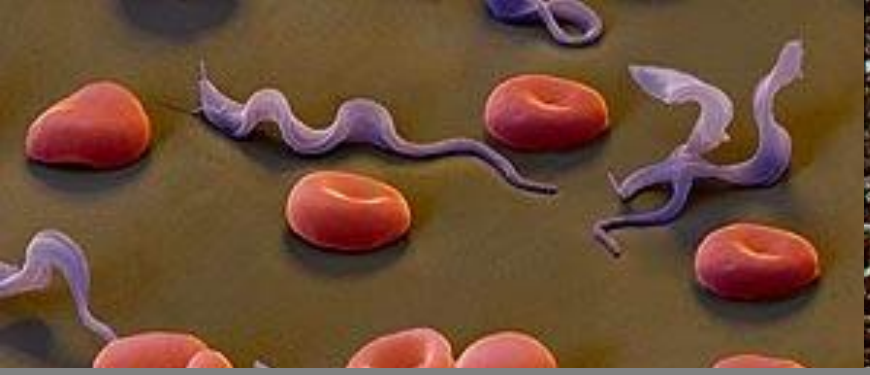
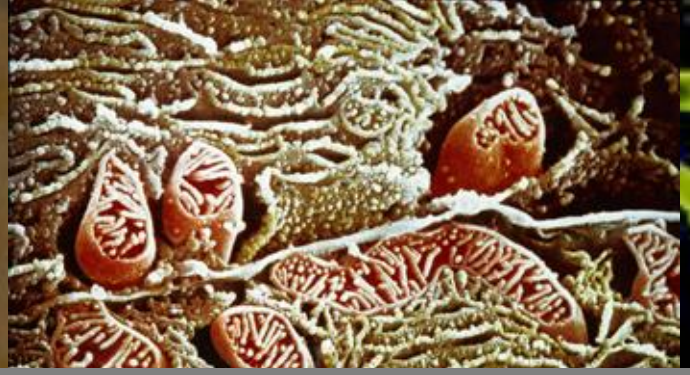
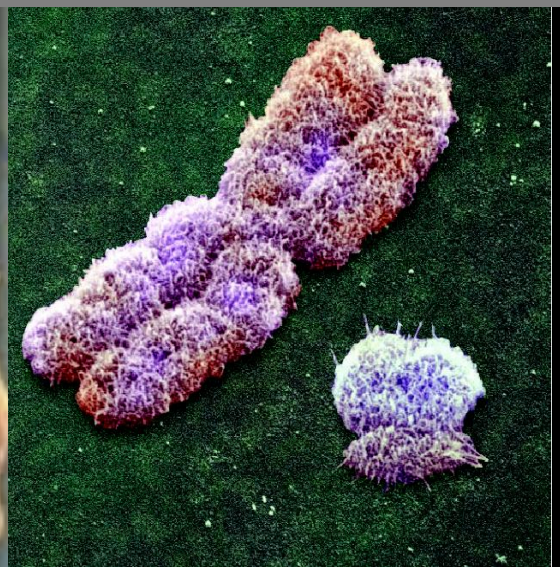



Human Genome Diversity:
of human evolution and natural selection

Guillaume LAVAL
CNRS URA3012, Unité de Génétique Evolutive Humaine, Institut Pasteur



Human Genetics

Has this mutation a phenotypic effect ?

When did this mutation appear?
Why it is present a such a frequency?

→ Medical Genetics

→ Epidemiological genetics

→ Population Genetics

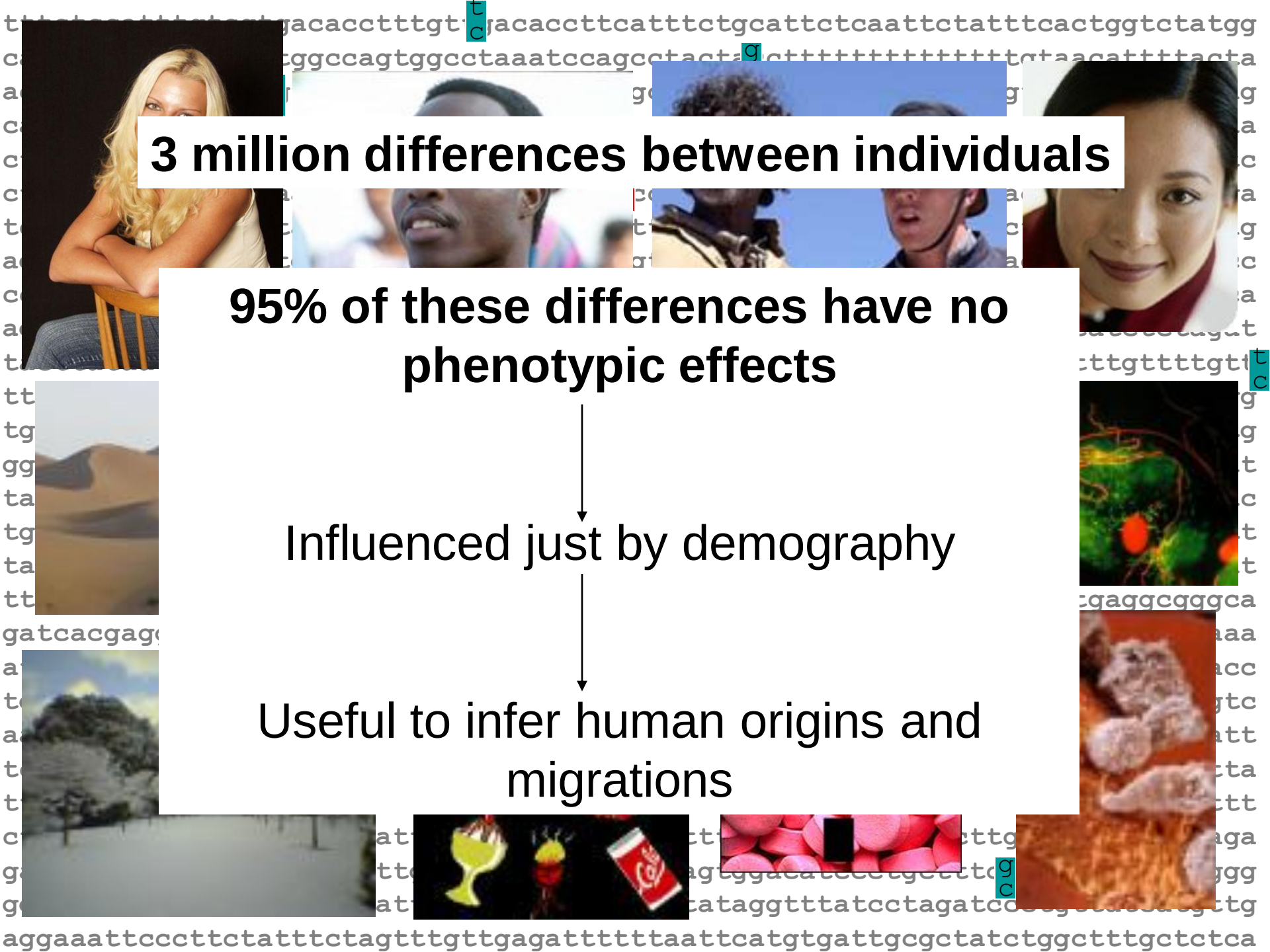
→ Evolutionary Genetics

Recent major steps in human genetics

1. The complete sequence of the human genome
2. SNP Discovery and Characterization

3. The HapMap Project (now Phase III)

4. The HGDP-CEPH
5. The 1,000 genomes project



3 million differences between individuals

95% of these differences have no phenotypic effects

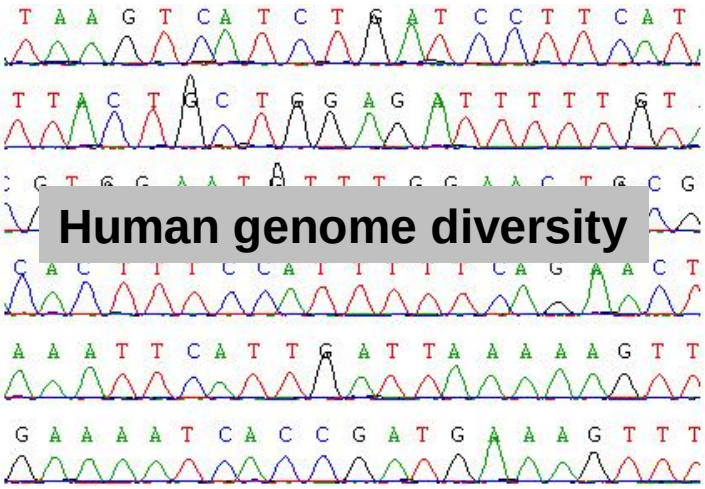
↓

Influenced just by demography

↓

Useful to infer human origins and migrations

Forces shaping genetic diversity



Demographic history

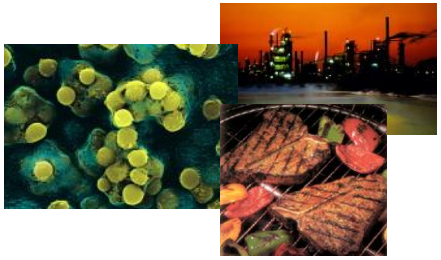
Expansions, bottlenecks

Cultural forces

Lifestyle, social organization

Natural selection

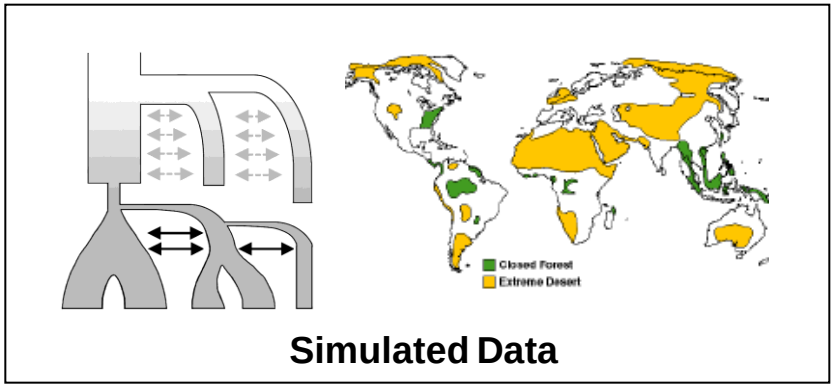
Adaptation to climate, pathogens, food

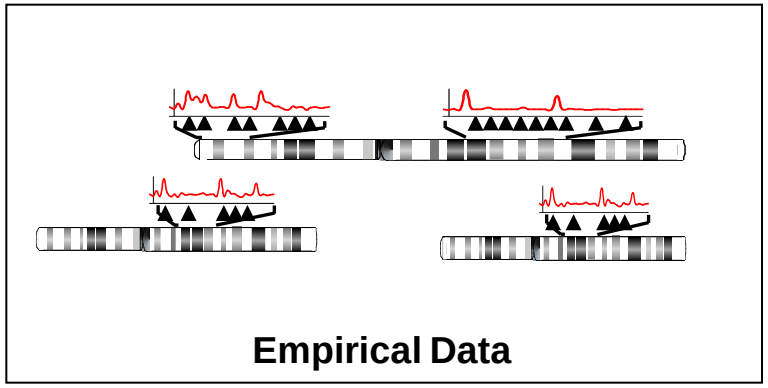


Genomic forces

Mutation, recombination

1st angle :
Influence of human demography on genome diversity





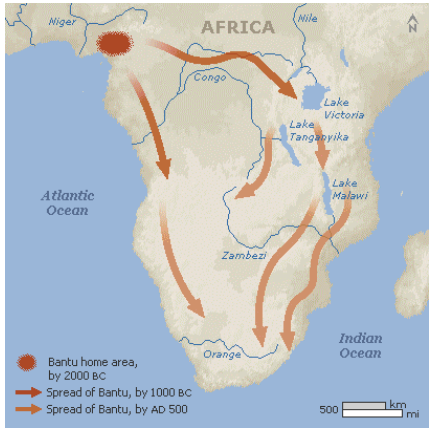
Model : sequencing of non-coding regions of the genome (autosomal, Y-chromosome, mtDNA, etc.) and comparison of observed data with simulated data (Coalescence, ABC)

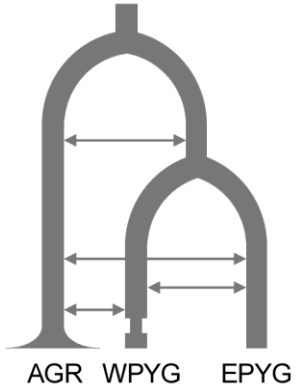
Objectives :

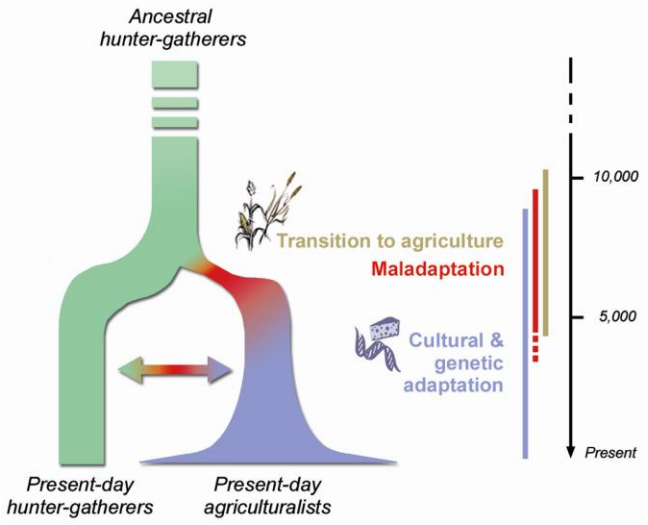
- better understand the demographic history of human populations using a multi-locus approach and a robust statistical framework

- determine the most likely model of human demography in order to « calibrate » the effects of demography on human genome variability and therefore, detect the effects of natural selection robustly.

2nd angle :
Influence of lifestyle on genome diversity







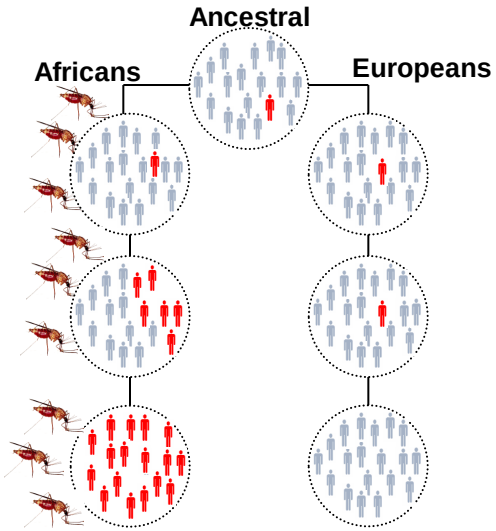
Patin & Quintana-Murci, ***Trends Ecol Evol*** 2008

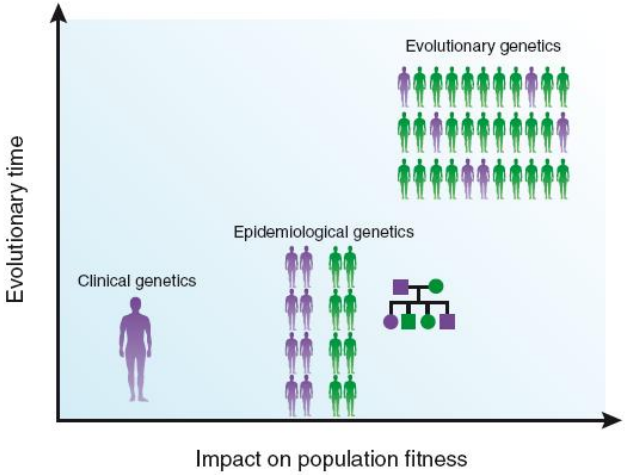
Model : comparison of genetic diversity in populations presenting completely different lifestyles: Pygmy hunter-gatherers and Bantu farmers

Objectives :

- determine the demographic history of these populations and their genetic interactions
- better understand how lifestyle and modes of subsistence have influenced population genetic diversity

3rd angle :
Influence of natural selection on genome diversity





Quintana-Murci et al., ***Nature Immunology*** 2008

General*:* the detection of natural selection at the genome-wide level

Model : different families of innate immunity receptors (TLRs, Nod-like, RIG-I like, type-C lectins) and their signaling pathways

Objectives :

- identify regions of the human genome having participated in our adaptation to the environment (climatic, nutritional, pathogenic)

- In the context of immunity, distinguish those genes having played a major role in human survival from those playing a more redundant role in host defense against pathogens

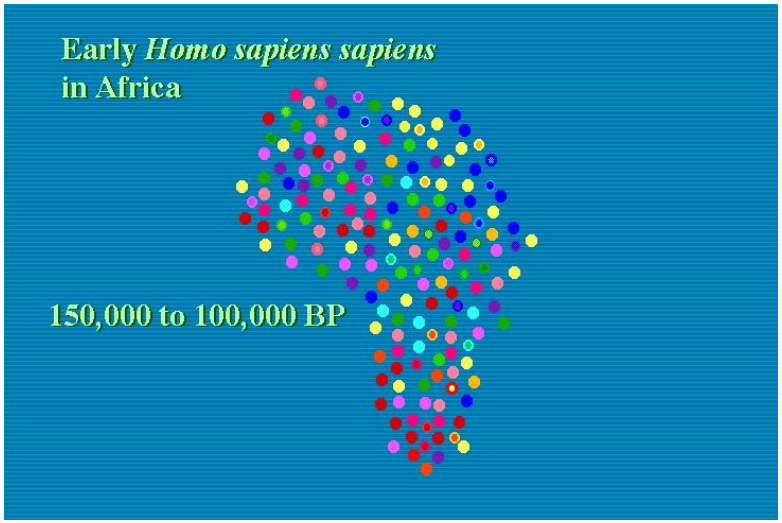
Examples

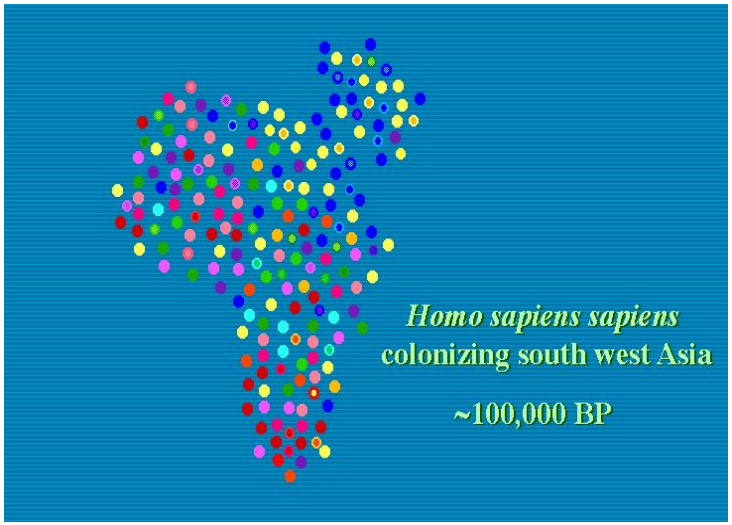
- Demographic Inference
  - Worldwide inference of past demographic events
  - History, population history and lifestyle
- Selective Inference
  - Genomewide detection of natural selection
  - Natural selection and immunity to infection

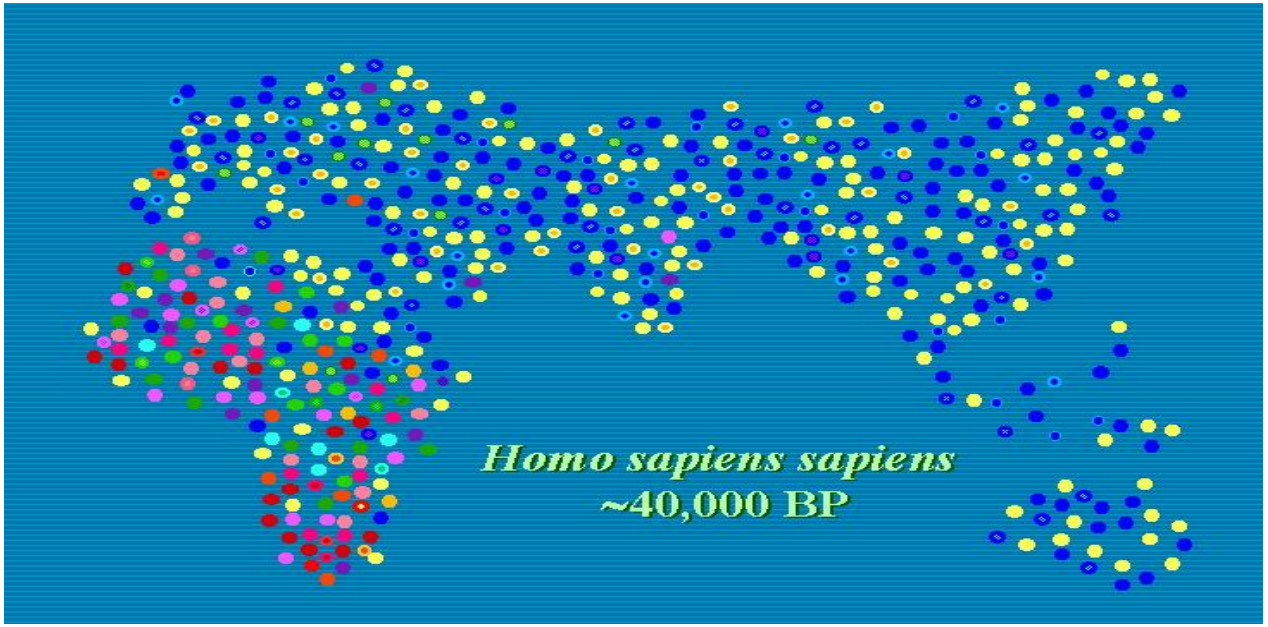
Examples

- Demographic Inference
  - Worldwide inference of past demographic events
  - History, population history and lifestyle
- Selective Inference
  - Genomewide detection of natural selection
  - Natural selection and immunity to infection

Population history and genetic diversity



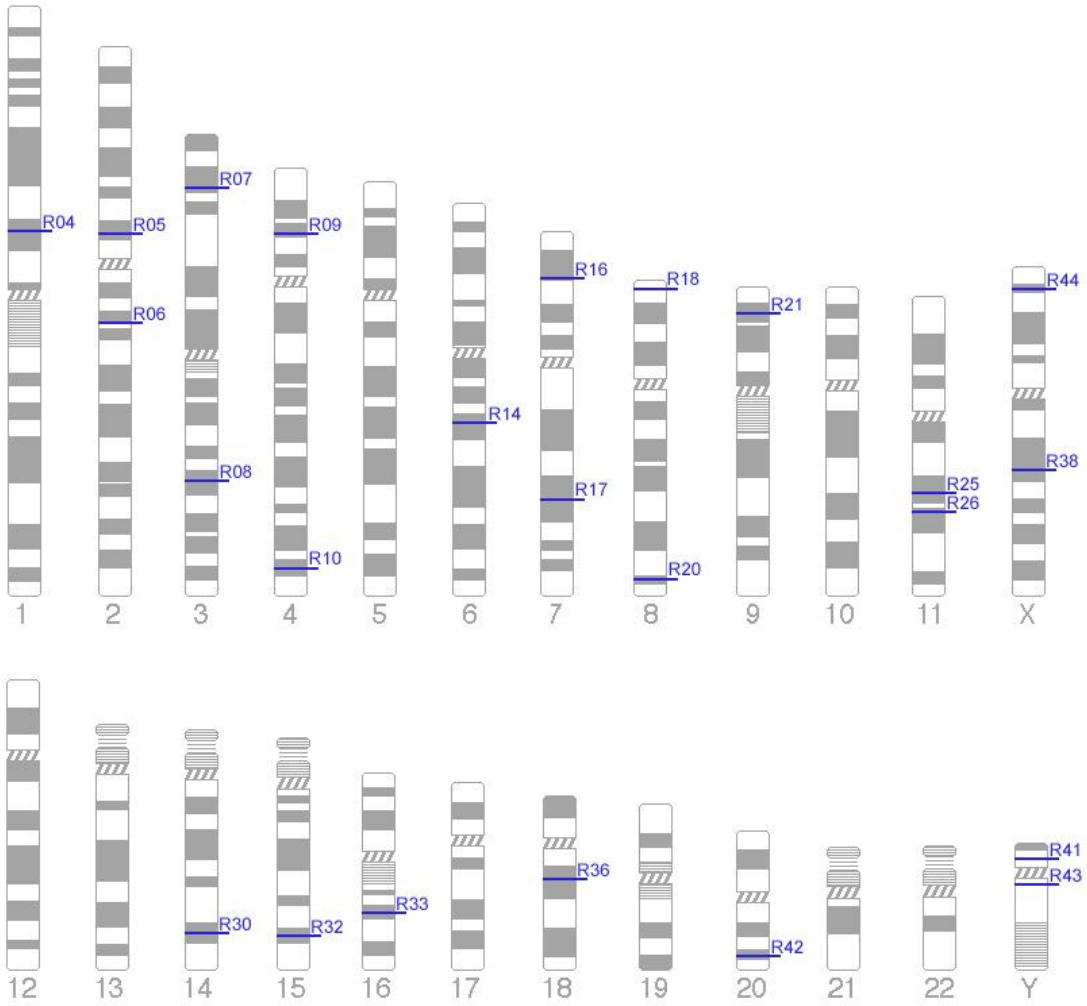




Questions

- When humans left Africa?
- How was the intensity of the bottleneck out of Africa?
- How intense have been the population expansions of modern humans?
- When the ancestors of contemporary Europeans and Asians started to diverge?

Demographic inference: non-coding regions of the genome

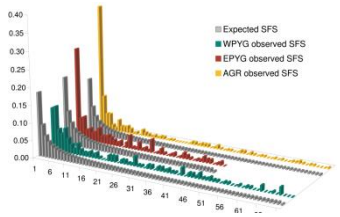


213 individuals from Africa
Europe east-Asia

Total sequence per individuals of ~27 kb of diploid

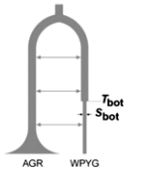
A best-fit model for the historical and demographic history

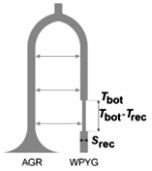
Empirical data



Observed summary statistics

Simulated data

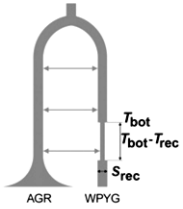




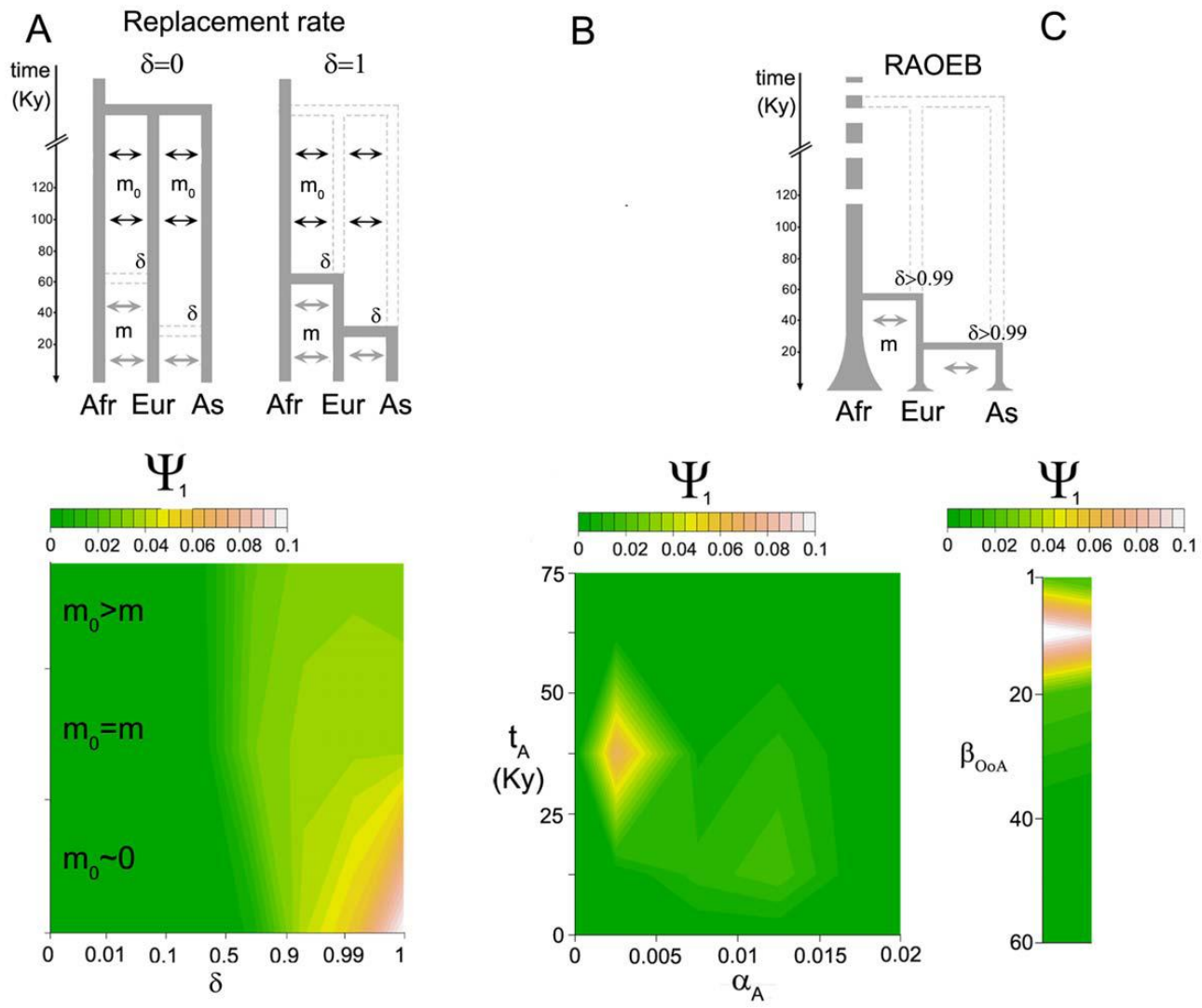
Simulated summary statistics

Assessment of weakest distances

Best-fit model

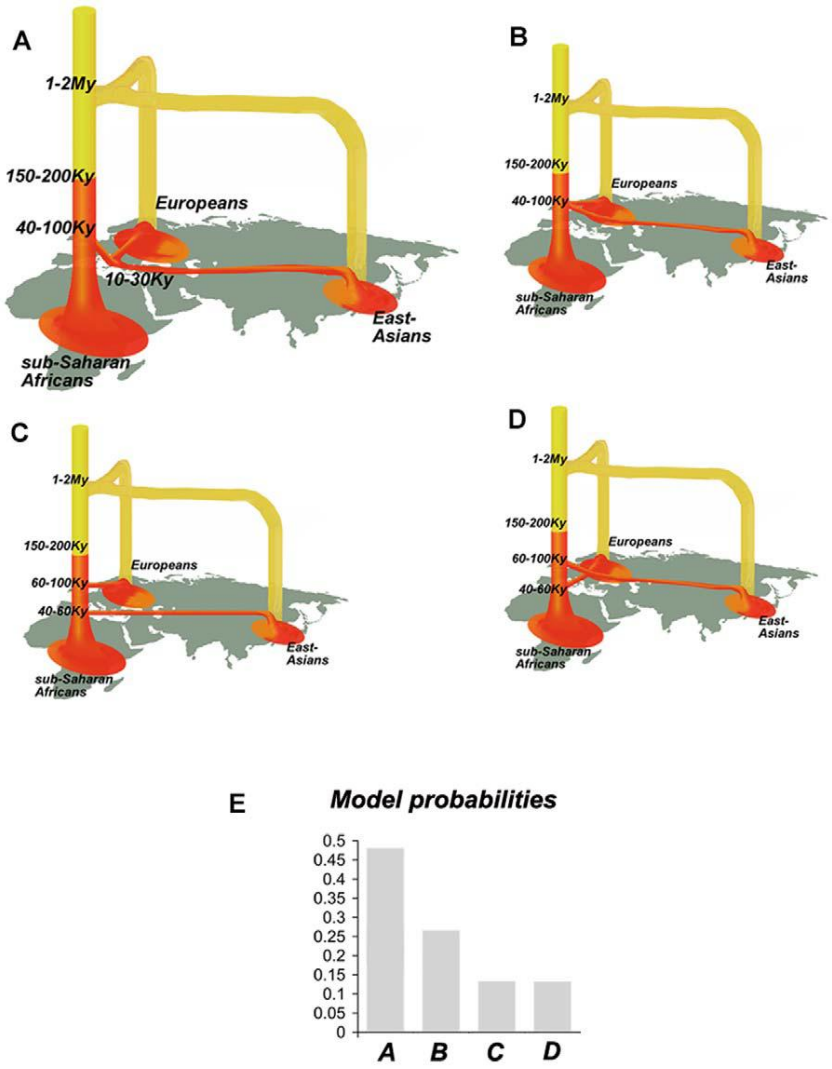


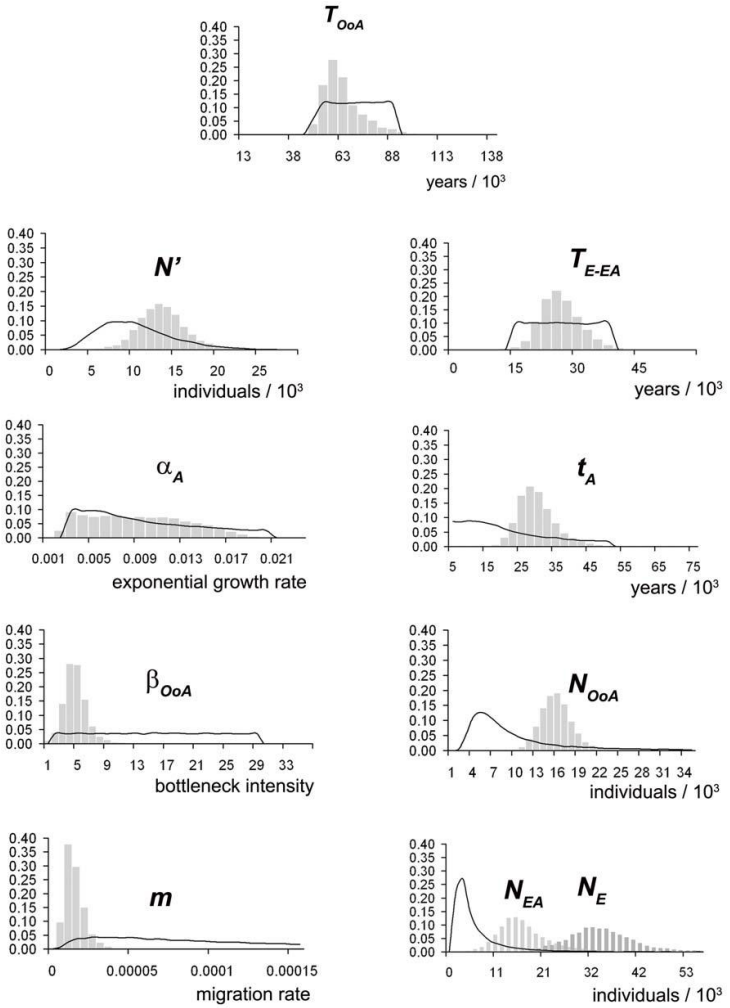
Model and parameter best-fitted estimations.



Laval et al. ***PLoS One*** 2010

Estimation of historical and demographic parameters



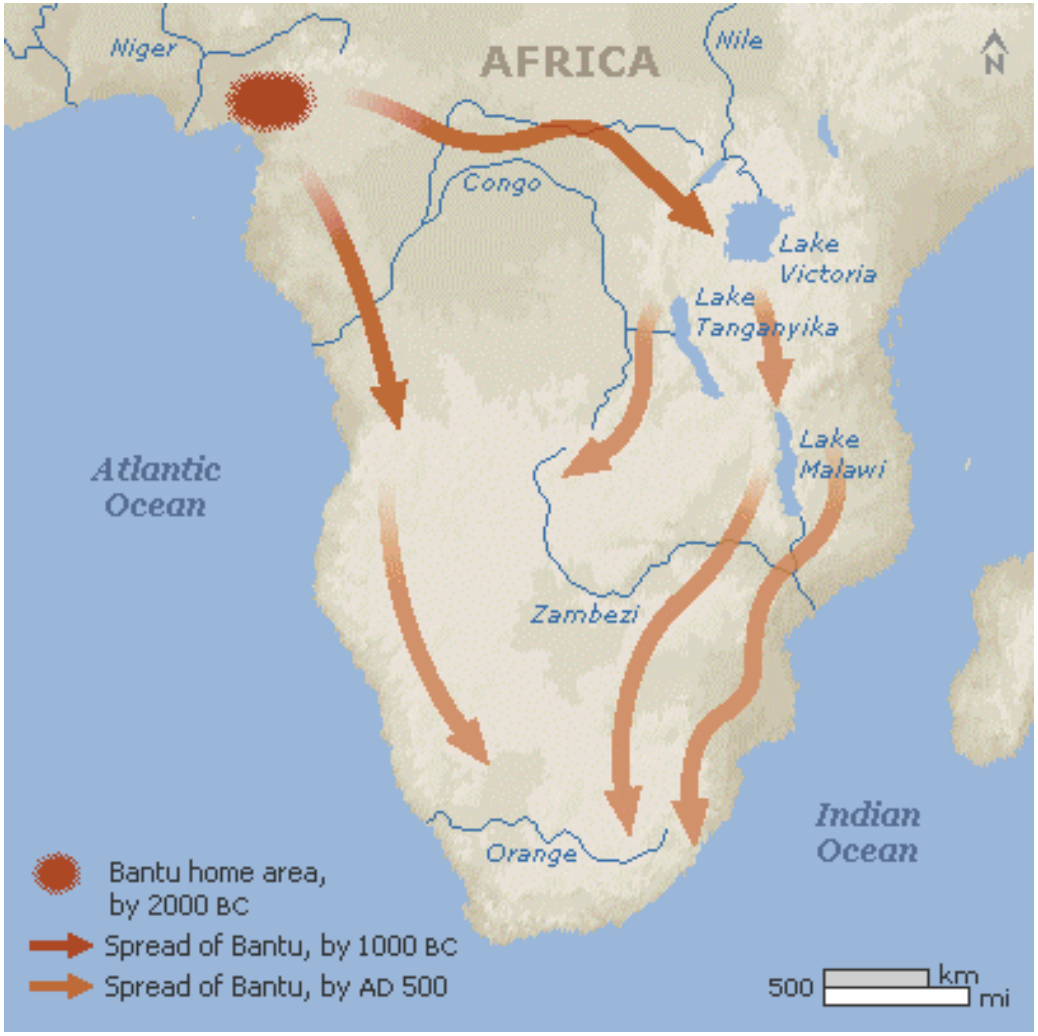


Laval et al. ***PLoS One*** 2010

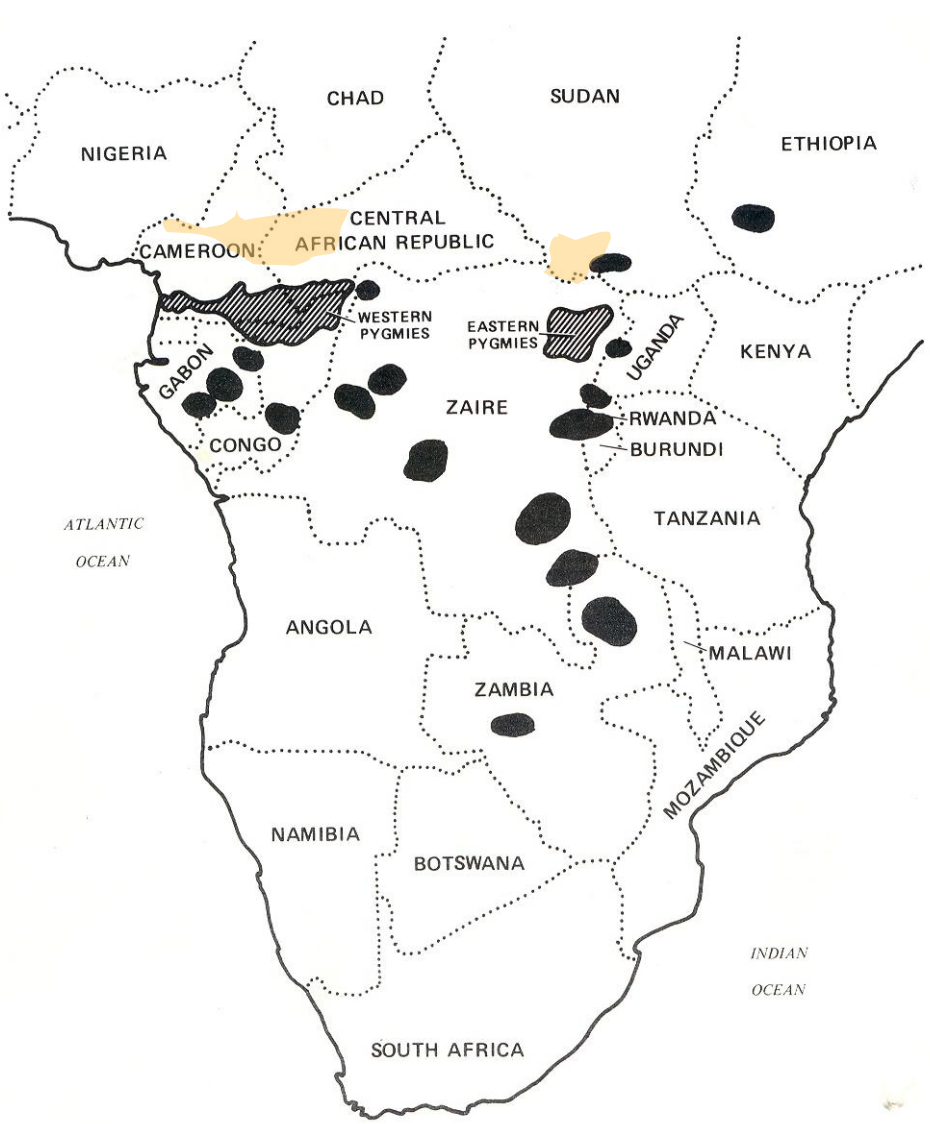
Examples

- Demographic Inference
  - Worldwide inference of past demographic events
  - History, population history and lifestyle
- Selective Inference
  - Genomewide detection of natural selection
  - Natural selection and immunity to infection

Our model: Pygmy hunter-gatherers and Bantu farmers



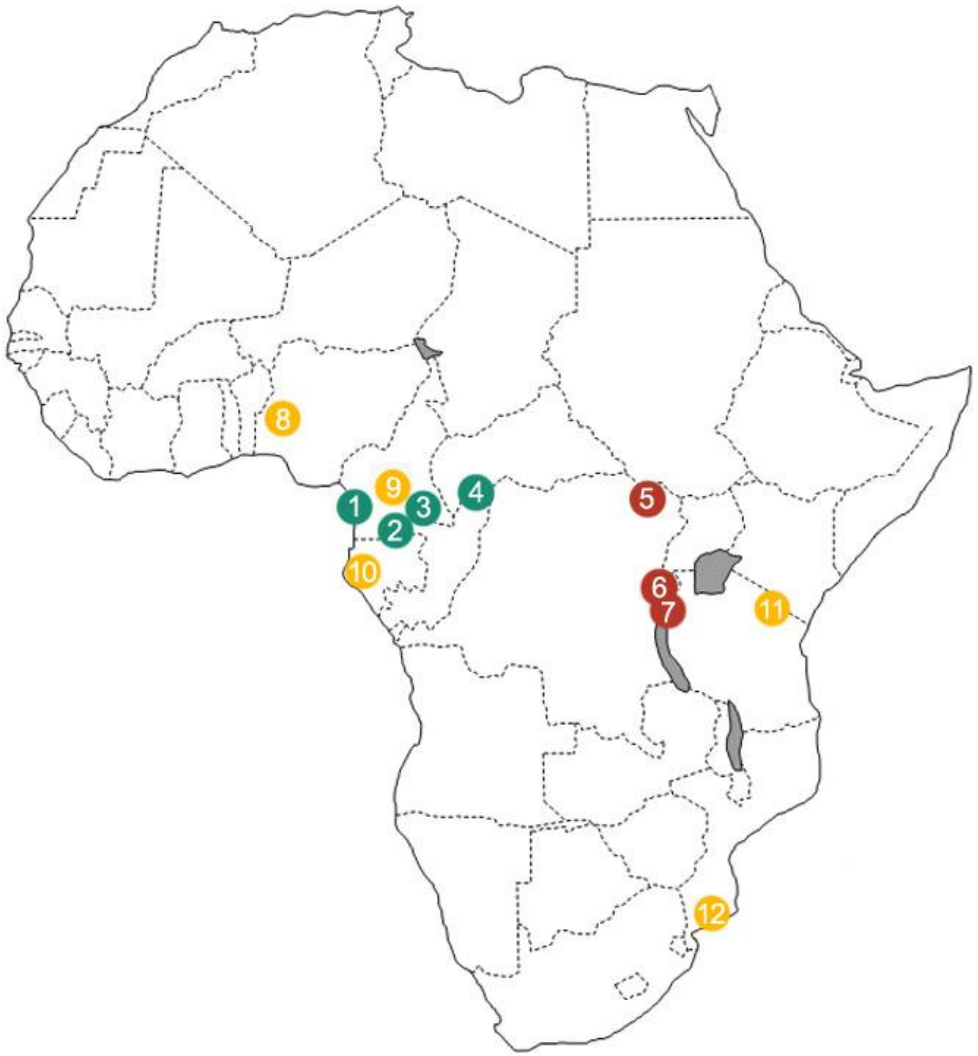
Our model: Pygmy hunter-gatherers and Bantu farmers



Bantu village

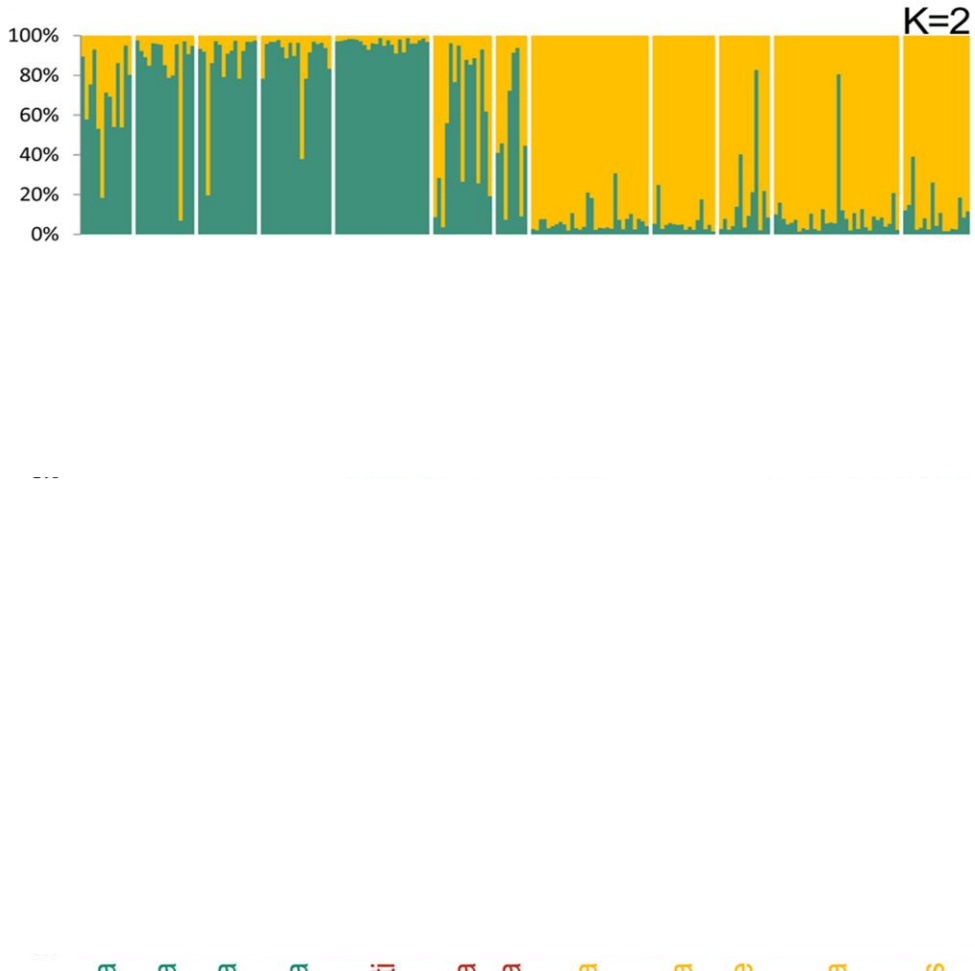
Pygmy settlement in forest

Population collection



Pygmies (PYG)		**118**
1.	Bakola (Cameroon)	16
2.	Baka (Cameroon)	15
3.	Baka (Gabon)	16
4.	Biaka (RCA)	24
5.	Mbuti (RDC)	24
6.	Twa (Rwanda north)	15
7.	Twa (Rwanda south)	8
Farmers (AGR)		**118**
8.	Yoruba (Nigeria)	31
9.	Akele (Gabon)	16
10.	Ngumba (Cameroon)	16
11.	Chagga (Tanzania)	32
12.	Mozambicans	23
		236

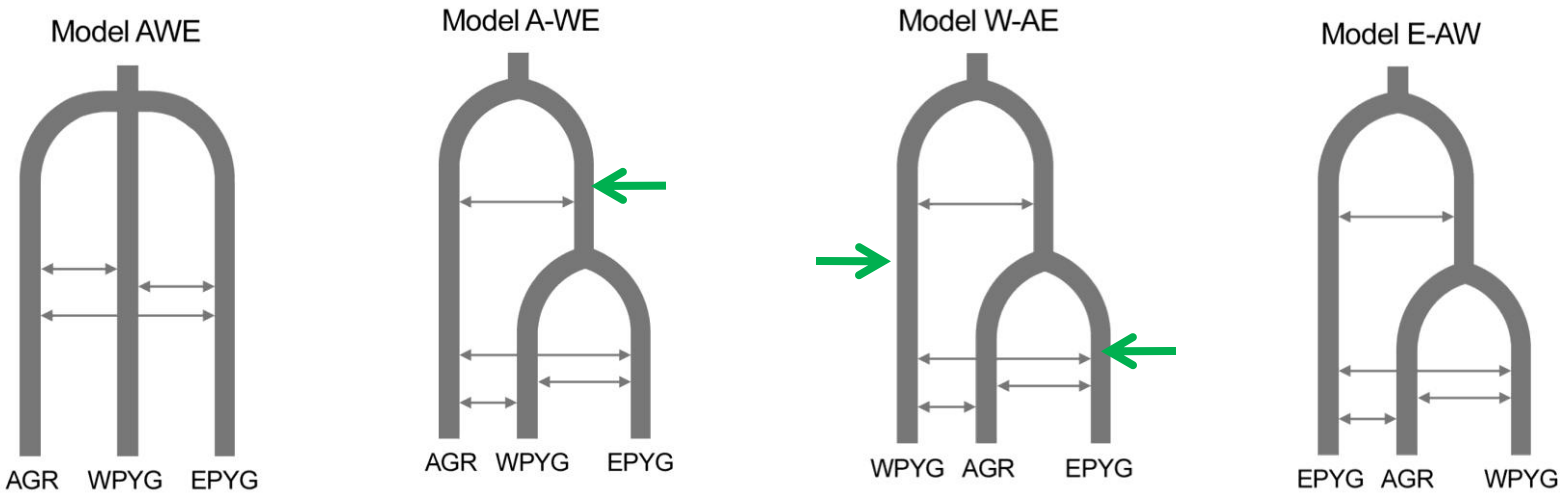
Population structure



Patin et al. ***PLoS Genetics*** 2009

Our model: Pygmy hunter-gatherers and Bantu farmers

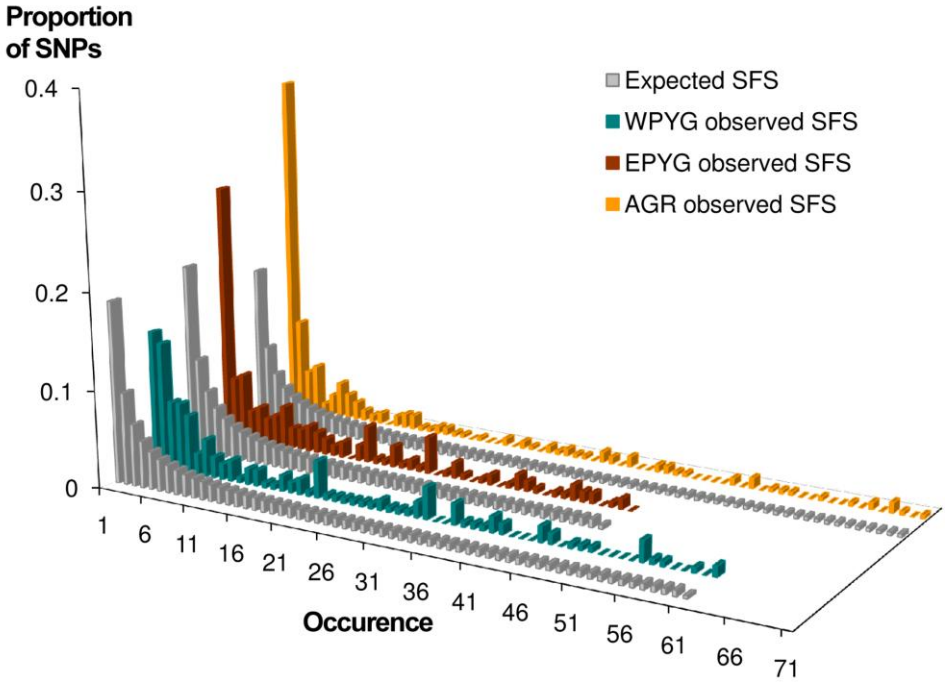
1. When AGR, WPYG and EPYG diverged ?

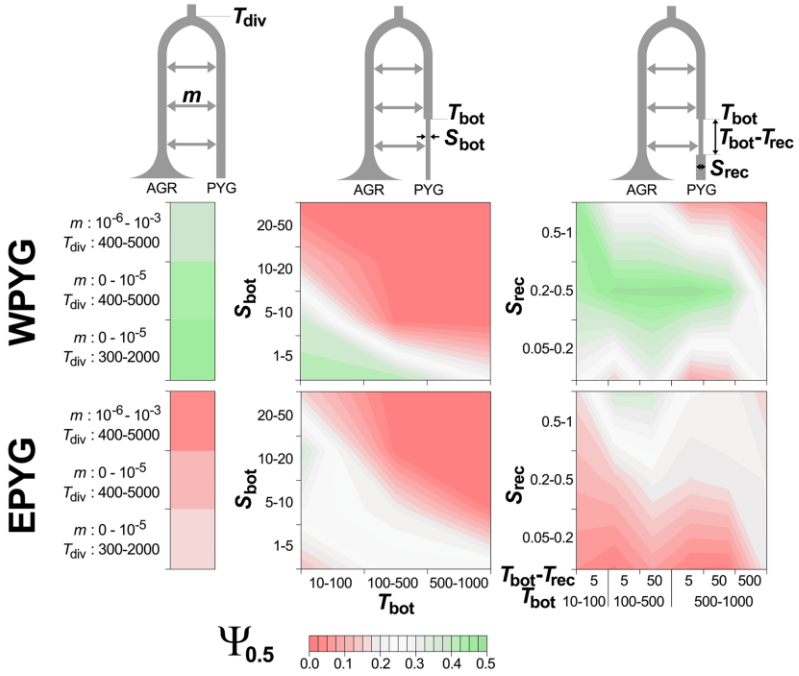


- Have the pygmy traits originated (→) only once or do they have appeared independently (convergent evolution)??

- Has the transition to agriculture involved the divergence between PYG and AGR ?

Demographic regimes of each population groups



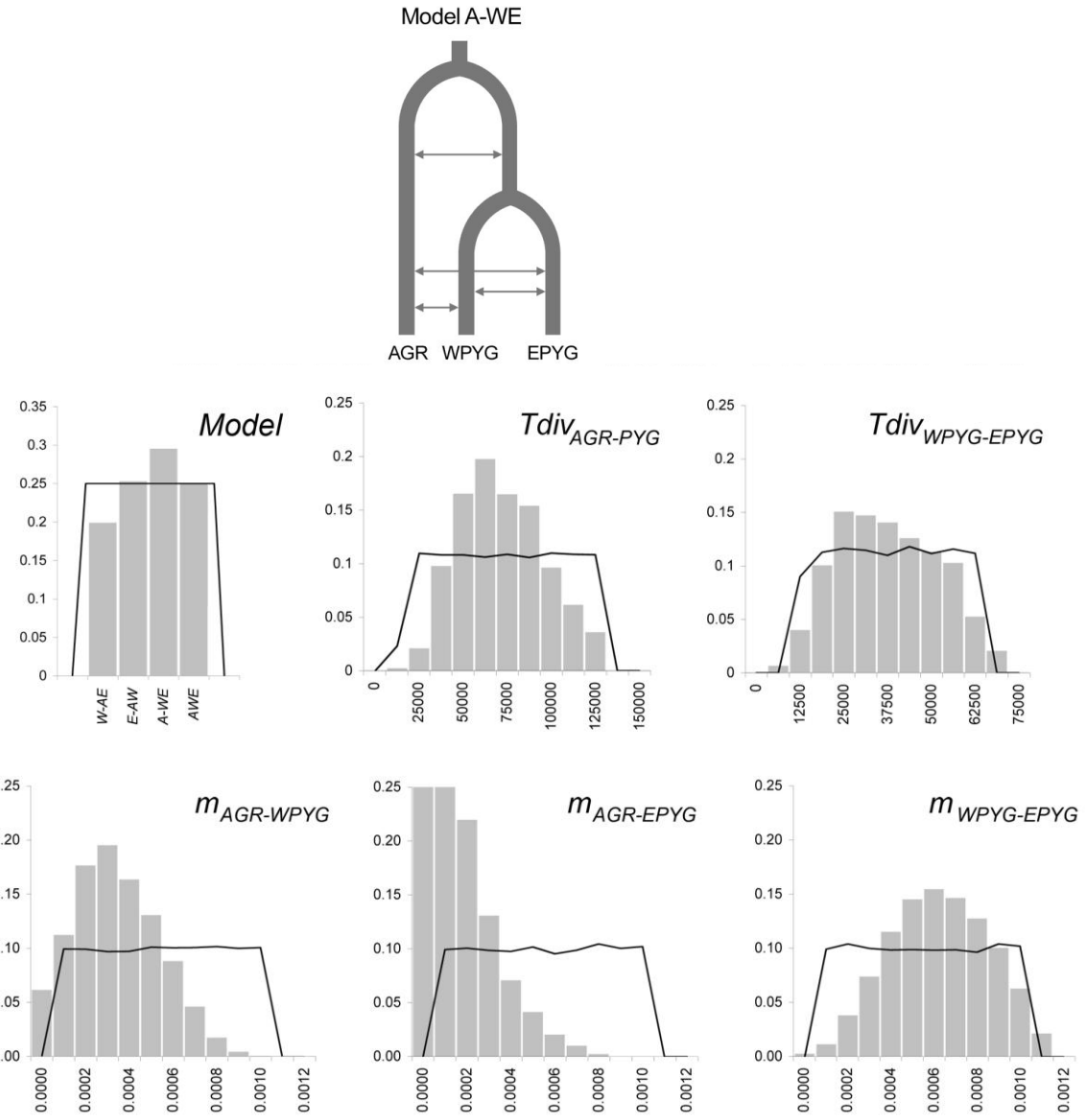


- Farmer populations have experienced population expansions

- Both Western and Eastern Pygmies have undergone reductions in population size (bottlenecks followed by recovery)

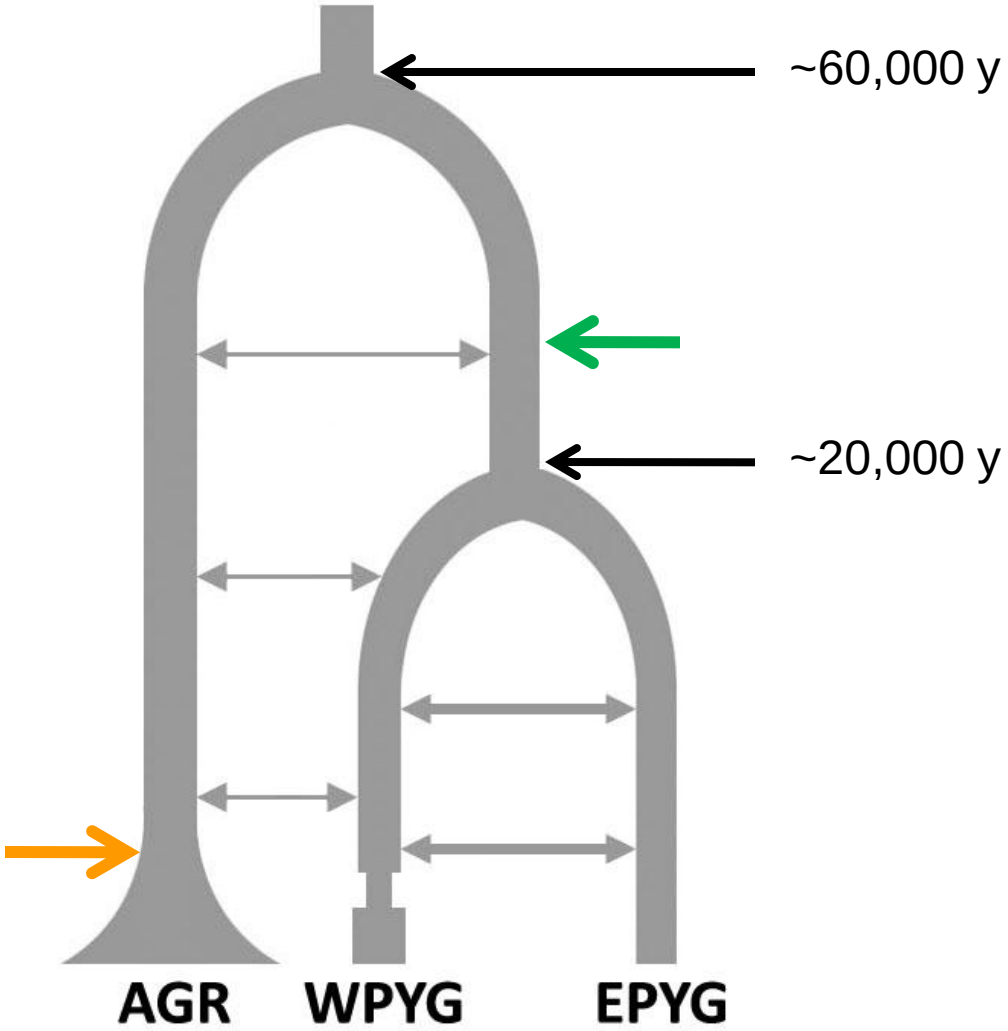
Patin et al. ***PLoS Genetics*** 2009

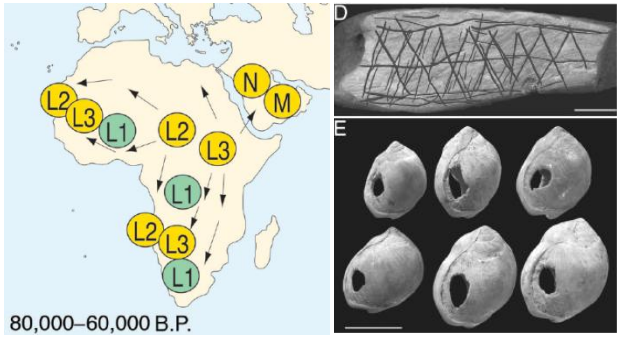
A best-fit model for the demographic history of PYG and AGR



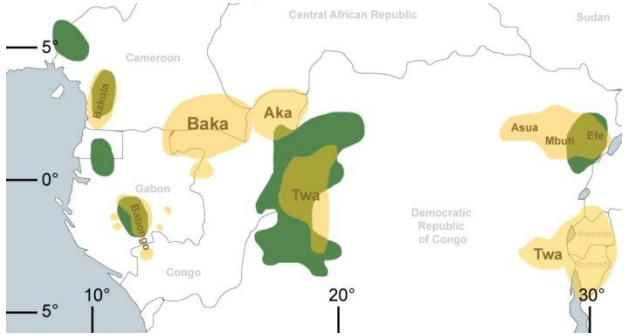
Patin et al. ***PLoS Genetics*** 2009

Implications for the history of hunter-gatheres and farmers in Africa





Mellars P., *PNAS* 2006

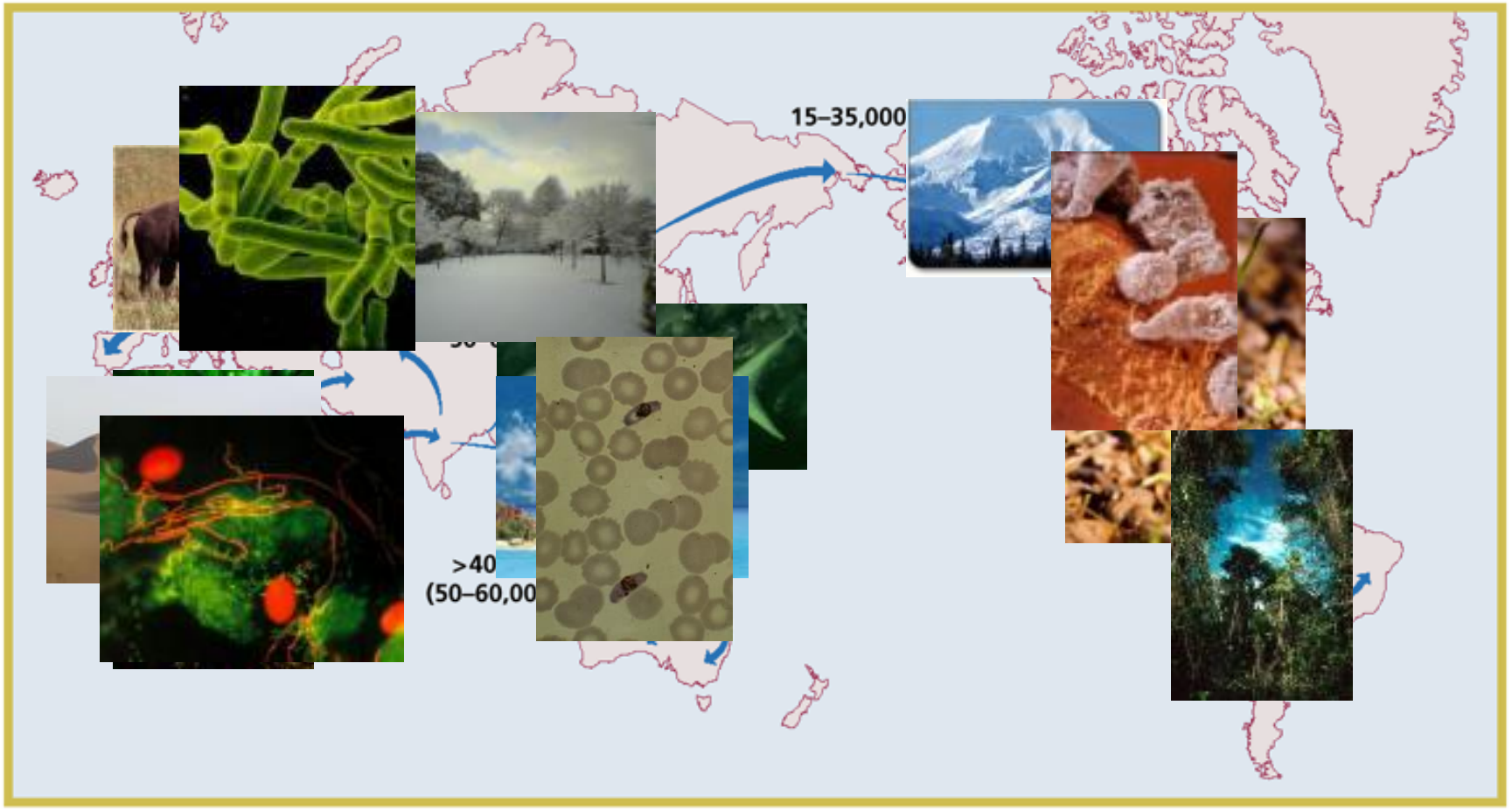


From S. Bahuchet, 1996

Examples

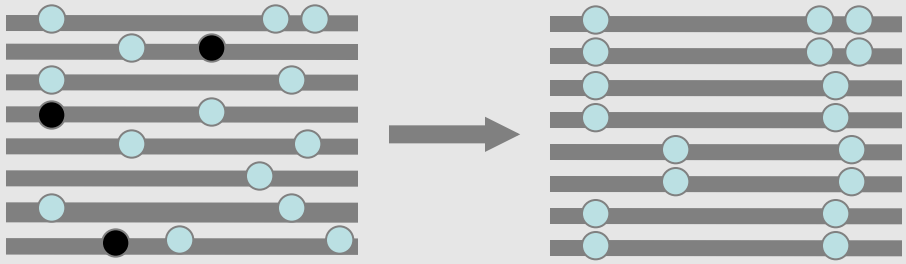
- Demographic Inference
  - Worldwide inference of past demographic events
  - History, population history and lifestyle
- Selective Inference
  - Genomewide detection of natural selection
  - Natural selection and immunity to infection

Migration, adaptation and natural selection



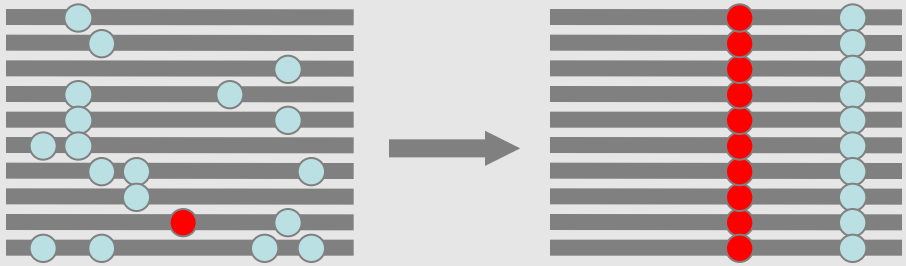
Genetic variants conferring an advantage to better adaptation will be selected

Different types of natural selection



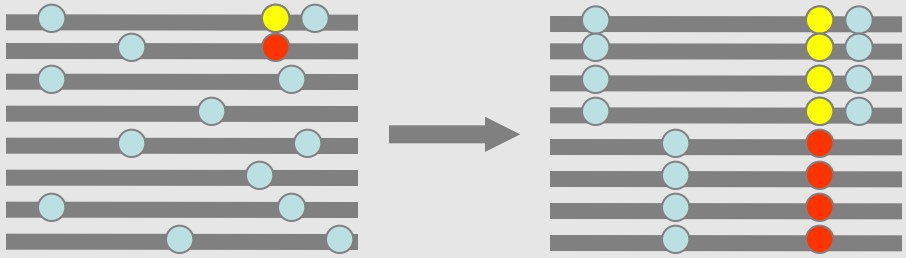
PURIFYING/NEGATIVE SELECTION

Ex. many human genes



POSITIVE SELECTION

Ex. G6PD, CD40 protection against malaria in Africa



BALANCING SELECTION

Ex. MHC worldwide, HbS in Africa (malaria)

Advantageous mutation

Deleterious mutation

"Balanced" mutation

Neutral mutation

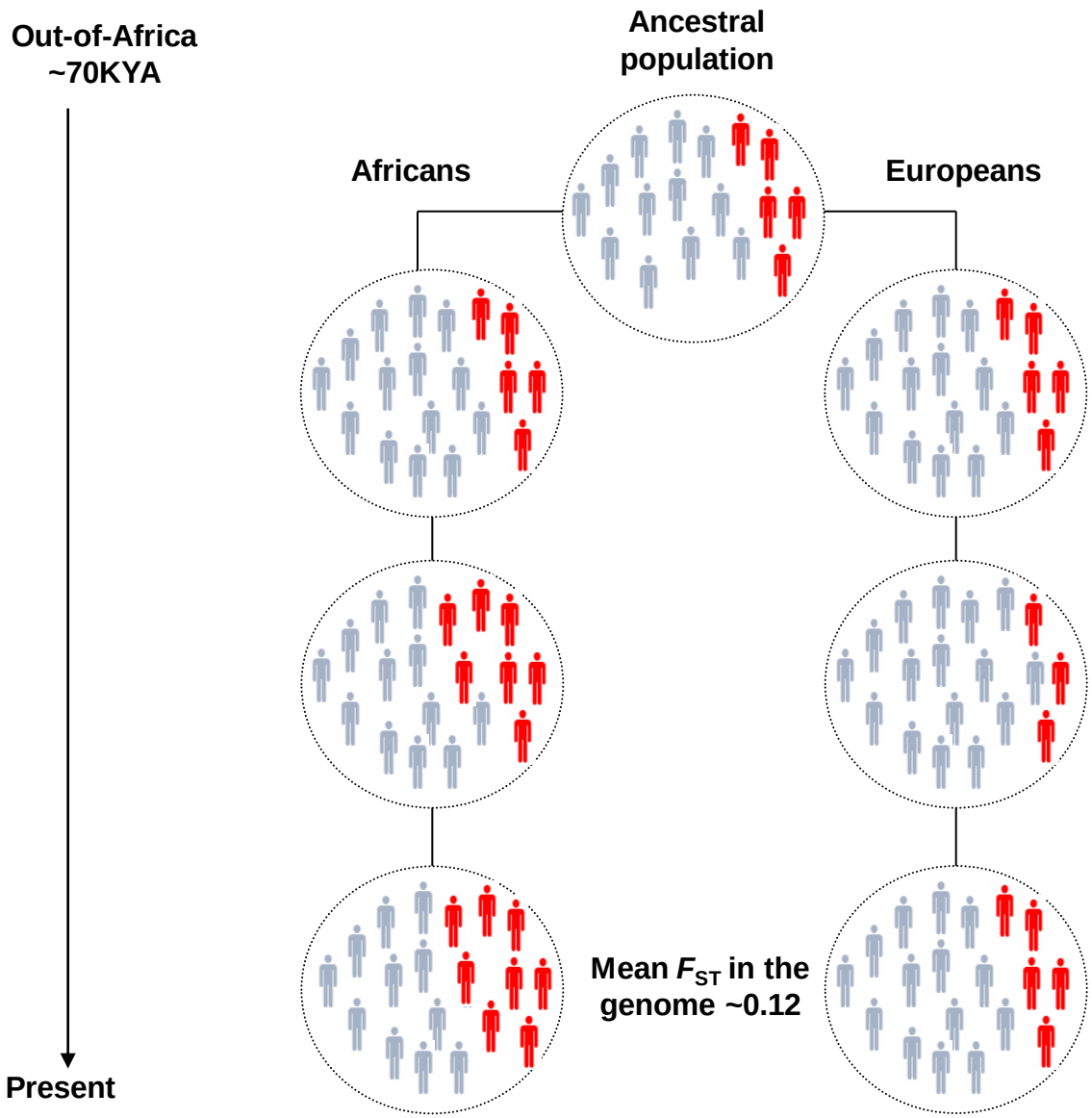
Selection at the Genome-wide Level?

- To which extent has natural selection influenced, at the scale of the entire genome, the degree of population differentiation in modern humans?

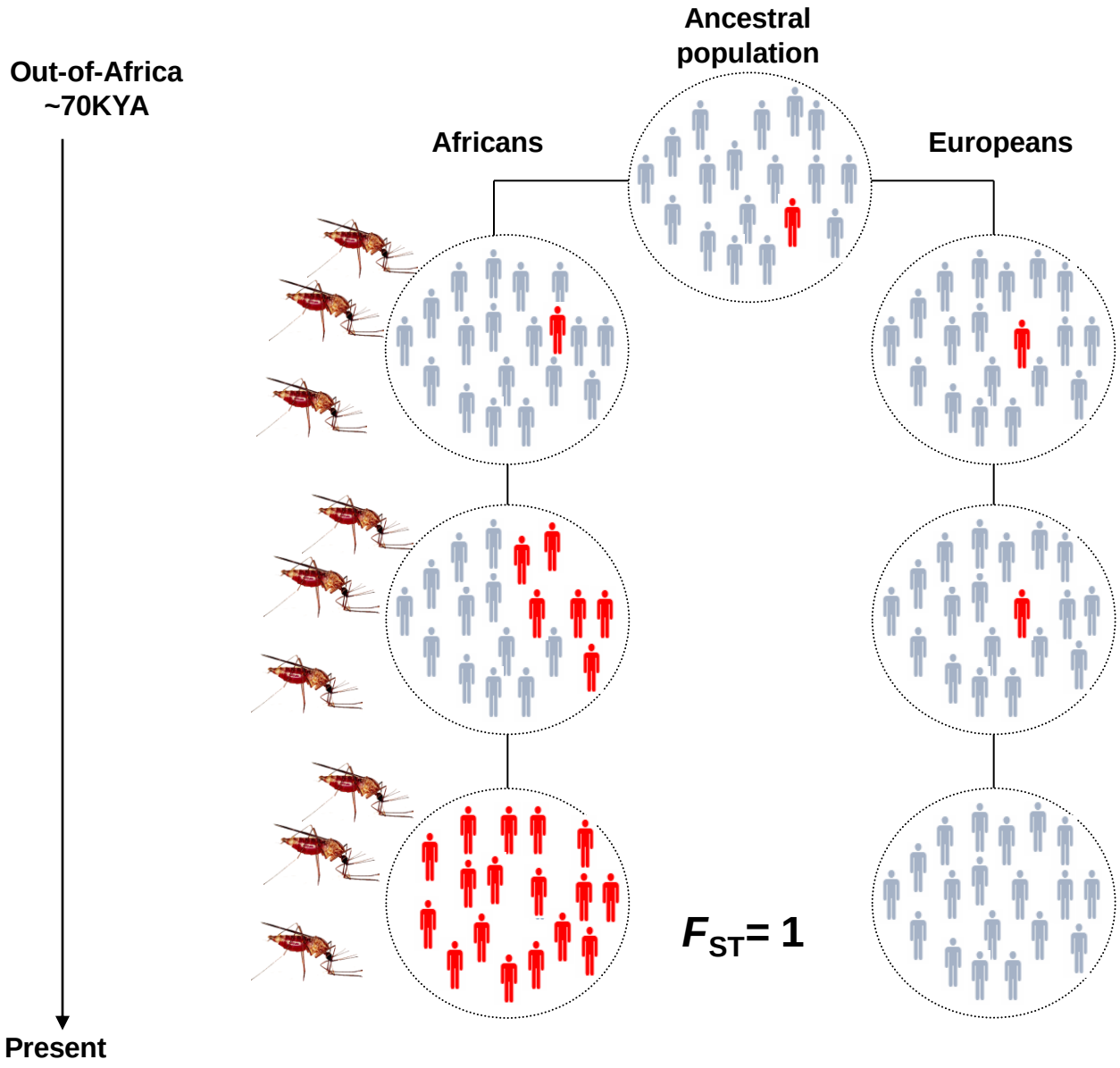
- Which type of genetic variants have been preferentially targeted by selection?

- Genes and gene variants under strong selective pressures can highlight regions of the genome explaining the current population phenotypic variation

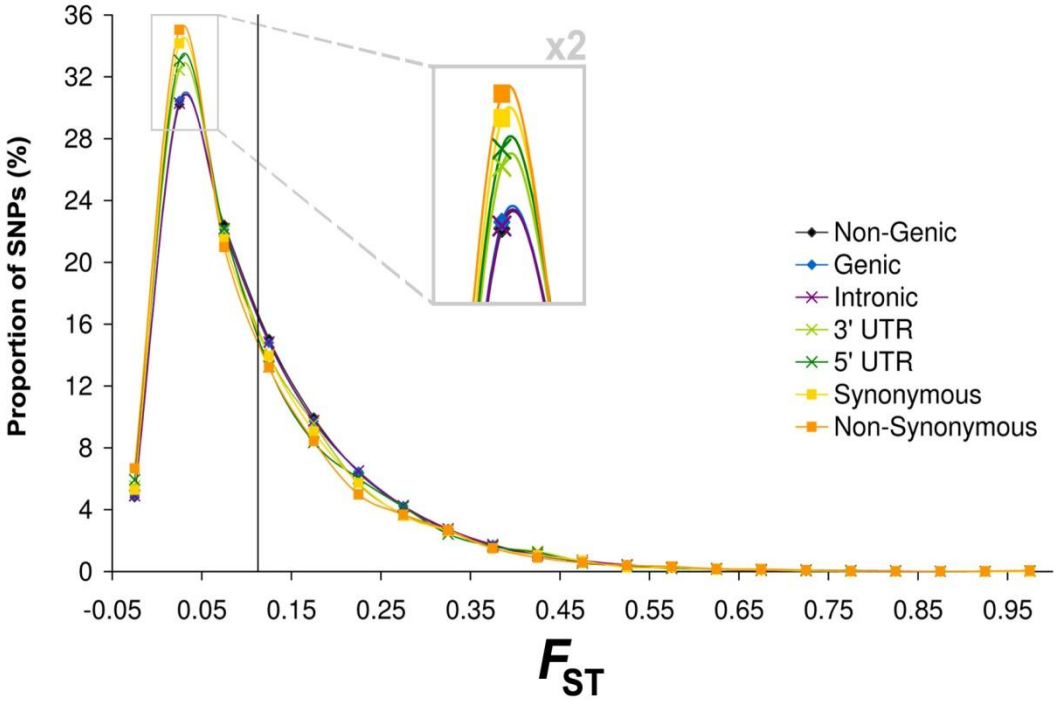
Levels of population differentiation under neutrality

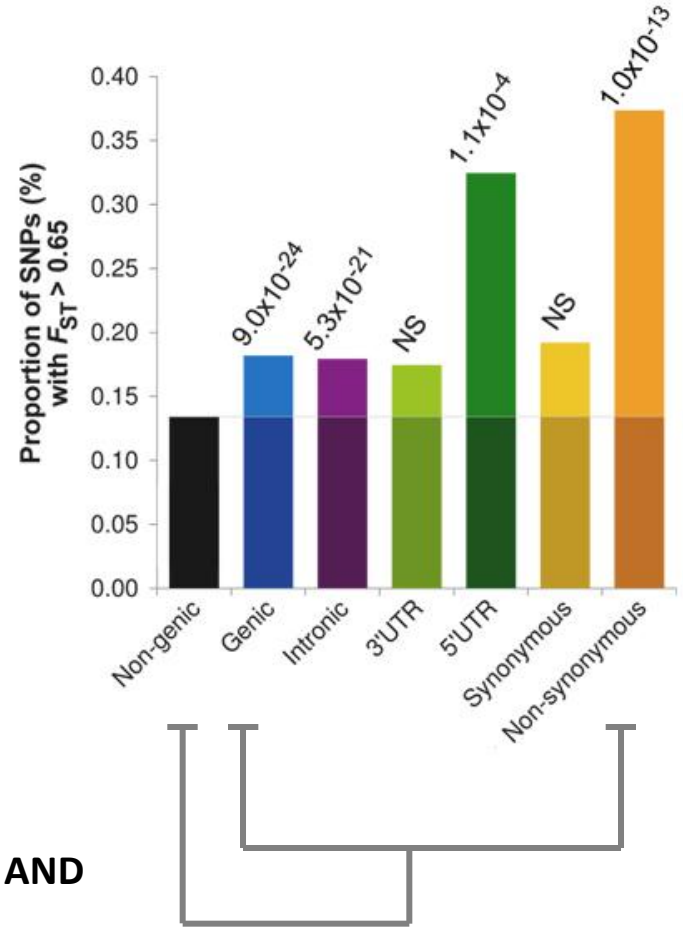


Positive selection increase population differentiation



F_{ST} estimation of 2.8 million HapMap SNPs





POSITIVE SELECTION HAS TARGETED AMINOACID ALTERING AND VARIANTS LOCATED IN 5' UTR

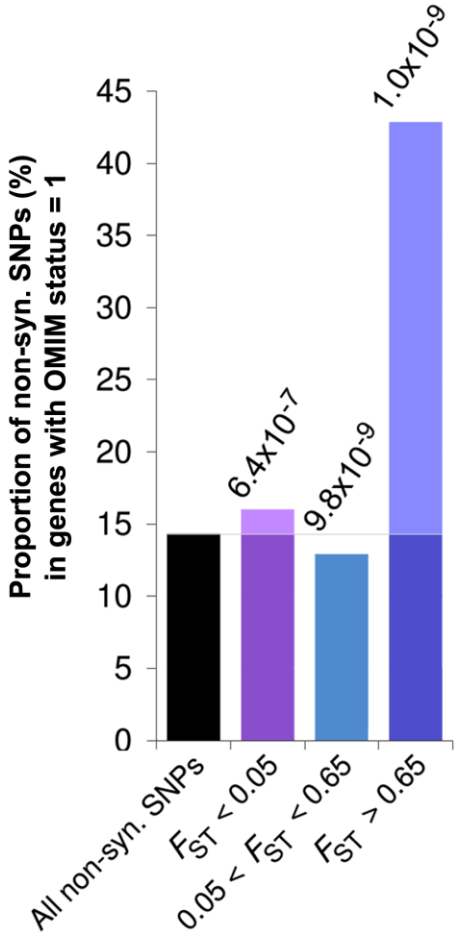
Barreiro et al. ***Nature Genetics*** 2008

Conservative list of targets of positive selection
Extended list of 585 genes

Phenotype category	Genes under strong positive selection
Morphological traits (skin pigmentation, hair development, etc.)	*ABCC11*, ***EDAR***, *MATP, PKP1, PLEKHA4, SLC24A5*
Immune response to pathogens	*CEACAM1, CR1,* ***DUOX2***, *VAV2*
DNA repair and replication	*MPG, POLG2, TDP1*
Sensory functions (olfaction, eye development…)	*COL18A1, OR52K2, RP1L1*
Insulin regulation, metabolic syndrome (obesity, diabetes, hypertension)	*ALMS1, CEACAM1, ENPP1*
Various metabolic pathways (ethanol, intestinal zinc, citrulline…)	***ADH1B***, *ASS, SLC39A4*
Miscellaneous	*SPAG6, FBX031, RTTN*
Unknown	*ABCC12,* ***ADAT1***, *AK127117, C17orf46, C8orf14, COLEC11,* ***CPSF3L***, *DNAJC5B, DNHD1, ETFDH, EXOC5,* ***FAIM***, *FLJ37464, FXR1, GCN5L2, KIAA0984, LAMB4, LOC648511,* ***MGC72127***, ***PCGF1***[1], *PLEKHG4, POL3S*[2], ***RNF135***, ***SLC30A9***, *SYTL3, TEX15,* ***TTC31***[1], ***FLJ14397***[1], *VPS33B, ZNF646*[2]

Barreiro et al. (2008) **Nat Genet**

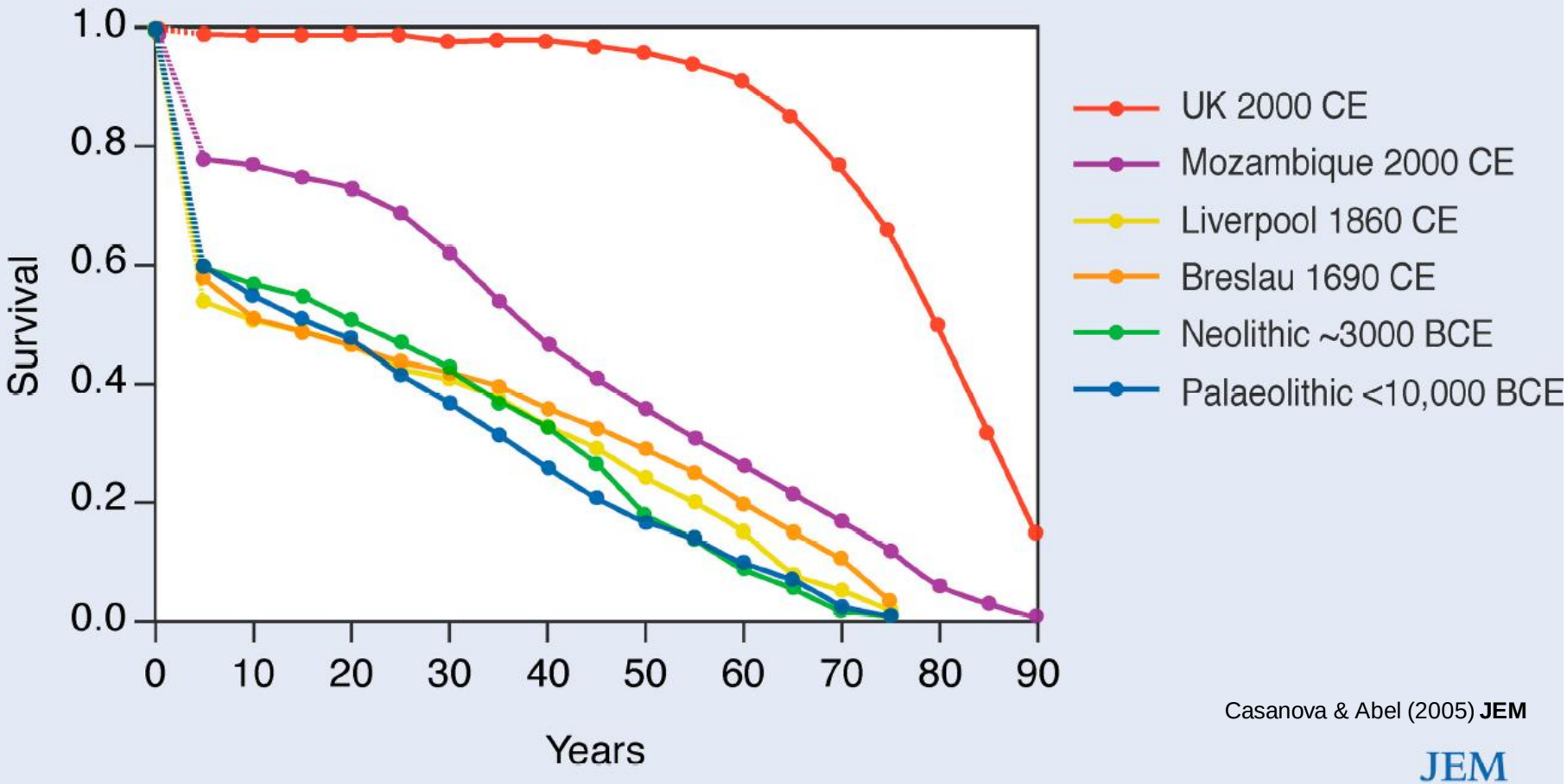
Natural selection and disease



Barreiro et al., ***Nature Genetics*** 2008

Non-synonymous variants that are highly differentiated among populations are more represented among genes known to be involved in human disease

Infectious diseases have imposed a tremendous selective pressure

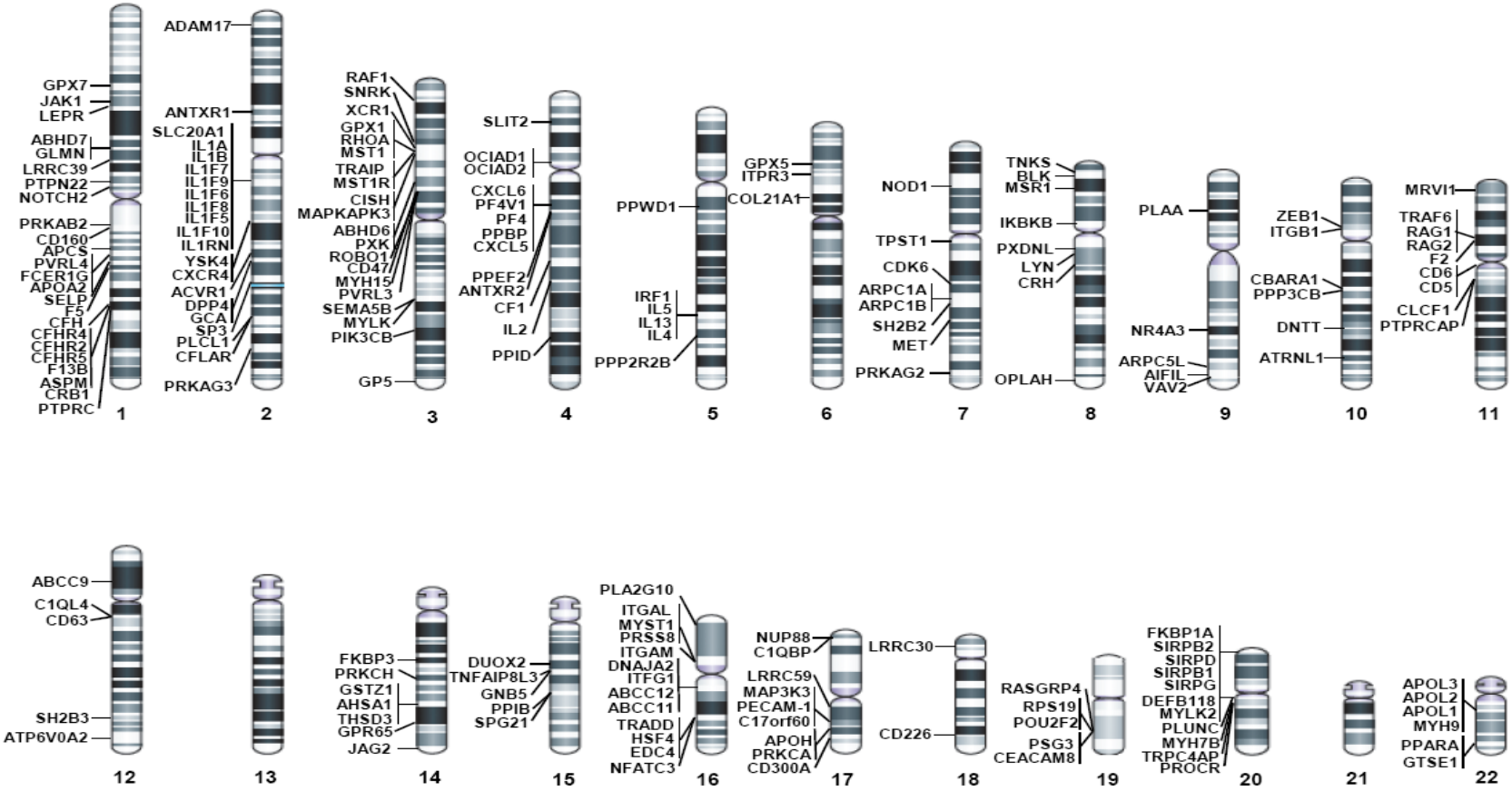


Casanova & Abel (2005) **JEM**

JEM

Infectious diseases have been a major force driving the evolution of our genomes

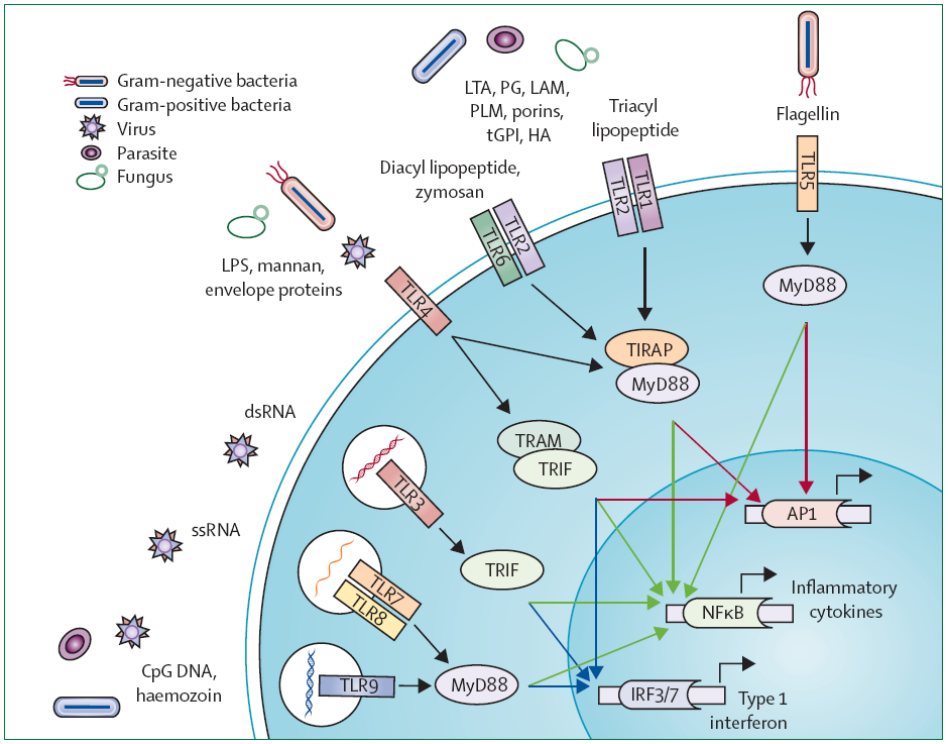
Whole genome approaches (GWSS)



Immunity-related genes are over-represented when restricting the analyses to more recent selective events (< 30,000 years)

Barreiro & Quintana-Murci ***Nat Rev Genet*** 2009

The human Toll-like receptors



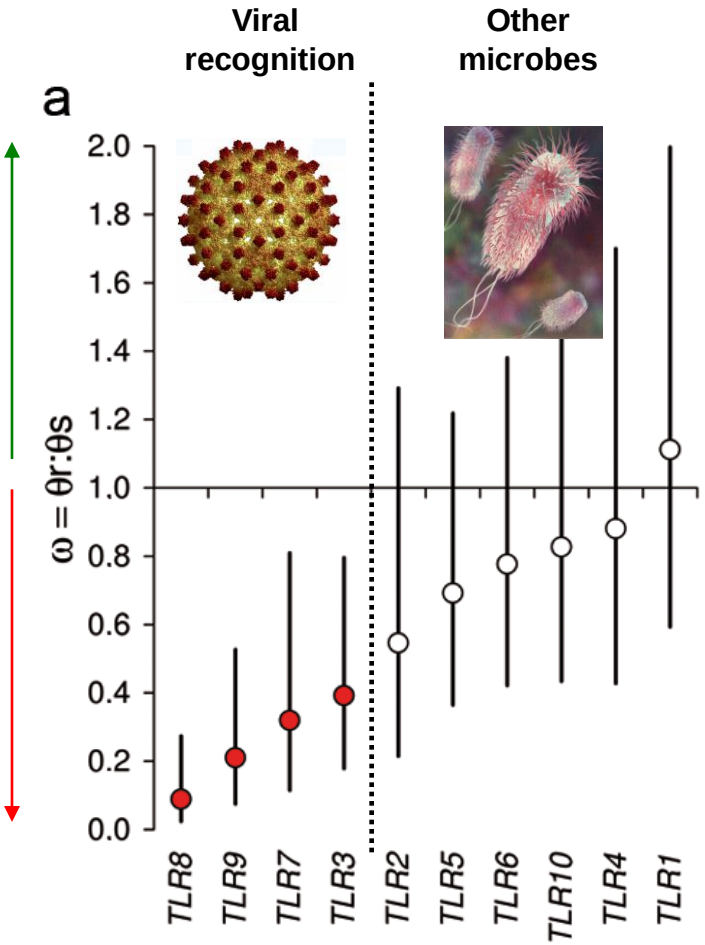
- Determine whether and how natural selection has acted upon the individual TLR members

- Sequence of the ten TLRs in a multi-ethnic panel of individuals worldwide

The effects of purifying selection on TLRs

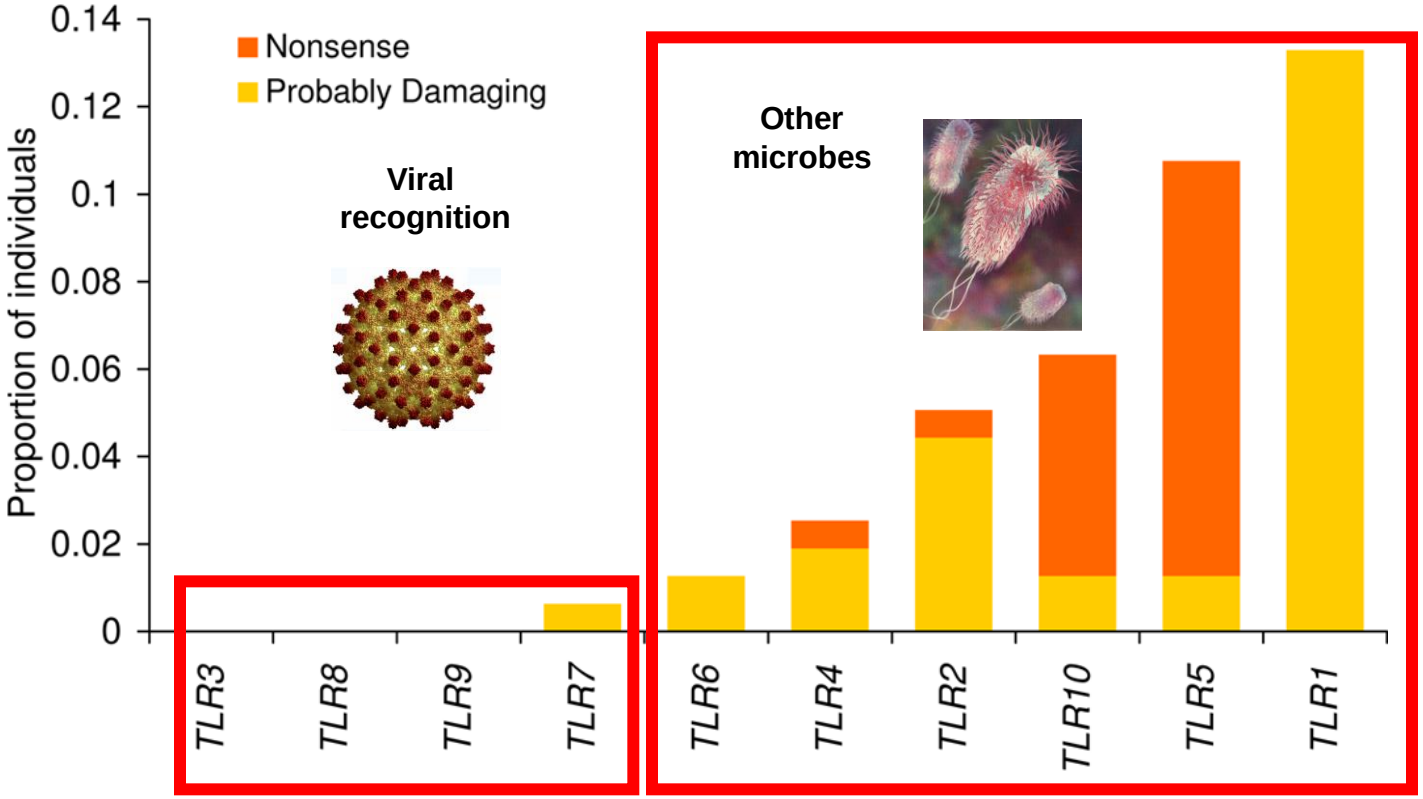
Positive selection

Purifying selection



Barreiro et al., *PLoS Genetics* 2009

Functional diversity at the individual TLR level



The high frequencies of damaging and nonsense mutations in “non viral” TLRs illustrate their greater redundant role in protective immunity

Conclusions

- **TLRs sensing virus are under strong purifying selection, indicating strong evolutionary constraints**
- **Virus have exerted strong selective pressures than bacteria and parasites on the TLR system**
- **Non-viral TLRs are more “dispensable” in host defense**
- **Some of them, TLR5, are completely redundant in host defense**

Some ongoing research

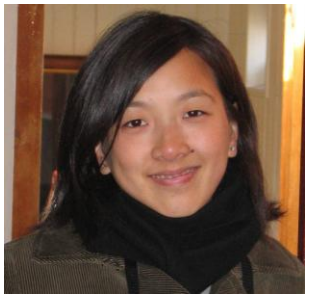
- **Genome-wide**:
  - Demographic inferences using genomewide data and complete human genome sequences
  - Detection of the (difficult) signatures of balancing selection

- **Immunity and infectious disease**:
    - The Nod-like receptor family
    - RIG-I like receptor family
    - The C-type lectin family
    - The IFN gene family

People involved in these projects, collaborations, funding

Lluis Quintana-Murci

Guillaume Laval



Helene Quach

Luis Barreiro

Etienne Patin

Meriem Ben Ali

National Collaborations

Evelyne Heyer	MNHN, Paris
Alain Froment	MNHN, Paris
Olivier Neyrolles	Toulouse
J-M Hombert	ISH, Lyon
JL Casanova	Necker, Paris
Laurent Abel	Necker, Paris
François Renaud	IRD, Montpellier
Antoine Gessain	Pasteur

International Collaborations

Ken Kidd	Yale Univ, USA
Chris Tyler-Smith	Sanger Center, UK
Rosalind Harding	Oxford Univ, UK
Pardis Sabeti	MIT, USA
Doron Behar	Technion, Israel
Antti Sajantila	Helsinki, Finland
Eileen van Helden	Stellenbosch, SA