Genetic isolation in human populations

Isolation in natural and human population

"An isolate is a population to which an external barrier prevents the free genetic exchange with other populations of the same species "

(Mayr 1963)

Isolation in natural and human population

NATURAL POPULATIONS

Geographical barriers

↓

Gene Flow
Inbreeding
Genetic Drift
Natural Selection

↓

Intra and Inter-pop diversity

↓

Reproductive isolation

↓

Speciation

Cultural barriers

HUMAN POPULATIONS

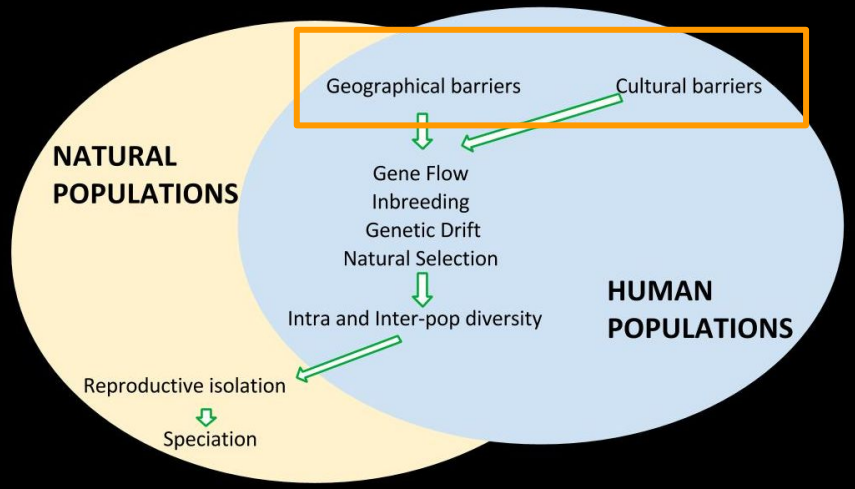
Reproductive isolation?

↓

races ? no!

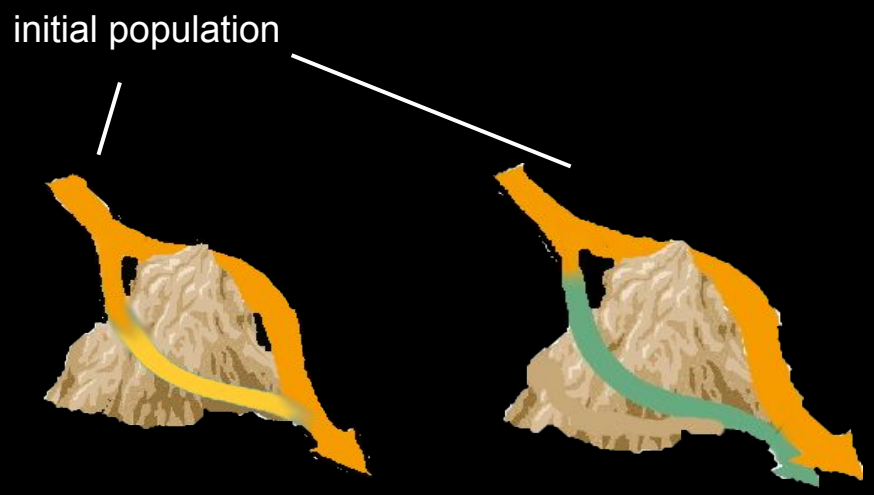
Isolation in natural and human population

Geographical barriers



Cultural barriers

physical barrier



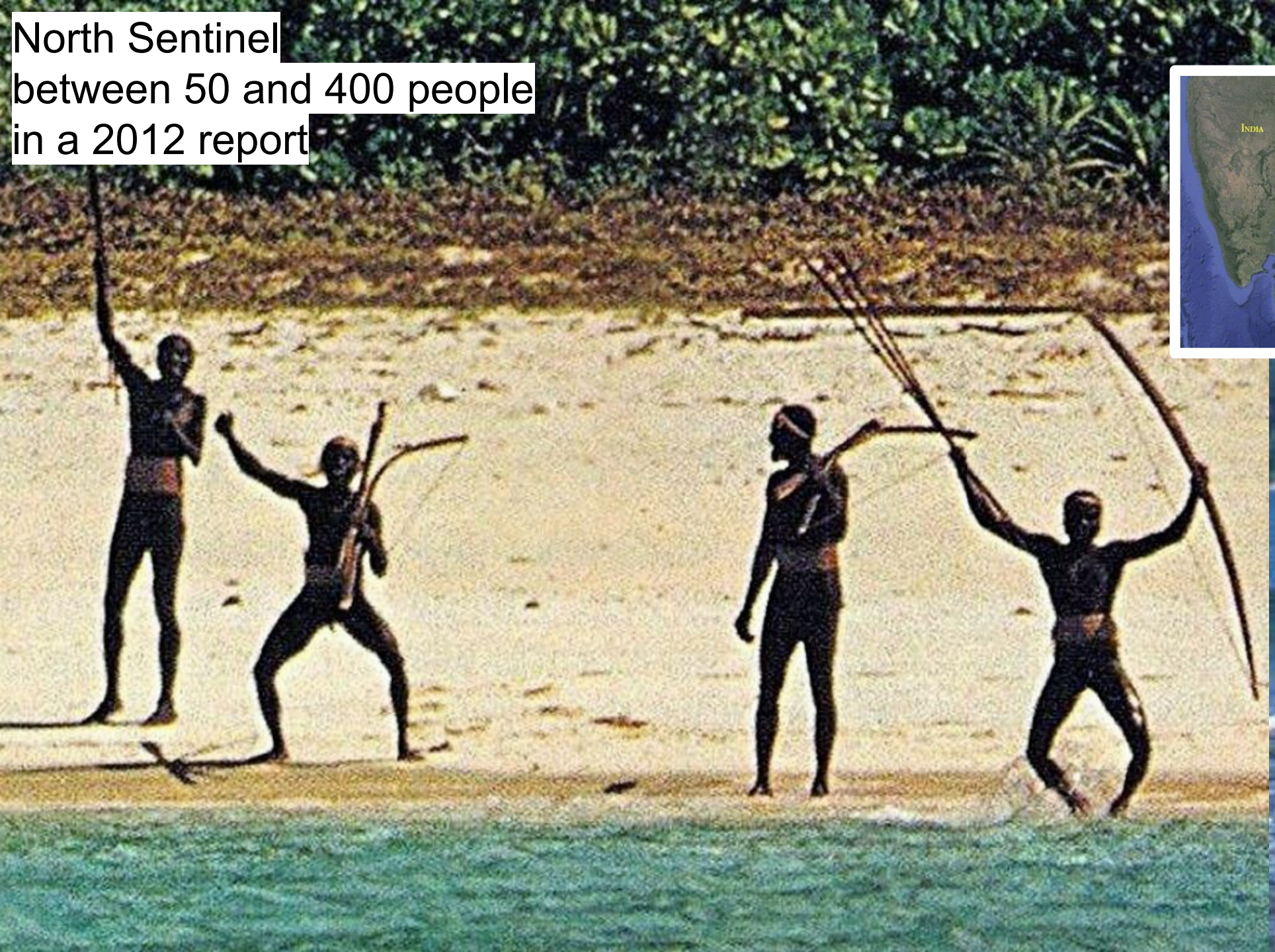
The extent of diversity of isolates depends on their time and completeness of isolation from the neighbouring populations.

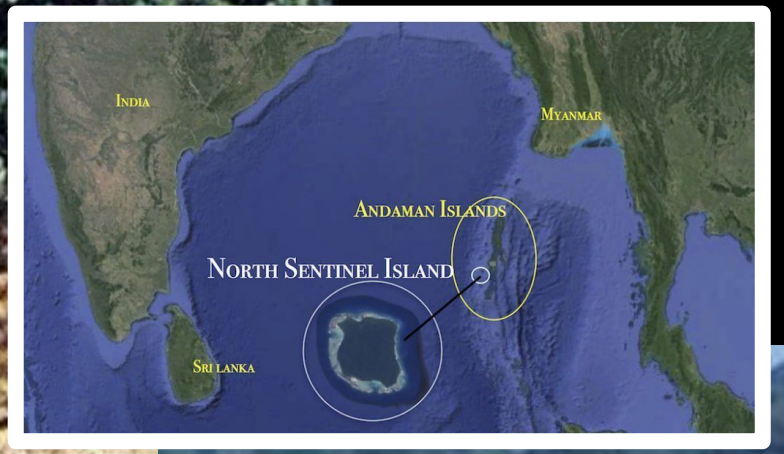
cultural barrier

differences in

culture, religion, language, social class

North Sentinel
between 50 and 400 people
in a 2012 report





361 635 (2021)

discendono da un numero molto piccolo di fondatori (circa 200-500 individui) emigrati dall'Europa centrale in Nord America nel XVIII secolo.
Altamente endogamici: sindrome di Ellis-van Creveld, la malattia di Tay-Sachs, e vari disturbi metabolici rari.

Genoma miscela di antiche popolazioni siberiane/uraliche e mescolamento con gruppi europei successivi.adattamenti a climi freddi,

Sami

c.ca 100.000 ab. (2021)

Il genoma sardo deriva in gran parte (circa il 62.5%) da agricoltori europei antichi

c.ca 1.600.000)

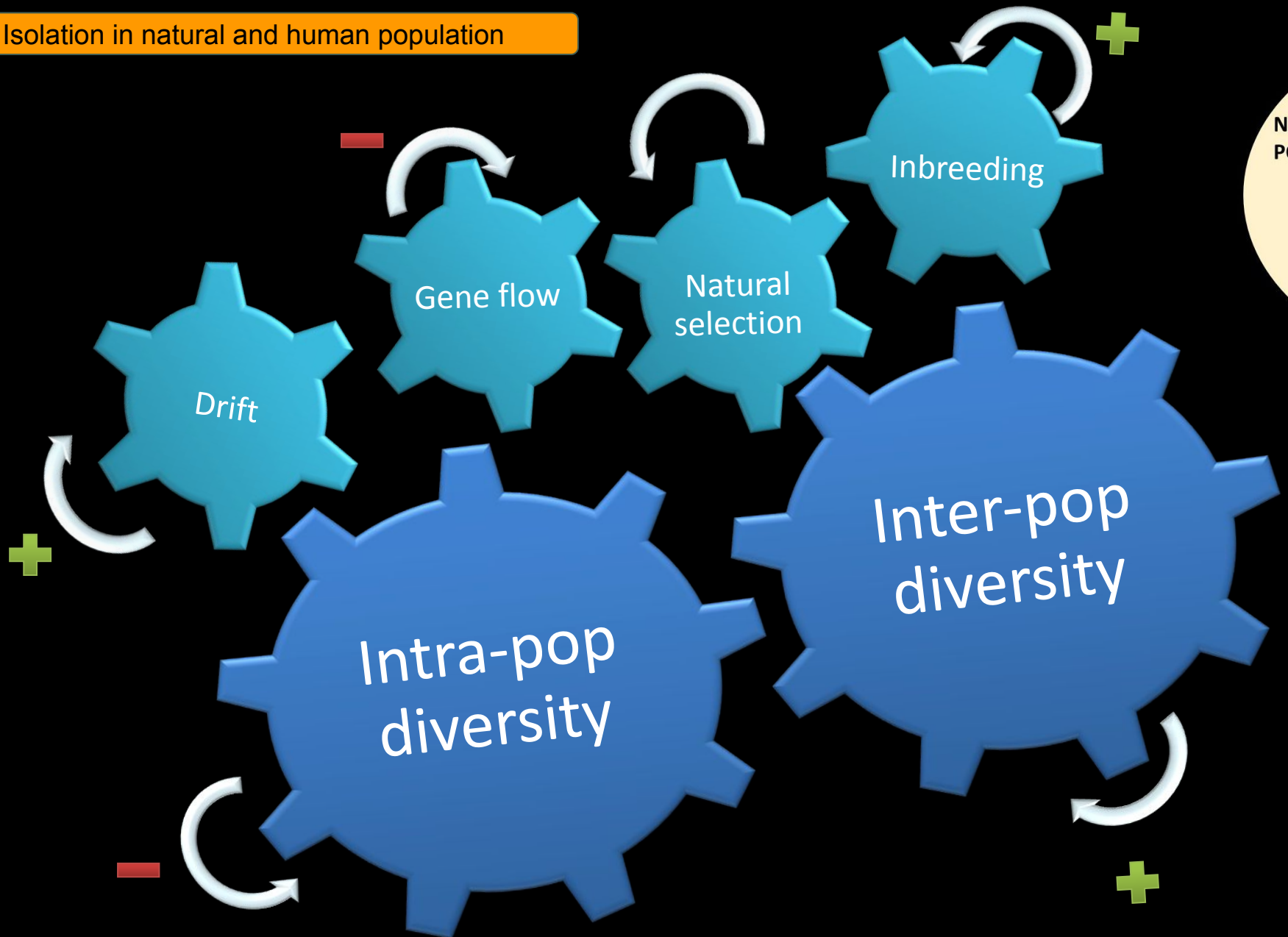
c.ca 2.800.000)

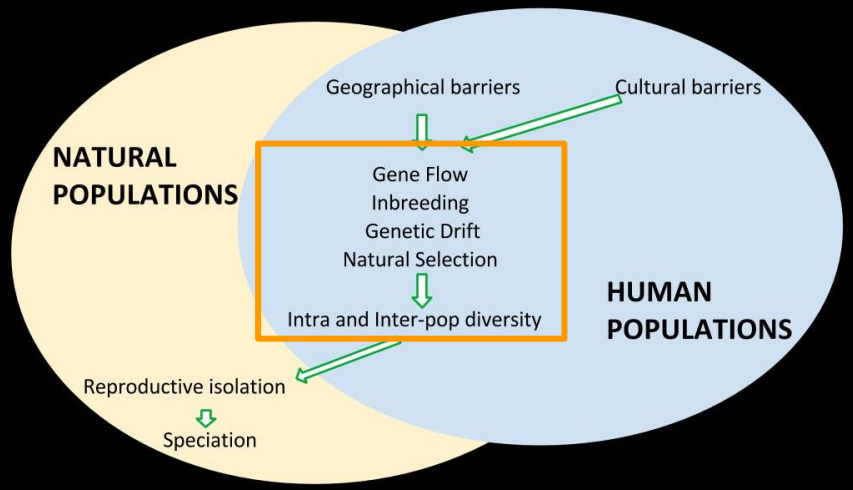
discendono in gran parte da popolazioni pre-agricole dell'Europa occidentale (cacciatori-raccoglitori del Mesolitico

Sistema delle caste, che si è consolidato circa 1.600-1.900 anni fa (periodo Gupta), ha imposto una rigorosa endogamia1. **Dalit (ex-caste degli intoccabili)**: circa il 16-20% della popolazione indiana, con una variazione regionale significativa.

2. **Caste degli Adivasi (tribali)**: Comprendono circa il 8-10% della popolazione indiana.
3. **Caste degli OBC (Other Backward Classes)**: Questo è un gruppo ampio e variegato che costituisce circa il 40-45% della popolazione.
4. **Caste dei Brahmini**: Costituiscono una piccola percentuale della popolazione totale, stimata intorno al 5-6%.
5. **Caste dei Kshatriya e Vaishya**: generalmente inferiore al 20%.

Isolation in natural and human population





Human isolates

1.tribal populations of presumably very ancient origin which since they emerged as distinct entities have had relatively little biological exchange with other similar groups: **Amerindians**

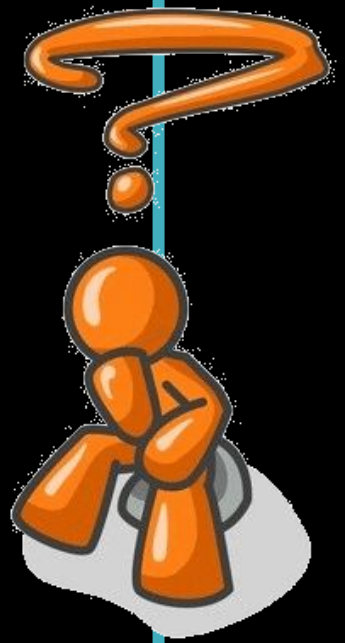
2.relatively large ethnic groups that are sub sample of an established, national population formed by the coalescence of tribes in a period of two-four thousand years **Pakistani in GB, Chinese in European and American towns**

3.small populations that show a slow demographic expansion and minimum immigrant rate: the allelic profile about these populations is largely influenced by the case: **Norfolk islands, most Italian linguistic isolates**

(Neel, 1992)

Background

Open populations



Isolated populations

+	Gene.ic flow	-
-	Genetic drift	+
-	Inbreeding	+
+	Intra-population diversity	-
-	Inter-population diversity	+

SAPIENZA
UNIVERSITÀ DI ROMA

Isolation in natural and human population

HUMAN POPULATION
(Neel's definition)

Geographical and/or cultural barriers

	Continuity of settlement (**primary isolates**) e.g. Amerindians	Relocation (**secondary isolates**): Large founding group (type I) e.g. Pakistani of London	Relocation (**secondary isolates**): Small founding group (type II) e.g. Norfolk island	
Gene flow	↓	↓	↓	Micro evolutionary forces
Genetic drift	↑	no	↑	
Natural selection	↑	≈/≠	≈/≠	
Effective population size	↓	↓	↓	Demography
Growth rate	↓	≈	↓	
Inbreeding rate	↑	↑	↑	Reproductive behaviour
Intra-pop diversity	↓	≈	↓	Genetic structure
Inter-pop diversity	↑	↑	↑	

Population history

Caratteristica/Aspetto	Popolazioni Isolate (e.g., Sardi, Amish, Sami, Baschi, Finlandesi, Hutterite)	Popolazioni Mescolate (e.g., Afroamericani, Latinoamericani, Indiani delle Caste)
Definizione	Popolazioni che hanno sperimentato un isolamento genetico prolungato (geografico, culturale, religioso) e/o un forte effetto del fondatore.	Popolazioni formate dalla recente (nelle ultime centinaia di generazioni) mescolanza di due o più popolazioni ancestrali geneticamente distinte.
Struttura Genetica Generale	Omogeneità genetica elevata all'interno del gruppo. Bassa variabilità genetica generale.	Eterogeneità genetica, con un mosaico di segmenti genomici provenienti da diverse popolazioni ancestrali.
Linkage Disequilibrium (LD)	**LD esteso su lunghe distanze cromosomiche.** La ricombinazione ha avuto meno tempo o opportunità di rompere i blocchi di LD ancestrali.	**LD esteso su lunghe distanze cromosomiche**, ma per meccanismi diversi (eredità di lunghi segmenti ancestrali non ricombinati). Decadimento del LD più rapido rispetto alle popolazioni isolate molto antiche.
Effetto del Fondatore	**Molto pronunciato.** Origine da un piccolo numero di individui, con successiva crescita in isolamento.	Meno rilevante come fattore primario di *creazione* di specifici blocchi di LD, ma le popolazioni ancestrali che si mescolano potrebbero aver avuto effetti del fondatore al loro interno.
Prevalenza Malattie Rare/Monogeniche	**Aumentata incidenza di specifiche malattie genetiche recessive** dovute alla deriva genetica e all'effetto fondatore.	Anche qui si possono trovare un aumento di malattie recessive, ma spesso legate a specifiche **ascendenze locali** o a colli di bottiglia all'interno di specifici sub-gruppi ancestrali.

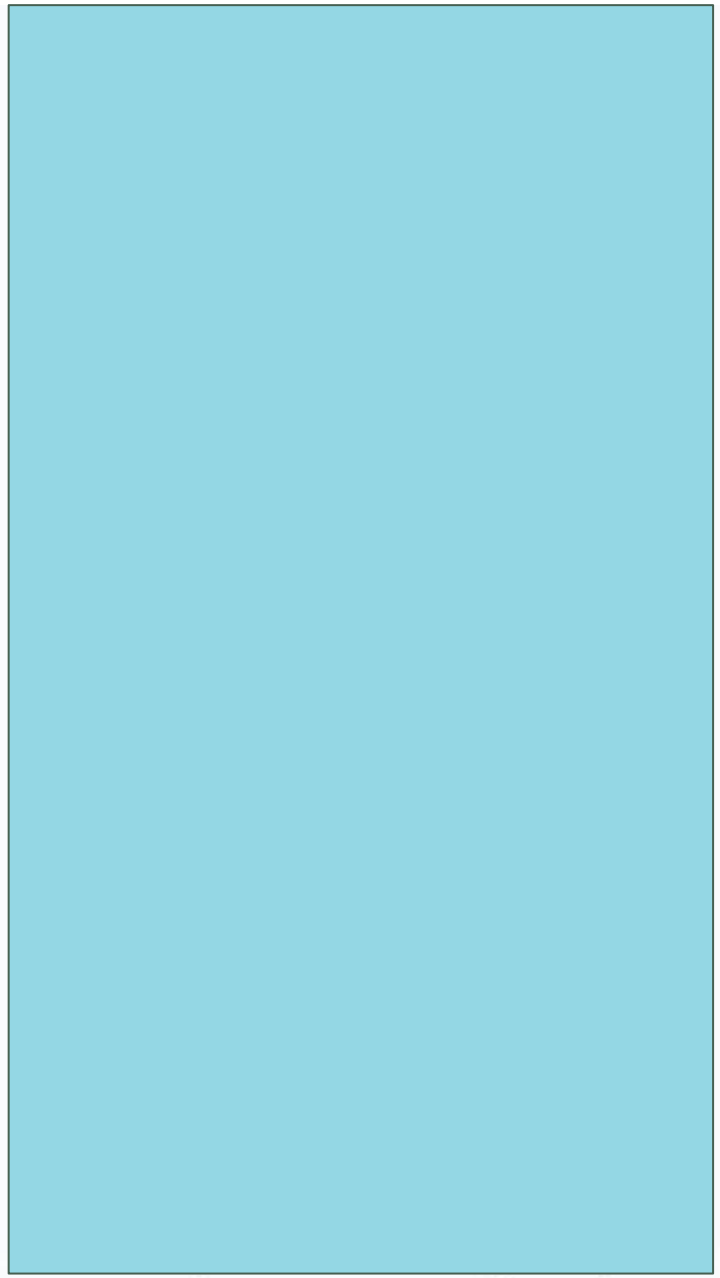
Caratteristica/Aspetto	Popolazioni Isolate (e.g., Sardi, Amish, Sami, Baschi, Finlandesi, Hutterite)	Popolazioni Mescolate (e.g., Afroamericani, Latinoamericani, Indiani delle Caste)
Vantaggi per Studi di Associazione Geni-Malattia	- **Mappatura efficiente:** Il LD esteso significa che meno marcatori sono necessari per coprire il genoma. Un singolo marcatore può essere in LD con un gene causativo più distante. - **Omogeneità genetica:** Minore eterogeneità genetica significa meno rumore di fondo nelle analisi e maggiore potere statistico per rilevare associazioni. - **Malattie recessive:** Ideali per identificare i geni responsabili di malattie recessive rare accumulate nella popolazione. - **Forni Unique Allele:** Possibilità di identificare alleli mutati unici e rari.	- **Mappatura dell'ascendenza (Admixture Mapping):** Permettono di identificare loci che contribuiscono a differenze di rischio di malattia tra popolazioni ancestrali. - **Riduzione della stratificazione della popolazione (se gestita correttamente):** La mescolanza può rompere associazioni spurie presenti nelle popolazioni non mescolate. - **Ampio spettro di varianti:** Offrono una maggiore diversità di varianti genetiche ereditate da diverse popolazioni ancestrali, utili per studi più ampi. - **Rilevanza per la salute pubblica:** Spesso rappresentano popolazioni di grande interesse per la salute pubblica (es. diabete, ipertensione).
Svantaggi per Studi di Associazione Geni-Malattia	- **Generalizzabilità limitata:** I risultati potrebbero non essere facilmente generalizzabili a popolazioni più eterogenee. - **Malattie comuni/poligeniche:** Non sempre ideali per tutti i tipi di malattie comuni e poligeniche, a meno che non ci siano specifici fattori di rischio accumulati. - **Omogeneità patologica:** Se la popolazione è molto piccola e isolata, potrebbe esserci meno variabilità clinica all'interno di una malattia, rendendo difficile identificare fattori modificatori.	- **Complessità analitica:** Richiedono modelli statistici più sofisticati per controllare l'ascendenza locale e la stratificazione genetica (che può portare a falsi positivi). - **Decadimento del LD:** Il LD decade più rapidamente rispetto a popolazioni isolate, richiedendo più marcatori per una copertura sufficiente (a meno che non si faccia Admixture Mapping). - **Difficoltà di replicazione:** Le associazioni trovate in una popolazione mescolata potrebbero non replicarsi in altre popolazioni mescolate se le proporzioni ancestrali o la storia del mescolamento sono diverse.
Esempio di Applicazione	Identificazione del gene causativo per la sindrome di Ellis-van Creveld negli Amish. Mappatura della beta-talassemia nei Sardi.	Identificazione di varianti che influenzano il rischio di diabete di tipo 2 negli Afroamericani o nei Latinoamericani, dove le varianti possono essere state ereditate da diverse ascendenze ancestrali

Comparison of Isolated vs. Admixed Populations in Genetic Studies

Feature	Isolated Populations	Admixed Populations
Genetic Diversity	Lower heterozygosity, reduced number of rare alleles	
Linkage Disequilibrium (LD)	Longer stretches of LD across chromosomes	
Allelic Heterogeneity	Often reduced; fewer causal mutations for a given trait	
Founder Effect	Pronounced; specific rare alleles can be enriched	
Genetic Drift	More significant due to smaller effective population size	
Homozygosity	Increased due to limited gene flow and consanguinity	
Pedigree Data	Often extensive and well-documented	

Comparison of Isolated vs. Admixed Populations in Genetic Studies

Environmental/Lifestyle Homogeneity	Often high, reducing confounding factors
Population Stratification	Generally lower, but can exist within sub-isolates
Power for Rare Variants	Potentially higher for specific enriched rare variants
Fine-Mapping	Initially challenging due to long LD, but can be improved with specific strategies
Replication	Crucial due to unique genetic architecture
Reference Panels for Imputation	Often require isolate-specific panels
Ethical Considerations	Specific attention to community engagement and benefit sharing



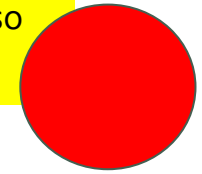
BRIEFINGS IN FUNCTIONAL GENOMICS. VOL 13. NO 5. 371–377 doi:10.1093/bfgp/elu022

Using population isolates in genetic association studies

Konstantinos Hatzikotoulas, Arthur Gilly and Eleftheria Zeggini

Advance Access publication date 9 July 2014

cliccare sul tondo rosso per aprire il file



Abstract

The use of genetically isolated populations can empower next-generation association studies. In this review, we discuss the advantages of this approach and review study design and analytical considerations of genetic association studies focusing on isolates. We cite successful examples of using population isolates in association studies and outline potential ways forward.

Keywords: *isolated populations; rare variants; complex disease; genetic association studies*

Purging selection
https://www.biorxiv.org/content/10.1101/022947v2.full

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168662/pdf/elu022.pdf

Population Isolate	Key Characteristics	Significant Disease Associations & Genes Discovered (Examples)
Finnish	Historical bottlenecks, relative isolation. Enriched rare disease alleles, reduced genetic heterogeneity.	**"Finnish Heritage Diseases"**: Congenital nephrosis (NPHS1), Diastrophic dysplasia (DTDST), Salla disease (SLC17A5), Progressive myoclonus epilepsy (CSTB). **Complex Diseases (FinnGen study)**: Hundreds of novel genetic risk factors for common diseases like Inflammatory bowel disease (e.g., variant in TNRC18), Hypothyroidism, Hearing loss, Endometriosis, protection from arthrosis, glaucoma, or heart disease.
Amish & Hutterite (Anabaptist groups)	Strong founder effects, large family sizes, detailed genealogical records.	**Maple Syrup Urine Disease** (more common in Old Order Mennonites). **Cystic Fibrosis** (higher rate in modern Hutterites). **Cardiovascular and Metabolic Traits**: Genetic variants associated with quantitative phenotypes like carotid intima-media thickness, left atrial volume index, and asthma (e.g., related to CHI3L1 gene). **Longevity and Metabolic Health**: *APOC3* null mutation associated with favorable lipid profiles and extended longevity in some Amish communities.

Norio R. Finnish Disease Heritage I: characteristics, causes, background. Hum Genet. 2003 May;112(5-6):441-56. doi: 10.1007/s00439-002-0875-3. Epub 2003 Mar 8. PMID: 12627295.

Puffenberger EG. Genetic heritage of the Old Order Mennonites of southeastern Pennsylvania. Am J Med Genet C Semin Med Genet. 2003 Aug 15;121C(1):18-31. doi: 10.1002/ajmg.c.20003. PMID: 12888983.

Population Isolate	Key Characteristics	Significant Disease Associations & Genes Discovered (Examples)
Icelandic (deCODE Genetics)	Relatively homogenous population, comprehensive genealogical database.	**Common Diseases:** Type 2 Diabetes (e.g., CDKAL1 gene), Myocardial Infarction (e.g., 9p21 locus), Atrial Fibrillation (e.g., 4q25 locus), Breast Cancer (e.g., 5p12, 2q35, 16q12 loci), Prostate Cancer (e.g., Xp11.22, 2p15, 17q loci), Exfoliation Glaucoma (LOXL1 gene). **Quantitative Traits:** Loci influencing traits like height and pigmentation.
Ashkenazi Jewish	Significant bottleneck and subsequent expansion, leading to higher prevalence of certain recessive disorders.	**Recessive Disorders:** Tay-Sachs disease (HEXA gene), Gaucher disease (GBA gene), Canavan disease (ASPA gene), Familial Dysautonomia (ELP1 gene), Factor XI deficiency (F11 gene).
Sardinian	Mediterranean isolate.	**Complex Traits and Common Diseases:** Asthma (e.g., IRAK-M gene), Obesity (e.g., FTO, PFKP genes), Fasting glucose levels (e.g., G6PC2/ABCB11 region).

Gudbjartsson, D. F., et al. (2007). Sequence variants affecting lipoprotein lipase, high-density lipoprotein cholesterol, and risk of myocardial infarction. Nature Genetics, 39(12), 1435-1442.

Gross, S. J., et al. (2008). Carrier screening in individuals of Ashkenazi Jewish descent. Genetics in Medicine, 10(1), 50-52.

Sidore, C., et al. (2015). Genome sequencing elucidates Sardinian genetic architecture and disease susceptibility. Science, 347(6224), 1152-1157.

Geographic Isolation	Large Mediterranean island, historically limited gene flow with mainland Europe and other regions.	Contributes to genetic distinctiveness and reduced admixture over time.
Ancient Neolithic Ancestry	High genetic similarity to early Neolithic European farmers; preserved significant ancient ancestry.	Provides insights into early European population migrations and the genetic landscape before later admixtures.
Genetic Homogeneity (with substructure)	Overall genetically more uniform than outbred populations, but with internal regional variations (e.g., mountain vs. coastal areas).	Facilitates gene mapping by reducing genetic noise; substructure allows for exploration of localized genetic adaptation.
Enriched Specific Variants	Higher frequencies of some genetic variants that are rare in other European populations (e.g., due to founder effects or historical selection).	Enables the discovery of associations with rare or otherwise hard-to-detect variants.
Extended Linkage Disequilibrium (LD)	Longer stretches of co-inherited genetic blocks across chromosomes.	Can simplify the initial localization of disease-causing genes (linkage analysis) by requiring fewer markers.
Comprehensive Resources	Extensive genetic studies, detailed genealogies, and rich health data available for research.	Powerful resource for robust genetic association studies and understanding disease etiology.
High Longevity ("Blue Zone")	Certain Sardinian regions show exceptional rates of centenarians.	Ideal for genetic studies focused on healthy aging, longevity, and resistance to age-related diseases.

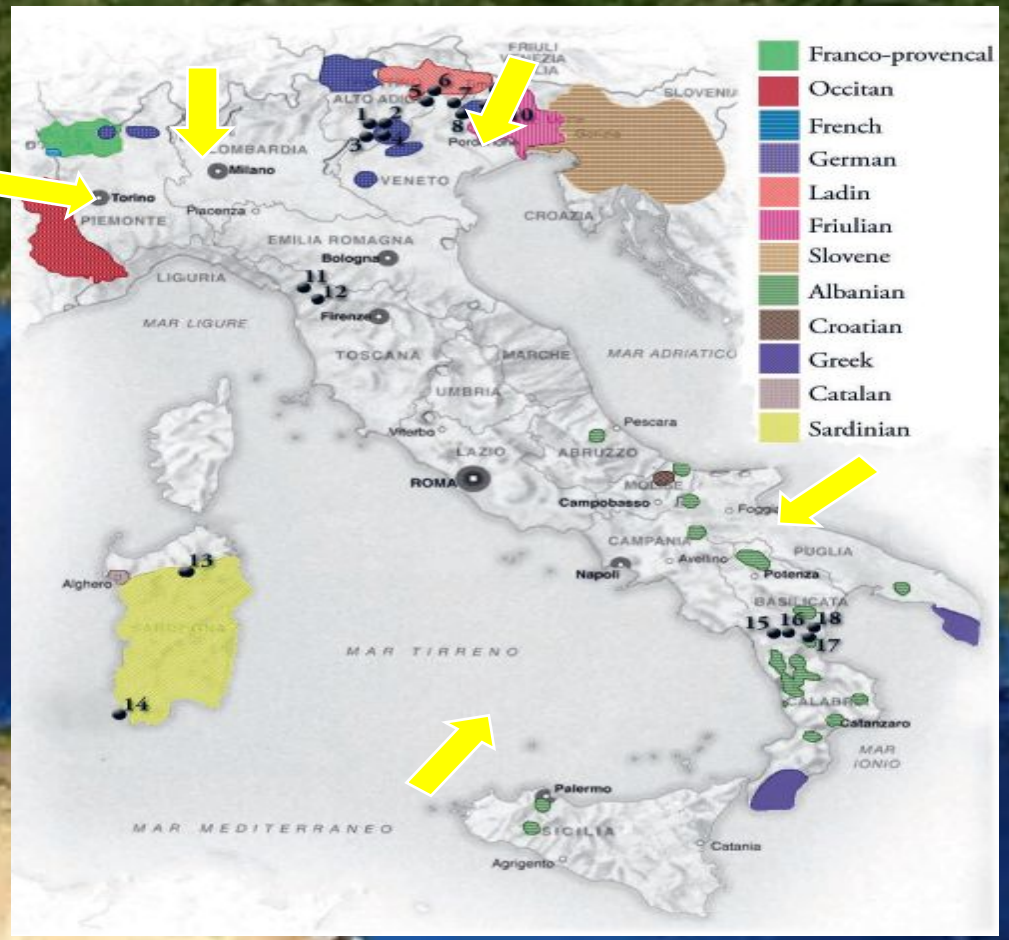
but not negligible

1. L’Italia ha una grande diversità linguistica…

35 lingue diverse (Germania 27, Francia 23, Spagna 15, UK 13)

indice di diversità 47%

…e quella genetica?


Franco-provencal
Occitan
French
German
Ladin
Friulian
Slovene
Albanian
Croatian
Greek
Catalan
Sardinian
Torino
Milano
Piacenza
Bologna
Firenze
Pescara
Viterbo
ROMA
Campobasso
Napoli
Avellino
Potenza
Foggia
Alghero
Catanzaro
Palermo
Catania
Agrigento
PIEMONTE
LOMBARDIA
VENETO
EMILIA ROMAGNA
LIGURIA
TOSCANA
MARCHE
UMBRIA
LAZIO
ABRUZZO
CAMPANIA
PUGLIA
BASILICATA
CALABRIA
SICILIA
SLOVENIA
CROAZIA
MAR LIGURE
MAR TIRRENO
MAR ADRIATICO
MAR IONIO
MAR MEDITERRANEO

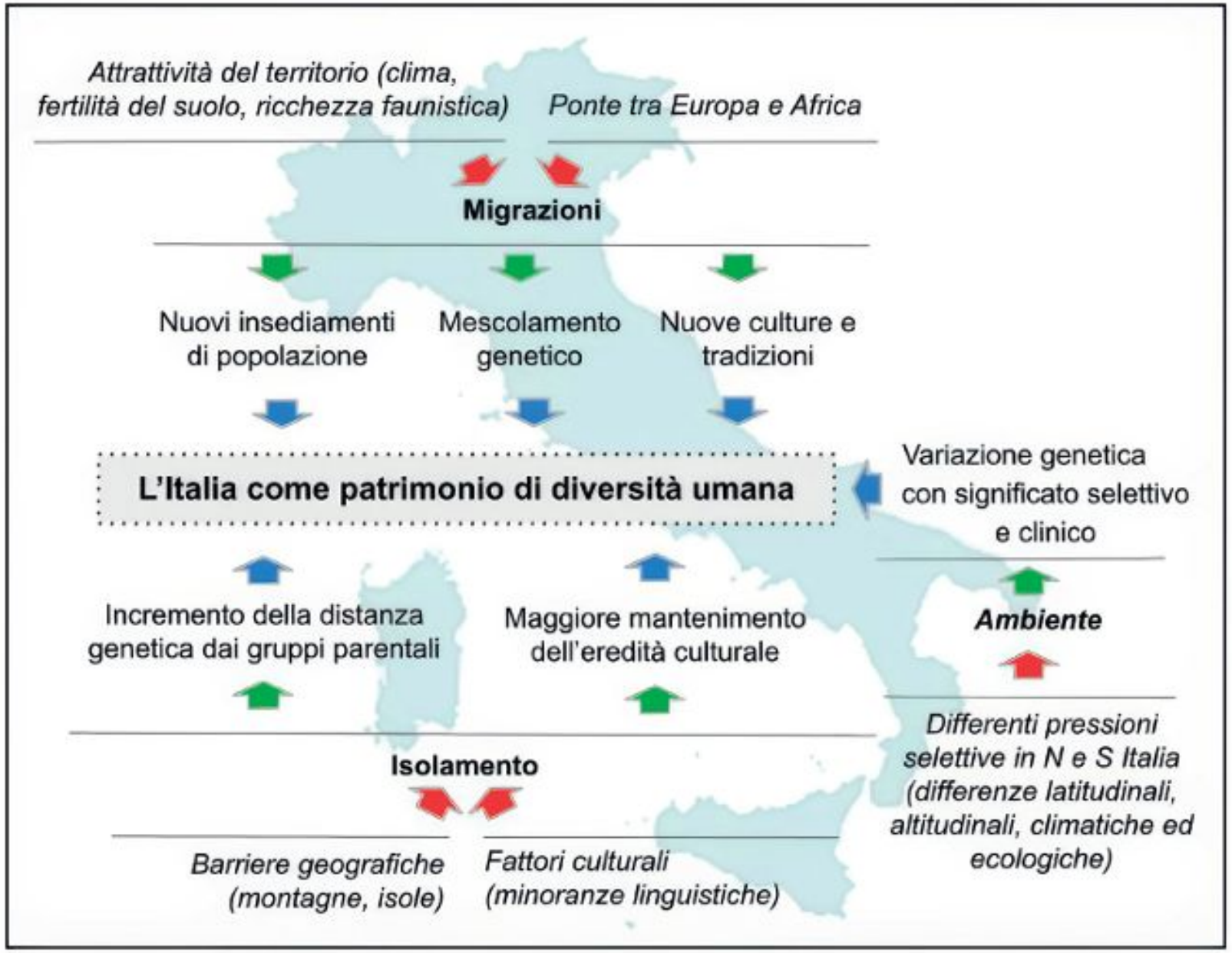
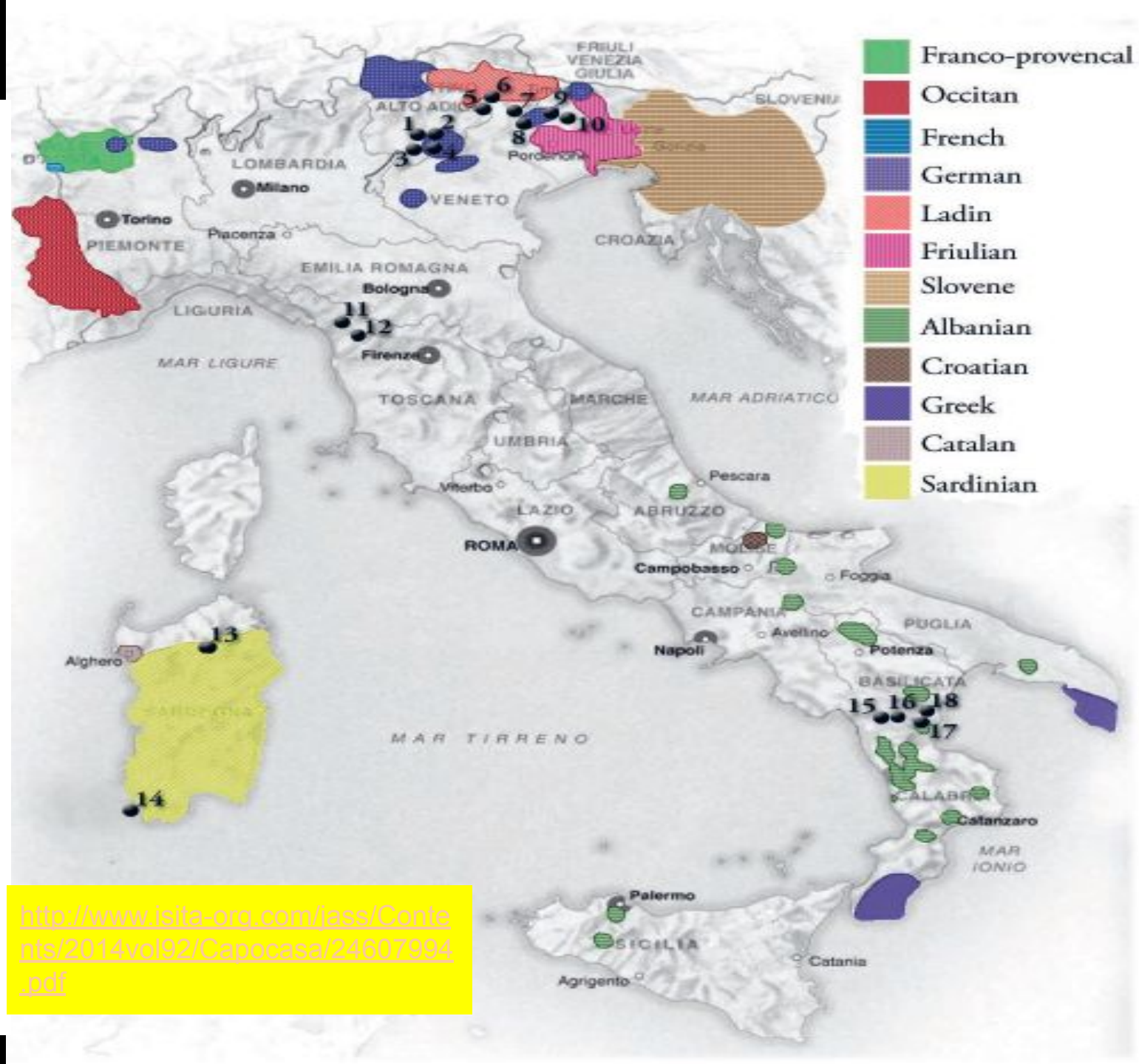


Figura 6. *L'azione combinata di migrazioni, isolamento, sia genetico sia culturale, e pressioni selettive ha reso l'Italia un patrimonio di diversità umana.*

POPULATION	ABB.	STATUS	SAMPLING STRATEGY	CENSUS SIZE*	ALT.°	MTDNA N	MTDNA REFERENCE	Y CHROMOSOME N	Y CHROMOSOME REFERENCE
Arbereshe (Calabria)	ARC	GL	FS	28034	535	87	This study	87	Boattini *et al.*, 2011; Pettener, *pers. com.*
Arbereshe (Sicily)	ARS	L	FS	7875	645	40	This study	42	Pettener, *pers. com.*
Benetutti	BEN	G	FS	2010	406	50	This study	45	Calò *et al.*, 2013
Cadore	CAD	O	GP	10797	984	32	Caglià, *pers. com.*	52	Caglià, *pers. com.*
Carloforte	CFT	GL	FS/GP	6420	10	51	Calò *et al.*, 2012	41	Robledo *et al.*, 2012
Cimbrians (Lessinia)	LES	GL	GP	13455	758	40	Capocasa *et al.*, 2013	29	Coia *et al.*, 2013
Cimbrians (Luserna)	LUS	GL	GP	286	1333	21	Coia *et al.*, 2012	25	Coia *et al.*, 2013
Circello	CIR	G	FS	2501	700	27	This study	34	Tofanelli, *pers. com.*
Cosenza	COS	O	GP	69131	238	42	Pettener, *pers. com.*	28	Pettener, *pers. com.*
Fiemme Valley	FIE	G	GP	18990	1033	41	Coia *et al.*, 2012	41	Coia *et al.*, 2013
Ladins (Fassa Valley)	LVF	GL	GP	9894	1345	47	Coia *et al.*, 2012	47	Coia *et al.*, 2013
Lucca plain	PIL	O	GP	154928	81	50	This study	50	This study
North Sardinia	NSA	O	GP	67253	440	40	This study	47	Calò *et al.*, 2013
Sappada	SAP	GL	GP	1307	1217	59	Capocasa *et al.*, 2013b	36	Coia *et al.*, 2013
Sauris	SAU	GL	GP	429	1212	48	Capocasa *et al.*, 2013b	29	Coia *et al.*, 2013
Sulcis-Iglesiente	SGL	O	GP	128614	96	50	Robledo *et al.*, 2012	46	Robledo *et al.*, 2012
Timau	TIM	GL	GP	500	830	46	Capocasa *et al.*, 2013b	22	Coia *et al.*, 2013
Trapani-Enna	TEN	O	GP	96834	-	80	Pettener, *pers. com.*	71	Pettener, *pers. com.*
Vagli	VAG	G	FS	995	575	22	This study	23	Tofanelli *pers. com.*

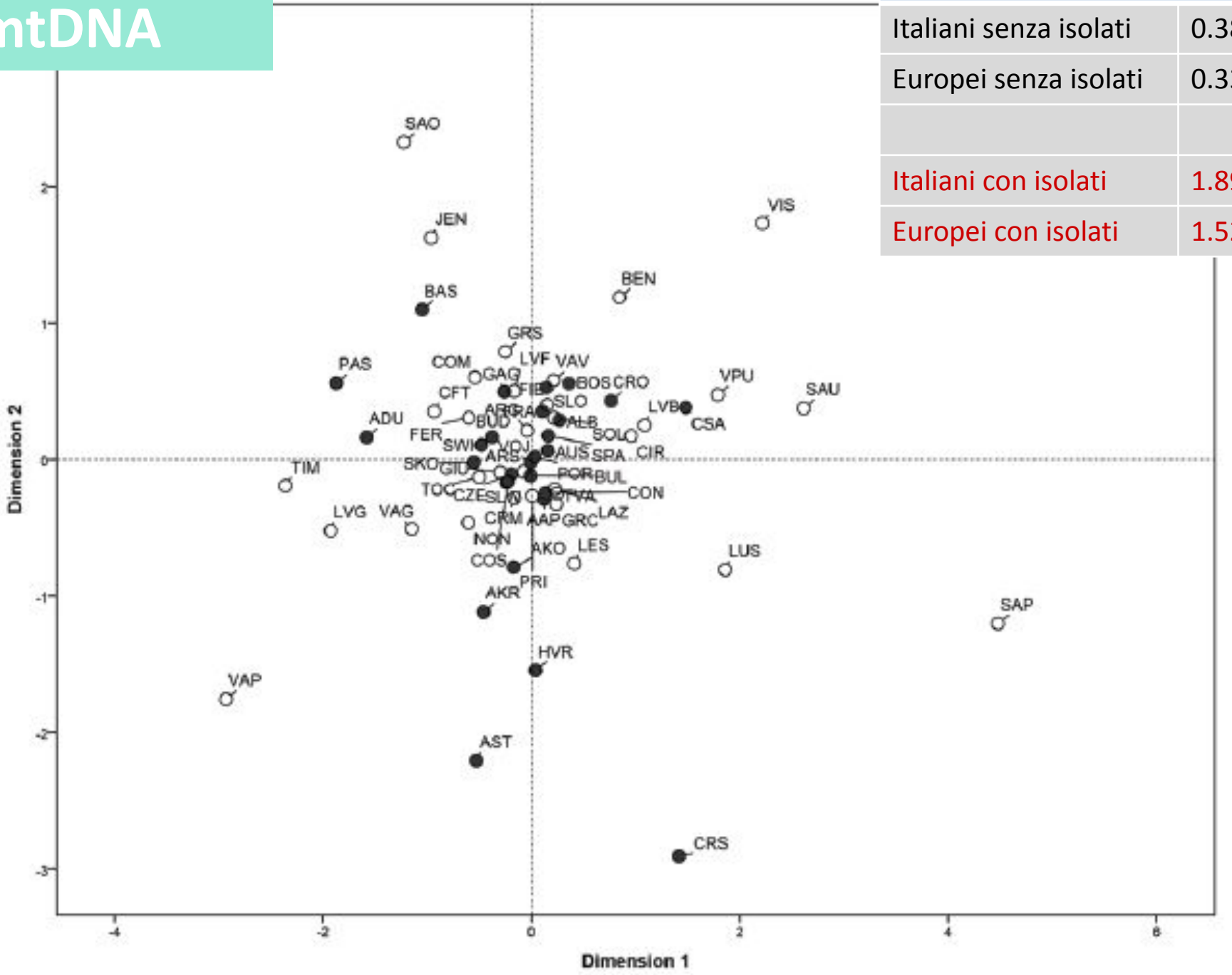
* Source: ISTAT (2011) http://demo.istat.it.

° meters above sea level.



http://www.isita-org.com/iass/Contents/2014vol92/Capocasa/24607994.pdf

mtDNA

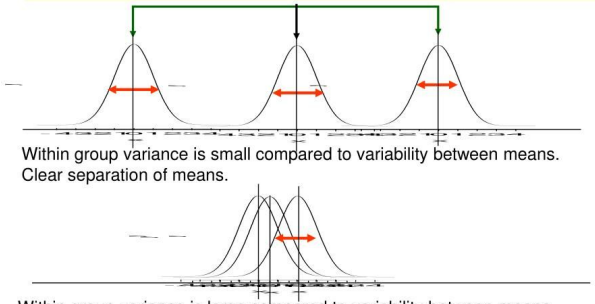


AMOVA (within group)	
Italiani senza isolati	0.38%
Europei senza isolati	0.33%
Italiani con isolati	1.89%
Europei con isolati	1.52%

Analysis of molecular variance (AMOVA)

- MOVA is a method for estimating molecular variation within a species
- It is not ANOVA :
  Don't require the assumption of a normal distribution (Pre-mutational methods)\
- This may contains regions within continent
- Individuals within population in area in region in continent

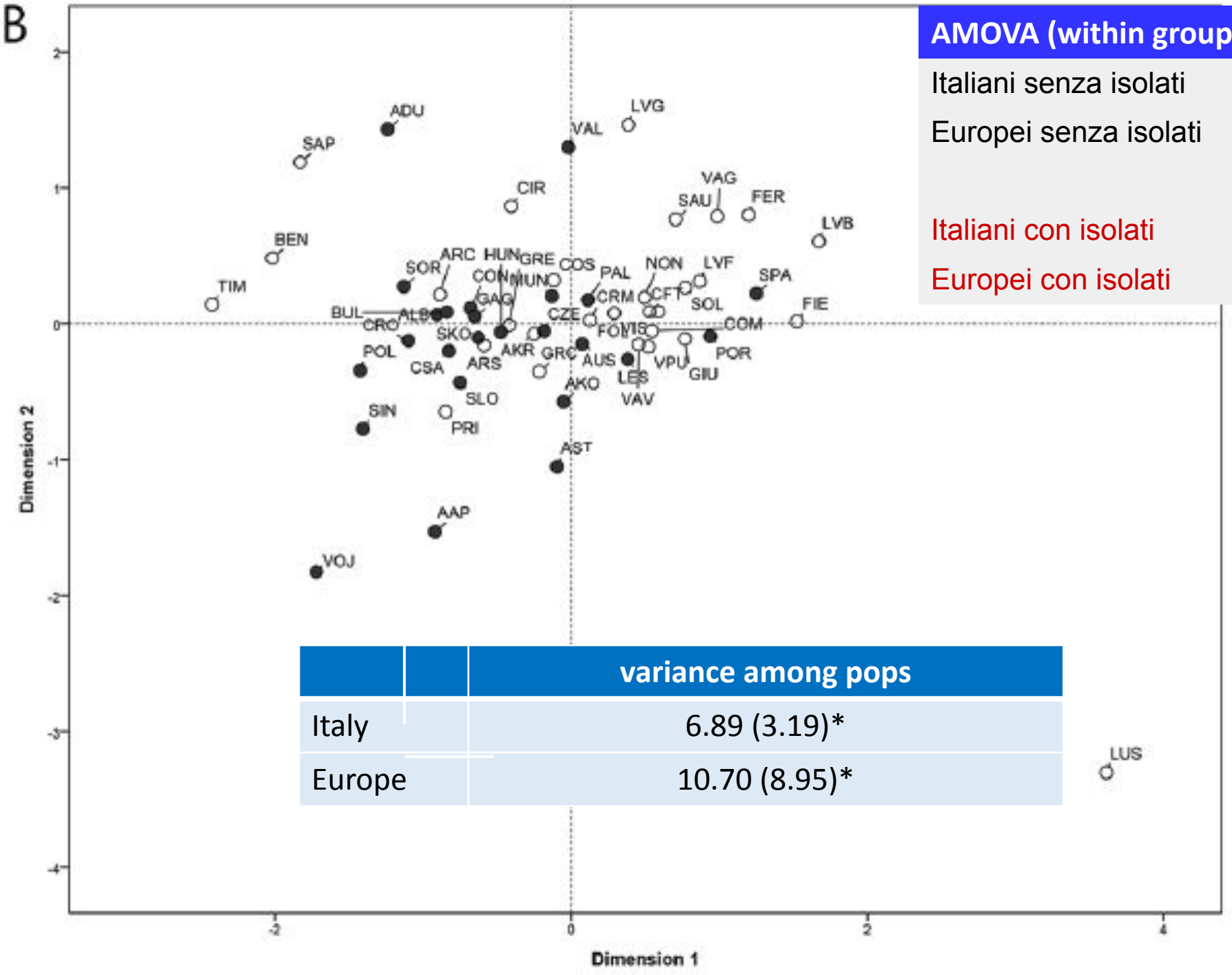
Extreme Situations



Within group variance is small compared to variability between means.
Clear separation of means.

Within group variance is large compared to variability between means.
Unclear separation of means.

B



AMOVA (within group)	
Italiani senza isolati	3.19%
Europei senza isolati	6.89%
Italiani con isolati	8.95%
Europei con isolati	10.60%

	variance among pops
Italy	6.89 (3.19)*
Europe	10.70 (8.95)*

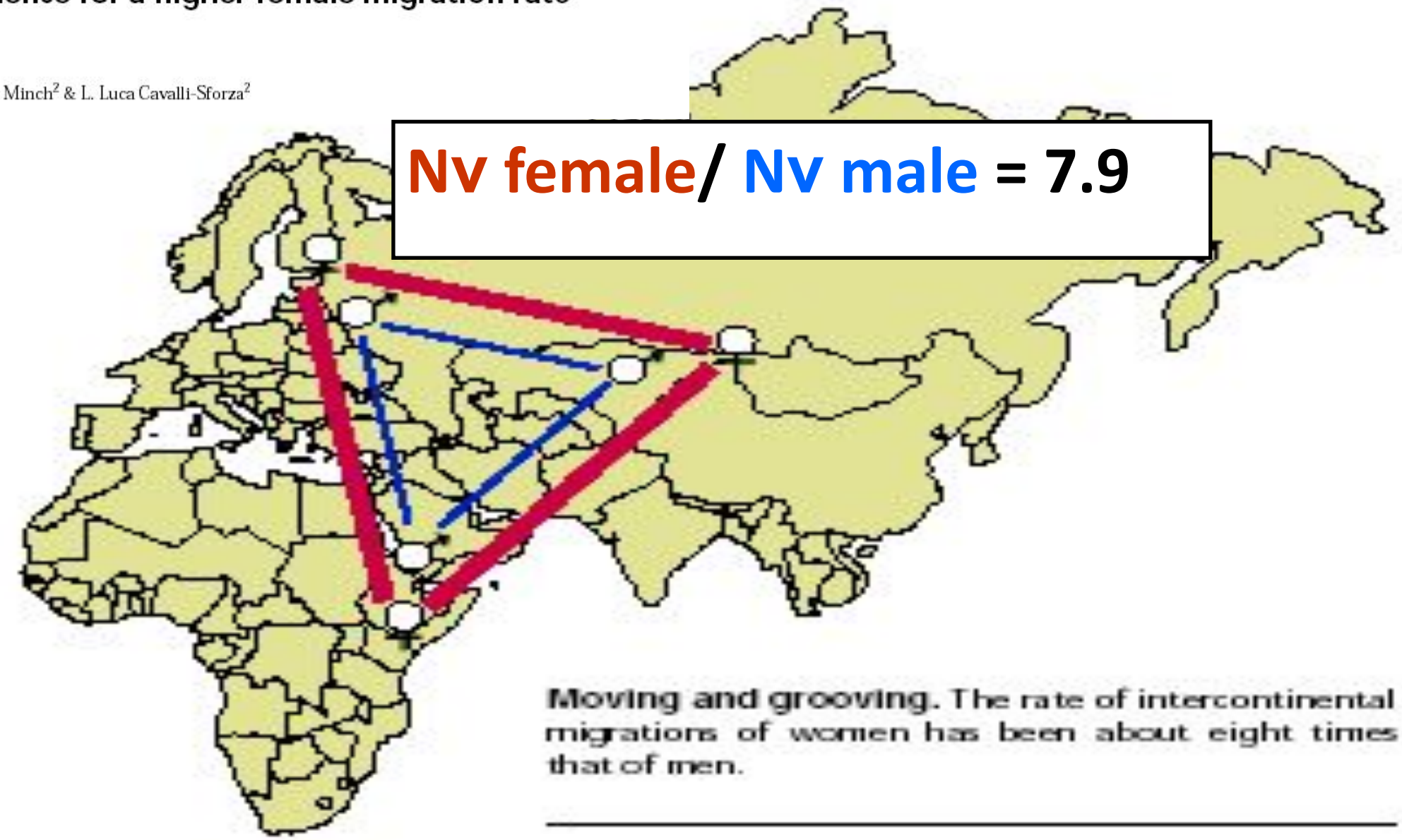
chr Y, 5 STRs

letter 1998 Nature America Inc. • http://genetics.nature.com

Genetic evidence for a higher female migration rate in humans

Mark T. Seielstad[1], Eric Minch[2] & L. Luca Cavalli-Sforza[2]

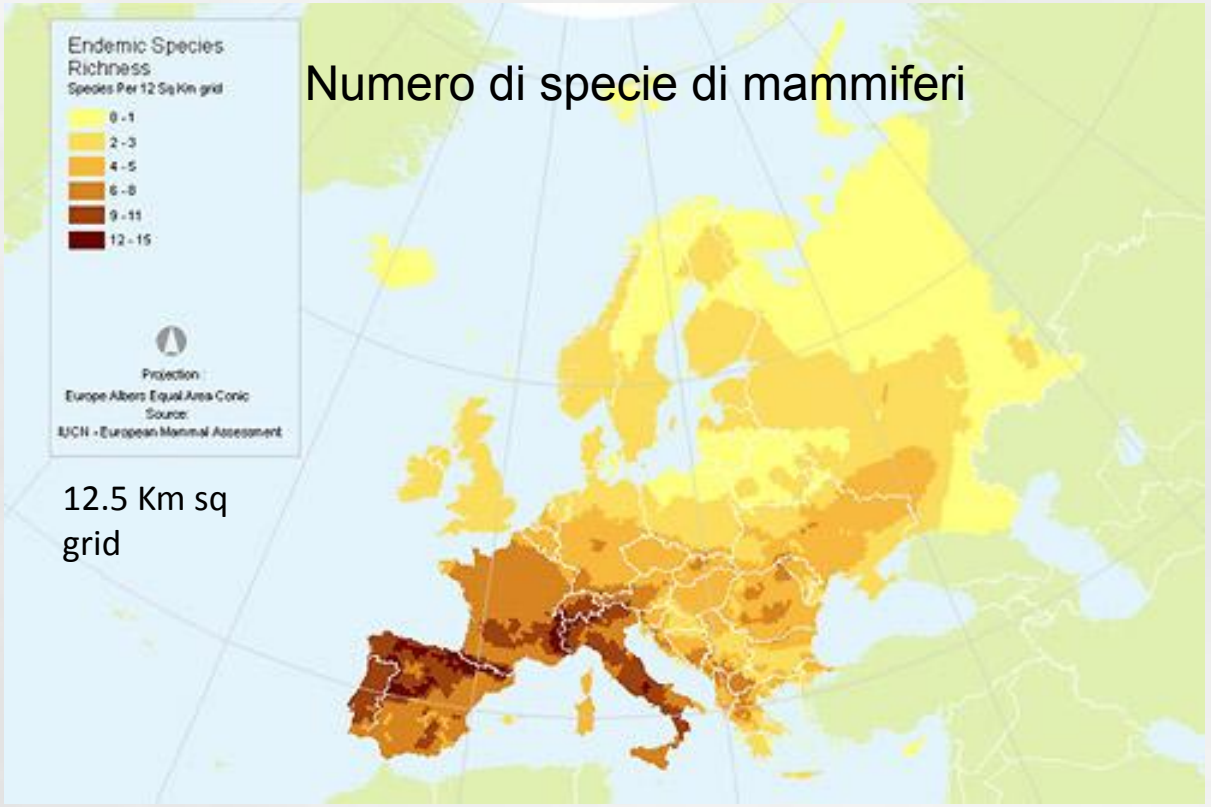
Nv female/ Nv male = 7.9



Moving and grooving. The rate of intercontinental migrations of women has been about eight times that of men.

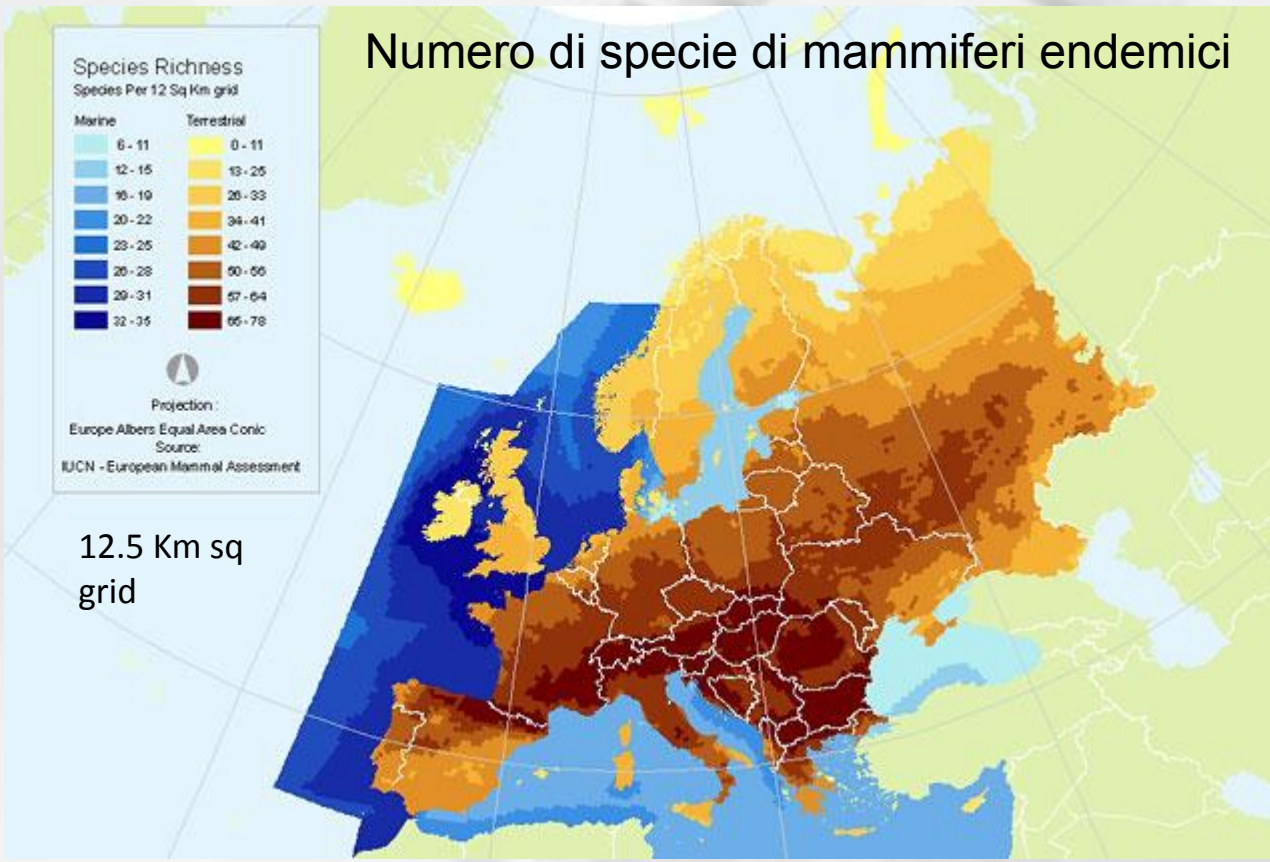
Isolamento nelle popolazioni umane

L'Italia è uno dei paesi europei più ricchi di biodiversità, sia animale che vegetale del bacino Mediterraneo e dell'Europa in generale.



Numero di specie di mammiferi

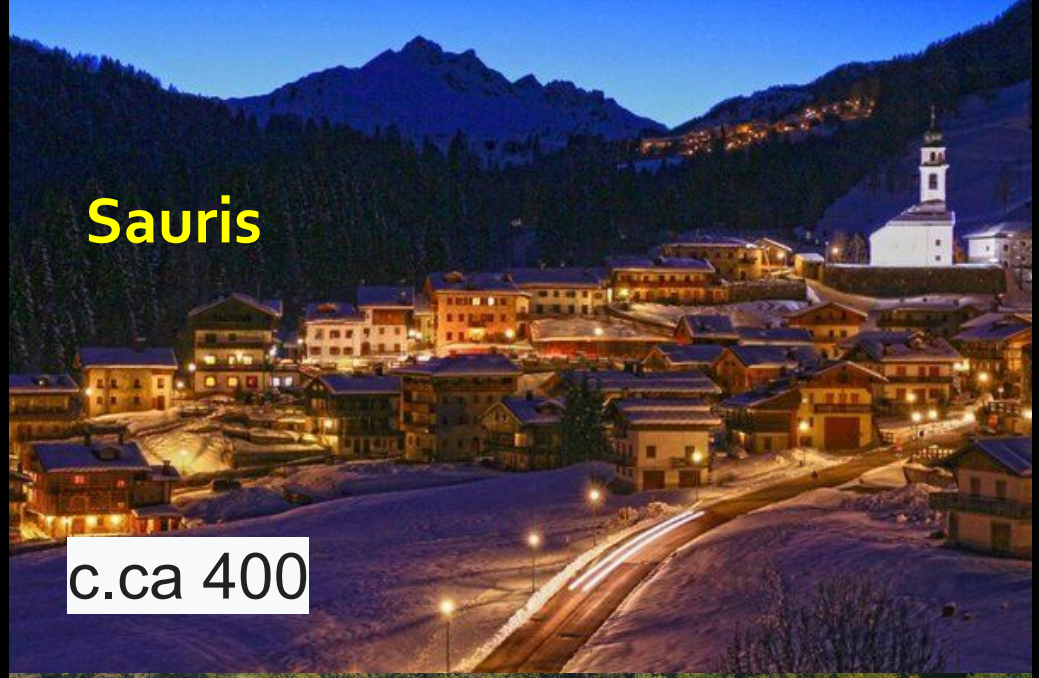
12.5 Km sq grid



Numero di specie di mammiferi endemici

12.5 Km sq grid

1. Isolation among isolates?


Sauris
c.ca 400


Sappada
c.ca 1.300


Timau
c.ca 350


Lessinia
c.ca 4.400

Autosomal markers

GenoChip 2.0

SNP array dedicated to genetic anthropology

150,000 SNPs

- mtDNA — 3,300 SNPS
- Y chr — 12,000 SNPS
- Autosomal — 130,000 SNPS

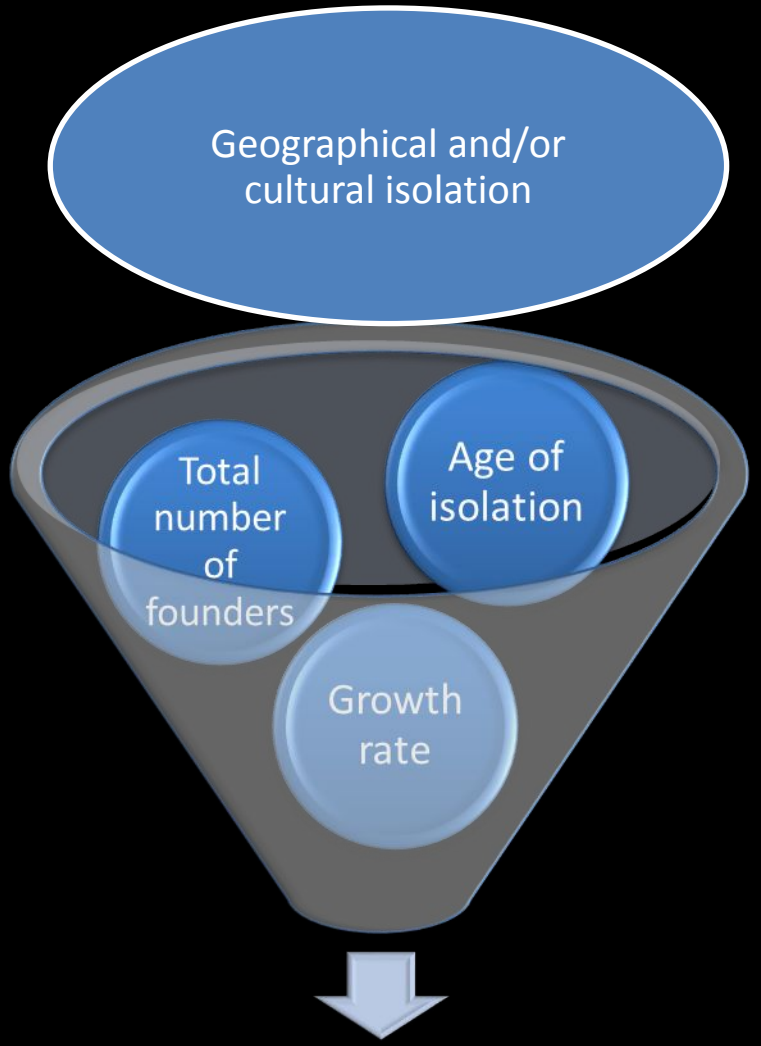
- SNPs: *genotyping success rate* > 90%
- Individuals: *genotyping success rate* >92 %
- IBD >0.185

Merging

- HGDP
- Hellenthal et al 2014
- Behar et al 2012

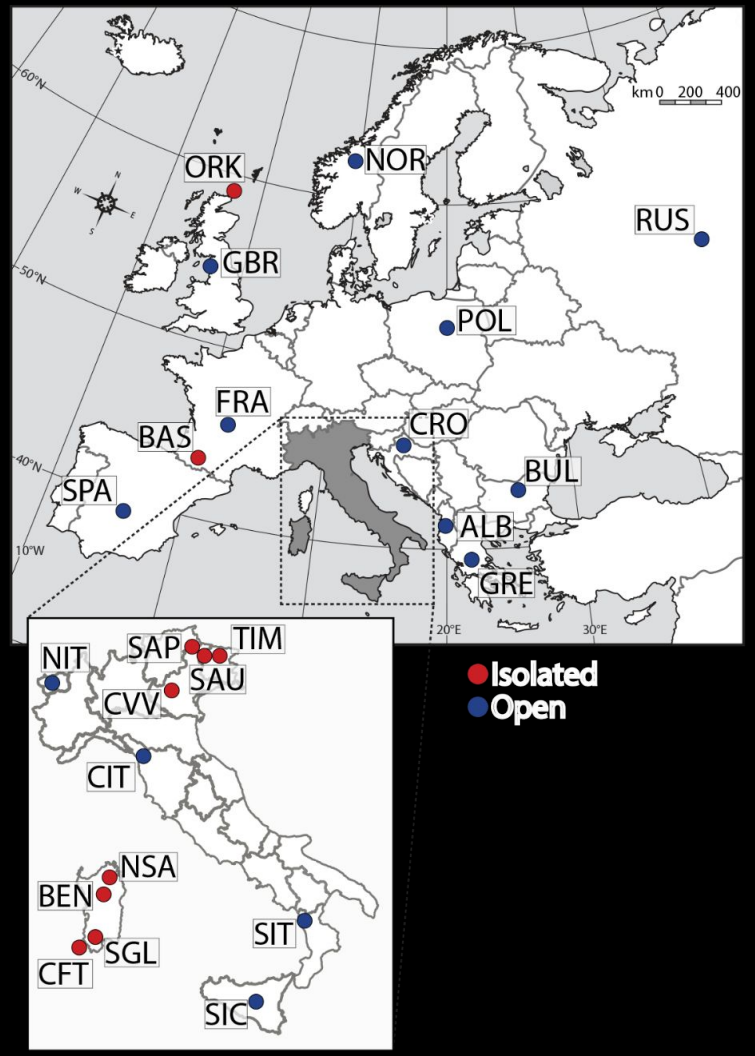
87,818 SNPs

Background



European country	Isolate population	Isolate type	Population size	Age
France	Basque	Cultural	730.000	Old
	Corsicans	Geographic	280.000	Old
Germany	Sorbs	Cultural	120.000	Young
Finland	Saami	Cultural	2.600	Very old
Iceland	Icelanders	Geographic	240.000	Old
Italy	Albanians	Linguistic	100.000	Young
	Croats	Linguistic	2.000	Young
	Greeks	Linguistic	20.000	Young
	Ladin	Geo-Linguistic	35.000	Young
	Sardinian	Geographic	1.500.000	Old
	Walser	Linguistic	3.500	Young
UK	Orcadians	Geographic	20.000	Young

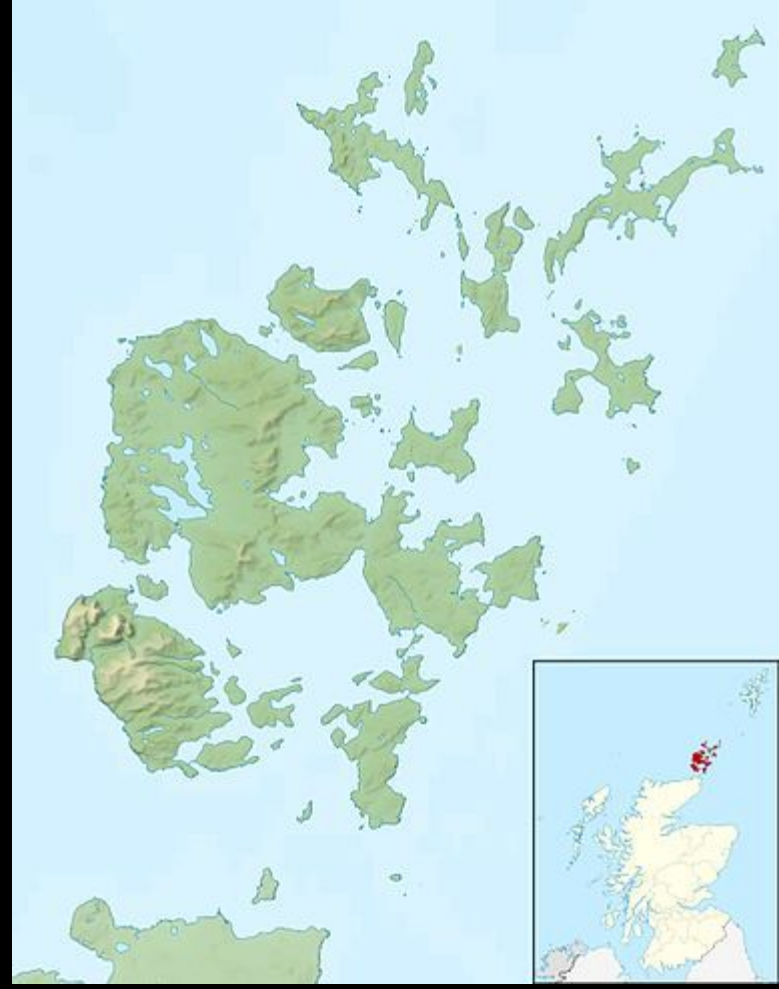
Samples



POPULATION	ACR	N	Reference
Alpine linguistic islands			
Sappada	SAP	24	This study
Sauris	SAU	10	This study
Timau	TIM	24	This study
Velo Veronese	CVV	33	This study
Sardinian isolates			
North Sardinia	NSA	25	This study
Sulcis Iglesiente	SGL	25	This study
Benetutti	BEN	25	This study
Carloforte	CFT	25	This study
European isolates			
Orkney	ORK	15	HGDP
Basque	BAS	24	HGDP

POPULATION	ACR	N	Reference
South Europe			
Albania	ALB	24	P.C.
Croatia	CRO	20	Behar et al 2012
Greece	GRE	20	Hellenthal et al 2014
Spain	SPA	34	Hellenthal et al 2014
East Europe			
Bulgary	BUL	31	Hellenthal et al 2014
Poland	POL	32	Hellenthal et al 2014
Russia	RUS	25	HGDP
West Europe			
France	FRA	29	HGDP
North Europe			
British isles	GBR	16	Hellenthal et al 2014
Norway	NOR	18	Hellenthal et al 2014
Italy			
Aosta	NIT	22	This study
Piana di Lucca	CIT	25	P.C.
South Italy	SIT	18	Hellenthal et al 2014
Sicily	SIC	20	Hellenthal et al 2014

Orkney islanders

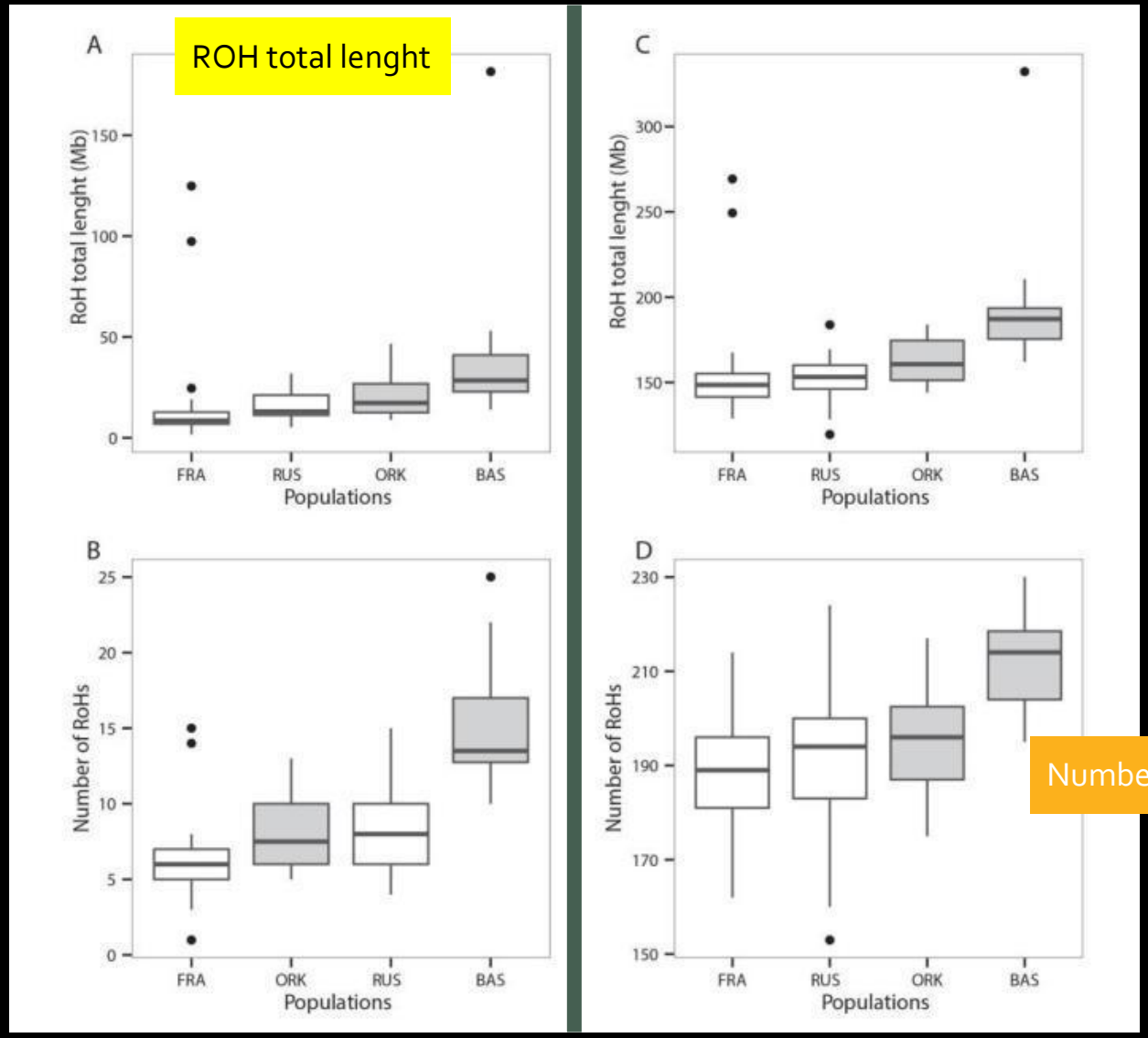


PRELIMINARY TEST Validation of the panel

Test the power of our set of of SNPs (reduced panel, 87,818 SNPs) in detecting signals of isolation

- Human Genome Diversity Project Panel: 647.789 SNPs
- 4 populations:
  - Basque e Orkney (isolates)
  - France e Russia (open)
- 2 measures of intra-population variation:
  - *Runs of Homozygosity* (RoHs)
  - *Identity by State* (IBS)

Validazione del pannello di marcatori

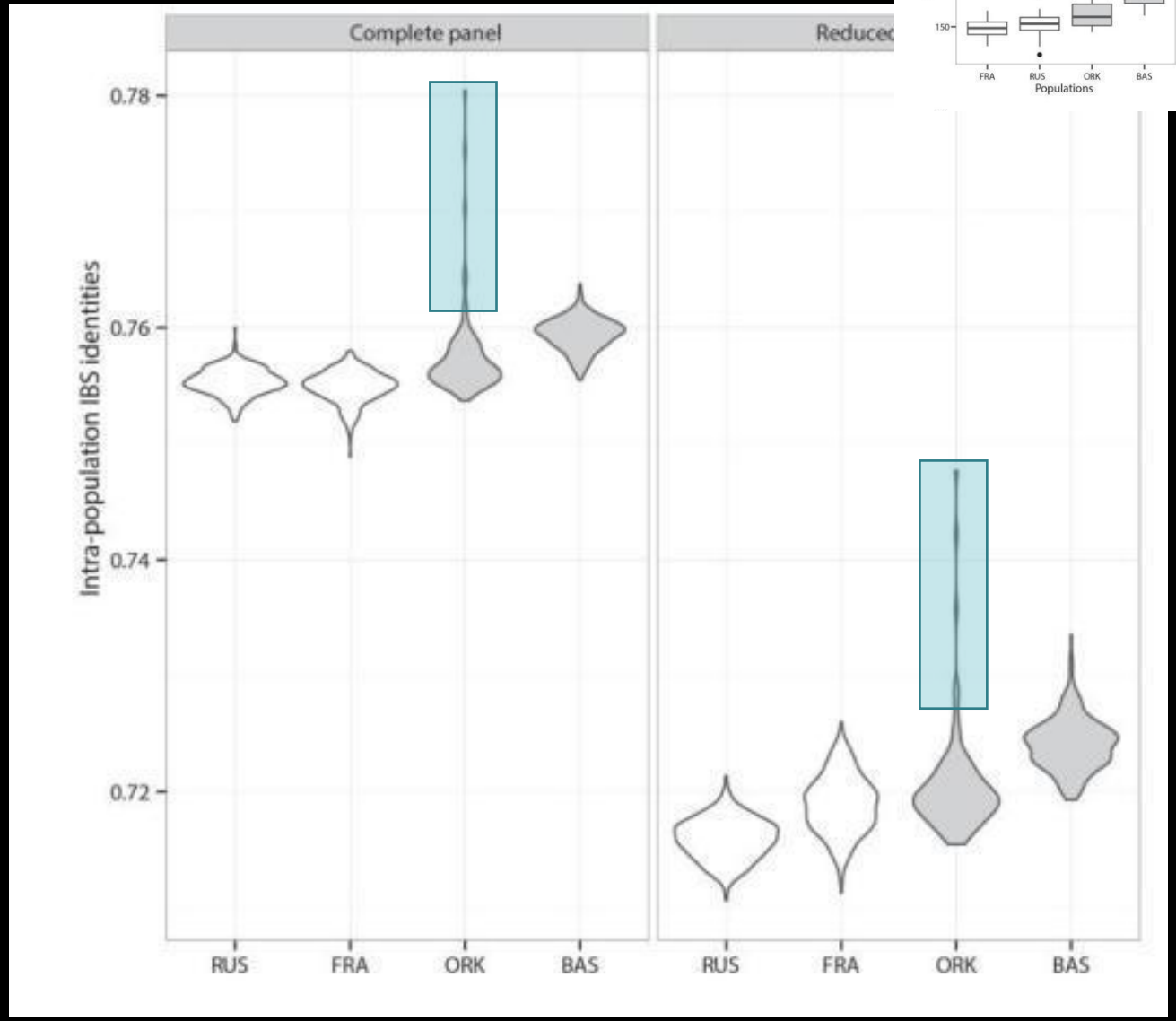


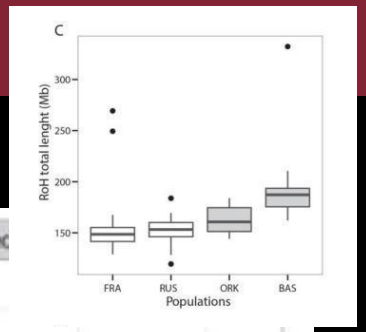
Number of ROH

Distributions of the total length and number of RoH found in the four HGDP populations using the reduced loci set (87,818 SNPs), frames A and B, and the complete loci set (647,789 SNPs), frames C and D.

- values different in absolute terms
- pattern of isolation remained substantially unchanged

Validazione del pannello di marcatori





Violin plots of the distributions of intra-population pairwise IBS identities in the four populations analysed using the complete and reduced SNP panels

- isolated population show higher average values than the opens in both cases
- isolated population exhibit the upper tail with both panels

IBS (Identity-by-State) refers to the state where two individuals share the same allele at a particular locus (genomic position), regardless of whether they inherited that allele from a common ancestor.

Analysis of genomic diversity

Intra-popolazione

- *Runs of Homozygosity* (RoHs)
- *Identity by State* (IBS)
- Shared *identical-by-descent* (IBD)
- *Linkage Disequilibrium* (LD) blocks

Inter-popolazione

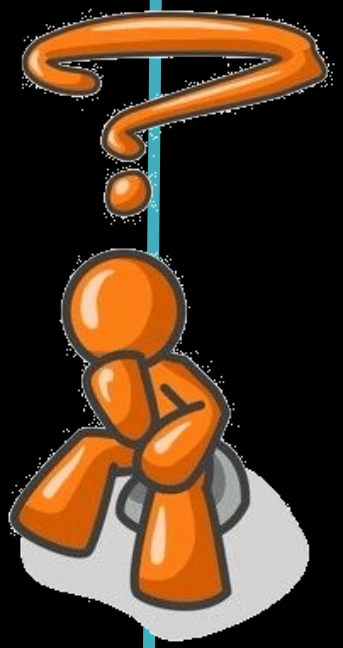
- Estimation of genetic ancestry (ADMIXTURE)
- Genetic structure and gene flow
- MDS based on IBS distance

To what extent are populations with and without isolation factors different from each other

https://www.nature.com/articles/srep41614

Background

Open populations



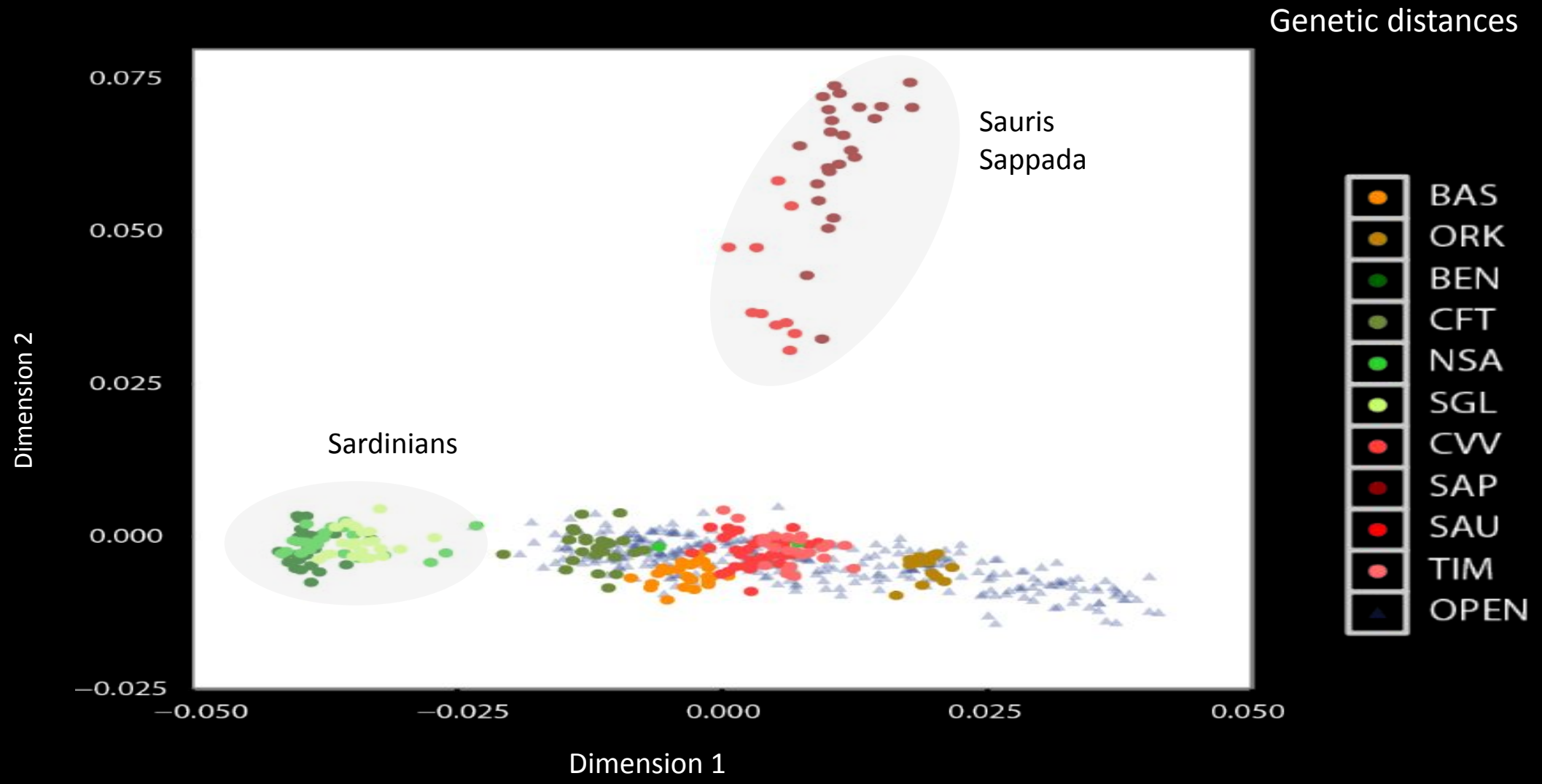
Isolated populations

+	Gene flow	-
-	Genetic drift	+
-	Inbreeding	+
+	Intra-population diversity	-
-	Inter-population diversity	+

THE GENOGRAPHIC PROJECT

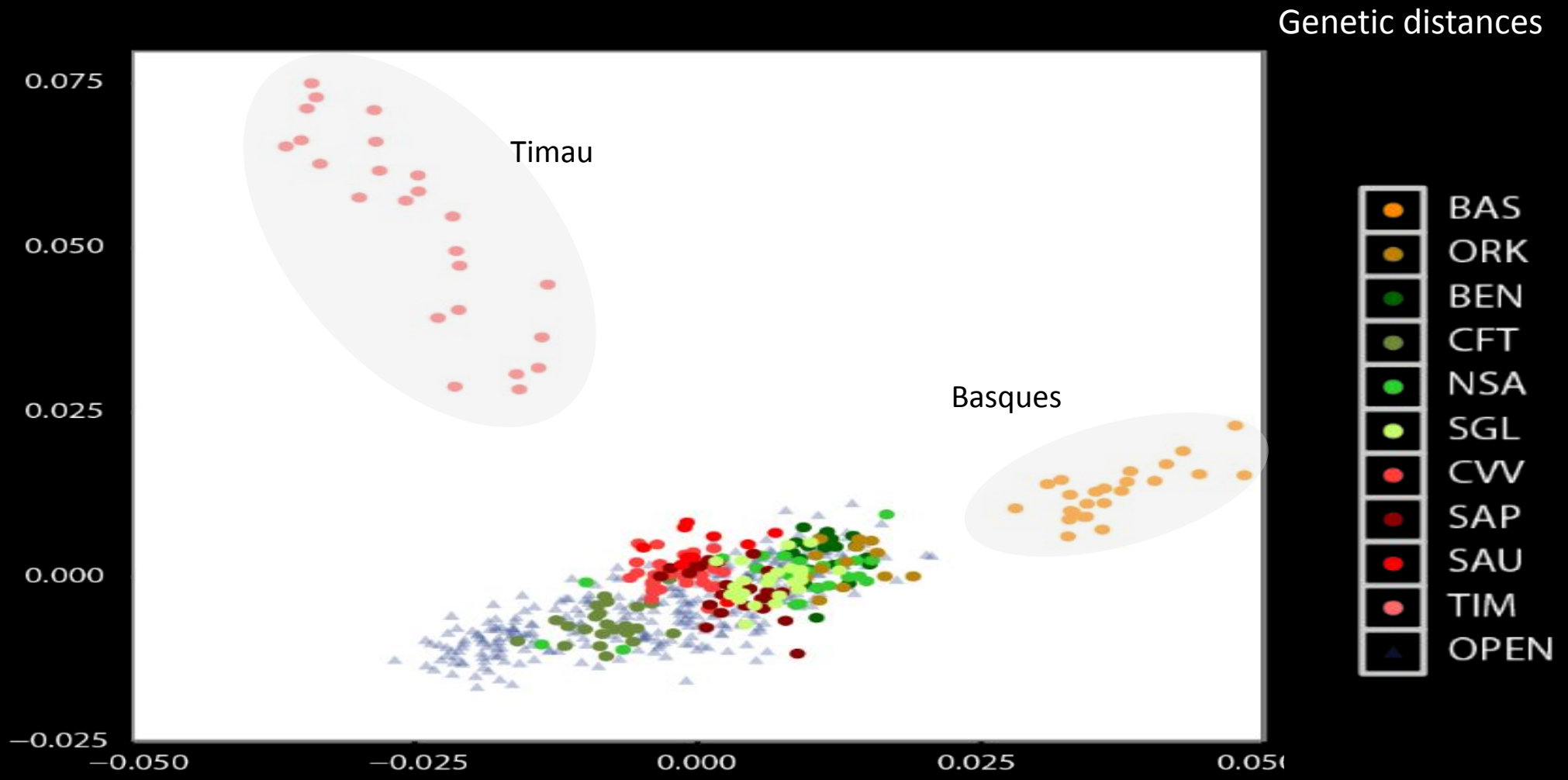
NATIONAL GEOGRAPHIC

GenoChip 2.0 87,818 SNPs



Anagnostou et al. submitted

GenoChip 2.0 87,818 SNPs



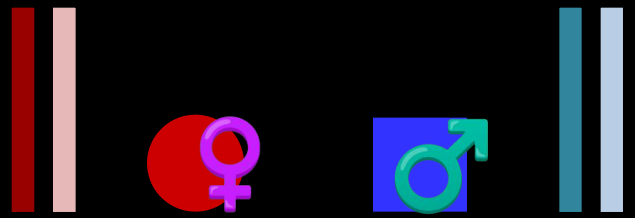
Anagnostou et al. submitted

WHAT are genomic features

- Runs of homozygosity (RoH)
- Identity dy descent (IBD)
- STRUCTURE

WHAT are genomic features: RoH

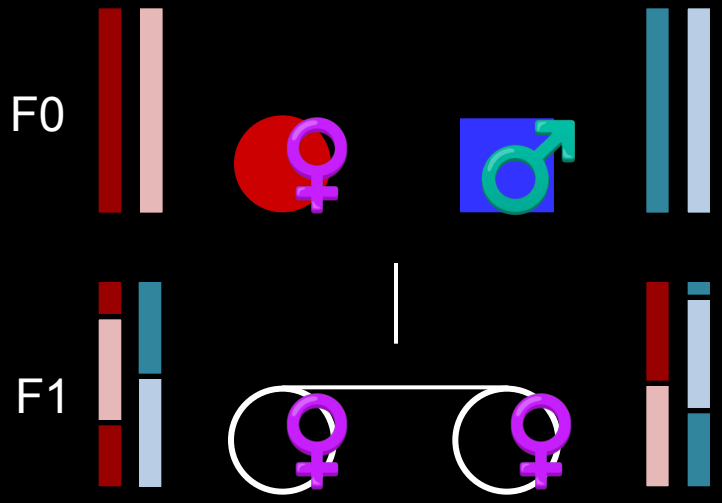
F0



The female common ancestor is red. The chromosome inherited from one of her parents is colored red, and the chromosome inherited from her other parent is colored pink. The male common ancestor is blue. The chromosome inherited from one of his parents is colored dark blue, and the chromosome inherited from his other parent is colored light blue.

Modified from McQuillan et al 2008

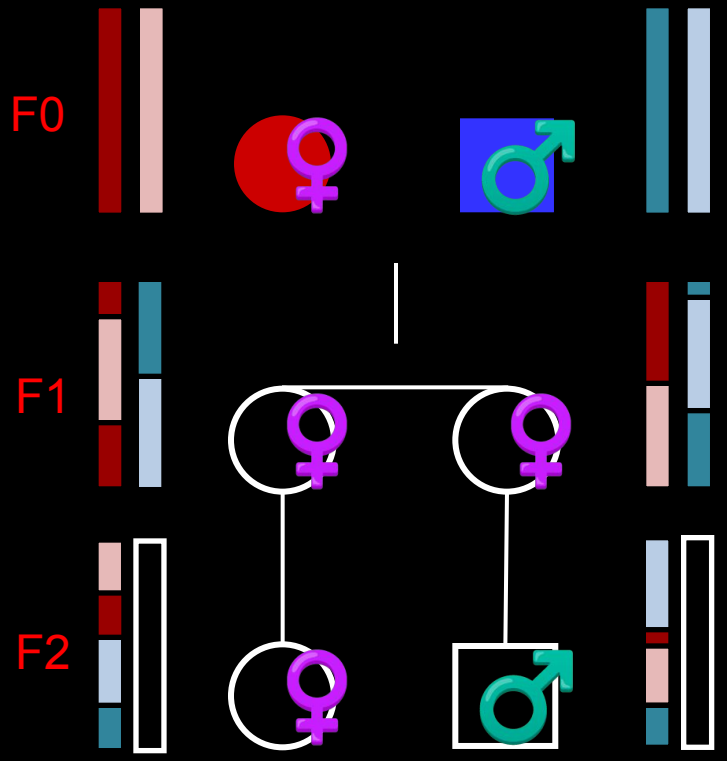
WHAT are genomic features: RoH



The second generation are sisters. They share around 50% of their chromosomes IBD. The segments colored red and pink are segments inherited from their mother, and the segments colored dark and light blue are segments inherited from their father.

Modified from McQuillan et al 2008

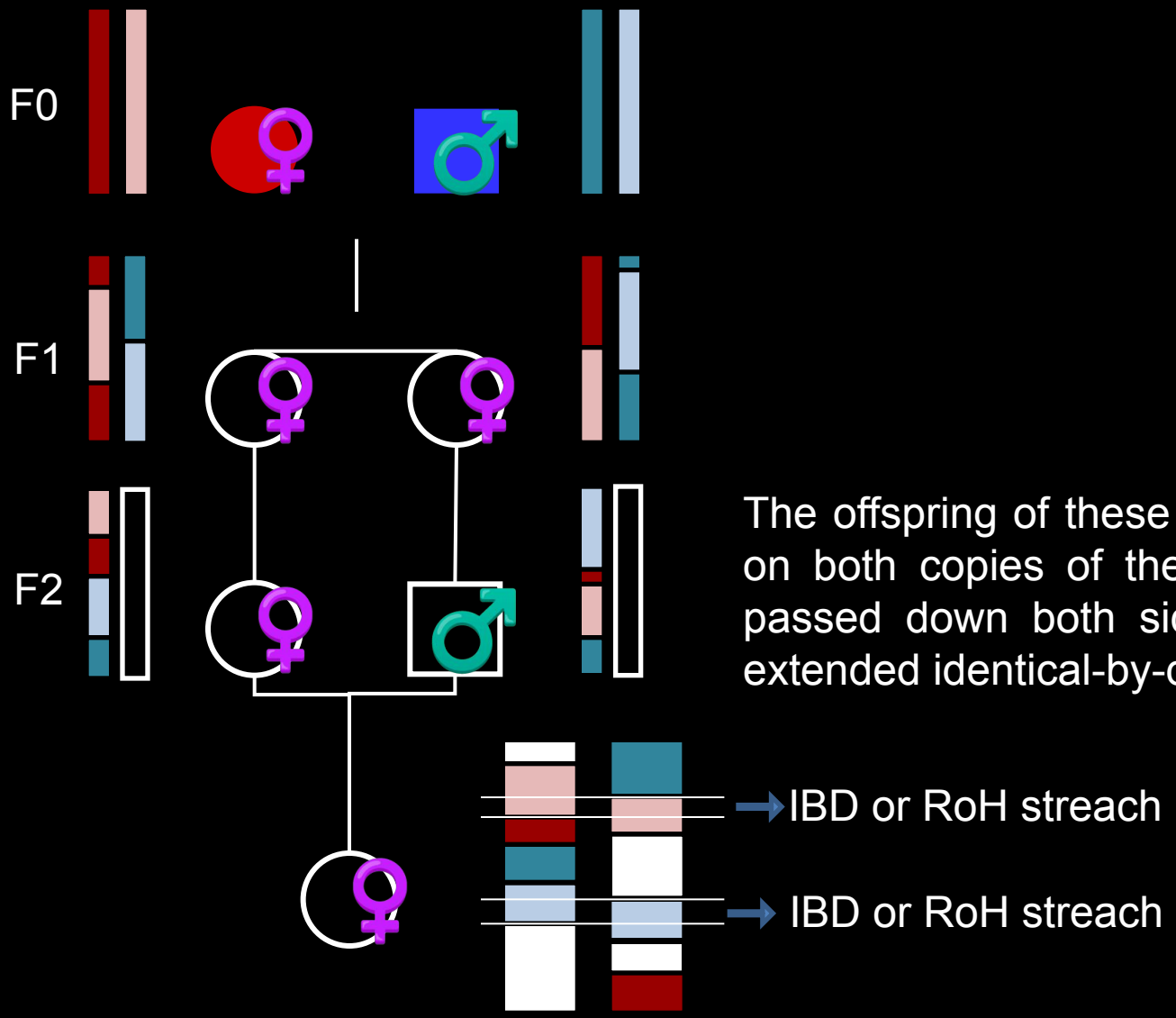
WHAT are genomic features: RoH



The third generation are first cousins. In each case, the second (white) chromosome derives from their fathers (not shown), the red and pink segments are inherited from their maternal grandmother, and the dark and light blue segments are inherited from their maternal grandfather.

Modified from McQuillan et al 2008

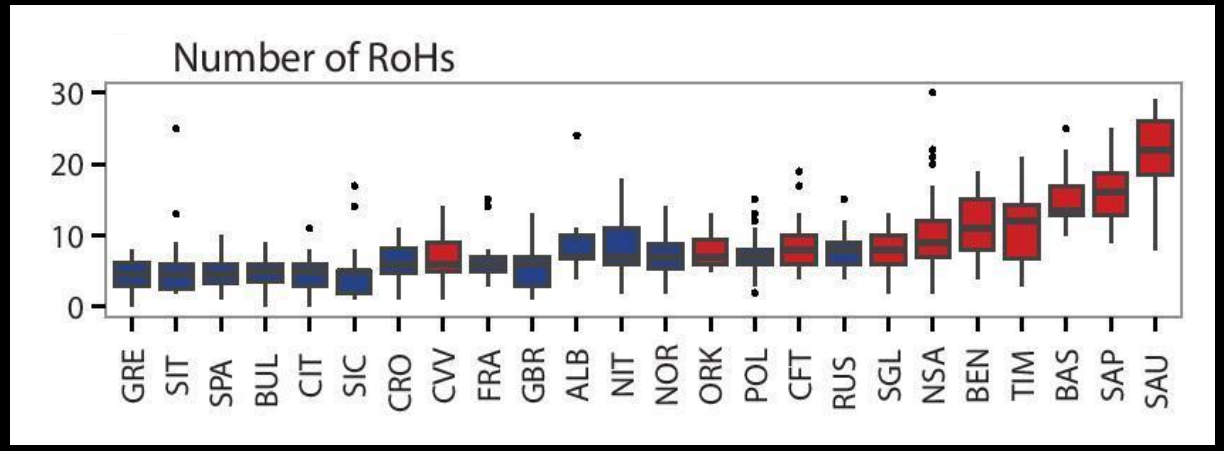
WHAT are genomic features: RoH

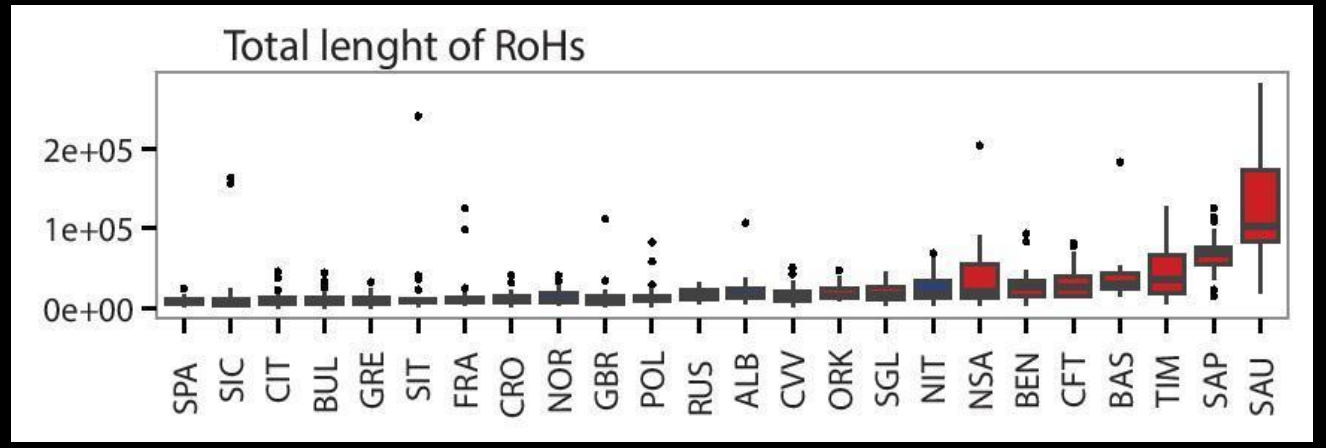


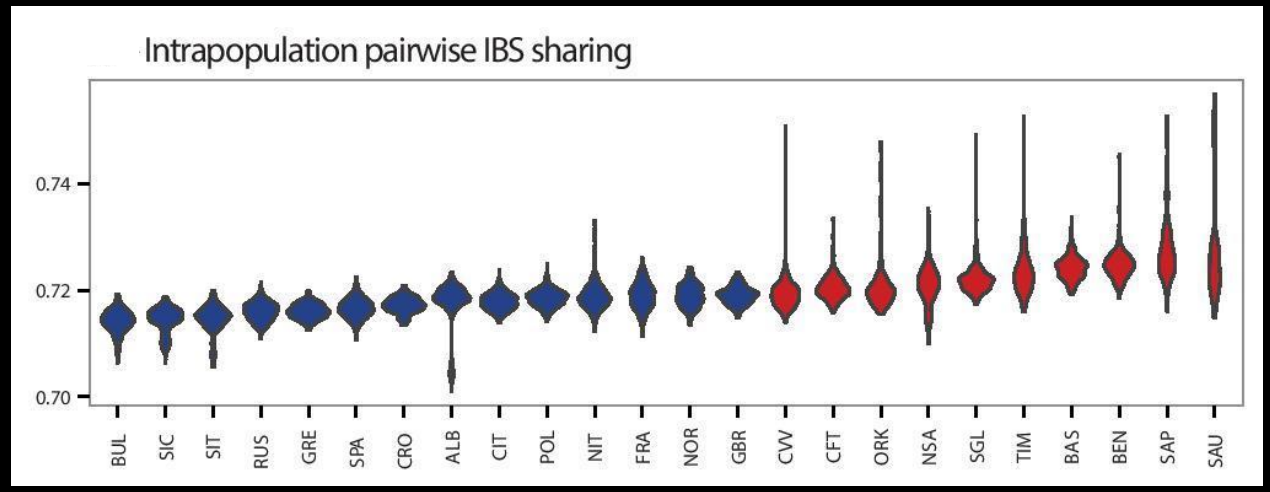
The offspring of these first cousins has segments inherited from both founders on both copies of the chromosome. Where the same segments have been passed down both sides of the pedigree, the offspring of first cousins has extended identical-by-descent tracts or runs of homozygosity.

Modified from McQuillan et al 2008

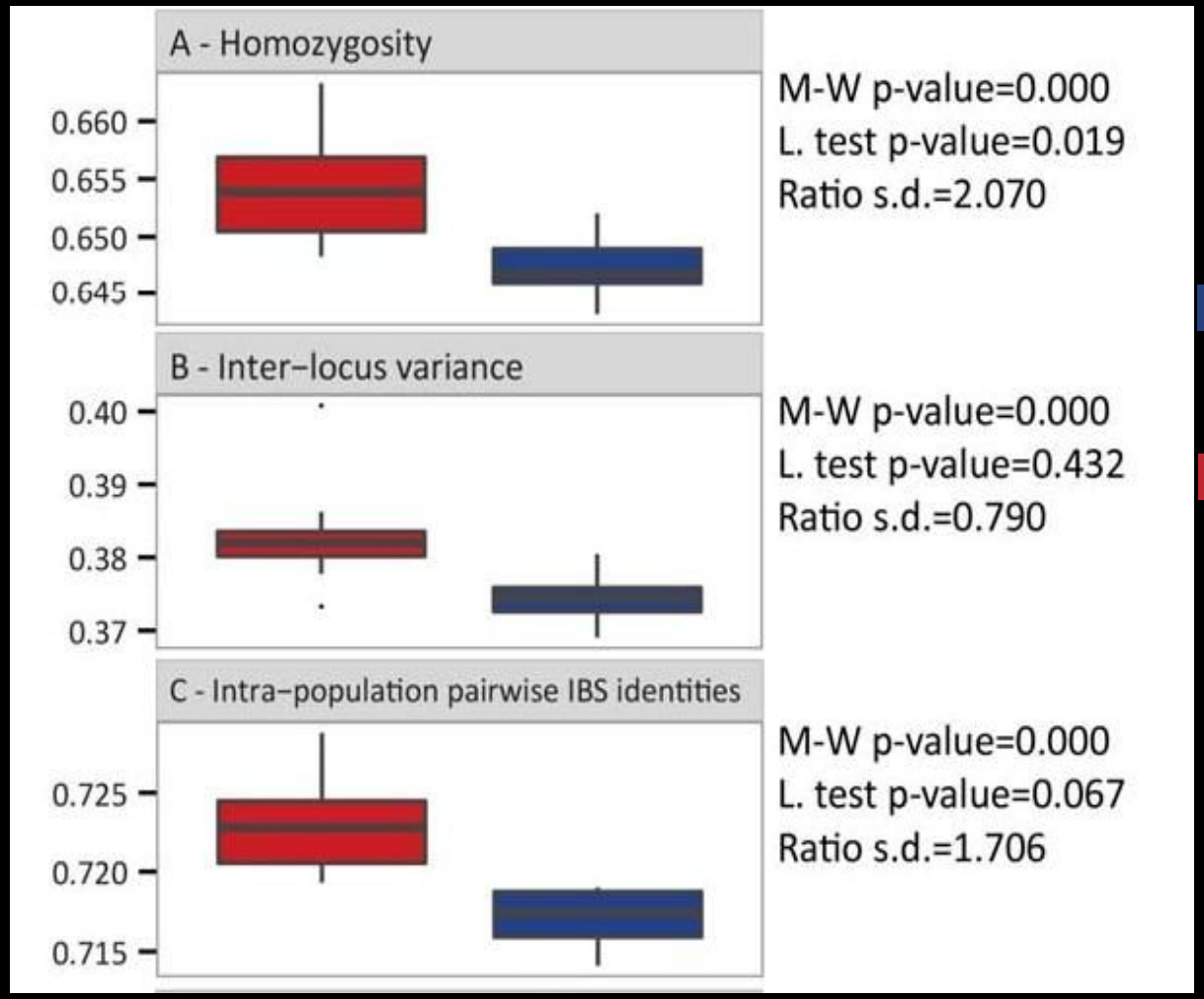
Genomic variation in open and isolated populations





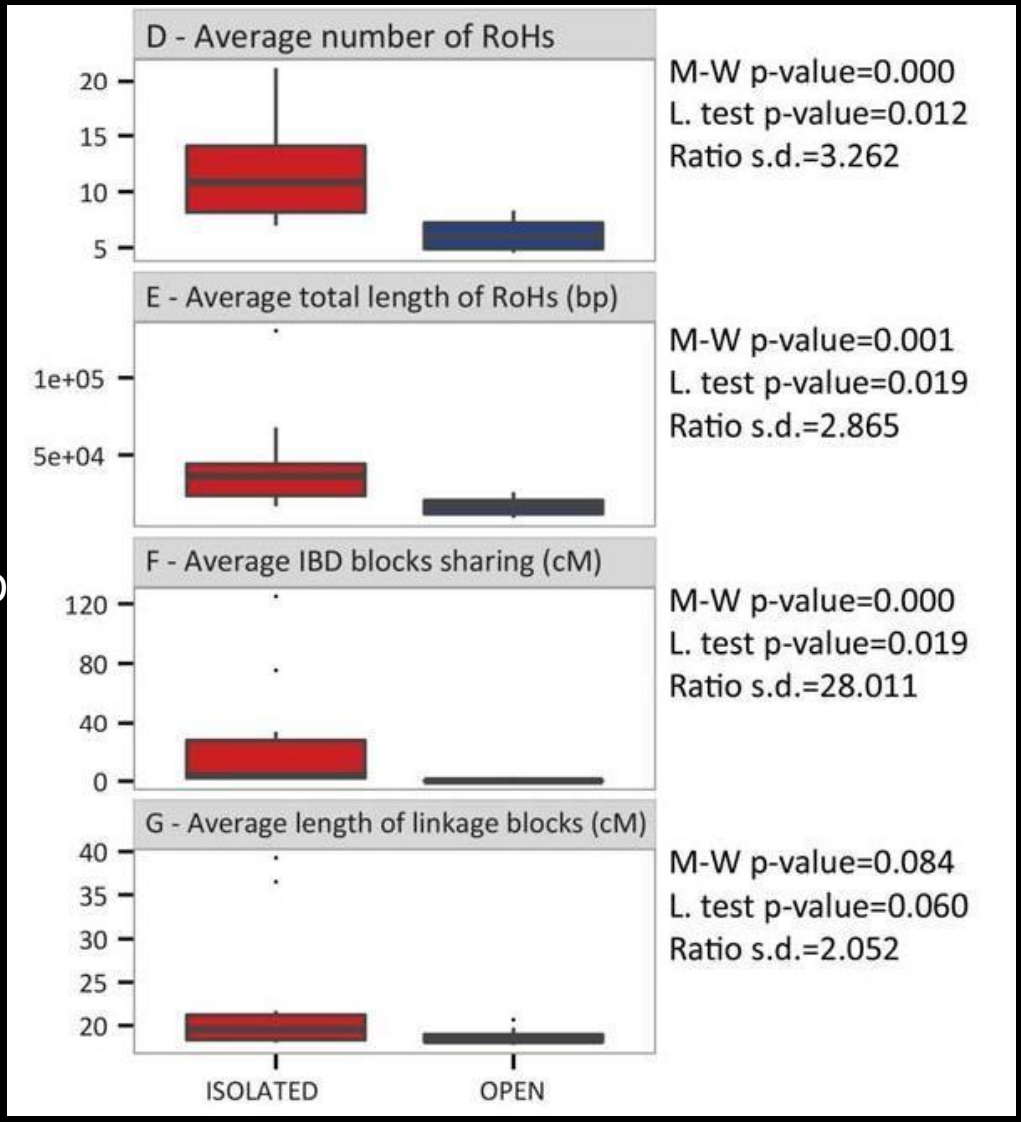


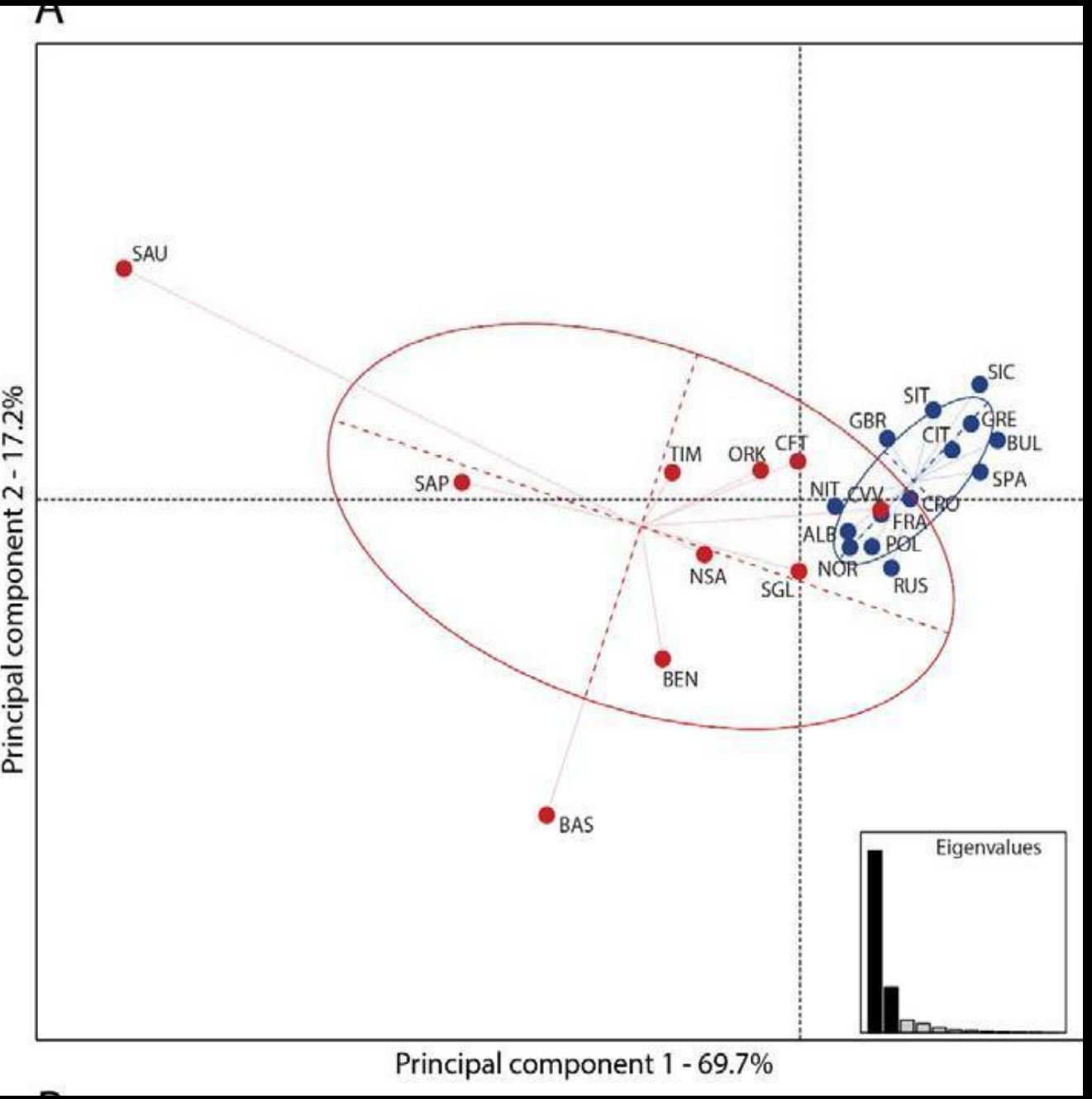
Genomic variation in open and isolated populations



OPEN

ISOLATED





Principal component plots based on intra-population measures

1st component

- separates isolates from open pop
- correlated with *inbreeding* and *drift*

2nd component

- correlated with the effective population size of the isolated population

Conclusions

- *New insights into the genomic diversity of isolated populations:*
  - ✓ Greater dispersion of isolated than open populations in the distribution of intra and inter-population analysis
  - ✓ Local pattern of isolation
- *Continuity rather than dichotomy:*
  - ✓ Our analysis highlights a continuous pattern of genomic variation among populations that have been categorized as ‘open’ and ‘isolated’
- *Heterogeneity within isolated population*:
  - ✓ Observation that contrasts the classic definition of "isolated" as a highly homogeneous population within it. It is necessary to test models that can explain it.

Anagnostou P, Dominici V, Battaggia C, Lisi A, Sarno S, Boattini A, et al. (2019) Inter-individual genomic heterogeneity within European population isolates. PLoS ONE 14(10): e0214564. https://doi.org/10.1371/journal.pone.0214564

A number of studies carried out since the early ‘70s has investigated the effects of isolation on genetic variation within and among human populations in diverse geographical contexts. However, no extensive analysis has been carried out on the heterogeneity among genomes within isolated populations. This issue is worth exploring since events of recent admixture and/or subdivision could potentially disrupt the genetic homogeneity which is to be expected when isolation is prolonged and constant over time.

what happens with patterns of intra-population diversity passing from summary statistics to interindividual heterogeneity?

Distribution of inter-individual heterogeneity values across populations and Mann-Whitney U test.

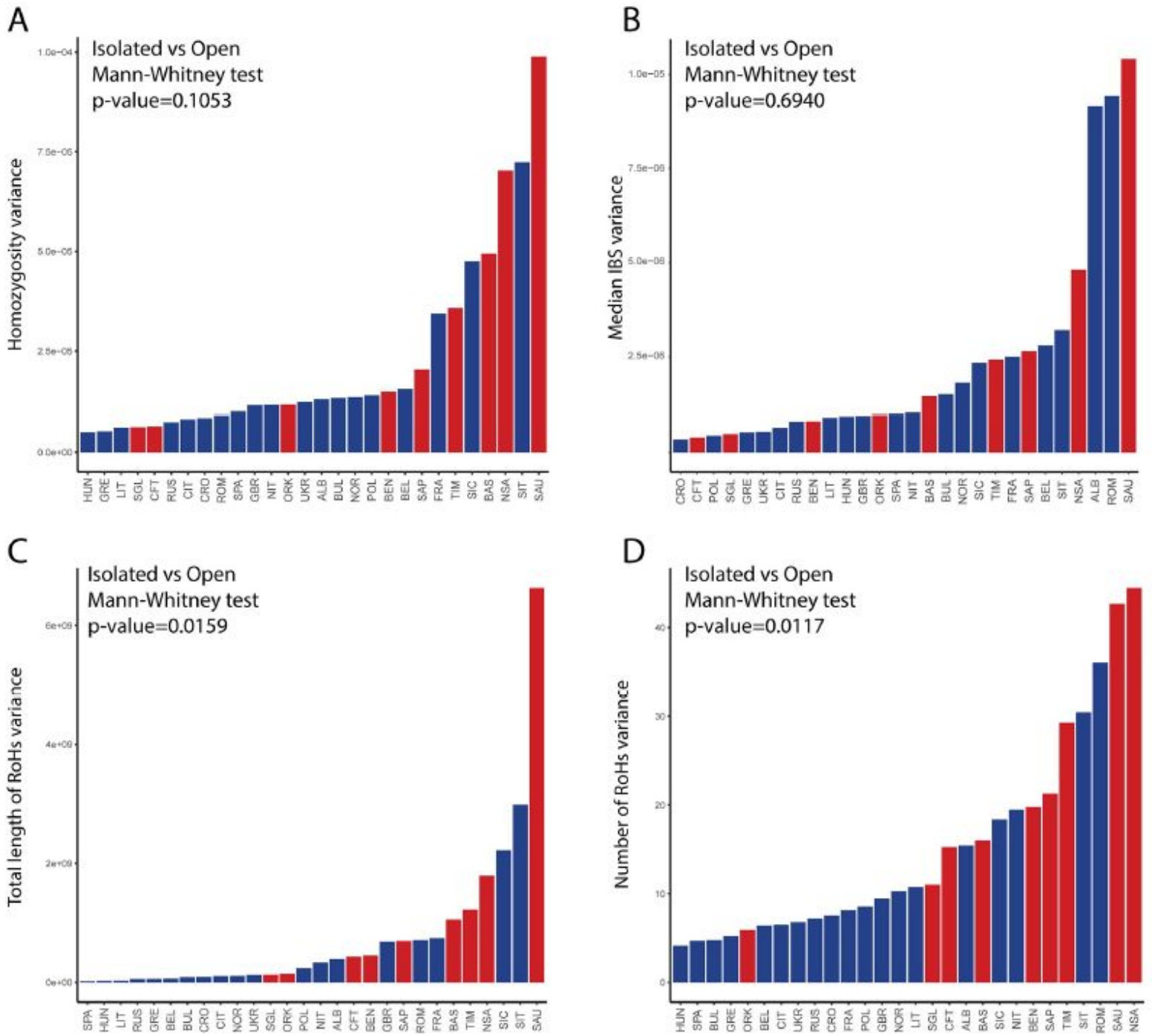
Comparison between isolated (red) and open (blue) populations for

homozygosity (A)

median values of intra-population IBS (B)

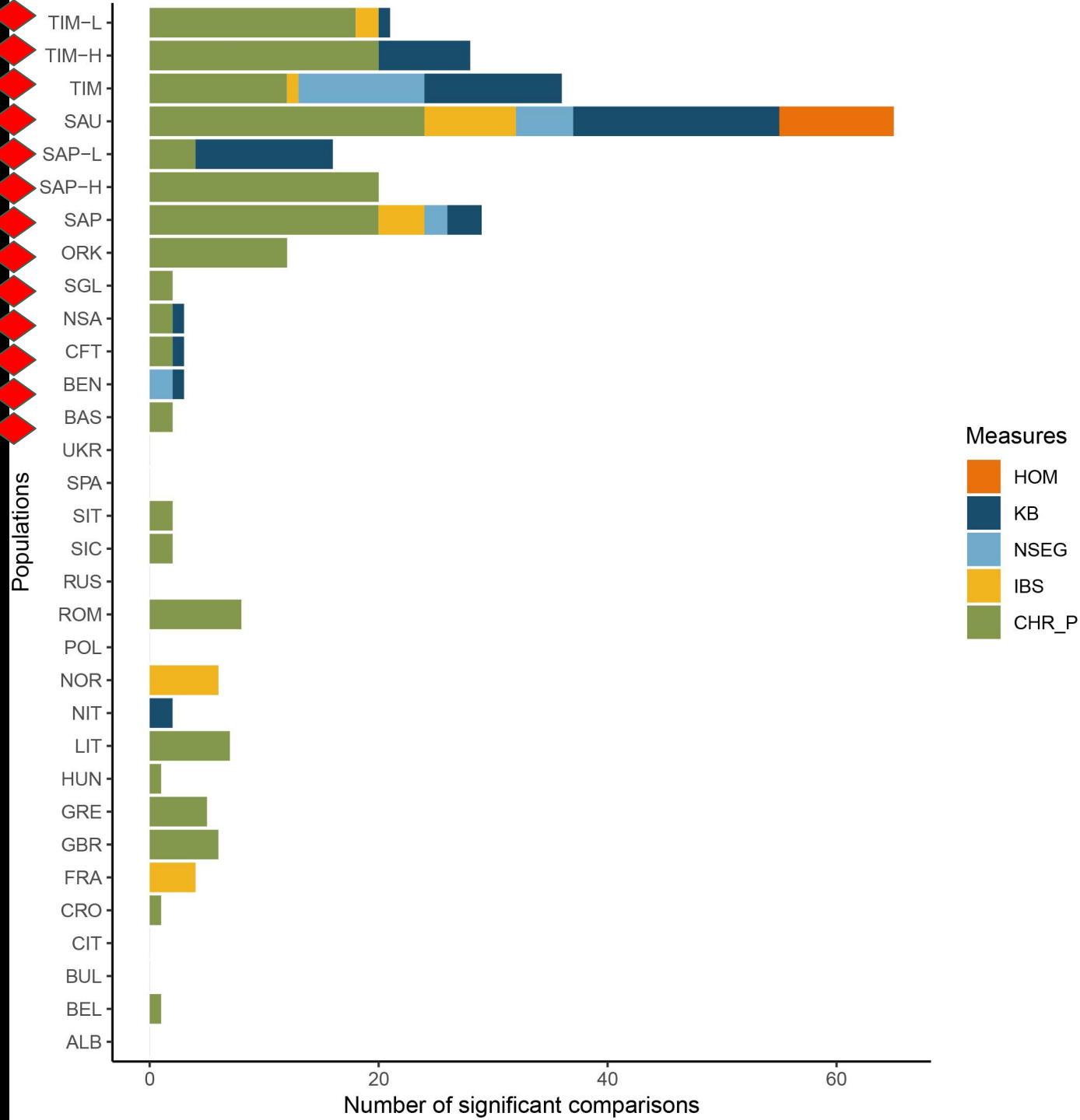
number of RoHs (C)

total length of RoHs (D).



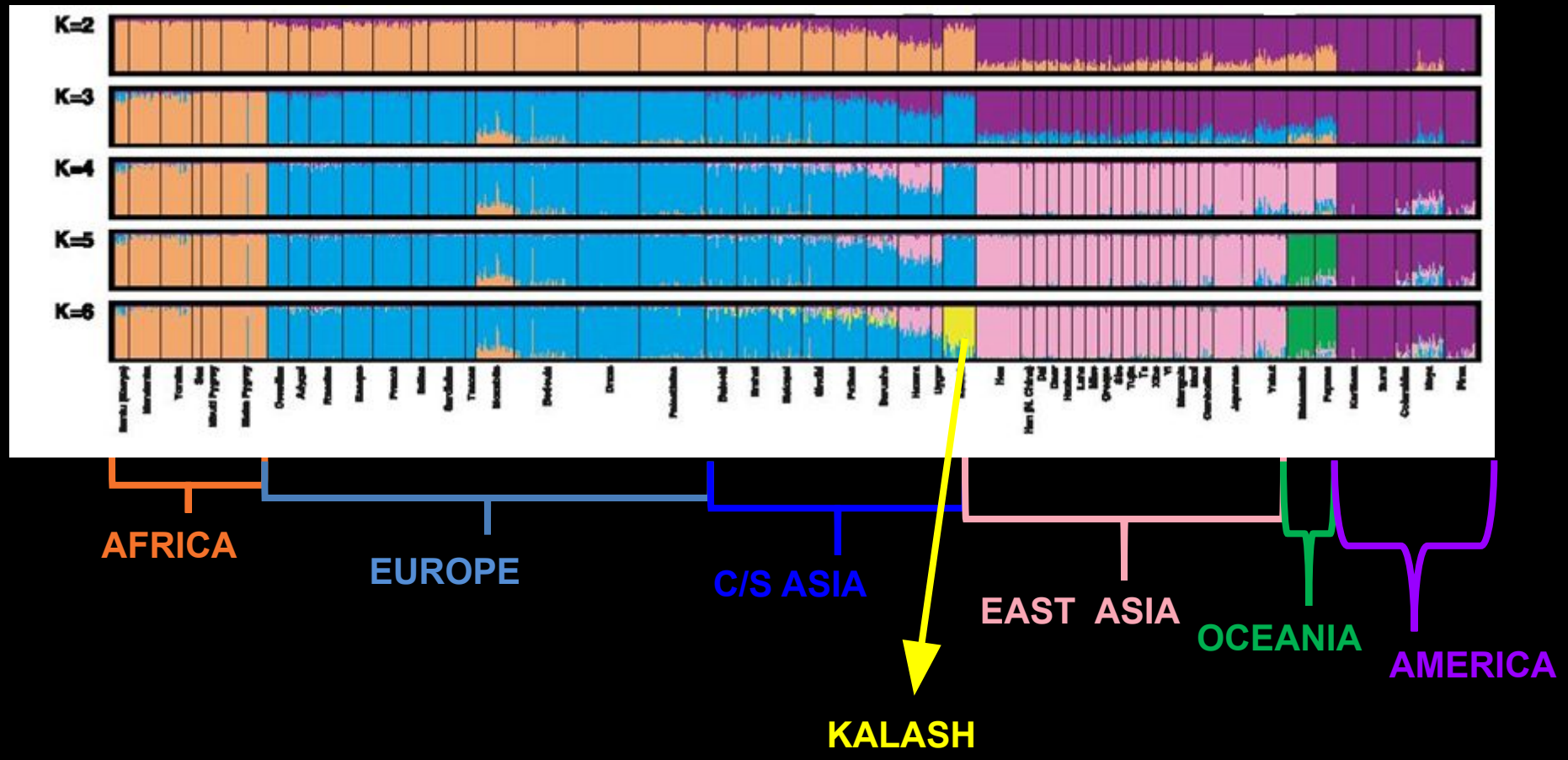
Number of statistically significant pairwise comparisons with a ratio between standard deviations >1 after Bonferroni correction.

For the measures based on pairwise comparisons (IBS and CHR_P), population variance was calculated using the individual median values.

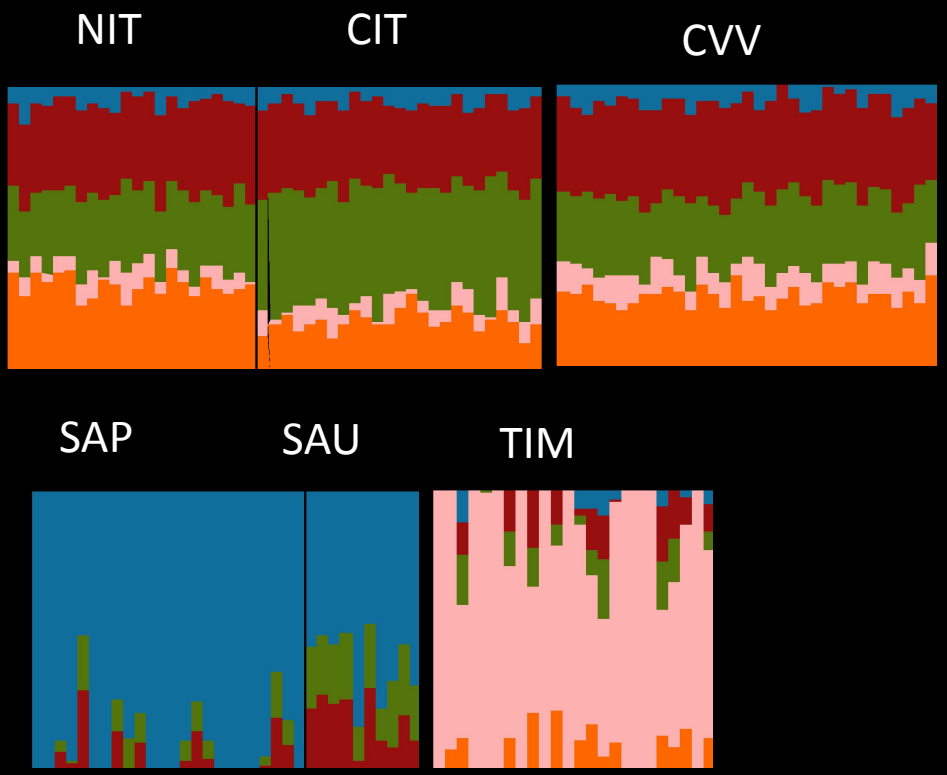


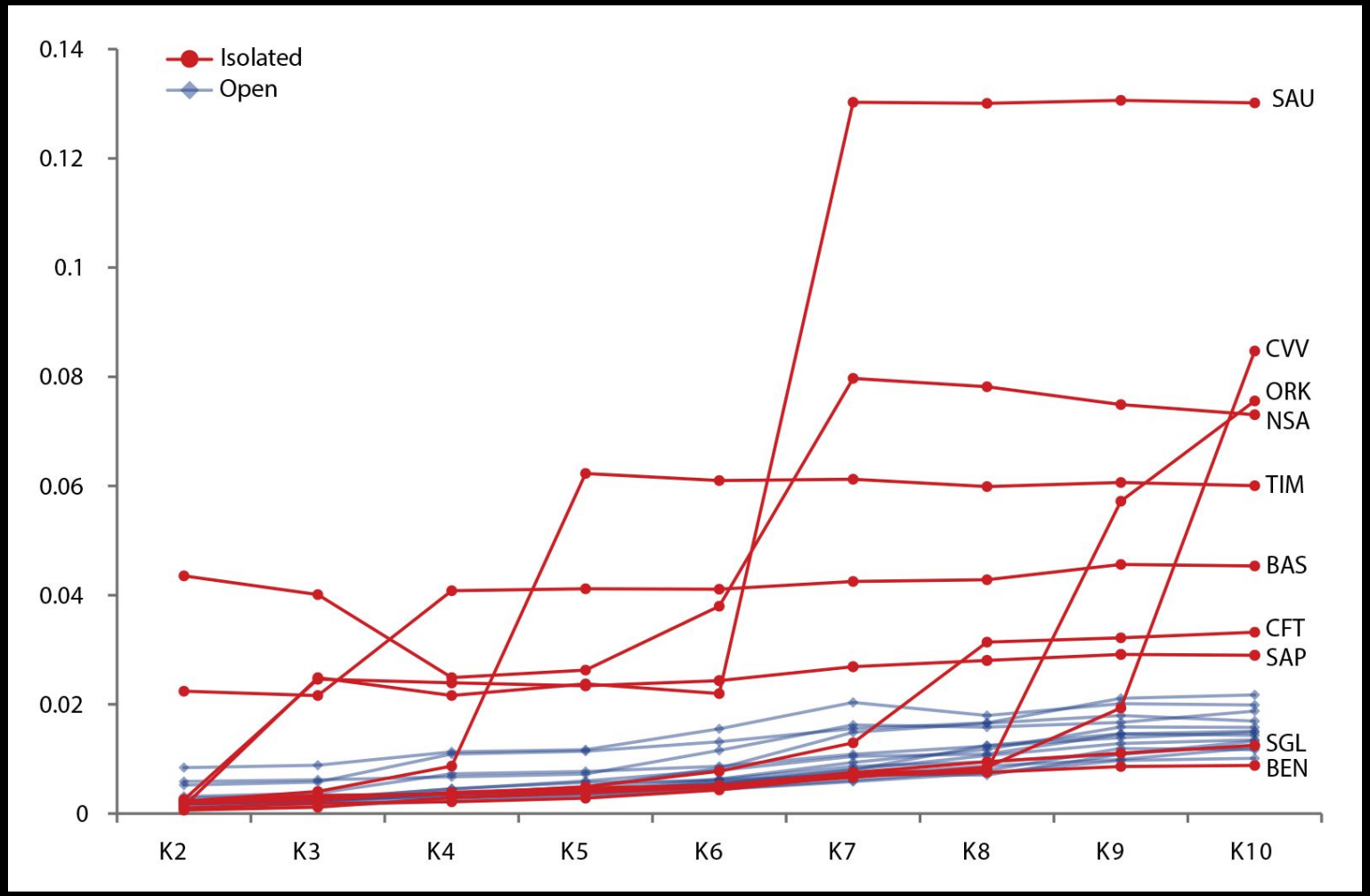
WHAT are genomic features: inter-pop analyses

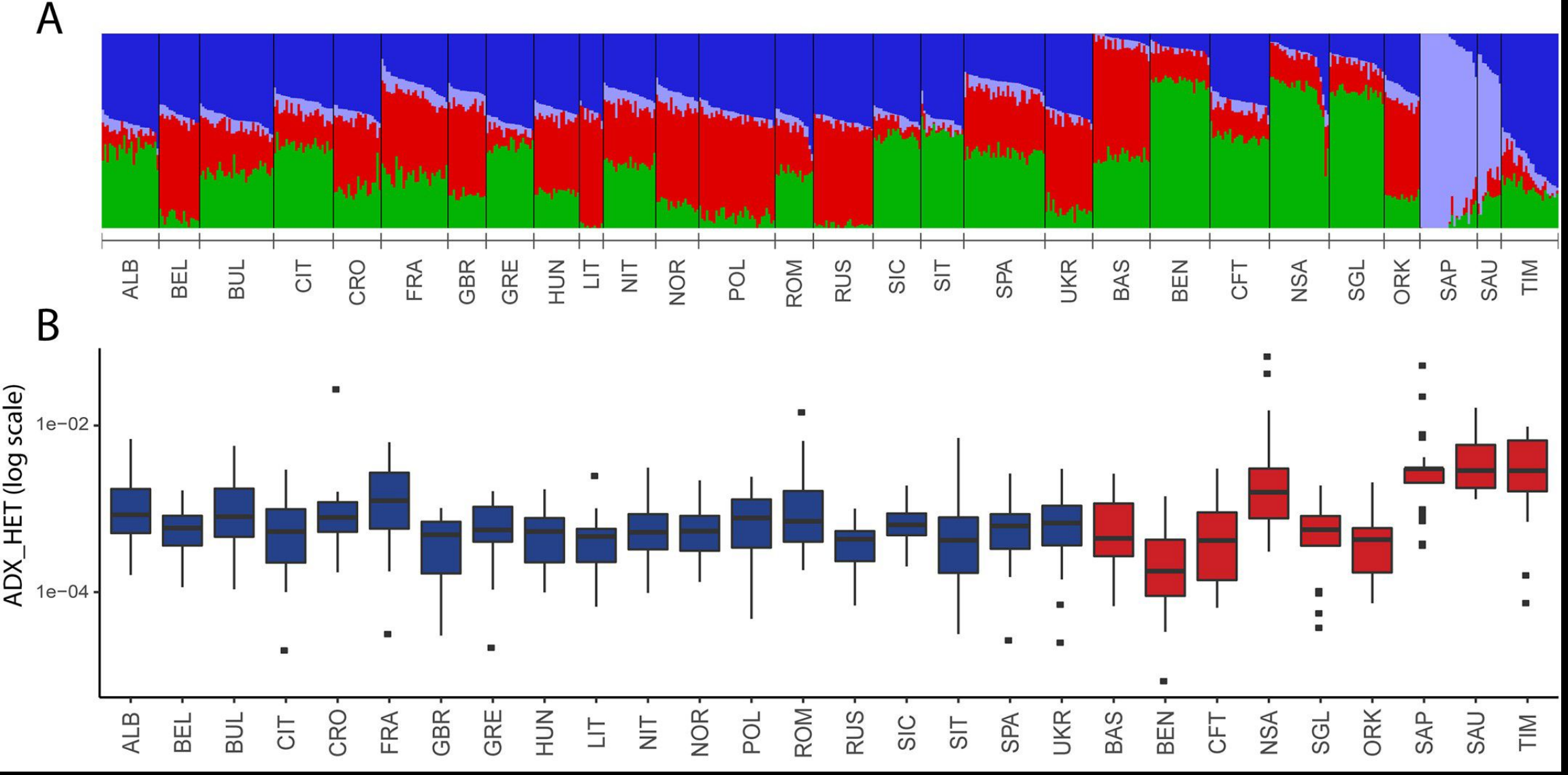
ADMIXTURE analysis



Hidden heterogeneity within population isolates







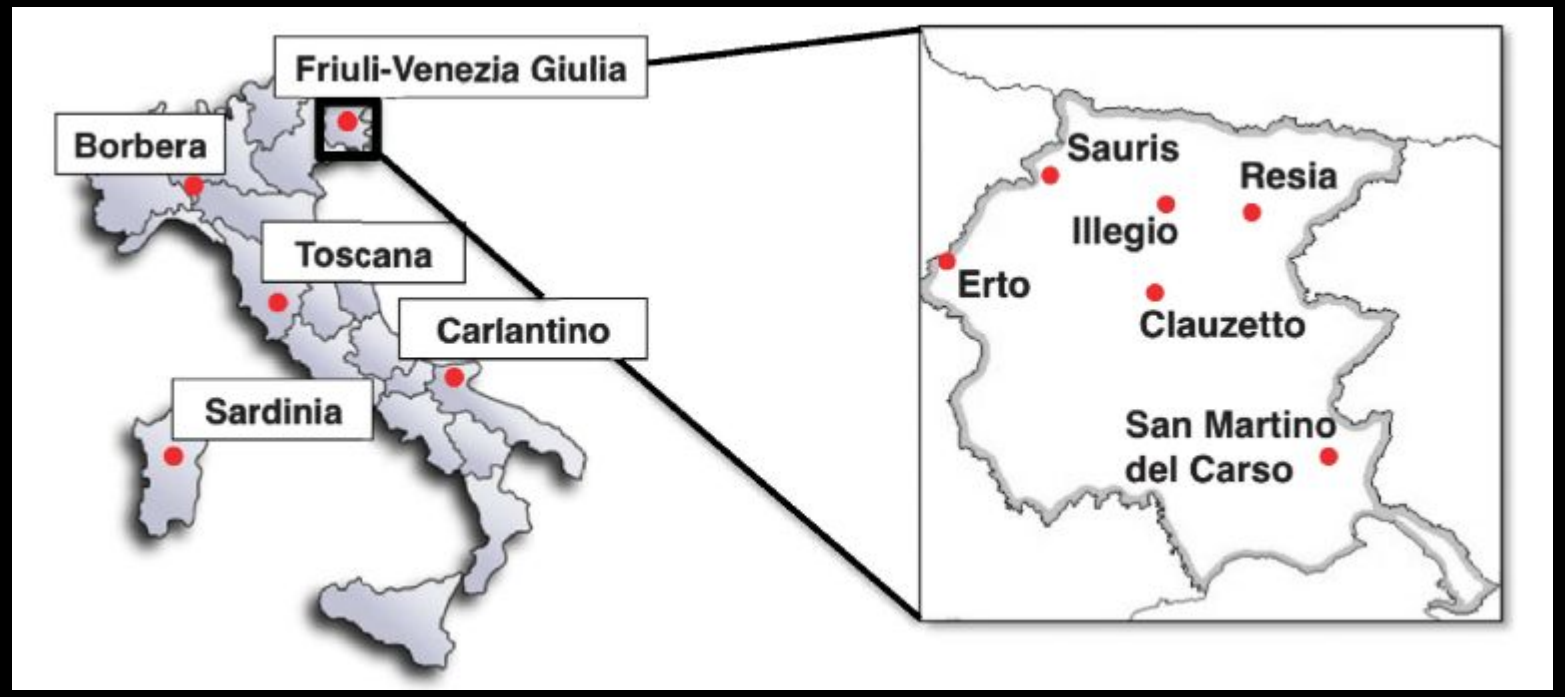
(A) Maximum likelihood estimates of individual ancestries (K = 4) for the 28 populations under study; **(**B) intra-population distribution of the admixture heterogeneity measure (y axis log scale);

Detecting isolation: autosomal markers

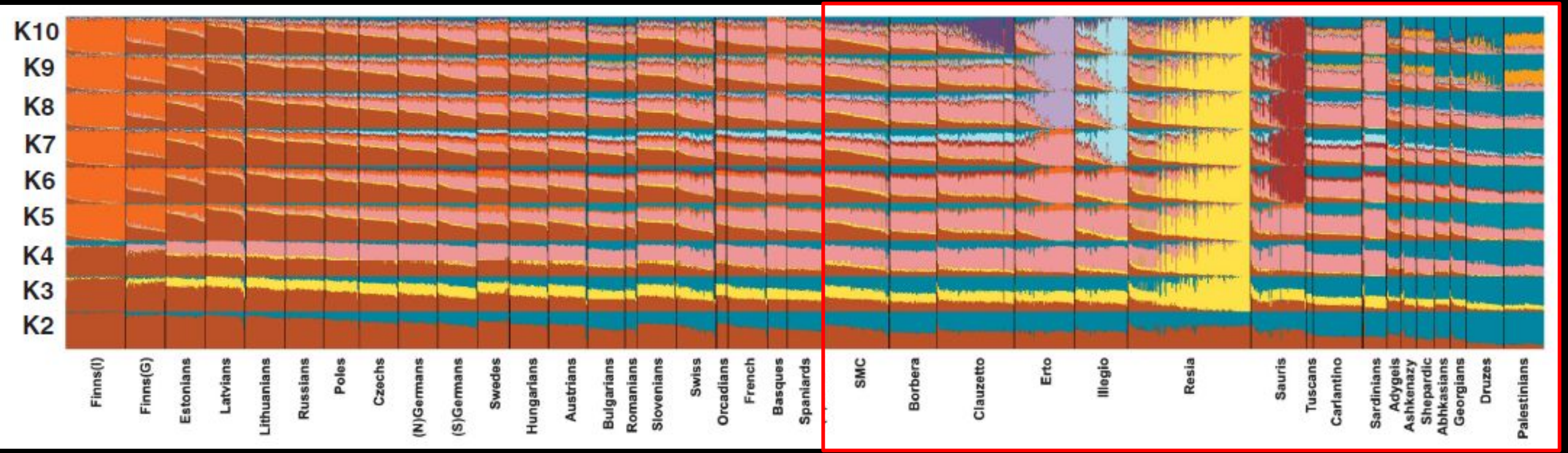
Genetic characterization of northeastern Italian Population isolates in the context of broader European genetic diversity. Esko et al. 2013

Aim:

- Relationship between villages and other Italian/European population
- Evaluate if any village represent a genuine population isolate



Detecting isolation: autosomal markers



Ancestry proportions of the studied 1008 individuals from 39 European and Near-Eastern populations (including the six FVG village populations) as revealed by the ADMIXTURE program with K=2 to K=10. A stacked column of the K proportions represents each individual, with fractions indicated on the y axis. From all non-FVG populations a subset of 24 randomly drawn individuals (if applicable) was used.

Detecting isolation: autosomal markers

RoHs analy

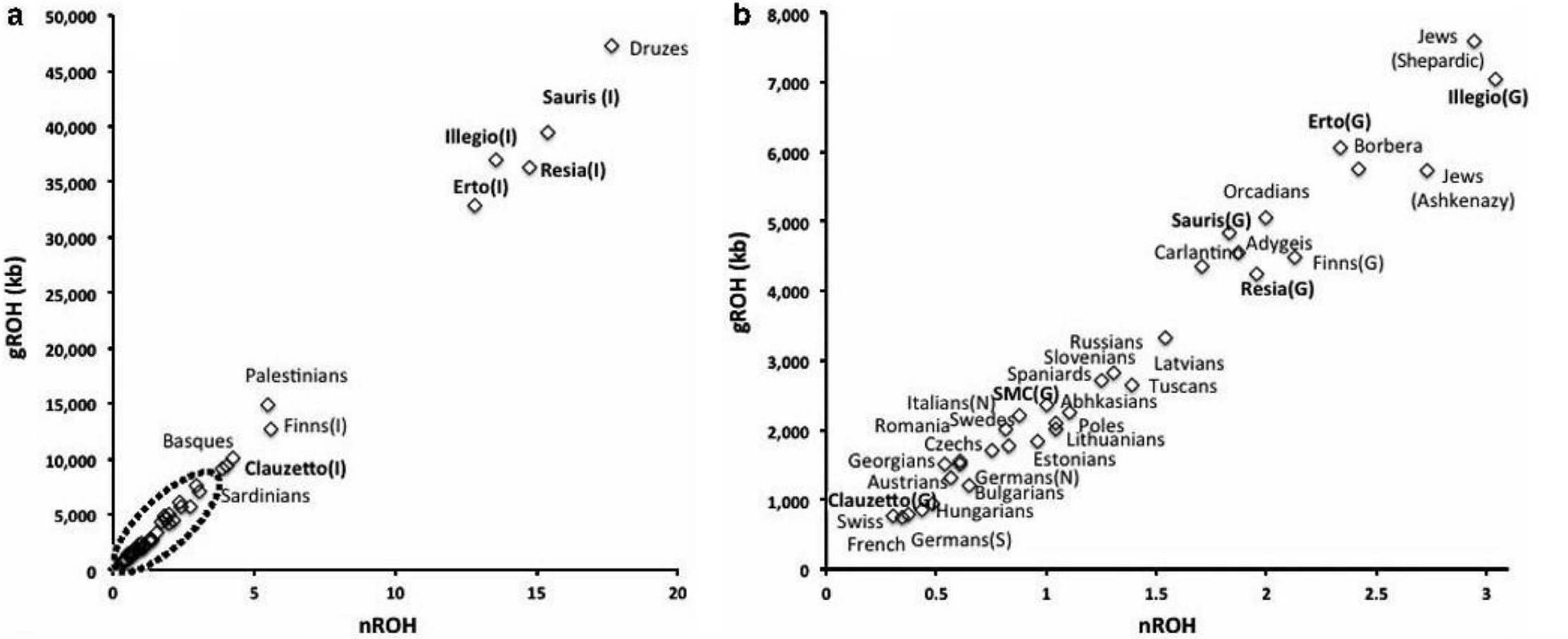


Figure 5 The genomic runs of homozygosity (gROH) based on 145 000 SNPs. (**a**) Distribution of population mean homozygous segments and mean count of gROH > 1.5 Mb per sample. (**b**) Zoom in view for the region indicated with a dashed ellipse on (**a**). FVG sub-populations are indicated in bold letters. In population names: I, a more homogeneous sub-population; G, a more general sub-population.

Detecting isolation: autosomal markers

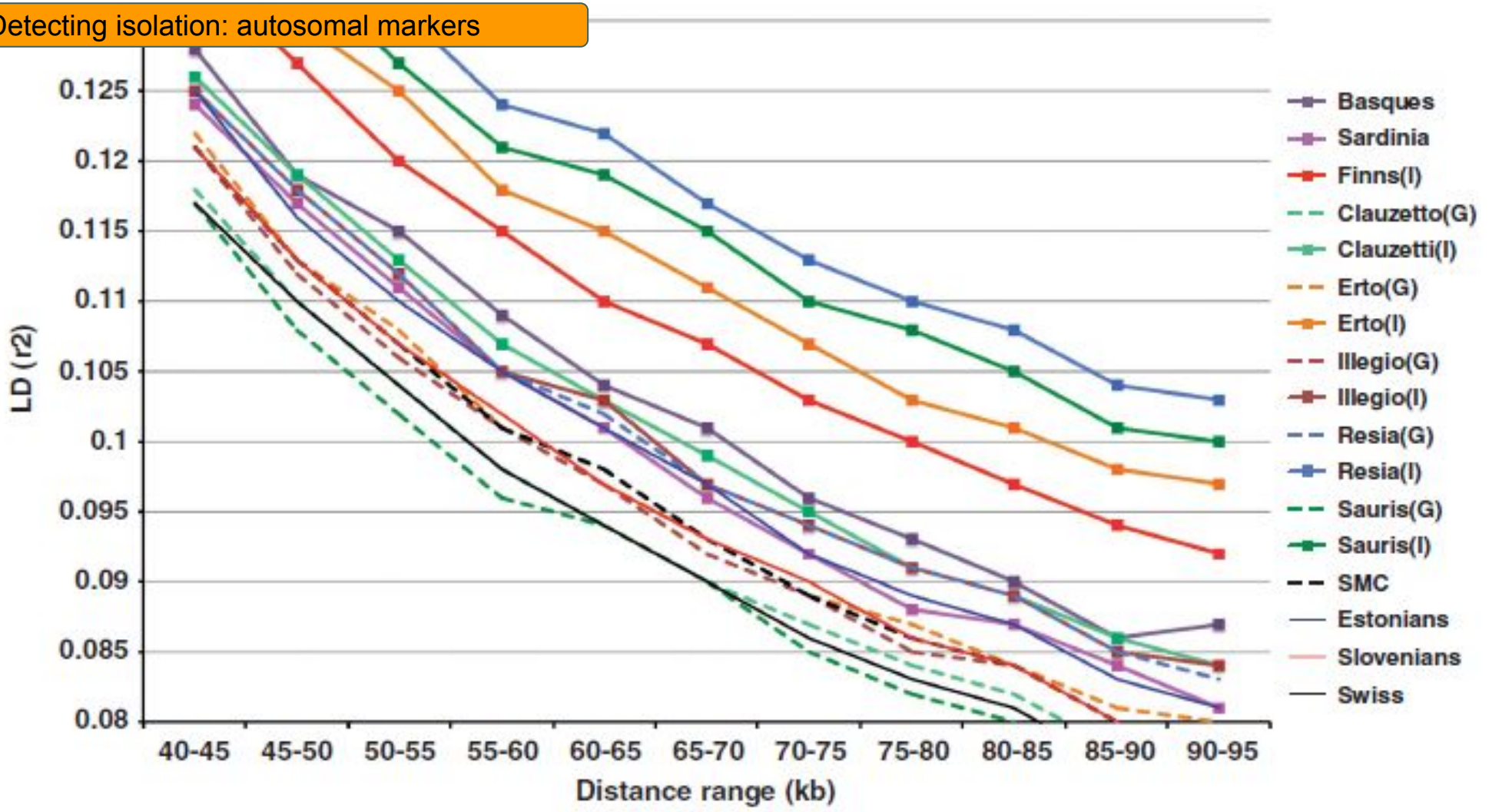
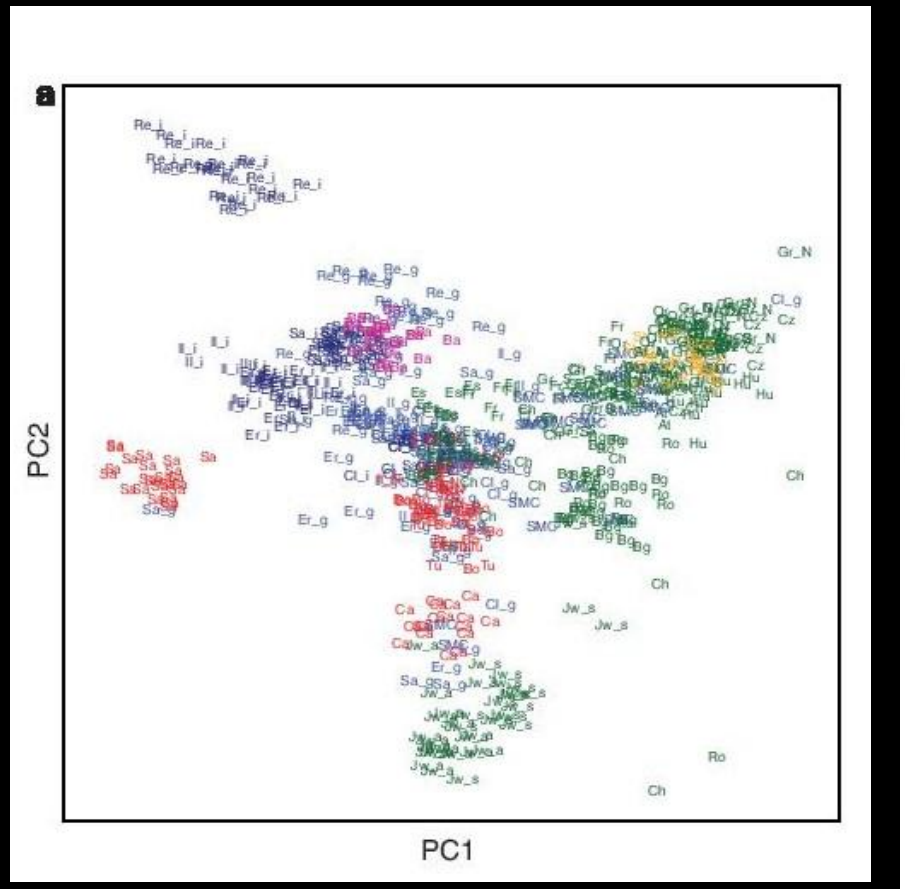
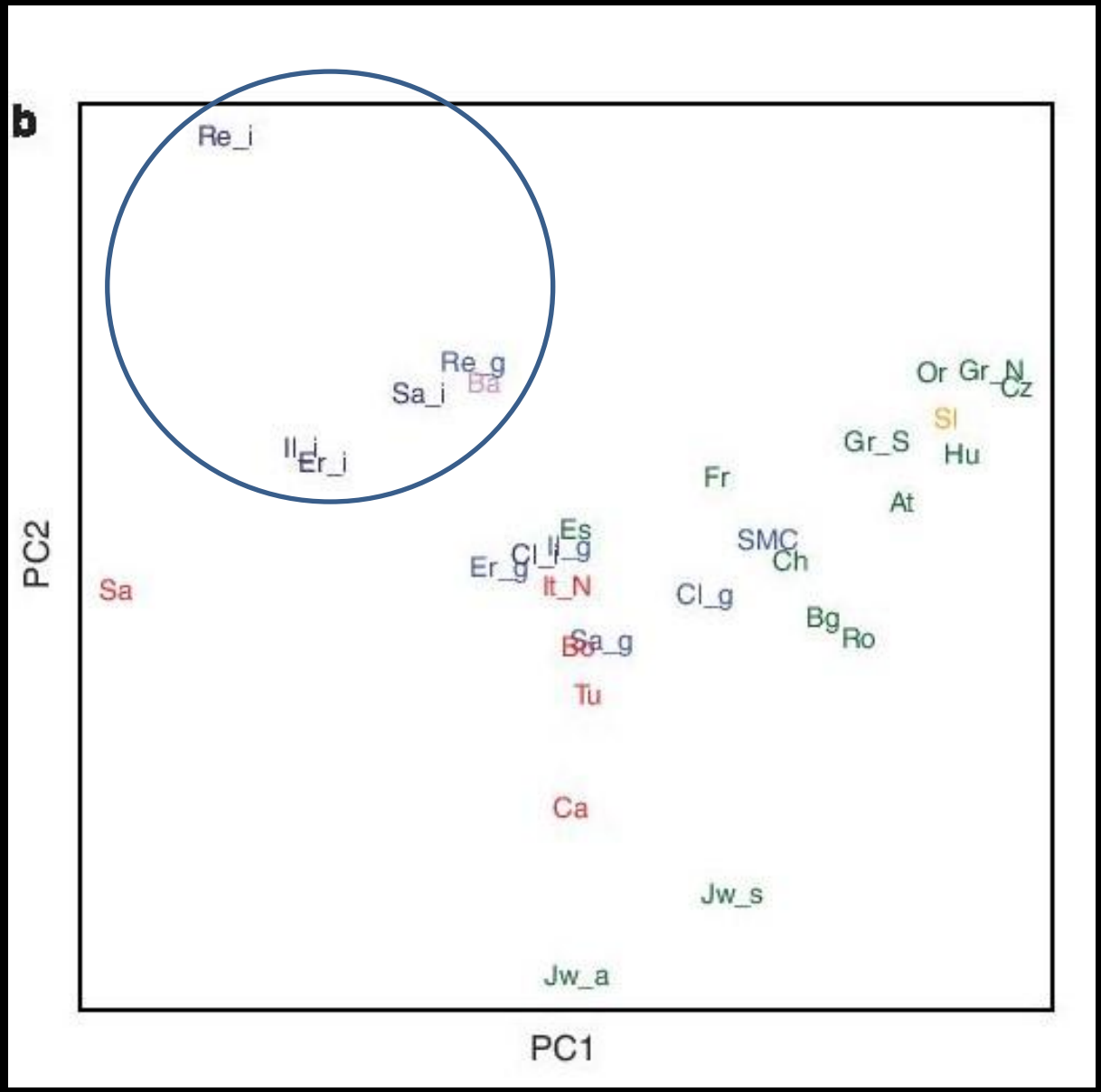


Figure 4 Genome-wide LD length based on 145 000 SNPs. Each line represents the LD decay averaged across populations and sub-populations and LD (r^2) between SNPs shown in 5 kb bins. Sub-populations drawn from same population have the same color coding. A dashed line represents the more general sub-populations and the solid line with rectangles represents the more homogeneous sub-populations respectively. A solid thin line represents the European reference populations (Estonians, Slovenians and Swiss). The LD extent in reference population isolates (Sardinians, French Basques and northern Finns) is also shown with a solid line with rectangles. In population names: I, a more homogeneous sub-population; G, a more general sub-population.

Detecting isolation: autosomal markers





Implications for association studies

"**what do our results imply for the way in which bio-medical studies are carried out in population isolates**?". Although, the most robust evidence was noticed in some young and small-sized population isolates—which are less used in association studies than the older and larger ones [39]—our results are worthy of attention since they highlight **a confounding factor** which has not been yet adequately taken into account. In fact, to the best of our knowledge, **the effect of increased allelic and haplotypic heterogeneity** has been investigated only in relation to the issue of undetected population structure in large scale association studies [40], whereas we argue that it may represent a drawback also for genetic investigations of population isolates.

We suggest that genetic clustering algorithms may be used to test for the presence of individuals with different ancestry proportions within isolated populations, similarly to what has been previously done by Esko et al. [36] (see also [41]). **Whenever genomes with substantially more heterogeneous ancestry are detected, it would be worth removing them, re-estimating the parameters of gene-disease association and comparing the new results with those obtained using the whole sample.** This could help **evaluate whether the genomes with mixed ancestry—in which the reduction of the haplotypic and allelic diversity produced by the effects of the founders and inbreeding should be less detectable—may have acted as confounding factors**. For each dataset, different ancestry proportions could be tried as thresholds, and the one able to reduce inter-individual heterogeneity without leading to a significant loss of power should be used.

Conclusions

In this study we have shed light on the occurrence of relatively high levels of inter-individual heterogeneity in population isolates and proposed a way to monitor their effects on the inferences of association between genes and diseases. This research work **challenges the traditional paradigm which considers population isolates as genetically uniform entities, providing further evidence that dichotomizing human populations into open and isolated groups fails to capture the actual relations among their genomic features**