

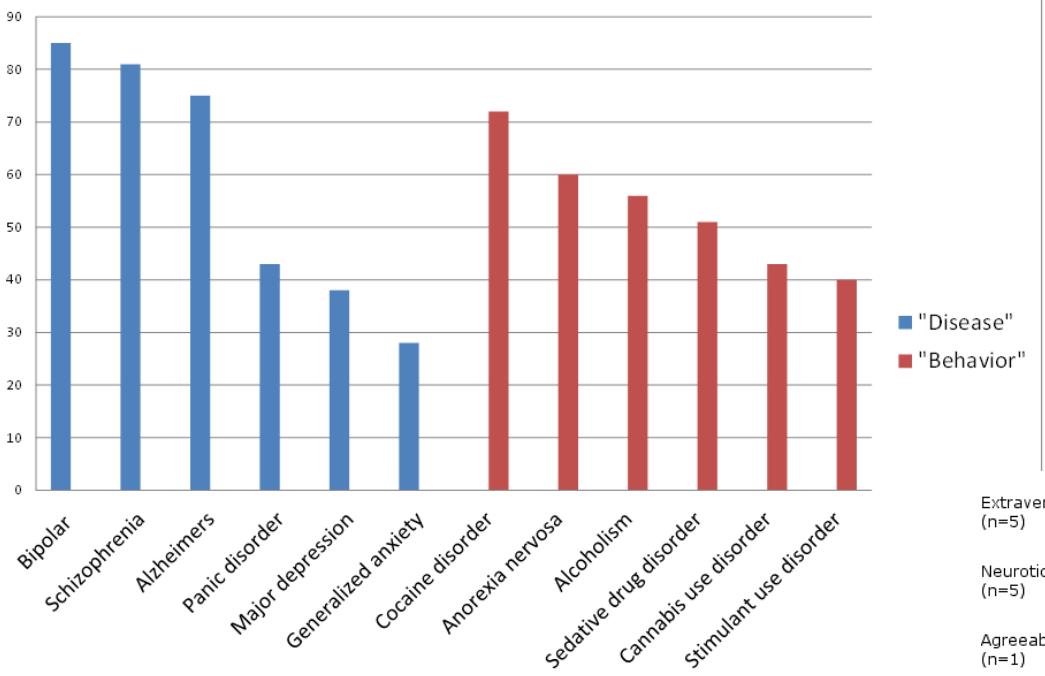


# Human Population Genomics

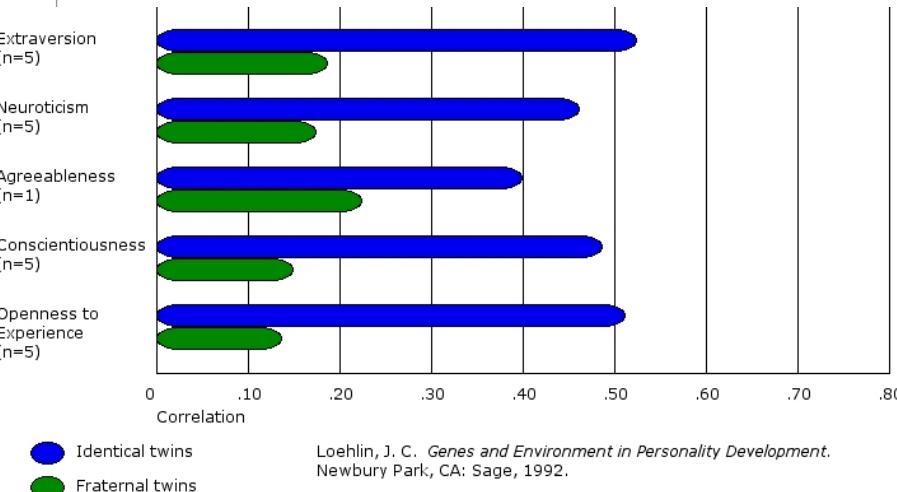
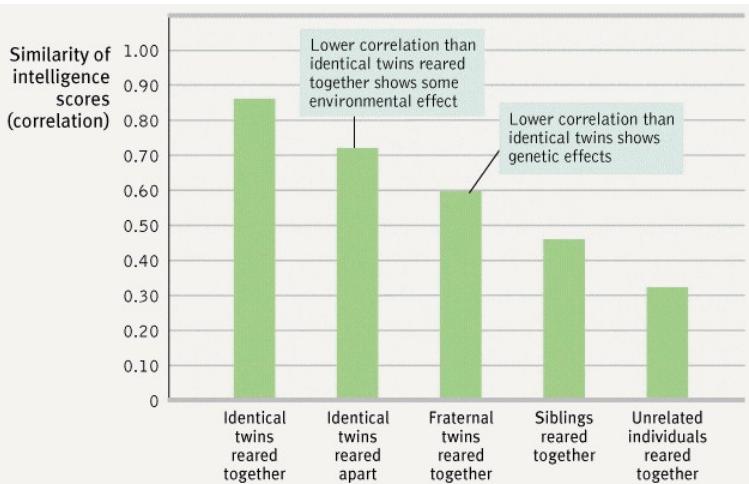
# Heritability & Environment



## Heritability of Disorders in Twin Studies



Bienvenu OJ, Davydow DS, &  
Kendler KS (2011).  
*Psychological medicine*,  
41 (1), 33-40 PMID:



Loehlin, J. C. *Genes and Environment in Personality Development*. Newbury Park, CA: Sage, 1992.

# Heritability & Environment



Disease	# of Loci	Heritability Explained	Heritability Estimated	Measure of Heritability
Age related macular degeneration	5	50%	46-71%	Sibling recurrent risk
Crohn's Disease	32	20%	50-60%	Genetic risk (liability)
Systemic Lupus Erythematosus	6	15%	44-66%	Sibling recurrent risk
Type 2 diabetes	18	6%	26%	"
HDL Cholesterol	7	5.2%		"
Height	40	5%	81%	Phenotypic Variance
Fasting glucose	4	1.5%		"

lio, Teri A., et al. "Finding the missing heritability of complex diseases." Nature 461.7265 (2009): 747-753.

# Heritability & Environment

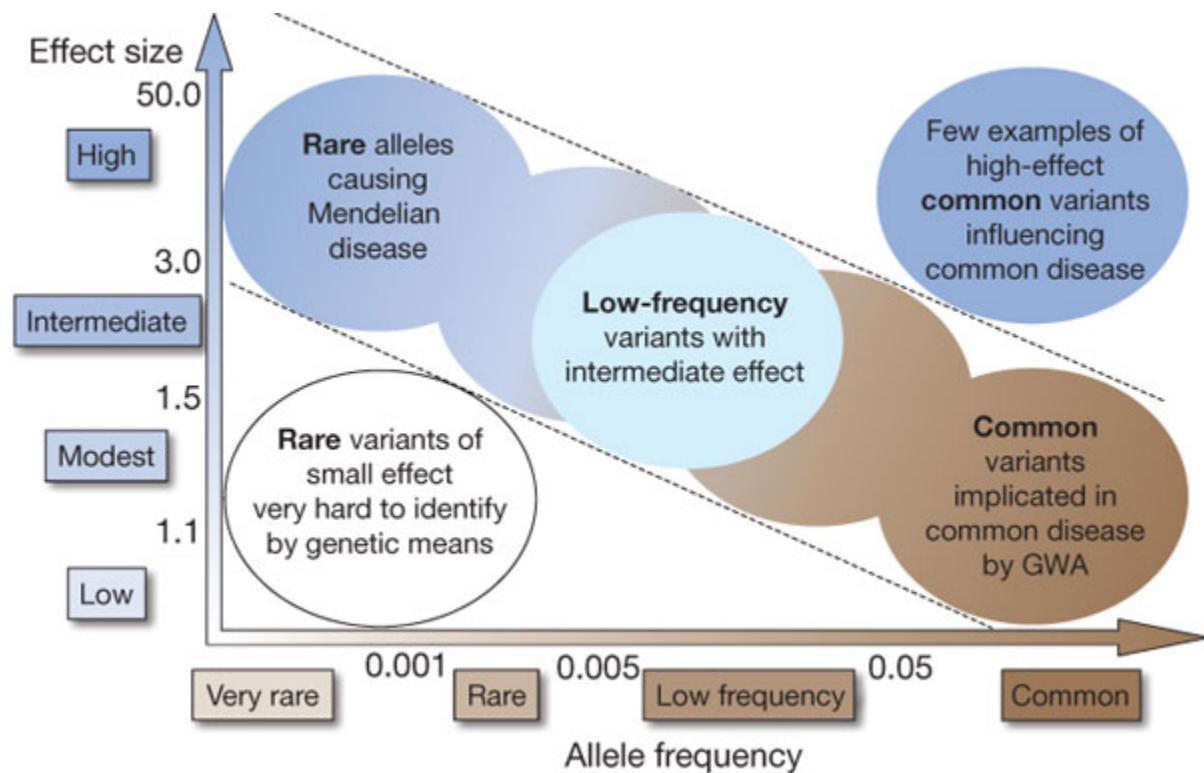


Trait/Disease	Estimated heritability
Alcoholism	50-60%
Alzheimers	58-79%
Asthma	30%
Bipolar Disorder	70%
Depression	50%
Hair Curliness	85-95%
Lung Cancer	8%
Height	81%
Obesity	70%
Longevity	26%
Sexual Orientation	60%
Schizophrenia	81%
Type 1 diabetes	88%
Type 2 diabetes	26%

# Heritability & Environment



Feasibility of identifying genetic variants by risk allele frequency and strength of genetic effect (odds ratio).



TA Manolio *et al.* *Nature* **461**, 747-753 (2009) doi:10.1038/nature08494

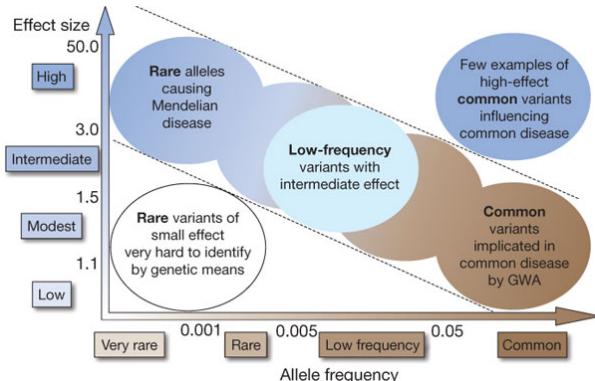
nature

# Where is the missing heritability?



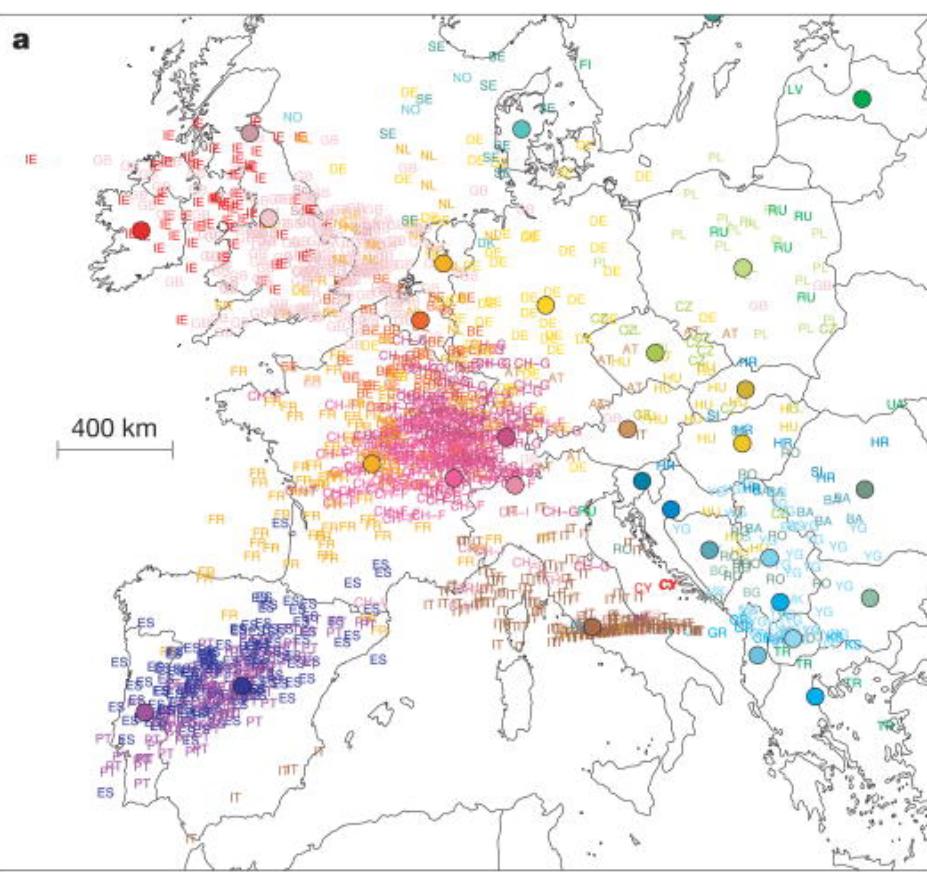
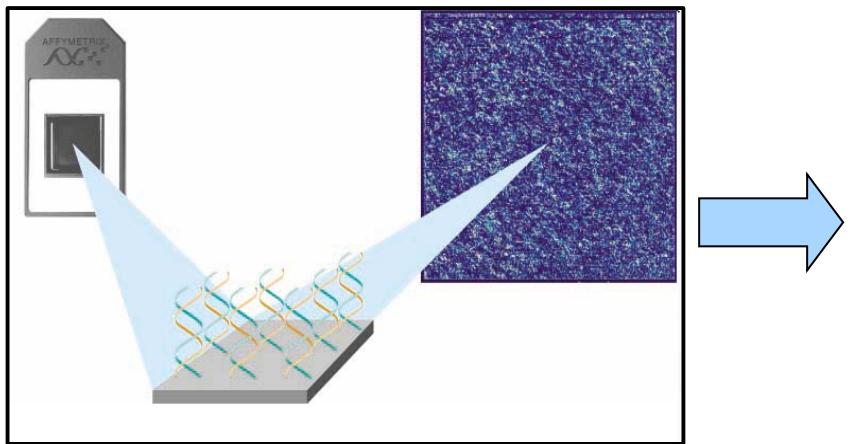
## Some plausible explanations

- **Rare variants** not captured in genotyping microarrays
- **Many variants** of small effect
- **Structural variants** not captured in short read sequencing
- **Epistatic effects:** non-linear gene-gene interactions
- **???**



TA Manolio *et al. Nature* **461**, 747-753 (2009) doi:  
10.1038/nature08494  
**nature**

# Global Ancestry Inference

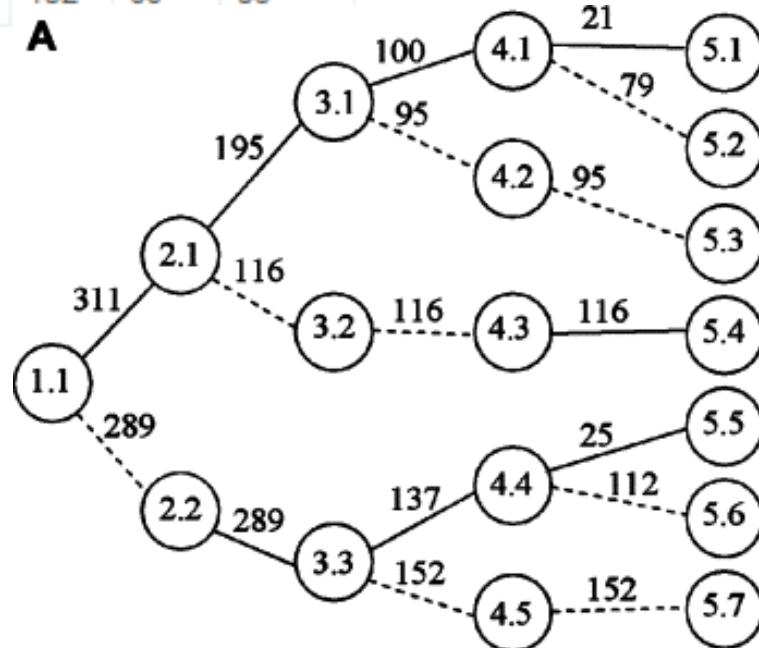


# Modeling population haplotypes – VLMC



HAPLOTYPE	COUNT		
	Total	Case	Control
0000	21	12	9
0001	79	43	36
0011	95	43	52
0110	116	59	57
1000	25	14	11
1001	112	60	52
1011	152	69	83

**A**

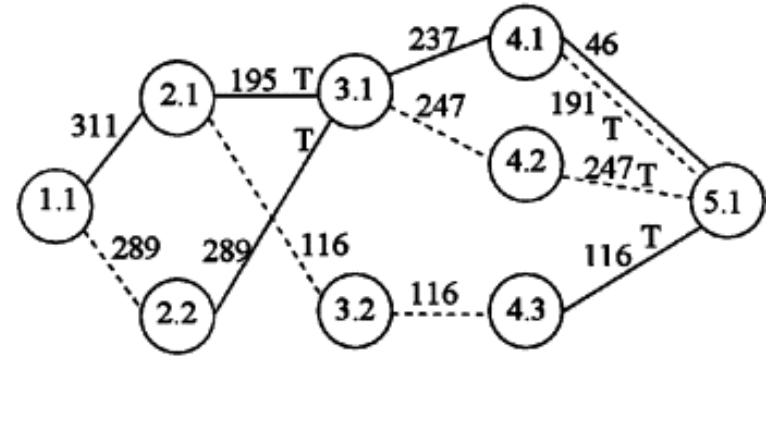


$$G_1, \dots, G_N; \quad G_i = g_{i1} \dots g_{in}; \quad g_{ij} \in \{0, 1, 2\}$$

$$G_i = H_{i1} + H_{i2}, \text{ where,}$$

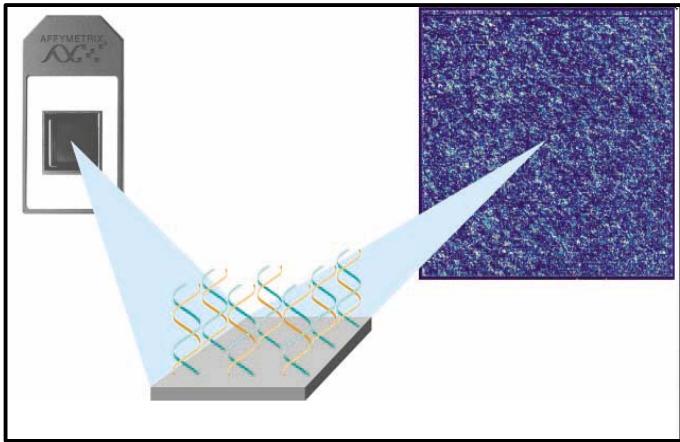
$$H_i = h_{ij1} \dots h_{ijn}; \quad h_{ijk} \in \{0, 1\}$$

**B**



[Browning, 2006](#)

# Phasing



## Haplotype Phasing

Haplotypes

ATCCGA  
AGACGC

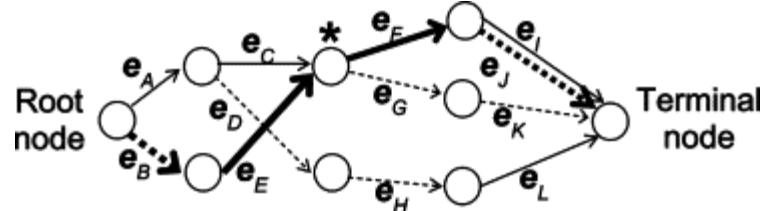
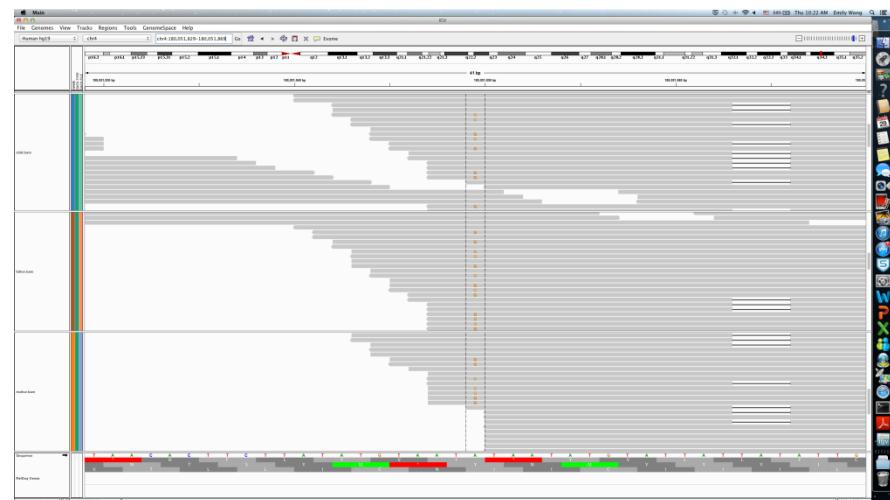
Genotype

A{T}{C}{C}  
G{A}{A}{A}

- High throughput cost effective sequencing technology gives genotypes and not haplotypes.

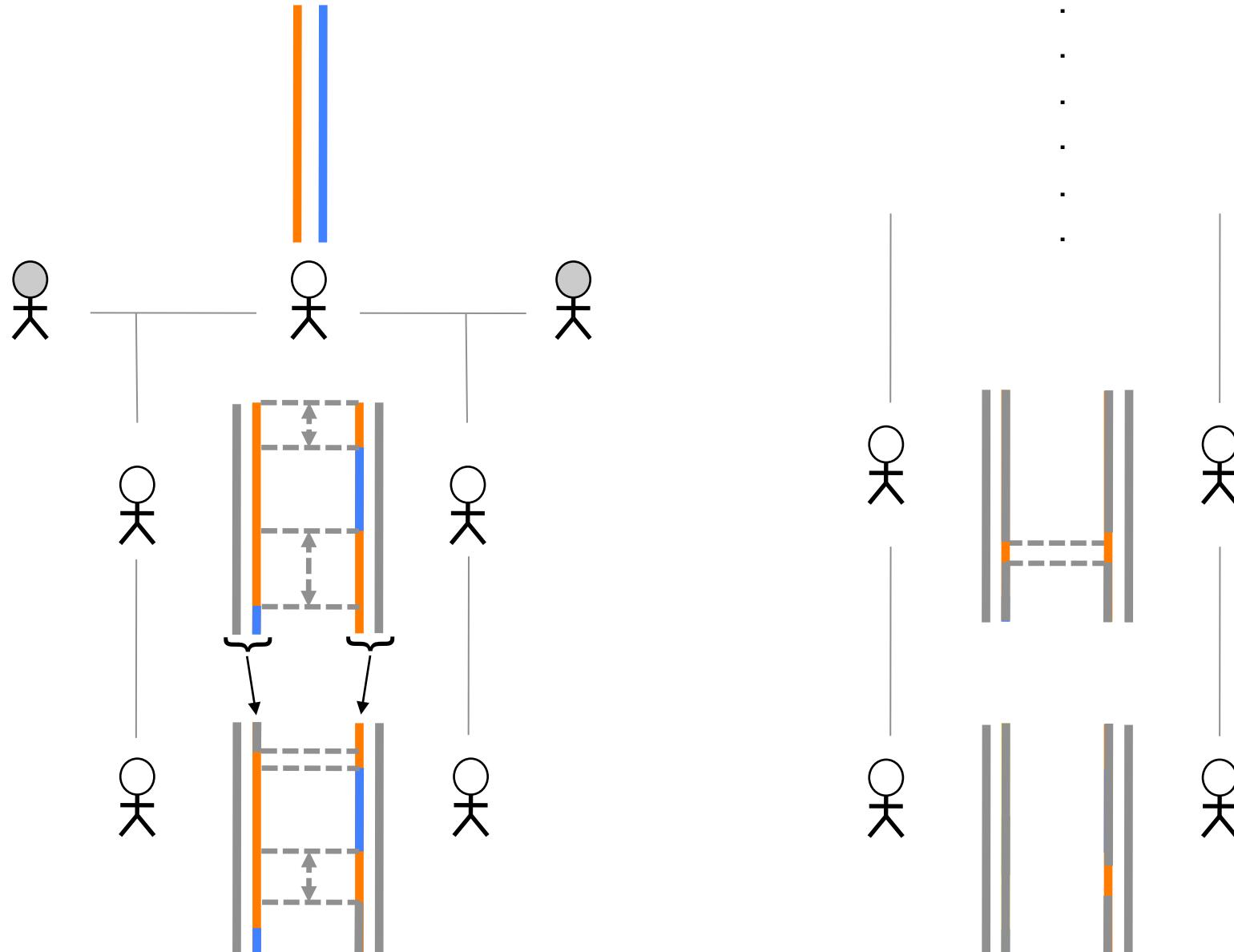
Possible phases:  
AGACGA  
AGCCGC

ATACGA  
ATCCGC ....



[Browning & Browning, 2007](#)

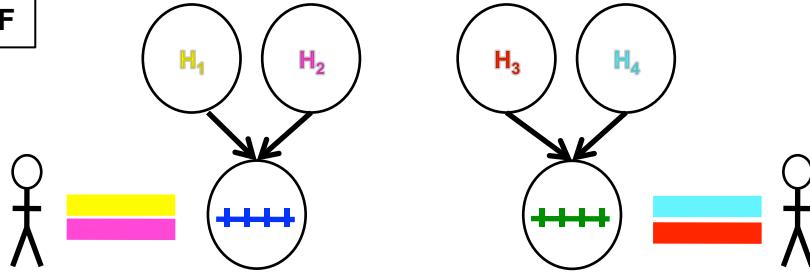
# Identity By Descent



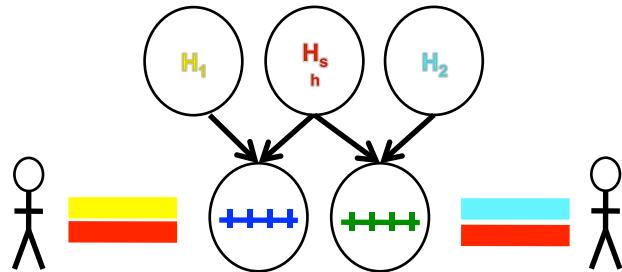
# IBD detection



IBD = F

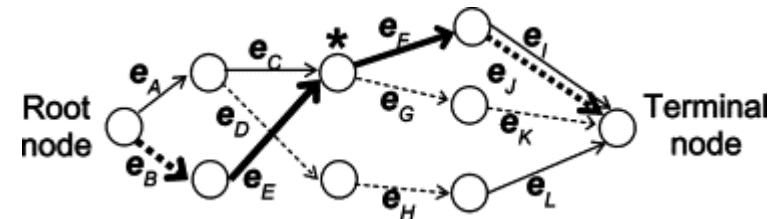
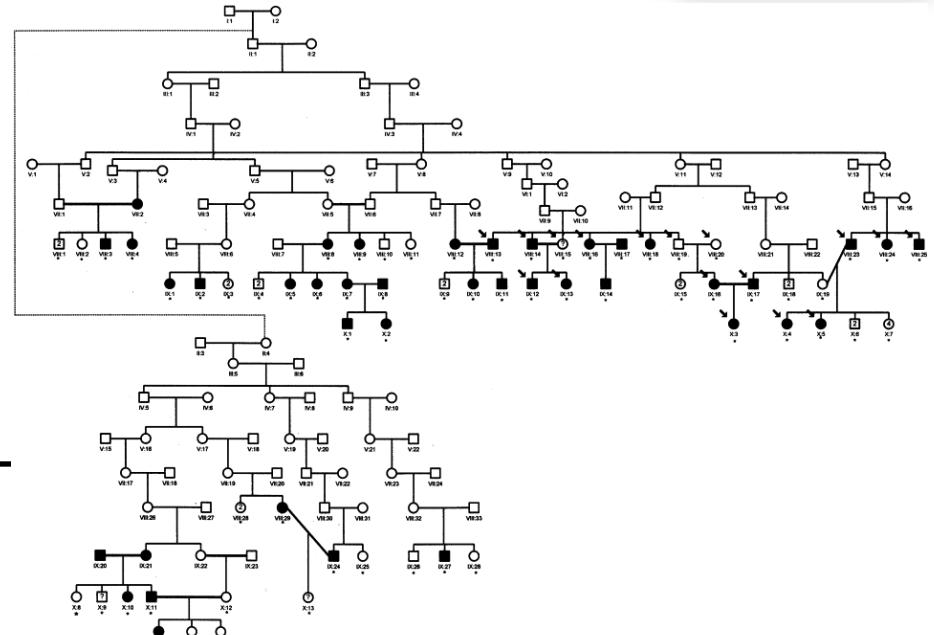


IBD = T



Parente

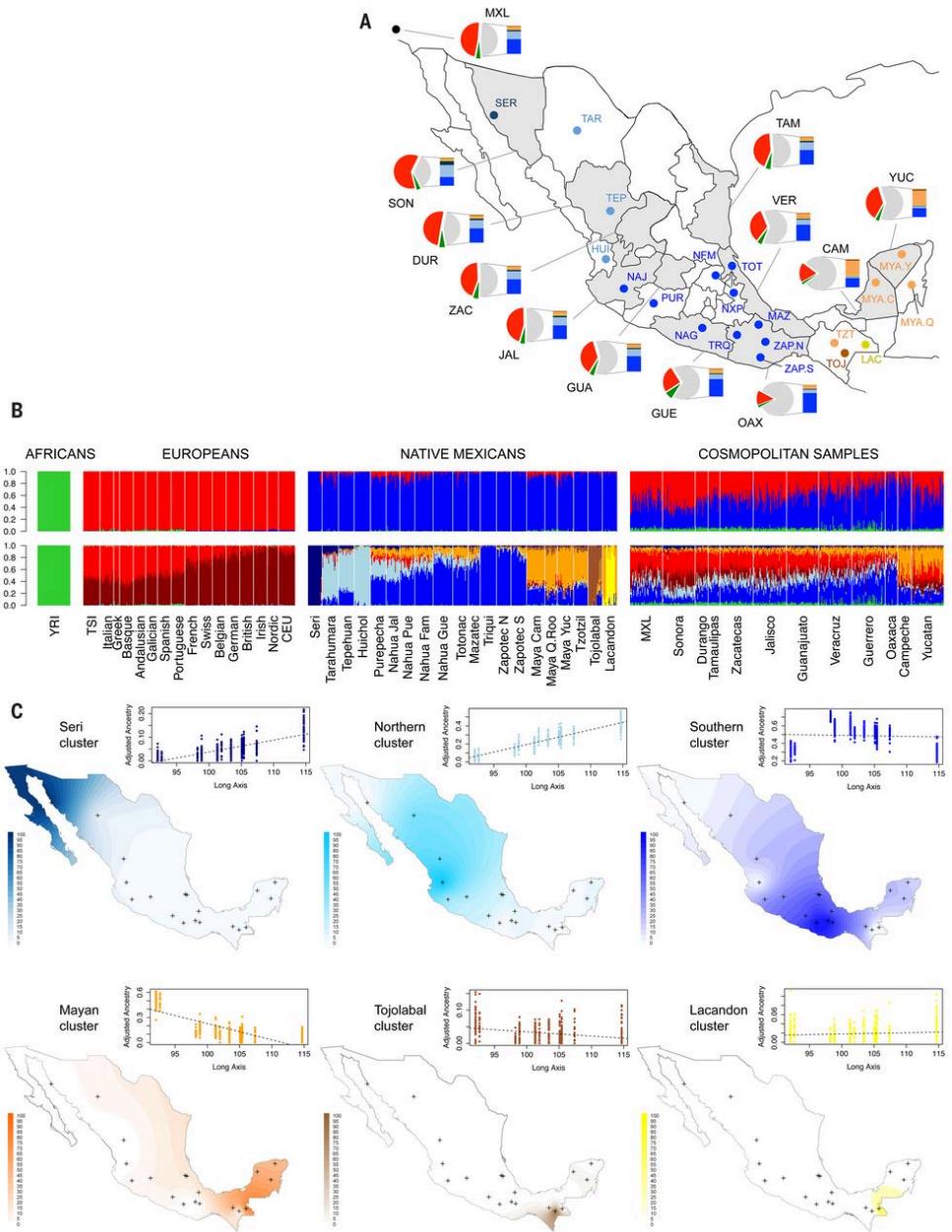
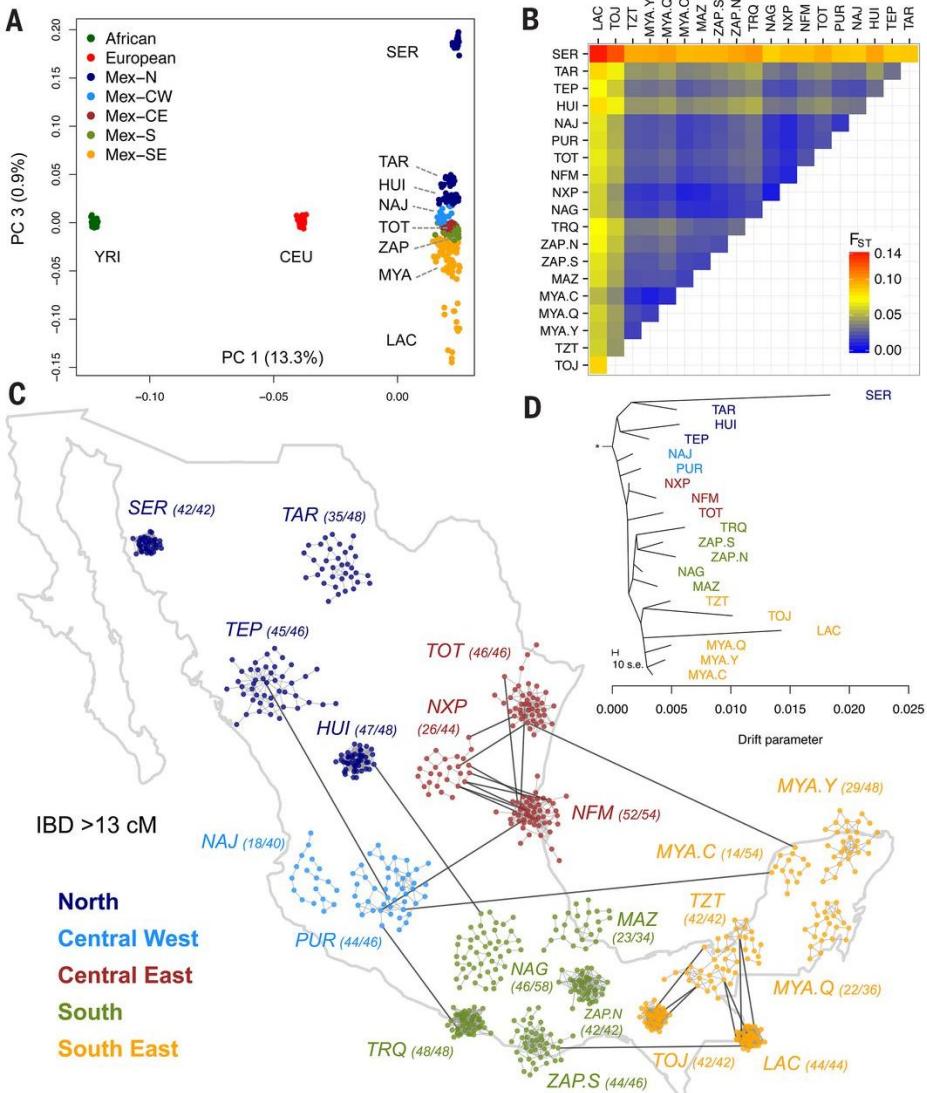
[Rodriguez et al. 2013](#)



FastIBD: sample haplotypes for each individual, check for IBD

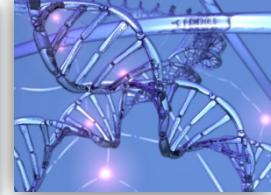
[Browning & Browning 2011](#)

# Mexican Ancestry

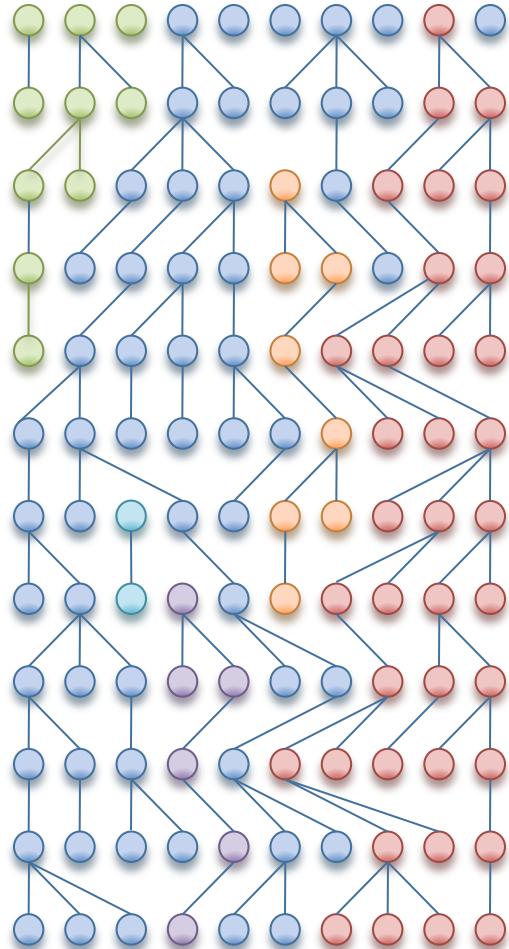


The genetics of Mexico recapitulates Native American substructure and affects biomedical traits, Moreno-Estrada et al. Science, 2014.

# Fixation, Positive & Negative Selection



Negative Selection



Neutral Drift



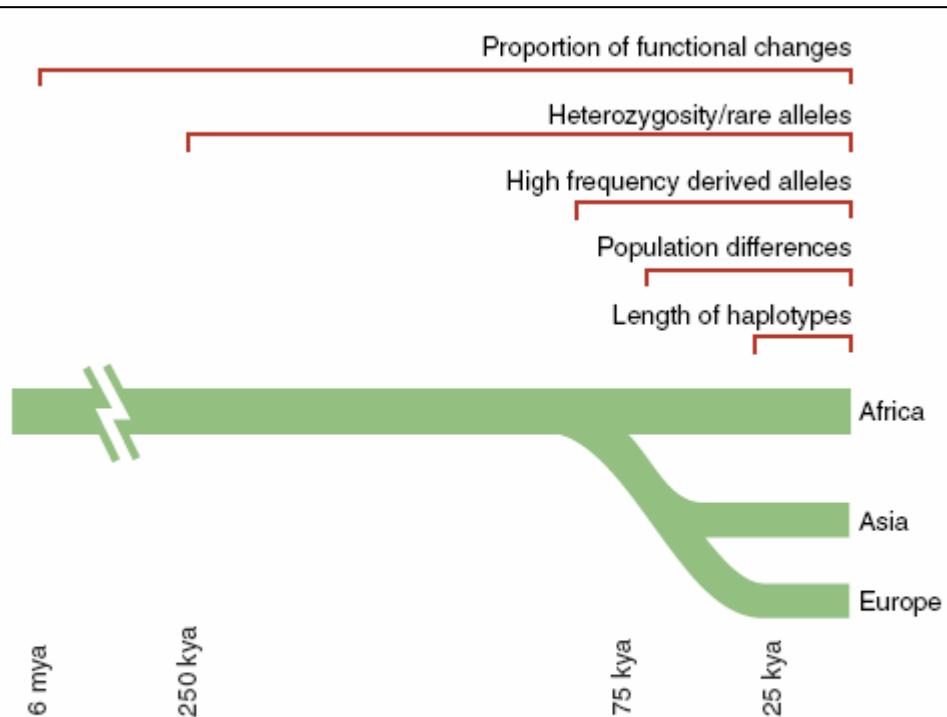
Positive Selection

How can we  
detect negative  
selection?

How can we  
detect positive  
selection?



# How can we detect positive selection?



**Fig. 1.** Time scales for the signatures of selection. The five signatures of selection persist over varying time scales. A rough estimate is shown of how long each is useful for detecting selection in humans. (See fig. S1 for details on how the approximate time scales were estimated).

## Ka/Ks ratio:

Ratio of nonsynonymous to synonymous substitutions

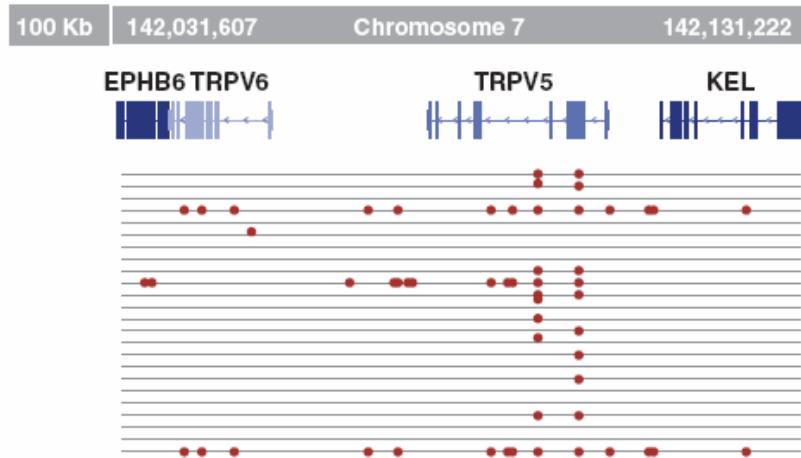
Very old, persistent, strong positive selection for a protein that keeps adapting

**Examples:** immune response, spermatogenesis

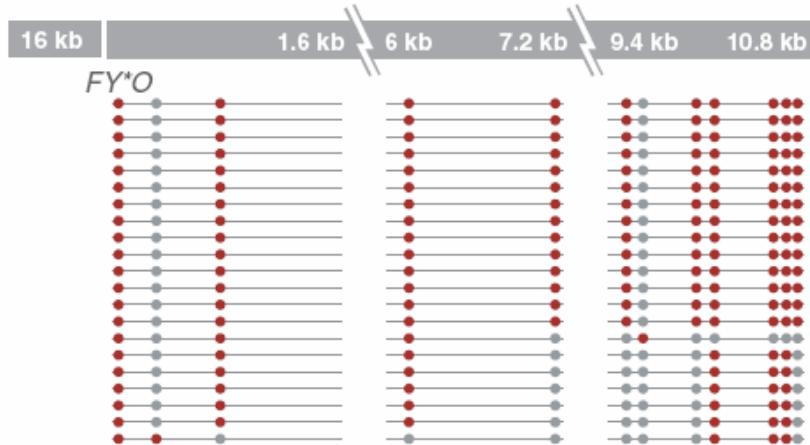
PRM1 Exon 2											
44 bp	11,341,281 Chromosome 16 11,341,324										
Human	STOP	H	R	R	C	R	P	R	Y	R	P
	AATCACAGAAGATGTAG	CGCC	AGAC	ATGGAC	CCGCCGCTGTGG						
Chimp	STOP	H	R	R	R	M	R	S	R	R	C
	AATCACAGAAGATGCA	GAG	TAAG	ACCTGGAC	CCGCCGCTGTGG						

**Fig. 2.** Excess of function-altering mutations in *PRM1* exon 2. The *PRM1* gene exon 2 contains six differences between humans and chimpanzees, five of which alter amino acids (7, 8).

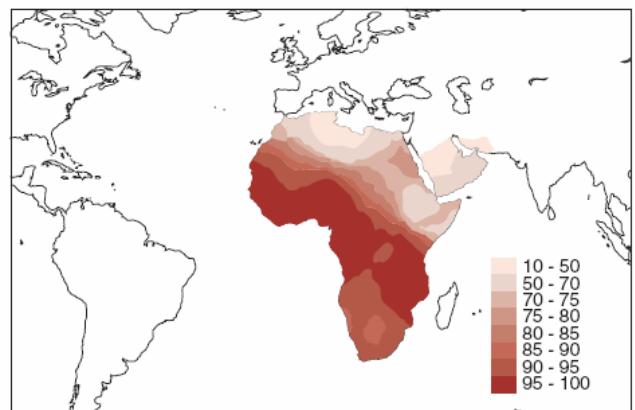
# How can we detect positive selection?



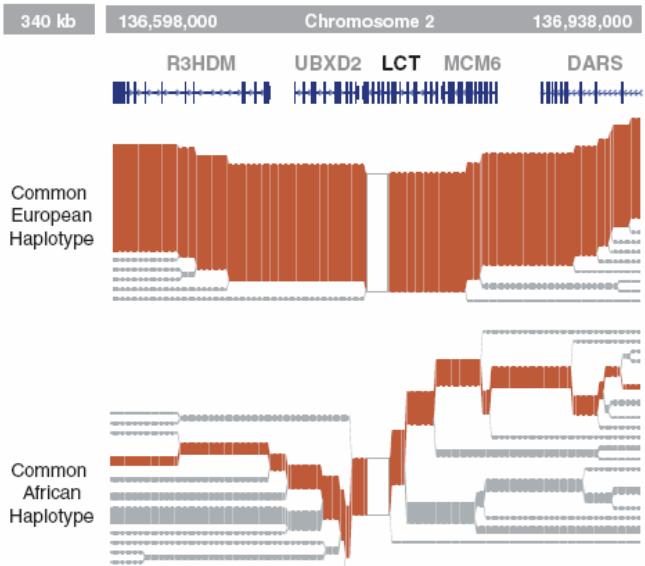
**Fig. 3.** Low diversity and many rare alleles at the Kell blood antigen cluster. On the basis of three different statistical tests, the 115-kb region (containing four genes) shows evidence of a selective sweep in Europeans (28).



**Fig. 4.** Excess of high-frequency derived alleles at the Duffy red cell antigen (*FY*) gene (34). The 10-kb region near the gene has far greater prevalence of derived alleles (represented by red dots) than of ancestral alleles (represented by gray dots).



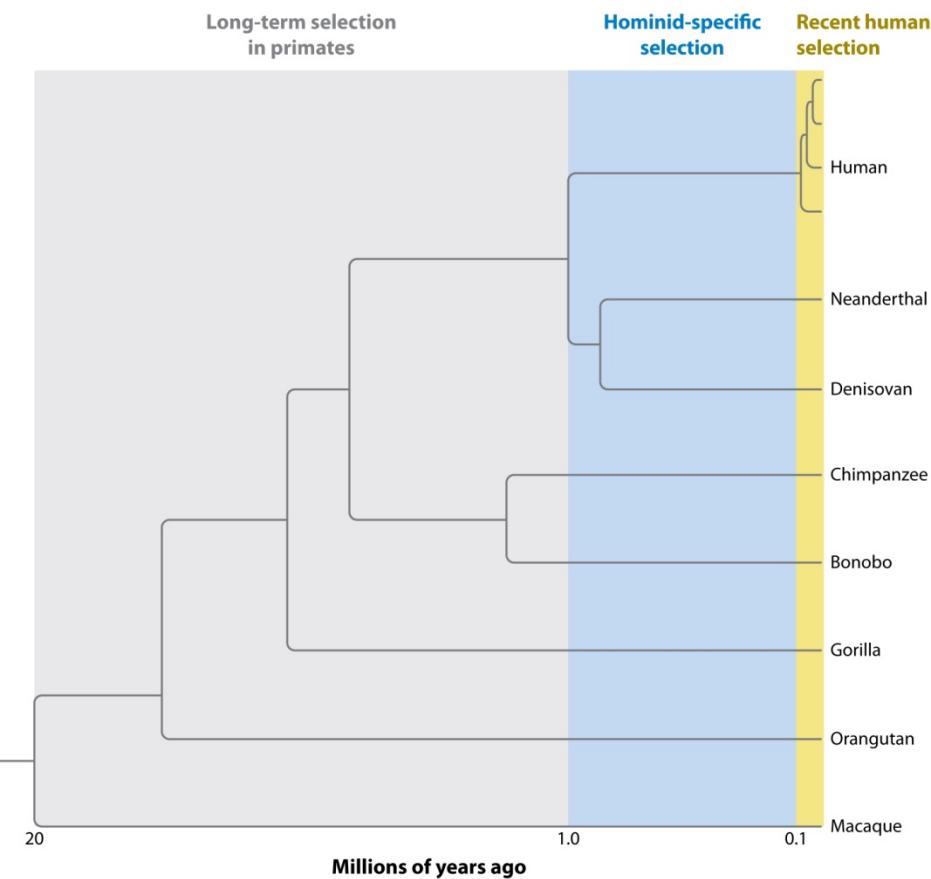
**Fig. 5.** Extreme population differences in *FY\*O* allele frequency. The *FY\*O* allele, which confers resistance to *P. vivax* malaria, is prevalent and even fixed in many African populations, but virtually absent outside Africa (38).



**Fig. 6.** Long haplotype surrounding the lactase persistence allele. The lactase persistence allele is prevalent (~77%) in European populations but lies on a long haplotype, suggesting that it is of recent origin (6).



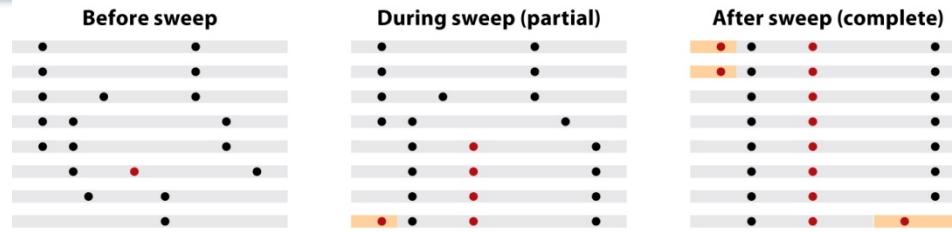
# Positive Selection in Human Lineage



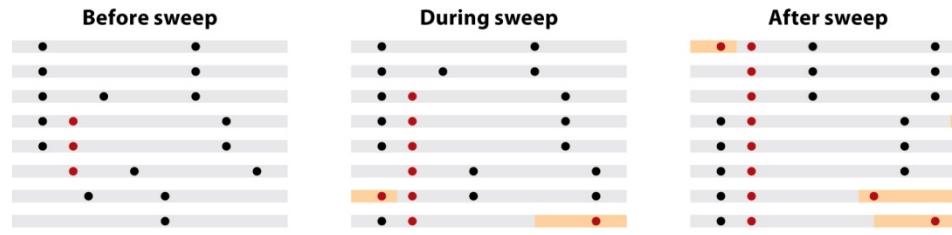
Fu W, Akey JM. 2013.

Annu. Rev. Genomics Hum. Genet. 14:467–89

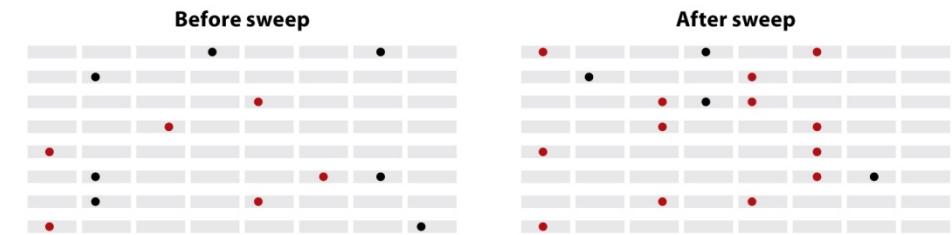
## a Hard sweep



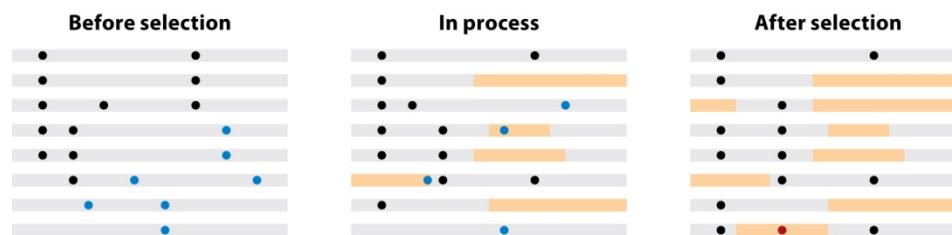
## b Positive selection on standing variation



## c Polygenic selection (adaptation)



## d Purifying selection

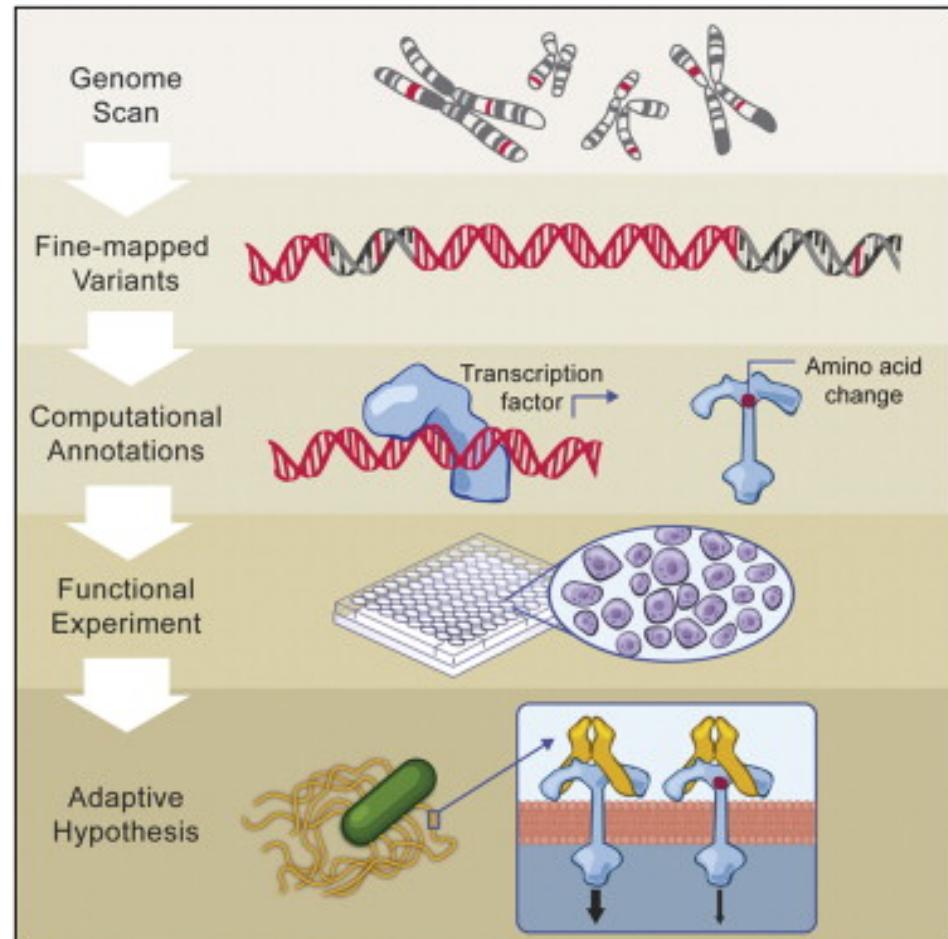
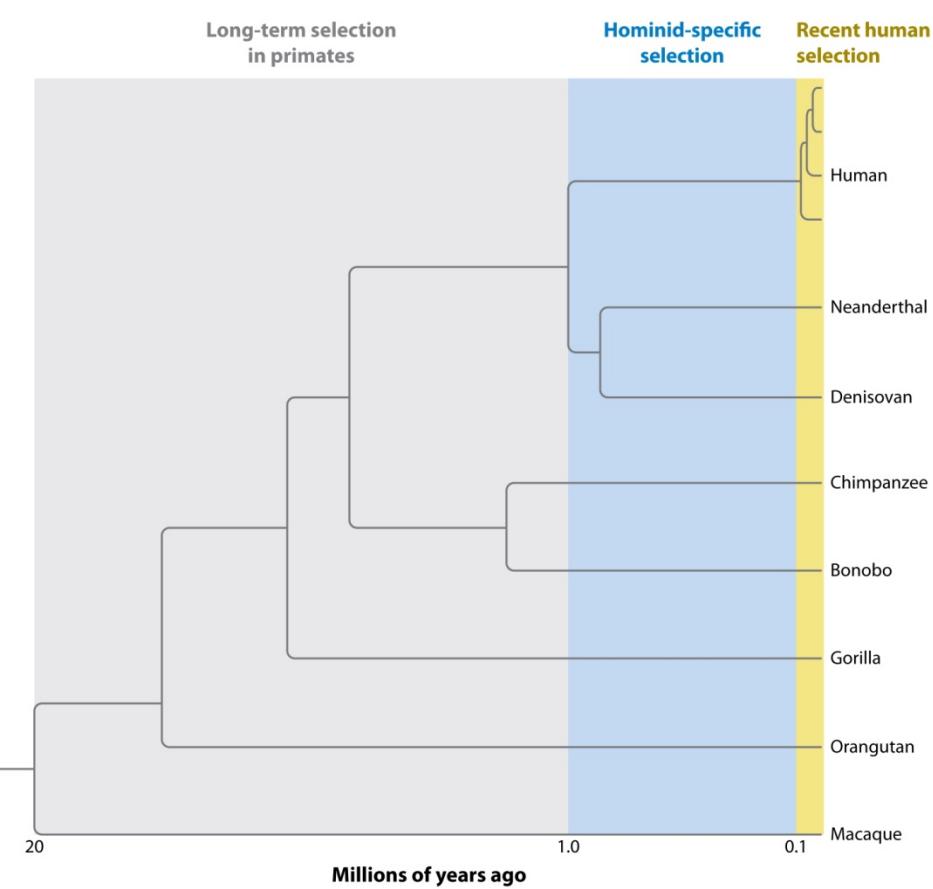


Fu W, Akey JM. 2013.

Annu. Rev. Genomics Hum. Genet. 14:467–89



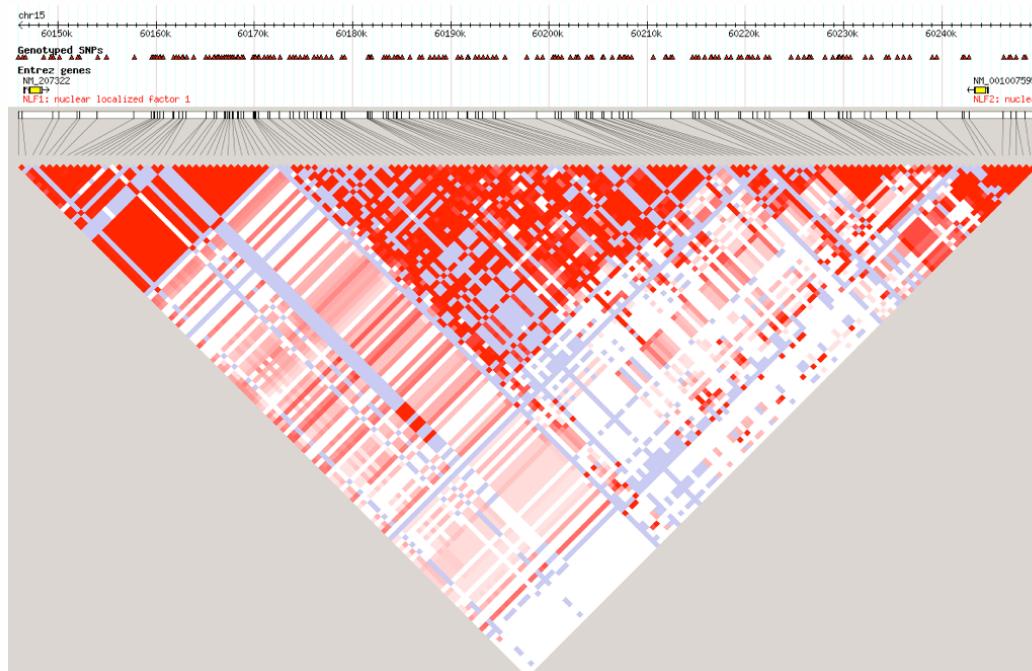
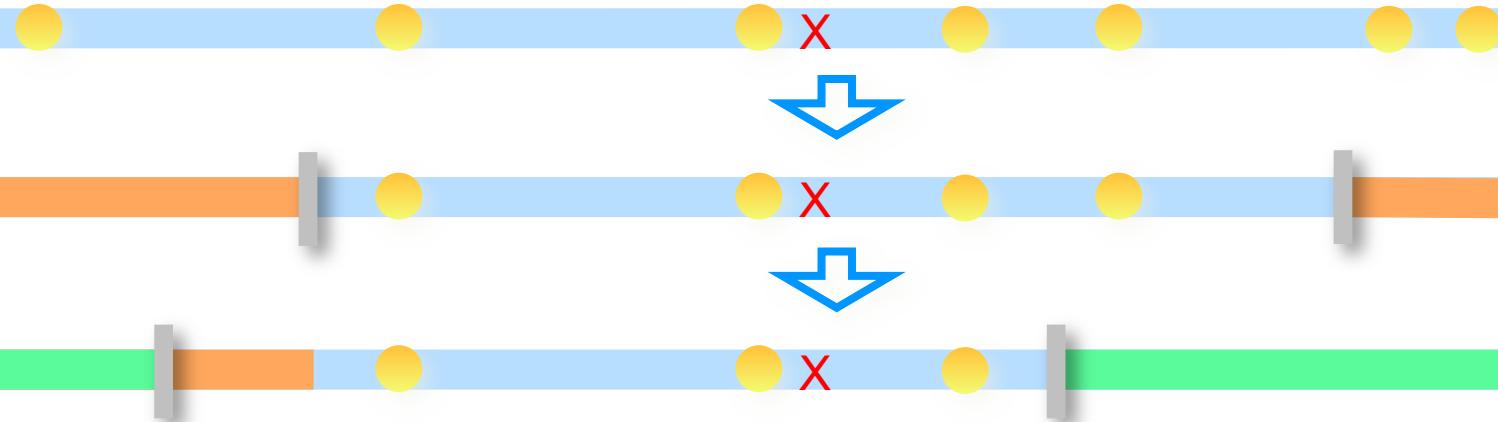
# Positive Selection in Human Lineage



Fu W, Akey JM. 2013.

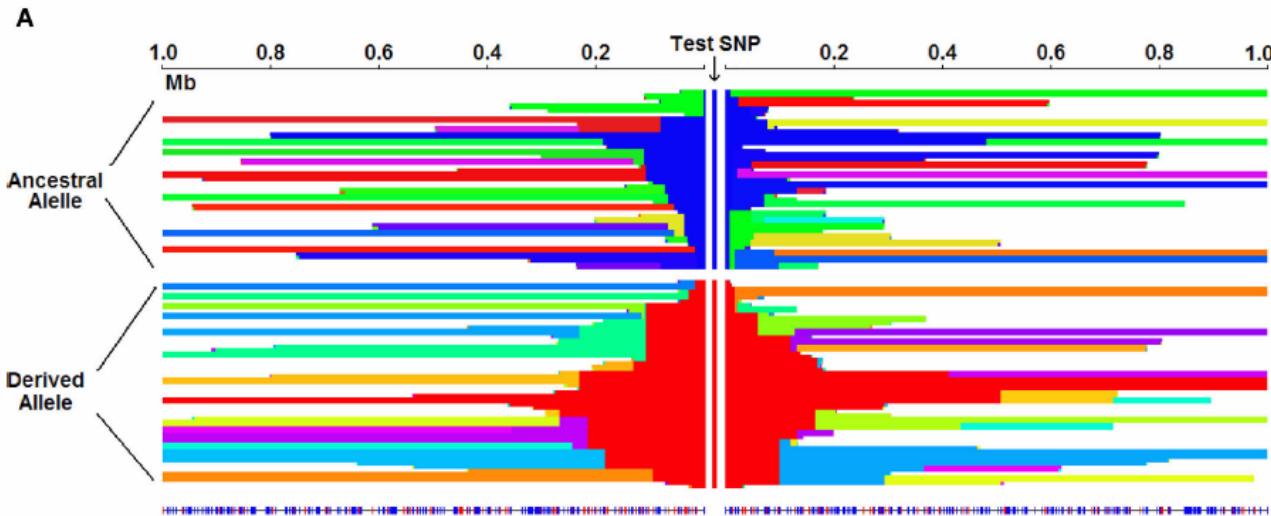
Annu. Rev. Genomics Hum. Genet. 14:467–89

# Mutations and LD



Slide Credits:  
Marc Schaub

# Long Haplotypes -EHS, iHS tests



$$iHS = \ln\left(\frac{iHH_A}{iHH_D}\right)$$

Less time:  
• Fewer mutations  
• Fewer recombinations

# Application: Malaria



- Study of genes known to be implicated in the resistance to malaria.
- Infectious disease caused by protozoan parasites of the genus *Plasmodium*
- Frequent in tropical and subtropical regions
- Transmitted by the *Anopheles* mosquito



Slide Credits:  
Image source: wikipedia.org  
Marc Schaub

# Application: Malaria

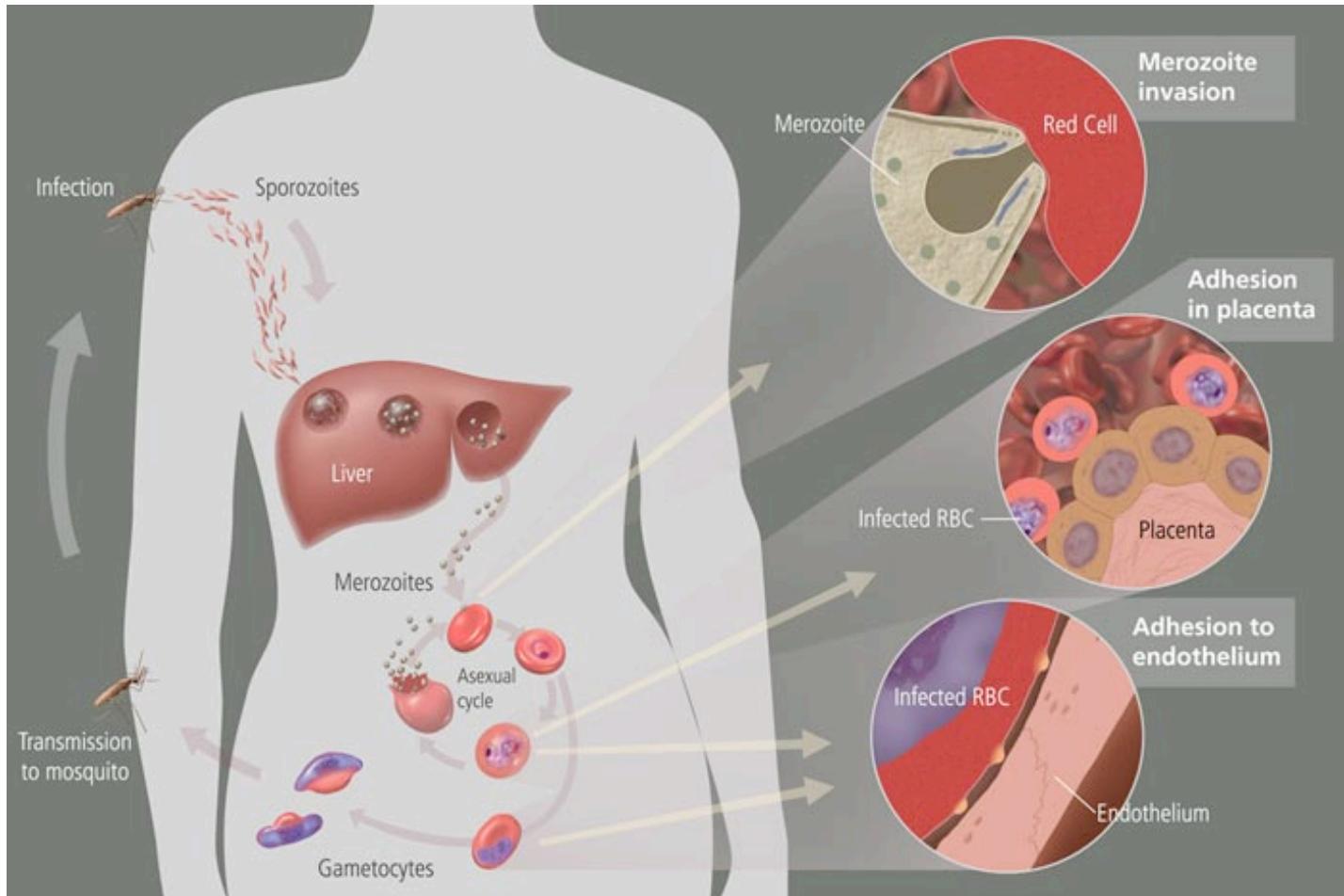


Image source:

NIH - <http://history.nih.gov/exhibits/bowman/images/malariacycleBig.jpg>

Slide Credits:  
Marc Schaub

# Application: Malaria



## Malaria Endemic Countries, 2003

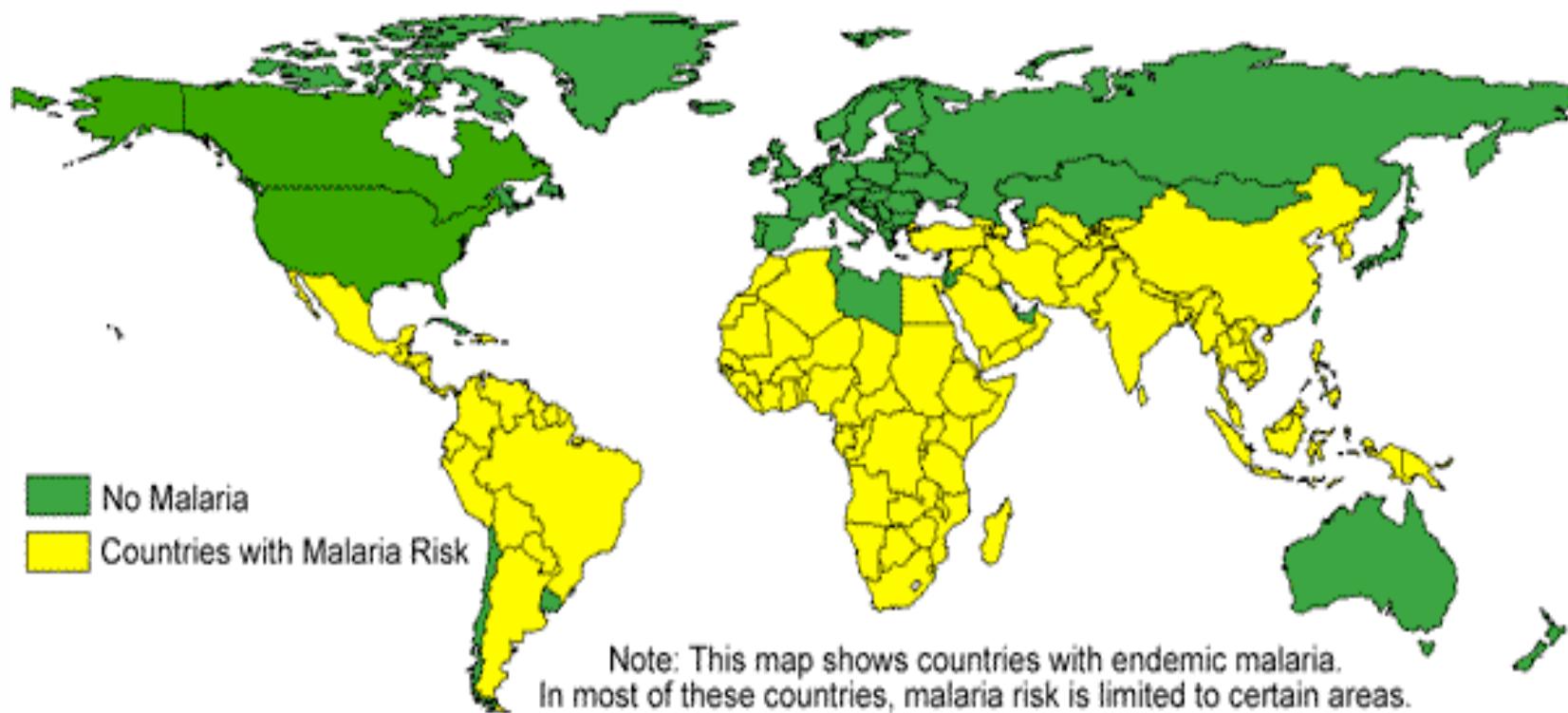
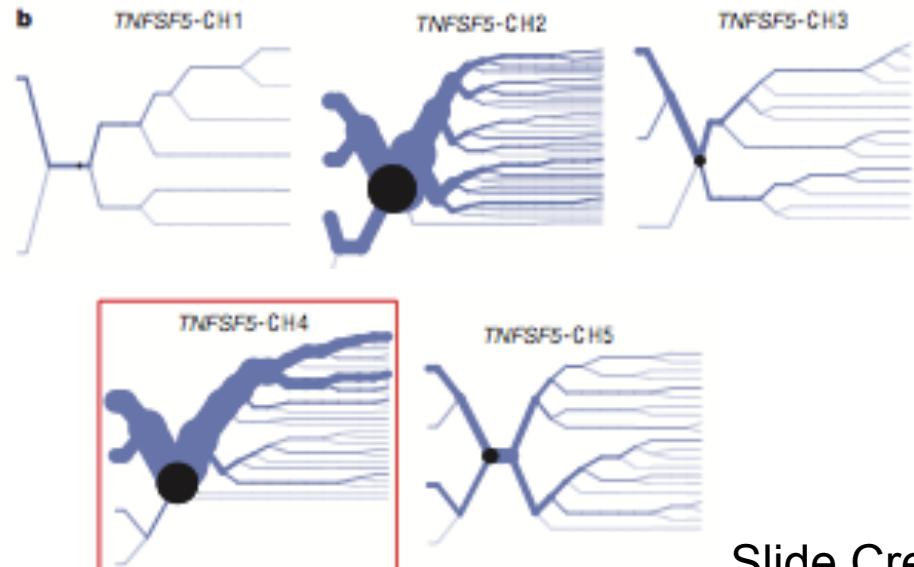
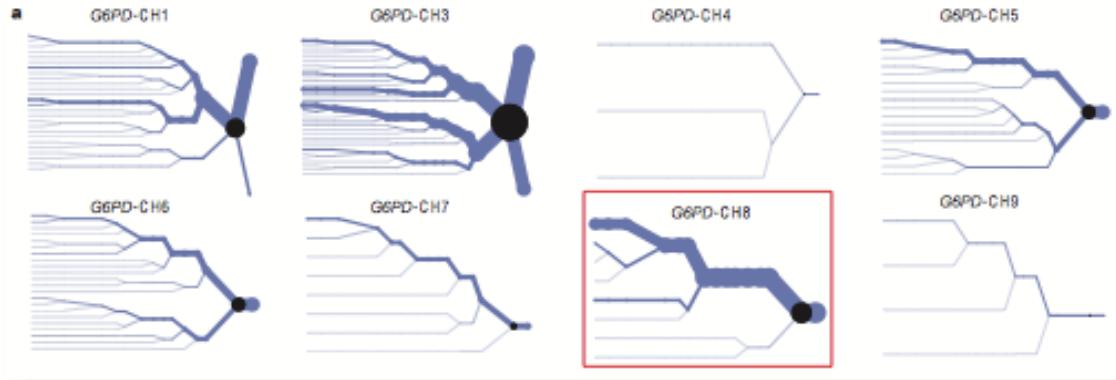


Image source: CDC -

[http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/M-R/Malaria/  
malaria\\_risk\\_2003.gif](http://www.dpd.cdc.gov/dpdx/images/ParasiteImages/M-R/Malaria/malaria_risk_2003.gif)

Slide Credits:  
Marc Schaub

# Results: G6PD, TNFSF5



Source: Sabeti et al. Nature 2002.

Slide Credits:  
Marc Schaub

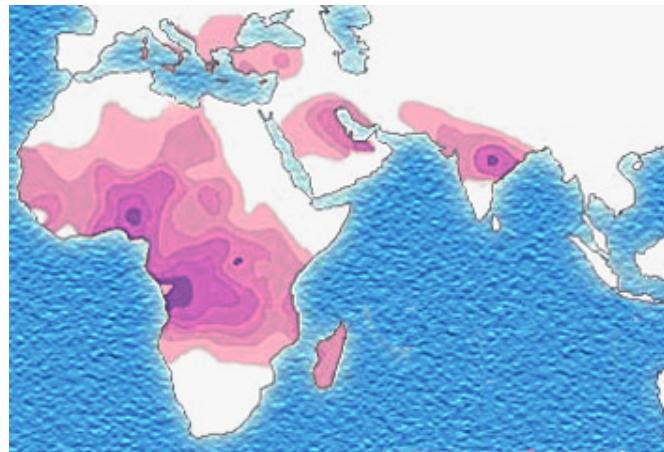
# Malaria and Sickle-cell Anemia



- Allison (1954): Sickle-cell anemia is limited to the region in Africa in which malaria is endemic.



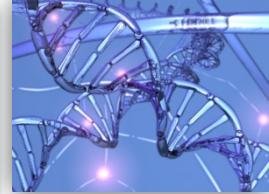
Distribution of malaria



Distribution of sickle-cell anemia

Image source: wikipedia.org

Slide Credits:  
Marc Schaub



# Lactose Intolerance

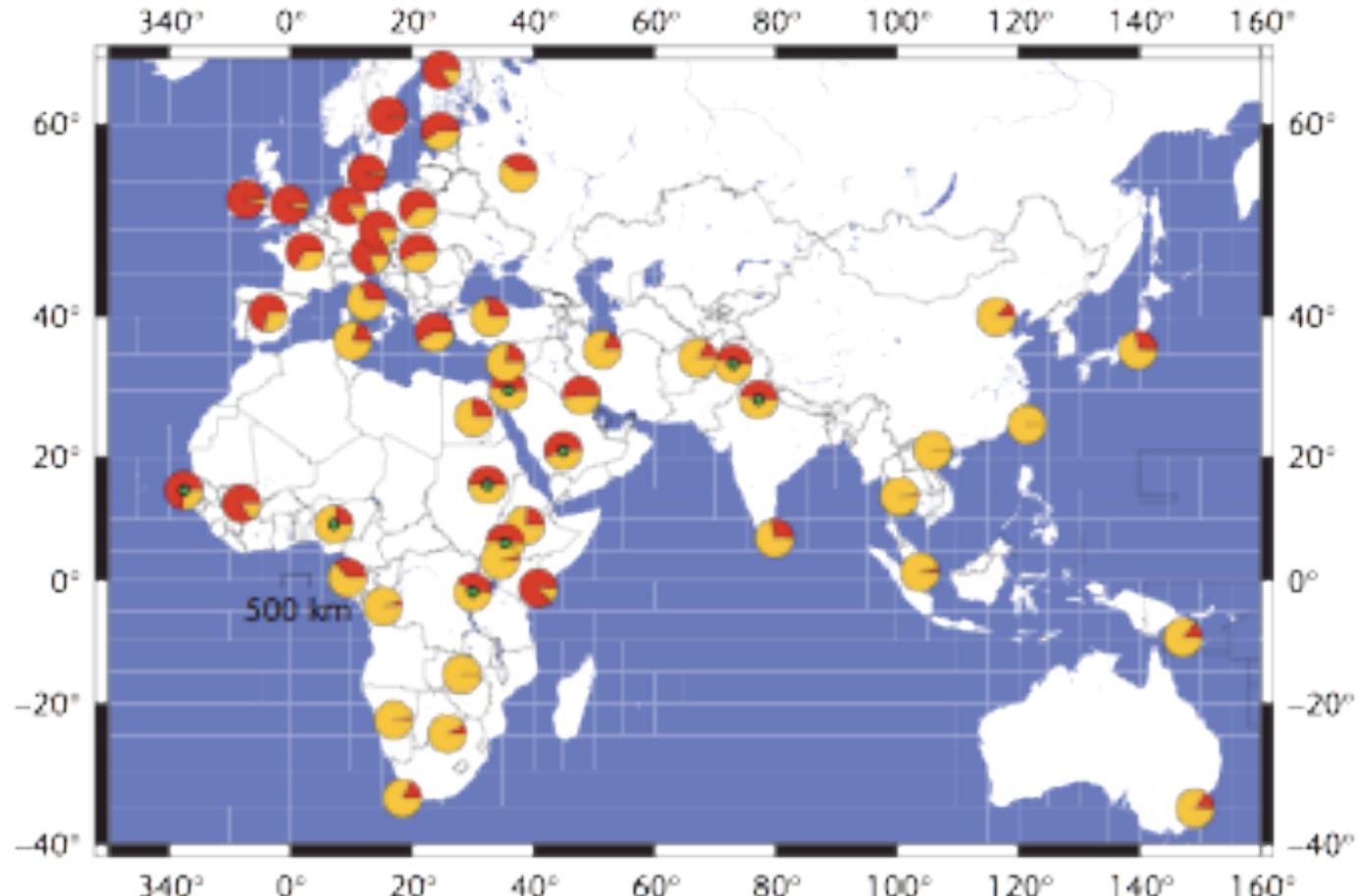
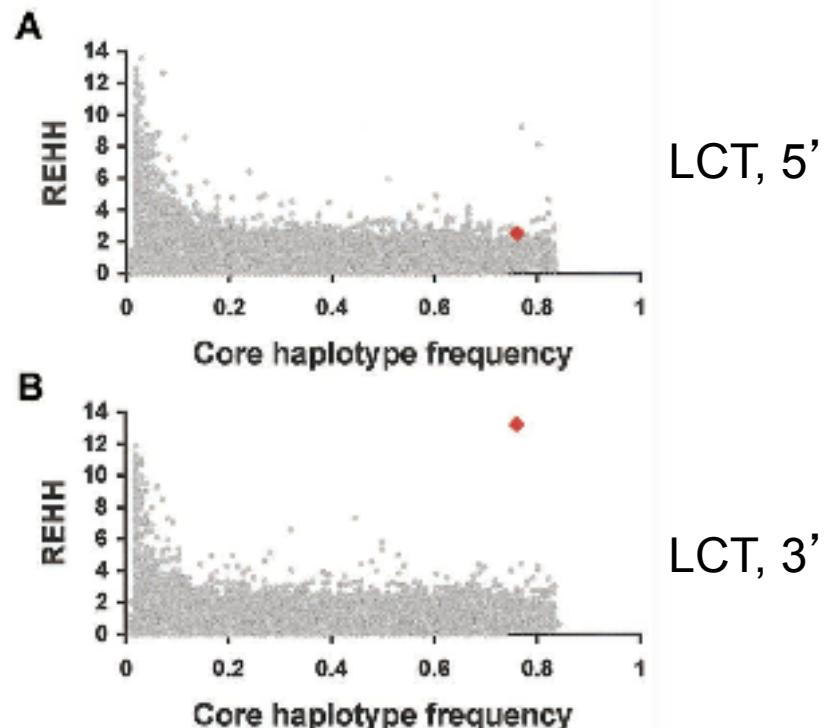
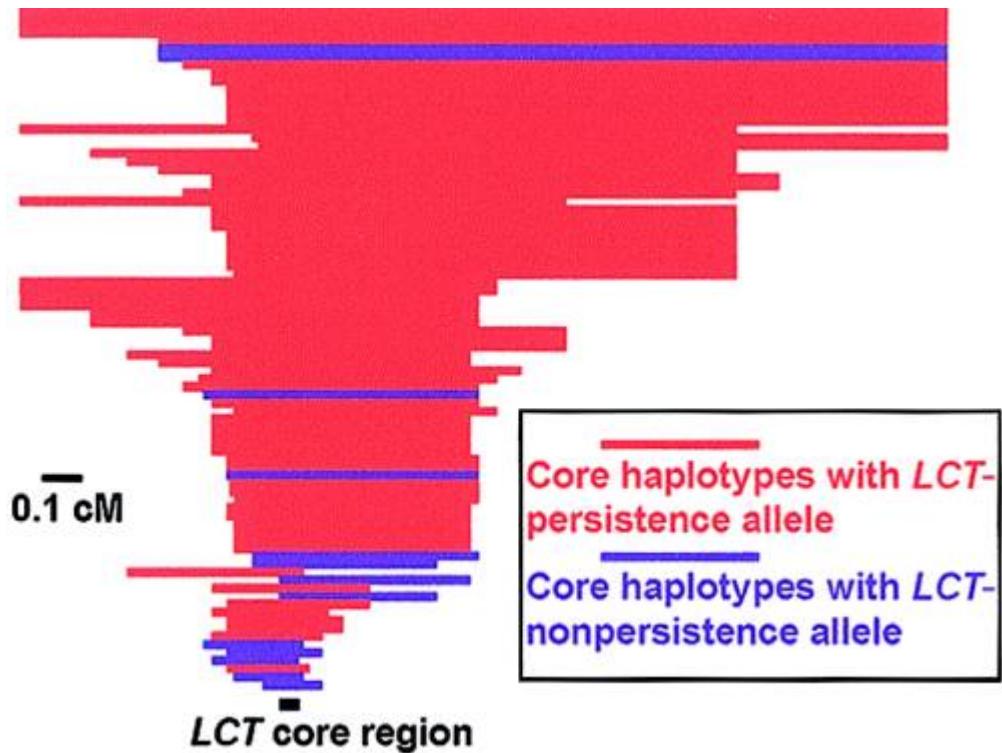


Figure 1 Old world distribution of frequency of lactase persistence (lactose digesters) taken from available published data. Red indicates the proportion of lactose digesters in a given population with yellow representing maldigesters. Charts with a green central circle indicate that the overall published frequency for a country is comprised of different ethnic groups with very different phenotype frequencies. Data compiled by Ingram 2007.

Source: Ingram and Swallow. Population Genetics of Lactose Persistence. Encyclopedia of Life Sciences. 2007.

Slide Credits:  
Marc Schaub

# Lactose Intolerance



Source: Bersaglieri *et al.* Am. J. Hum. Genet. 2004.

Slide Credits:  
Marc Schaub

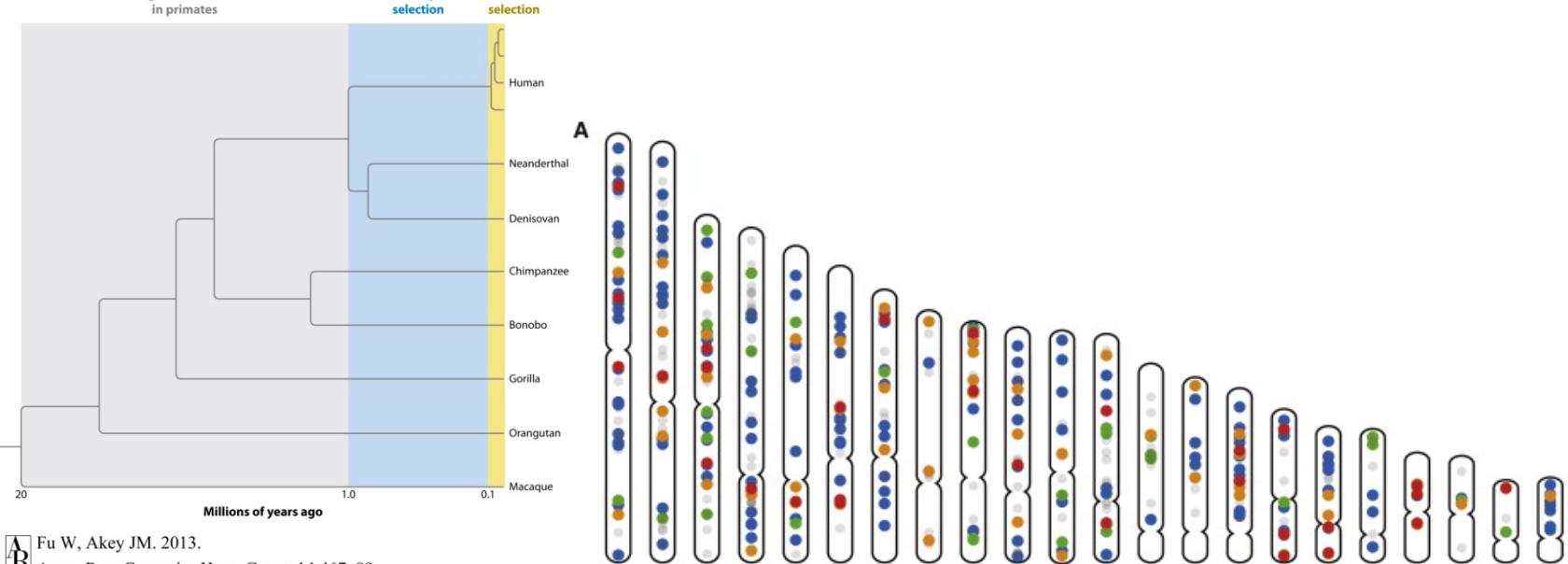


# Positive Selection in Human Lineage

Long-term selection  
in primates

Hominid-specific  
selection

Recent human  
selection



Fu W, Akey JM. 2013.

Annu. Rev. Genomics Hum. Genet. 14:467–89

**B**

