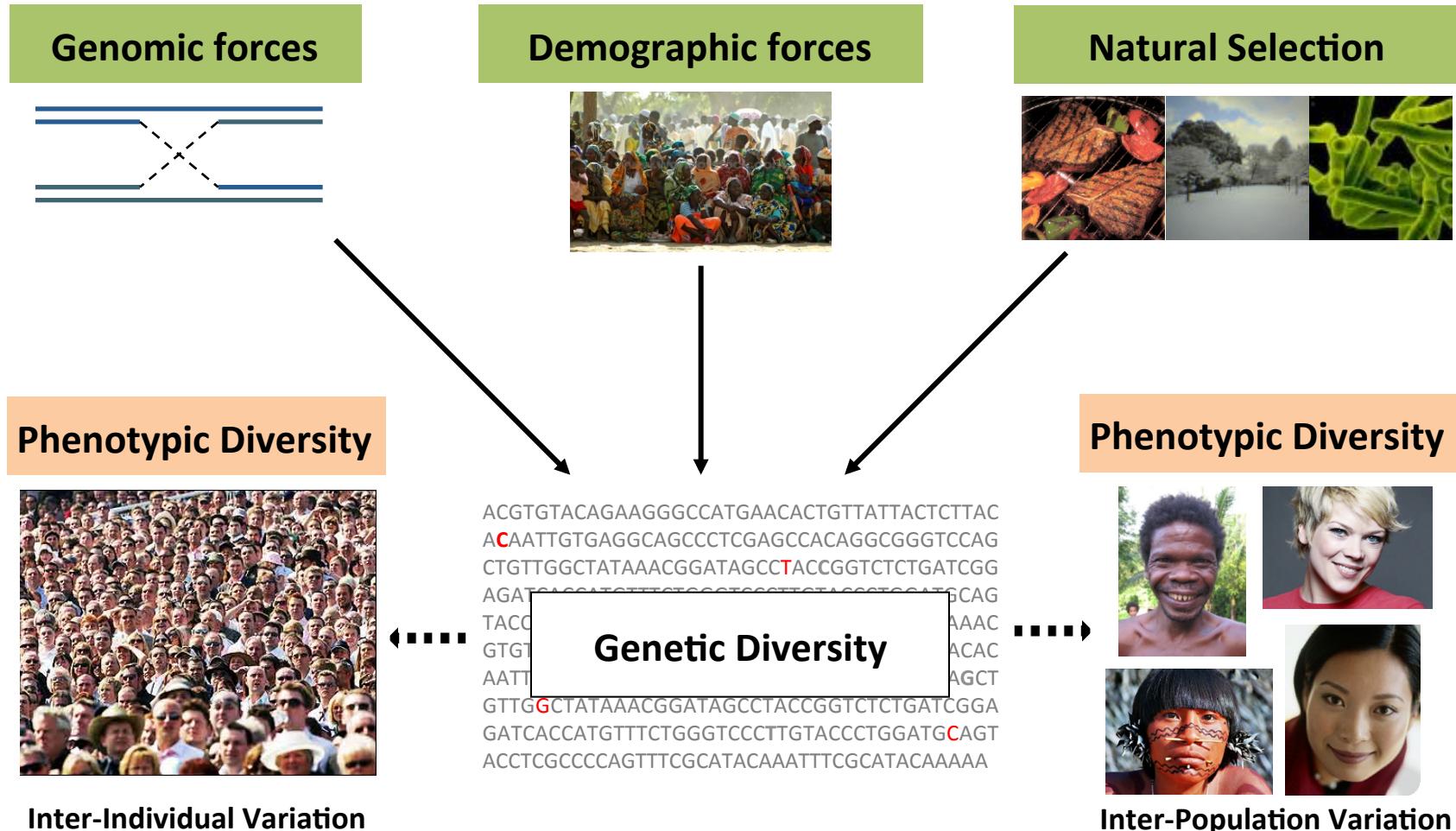


# Genomics: Human Evolutionary Perspectives

Maud Fagny, PhD  
Quackenbush lab

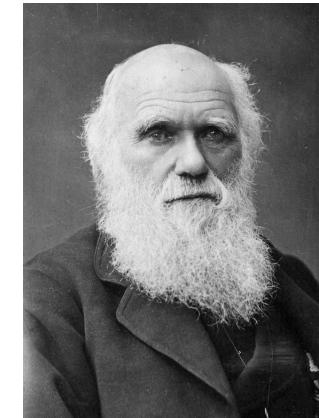
# Evolution: studying the diversity of life



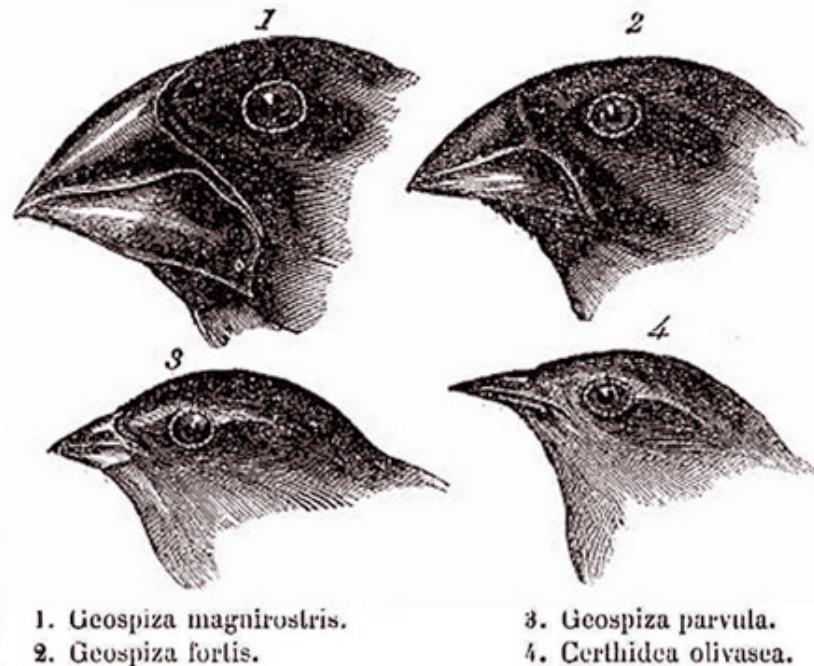
# Evolutionary biology: a short history

## The theory of evolution by natural selection - Charles Darwin (1809-1882)

- Based on his observation of wildlife during his 5-years voyage on the HMS Beagle.
- Publish his theory in 1859 in his book *On the origin of species*.



**Theory:** "As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequently recurring struggle for existence, it follows that any being, if it vary however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be naturally selected."

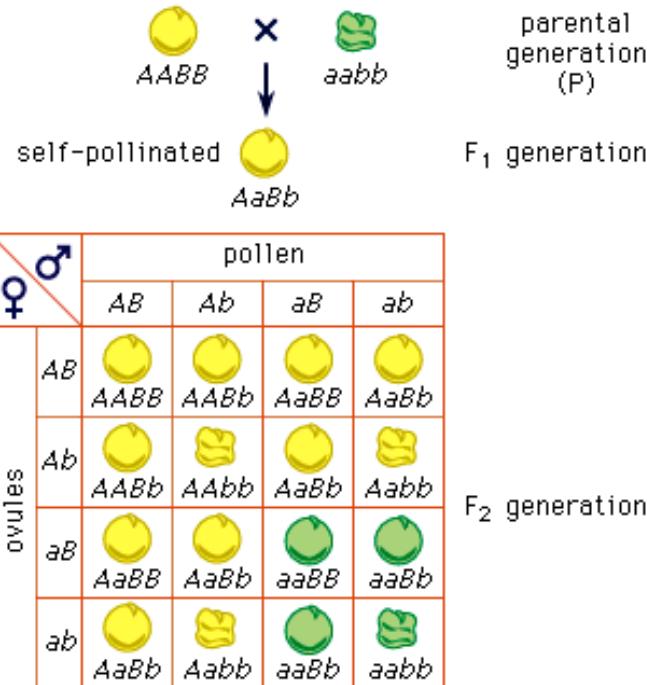


Darwin's drawings of Galapagos finches

# Evolutionary biology: a short history

## Mendel's laws on inheritance – Gregor Mendel (1822-1884)

- Law of dominance
- Law of segregation
- Law of independent assortment



© 2006 Encyclopædia Britannica, Inc.

## Chromosome theory of inheritance - Walter Sutton (1877-1916) and Theodor Boveri (1862-1915):

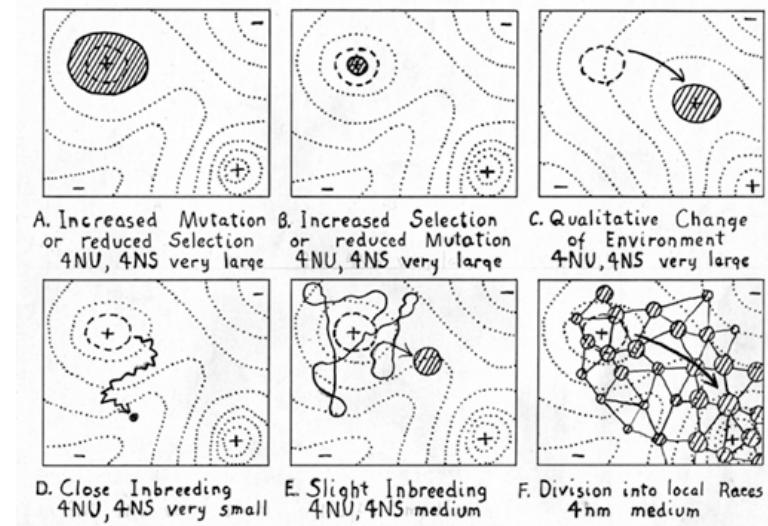
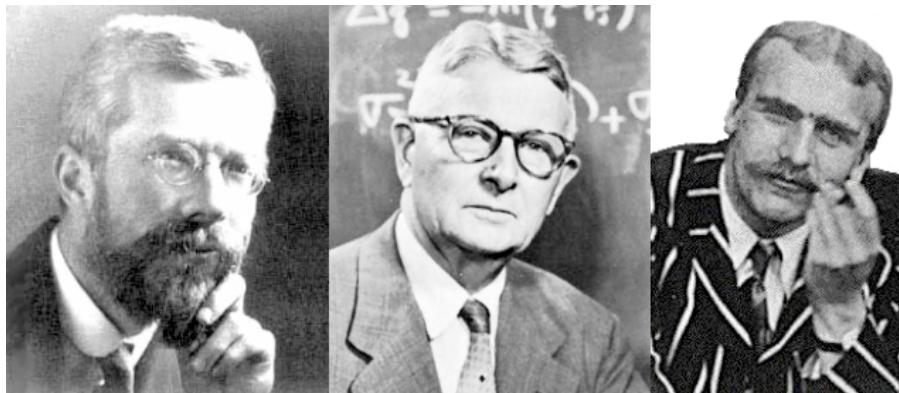
- Identify chromosomes as "the physical basis of Mendelian inheritance".



# Evolutionary biology: a short history

**Modern evolutionary synthesis - Ronald Fisher (1890-1962), John Haldane (1892-1964) and Sewall Wright (1889-1988):**

- Founders of modern population genetics and evolutionary biology
- Reconcile Darwin's theory of gradual evolution by natural selection and Mendelian genetics:
  - Show that the action of several independent genetic loci could lead to continuous phenotypic variations.
  - Analyze natural selection examples using mathematical models
  - Introduce the notion of adaptive landscape and genetic drift

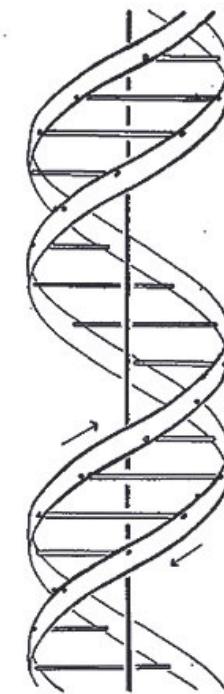
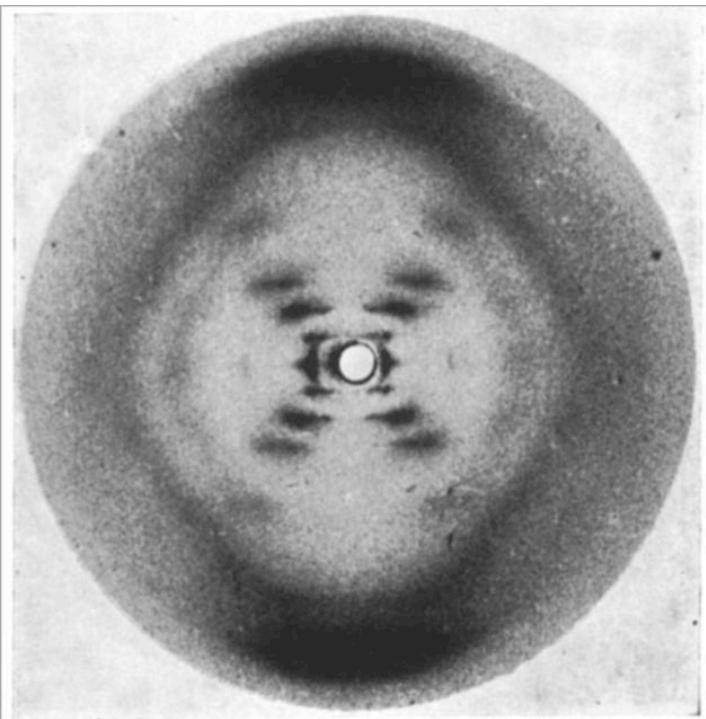


Wright, 6<sup>th</sup> International Congress of Genetics (1932)

# Evolutionary biology: a short history

**Discovery of DNA double helix: James Watson (1928-), Francis Crick (1916-2004) and Rosalind Franklin (1920-1958):**

- Explain the mechanism of base-pairing.
- Allow to understand how genetic information is stored and replicated during cell division.



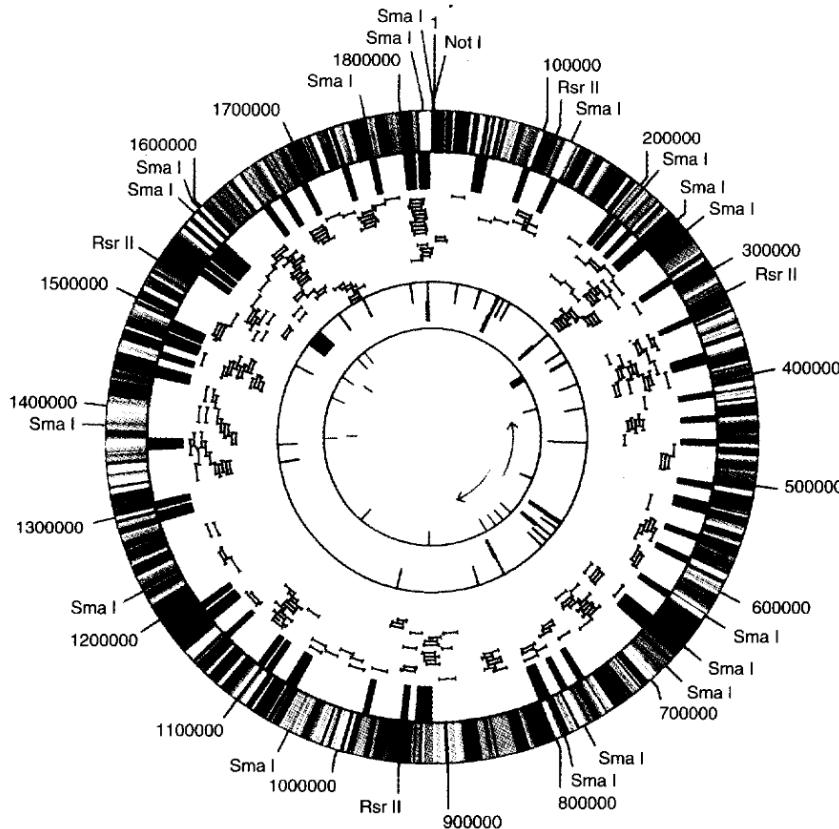
Franklin & Gosling, *Nature* (1953)

Watson & Crick, *Nature* (1953)

# Evolutionary biology: a short history

The *Haemophilus influenza* genome - Fleischman et al., *Science*, 1995:

- First non-viral genome sequenced
- Open the era of Comparative Genomics



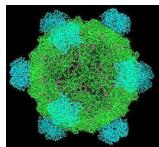
[www.nhs.uk/conditions/Hib/Pages/Introduction.aspx](http://www.nhs.uk/conditions/Hib/Pages/Introduction.aspx)

## Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,\* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghegan, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

# Comparative Genomics

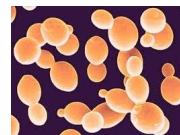
1977 - *ΦX174*



1995 - *Haemophilus influenza*  
(First bacteria)



1996 - *Saccharomyces cerevisiae*  
(First yeast)



1998 - *Caenorhabditis elegans*  
(First invertebrate)



2000 - *Drosophila melanogaster*



2000 - *Arabidopsis thaliana*  
(First plant)



2001 - *Homo sapiens*  
(First Mammals)



2002 - *Mus musculus*



2002 - *Takifugu rubripes*  
(First fish)



2005 - *Pan troglodytes*



2010 - *Palaeo-Eskimo*  
(First ancient genome)



2010 - *Homo neanderthalensis*  
(First extinct species)

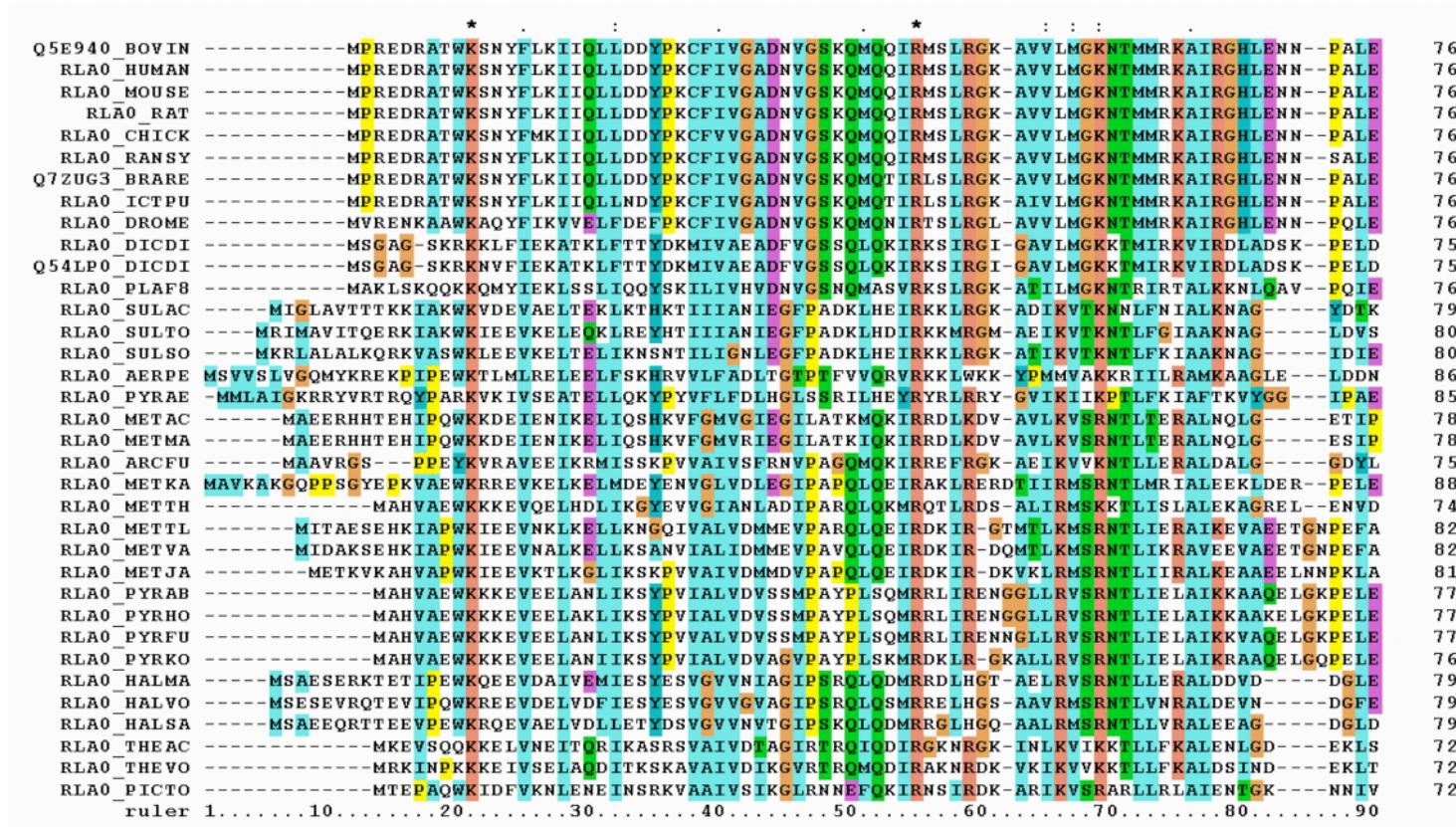


2013 - *Danio rerio*



# Comparative Genomics

- Aligning genomes from different species:



- Where to find DNA alignment tools:
  - <https://omictools.com/whole-genome-alignment-category>
  - [https://en.wikipedia.org/wiki/List\\_of\\_sequence\\_alignment\\_software](https://en.wikipedia.org/wiki/List_of_sequence_alignment_software)

# Comparative Genomics: What can we learn?

Perspective

---

## Comparative genomics as a tool to understand evolution and disease

Jessica Alföldi<sup>1</sup> and Kerstin Lindblad-Toh<sup>1,2,3</sup>

### Annotation of new genomes

- Estimation of number of genes (e.g. in human genome)
- Annotation of regulatory elements

### Comparison of genome structure

- Evolution of chromosomal rearrangements
- Identification of new genes (brain function in humans)

### Evolution at different time-scales by comparing

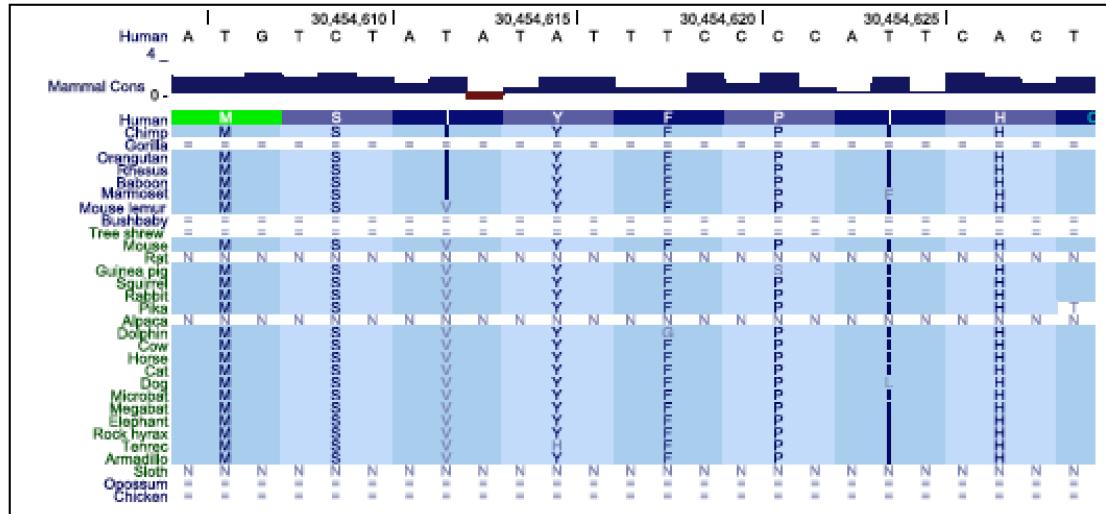
- Different species at different level of the tree of life (eukaryotes, primates)
- Different populations within the same species

### Evolution of specific biological function:

- Species specific phenotypes (language in humans)
- Population specific phenotypes (lactose tolerance in some human populations)

# Comparative Genomics

- Comparing sequences at the global and local level

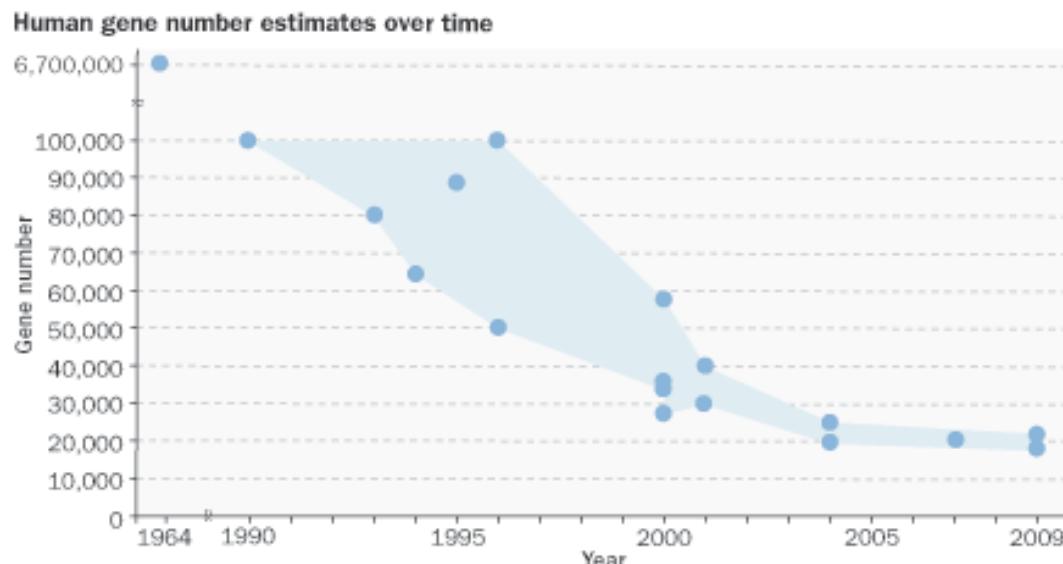


LBH gene – Alf.Idi & Lindblad-Toh, Genome Res. (2016)

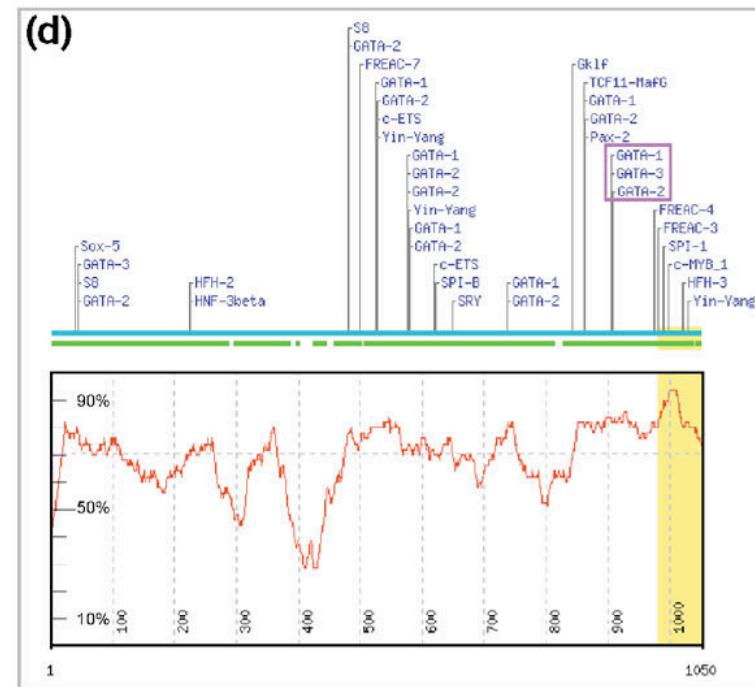
- Gene duplications and loss
- Inference of more accurate phylogenetic trees
- Inference of population demographic history
- Inference of common ancestor DNA sequence
- Evolution of a genomic locus between species (gene, regulatory element)
- Evolution of a genomic locus within species (gene, regulatory element)

# Annotation of genome

## Annotation of human genome: evolution of the number of genes



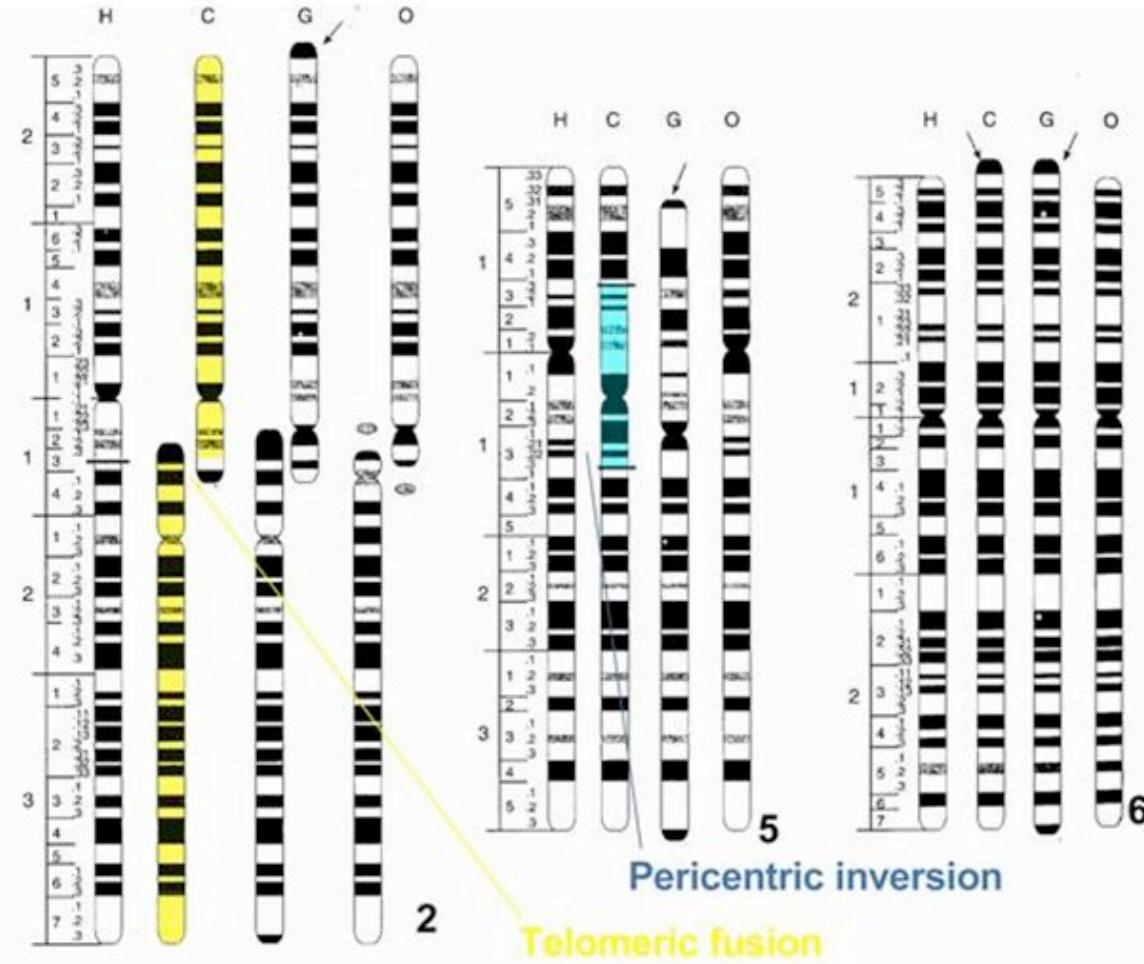
## Identification of regulatory sequence (comparison with cow)



Lenhard et al, & Salzberg, J. of Biology, 2002

- Assessing the total number of genes in the human genome
- Identifying regulatory regions by studying sequence conservation across species.

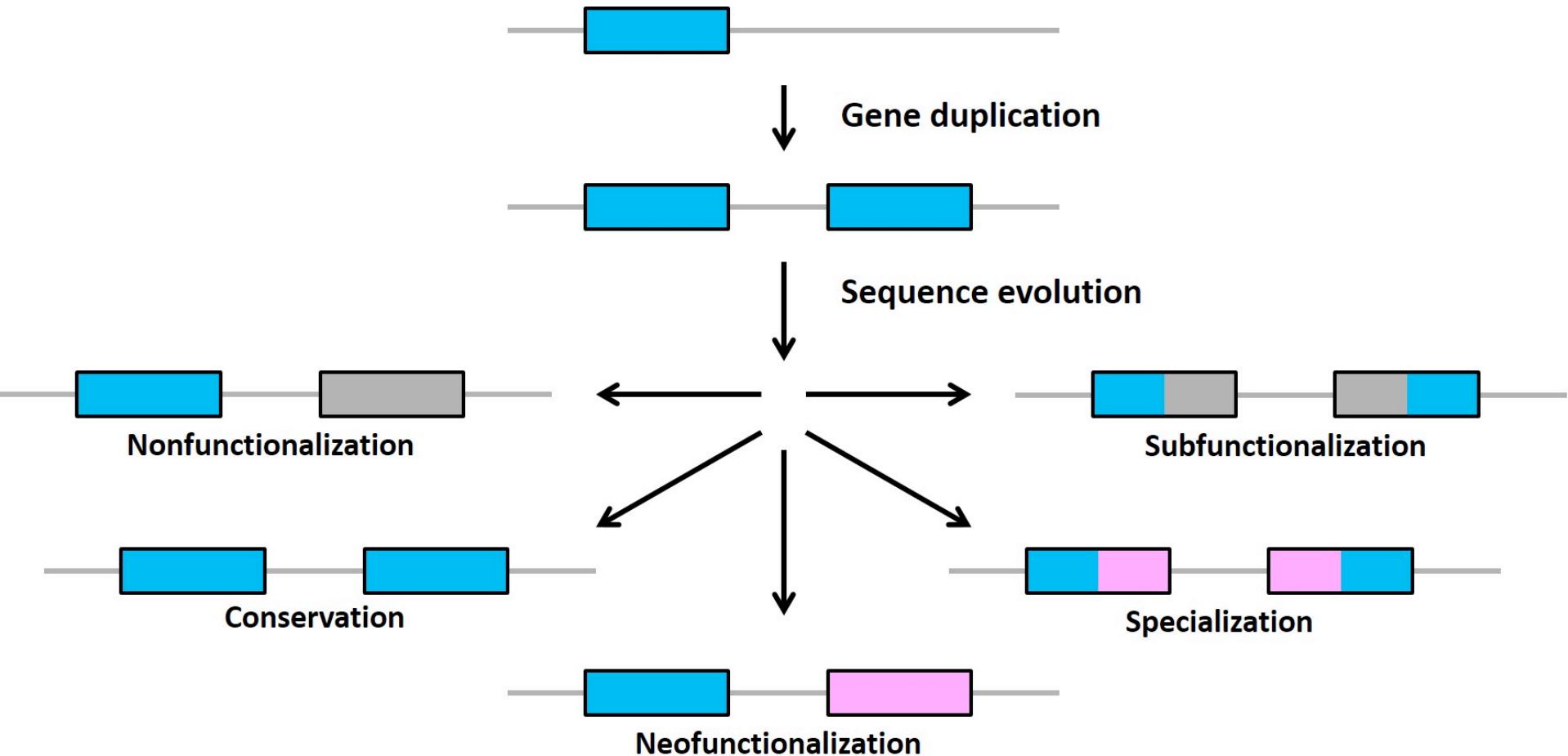
# Chromosomal rearrangement



- A role in the speciation (Ayala & Coluzzi, PNAS, 2005)
- Influence the evolution of species-specific traits? Sequencing revealed that it wasn't the main driver of Human/Chimp phenotypic divergence. (Vallander & Lahn, Genomics, 2004)
- Importance of local duplications and deletions (Cheng et al., Nature, 2005).

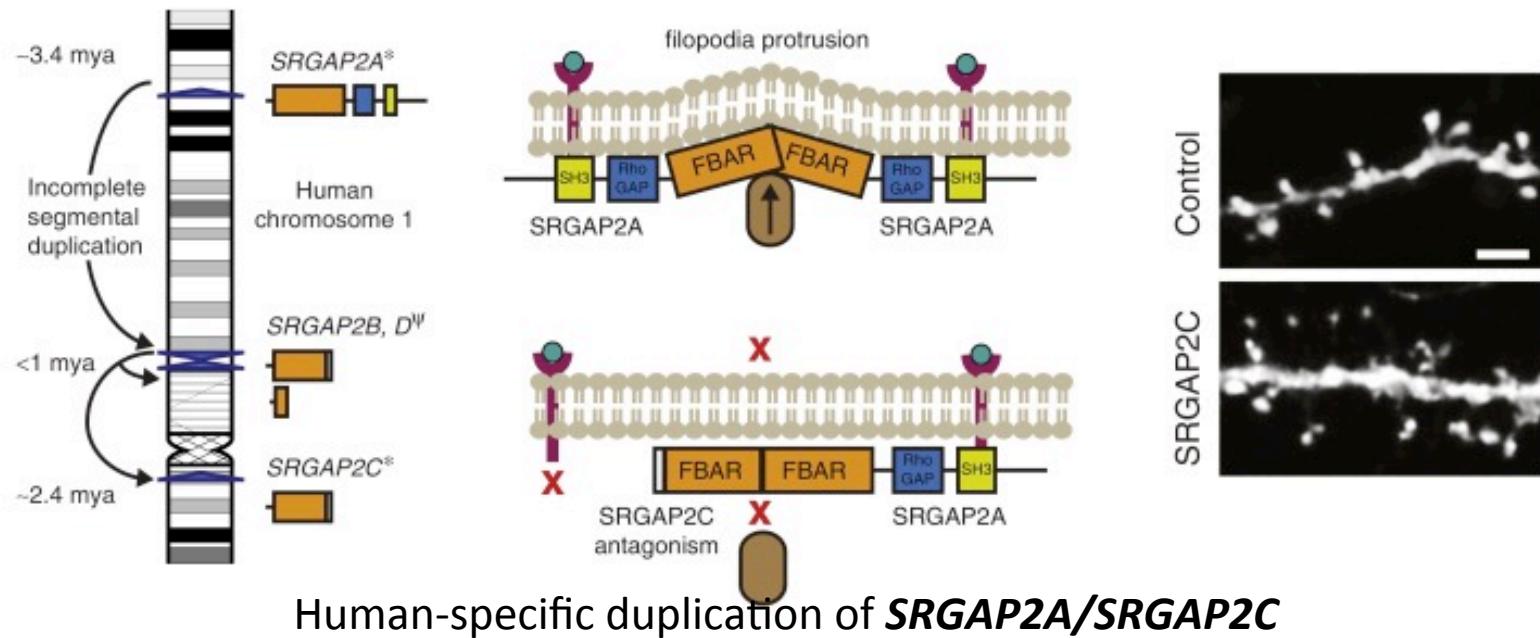
Chromosomal rearrangement between Human and Chimp - From Strachan & Dean 1998 p 344

# Gene duplication/loss



From Assis lab website: <http://www.personal.psu.edu/rua15/research.html>

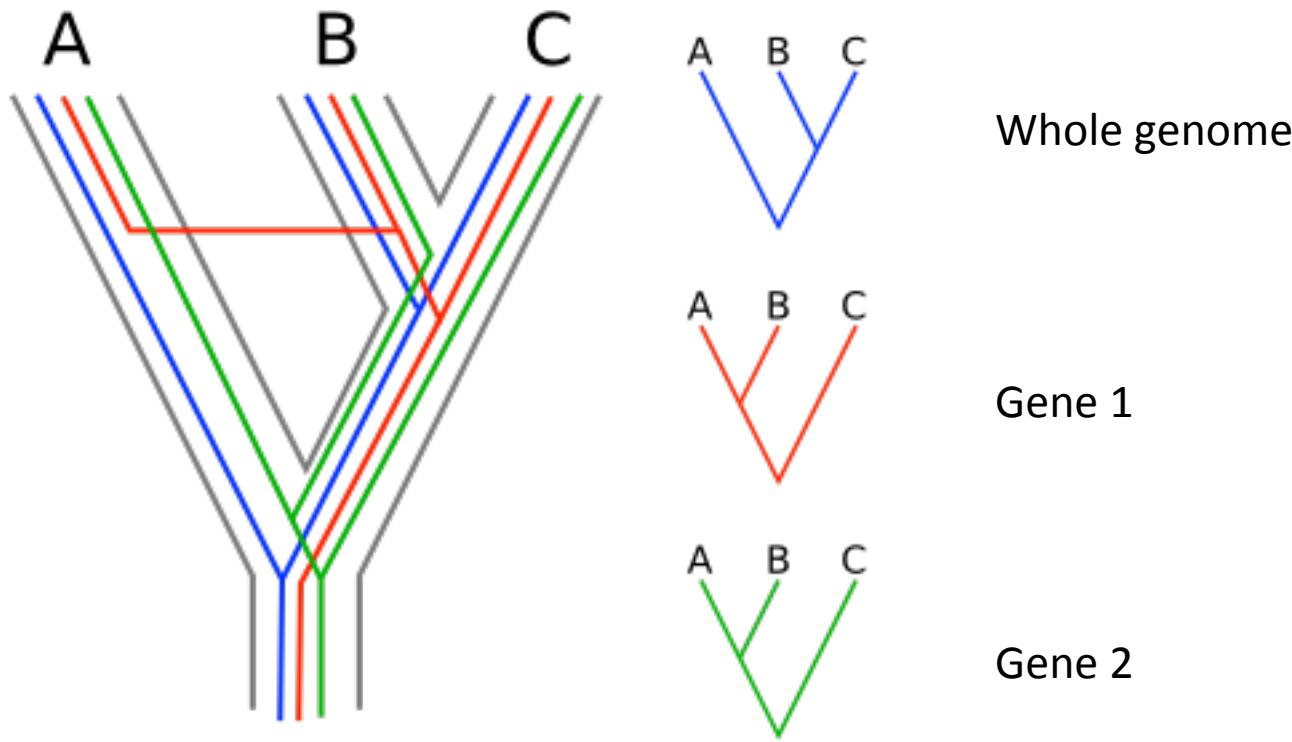
# Genes duplication/loss: evolution of brain



- *SRGAP2C* is fixed in humans
- *SRGAP2C* is expressed in human embryonic cortex / brain
- Over-expression of *SRGAP2C* in mice leads to human-specific features:
  - neoteny
  - longer spines at a greater density
  - sustained radial migration in the developing mouse neocortex

# Inferring phylogenetic trees

- Importance of having information of the whole genome to obtain accurate phylogenetic trees

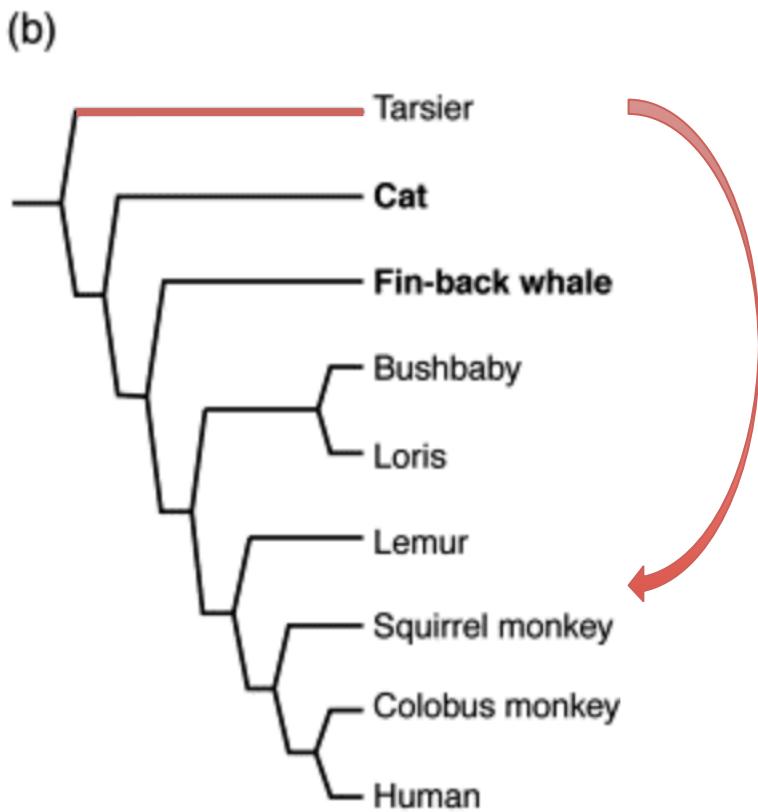


Where to find tools to obtain phylogenetic trees:

- <http://molbiol-tools.ca/Phylogeny.htm>
- [https://en.wikipedia.org/wiki/List\\_of\\_phylogenetics\\_software](https://en.wikipedia.org/wiki/List_of_phylogenetics_software)

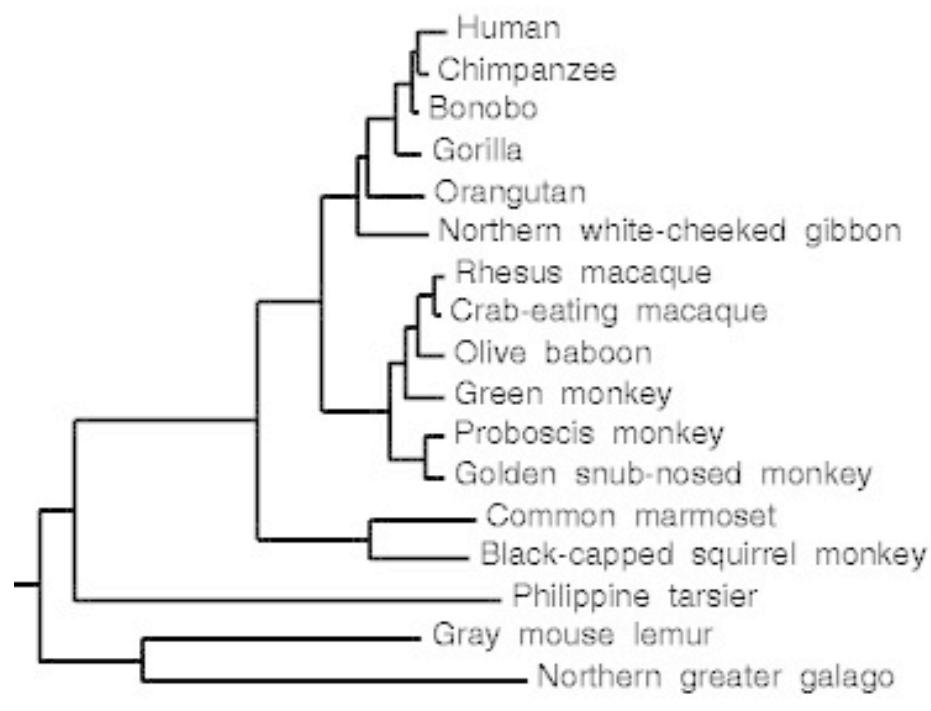
# Inferring phylogenetic trees

Cytochrome b phylogenetic tree



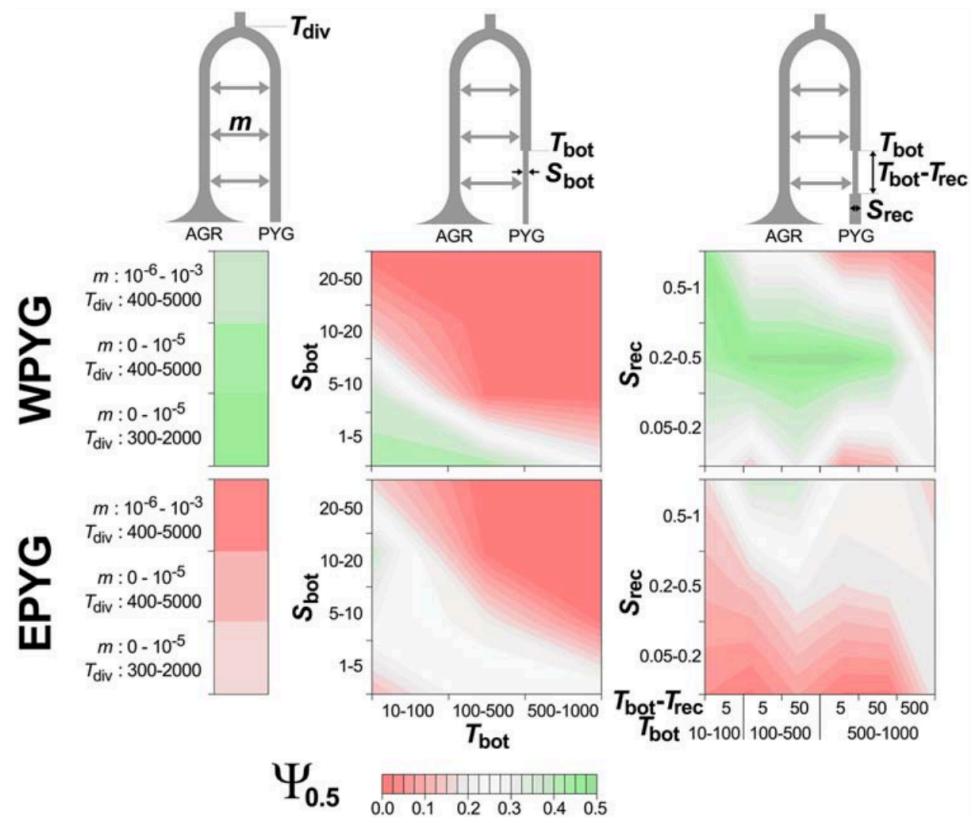
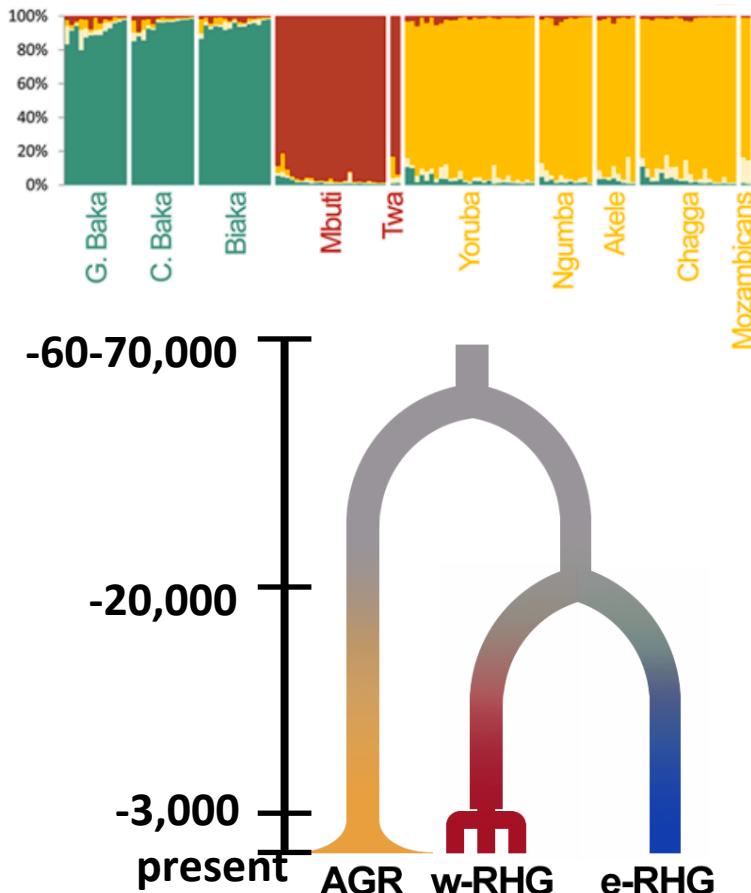
Whole genome phylogenetic tree

a) UCSC genome browser



# Inferring demographic history

- Estimating population structure from allele frequencies (e.g. ADMIXTURE software <https://www.genetics.ucla.edu/software/admixture/>).
- Estimating population branching and demographic history (e.g. using ABC )



Adapted from Patin et al., PLoS Genet (2009)

# Neanderthal, Denisova and modern humans

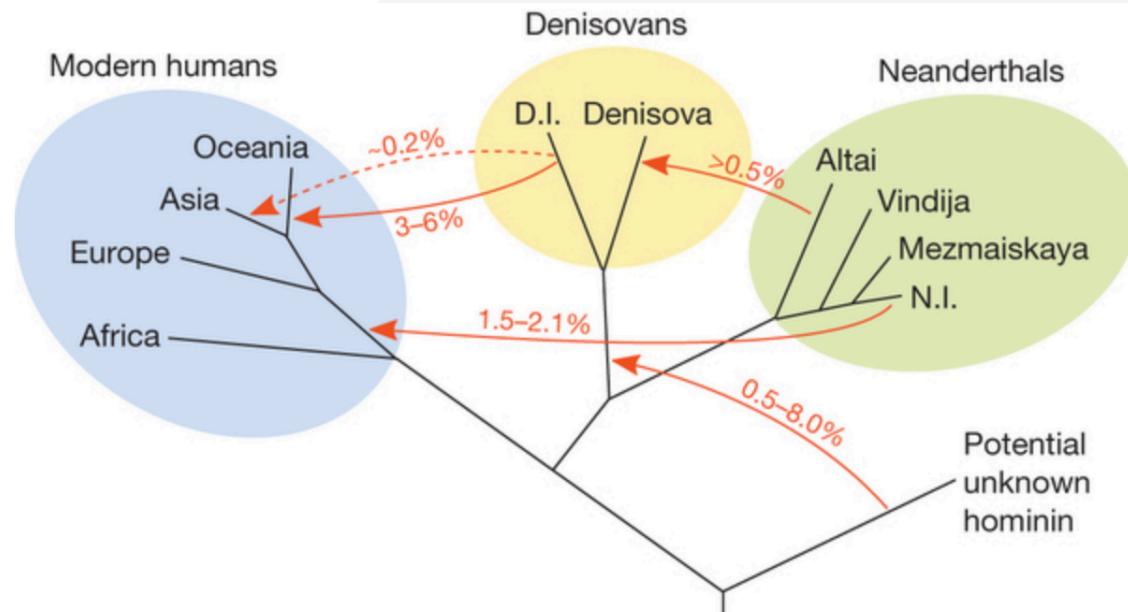
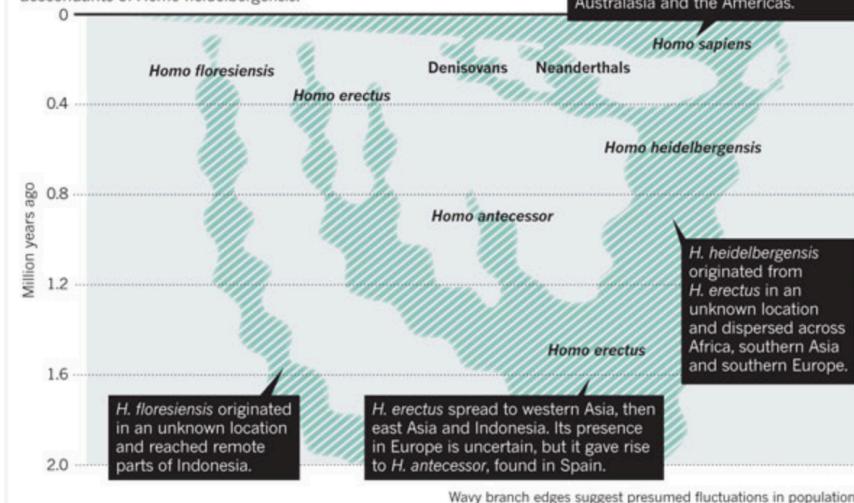
## A Draft Sequence of the Neandertal Genome

Richard E. Green,<sup>1,\*†‡</sup> Johannes Krause,<sup>1,†§</sup> Adrian W. Briggs,<sup>1,†§</sup> Tomislav Maricic,<sup>1,†§</sup> Udo Stenzel,<sup>1,†§</sup> Martin Kircher,<sup>1,†§</sup> Nick Patterson,<sup>2,†§</sup> Heng Li,<sup>2,†</sup> Weiwei Zhai,<sup>3,†||</sup> Markus Hsi-Yang Fritz,<sup>4,†</sup> Nancy F. Hansen,<sup>5,†</sup> Eric Y. Durand,<sup>3,†</sup> Anna-Sapfo Malaspinas,<sup>3,†</sup> Jeffrey D. Jensen,<sup>6,†</sup> Tomas Marques-Bonet,<sup>7,13,†</sup> Can Alkan,<sup>7,†</sup> Kay Prüfer,<sup>1,†</sup> Matthias Meyer,<sup>1,†</sup> Hernán A. Burbano,<sup>1,†</sup> Jeffrey M. Good,<sup>1,8</sup> Rigo Schultz,<sup>1</sup> Ayinuer Aximu-Petri,<sup>1</sup> Anne Butthof,<sup>1</sup> Barbara Höber,<sup>1</sup> Barbara Höffner,<sup>1</sup> Madlen Siegemund,<sup>1</sup> Antje Weihmann,<sup>1</sup> Chad Nusbaum,<sup>2</sup> Eric S. Lander,<sup>2</sup> Carsten Russ,<sup>2</sup> Nathaniel Novod,<sup>2</sup> Jason Affourtit,<sup>9</sup> Michael Egholm,<sup>9</sup> Christine Verna,<sup>21</sup> Pavao Rudan,<sup>10</sup> Dejana Brajkovic,<sup>11</sup> Željko Kucan,<sup>10</sup> Ivan Gušić,<sup>10</sup> Vladimir B. Doronichev,<sup>12</sup> Liubov V. Golovanova,<sup>12</sup> Carles Lalueza-Fox,<sup>13</sup> Marco de la Rasilla,<sup>14</sup> Javier Fortea,<sup>14,||</sup> Antonio Rosas,<sup>15</sup> Ralf W. Schmitz,<sup>16,17</sup> Philip L. F. Johnson,<sup>18,†</sup> Evan E. Eichler,<sup>7,†</sup> Daniel Falush,<sup>19,†</sup> Ewan Birney,<sup>4,†</sup> James C. Mullikin,<sup>5,†</sup> Montgomery Slatkin,<sup>3,†</sup> Rasmus Nielsen,<sup>3,†</sup> Janet Kelso,<sup>1,†</sup> Michael Lachmann,<sup>1,†</sup> David Reich,<sup>2,20,\*†</sup> Svante Pääbo<sup>1,\*†</sup>

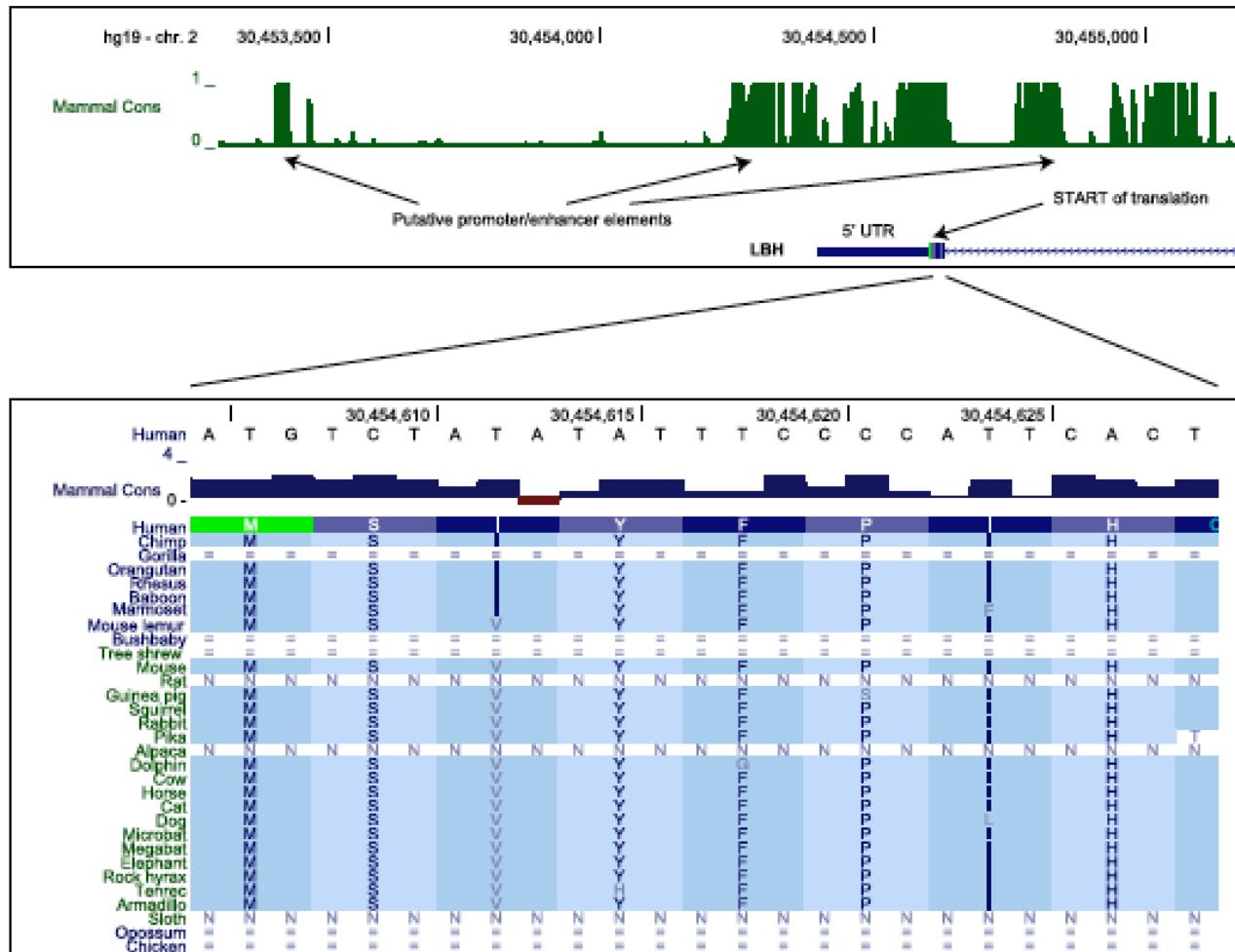
### A WINDING PATH

After early modern humans left Africa around 60,000 years ago (top right), they spread across the globe and interbred with other descendants of *Homo heidelbergensis*.

*H. sapiens* spread from Africa to western Asia and then to Europe and southern Asia, eventually reaching Australasia and the Americas.

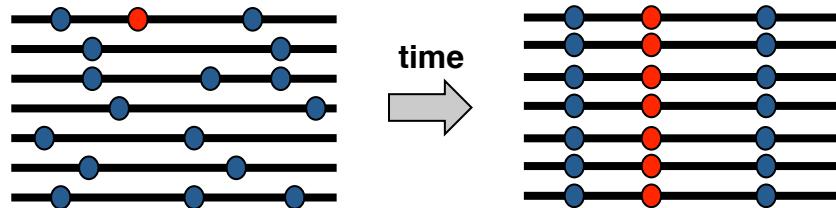


# Natural selection of genomic loci



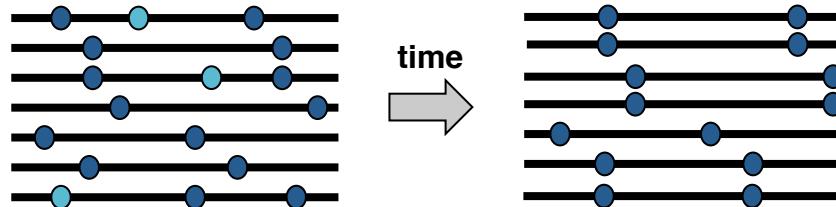
# Evolution of genomic loci: the different types of selection

Positive selection



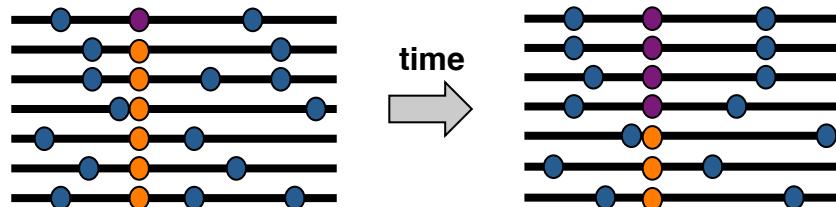
Rapid increase in frequency of advantageous mutation

Negative selection



Remove deleterious mutations

Balancing selection

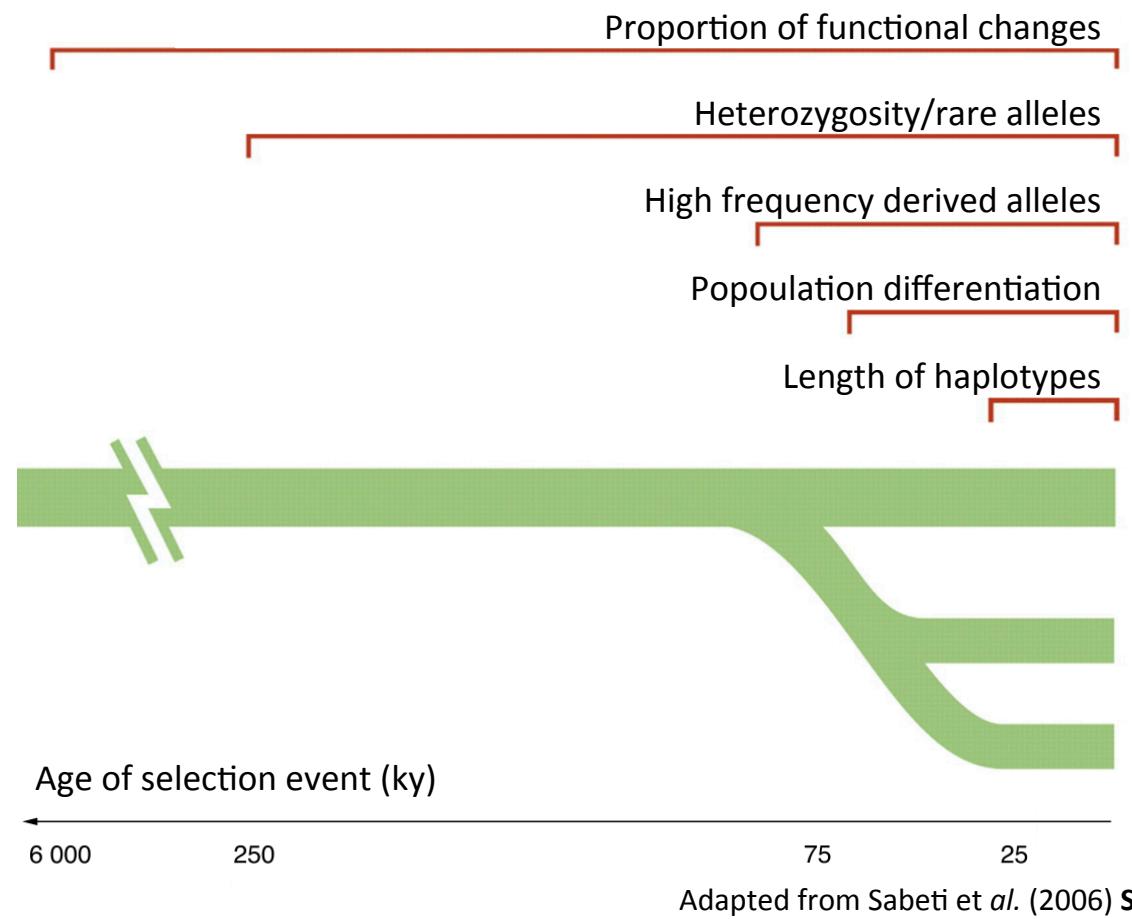


Maintain several alleles in the genome.

These effect can be mimicked by demographic events:

- Demography = global effect on the genome
- Selection = local effect

# Detecting selection at different time scales



**McDonald-Kreitman, Ka/Ks tests**

**Tajima' s D, Fu and Li' s D\***

**Fay and Wu' s H**

$F_{ST}$

**LRH, iHS, DIND, XP-EHH**

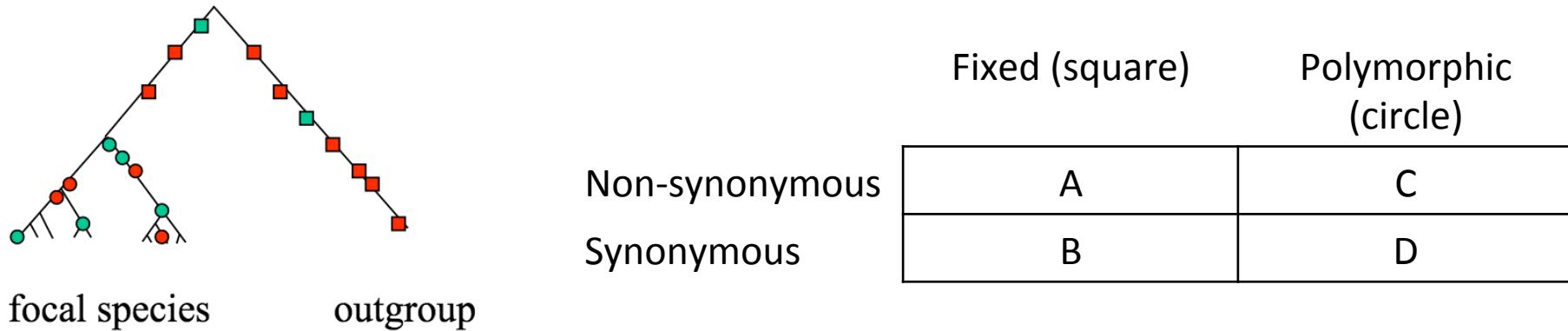
Detecting Natural Selection  
in Genomic Data

Joseph J. Vitti,<sup>1,2</sup> Sharon R. Grossman,<sup>2,4</sup>  
and Pardis C. Sabeti<sup>1,2</sup>

# Interspecies evolution of genomic loci

## Inter-species evolution:

- Mainly based on comparison of divergences and polymorphisms between the studied species and an outgroup.
- The McDonald-Kreitman test:

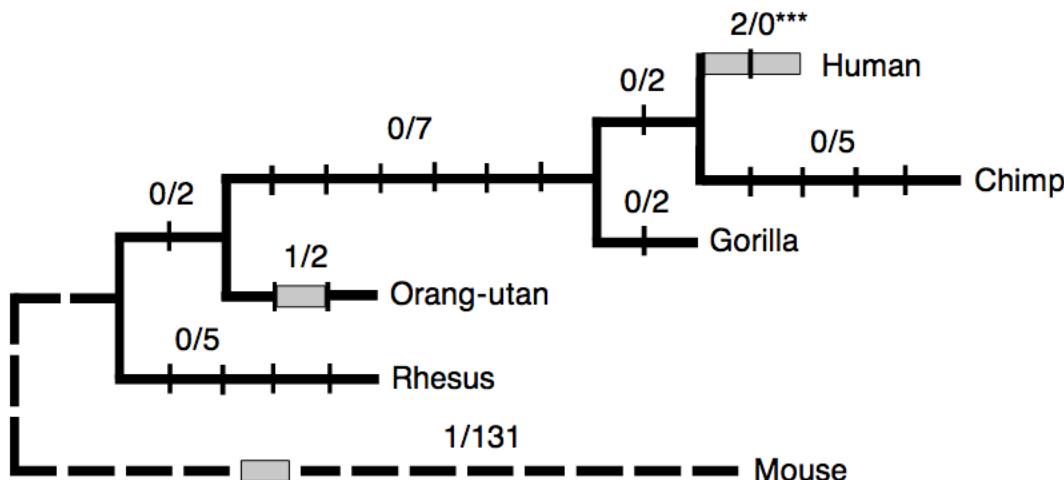


- If  $A/C = B/D$ : Neutrality
- If  $A/C > B/D$ : Positive selection
- If  $A/C < B/D$ : Negative selection OR balancing selection (need to look at SNP frequency)

# The example of *FOXP2* and the language

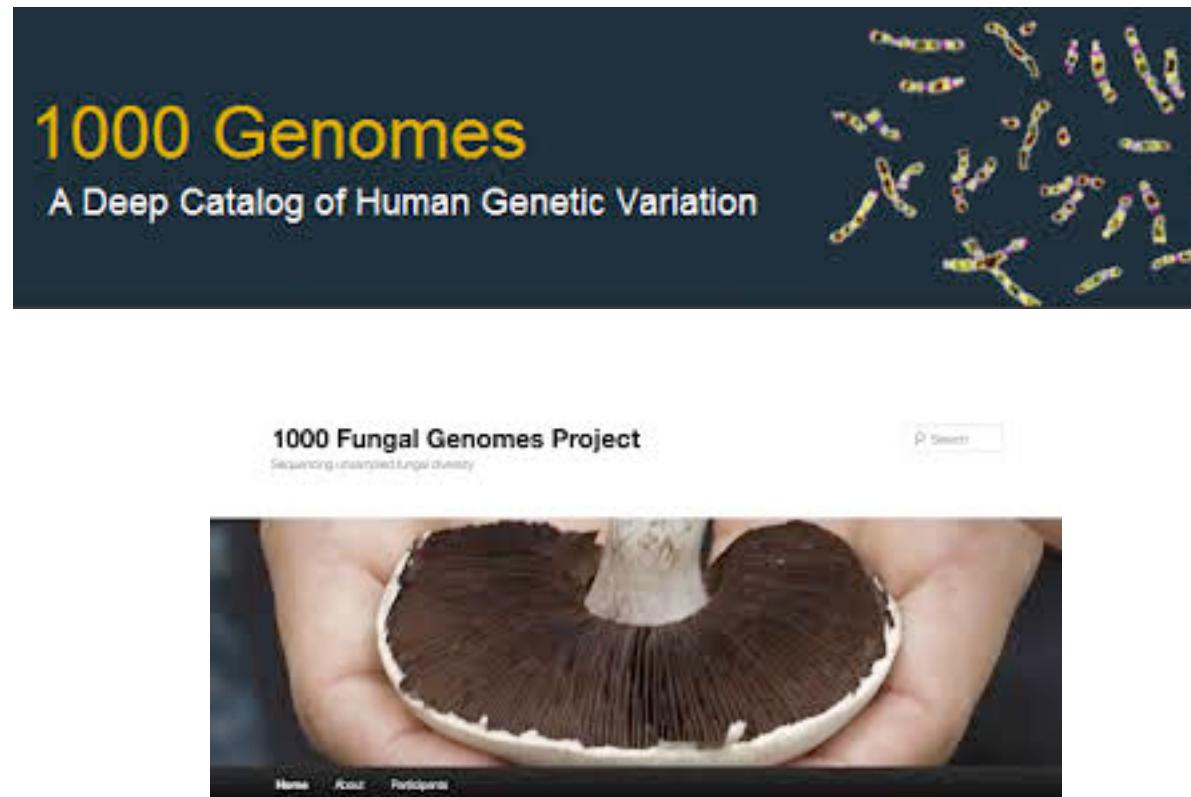
## Molecular evolution of *FOXP2*, a gene involved in speech and language

Wolfgang Enard\*, Molly Przeworski\*, Simon E. Fisher†, Cecilia S. L. Lai†,  
Victor Wiebe\*, Takashi Kitano\*, Anthony P. Monaco† & Svante Pääbo\*



- Gene involved in human ability to acquire language
- Dominant mutation in this gene lead to severe articulation difficulties and linguistic impairment.
- Under positive selection in humans.

# Intra-species selection: population-specific selection



- Advent of next-generation sequencing technologies allow to generate whole genome sequences at low cost.
  - Makes comparison of whole genome sequences between individuals and populations possible.

## Genome-wide scans for footprints of natural selection

Taras K. Oleksyk, Michael W. Smith and Stephen J. O'Brien



Population differentiation as a test for selective sweeps  
Hua Chen, Nick Patterson and David Reich

Natural selection has driven population differentiation in modern humans

Identifying Recent Adaptations in Large-Scale Genomic Data

Sharon R. Grossman,<sup>1,2,4,14,\*</sup> Kristian G. Andersen,<sup>1,6,14</sup> Ilya Shlyakhter,<sup>1,6,14</sup> Shervin Tabrizi,<sup>1,6,14</sup> Sarah Winnicki,<sup>1,6</sup> Elinor K. Karlsson,<sup>1,6</sup> Sunny H. Wong,<sup>1,6</sup> Moran Cabili,<sup>1,6</sup> Angela Yen,<sup>1,3</sup> Daniel J. Park,<sup>1,6</sup> Dustin Griesemer,<sup>4,6</sup> Elinor K. Karlsson,<sup>1,6</sup> Sunny H. Wong,<sup>1,6</sup> Moran Cabili,<sup>1,6</sup> Richard A. Adesigboya,<sup>8</sup> Rameshwar N.K. Bamezai,<sup>10</sup> Adriam V.S. Hill,<sup>8</sup> Fredrik O. Vannberg,<sup>9</sup> John L. Rinn,<sup>1,7,12</sup> 1000 Genomes Project, Eric S. Lander,<sup>1,2,5</sup> Stephen F. Schaffner,<sup>1</sup> and Pardis C. Saberi,<sup>1,6,13,\*</sup>



Signals of recent positive selection in a worldwide sample of human populations

Joseph K. Pickrell, Graham Coop, John Novembre, et al.



Genomic regions exhibiting positive selection identified from dense genotype data

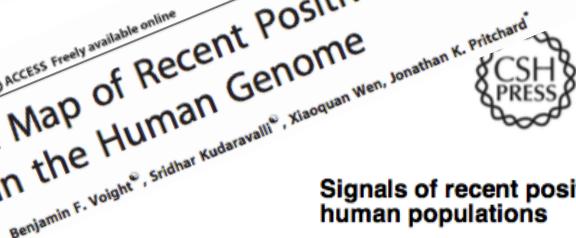
Christopher S. Carlson, Daryl J. Thomas, Michael A. Eberle, et al.

Genome-wide detection and characterization of positive selection in human populations

Pardis C. Saberi<sup>1,\*</sup>, Patrick Varilly<sup>1,\*</sup>, Ben Fry<sup>1</sup>, Jason Lohmueller<sup>1</sup>, Elizabeth Hostetter<sup>1</sup>, Chris Cotsaris<sup>1</sup>, Xiaohui Xie<sup>1</sup>, Elizabeth H. Byrne<sup>1</sup>, Steven A. McCarroll<sup>1,2</sup>, Rachelle Gaudet<sup>3</sup>, Stephen F. Schaffner<sup>1</sup>, Eric S. Lander<sup>1,2</sup>, Luis B Barreiro<sup>1,2</sup>, Guillaume Laval<sup>1,2</sup>, Hélène Quach<sup>1</sup>, Etienne Patin<sup>1</sup> & Lluís Quintana-Murci<sup>1</sup>  
& The International HapMap Consortium

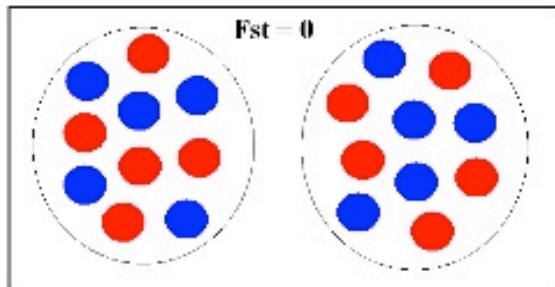
OPEN ACCESS Freely available online

A Map of Recent Positive Selection in the Human Genome

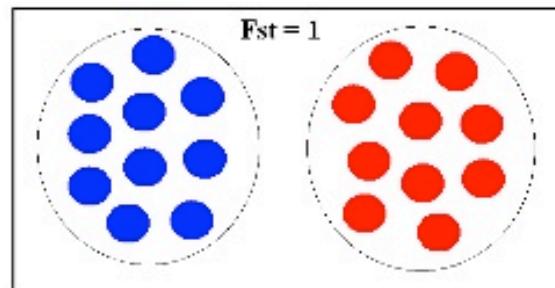


Benjamin F. Voight,<sup>\*</sup> Sridhar Kudaravalli,<sup>8</sup> Xiaoquan Wen, Jonathan K. Pritchard,<sup>\*</sup>

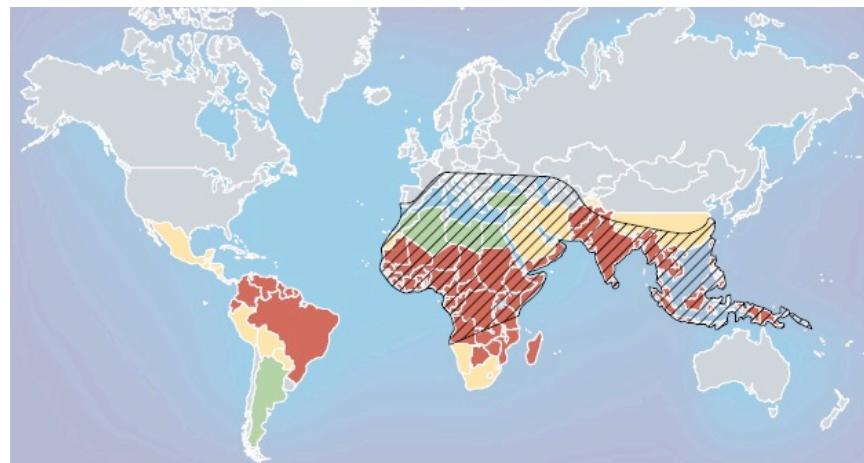
# Population differentiation: the example of DARC



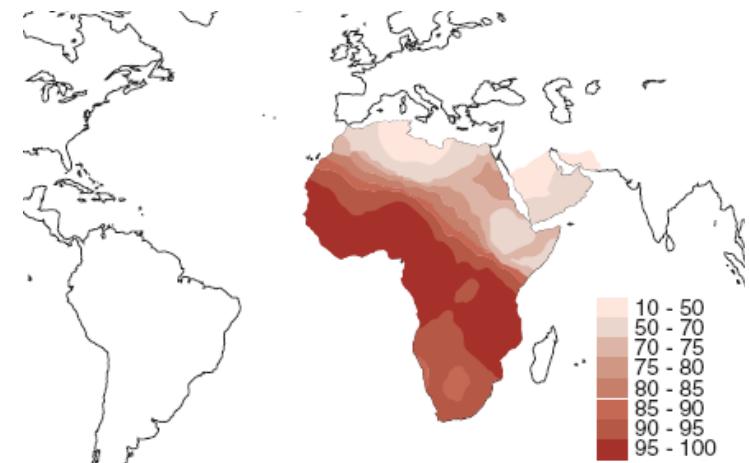
➤ If population-specific selection: local increase in  $F_{ST}$ .



Global distribution of malaria (green to red)

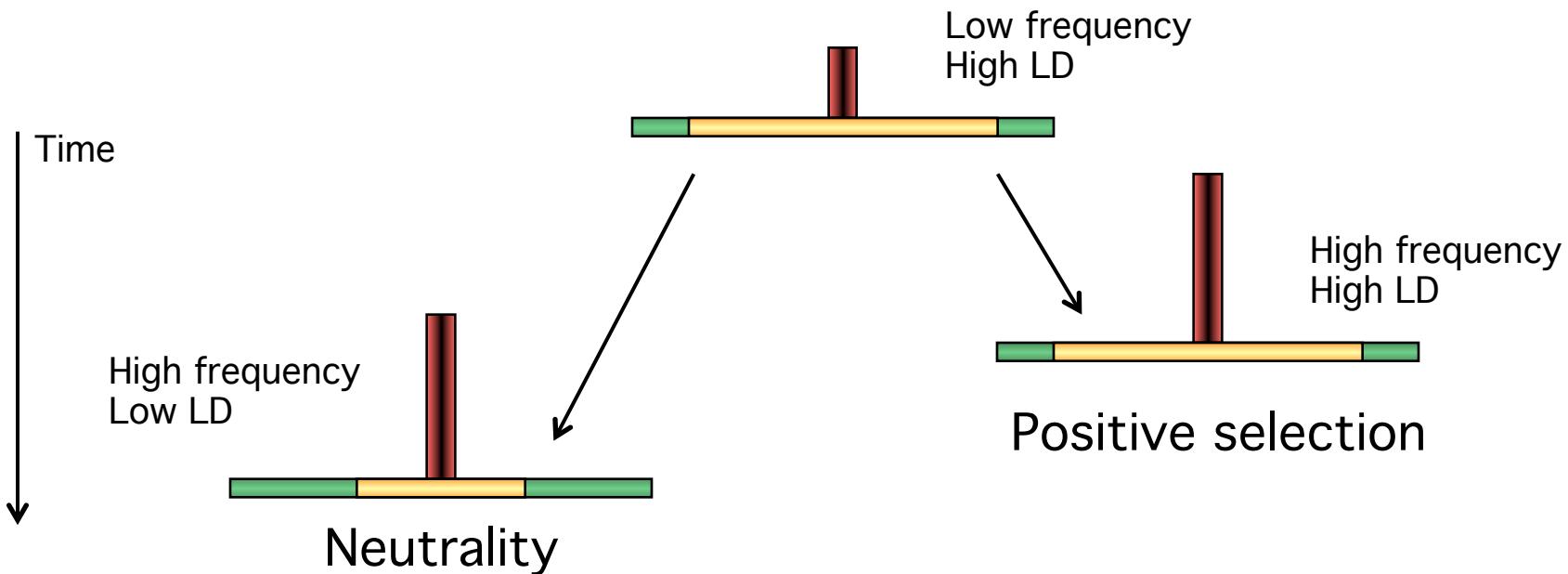


Extreme population differences in FY\*O frequency



Positive selection for the DARC null allele in Africa

# Haplotype-based statistics



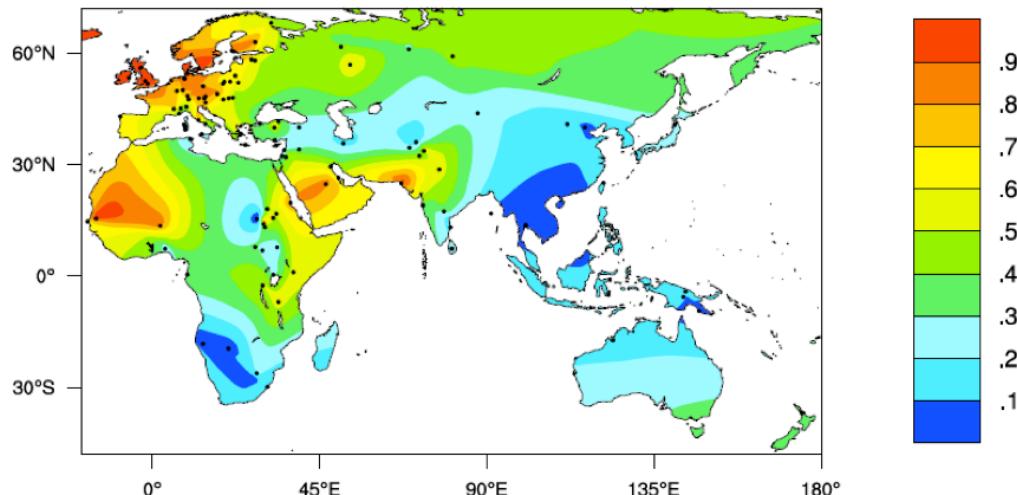
Sabeti et al. Detecting recent positive selection from haplotype block structure.  
Nature (2002)

Examples of haplotype reconstruction software from WGS data:

- PHASE (<http://stephenslab.uchicago.edu/software.html#phase>)
- SHAPEIT (<http://www.shapeit.fr/>)

# Population Genetics: Example of *LCT* in humans

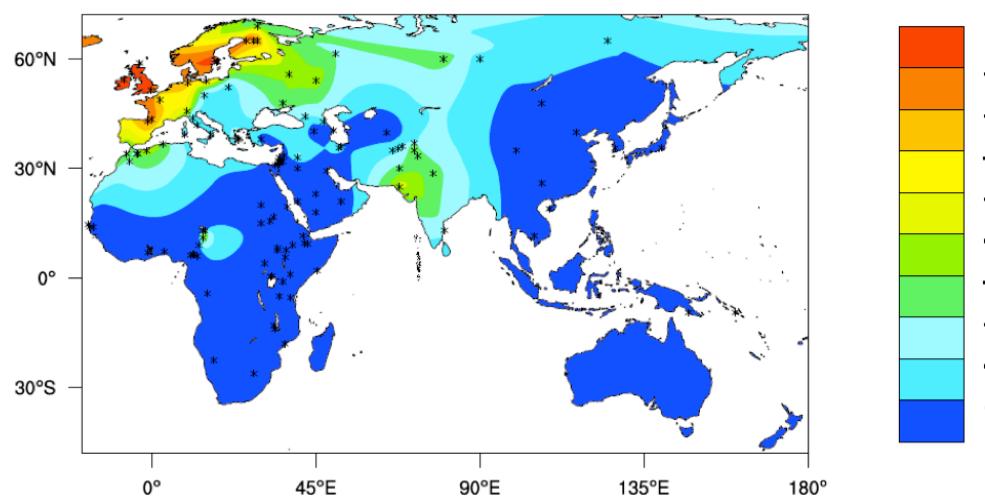
Lactase persistence map



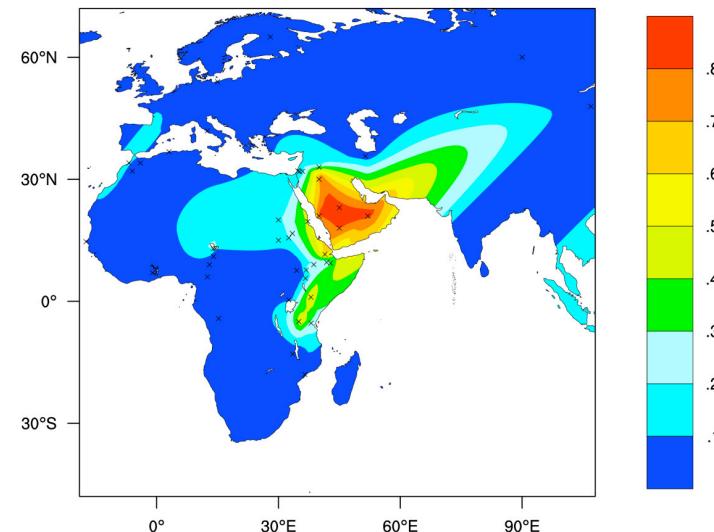
***LCT*: Gene coding for lactase enzyme.**

- Mutations in the promoter of this gene allow persistence of lactase in adulthood.
- A recent positive selection event (~4,000-7,000 y)
- An example of convergent evolution.

LP inferred from -13910\*T allele frequency



LP inferred from 3 other LP alleles



# Advantages of WGS in population genetics

- **What did genomics bring to the field of evolution?**
  - separate population demographic history and selection signatures.
  - study selection without using any candidate-gene approach.
  - access to sequences of non-coding region.
- **Main findings using genomics in human evolutionary history:**
  - Admixture event between Denisova, Neanderthal and Asian and European modern human populations.
  - 70–75% of the non-synonymous mutations are under moderate to strong negative selection.
  - Positive selection largely targets regulatory regions.
  - ~100 major selective sweeps in human populations in the past 100,000 years.

## Exploring the Occurrence of Classic Selective Sweeps in Humans Using Whole-Genome Sequencing Data Sets

Maud Fagny,<sup>1,2,3</sup> Etienne Patin,<sup>1,2</sup> David Enard,<sup>4</sup> Luis B. Barreiro,<sup>5</sup> Lluis Quintana-Murci,<sup>\*1,2</sup> and Guillaume Laval<sup>\*1,2</sup>

Research

## Genome-wide signals of positive selection in human evolution

David Enard,<sup>1</sup> Philipp W. Messer, and Dmitri A. Petrov<sup>1</sup>

Department of Biology, Stanford University, Stanford, California 94305, USA

# Evolutionary biology: future directions

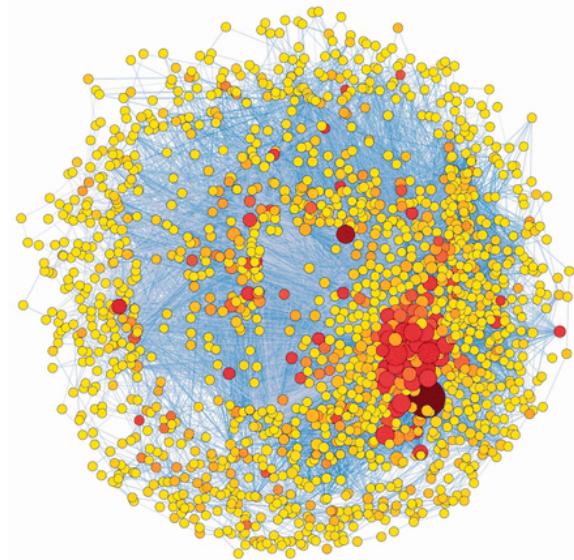
## Evolution of gene networks

### Revealing the architecture of gene regulation: the promise of eQTL studies

Yoav Gilad<sup>1</sup>, Scott A. Rifkin<sup>2</sup> and Jonathan K. Pritchard<sup>1</sup>

<sup>1</sup> Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

<sup>2</sup> Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA



Aluru et al., Nucl. Acid Res. (2012)

#### Research Article

### Positive Selection and Centrality in the Yeast and Fly Protein-Protein Interaction Networks

Sandip Chakraborty and David Alvarez-Ponce

Trends in Plant Science

CelPress

### Evidence for Polygenic Adaptation to Pathogens in the Human Genome

Josephine T. Daub,<sup>\*1,2</sup> Tamara Hofer,<sup>1,2</sup> Emilie Cutivet,<sup>1</sup> Isabelle Dupanloup,<sup>1,2</sup> Lluis Quintana-Murci,<sup>3,4</sup> Marc Robinson-Rechavi,<sup>2,5</sup> and Laurent Excoffier<sup>\*1,2</sup>

#### Opinion

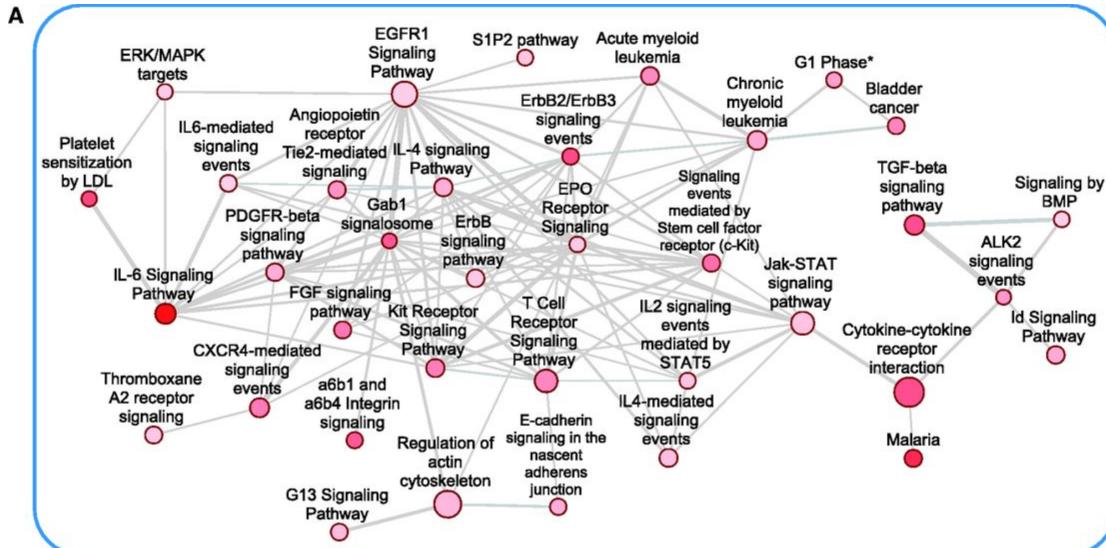
Beyond Genomics: Studying Evolution with Gene Coexpression Networks

Colin Ruprecht,<sup>1</sup> Neha Vaid,<sup>2</sup> Sebastian Proost,<sup>2</sup> Staffan Persson,<sup>3,4</sup> and Marek Mutwil<sup>2,\*</sup>

# Evolution of complex phenotypes

- GWAS showed that most phenotypes are polygenic.
- Evolution by **Polygenic adaptation** (small shift in frequency of several mutations at different loci)
- Importance of **taking into account the complex relationship between genes (genes networks)** in evolutionary biology

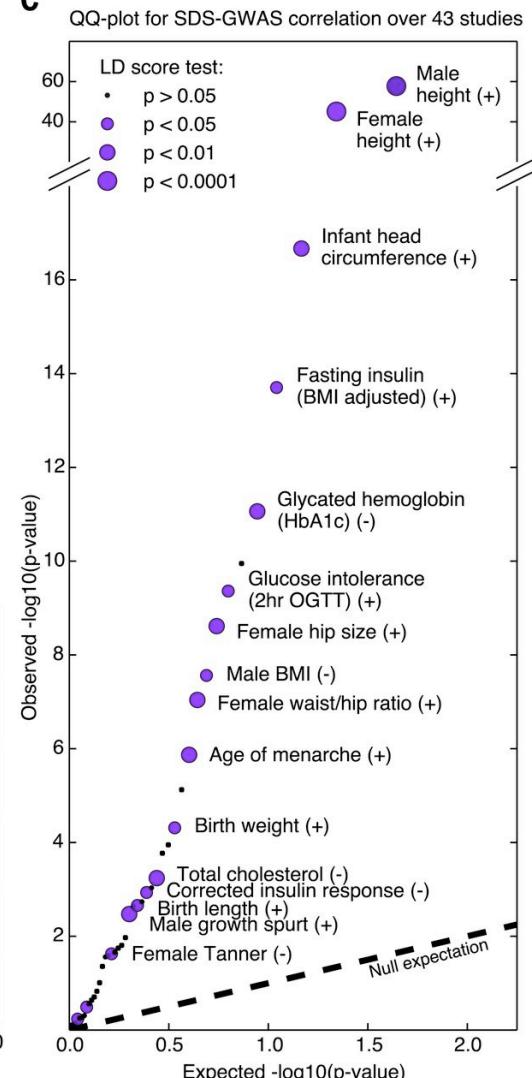
## Positive selection in pathways related to immunity and inflammation



Daub et al., MBE, 2013

## Polygenic selection of anthropomorphic traits

**C**

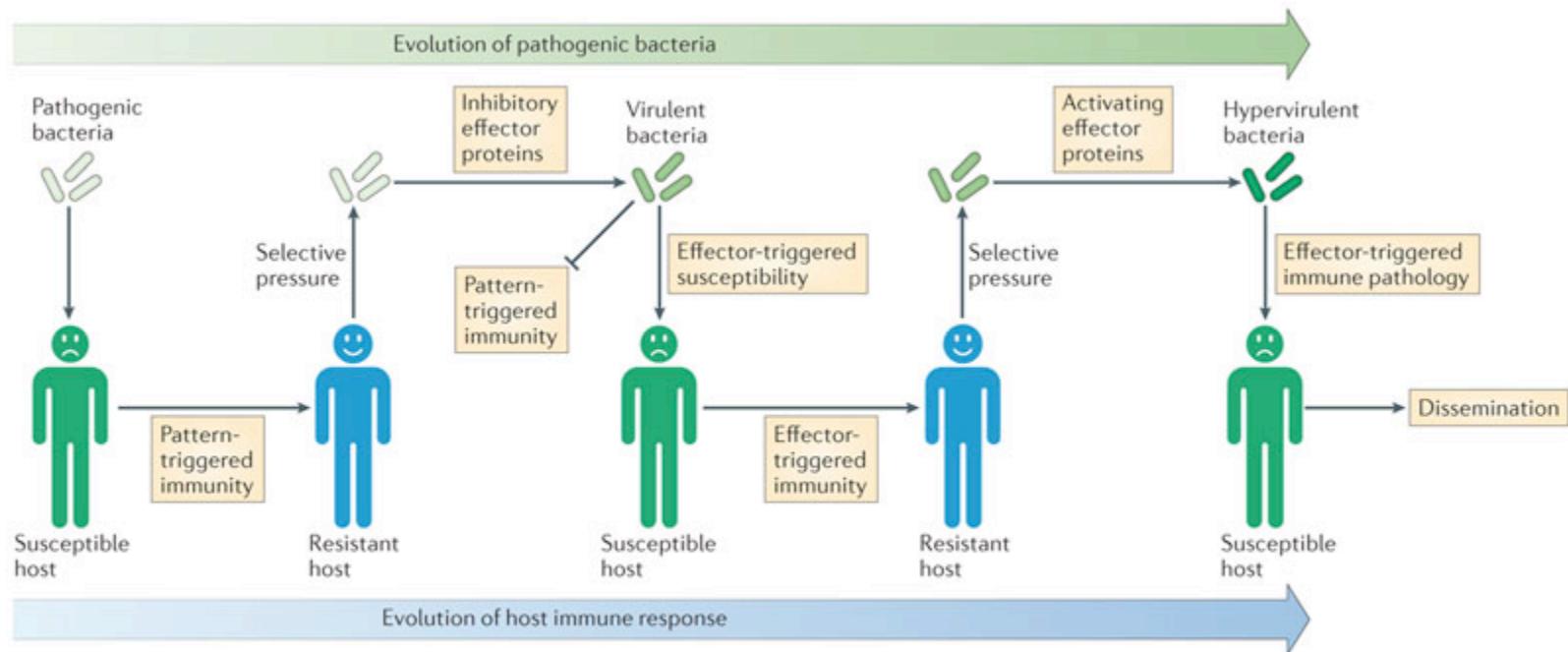


Field et al., Science, 2016

# Evolutionary biology: future directions

## Co-evolution of species

- Hosts and pathogens
- Hosts and microbiomes



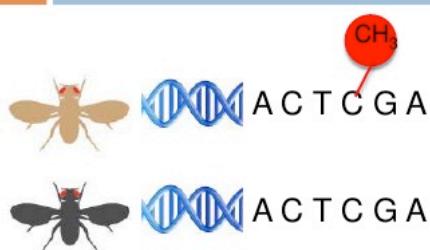
# Evolutionary biology: future directions

## Role of epigenetics in short-term adaptation?

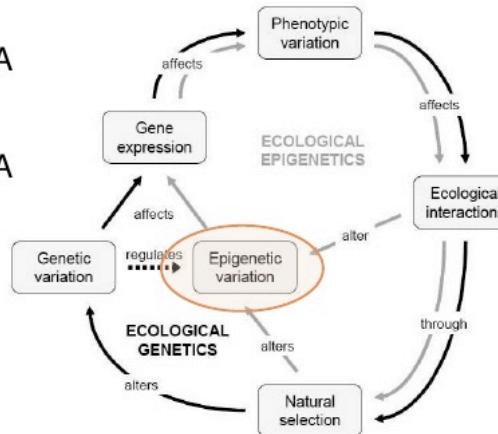
### Epigenetics in ecology and evolution: what we know and what we need to know

KOEN J. F. VERHOEVEN,\* BRIDGETT M. VONHOLDT† and VICTORIA L. SORK‡§

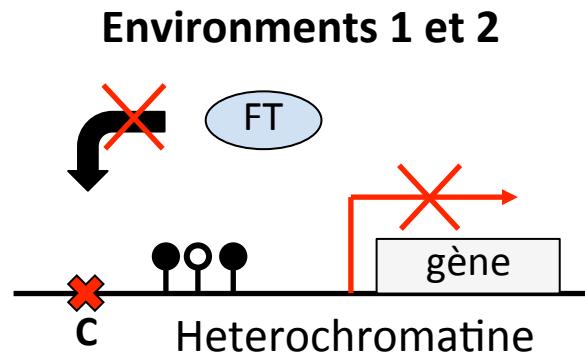
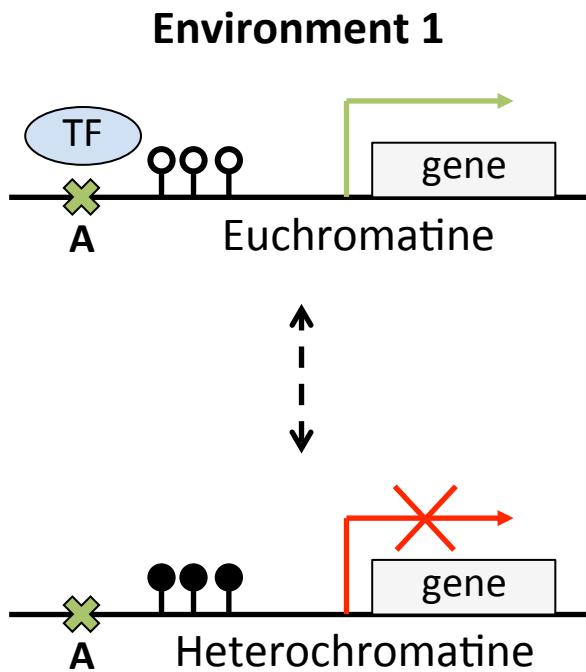
#### Epigenetic variation



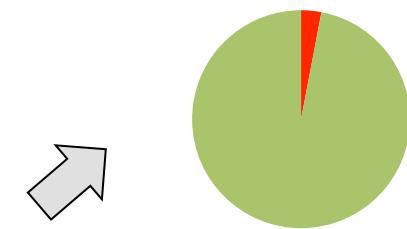
Epigenetic mechanisms can alter gene expression and phenotypes without any change in the DNA sequence



# Epigenetics and evolution

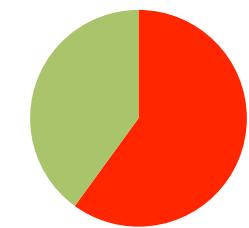


**Environment 1**



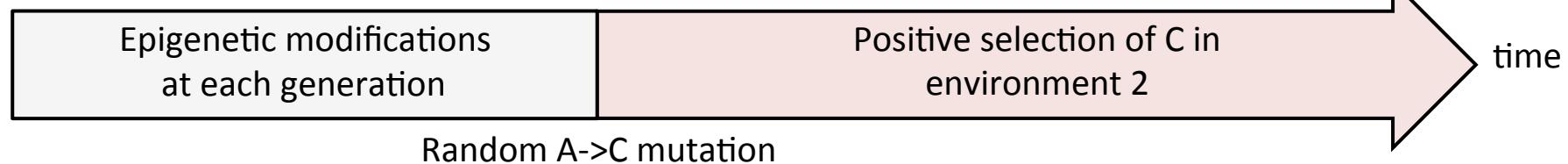
Allele frequencies

■ C ■ A



**Environment 2**

**Environment 2**



FT Transcription factor

SNP

DNA methylation site

● Methylated ○ Unmethylated