Human Variation, Natural Selection and Network Analysis

Hafid Laayouni

ESCI-UPF, 8 February 2021

What we would like to do:

correctly describe and quantify differences in phenotypes -susceptibility to disease as example-

interpret-them: why are they and why are they?

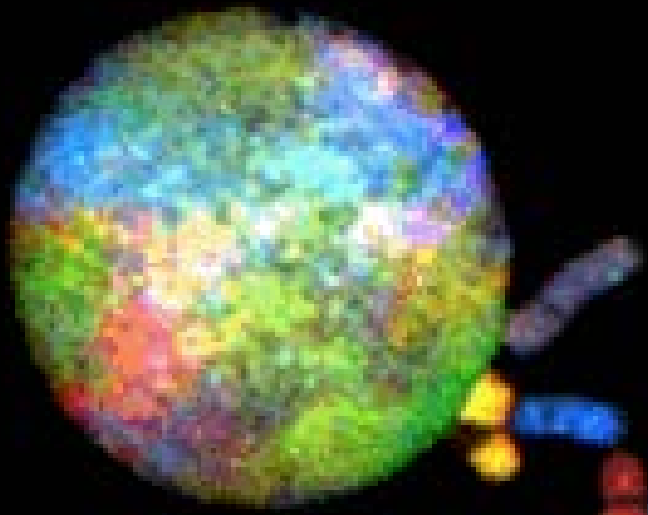
Answers from biology and from evolution



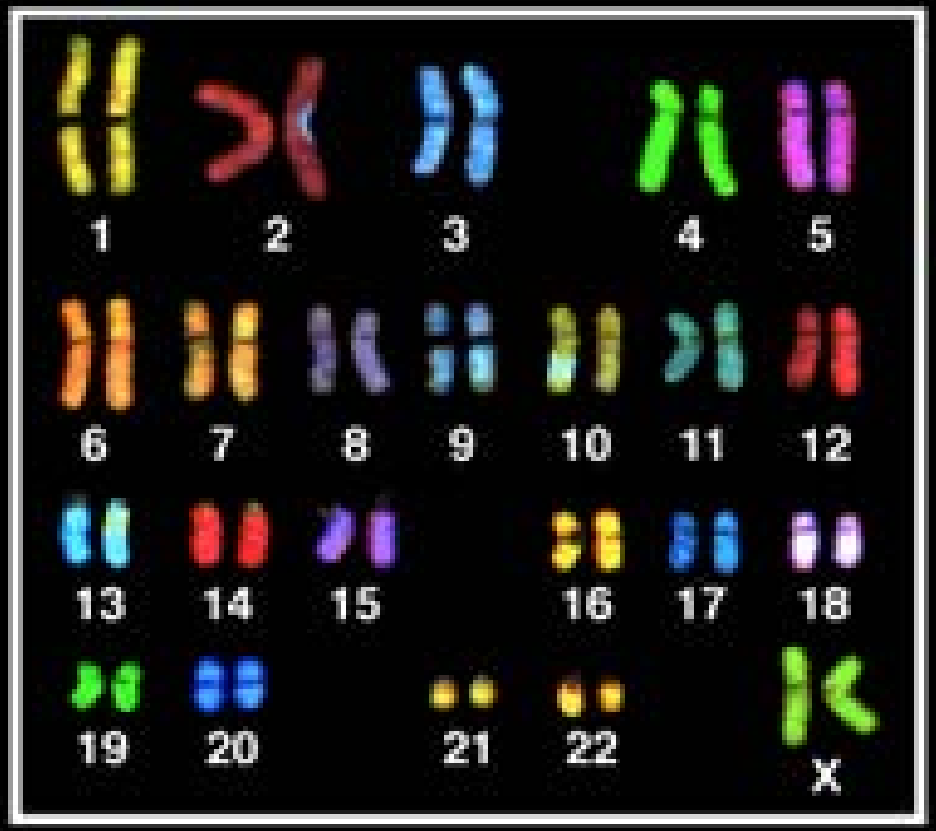

The information is contained in the DNA, which forms genes, and altogether is called the genome

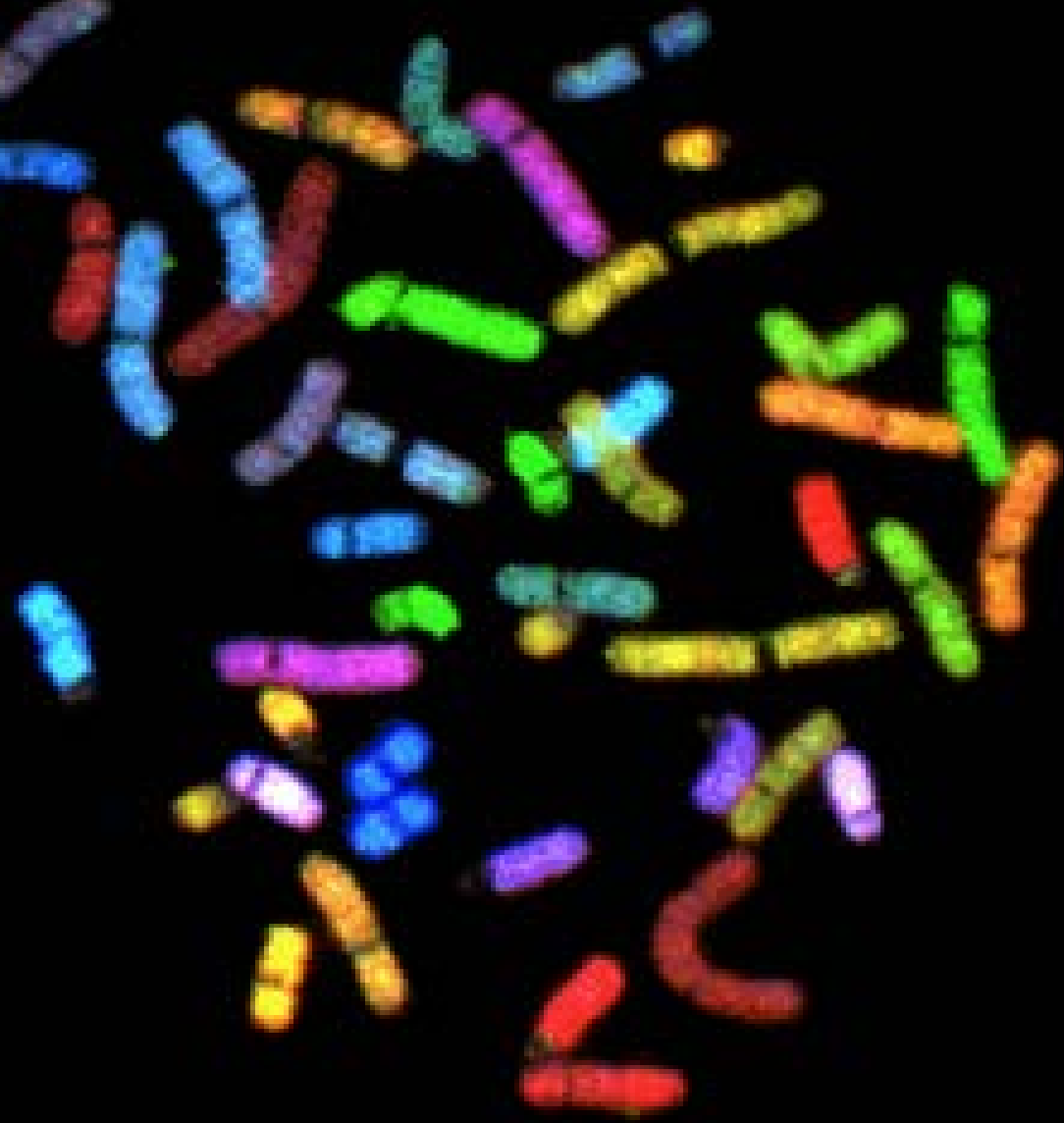


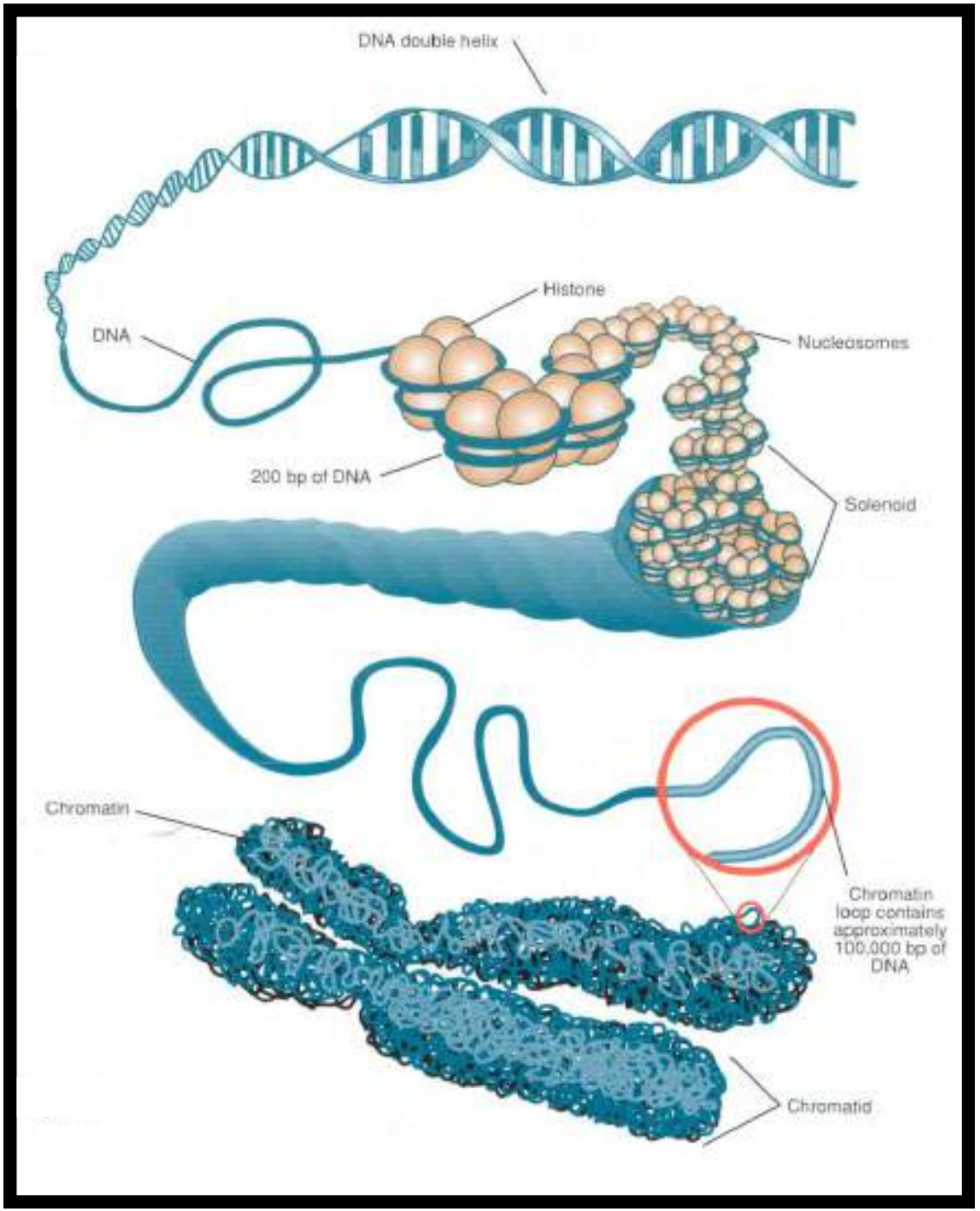
Chromosomes



45mb - 280mb



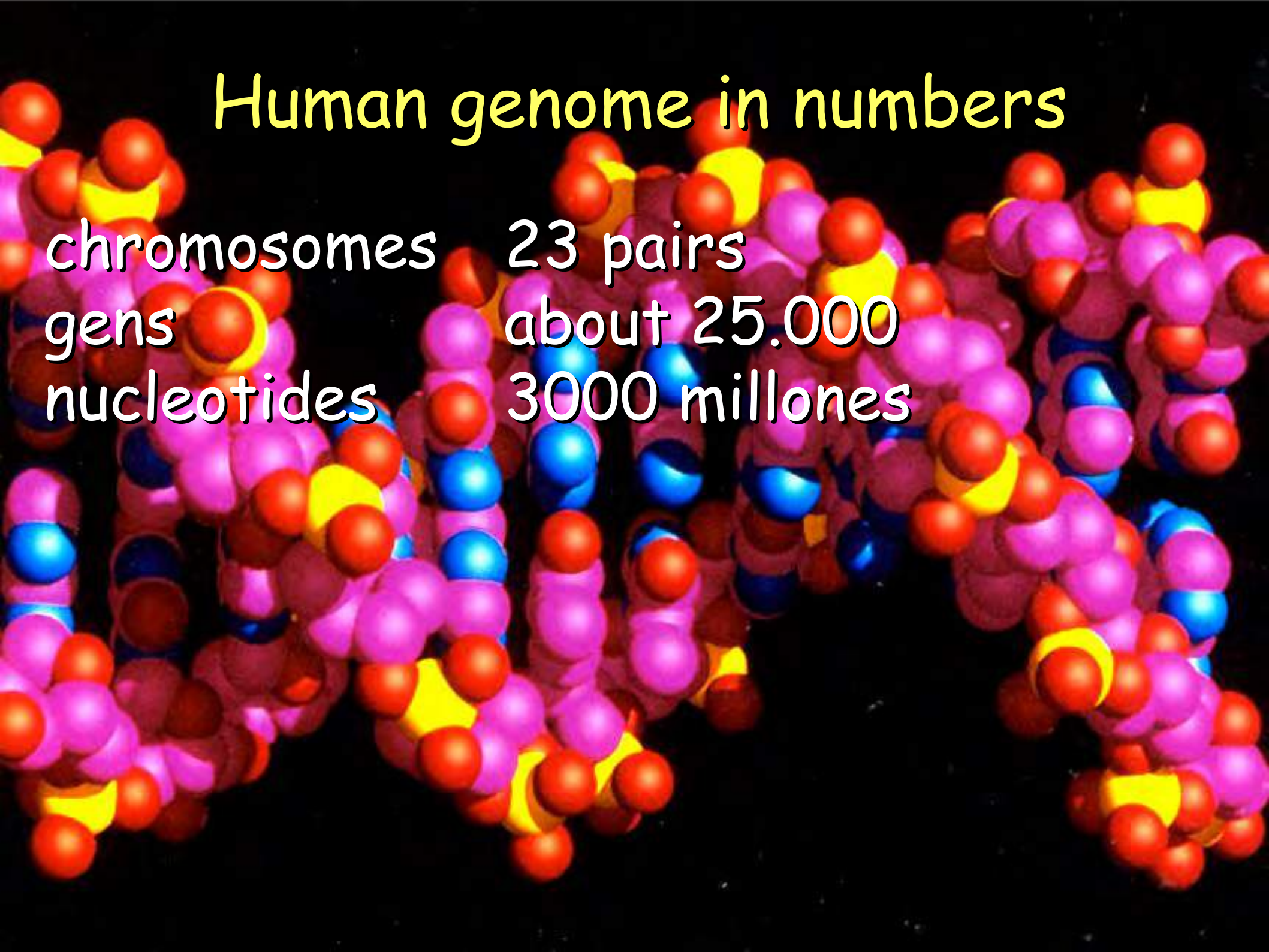



DNA double helix
Histone
DNA
Nucleosomes
200 bp of DNA
Solenoid
Chromatin
Chromatin loop contains approximately 100,000 bp of DNA
Chromatid

ATGCGTCCTGAGAGAGCCTGTGATATAAAGGTGTGTGAAACCA
GATGACAGATGATCCCCAGATTGATTAGACACAGATAGGACAC
ACAGAGATAGAGACACACCAAGGATATCCGTCCTGAGAGAGCC
TGTGATATAAAGGTGTGTGAAACCAGATGACAGATGATCCCC
AGATTGATTAGACACAGATAGGACACACAGAGATAGAGACAC
ACCAAGGATATCCGTCCTGAGAGAGCCTGTGATATAAAGGTGT
GTGAAACCAGATGACAGATGATCCCCAGATTGATTAGACACAG
ATAGGACACACAGAGATAGAGACACACCAAGGATATCCGTCCT
GAGAGAGCCTGTGATATAAAGGTGTGTGAAACCAGATGACAG
ATGATCCCCAGATTGATTAGACACAGATAGGACACACAGAGA
TAGAGACACACCAAGGATATCCGTCCTGAGAGAGCCTGTGATA
TAAAGGTGTGTGAAACCAGATGACAGATGATCCCCAGATTGA
TTAGACACAGATAGGACACACAGAGATAGAGACACACCAAGG
ATATCCGTCCTGAGAGAGCCTGTGATATAAAGGTGTGTGAAAC
CAGATGACAGATGATCCCCAGATTGATTAGACACAGATAGGAC
ACACAGAGATAGAGACACACCAAGGATATTTTCCGATGCCCAA
TCCGTCCTGAGAGAGCCTGTGATATAAAGGTGAGATAGGACAC
ACAGAGATAGAGACACACCAAGGATCCCCCATGGAACTGA

Human genome in numbers

chromosomes	23 pairs
gens	about 25.000
nucleotides	3000 millones



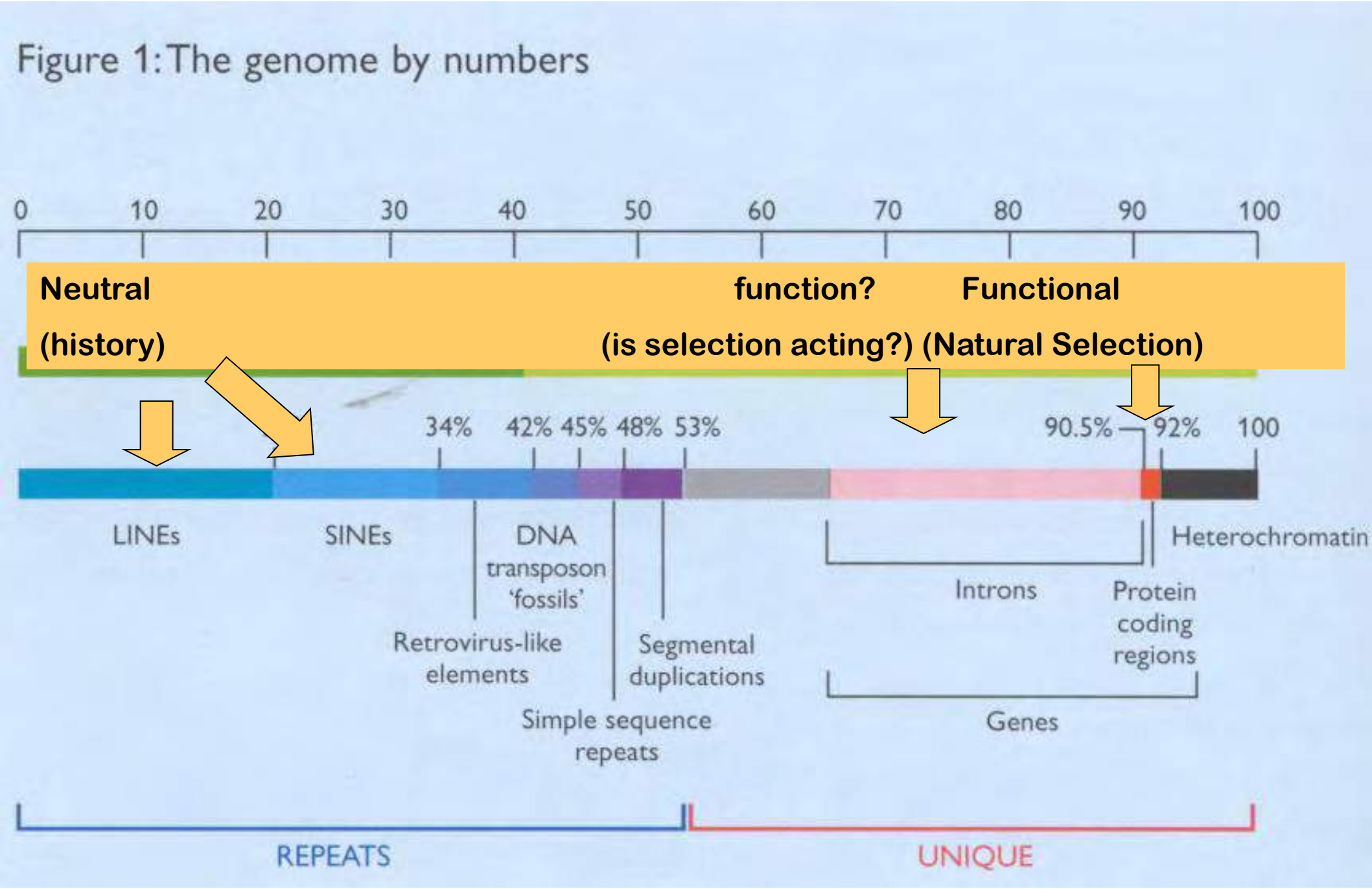


Figure 1: The genome by numbers
0
10
20
30
40
50
60
70
80
90
100
Neutral
(history)
function?
(is selection acting?)
Functional
(Natural Selection)
34%
42%
45%
48%
53%
90.5%
92%
100
LINEs
SINEs
DNA transposon 'fossils'
Retrovirus-like elements
Segmental duplications
Simple sequence repeats
Introns
Protein coding regions
Heterochromatin
Genes
REPEATS
UNIQUE

Can we measure genetic differences?

We know how genetic makeup changes over time

- Mutation
- Natural selection
- Genetic drift (small populations)

Can we reconstruct evolution / history?

1.- Humans among primates

2.- The origin of modern humans

3.- The history of human populations

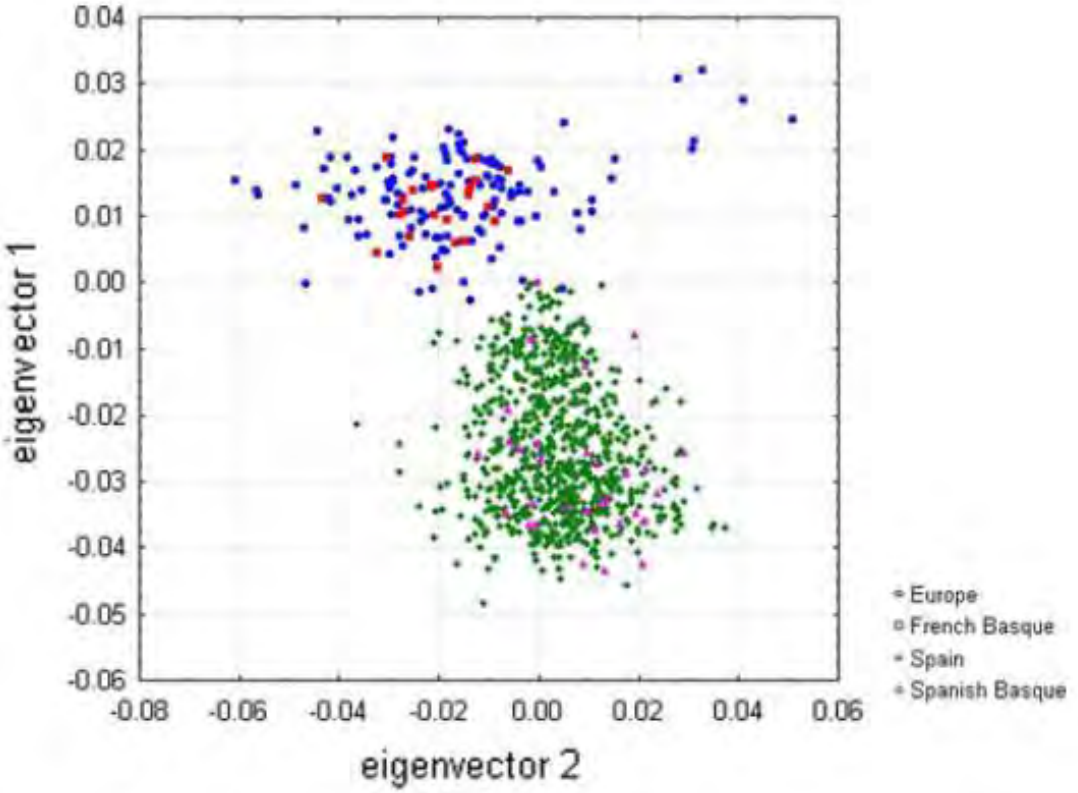


Fig. 2 Spanish and European populations from HGDP samples plotted for the first two principal components obtained by PCA analysis using 109 highly informative SNPs from genotyping data. (Spain: Spanish non-Basque; Europe comprise individuals in HGDP from various locations: French, Sardinian, North Italian, Orcadian, Adygei and Russian)

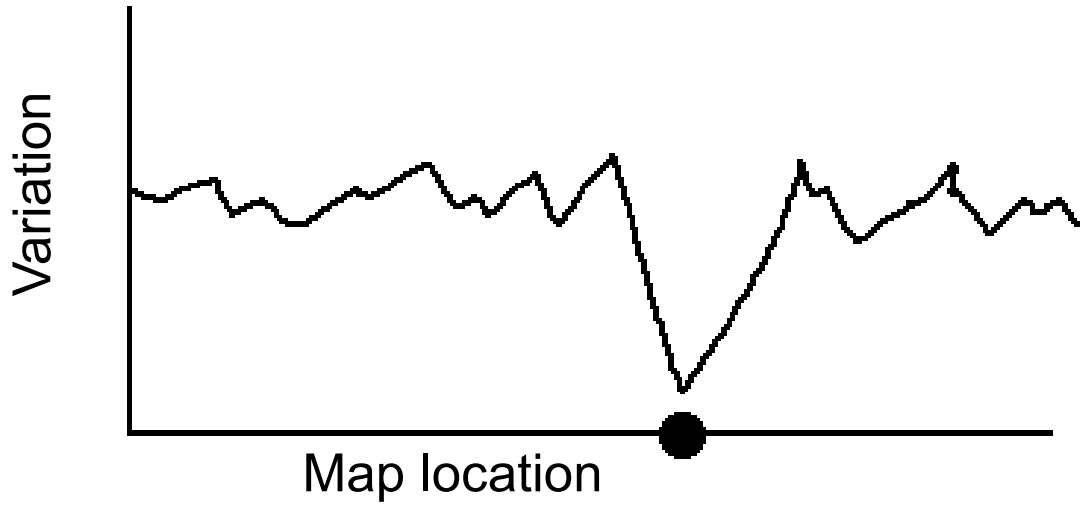
Hum Genet (2010) 127:455–458
DOI 10.1007/s00439-010-0798-3

SHORT REPORT

A genome-wide survey does not show the genetic distinctiveness of Basques

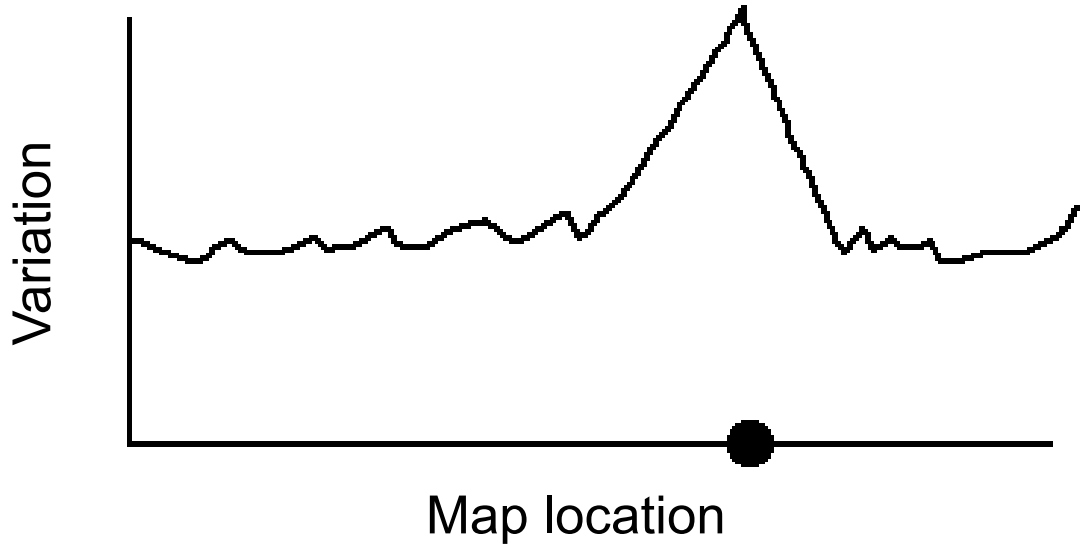
Hafid Laayouni · Francesc Calafell · Jaume Bertranpetit

A scan of levels of polymorphism can thus suggest sites under selection



Directional selection
(selective sweep)

Local region with
reduced mutation rate



Balancing selection

Local region with
elevated mutation rate

Two typical classes of departures are seen with polymorphism data

1: An excess of rare alleles, a deficiency of intermediate frequency alleles (alleles younger than expected)

2: An excess of intermediate frequency alleles, a deficiency of rare alleles (alleles older than expected)

Pattern 1 expected under a selective sweep, when coalescent times are shorter than expected

Pattern 2 expected under balancing selection, when coalescent times are longer than expected

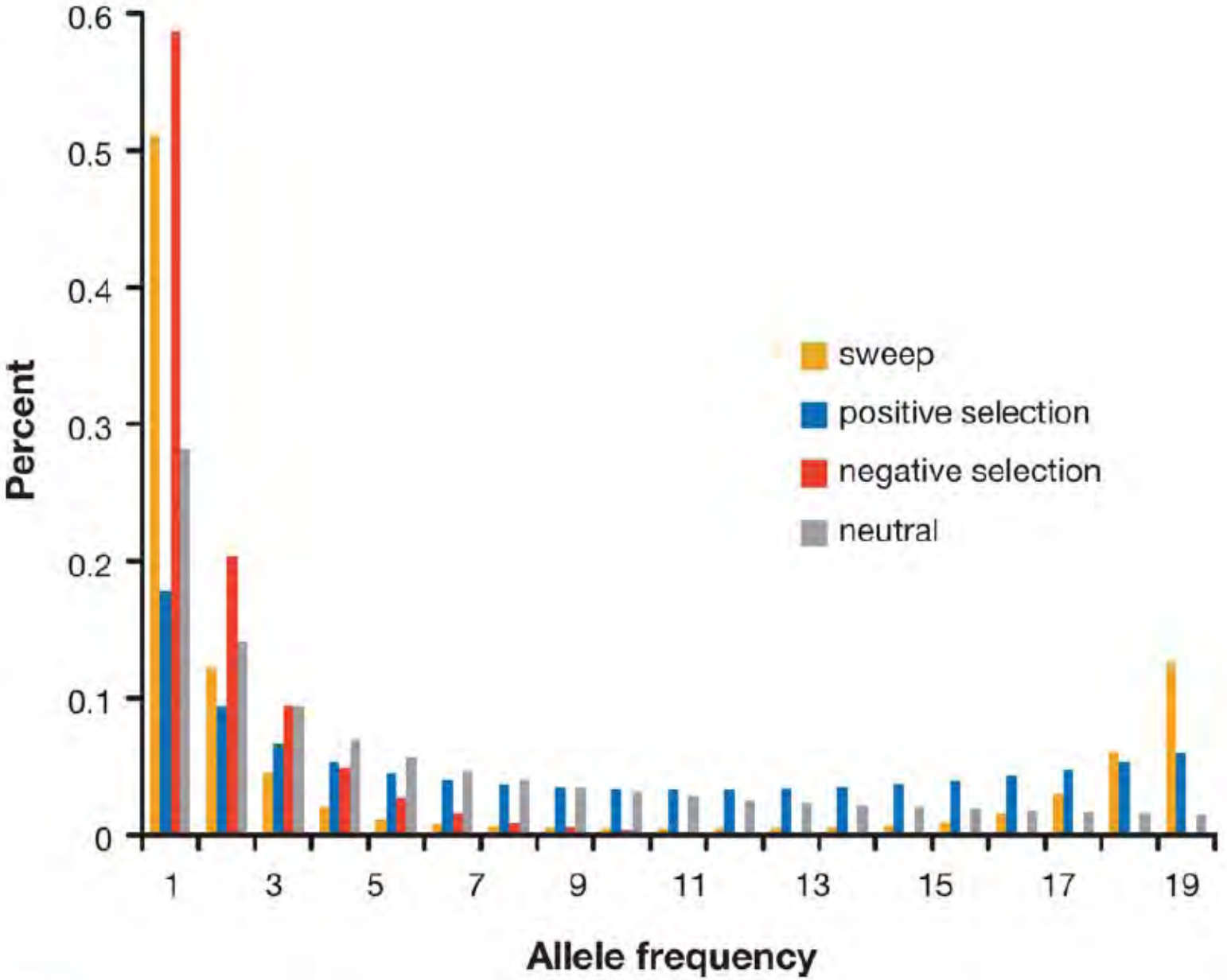


Figure 2

The frequency spectrum under a selective sweep, negative selection, neutrality, and positive selection. The frequency spectra under negative and positive selection are calculated using the PRF model by Sawyer & Hartl (88) for mutations with $2Ns = -5$ and 5, respectively, where N is the population size and s is the selection coefficient. For the selective sweep, the frequency spectrum is calculated in a window around the location of the adaptive mutation immediately after it has reached fixation in the population. In all cases, a demographic model of a population of constant size with no population subdivision is assumed.

Tajima's D test

One of the first, and most popular, polymorphism tests was Tajima's D test (Tajima 1989)

D contrasts estimates of θ based on S vs. k (π)

$$D = \frac{k - \theta_W}{\sqrt{Var\left(k - \theta_W\right)}} \Rightarrow D = \frac{\hat{\theta}_k - \hat{\theta}_S}{\sqrt{Var\left(\hat{\theta}_k - \hat{\theta}_S\right)}}$$

Under neutrality D=0

$$\theta_W = k = \theta = 4N_e\mu$$

Idea: For S we simply count sites, independent of their frequencies. Hence, S rather sensitive to changes in the frequency of rare alleles.

Major Complication With Polymorphism-based tests

Demographic factors can also cause these departures from neutral expectations!

Too many young alleles -> recent population expansion

Too many old alleles -> population substructure

Thus, there is a composite alternative hypothesis, so that rejection of the null does not imply selection. Rather, selection is just one option.

Can we overcome this problem?

It is an important one, as only polymorphism-based tests can indicate on-going selection

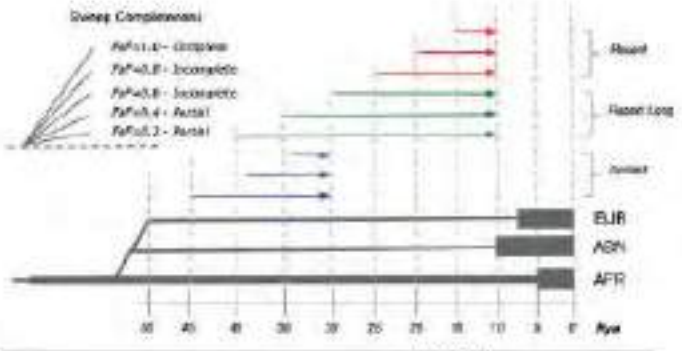
Solution: demographic events should leave a constant signature across the genome

Essentially, all loci experience common demographic factors

Genome scan approach: look at a large number of markers. These generate null distribution (most not under selection), outliers = potentially selected loci (genome wide polymorphism tests)

Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations

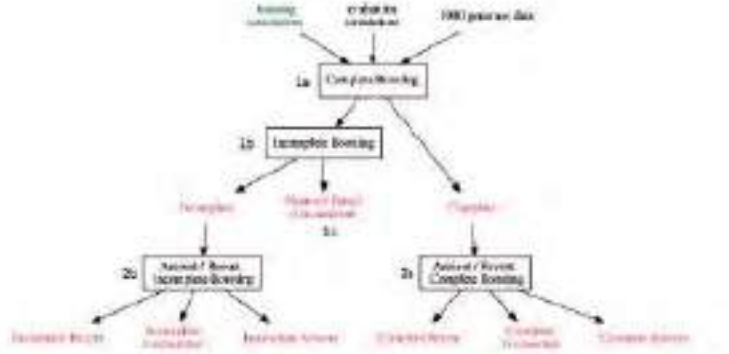
Marc Pybus[1,†], Pierre Luisi[1,2,†], Giovanni Marco Dall'Olio[1,3,†], Manu Uzkudun[1], Hafid Laayouni[1,4], Jaume Bertranpetit[1,*] and Johannes Engelken[1]



CLASSES	Final Allele Frequency	TYPES: Recent (25 to 10 Kya)	Recent Long (40 to 10 Kya)	Ancient (45 to 30 Kya)
Complete Sweep	1.0	300 replicates	300 replicates	300 replicates
Incomplete Sweep	0.8	300 replicates	300 replicates	300 replicates
	0.6	300 replicates	300 replicates	300 replicates
Partial Sweep	0.4	300 replicates	300 replicates	300 replicates
	0.2	300 replicates	300 replicates	300 replicates
Neutral		3600 replicates		

* This simulation scheme was duplicated for evaluation purposes.
* Selection simulations were run for each population demography independently

Fig. 1. Coalescent simulations were run following a calibrated human demographic model (Schaffner *et al.*, 2005) mimicking population genetic data from three reference continental populations (YRI, CEU and JPT/CHB). Nine different time-spanning selective sweeps were simulated (grouped as Neutral, Recent, Recent Long and Ancient) allowing for five different FAF (FAF = 0.2, 0.4, 0.6, 0.8 and 1.0)



Complete Boosting	Complete	ns	Incomplete **Partial** Neutral
Incomplete Boosting	Incomplete	ns	**Partial** Neutral
Ancient/Recent Complete Boosting	Complete Ancient	ns	Complete Recent
Ancient/Recent Incomplete Boosting	Incomplete Ancient	ns	Incomplete Recent

Fig. 2. The implemented 'Hierarchical Boosting' classification tree

EHH (Sabeti *et al.*, 2002b), dDAF (Hofer *et al.*, 2009), diHH (Voight *et al.*, 2006), Fay and Wu's *H* (Fay and Wu, 2000), Omega (Pavlidis *et al.*, 2010), EHH Av (Sabeti *et al.*, 2002b), Fu and Li's *D* (Fu and Li, 1993) and Tajima's *D* (Tajima, 1989).

> Nucleic Acids Res. 2018 Jan 4;46(D1):D1003-D1010. doi: 10.1093/nar/gkx943.

PopHuman: the human population genomics browser

Sònia Casillas [1], Roger Mulet [1], Pablo Villegas-Mirón [2], Sergi Hervas [1], Esteve Sanz [3], Daniel Velasco [1], Jaume Bertranpetit [2], Hafid Laayouni [2 4], Antonio Barbadilla [1 3]

Affiliations + expand

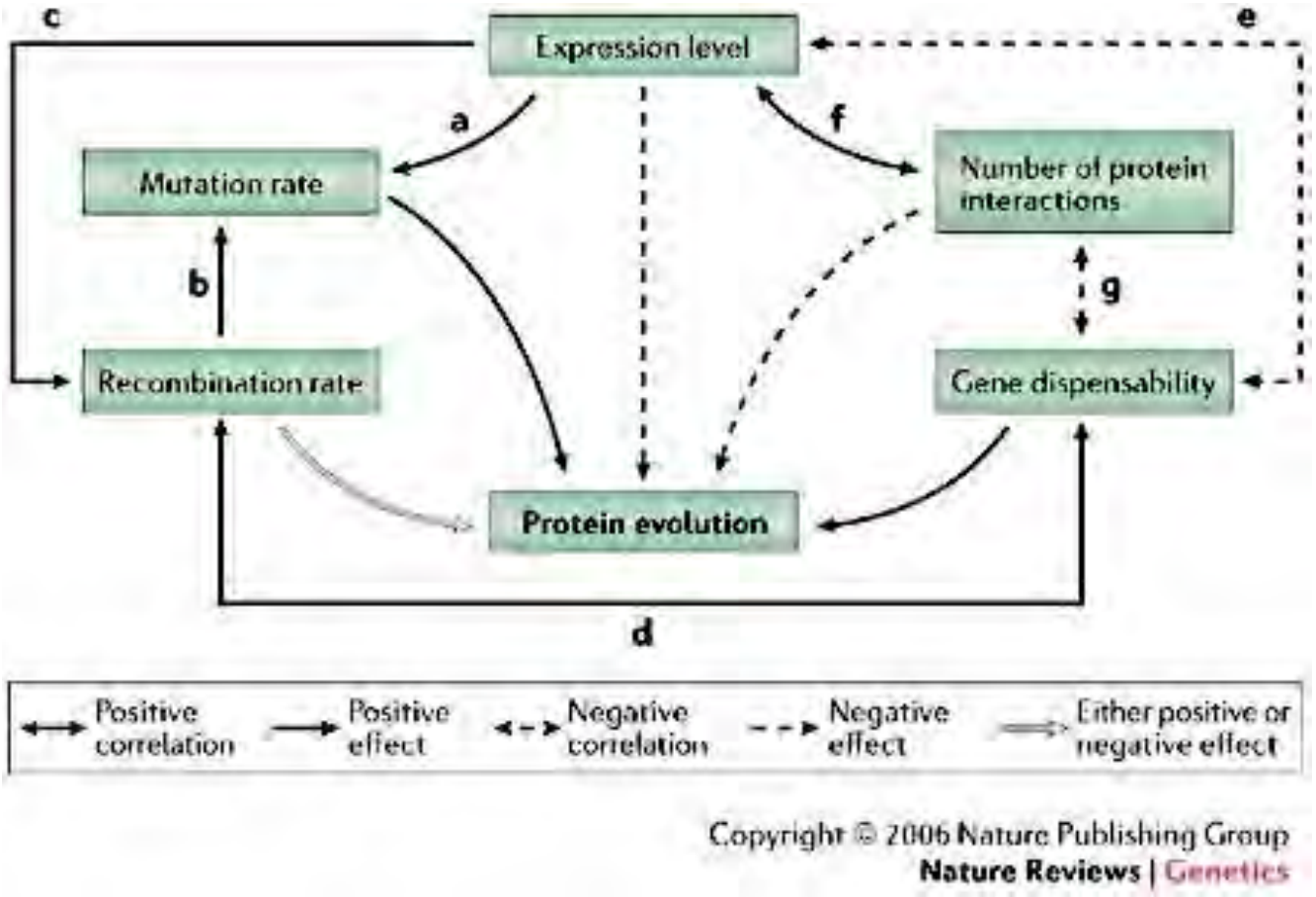
PMID: 29059408 PMCID: PMC5753332 DOI: 10.1093/nar/gkx943

Free PMC article

Abstract

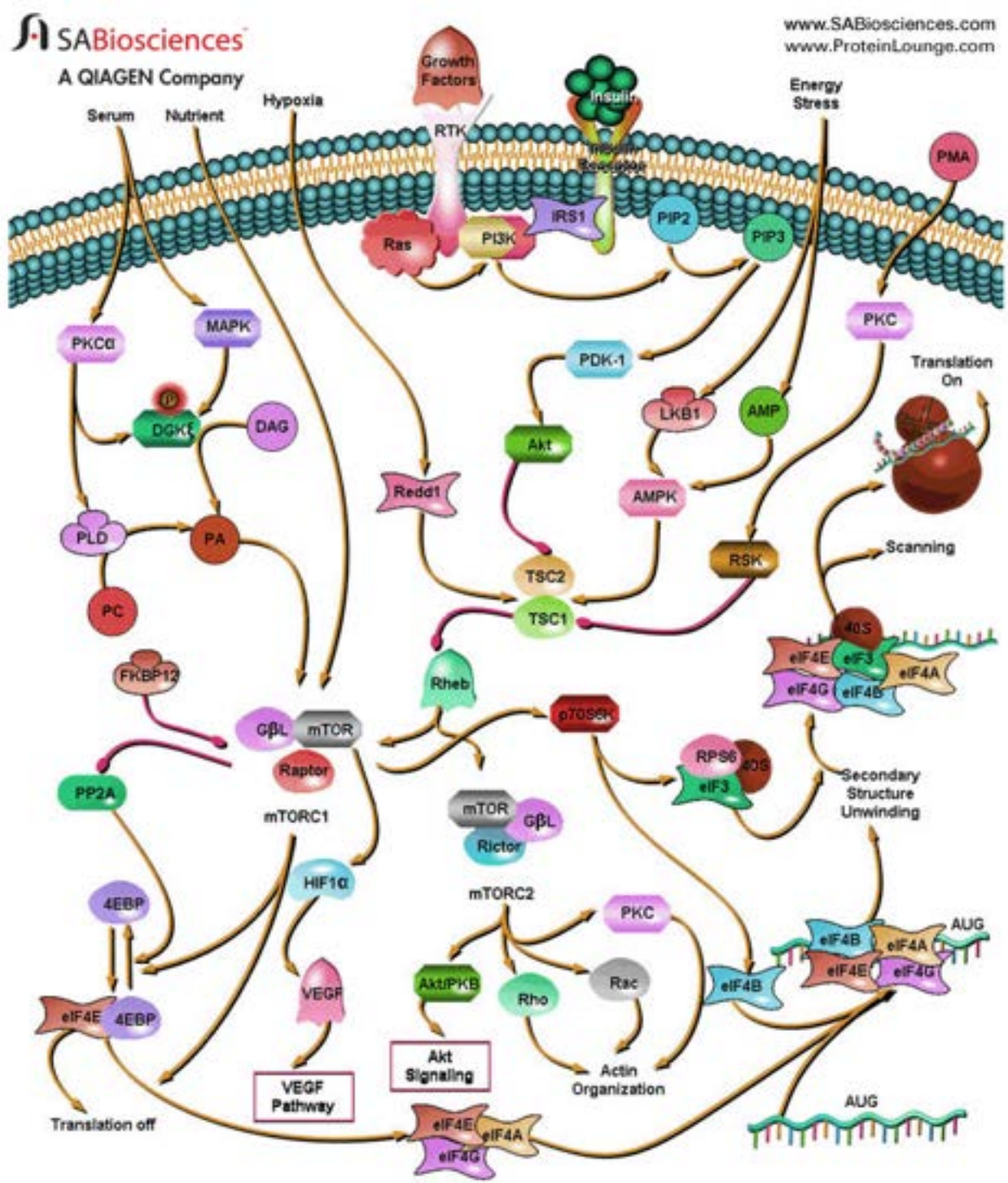
The 1000 Genomes Project (1000GP) represents the most comprehensive world-wide nucleotide variation data set so far in humans, providing the sequencing and analysis of 2504 genomes from 26 populations and reporting >84 million variants. The availability of this sequence data provides the human lineage with an invaluable resource for population genomics studies, allowing the testing of molecular population genetics hypotheses and eventually the understanding of the evolutionary of genetic variation in human populations. Here we present PopHuman, a new population oriented genome browser based on JBrowse that allows the interactive visualization and retrieval of an extensive inventory of population genetics metrics. Efficient and reliable parameter estimates have been computed using a novel pipeline that faces the unique features and limitations

Measuring selective pressures at **protein** level



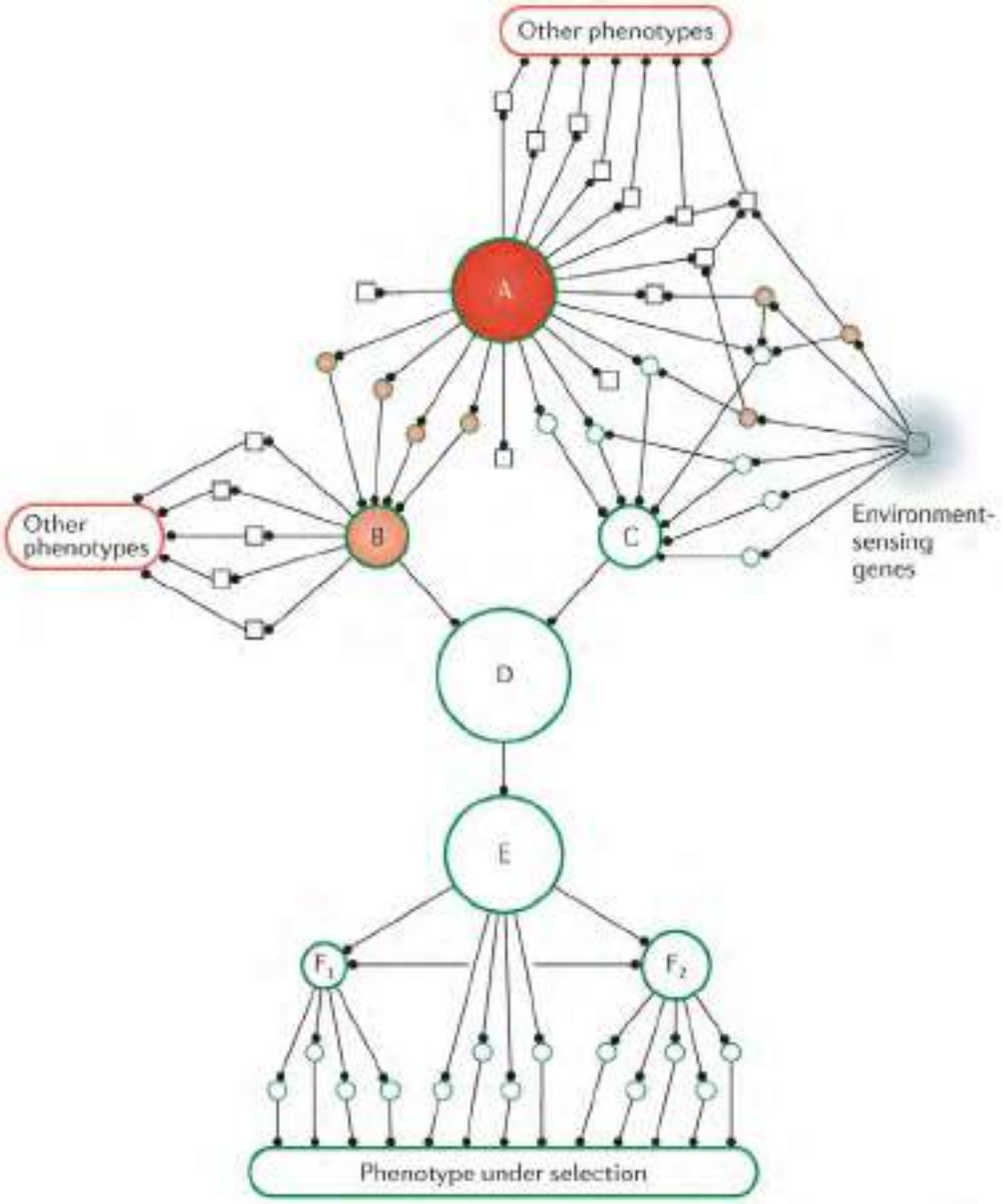
Pál *et al.* 2006 *Nature Reviews Genetics* **7**, 337–348

Why to care about Network topology when we study Natural Selection?



www.sabiosciences.com/pathway.php?sn=mTOR_Pathway

Importance of network effects for adaptation



Nature Reviews Genetics 13, 867-877 (December 2012)

Goal

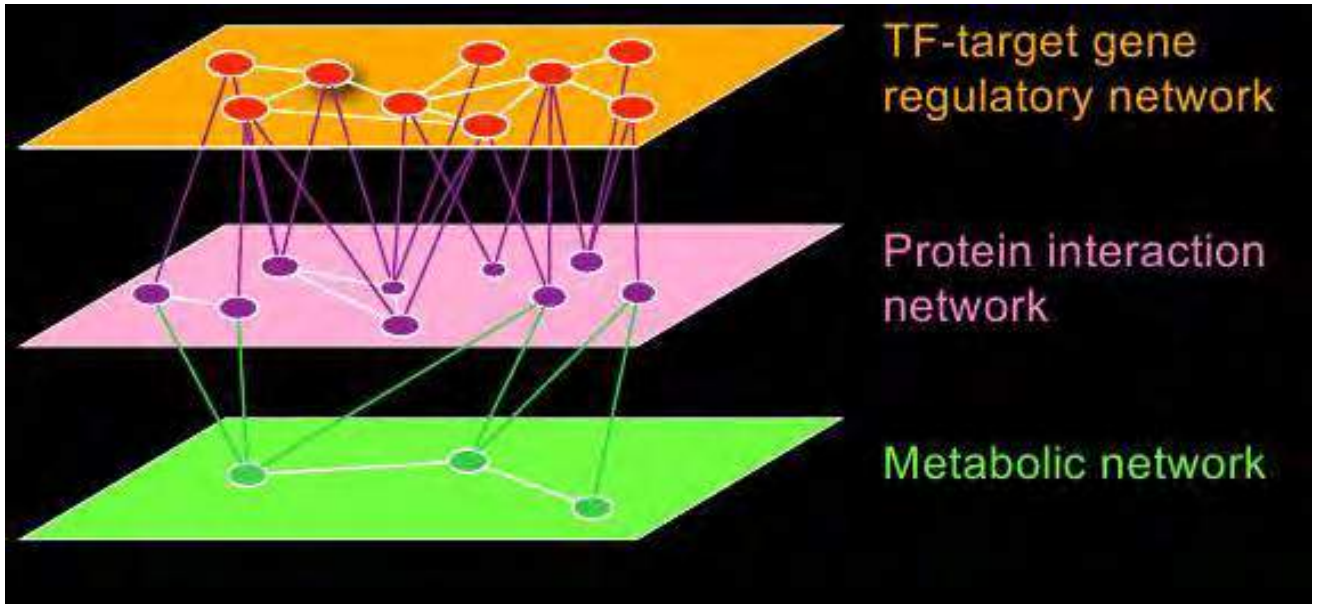
- Analyze the relationship between signals of selection and the topology of the network to which genes belong.
- Do central genes evolve faster / slower than peripheral genes?
- Constraints on the rates of evolution given the position of a gene in a network?

Network Representation of Biological Interactions

(1) Which types of interactions to consider

(2) How to represent them

(3) Large-scale / Small-scale networks



Small scale vs. genome wide scale networks?

- Small scale
  - small number of genes functionally related .
  - interactions are determined on the basis of established biological knowledge on the process.
  - likely to reveal pathway specific patterns that may not be generalized
  - lack statistical power to detect correlations with topological parameters
- GW scale Network
  - High throughput techniques, global perspectives.
  - hardly interpretable in the light of specific biological functions
  - Low quality

Covariates to be taken into account

- Highly expressed proteins tend to evolve slowly (Drummond et al. 2006).
- Translational robustness hypothesis: low rate of translational errors will constrain sequence evolution.
- Protein length, protein structure, essentiality and protein dispensability (Bloom et al. 2006)

Mol Biol Evol. 2012 May;29(5):1379-92. doi: 10.1093/molbev/msr298. Epub 2011 Dec 1.

Network-level and population genetics analysis of the insu transduction pathway across human populations.

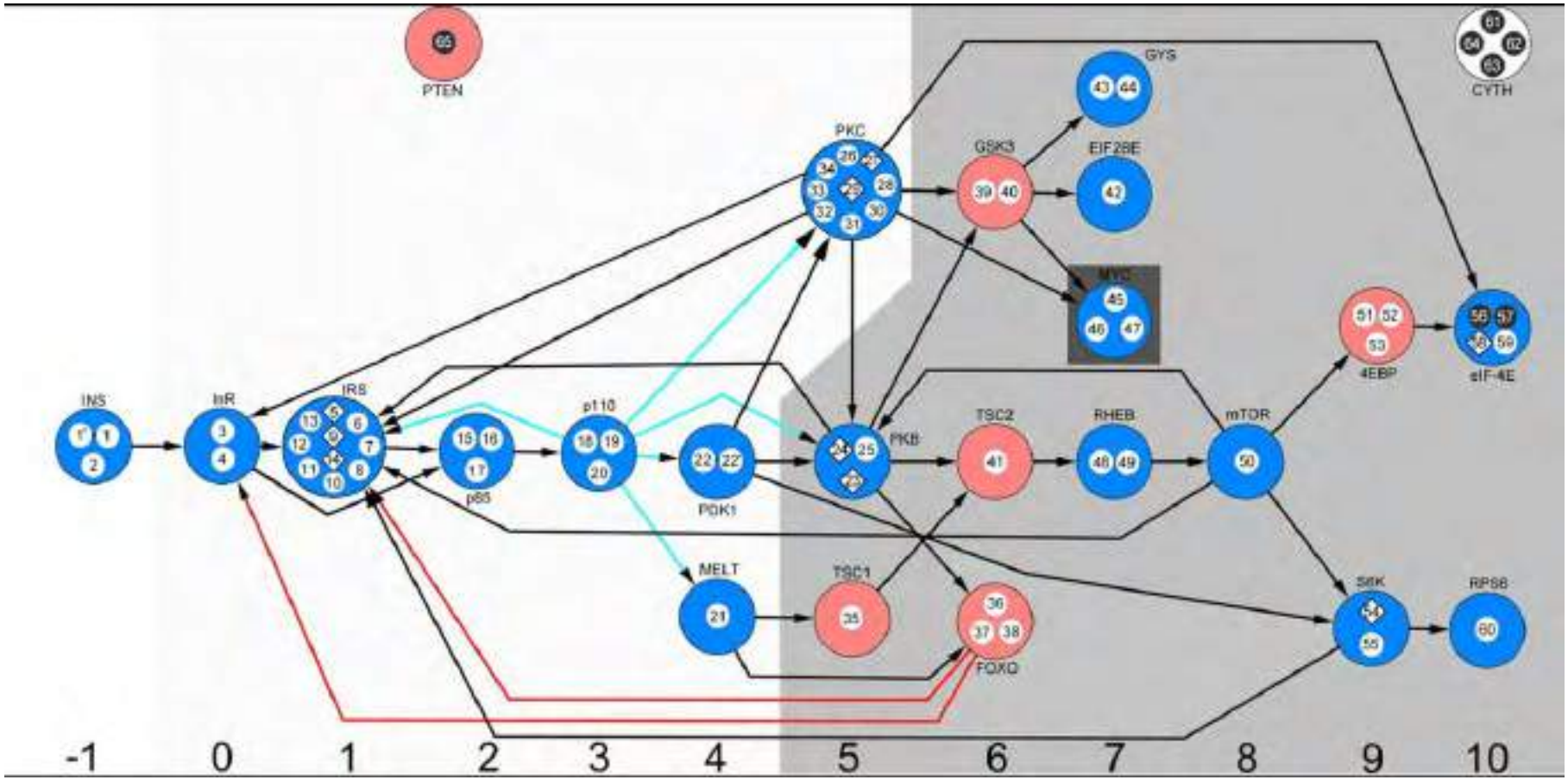
Luisi P, Alvarez-Ponce D, Dall'Olio GM, Sikora M, Bertranpetit J, Laayouni H.

Institute of Evolutionary Biology CEXS-UPF-PRBB, Barcelona, Catalonia, Spain.

Abstract

Genes and proteins rarely act in isolation, but they rather operate as compone

Structure of the insulin/TOR signal transduction pathway.



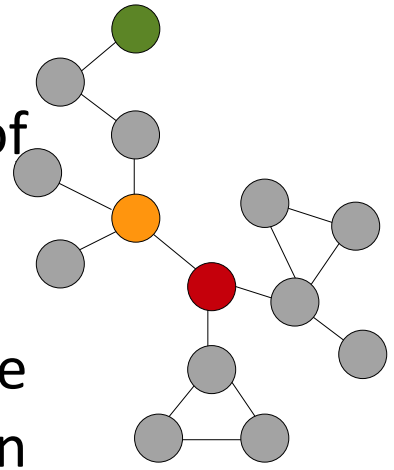
Luisi P et al. Mol Biol Evol 2012;29:1379-1392

MBE MOLECULAR BIOLOGY AND EVOLUTION

Distribution of selection over biomolecular networks

Centrality measures

To characterize the position of a node with respect of the whole network:



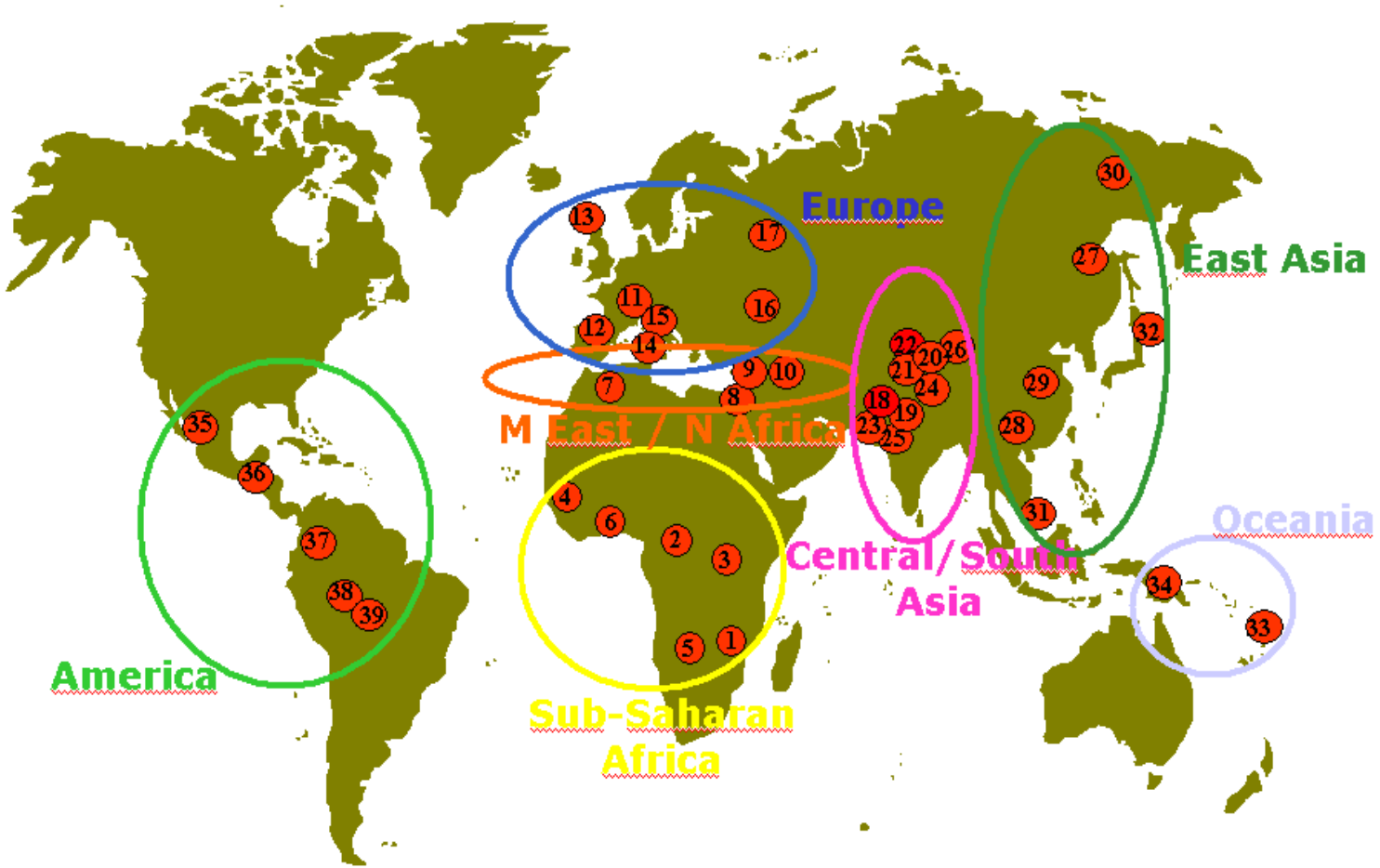
- **Degree centrality** of a node is defined as the number of its connections (also named connectivity of a node).

- **Betweenness centrality** of a node is defined as the fraction of all shortest paths between all pairs of nodes in the network that passes through that node. (“bridge” positions).

- **Closeness centrality** of a node is defined as its average distance between all the nodes in the network. (center or in the periphery)

Genotyping data: HGDP populations

- Samples: 1049 individuals grouped in 39 populations



Population differentiation – Method

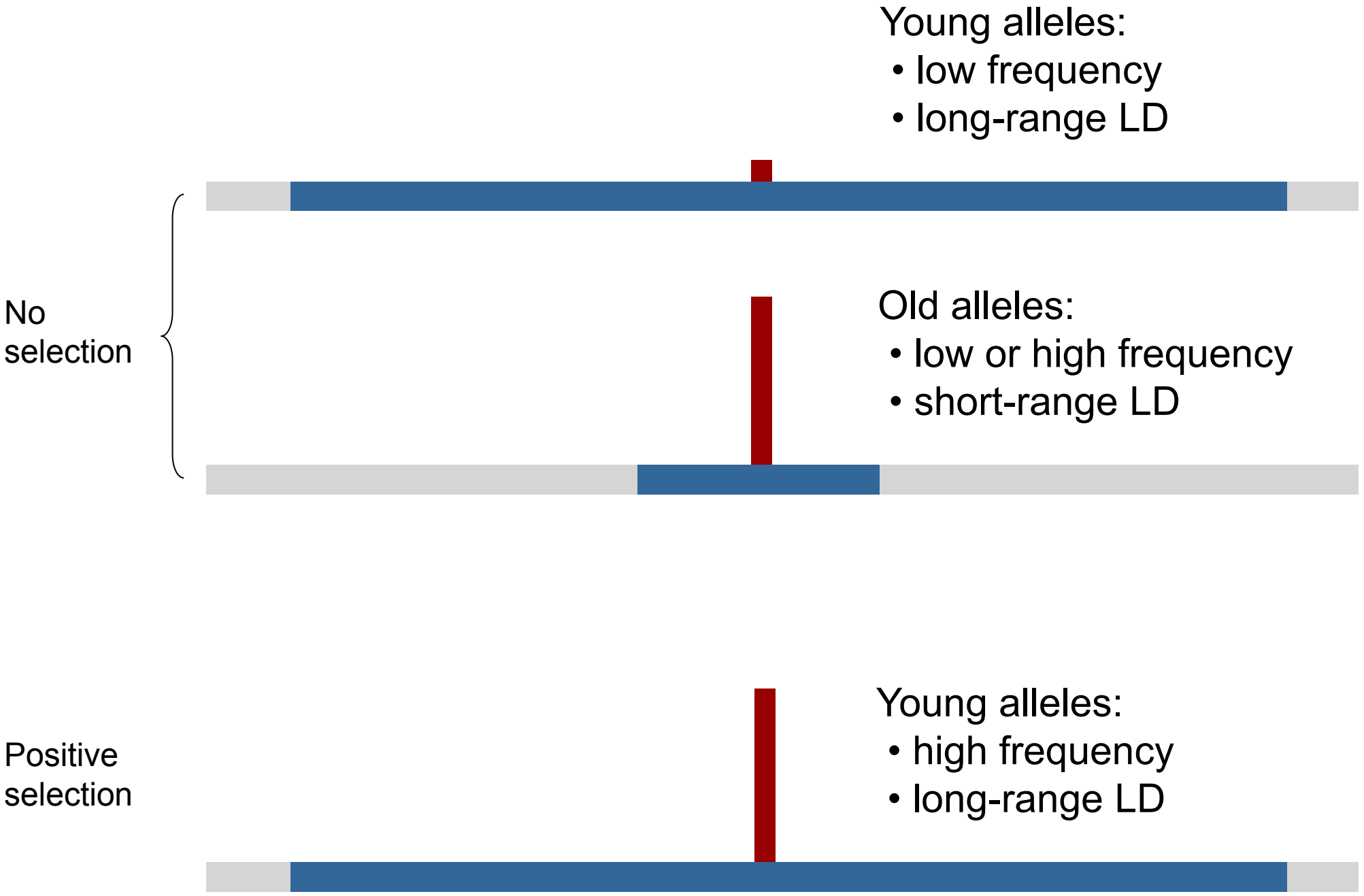
Principle

Local adaptation causes differences in allele frequencies between populations

Method

Calculate molecular fixation index F_{ST} and identify regions with extreme values

Long range haplotypes – Method



SNPs to genes– Method

3 methods based on

- Haplotype Structure: iHS
- genetic differentiation: Fst, dDAF (differences in Derived Allele Frequenc among populations)

Compute empirical p-values

- outlier approach from GW distribution
- Correction for Minor Allele Frequency

Combine at genomic region level

- Fisher combination

$$-2\sum_{i=1}^{K}\log P_i \sim \chi^2_{2K}$$

Table 4. Relationship between the Structure of the Insulin/TOR Pathway and the Impact of Positive Selection.

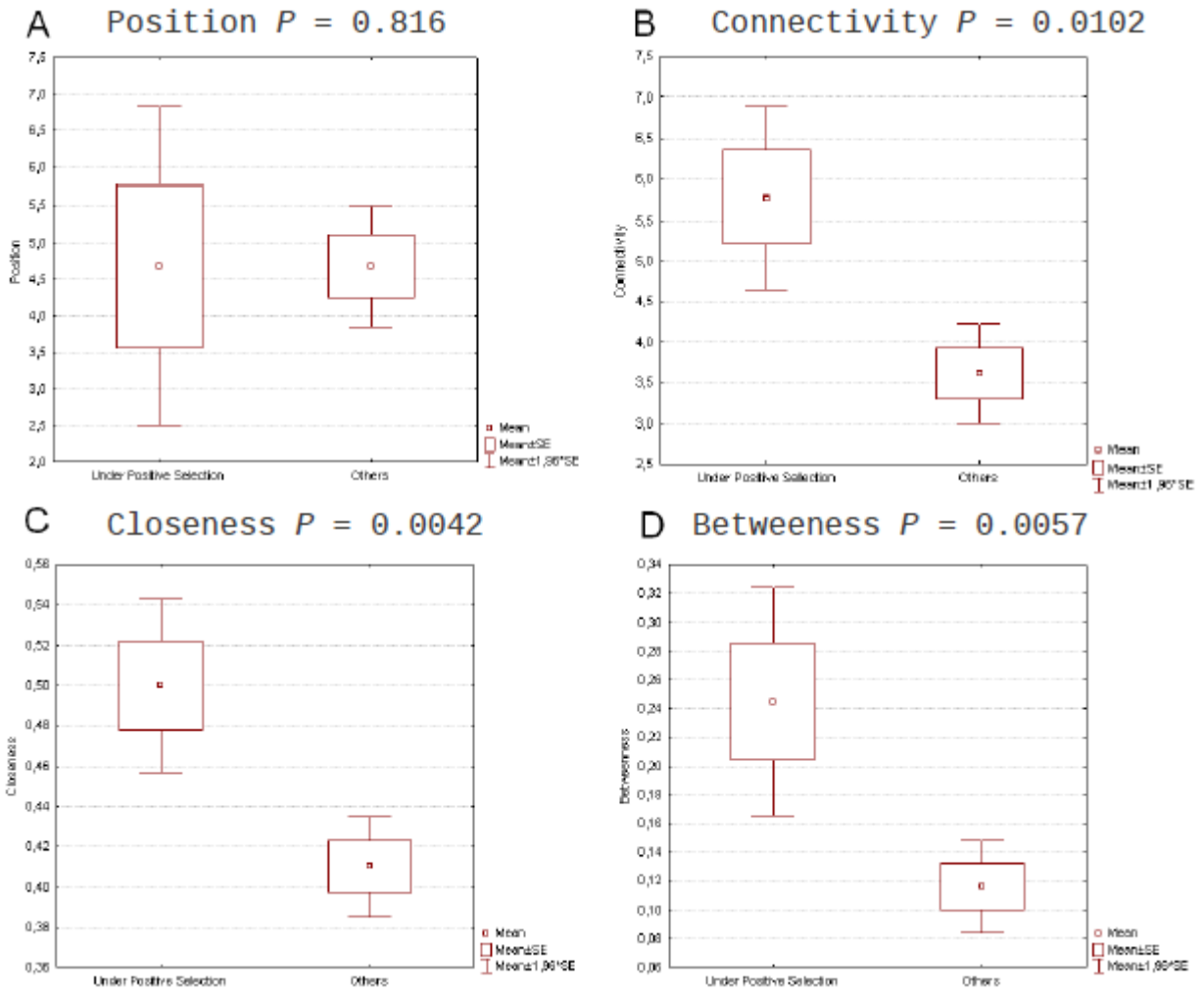
				Mann-Whitney test		Partial correlation[a]	
Network[b]	**Parameter**	**Mean for genes under positive selection**	**Mean for genes without signal of positive selection**	***U***	***P***	ρ	***P***
-	Position	4.667	4.673	209.5	0.816	-0.203	0.167
(i)	Connectivity	5.778	3.612	103	0.010*	0.397	0.005**
	Closeness	0.500	0.410	91	0.004**	0.417	0.003**
	Betweenness	0.244	0.116	95.5	0.006**	0.412	0.004**
(ii)	Connectivity	6.889	4.571	116	0.024*	0.432	0.002**
	Closeness	0.536	0.455	101	0.009**	0.450	0.001**
	Betweenness	0.188	0.098	101.5	0.009**	0.419	0.003**
(iii)	Connectivity	35.444	38.087	186	0.646	-0.132	0.372
	Closeness	0.291	0.294	179	0.538	-0.170	0.250
	Betweenness	0.0115	0.006	153	0.228	-0.074	0.618

[a]Spearman's partial correlation between the impact of positive selection and network parameters controlling for gene expression level and breadth and length of the encoded proteins (see *Materials and Methods*).

[b]Three different sets of interactions were used for centrality calculations: (i) protein-protein interactions within the insulin/TOR pathway; (ii) all kinds of interactions within the pathway (i.e., protein-protein, metabolic, and transcriptional activation interactions); and (iii) protein-protein interactions from the whole interactome.

* $P < 0.05$; ** $P < 0.01$.

Comparison between genes evolving under positive selection and the remaining ones.



Luisi P et al. Mol Biol Evol 2012;29:1379-1392

MBE MOLECULAR BIOLOGY AND EVOLUTION

Conclusion

- We found that positive selection preferentially targets the most central elements in the pathway, in contrast to previous observations in the whole human interactome. This observation indicates that the impact of positive selection on genes involved in the insulin/TOR pathway is affected by the pathway structure.

BMC Evol Biol. 2012 Jun 25;12:98. doi: 10.1186/1471-2148-12-98.

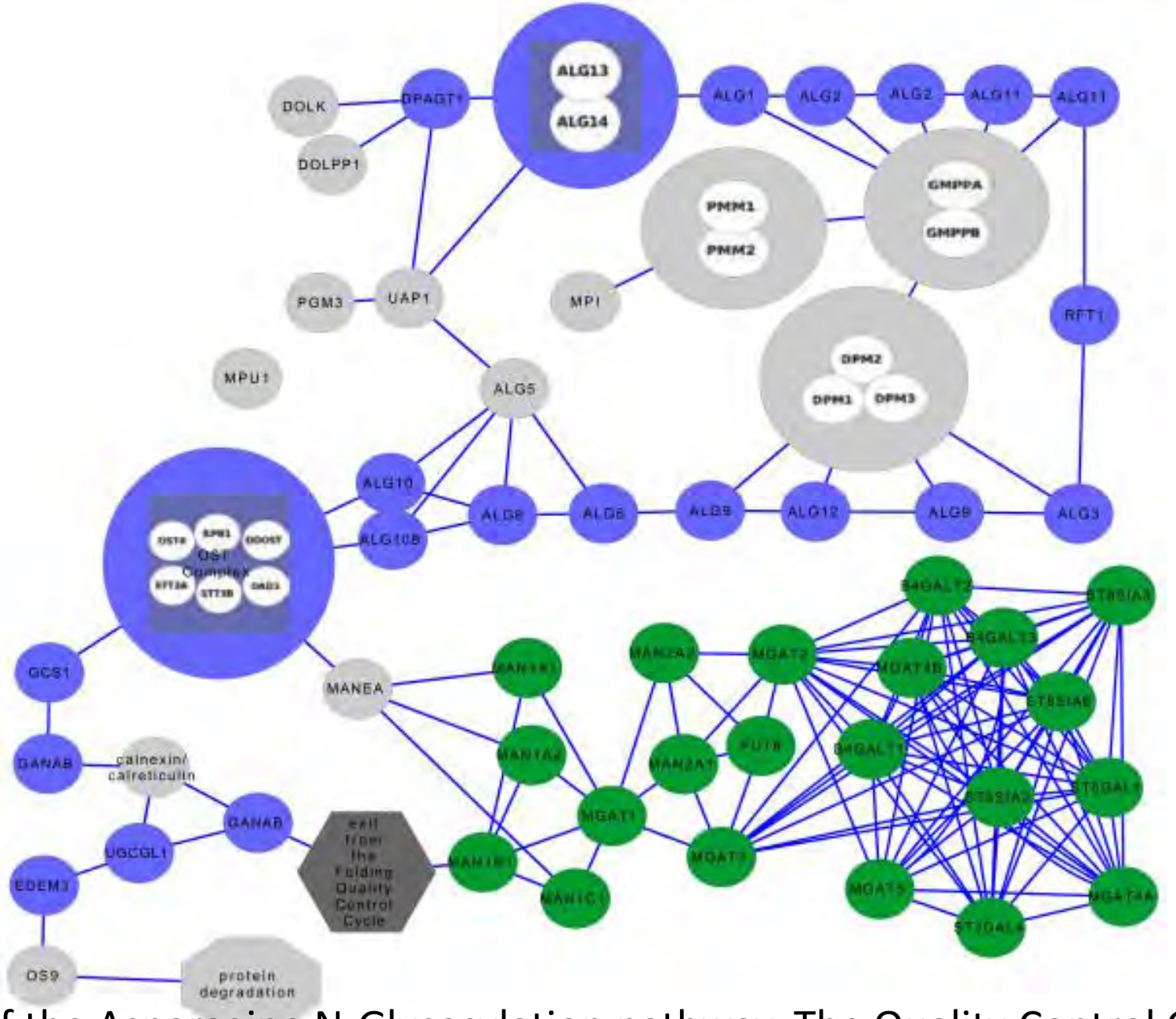
Distribution of events of positive selection and popula
a metabolic pathway: the case of asparagine N-glycosy

Dall'Olio GM, Laayouni H, Luisi P, Sikora M, Montanucci L, Bertranpetit J.

IBE, Institut de Biologia Evolutiva (UPF-CSIC), Parc de Recerca Biomèdica de Barcelo
08003, Barcelona, Catalonia, Spain.

Abstract

BACKGROUND: Asparagine N-Glycosylation is one of the most importan



Overview of the Asparagine N-Glycosylation pathway. The Quality Control Cycle, which divides the pathway into two parts, is shown as an octagon. Genes classified as ‘upstream’ in the analysis are in blue; genes classified as ‘downstream’ are in green. Genes in gray have been excluded from the network analysis

Glycobiology. 2011 Nov;21(11):1395-400. doi: 10.1093/glycob/cwq215. Epub 2011 Jan 2.

The annotation of the asparagine N-linked glycosylation pathway in the Reactome database.

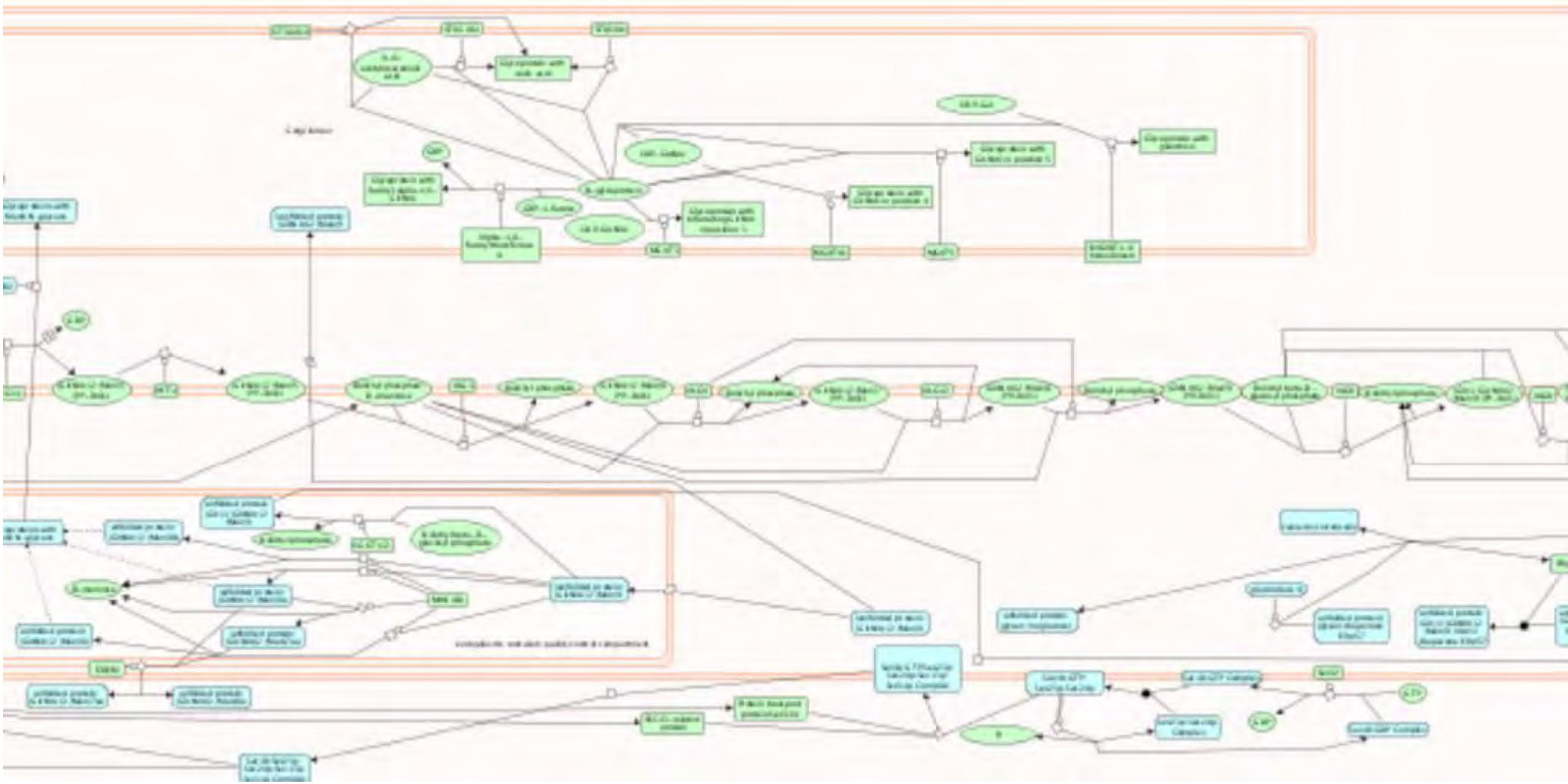
Dall'Olio GM, Jassal B, Montanucci L, Gagneux P, Bertranpetit J, Laayouni H.

Institute of Evolutionary Biology, Carrer Doctor Aiguader 88, Barcelona, Catalonia, Spain.

Abstract

Asparagine N-linked glycosylation is one of the most important forms of protein post-translational modification in eukaryotes and is one of the first metabolic pathways described at a biochemical level. Here, we report a new annotation of this pathway for the Human species, published after passing a peer-review process in Reactome. The new annotation presented here offers a high level

REACTOME



Database (Oxford). 2010 Dec 23;2010:baq035. Print 2010.

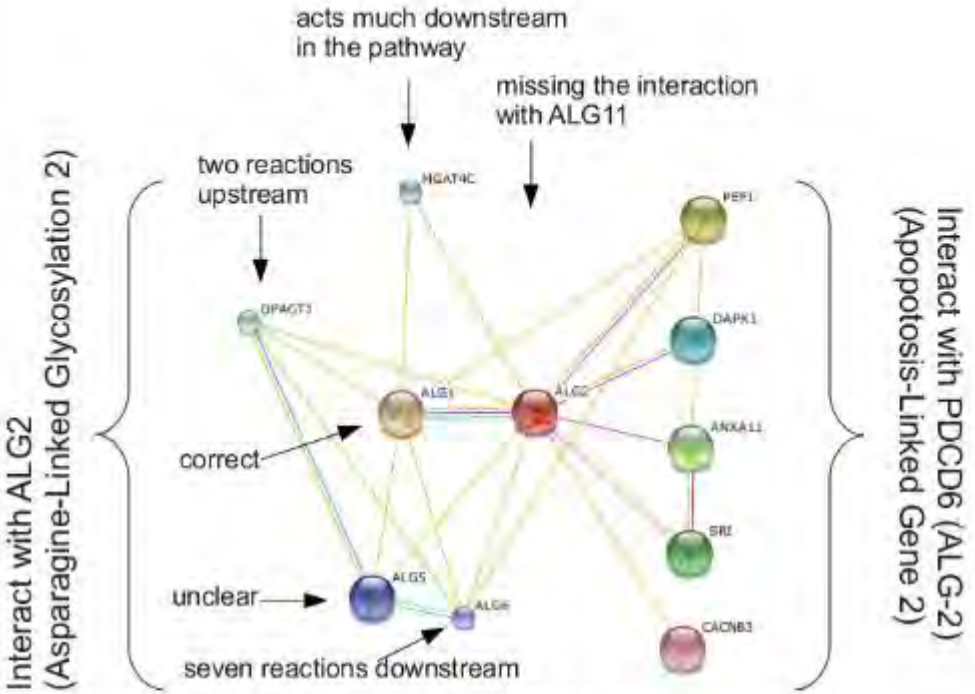
The annotation and the usage of scientific databases could be improved with public issue tracker software.

Dall'Olio GM, Bertranpetit J, Laayouni H.

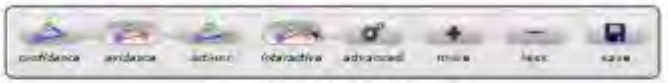
Institute of Evolutionary Biology, UPF-CSIC, CEXS-UPF-PRBB, Barcelona, Catalonia, Spain.

Abstract

Since the publication of their longtime predecessor The Atlas of Protein Sequences and Structures i 1965 by Margaret Dayhoff, scientific databases have become a key factor in the organization of modern science. All the information and knowledge described in the novel scientific literature is translated into entries in many different scientific databases, making it possible to obtain very accurate information on a biological entity like genes or proteins without having to manually review



This is the **evidence view**. Different line colors represent the types of evidence for the association



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	Homology	Score
ALG2	Alpha-1,3-mannosyltransferase ALG2 (EC 2.4.1.-) (GDP- Man-Man(1)GlcNAc(2)-PP-dolichol mannosyltransferase); Mannosylates Man(2)GlcNAc(2)-dolichol diphosphate and Man(1)GlcNAc(2)-dolichol diphosphate to form Man(3)GlcNAc(2)- dolichol diphosphate (416 aa) *(Homo sapiens)*									

Predicted Functional Partners:

ALG1	Chitobiosyldiphosphodolichol beta-mannosyltransferase (EC 2.4.1.142) (GDP-mannose-dolichol diph [...] (464 aa)						•	•	•	0.990
PEF1	Peflin (PEF protein with a long N-terminal hydrophobic domain) (Penta- EF hand domain-containin [...] (284 aa)						•		•	0.905
ANXA11	Annexin A11 (Annexin XI) (Calcyclin-associated annexin 50) (CAP-50) (56 kDa autoantigen) (505 aa)						•			0.849
SRI	Sorcin (22 kDa protein) (CP-22) (V19) (198 aa)						•		•	0.820
DPAGT1	UDP-N-acetylglucosamine--dolichyl-phosphate N- acetylglucosaminephosphotransferase (EC 2.7.8.15 [...] (408 aa)								•	0.802
DAPK1	Death-associated protein kinase 1 (EC 2.7.11.1) (DAP kinase 1); Calcium/calmodulin dependent se [...] (1455 aa)						•		•	0.795
HGAT4C	UDP-N-acetylglucosamine- alpha-1,3-D-mannoside beta-1,4-N-acetylglucosaminyltransferase IV; Gly [...] (478 aa)								•	0.762
ALG5	Dolichyl-phosphate beta-glucosyltransferase (EC 2.4.1.117) (DolP- glucosyltransferase) (324 aa)								•	0.700
ALG6	Dolichyl pyrophosphate Man9GlcNAc2 alpha-1,3-glucosyltransferase (EC 2.4.1.-) (Dolichyl-P-Glc-M [...] (511 aa)								•	0.688
CACNB3	Voltage-dependent L-type calcium channel subunit beta-3 (CAB3) (Calcium channel voltage-depende [...] (484 aa)								•	0.651

Recent Positive selection targets the center of the human protein-protein interaction network

Pierre Luisi, David Alvarez-Ponce, Marc Pybus, Mario A. Fares, Jaume Bertranpetit and Hafid Laayouni

Motivation - Objectives

- Analyze the relationship between signals of selection (Positive and Purifying) and the topology of the Protein interaction network.
- Do central genes evolve faster / slower than peripheral genes?
- Constraints on the rates of evolution given the position of a gene in a network?


BioGRID | Database of Protein and Genetic Interactions - Mozilla Firefox
thebiogrid.org
BioGRID 3.1
home help wiki tools contribute statistics downloads partners about us
Welcome to the Biological General Repository for Interaction Datasets
BioGRID is an online interaction repository with data compiled through comprehensive curation efforts. Our current index is version 3.1.94 and searches 36,076 publications for 564,472 raw protein and genetic interactions from major model organism species. All interaction data are freely provided through our search index and available via download in a wide variety of standardized formats.
INTERACTION STATISTICS
LATEST DOWNLOADS
Search the BioGRID
Search by identifiers, keywords, and gene names...
All Organisms
SUBMIT GENE SEARCH
By Gene
By Publication
Advanced Search
Search Tips
Featured Datasets
AREAS OF INTEREST TO HELP YOU GET STARTED
Build and Download Interaction Datasets
Create custom interaction datasets by protein or by publication. You can also download our entire dataset in a wide variety of standard formats.
Link To Us or Submit Interactions
Send us your datasets or link to the BioGRID directly from your own website or database. Full details on how to contribute are available here.
Online Tools and Resources
We've developed tools that make use of BioGRID data. Check out the list of tools to see if we can help you work with our data.
View Our Interaction Statistics
Find out how many organisms, proteins, publications, and interactions are available in the current release of the BioGRID.
BIOGRID FUNDING AND PARTNERS
CIHR IRSC
BBSRC
MOUNT SINAI HOSPITAL
PRINCETON UNIVERSITY
University of Edinburgh
SGD
LAMHDI
IMEx
more partners
LATEST NEWS

Material And Methods: Interaction data

PIN0: Protein Interaction Network from BioGrid 3.1.81 (Stark et al. 2011) for Human

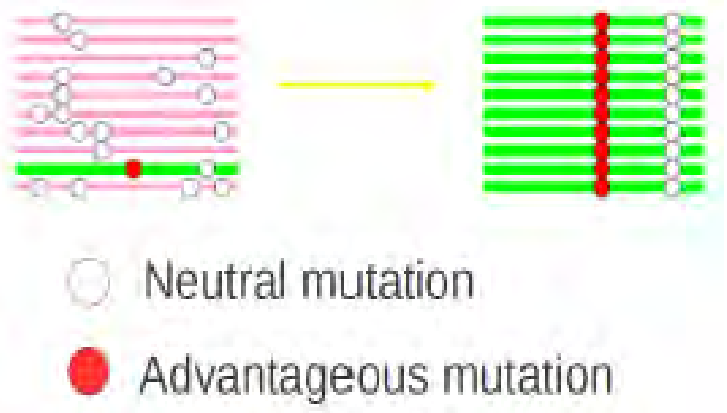
- Curated interactions supported by published experimental evidence
- Use only non-redundant physical interactions among pairs of proteins with an Ensembl ID

==> 9046 proteins connected by 39,546 interactions

Degree, betweenness and closeness centralities computed using the NetworkX package

Material and Methods: Detect positive selection within humans

3 Tests based on:
- LD (iHS; Voight et al. 2006)
- Genetic differentiation (XP-CLR; Chen et al. 2010)
- SFS (DH; Zeng 2006)



Combine at genomic region level
- Fisher combination

$$-2\sum_{i=1}^{K}\log P_i \sim \chi^2_{2K}$$

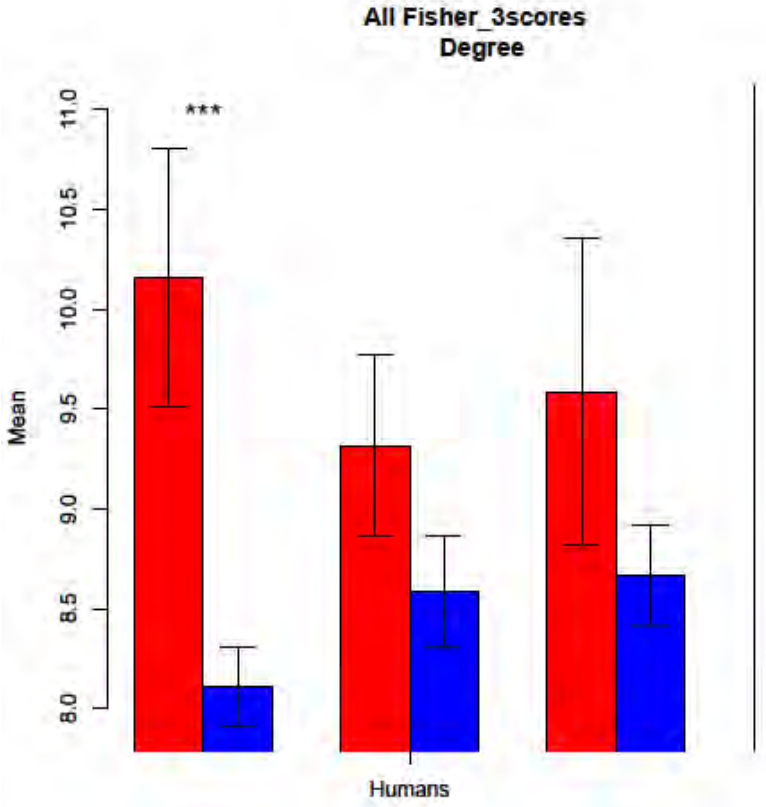
Results: Relationship between degree and the impact of purifying selection

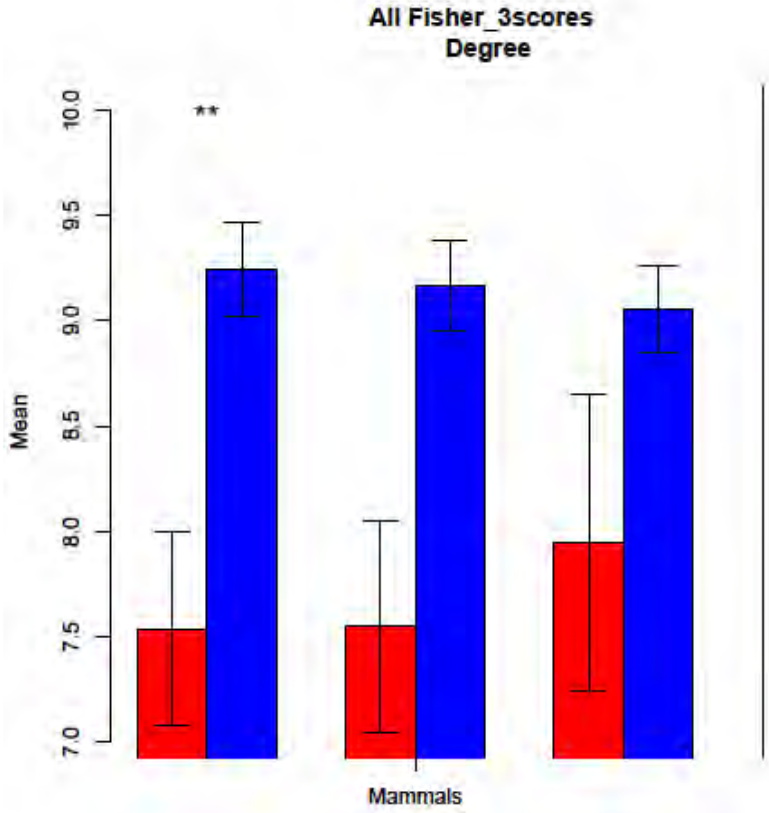
		Purifying Selection	
		Humans	Mammals
Spearman Correlation[a]	ρ	-0.0875	-0.2035
	P-value	5.75×10^{-16}***	1.14×10^{-55}***
Partial Spearman Correlation[b]	ρ	-0.0665	-0.1692
	P-value	2.69×10^{-09}***	5.01×10^{-37}***
ANOVA[c]	*F*	17.208	52.97
	P-value	3.88×10^{-11}***	9.05×10^{-34}***
Trend Test[c]	*F*	48.886	158.6
	P-value	2.92×10^{-12}***	6.73×10^{-36}***
Partial ANOVA[b]	*F*	5.314	40.23
	P-value	0.0012**	1.03×10^{-25}***
Partial Trend Test[b]	*F*	12.78	120.6
	P-value	0.0004***	8.77×10^{-28}***

Result: Relationship between degree and the impact of positive selection

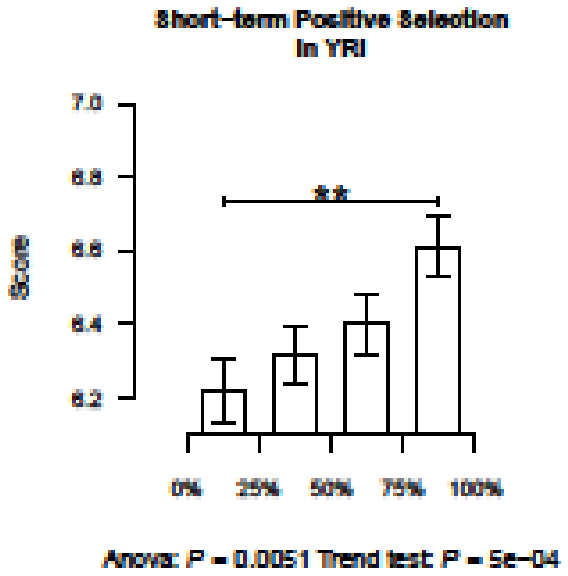
		Positive Selection			
		YRI	CEU	CHB	Mammals
Spearman Correlation[a]	ρ	0.0503	0.0411	0.0469	-0.0846
	P-value	1.04×10^{-5}***	0.0003***	3.84×10^{-5}***	9.37×10^{-11}***
Partial Spearman Correlation[b]	ρ	0.0453	0.0327	0.0372	-0.0424
	P-value	0.0001***	0.0057**	0.0016**	0.0016**
ANOVA[c]	*F*	4.746	3.977	3.880	1.2662
	P-value	0.0026**	0.0077**	0.0088**	0.2841
Trend Test[c]	*F*	13.78	5.633	9.734	3.576
	P-value	0.0002***	0.0177*	0.0018**	0.0587
Partial ANOVA[b]	*F*	2.639	2.611	2.161	1.711
	P-value	0.0478*	0.0497*	0.0904	0.1624
Partial Trend Test[b]	*F*	7.223	0.7061	3.990	4.475
	P-value	0.0072**	0.4008	0.0458*	0.0344*

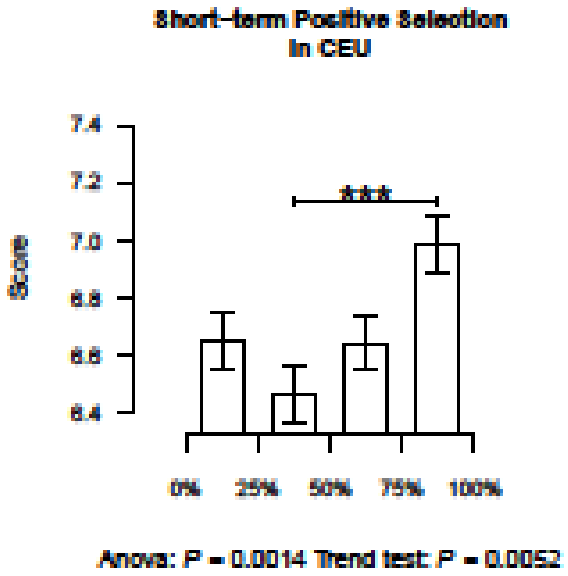
Result: Distribution of genes with putative signals of positive selection within the Protein Interaction Network.

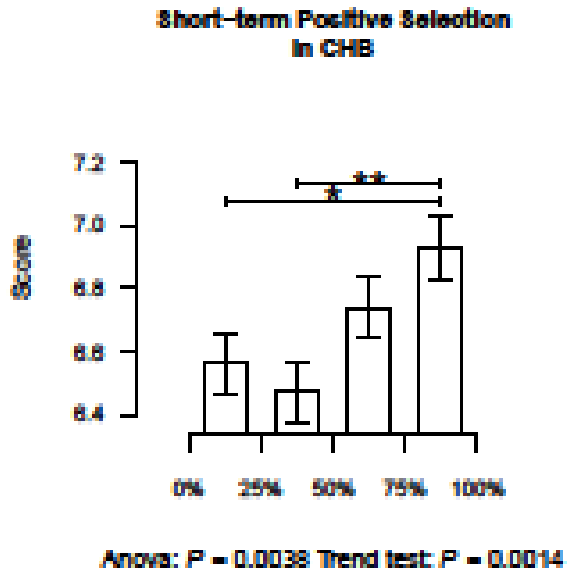


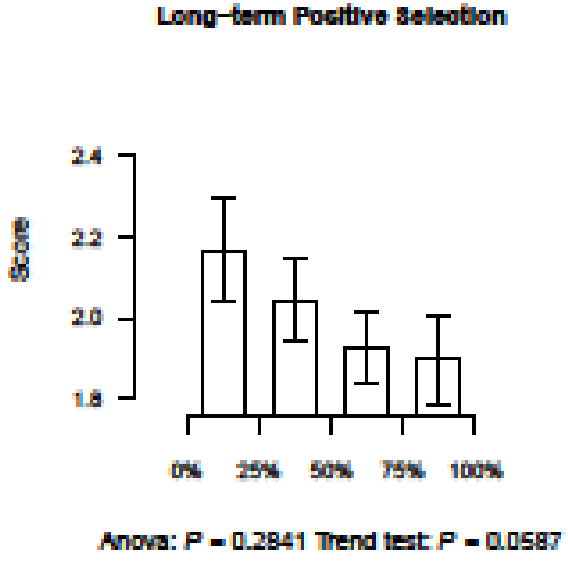


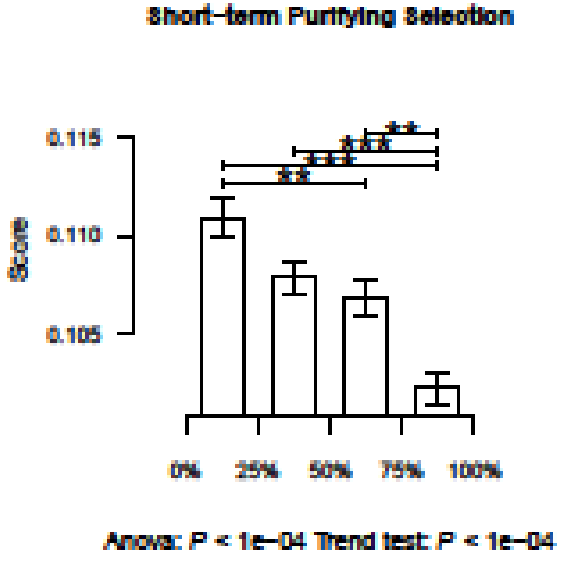
Result: Impact of natural selection among groups of genes divided according to the degree quartiles.

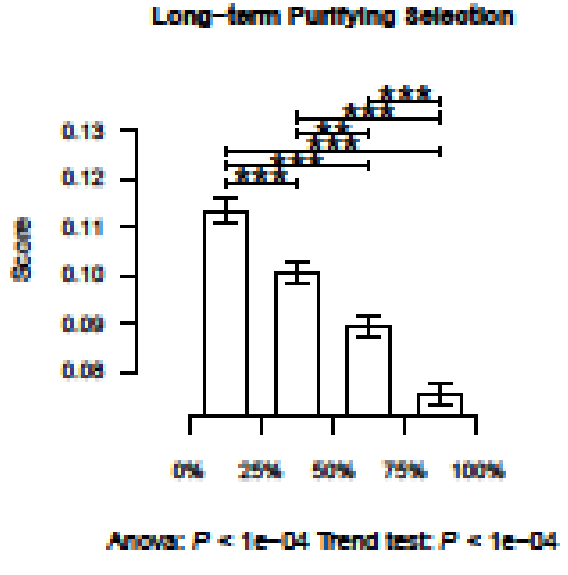












Relationship between degree and the impact of positive selection in humans correcting for ω in Mammals

		YRI	CEU	CHB
Spearman Correlation[a]	ρ	0.0433	0.0194	0.0412
	P-value	0.0020**	0.1653	0.0031**
ANOVA[b]	*F*	2.477	2.495	2.482
	P-value	0.0595	0.0580	0.0590
Trend Test[b]	*F*	6.716	0.5799	5.182
	P-value	0.0096**	0.4464	0.0228*

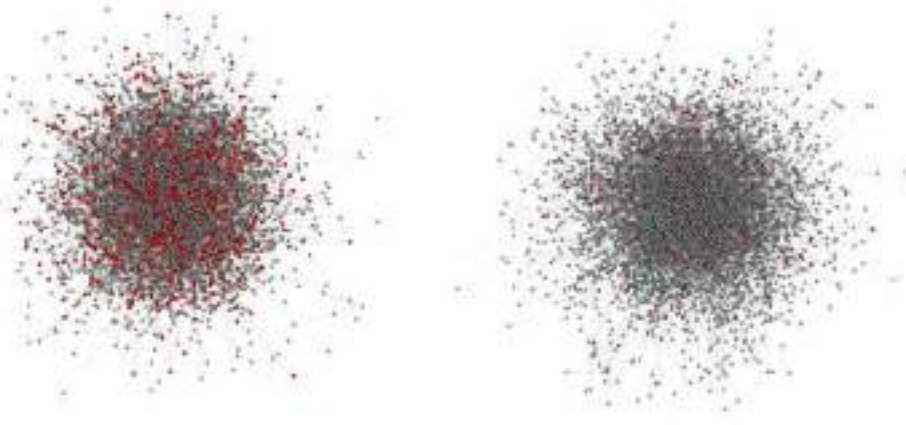
Association between gene essentiality and degree and the impact of positive and purifying selection

	Indispensability Score[b]	
	ρ	*P*-value
Degree	0.2311	3.02×10^{-107}***
Positive selection in YRI[c]	0.0473	4.34×10^{-05}***
Positive selection in CEU[c]	0.0695	2.01×10^{-09}***
Positive selection in CHB[c]	0.0379	0.0010**
Positive selection in Mammals[d]	-0.1373	1.79×10^{-25}***
Purifying selection in Humans[e]	-0.1131	5.14×10^{-25}***
Purifying selection in Mammals[f]	-0.2452	5.75×10^{-79}***

Final remarks

Central genes are more prone to the action of short term Positive Selection.

Short term positive selection



Long term positive selection

Innovation is a high-risk high-gain system and would follow a high stakes model for selection

> Biol Direct. 2019 Sep 3;14(1):17. doi: 10.1186/s13062-019-0248-7.

Gene connectivity and enzyme evolution in the human metabolic network

Begoña Dobon [1], Ludovica Montanucci [2], Juli Peretó [3], Jaume Bertranpetit [4], Hafid Laayouni [5 6]

Affiliations – collapse

Affiliations

1 Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Dr. Aiguader 88, 08003, Barcelona, Catalonia, Spain.
2 Dipartimento di Biomedicina Comparata e Alimentazione, Università degli Studi di Padova, Padua, Italy.
3 Institute for Integrative Systems Biology I2SysBio (University of Valencia-CSIC) and Department of Biochemistry and Molecular Biology, University of Valencia, Valencia, Spain.
4 Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Dr. Aiguader 88, 08003, Barcelona, Catalonia, Spain. jaume.bertranpetit@upf.edu.
5 Institut de Biologia Evolutiva (UPF-CSIC), Universitat Pompeu Fabra, Dr. Aiguader 88, 08003, Barcelona, Catalonia, Spain. hafid.laayouni@upf.edu.
6 Bioinformatics Studies, ESCI-UPF, Pg.Pujades 1, 08003, Barcelona, Catalonia, Spain. hafid.laayouni@upf.edu.

PREV RESULT
3 of 49

> PLoS One. 2018 Dec 14;13(12):e0208782. doi: 10.1371/journal.pone.0208782. eCollection 2018.

Influence of pathway topology and functional class on the molecular evolution of human metabolic genes

Ludovica Montanucci [1], Hafid Laayouni [1 2], Begoña Dobon [1], Kevin L Keys [1 3], Jaume Bertranpetit [1], Juli Peretó [4]

Affiliations + expand

PMID: 30550546 PMCID: PMC6294346 DOI: 10.1371/journal.pone.0208782

Free PMC article

PREV RESULT
4 of 49

Abstract

Metabolic networks comprise thousands of enzymatic reactions functioning in a controlled manner and have been shaped by natural selection. Thanks to the genome data, the footprints of adaptive selection are detectable, and the strength of purifying selection can be measured. This has made possible to know where, in the metabolic network, adaptive selection has acted and where purifying selection is more or less strong and efficient. We have carried out a comprehensive molecular evolutionary study of all the genes involved in the human metabolism. We investigated the

> BMC Genet. 2020 Oct 22;21(Suppl 1):108. doi: 10.1186/s12863-020-00908-5.

Positive selection in admixed populations from Ethiopia

Sandra Walsh [1], Luca Pagani [2 3], Yali Xue [4], Hafid Laayouni [1 5], Chris Tyler-Smith [6], Jaume Bertranpetit [7]

Affiliations + expand

PMID: 33092534 PMCID: PMC7580818 DOI: 10.1186/s12863-020-00908-5

Free PMC article

Abstract

Background: In the process of adaptation of humans to their environment, positive or adaptive selection has played a main role. Positive selection has, however, been under-studied in African populations, despite their diversity and importance for understanding human history.

Results: Here, we have used 119 available whole-genome sequences from five Ethiopian populations (Amhara, Oromo, Somali, Wolayta and Gumuz) to investigate the modes and targets of positive selection in this part of the world. The site frequency spectrum-based test SFselect was applied to idfentify a wide range of events of selection (old and recent), and the haplotype-based statistic integrated haplotype score to detect more recent events, in each case with evaluation of the significance of candidate signals by extensive simulations. Additional insights were provided by considering admixture proportions and functional categories of genes. We identified both individual

> Sci Rep. 2020 Sep 30;10(1):16134. doi: 10.1038/s41598-020-73182-1.

The shaping of immunological responses through natural selection after the Roma Diaspora

Begoña Dobon [1 2], Rob Ter Horst [3 4], Hafid Laayouni [1 5], Mayukh Mondal [6], Erica Bianco [1], David Comas [1], Mihai Ioana [7], Elena Bosch [1 8], Jaume Bertranpetit [9], Mihai G Netea [10 11 12]

Affiliations + expand

PMID: 32999407 PMCID: PMC7528012 DOI: 10.1038/s41598-020-73182-1

Free PMC article

Abstract

The Roma people are the largest transnational ethnic minority in Europe and can be considered the last human migration of South Asian origin into the continent. They left Northwest India approximately 1,000 years ago, reaching the Balkan Peninsula around the twelfth century and Romania in the fourteenth century. Here, we analyze whole-genome sequencing data of 40 Roma and 40 non-Roma individuals from Romania. We performed a genome-wide scan of selection comparing Roma, their local host population, and a Northwestern Indian population, to identify the selective pressures faced by the Roma mainly after they settled in Europe. We identify under recent selection

> Nat Biotechnol. 2019 Dec;37(12):1466-1470. doi: 10.1038/s41587-019-0333-6. Epub 2019 Dec 2.

Large multiple sequence alignments with a root-to-leaf regressive method

Edgar Garriga [1], Paolo Di Tommaso [1], Cedrik Magis [1], Ionas Erb [1], Leila Mansouri [1], Athanasios Baltzis [1], Hafid Laayouni [2 3], Fyodor Kondrashov [4], Evan Floden [5], Cedric Notredame [6 7]

Affiliations + expand

PMID: 31792410 PMCID: PMC6894943 DOI: 10.1038/s41587-019-0333-6

Free PMC article

Abstract

Comparative Study > Proc Natl Acad Sci U S A. 2014 Feb 18;111(7):2668-73.
doi: 10.1073/pnas.1317723111. Epub 2014 Feb 3.

Convergent evolution in European and Rroma populations reveals pressure exerted by plague on Toll-like receptors

Hafid Laayouni [1], Marije Oosting, Pierre Luisi, Mihai Ioana, Santos Alonso, Isis Ricaño-Ponce, Gosia Trynka, Alexandra Zhernakova, Theo S Plantinga, Shih-Chin Cheng, Jos W M van der Meer, Radu Popp, Ajit Sood, B K Thelma, Cisca Wijmenga, Leo A B Joosten, Jaume Bertranpetit, Mihai G Netea

Affiliations + expand

PMID: 24550294 PMCID: PMC3932890 DOI: 10.1073/pnas.1317723111

Free PMC article

PREV RESULT
23 of 49

Abstract

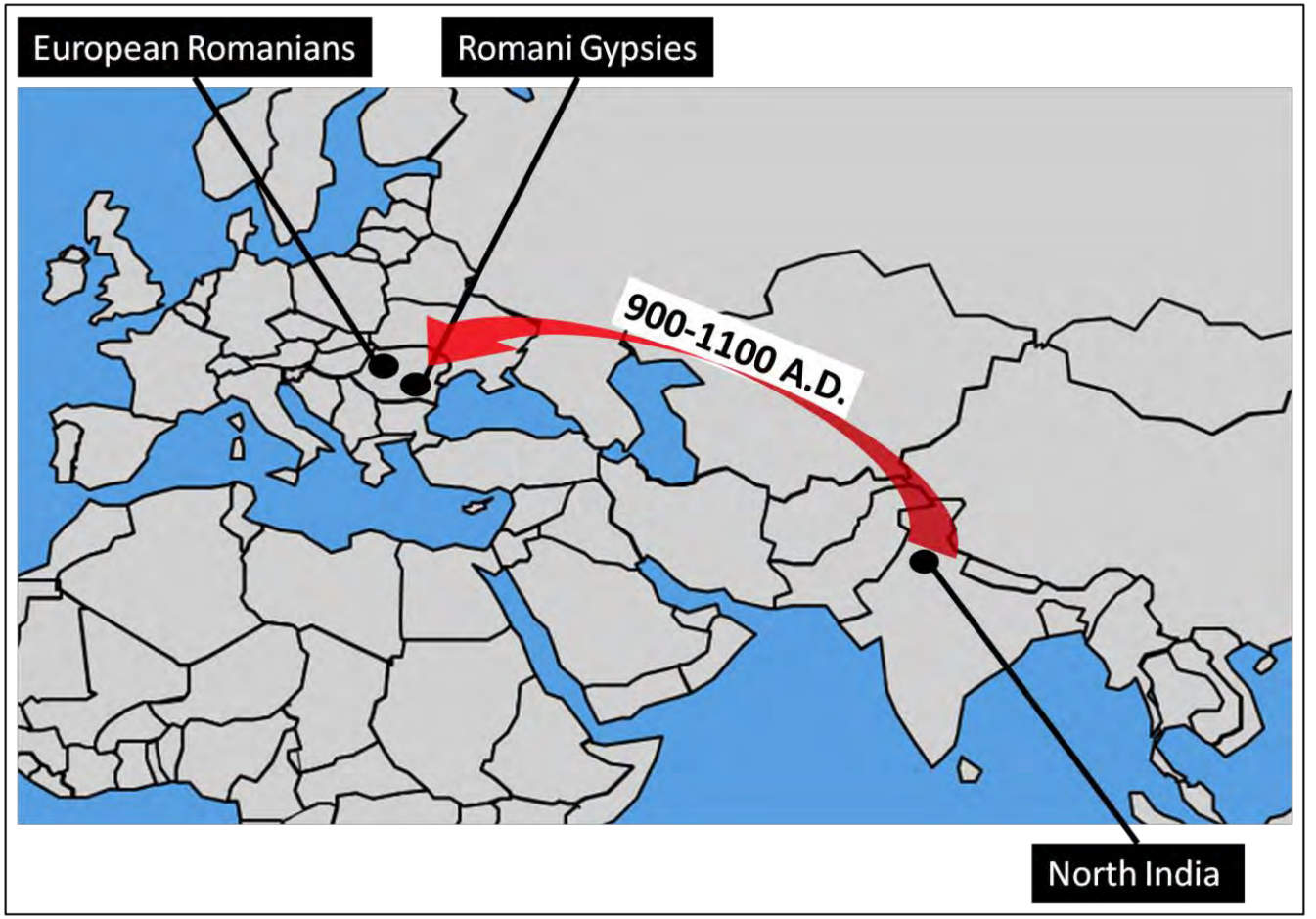
Recent historical periods in Europe have been characterized by severe epidemic events such as plague, smallpox, or influenza that shaped the immune system of modern populations. This study

Spread of the black death in Europe (1346–53)

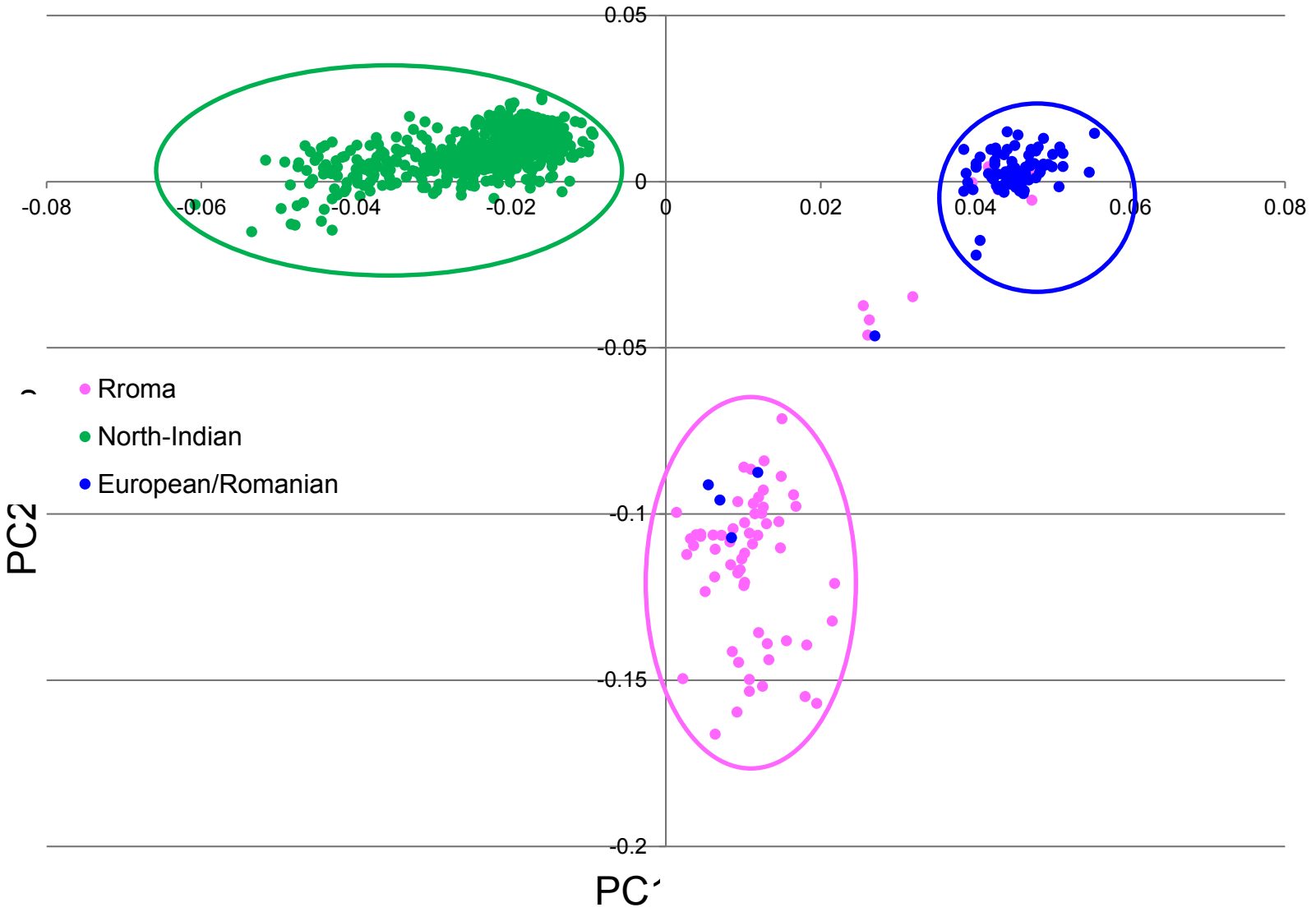
How the Black Death Worked

FLAGELLANTS

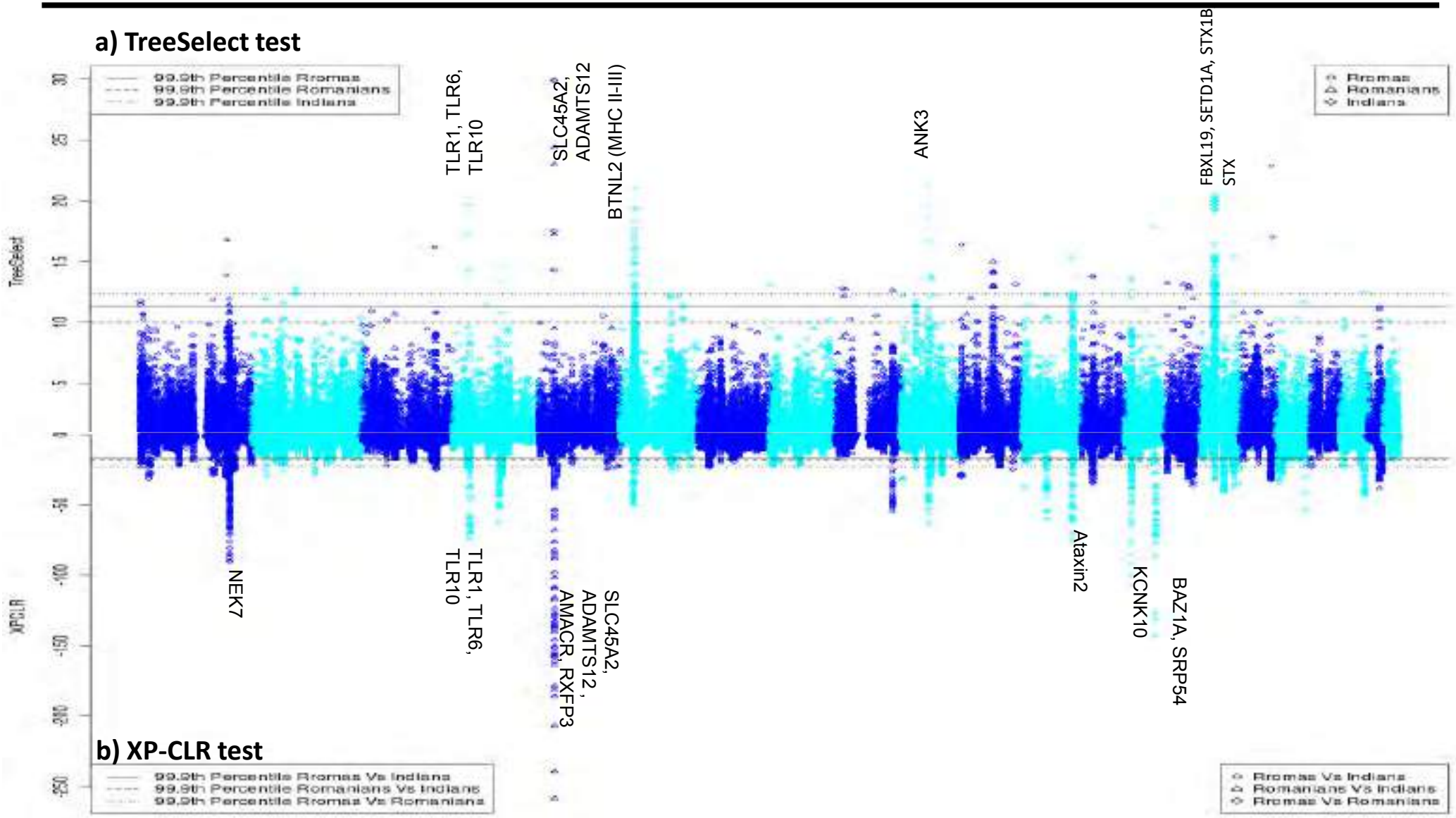
European/Romanians and Rroma/Gipsy share the same location, even if the origin of the latter is in North India



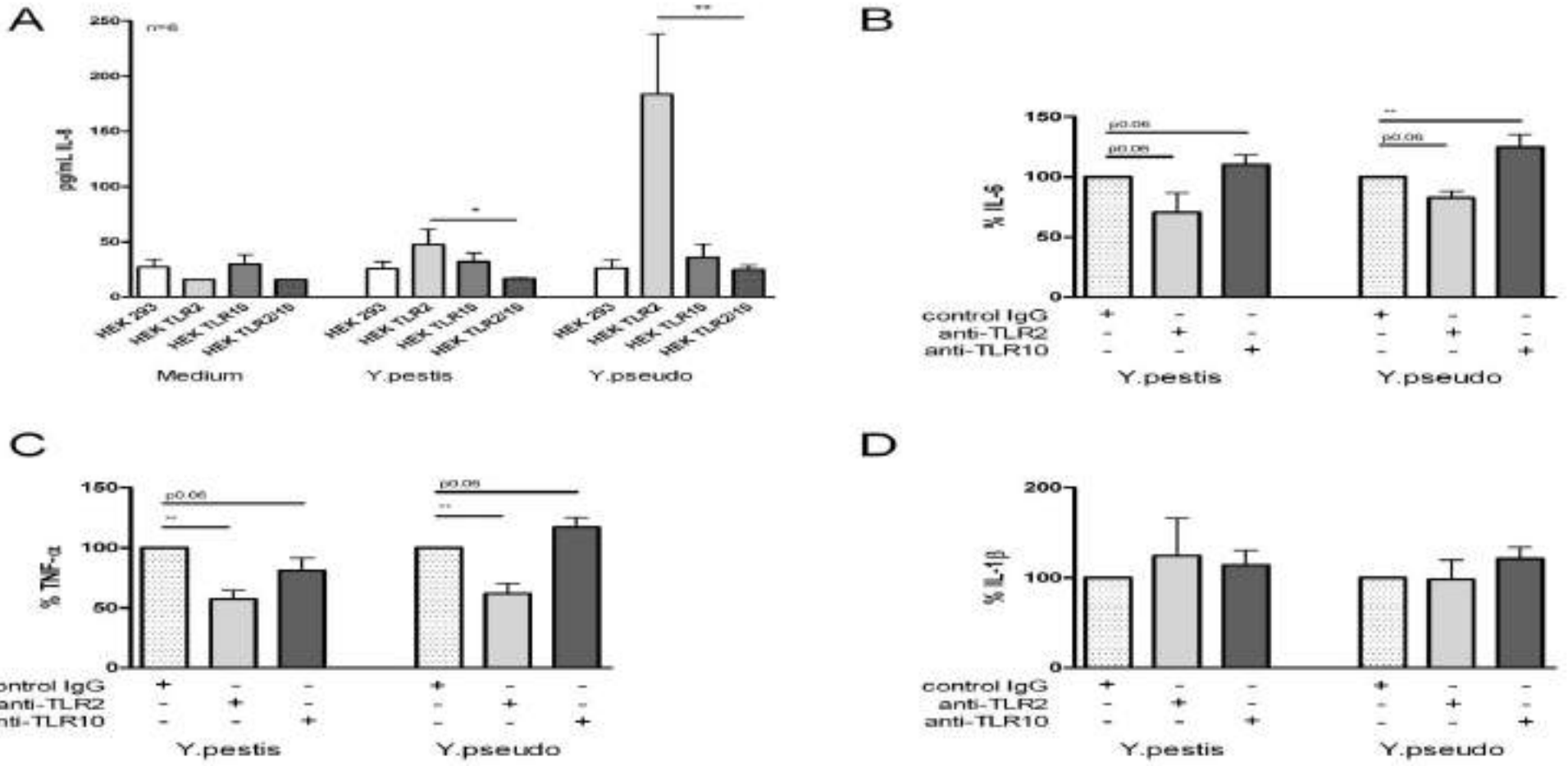
Data: Geographic origin of the three populations studied.

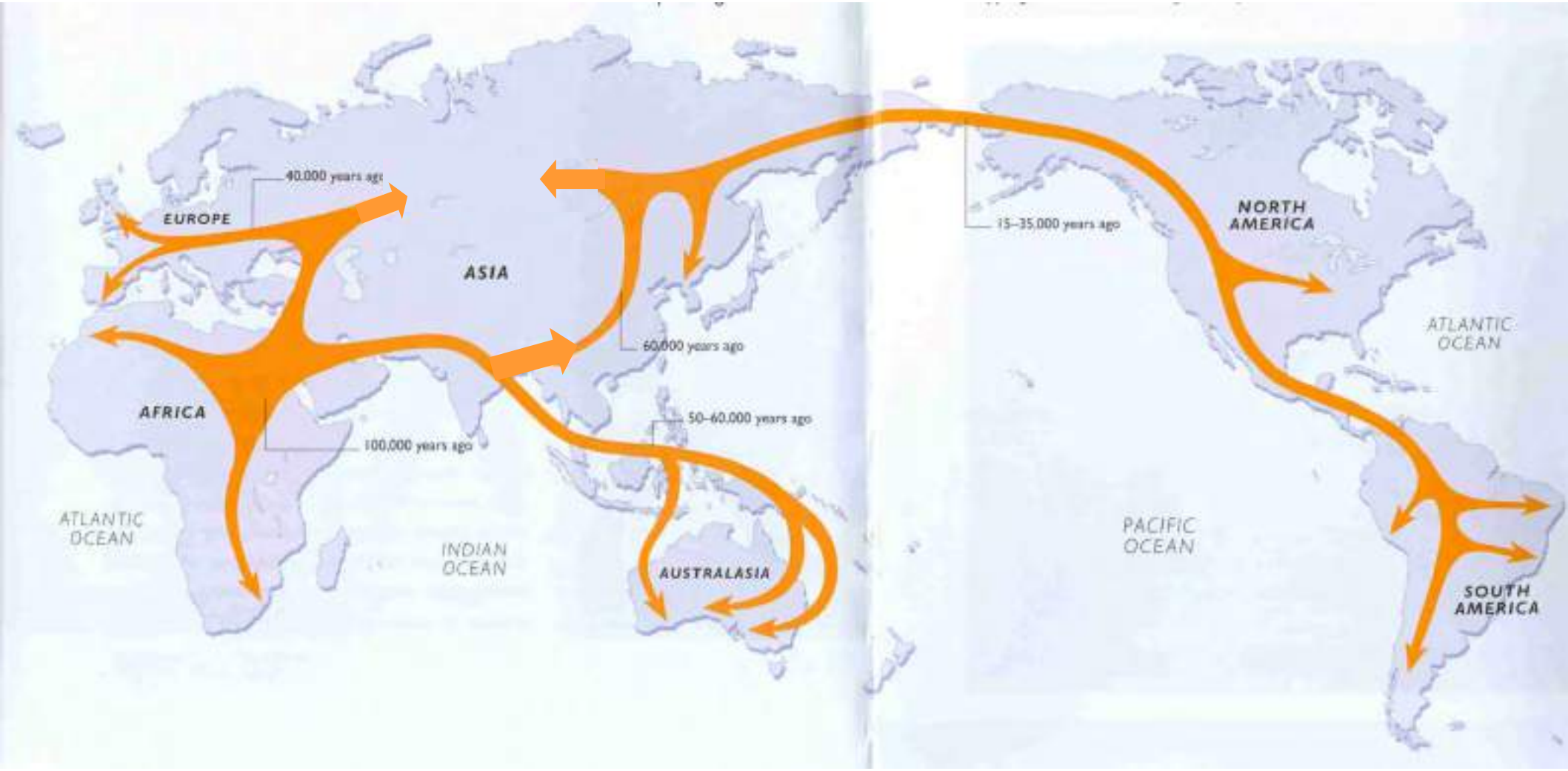


Results: Genes under Positive selection tests in Rroma, Romanians and Indians using Tree Select statistic and XP-CLR statistic



TLR2 cluster genes are involved in the recognition of Yersinia pestis and *Y. pseudotuberculosis: .*




40,000 years ago
EUROPE
ASIA
AFRICA
100,000 years ago
60,000 years ago
50–60,000 years ago
AUSTRALASIA
ATLANTIC OCEAN
INDIAN OCEAN
15–35,000 years ago
NORTH AMERICA
ATLANTIC OCEAN
PACIFIC OCEAN
SOUTH AMERICA

All equal, all different-

- Among humans, due to our evolutionary history, we have differences, although very few

- The difference is fundamental for life. Without difference there would be no evolution

- The sequence of the human genome can explain our particularity as a species

- Understanding the difference explains the evolutionary process and our history.

Implications of the (few) genetic differences between humans

- Few genetic differences give rise to the differences we observe in appearance, physiology, behavior

-These differences are at the base of the susceptibility to complex diseases, which have genetic influence and affect many people

- These differences may be used in diagnosis and therapy since they are the basis for individual differences in response to drugs.

ACKLNOWLEDGMENTS

Jaume Bertranpetit
Ludovica Montanucci, Ferran Casals
Giovanni M Dall'Olio, Pierre Luisi, Brandon Invergo, Marc Pybus, Martino Colombo