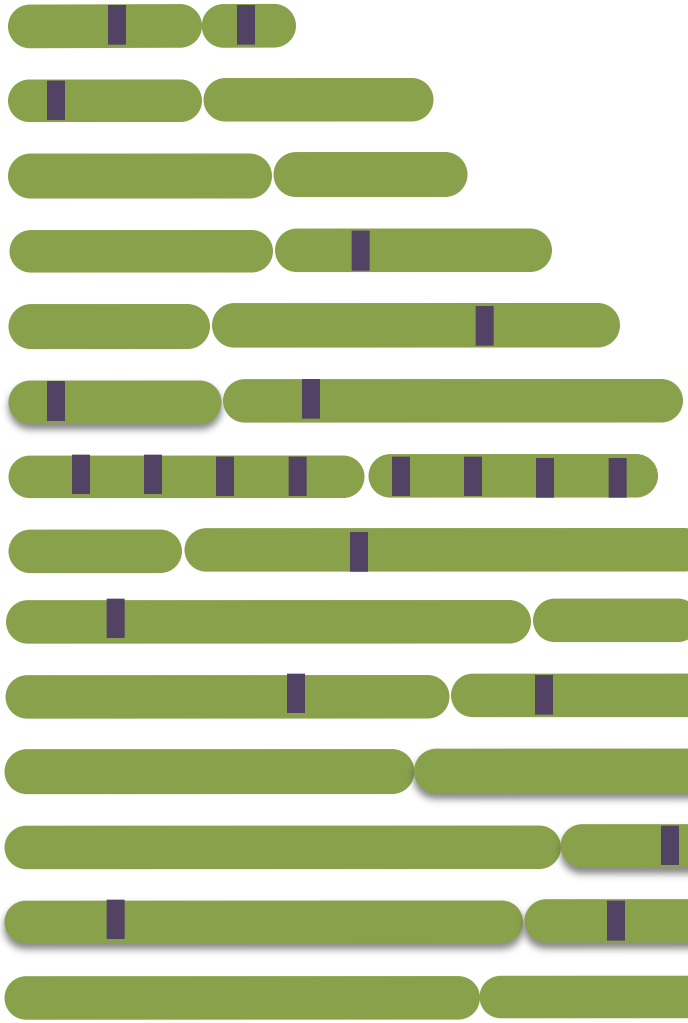
	1432				1444					1461
Dd2	a c a	t c t	c a t	t a t	c t t	g g a	a c a	a a t	t t t	g c a
	Thr	Ser	His	Tyr	Leu	Gly	Thr	Asn	Phe	Ala
HFGR I	a c a	t c t	c a t	t a t	c a t	g g a	a c a	a a t	t t t	g c a
	Thr	Ser	His	Tyr	His	Gly	Thr	Asn	Phe	Ala
HFGR II	a c a	t c t	c a t	t a t	t t t	g g a	a c a	a a t	t t t	g c a
	Thr	Ser	His	Tyr	Phe	Gly	Thr	Asn	Phe	Ala

Figure from Flannery et al., J Med Chem (2013).

Concept 3: Epidemiological Modeling: Detecting Transmission Declines and Rebounds



Molecular barcode to fingerprint parasites



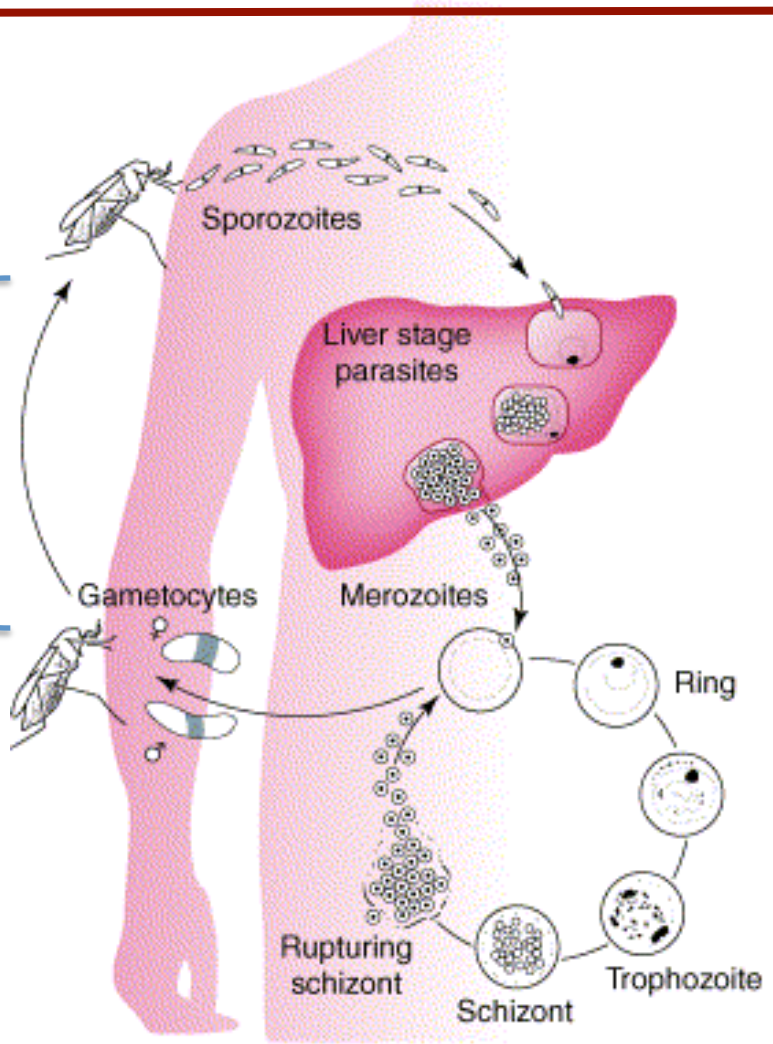
C G A G C T G A A T C G A C T C C T A G A G C T 7G8
T G G G T A G A G A C A A C T A T G C A A C C T D6
C A A A T A C G A T A G T A C A T T C A T G A A Dd2

Daniels et al, Malaria Journal 2008, 7:223

Plasmodium lifecycle

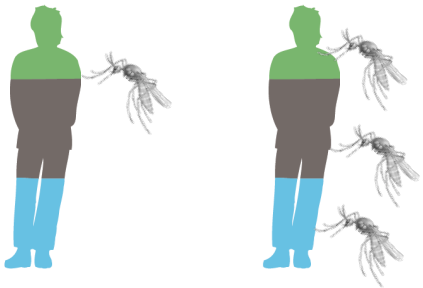
~10 days

Mosquitoes bite once per cycle



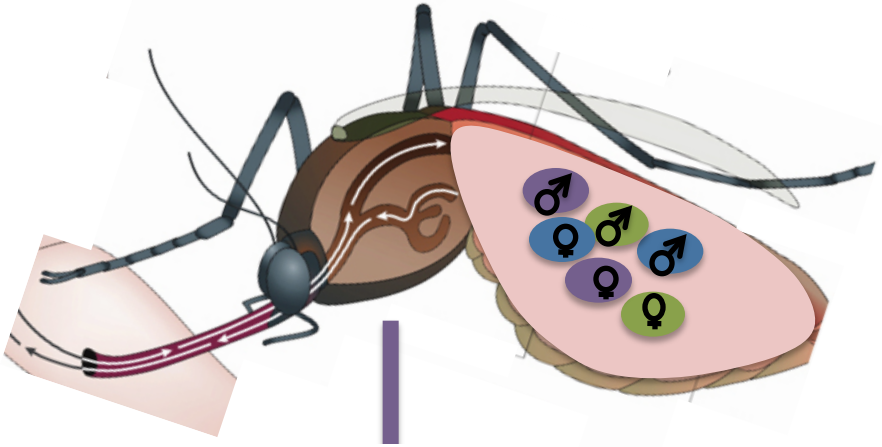
COI = 1

COI = 2



Co-transmission vs. Super-infection

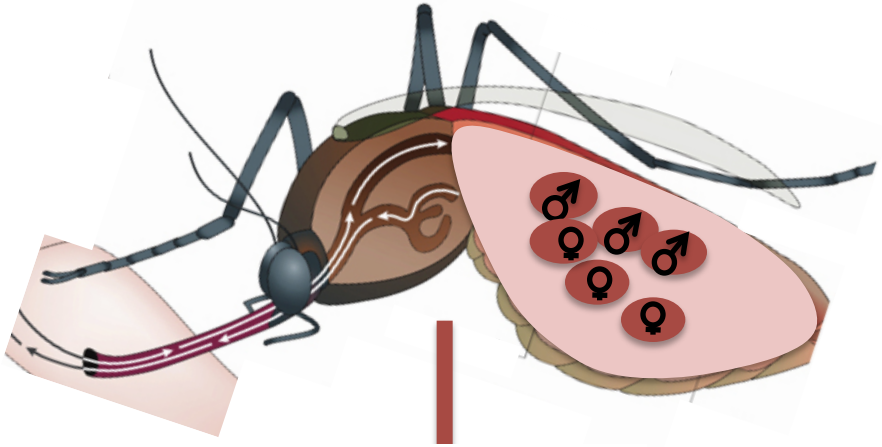
Genetic consequences of outcrossing vs. inbreeding



Outcrossing

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
I	T	G	C	C	C	C	A	G	A	T	C	A	C	A	A	C	T	A	A	G	A	T	T	T
II	T	A	T	C	C	G	A	A	T	T	T	A	T	C	A	A	T	A	C	A	A	C	G	T
III	C	A	C	C	C	G	G	G	A	T	T	A	C	A	A	A	C	A	A	G	C	T	T	T

COI = 3



Inbreeding

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
I	C	G	C	T	C	C	G	G	A	C	T	G	C	A	C	C	C	A	A	G	A	T	T	G
II	C	G	C	T	C	C	G	G	A	C	T	G	C	A	C	C	C	A	A	G	A	T	T	G
III	C	G	C	T	C	C	G	G	A	C	T	G	C	A	C	C	C	A	A	G	A	T	T	G

COI = 1

Molecular barcode data from Senegal

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
S1	T	A	C	T	C	C	G	A	T	T	C	A	C	A	A	C	T	T	C	G	A	C	T	G
S2	T	A	C	T	C	C	A	A	A	T	T	G	C	C	C	A	C	A	A	G	A	T	T	G
S3	T	A	T	T	C	C	A	G	A	T	C	G	C	A	A	C	C	A	A	G	A	T	T	G
S4	T	A	T	T	C	G	A	G	A	C	T	G	T	A	C	C	C	T	C	G	C	C	T	G
S5	T	G	C	T	C	C	A	G	A	T	T	A	T	A	A	C	T	T	C	G	A	C	T	G
S6	T	A	T	T	C	C	G	G	T	T	T	G	C	A	C	C	T	A	A	A	A	T	T	G
S7	T	A	C	T	G	C	A	G	A	T	T	G	T	C	A	C	T	T	A	A	A	C	T	G
S8	T	A	C	T	C	C	G	A	A	C	T	G	T	C	C	C	C	A	C	G	C	T	T	G
S9	T	A	C	T	C	C	G	G	T	T	T	A	T	A	C	A	C	T	A	G	C	T	T	G
S10	T	A	C	T	C	C	A	G	A	T	C	G	C	A	A	C	T	A	A	G	C	T	T	G
S11	T	A	C	T	C	G	G	G	T	C	C	G	T	A	C	C	C	A	A	G	A	C	T	G
S12	T	A	C	T	G	G	G	G	T	T	T	A	T	A	A	C	T	A	A	A	A	T	T	G
S13	T	A	C	T	C	C	A	G	A	C	T	G	C	A	A	C	T	T	C	G	C	T	T	G
S14	T	A	T	T	C	C	G	G	A	C	T	G	T	A	A	C	C	T	A	G	C	C	T	G
S15	C	A	C	T	C	C	A	G	T	T	T	A	C	A	A	C	C	T	C	G	C	C	T	G
S16	C	A	T	T	G	C	A	G	A	T	T	G	C	A	C	C	T	A	C	A	A	C	T	G
S17	C	A	C	T	G	G	A	A	A	T	C	G	T	A	C	C	C	A	A	A	A	T	T	G
S18	C	A	C	T	G	G	A	A	T	T	T	A	T	A	C	A	C	A	A	G	C	C	T	G
S19	T	A	C	T	C	C	A	G	A	C	T	G	C	A	A	C	T	T	C	G	C	T	T	G
S20	T	A	T	T	G	C	A	G	A	C	C	G	C	A	A	A	T	T	A	G	A	T	T	G

Molecular barcode data from Senegal—Clonal Parasites

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
S1	T	A	C	T	C	C	G	A	T	T	C	A	C	A	A	C	T	T	C	G	A	C	T	G
S2	T	A	C	T	C	C	A	A	A	T	T	G	C	C	C	A	C	A	A	G	A	T	T	G
S3	T	A	T	T	C	C	A	G	A	T	C	G	C	A	A	C	C	A	A	G	A	T	T	G
S4	T	A	T	T	C	G	A	G	A	C	T	G	T	A	C	C	C	T	C	G	C	C	T	G
S5	T	G	C	T	C	C	A	G	A	T	T	A	T	A	A	C	T	T	C	G	A	C	T	G
S6	T	A	T	T	C	C	G	G	T	T	T	G	C	A	C	C	T	A	A	A	A	T	T	G
S7	T	A	C	T	G	C	A	G	A	T	T	G	T	C	A	C	T	T	A	A	A	C	T	G
S8	T	A	C	T	C	C	G	A	A	C	T	G	T	C	C	C	C	A	C	G	C	T	T	G
S9	T	A	C	T	C	C	G	G	T	T	T	A	T	A	C	A	C	T	A	G	C	T	T	G
S10	T	A	C	T	C	C	A	G	A	T	C	G	C	A	A	C	T	A	A	G	C	T	T	G
S11	T	A	C	T	C	G	G	G	T	C	C	G	T	A	C	C	C	A	A	G	A	C	T	G
S12	T	A	C	T	G	G	G	G	T	T	T	A	T	A	A	C	T	A	A	A	A	T	T	G
S13	T	A	C	T	C	C	A	G	A	C	T	G	C	A	A	C	T	T	C	G	C	T	T	G
S14	T	A	T	T	C	C	G	G	A	C	T	G	T	A	A	C	C	T	A	G	C	C	T	G
S15	C	A	C	T	C	C	A	G	T	T	T	A	C	A	A	C	C	T	C	G	C	C	T	G
S16	C	A	T	T	G	C	A	G	A	T	T	G	C	A	C	C	T	A	C	A	A	C	T	G
S17	C	A	C	T	G	G	A	A	A	T	C	G	T	A	C	C	C	A	A	A	A	T	T	G
S18	C	A	C	T	G	G	A	A	T	T	T	A	T	A	C	A	C	A	A	G	C	C	T	G
S19	T	A	C	T	C	C	A	G	A	C	T	G	C	A	A	C	T	T	C	G	C	T	T	G
S20	T	A	T	T	G	C	A	G	A	C	C	G	C	A	A	A	T	T	A	G	A	T	T	G

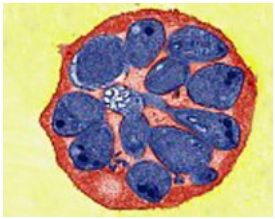
Genetic signals related to changing transmission

Range of Transmission Intensity

EIR	10	1	0.1	0.01	0.001	0
Prevalence	60%–30%	30%–10%	10%–1%	<2%	<0.3%	0

COI

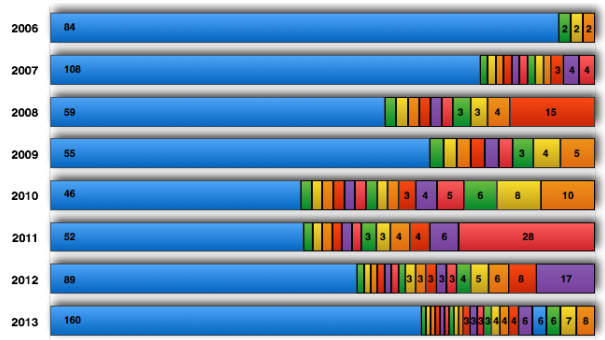
High	Low	1

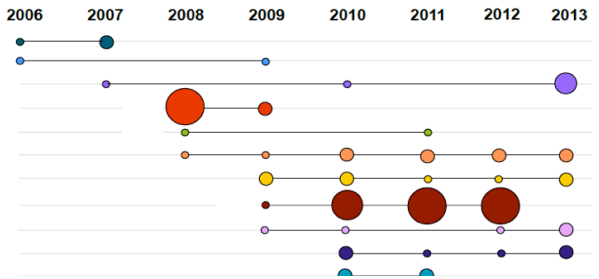


Genotyping

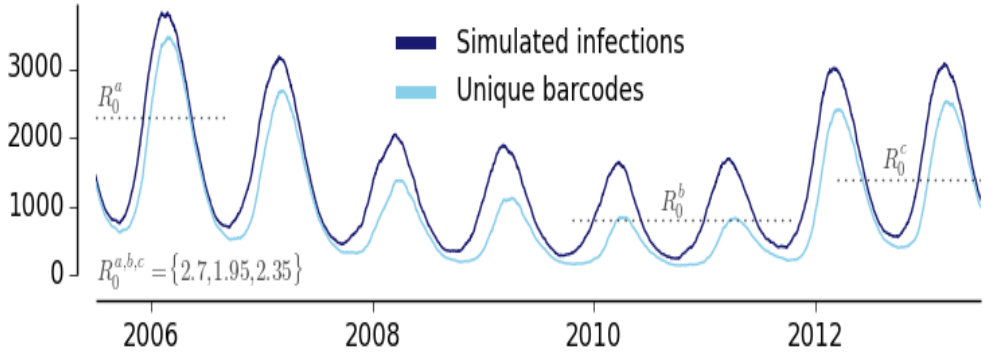
Polygenomic	Monogenomic	Clonal

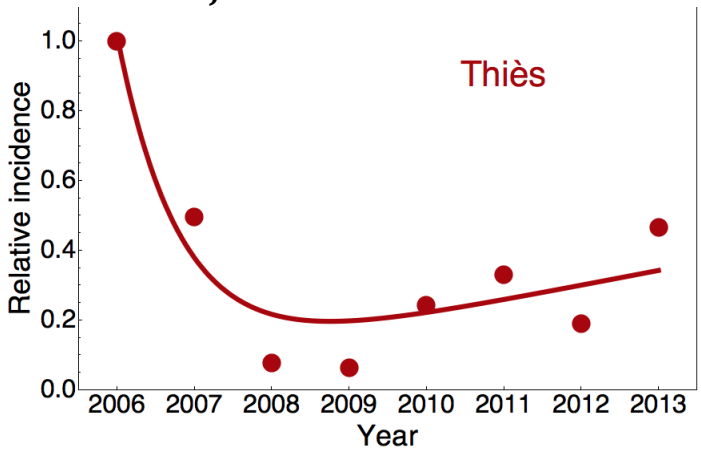
Genetics detects transmission interruption in Senegal



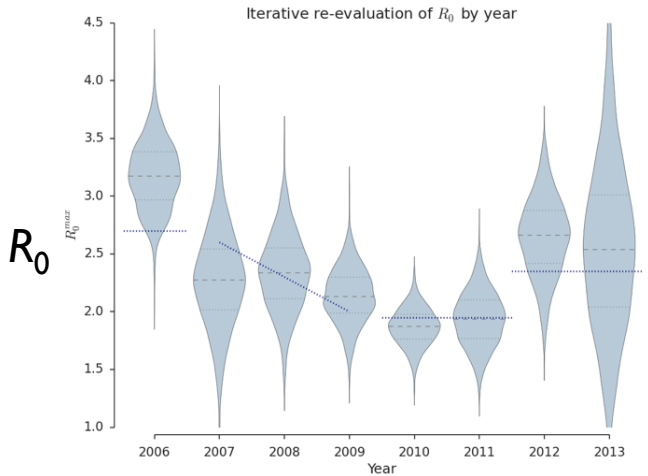


Daniels, PNAS 2015

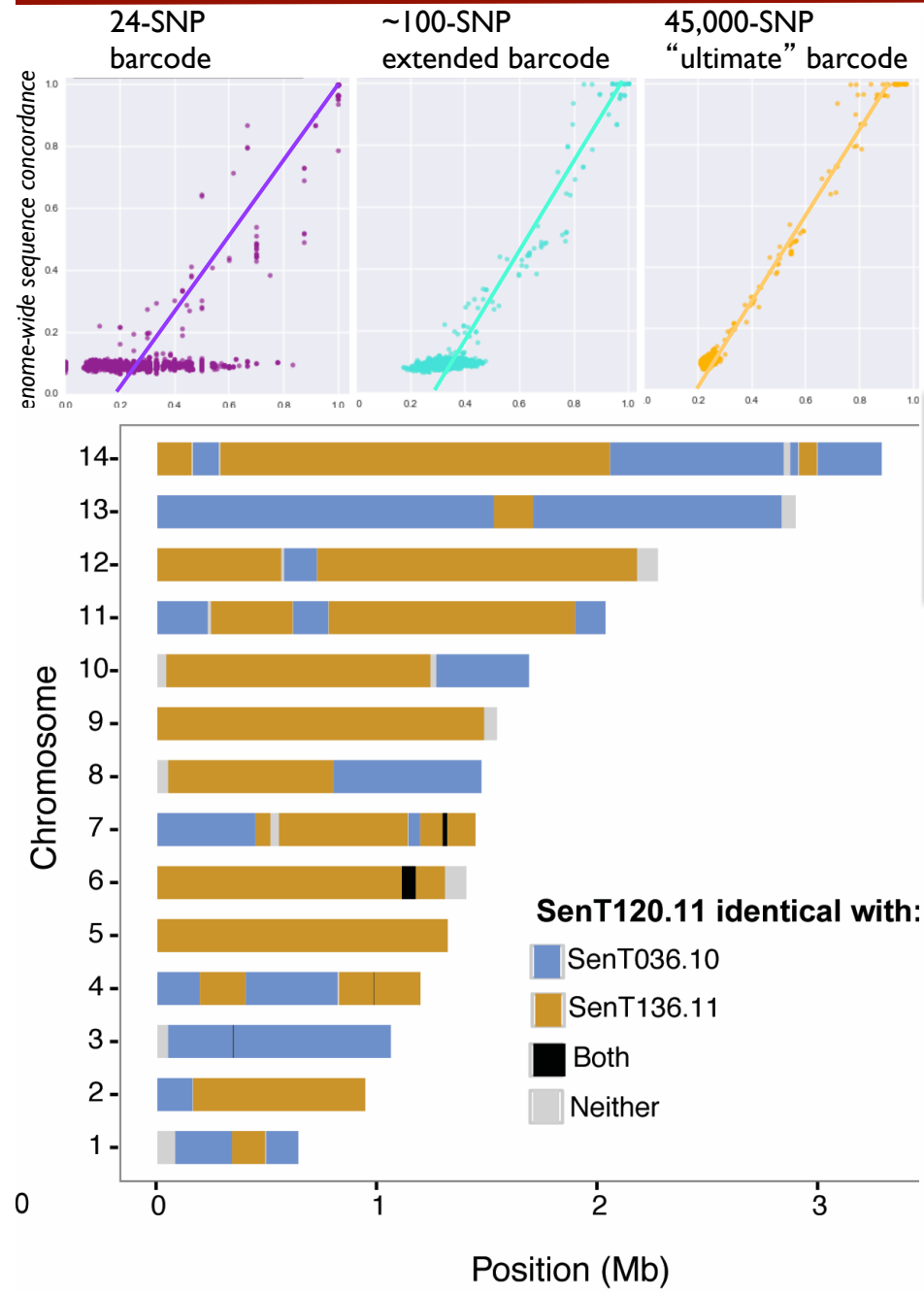


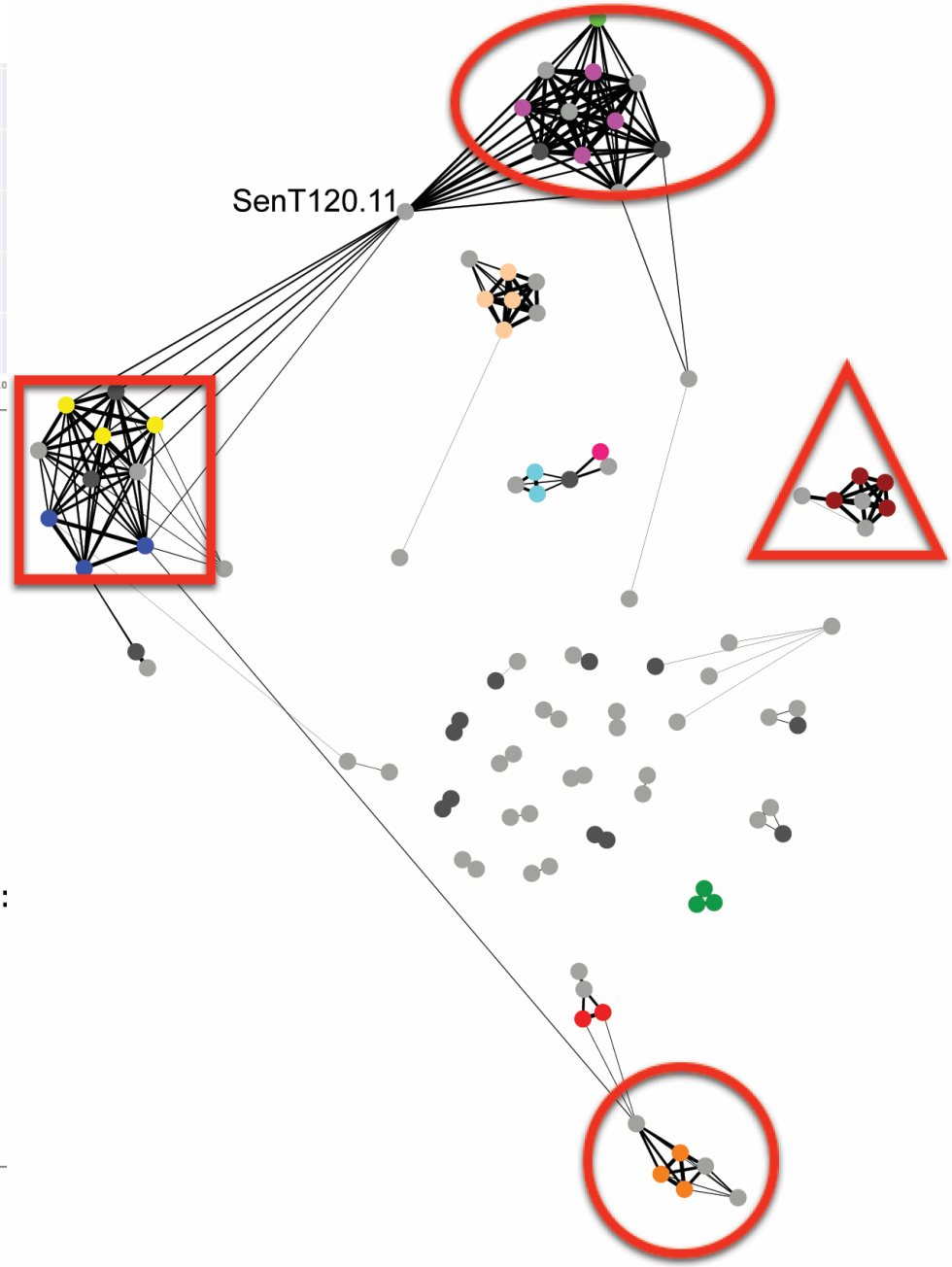


- *Temporal sampling from 2006*
- *Clonal parasites identified*
- *Parasite types persist across seasons*
- *Incidence decline, then increase*
- *Modeling genetic signals detects transmission decline and rebound*

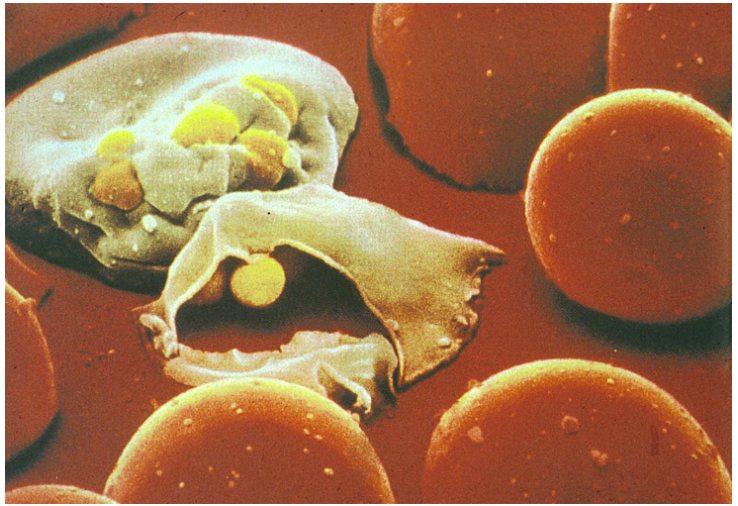


Genetics detects parasite relatedness





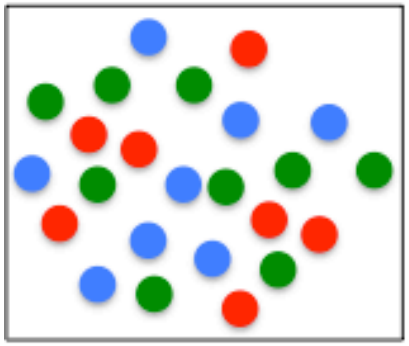
Concept 4: Intervention Impact Assessment



Intervention Impact Assessment

- Genotyping to interpret and evaluate implementation strategies
- Interventions for reduction or interruption of malaria transmission
- First genetic signature—reduction of transmission: reduction of COI, increase in relatedness, and eventually clonality
- Second genetic signature—consequences of selection based upon intervention type
- Take Home Lesson: Reduction in parasite population structure with specific interventions may select for parasite populations that are less responsive to that intervention.
- Example: transmission reduction with drugs may selective for drug resistant parasites.

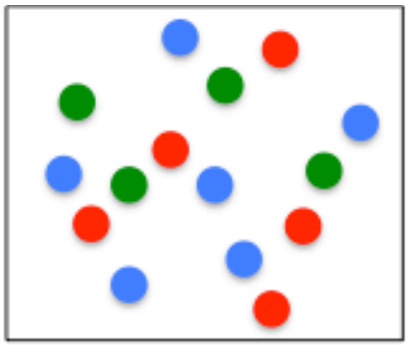
Intervention Impact—distinct genetic patterns expected



Apply Intervention

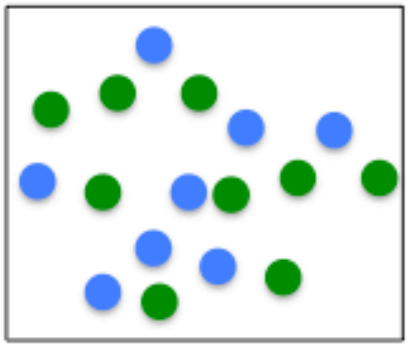
Reduce all parasite types

Bednet use may non-specifically reduce all parasite types



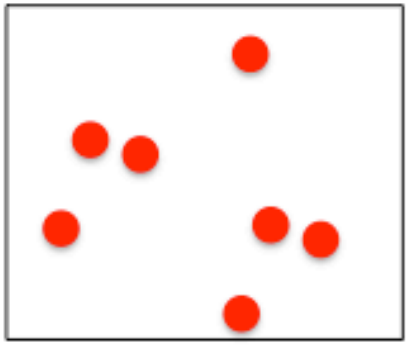
Reduce select parasite types

Vaccine use may reduce specific parasite types

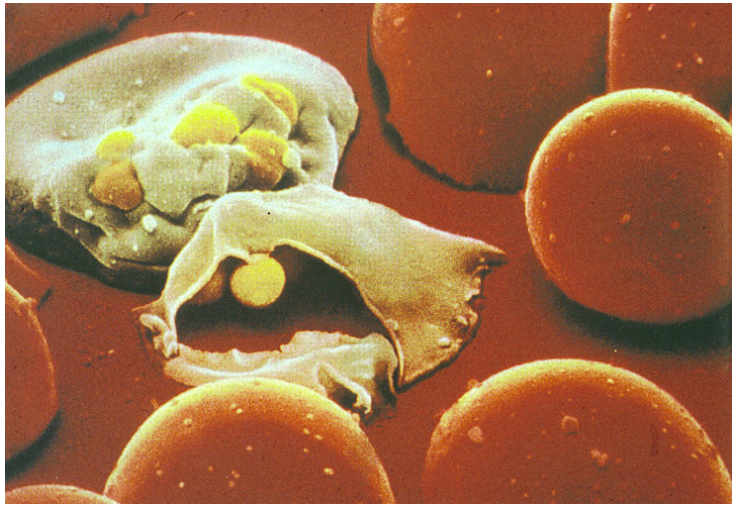


Reduce all except select parasite types

Drug pressure may select for resistant parasites



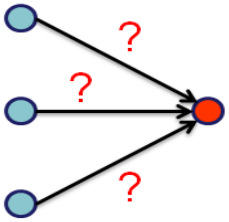
Concept 5: Tracking Parasites



Tracking Parasites

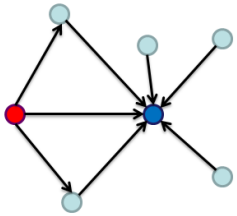
- Genotyping or fingerprinting to track parasites.
- Identify connectivity between infections—where do parasites come from (source) and where do they go (sink).
- Identify reservoirs of new infection and asymptomatic individuals contributing to incident infections.
- Apply to outbreak investigation in pre-elimination settings to prevent reintroduction of infection.

Transmission mapping critical for elimination



How many infections are being imported?

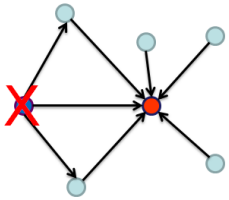
When? From where? By who?



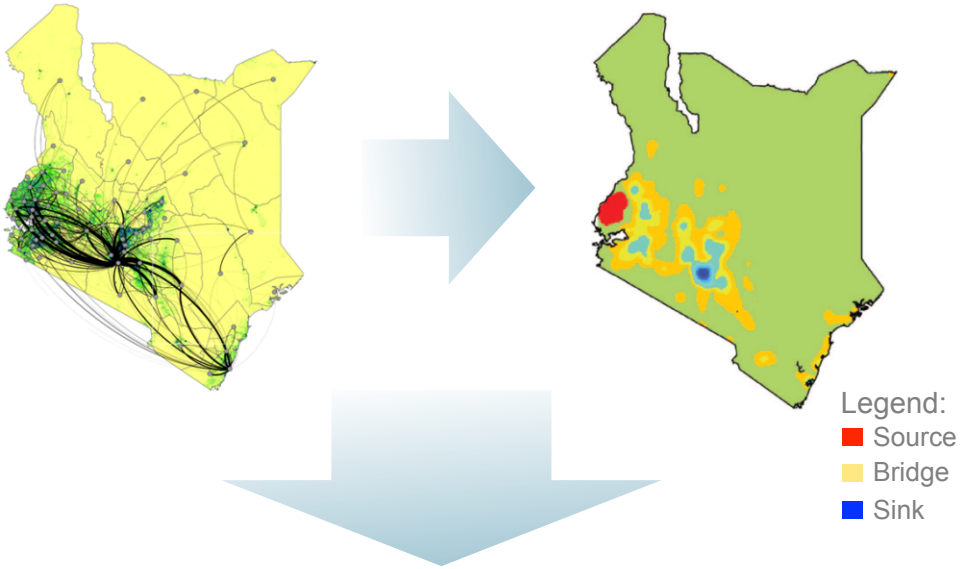
Where are the likely sources of imported infections?

Where are they going?

Who's bringing them?

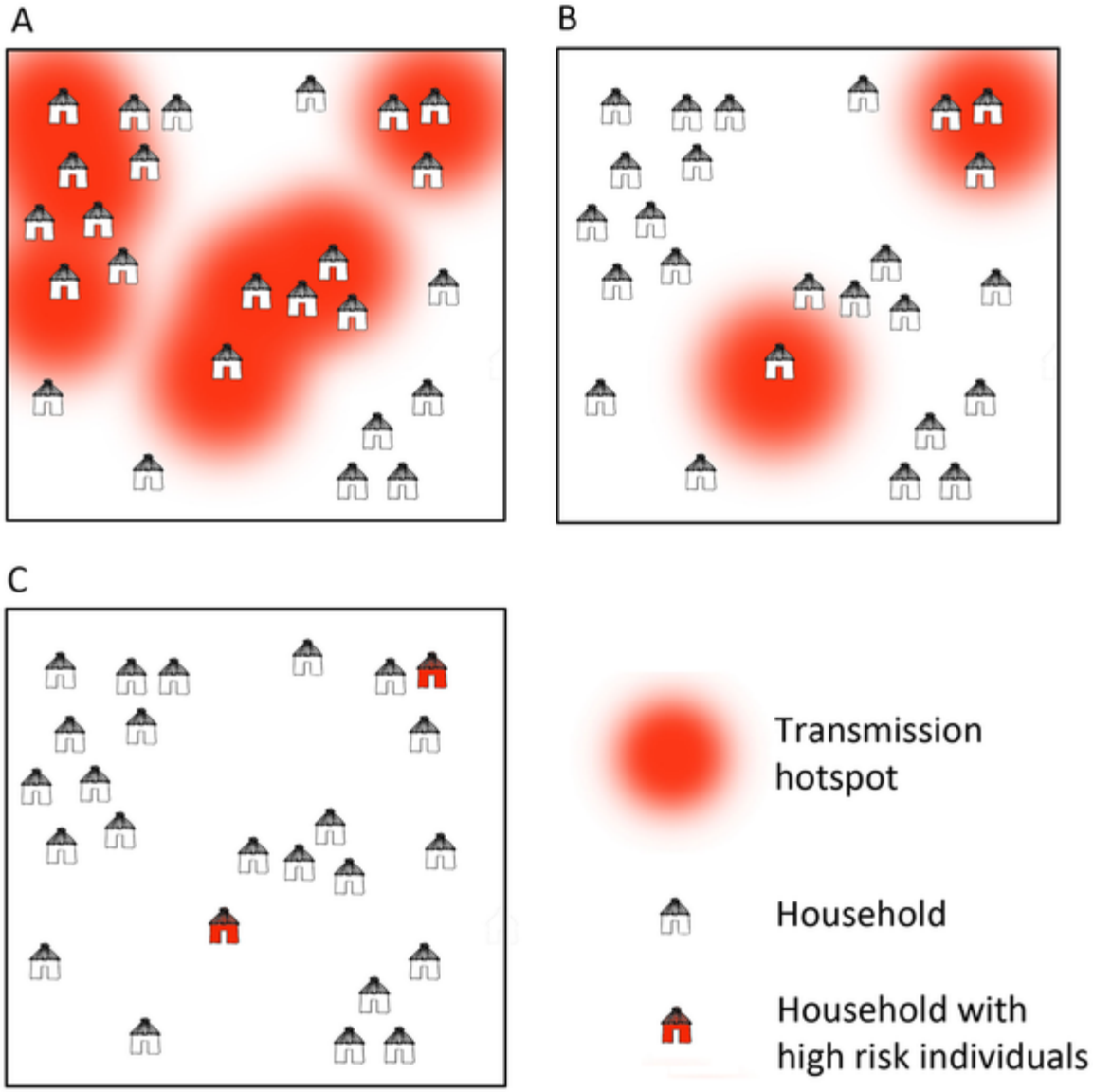


How can we make efficient use of connectivity data to target surveillance and control, plan an attack strategy?



Regional Stratification that takes into account connectivity between (asymptomatic) reservoirs of infection

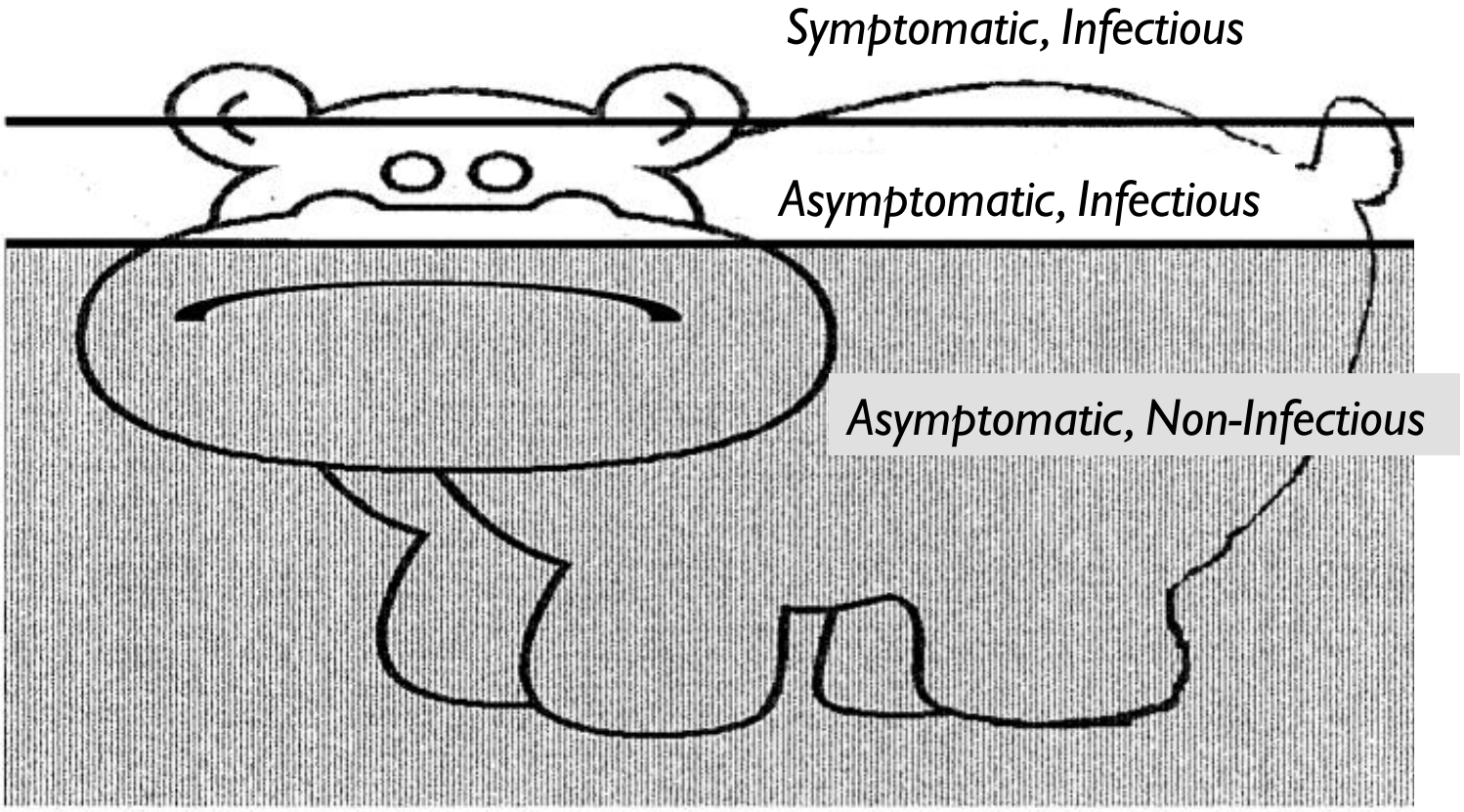
Genetics help map relationships between transmission foci



- Identify sources of malaria transmission
- Distinguish between local and imported infections
- Target transmission-based interventions to where transmission is occurring

Sturrock, PLoSMedicine 2013

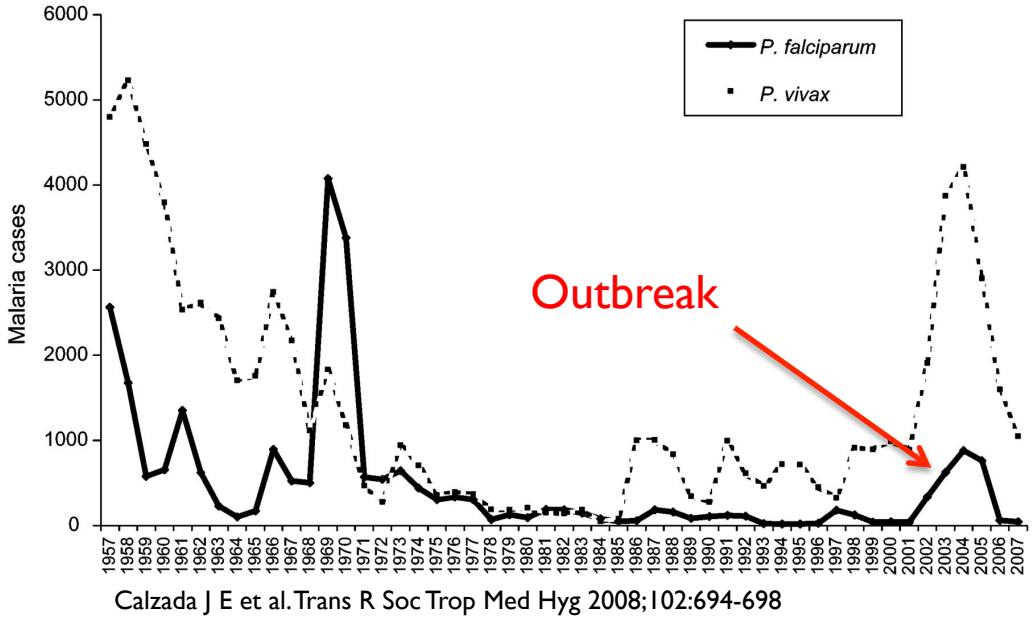
Tracking Parasites—Asymptomatic Reservoir



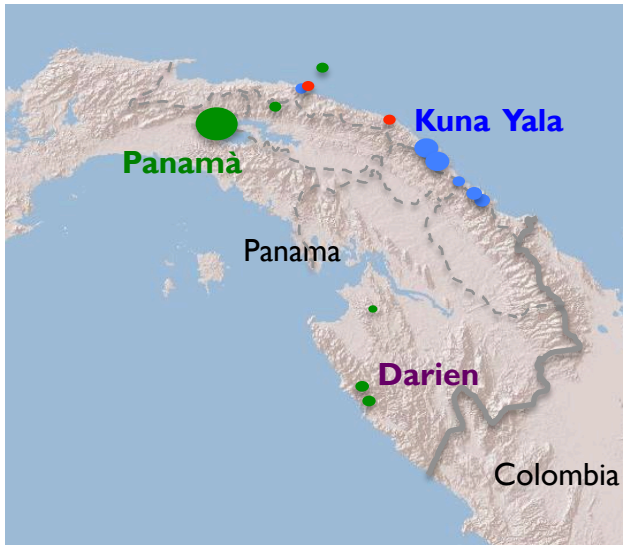
- *Detect who harbors infection*
- *Identify who contributes to onward infection (infectious)*
- *Strategize interventions to reduce onward transmission*

Breman, JASTMH 2001

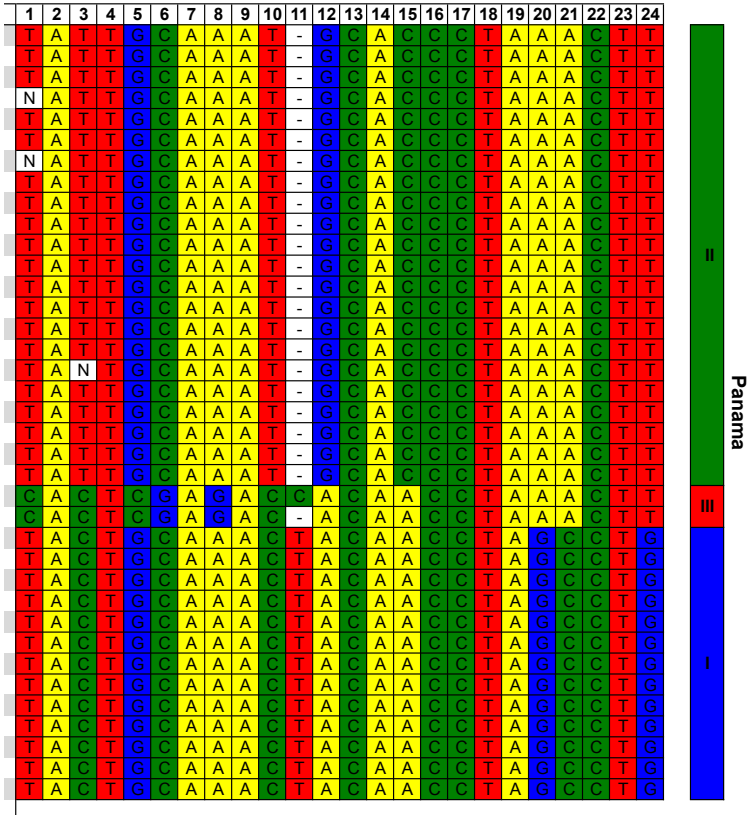
Outbreak investigation—example of Panama



Calzada J E et al. Trans R Soc Trop Med Hyg 2008;102:694-698

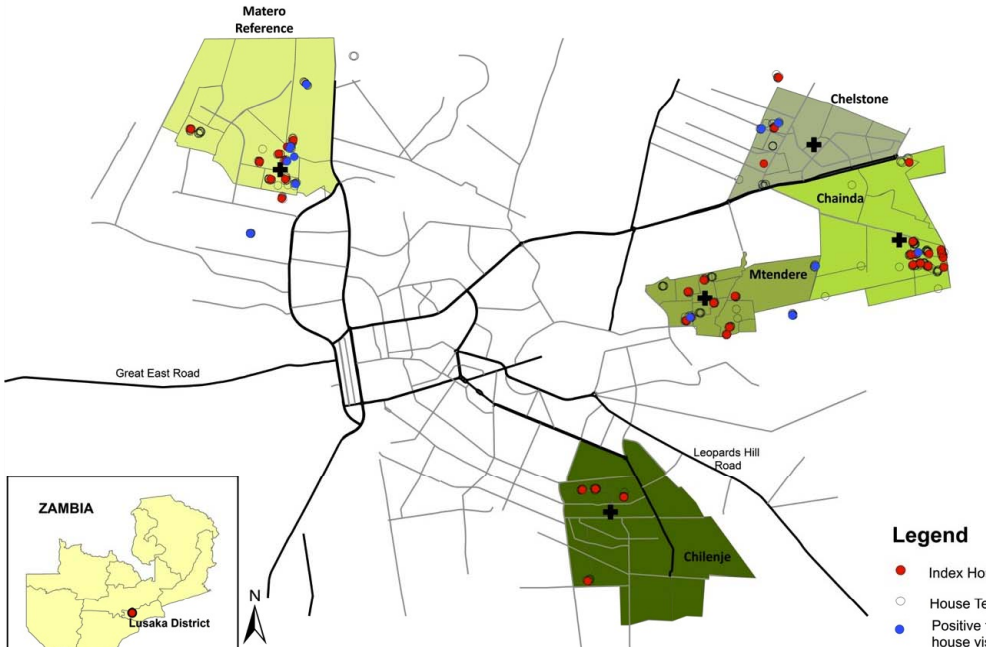


Outbreak dominated by two clonal parasite populations detected by molecular barcode



Obaldia, JID 2014

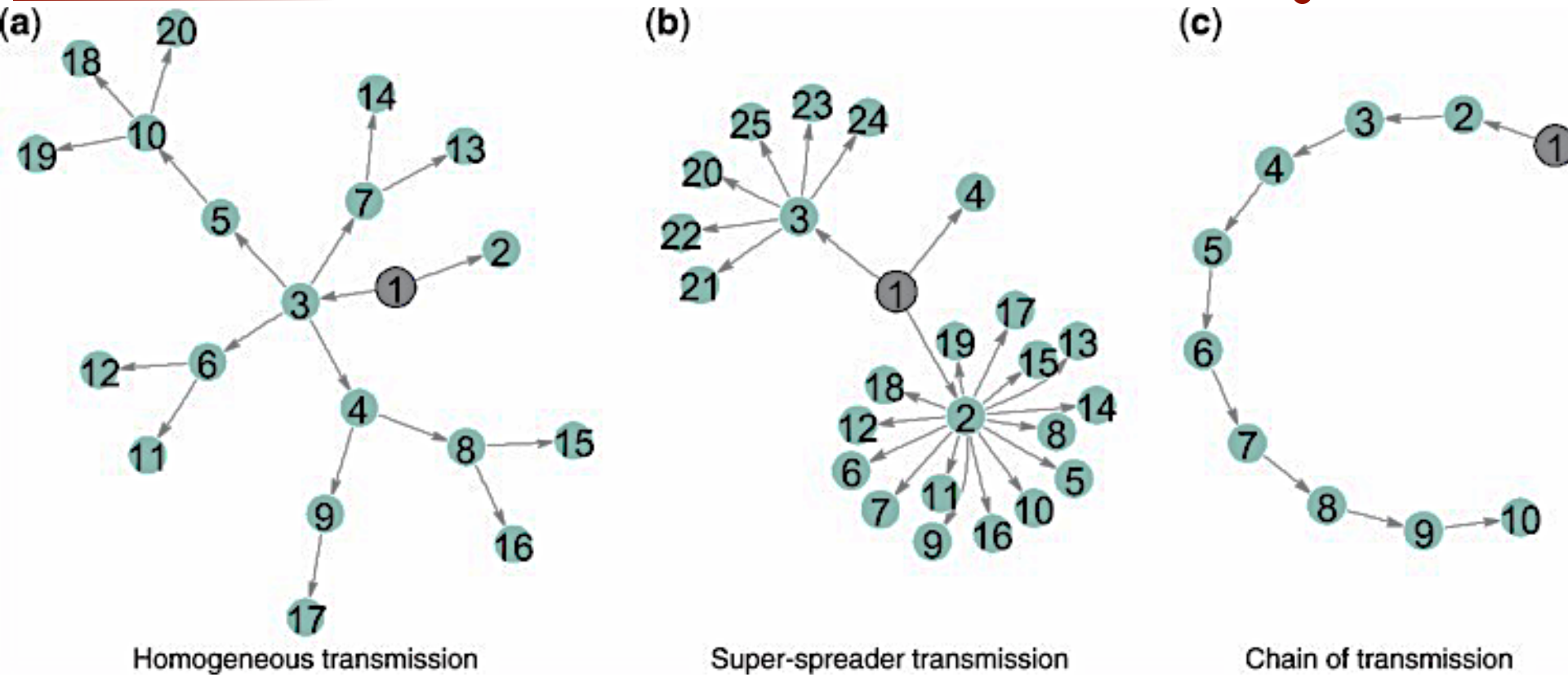
Reactive case detection—example of Zambia



T	A	T	C	C	C	G	G	A	T	C	G	C	A	C	A	C	-	A	G	A	T	T	G
T	A	T	C	C	C	G	G	A	T	C	G	C	A	C	A	C	-	A	-	A	T	T	G

- Lusaka District, Zambia
- Reactive case detection: passive case detection + active follow up
- Highly related parasites within households
- Foci investigation: space and time: transmission trees

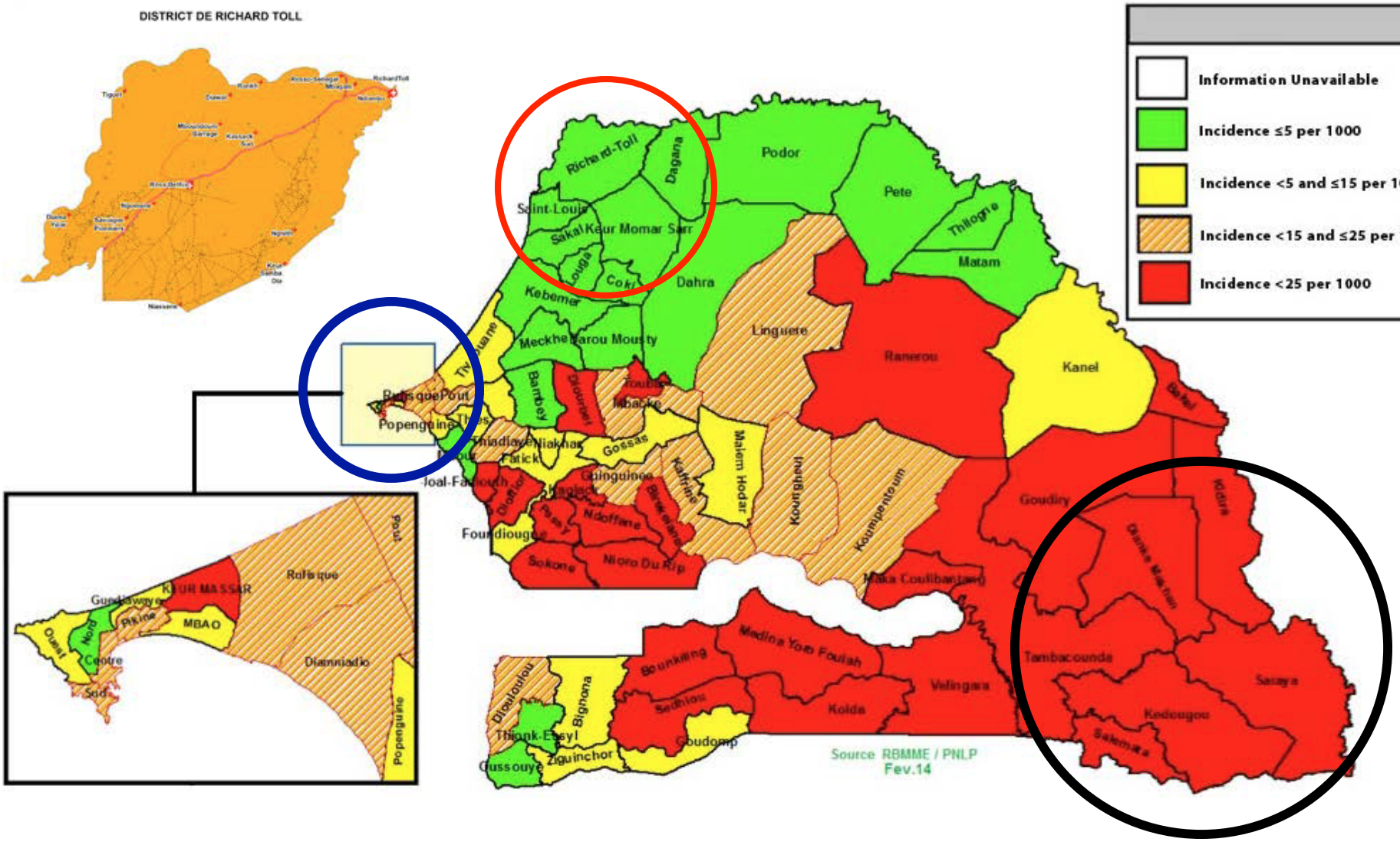
Genomics to inform transmission networks



- *Phylogenetic tree shapes resolve disease transmission patterns*
- *Development of tools to infer these relationships and understand the spatial-temporal relationships of infections*
- *Leverage knowledge from other infections (e.g., polio) to inform strategies for malaria*

Colijn & Gardy, Evolution, Medicine and Public Health, 2014

Senegal: opportunities for "sink", "source" and "bridge" testing



Senegal NMCP

Genetic evidence of parasite or population migration in Senegal

Identical barcode in Thiès and Pikine, 2013

T A T T G C G G T T T A T A C C C T A G C C T G

T A T T G C G G T T T A T A C C C T A G C C T G

Identical barcode in Thiès and Pikine, 2012 - 2013

C A T T C G A G T T C G T A C C C A C G C C T G

C A T T C G A G T T C G T A C C C A C G C C T G

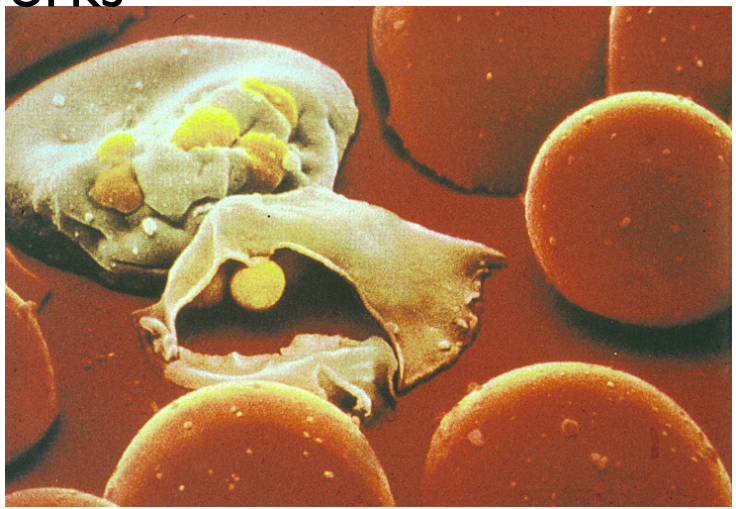
Genetics detects multiple species in Senegal

- *Detect infections among RDT negative samples*
- *Multiple species detected, several mixed infections*
- *Potential impact on diagnosis of malaria by RDT*
- *Evaluate species dynamics as interventions directed toward Plasmodium falciparum are applied*

	Sample ID	*P. ovale curtisi*	*P. ovale wallikeri*	*P. malariae*	*P. falciparum*
RDT (+)	61109				
	61113				
	62068				
	62256				
RDT (-)	61004				
	61017				
	61043				
	61203				
	62003				
	62036				
	62067				

Use genomics to infer field and operational questions

- Infer changing transmission dynamics
  - Genomic thermometer
  - Predict transmission declines and rebounds
- Assess intervention impact
  - Mass drug administration
  - Vaccines
- Track parasites in space and time
  - Outbreak Investigation
  - Sink-source
  - Asymptomatic reservoir
  - Transmission networks



Acknowledgements