Principali Risorse per lo studio della diversità genetica umana

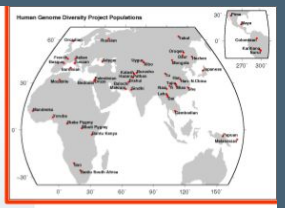
377 STRs
1056 individuals
HGDP

Genetic Structure of Human Populations

Noah A. Rosenberg,[1]* Jonathan K. Pritchard,[2] James L. Weber,[3] Howard M. Cann,[4] Kenneth K. Kidd,[5] Lev A. Zhivotovsky,[6] Marcus W. Feldman[7]

Methods

- HGDP (Human Genome Diversity Project) dataset
  - Populations from all continents
- Bedouin, Druze, Palestinians, Mozabite Berbers, Brahui, Baluchi, Makrani, Sindhi, and Pathan
- Analyzed data at the continental and population level



1043 individuals

- Continental groups
  - Europe
  - Sub-Saharan Africa
  - America (Native American)
  - Oceania
  - East Asia
  - Central/South Asia
  - West Asia

Caratteristica	**Human Genome Diversity Project (HGDP)**	**1000 Genomes Project**
Periodo di Avvio	Inizi anni '90 (raccolta dati dal 1994 circa)	2008
Obiettivo Primario	Studiare la storia delle migrazioni e le relazioni tra le popolazioni	Catalogare varianti genetiche comuni e rare (≥1% di frequenza) per facilitare gli studi sulle malattie.
Numero di Individui	Circa 1.050 individui non imparentati	Circa 2.504 individui non imparentati
Numero di Popolazioni	52 popolazioni distinte	26 popolazioni (in 5 "superpopolazioni" continentali)
Tipo di Popolazioni	popolazioni indigene e isolate con presunta bassa mescolanza genetica recente, per rappresentare la diversità globale in modo filogenetico.	popolazioni più ampie e clinicamente rilevanti, spesso con mescolanza genetica, per fornire una risorsa utile agli studi di associazione.
Campionamento Geografico	forte rappresentazione di regioni meno studiate (es. Medio Oriente, alcune Americhe, Oceania).	concentrato su Europa, Africa subsahariana, Asia Orientale, Asia Meridionale e Americhe
Tipo di Dato Primario	Inizialmente genotipizzazione (microarray), poi sequenziamento dell'esoma e del genoma intero ad alta copertura (circa 30x).	Inizialmente sequenziamento a bassa copertura, sequenziamento dell'esoma, poi esteso a sequenziamento del genoma intero ad alta copertura.
Formato dei Campioni	Principalmente linee cellulari immortalizzate	DNA da campioni di sangue/saliva.
Applicazioni Tipiche	Storia delle migrazioni, analisi delle relazioni tra gruppi etnici	Varianti genetiche associate a malattie complesse (GWAS), studi di associazione genetica
Relazione tra i Progetti	Complementari. **Non c'è una sovrapposizione significativa di campioni.** I dati di entrambi i progetti sono spesso combinati in analisi più recenti (es. nel gnomAD o HPRC) per creare risorse di riferimento ancora più complete.Sono stati un riferimento fondamentale l'uno per l'altro per lo sviluppo di metodologie e analisi, ma con scopi distinti.	

Iniziativa / Progetto	Obiettivo Principale	Popolazioni / Campioni Rappresentati	Tipo di Dati Generati	Applicazione Principale
International HapMap Project	Una mappa di varianti genetiche comuni (SNP e aplotipi) per studi di associazione.	Yoruba (Africa), Cinese Han, Giapponese, Europeo e poi altri gruppi.	Genotipizzazione (milioni di SNP)	Base per i primi GWAS, identificazione di aplotipi comuni.
Genome Aggregation Database (gnomAD)	Aggregare Centinaia di migliaia dati di sequenziamento da diverse iniziative	Genomi ed esomi da pop europee, asiatiche, africane, latinoamericane, etc.	Sequenziamento dell'esoma e del genoma intero (alta copertura).	Interpretazione della patogenicità, frequenza e genetica clinica.
Human Pangenome Reference Consortium (HPRC)	Genoma di riferimento più completo che rifletta la variazione globale.	Diversità etnica ampia, focus sulla sottorappresentazione storica.	Sequenziamento del genoma intero (alta qualità e copertura).	Ridurre il bias nelle analisi genomiche, migliorare medicina di precisione .
UK Biobank	risorsa su larga scala con dati genetici, sanitari e di stile di vita.	Oltre 500.000 individui principalmente di origine europea (UK).	Genotipizzazione, sequenziamento , dati clinici, immagini, questionari.	Studi di associazione gene-fenotipo, fattori di rischio genetici e ambientali.

Indagini Evoluzionistiche (Storia della Specie Umana)

HGDP (Human Genome Diversity Project)

1. **Le origini della nostra specie:** Da dove veniamo? (L'Africa!)
2. **Le grandi migrazioni:** Come si è diffusa l'umanità sul pianeta? (I modelli di migrazione fuori dall'Africa, il popolamento delle Americhe, ecc.).
3. **Le relazioni tra le popolazioni:** Chi è più "vicino" geneticamente a chi? Questo ci aiuta a capire la storia dei mescolamenti e delle separazioni tra gruppi umani.
4. **Adattamenti all'ambiente:** Come si è adattato il nostro DNA a climi diversi, diete particolari o agenti patogeni specifici? Ad esempio, la capacità di digerire il lattosio nell'età adulta, o la resistenza alla malaria in certe popolazioni africane.

Indagini Mediche (Associazione Geni-Malattie)

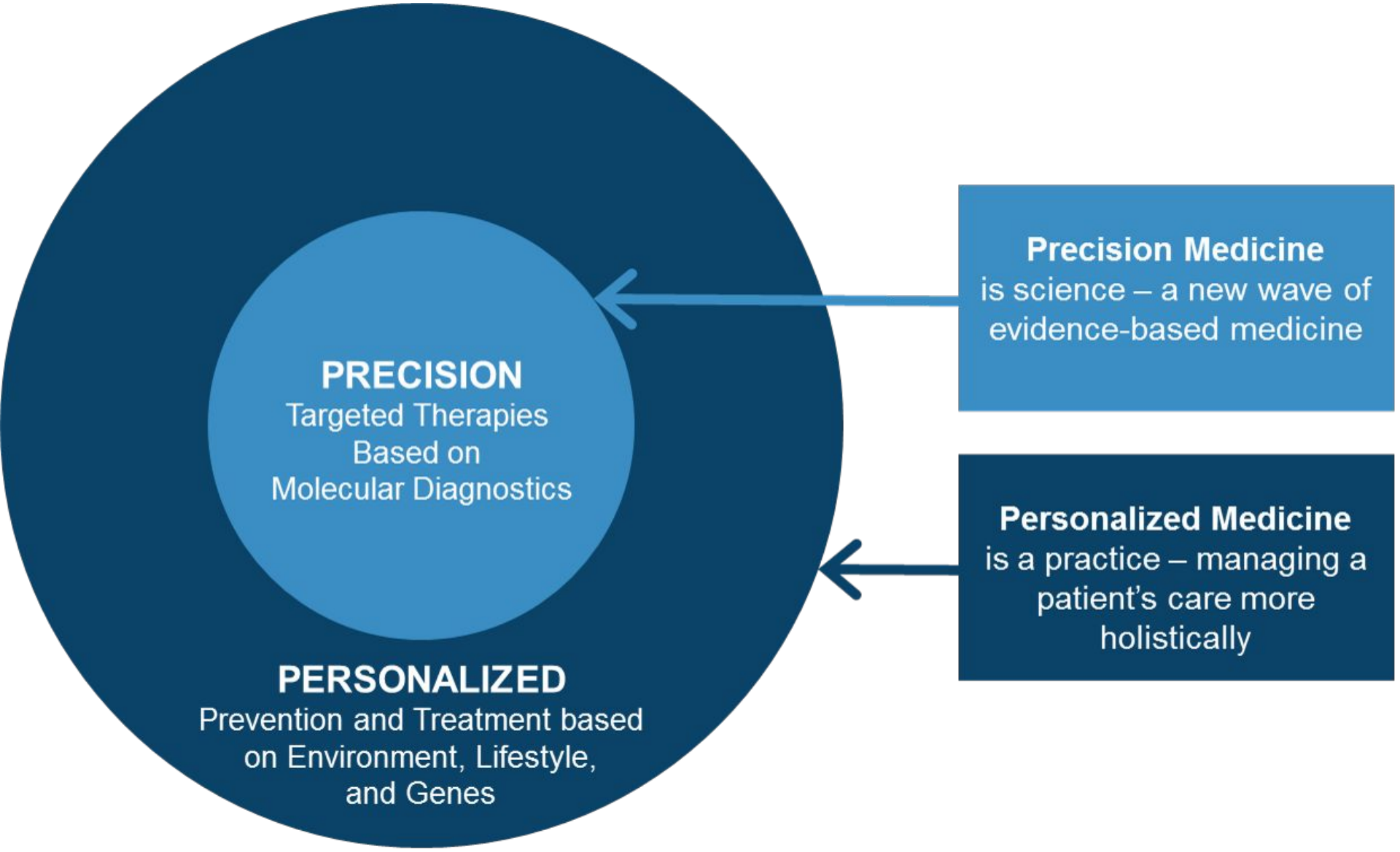
… medicina di precisione

Identificare i geni della malattia: Quali varianti genetiche sono associate a un rischio maggiore o minore di sviluppare una malattia come il diabete, l'Alzheimer o il cancro?

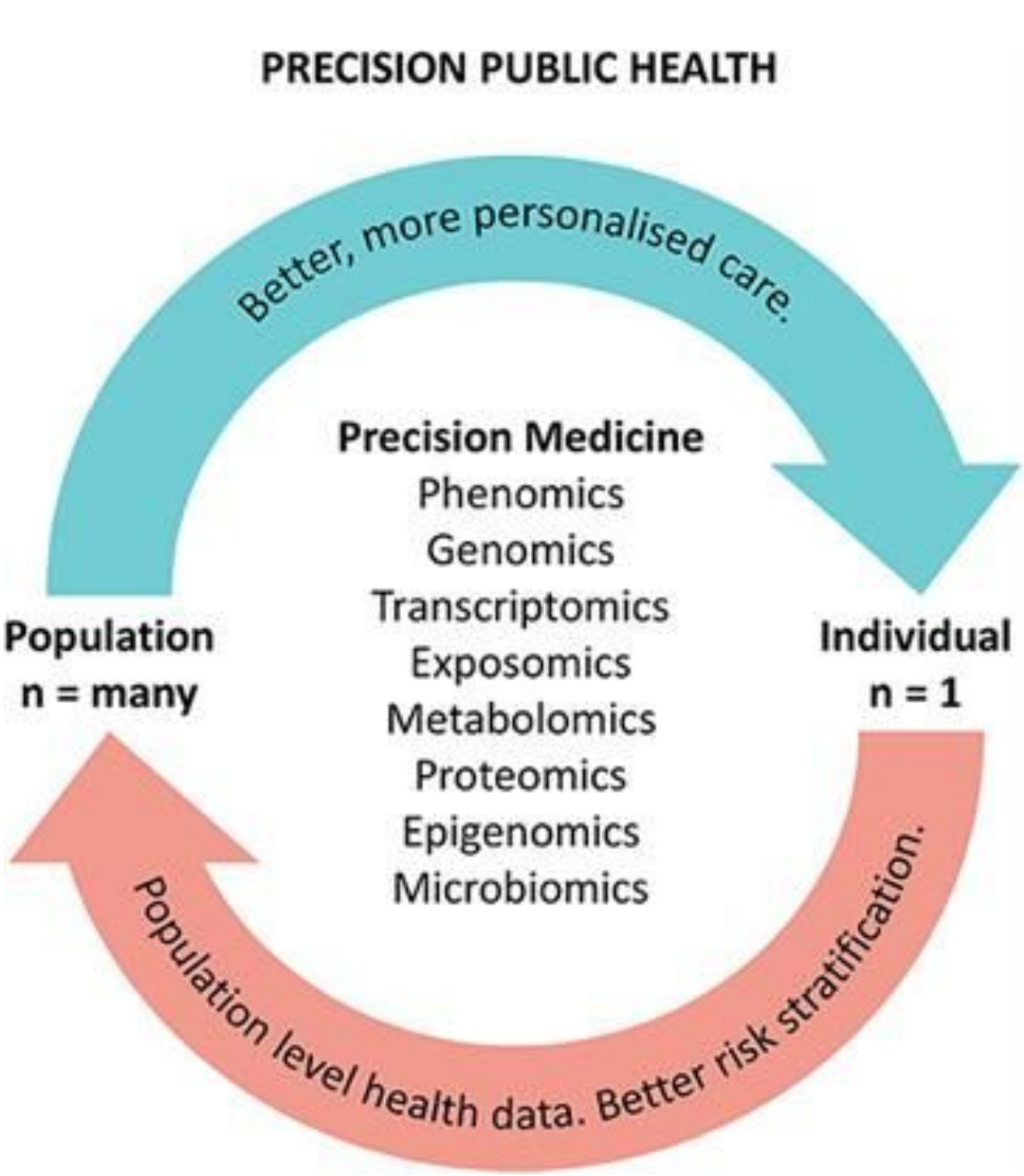
1. **Sviluppo di farmaci:** Una volta identificati i geni chiave, possiamo progettare farmaci che agiscano su quei meccanismi.
2. **Prevenzione e diagnosi precoce:** Se sappiamo che una persona ha una certa variante genetica, potremmo consigliarle stili di vita specifici o controlli più frequenti.
3. **Capire la variabilità della risposta ai farmaci:** Perché alcune persone rispondono bene a un farmaco e altre no? La genetica può dare risposte.

Il **1000 Genomes Project** e soprattutto il **gnomAD (Genome Aggregation Database)** hanno fornito un catalogo gigantesco di varianti genetiche e delle loro frequenze in diverse popolazioni, permettendo ai ricercatori di distinguere le varianti rare ma potenzialmente patogene da quelle comuni e innocue.

Iniziative come **UK Biobank** e **All of Us** vanno oltre, collegando questi dati genetici a cartelle cliniche e stili di vita, rendendo possibile identificare le associazioni tra geni e malattie su larga scala.


PRECISION
Targeted Therapies
Based on
Molecular Diagnostics
PERSONALIZED
Prevention and Treatment based
on Environment, Lifestyle,
and Genes
Precision Medicine
is science – a new wave of
evidence-based medicine
Personalized Medicine
is a practice – managing a
patient's care more
holistically

Precision medicine is "an emerging approach for disease treatment and prevention that takes into account **individual variability in genes, environment, and lifestyle for each person**." The new GHR primer stresses that, although the term *precision medicine* is relatively new, the concept has existed in different areas of medical practice for a long time.



Indagini Mediche (Associazione Geni-Malattie)

… **medicina di precisione**

Identificare i geni della malattia: Quali varianti genetiche sono associate a un rischio maggiore o minore di sviluppare una malattia come il diabete, l'Alzheimer o il cancro?

1. **Sviluppo di farmaci:** Una volta identificati i geni chiave, possiamo progettare farmaci che agiscano su quei meccanismi.
2. **Prevenzione e diagnosi precoce:** Se sappiamo che una persona ha una certa variante genetica, potremmo consigliarle stili di vita specifici o controlli più frequenti.
3. **Capire la variabilità della risposta ai farmaci:** Perché alcune persone rispondono bene a un farmaco e altre no? La genetica può dare risposte.

Il **1000 Genomes Project** e soprattutto il **gnomAD (Genome Aggregation Database)** hanno fornito un catalogo gigantesco di varianti genetiche e delle loro frequenze in diverse popolazioni, permettendo ai ricercatori di distinguere le varianti rare ma potenzialmente patogene da quelle comuni e innocue.

Iniziative come **UK Biobank** e **All of Us** vanno oltre, collegando questi dati genetici a cartelle cliniche e stili di vita, rendendo possibile identificare le associazioni tra geni e malattie su larga scala.

… medicina di precisione

- **Oncologia:** È il campo in cui la medicina di precisione ha fatto i passi più significativi ; grazie al sequenziamento genomico, è possibile identificare le mutazioni specifiche che guidano la crescita di un tumore e utilizzare terapie mirate che agiscono su quei bersagli molecolari basate su approcci come l'immunoterapia e le CAR-T cells stanno rivoluzionando il trattamento di diverse forme tumorali.
- **Malattie rare:** La medicina di precisione offre grandi speranze per le malattie rare, spesso causate da singole mutazioni genetiche. L'avanzamento delle tecnologie di editing genomico come CRISPR sta aprendo nuove frontiere per la cura di queste patologie, permettendo trattamenti più specifici e potenzialmente risolutivi.
- **Malattie cardiovascolari, il diabete e le malattie neurologiche:** l'obiettivo è identificare marcatori precoci per la prevenzione e terapie più efficaci.
- **"Big Data" e Intelligenza Artificiale:** I "Big Data" (genomici, clinici, di stile di vita) e l'AI stanno diventando strumenti essenziali per l'identificazione di pattern, la previsione di risposte ai trattamenti e la scoperta di nuove connessioni tra genetica e malattia.

Le risorse per la diversità genetica umana sono come un grande ponte che collega il nostro passato evolutivo con il nostro futuro sanitario. Non si può comprendere appieno l'uno senza l'altro, e le due prospettive si rafforzano continuamente a vicenda.

La genetica evoluzionistica illumina la medicina

- **Adattamenti e Malattie:** Se una variante genetica si è diffusa in una popolazione perché offriva un vantaggio (ad esempio, una maggiore resistenza a un'infezione), questa stessa variante potrebbe oggi essere associata a una malattia in un ambiente cambiato. Comprendere la sua storia evolutiva ci aiuta a capire perché è così comune in quella popolazione e quali potrebbero essere le sue implicazioni per la salute.

  Esempio: Ipercoagulazione da Neanderthal a Sapiens, CREB3, Anemia falciforme

- **Frequenza delle varianti:** I dati evolutivi ci spiegano perché certe varianti genetiche sono più comuni in alcune popolazioni e rare in altre. Questo è fondamentale per la medicina, perché sapere la "frequenza di base" di una variante in una popolazione aiuta i medici a decidere se una variante rara è probabilmente la causa di una malattia in un paziente.

  Esempio: Varianti di Suscettibilità al Cancro al Seno (BRCA1/BRCA2) e gnomAD (Genome Aggregation Database)

- **Origini della malattia:** Capire da dove provengono le mutazioni e come si sono diffuse (o non diffuse) attraverso le popolazioni può fornire indizi preziosi sulle origini di alcune malattie genetiche.

  Esempio: Fibrosi Cistica e la Mutazione DeltaF508

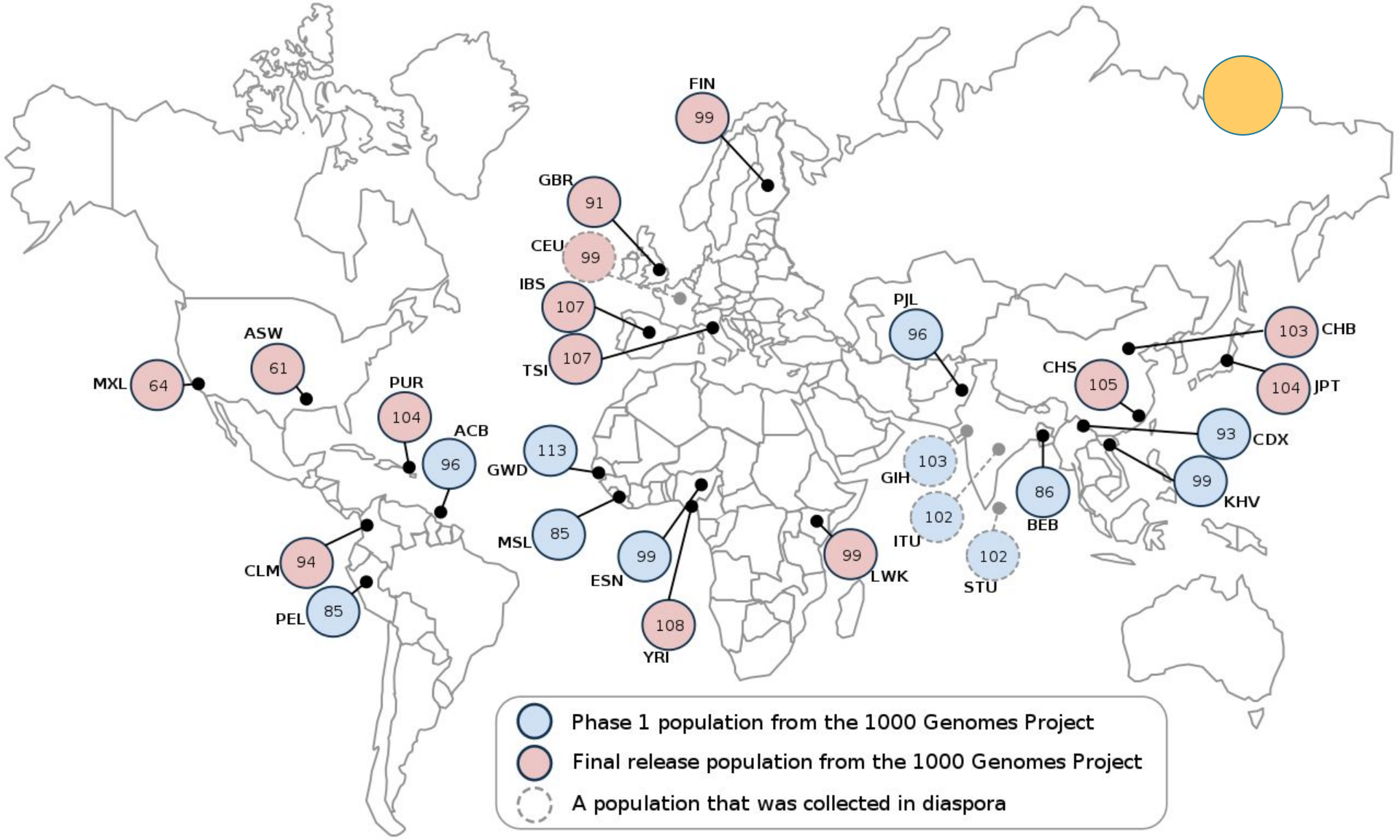
La genetica medica arricchisce la ricerca evoluzionistica

- **Dati su larga scala:** I progetti medici come UK Biobank o All of Us generano una quantità enorme di dati genomici da milioni di persone. Anche se il loro scopo primario è medico, questi dati sono una miniera d'oro per gli studi evoluzionistici. Possono essere usati per analizzare la diversità genetica a un livello di dettaglio senza precedenti all'interno di popolazioni molto ampie, rivelando modelli di selezione naturale recenti o eventi demografici che altrimenti sarebbero invisibili.
- **Identificazione di varianti funzionali:** Quando la ricerca medica identifica una variante genetica che ha un forte effetto su una malattia o un tratto (es. colesterolo alto, altezza), i genetisti evoluzionisti possono poi studiare la storia di quella variante: è stata selezionata naturalmente? È comparsa di recente? Questo aggiunge un pezzo al puzzle della nostra evoluzione.

L'iniziativa **H3Africa** **(Human Heredity and Health in Africa)** ha lo scopo medico di affrontare le malattie prevalenti in Africa, ma per farlo deve studiare la diversità genetica africana, che è la più ricca del mondo. Questo studio non solo aiuterà a sviluppare terapie e diagnosi più efficaci per le popolazioni africane (scopo medico), ma arricchirà enormemente la nostra comprensione dell'evoluzione umana e della diversità genetica globale (scopo evoluzionistico).

The 1000 Genomes Project Data

- An international research effort to establish by far the most detailed catalogue of human genetic variation.
- Scientists planned to sequence the genomes of at least one thousand anonymous participants from a number of different ethnic groups
- In 2012, the sequencing of 1092 genomes was announced in a Nature publication.
- In 2015, two papers in Nature reported results and the completion of the project and opportunities for future research


FIN
99
GBR
91
CEU
99
IBS
107
TSI
107
ASW
61
MXL
64
PUR
104
ACB
96
GWD
113
MSL
85
CLM
94
PEL
85
ESN
99
YRI
108
LWK
99
PJL
96
GIH
103
ITU
102
STU
102
BEB
86
CHS
105
CHB
103
JPT
104
CDX
93
KHV
99
Phase 1 population from the 1000 Genomes Project
Final release population from the 1000 Genomes Project
A population that was collected in diaspora

Goals

- Find > 95% of the SNPs with a frequency at least 1% (towards 0.1% in coding regions).
- Find short indels and larger structural variants.
- Provide genotypes, haplotypes, by sample.
- Provide cell lines for the samples.
- Release the data publicly and quickly.

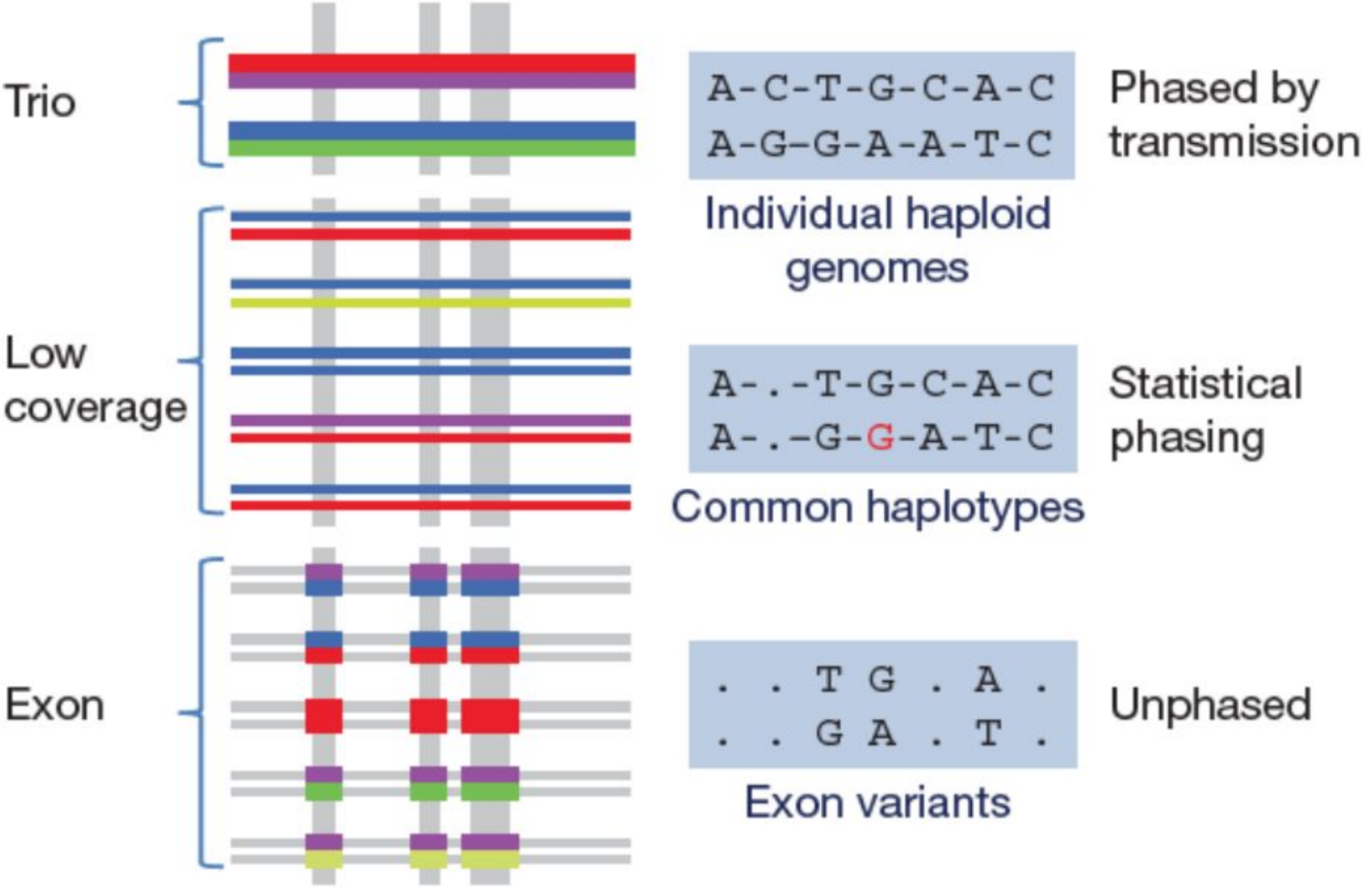
Plans for the Full-Scale Project

Sequence data

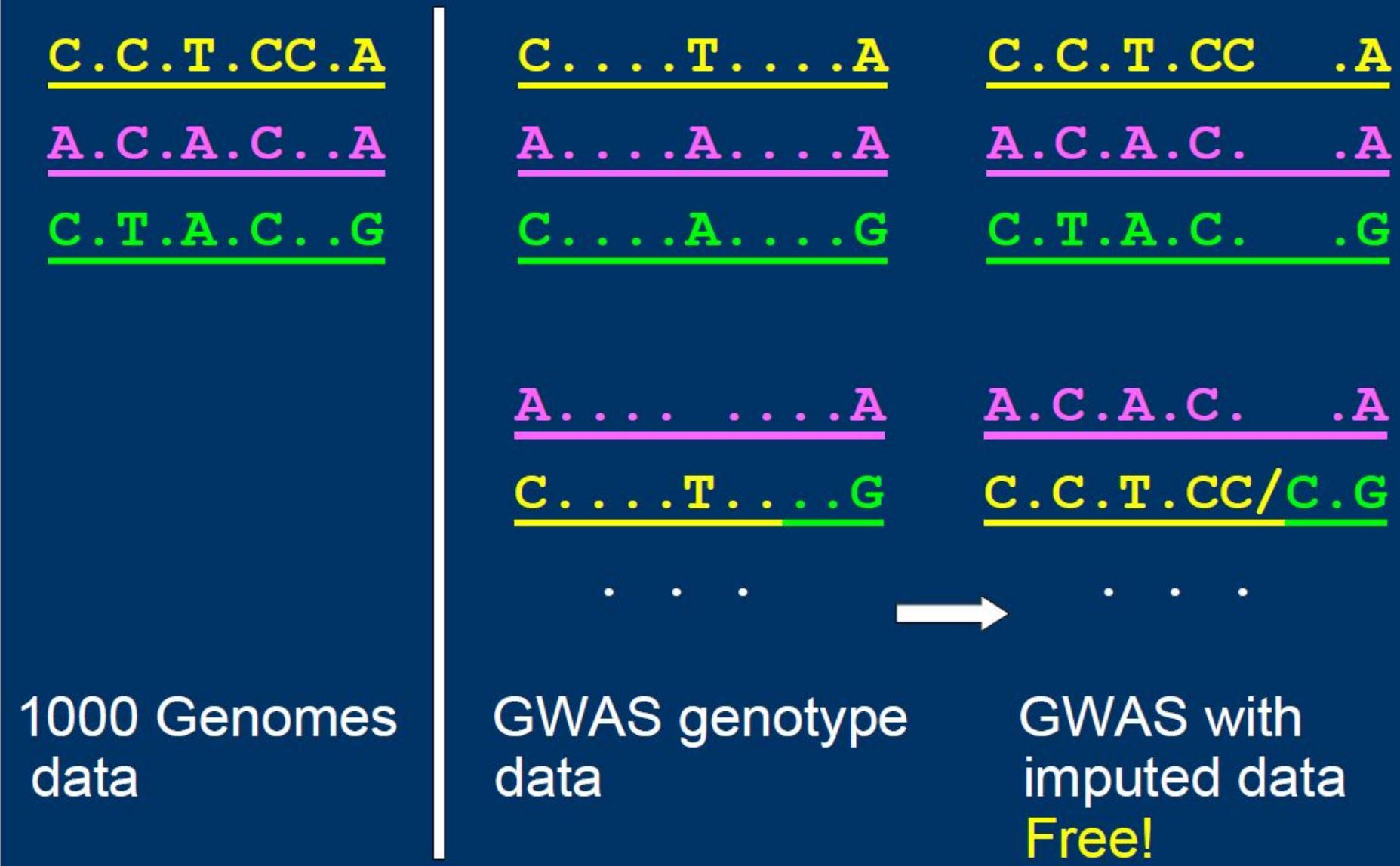
- Interim release with >25M SNPs on >600 samples today.
- 4-6X in 1100 samples (available) by Nov 2010
  - release 1st Q 2011.
- 4-6X in 600 samples (collecting) in 2010-11.
- 4-6X in 800 samples (plan collect) in 2011-12.
- Exome sequence for all of the samples.

Genotype data

- Up to 10 million SNPs with new Illumina arrays based on project data.
- Several types of structural variants in sets of the samples.


Trio
A-C-T-G-C-A-C
A-G-G-A-A-T-C
Phased by transmission
Individual haploid genomes
Low coverage
A-.-T-G-C-A-C
A-.-G-G-A-T-C
Statistical phasing
Common haplotypes
Exon
. . T G . A .
. . G A . T .
Unphased
Exon variants

Imputing Untyped Variants



Livelli di polimorfismo nell'uomo: genomi completi

Khoisan

D. Tutu

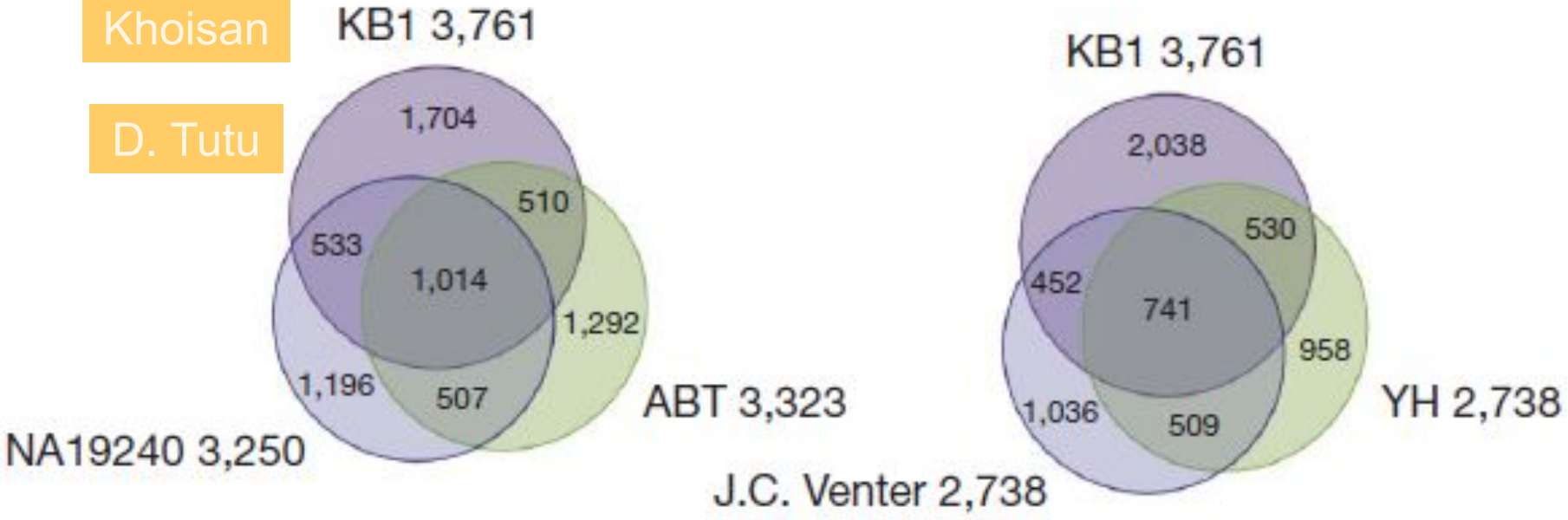


Figure 2 | Three-way relationships among SNPs. SNPs from KB1 are compared with those of the Yoruban NA19240 and ABT (left panel), and with an American of European descent (J. C. Venter) and a Chinese individual (YH) (right panel). Numbers are given in thousands. Variant positions that

ARTICLE

OPEN

doi:10.1038/nature15393

https://www.nature.com/articles/nature15393

A global reference for human genetic variation

2015

The 1000 Genomes Project Consortium*

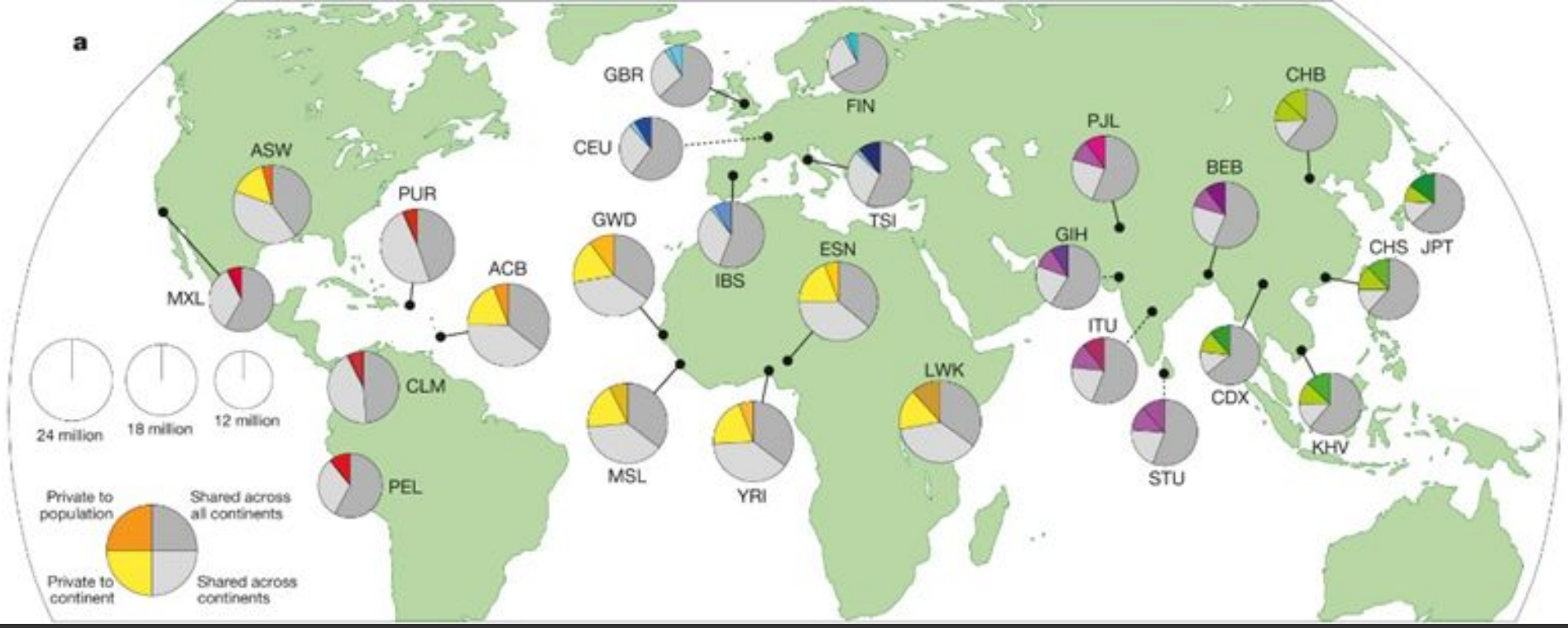
The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. We characterized a broad spectrum of genetic variation, in total over 88 million variants (84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all phased onto high-quality haplotypes. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

All individuals were sequenced using both whole-genome sequencing (mean depth = 7.4×) and targeted exome sequencing (mean depth = 65.7×). In addition, individuals and available first-degree relatives (generally, adult offspring) were genotyped using high-density SNP microarrays.

In contrast to earlier phases of the project, we expanded analysis beyond bi-allelic events to include multi-allelic SNPs, indels, and a diverse set of structural variants (SVs).

A typical genome

1. We find that a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites.
2. Although >99.9% of variants consist of SNPs and short indels, structural variants affect more bases: the typical genome contains an estimated 2,100 to 2,500 structural variants (~1,000 large deletions, ~160 copy-number variants, ~915 Alu insertions, ~128 L1 insertions, ~51 SVA insertions, ~4 NUMTs, and ~10 inversions), affecting ~20 million bases of sequence.
3. The total number of observed non-reference sites differs greatly among populations . Individuals from African ancestry populations harbour the greatest numbers of variant sites, as predicted by the out-of-Africa model of human origins. Individuals from recently admixed populations show great variability in the number of variants, roughly proportional to the degree of recent African ancestry in their genomes.

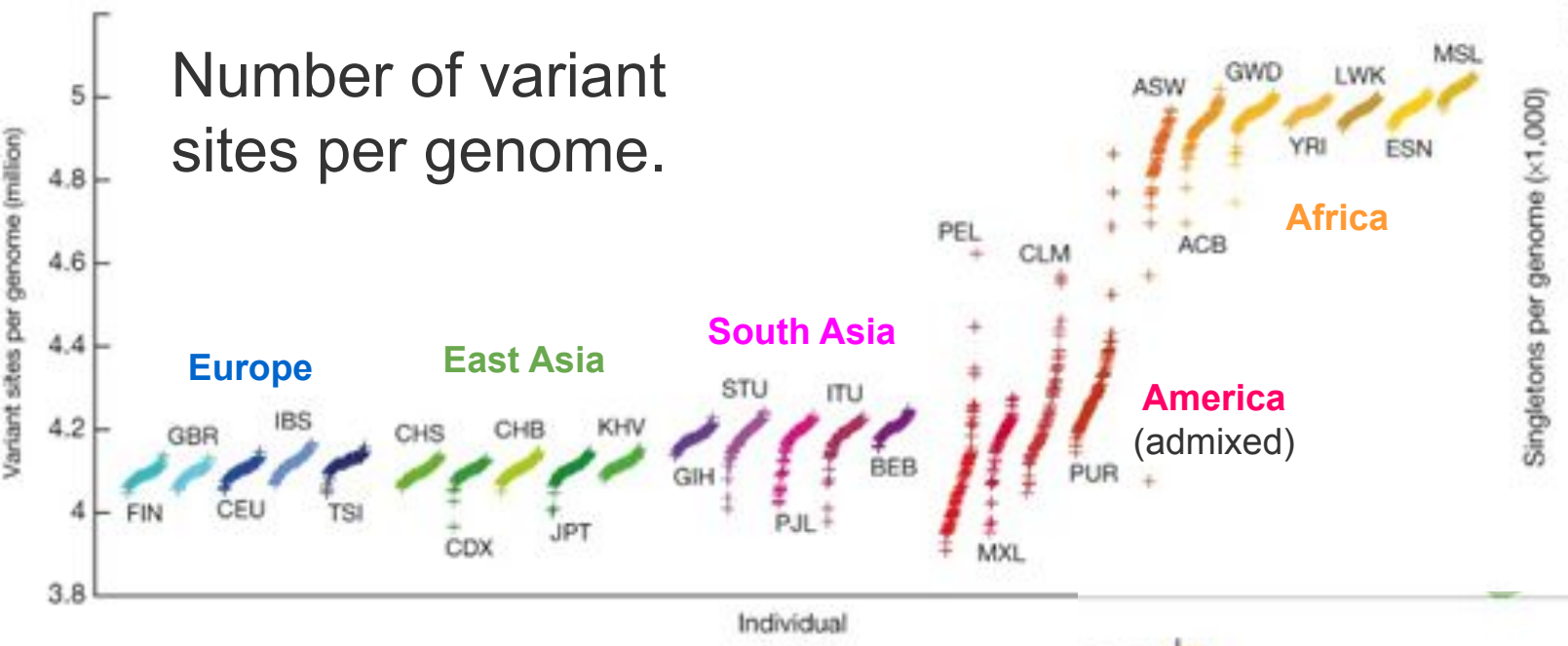


African ancestry **Americas** **East Asian an.** **European an.** **South Asian an.**

Polymorphic variants within sampled populations. The area of each pie is proportional to the number of polymorphisms within a population. Pies are divided into four slices, representing variants private to a population (darker colour unique to population), private to a continental area (lighter colour shared across continental group), shared across continental areas (light grey), and shared across all continents (dark grey). Dashed lines indicate populations sampled outside of their ancestral continental region.

Population		Code	Population Color	Continental Group Color
African ancestry				
Esan in Nigeria	Esan	ESN		
Gambian in Western Division, Mandinka	Gambian	GWD		
Luhya in Webuye, Kenya	Luhya	LWK		
Mende in Sierra Leone	Mende	MSL		
Yoruba in Ibadan, Nigeria	Yoruba	YRI		
African Caribbean in Barbados	Barbadian	ACB		
People with African Ancestry in Southwest USA	African-American SW	ASW		
Americas				
Colombians in Medellin, Colombia	Colombian	CLM		
People with Mexican Ancestry in Los Angeles, CA, USA	Mexican-American	MXL		
Peruvians in Lima, Peru	Peruvian	PEL		
Puerto Ricans in Puerto Rico	Puerto Rican	PUR		
East Asian ancestry				
Chinese Dai in Xishuangbanna, China	Dai Chinese	CDX		
Han Chinese in Beijing, China	Han Chinese	CHB		
Southern Han Chinese	Southern Han Chinese	CHS		
Japanese in Tokyo, Japan	Japanese	JPT		
Kinh in Ho Chi Minh City, Vietnam	Kinh Vietnamese	KHV		
European ancestry				
Utah residents (CEPH) with Northern and Western European ancestry	CEPH	CEU		
British in England and Scotland	British	GBR		
Finnish in Finland	Finnish	FIN		
Iberian Populations in Spain	Spanish	IBS		
Toscani in Italia	Tuscan	TSI		
South Asian ancestry				
Bengali in Bangladesh	Bengali	BEB		
Gujarati Indians in Houston, TX, USA	Gujarati	GIH		
Indian Telugu in the UK	Telugu	ITU		
Punjabi in Lahore, Pakistan	Punjabi	PJL		
Sri Lankan Tamil in the UK	Tamil	STU		
Total				

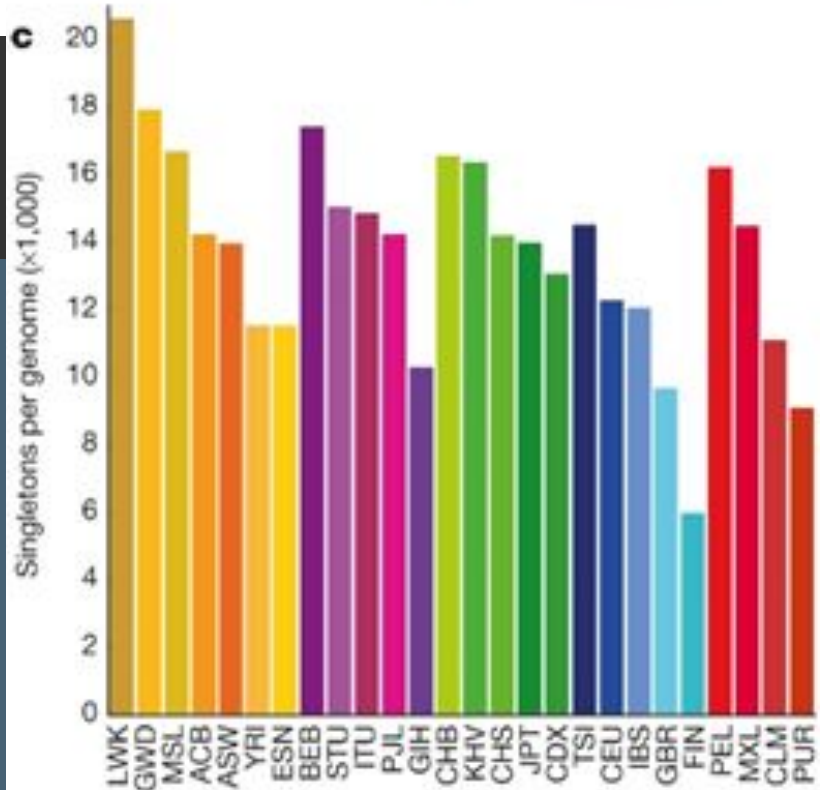
Number of variant sites per genome.

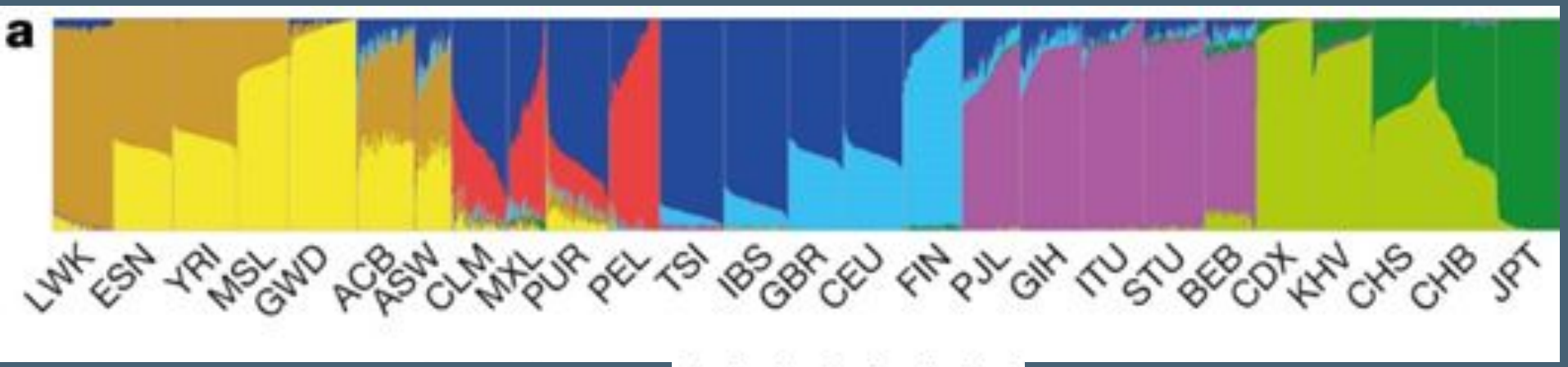


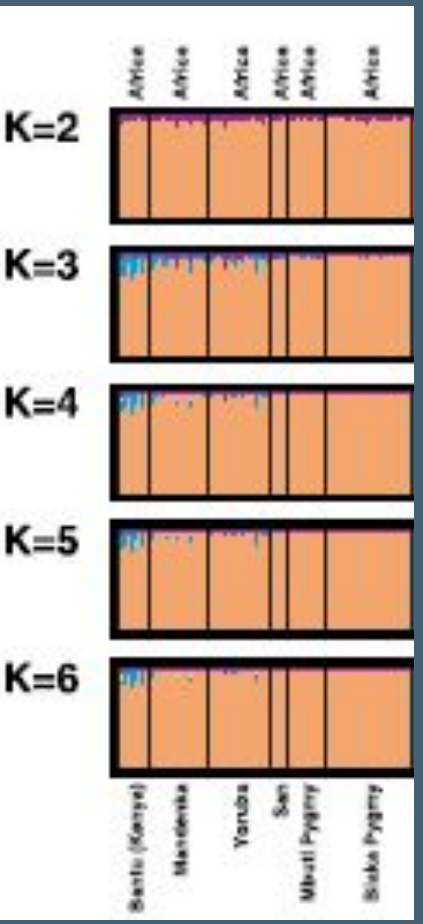
African ancestry **Americas**
East Asian ancestry
European ancestry **South Asian ancestry**

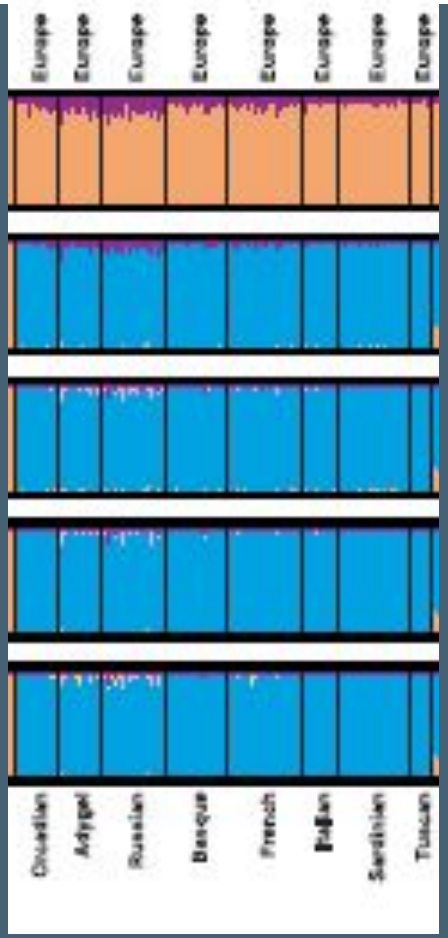
average number of singletons per genome

In the context of DNA and genomics, a "singleton" refers to a **genetic variant (like a single nucleotide polymorphism or SNP) that is observed only once in a given sample or population study.**




a
LWK
ESN
YRI
MSL
GWD
ACB
ASW
CLM
MXL
PUR
PEL
TSI
IBS
GBR
CEU
FIN
PJL
GIH
ITU
STU
BEB
CDX
KHV
CHS
CHB
JPT


K=2
K=3
K=4
K=5
K=6
Africa
Bantu (Kenya)
Mandenka
Yoruba
San
Mbuti Pygmy
Biaka Pygmy


Europe
Orcadian
Adygei
Russian
Basque
French
Italian
Sardinian
Tuscan

Earth's heart of iron begins to yield its secrets p. 18

Microglia in chronic pain recovery and relapse pp. 33 & 86

Particle acceleration in a nova explosion p. 77

Science

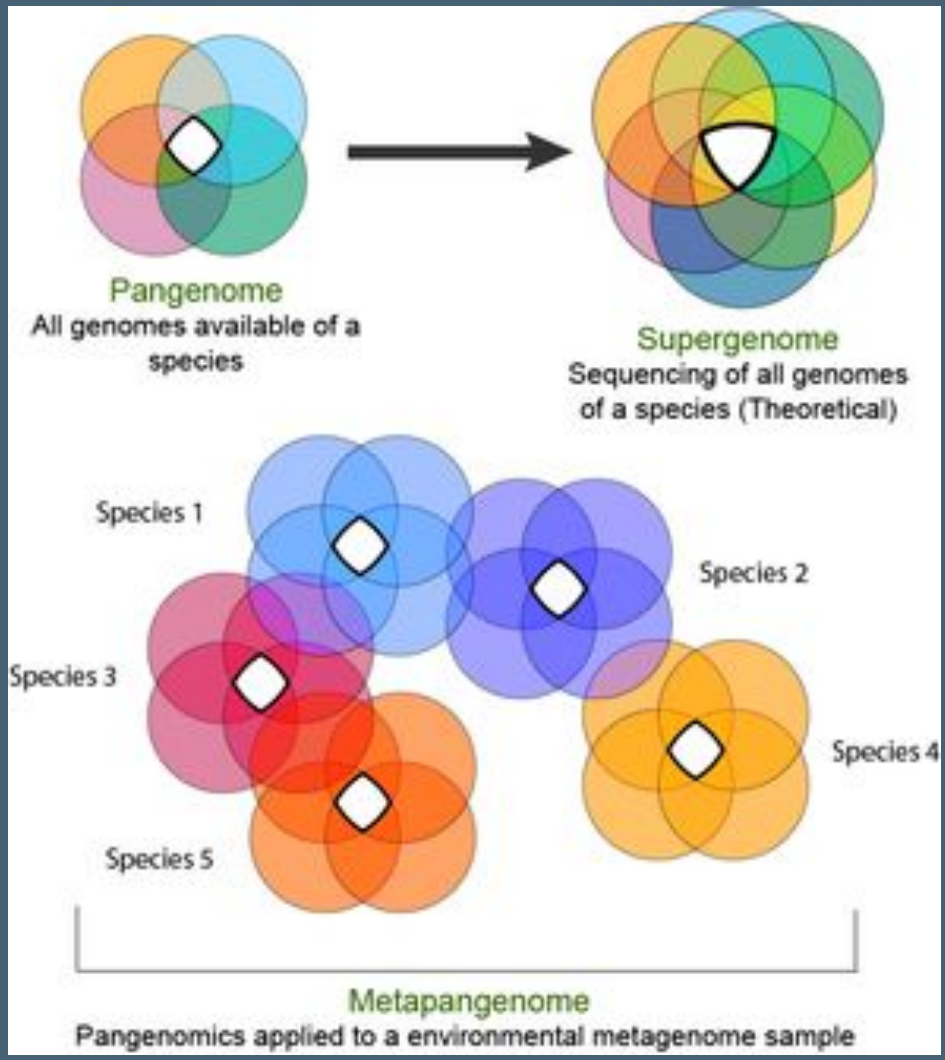
$15
1 APRIL 2022
SPECIAL ISSUE
science.org
AAAS

FILLING THE GAPS

Closing in on a complete human genome p. 42

a **pan-genome** (**pangenome** or **supragenome**) is the **entire set of genes** from all strains within a clade. More generally, it is the union of all the genomes of a clade.

The pan-genome can be broken down into a "core pangenome" that contains genes present in all individuals, a "shell pangenome" that contains genes present in two or more strains, and a "cloud pangenome" that contains genes only found in a single strain.



Genoma

Ilinsieme completo del materiale genetico (DNA o RNA, nel caso di alcuni virus) di **un organismo**, che include tutti i suoi geni, le sequenze regolatrici e le regioni non codificanti. Ogni specie, e ogni individuo all'interno di quella specie, ha un genoma specifico che codifica le istruzioni genetiche necessarie per lo sviluppo e il funzionamento.

Pangenoma

utilizzato principalmente in microbiologia e genetica delle piante, che si riferisce all'insieme completo dei geni trovati in tutte le **diverse ceppi o individui di una particolare speci**e. Il pangenoma è composto da un "core genome" (il set di geni presenti in tutti gli individui della specie) e da un "dispensable genome" (geni presenti solo in alcuni individui e non essenziali per la sopravvivenza di base, ma possono conferire vantaggi in specifici ambienti o condizioni). Questo concetto aiuta a comprendere la diversità genetica all'interno di una specie.

Supergenoma

può riferirsi all'insieme di **tutti i genomi all'interno di un gruppo ecologico o una comunità biologica.** In questo senso, potrebbe essere simile al concetto di pangenoma ma applicato a una scala più ampia, includendo molteplici specie all'interno di un ecosistema o una comunità

Metagenoma

Il **metagenoma** si riferisce all'insieme **di tutti i genomi di microbi presenti in un ambiente specifico.** Questo termine è usato soprattutto nello studio delle comunità microbiche, come quelle presenti nel suolo, nell'acqua o nel tratto gastrointestinale degli animali. La metagenomica è la tecnica attraverso la quale i ricercatori sequenziano il DNA di campioni ambientali senza la necessità di coltivare i microrganismi in laboratorio, permettendo l'analisi della diversità microbica e delle funzioni ecologiche in vari habitat.

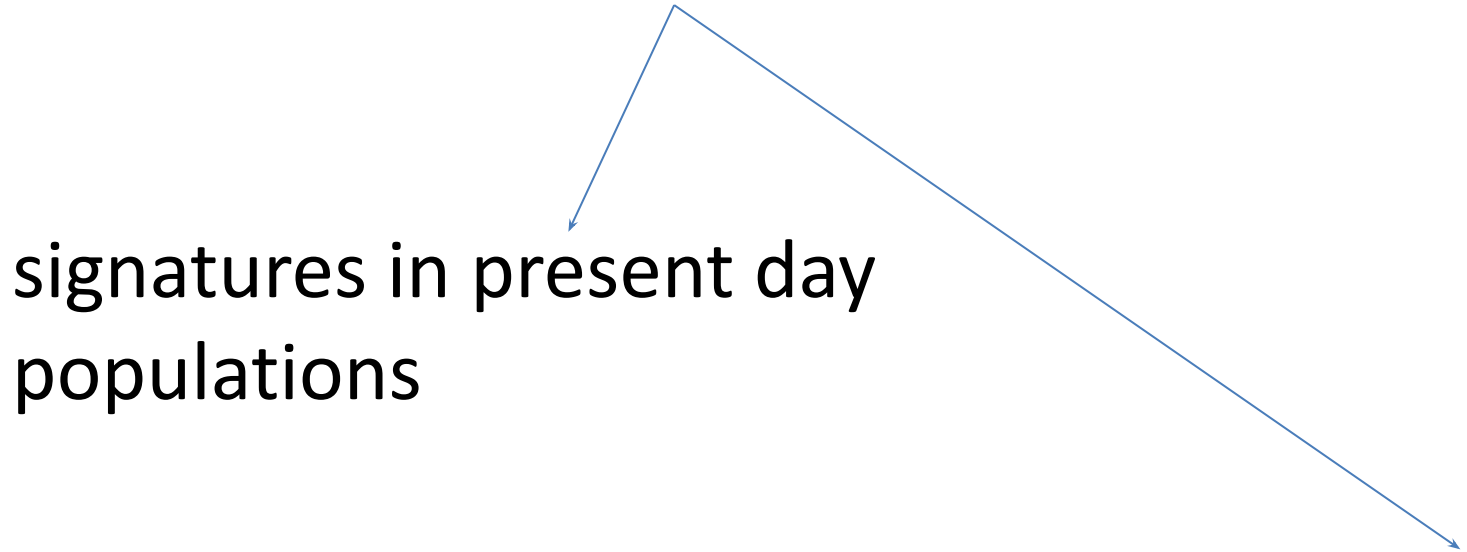
questi termini differiscono principalmente nel loro ambito e nella loro applicazione, che possono spaziare dal livello individuale al livello di specie, fino ad arrivare al livello di comunità o ecosistema.

Migration and human evolution

Migro ergo sum

Migration and human evolution

biological (genetic) changes during and after migrations

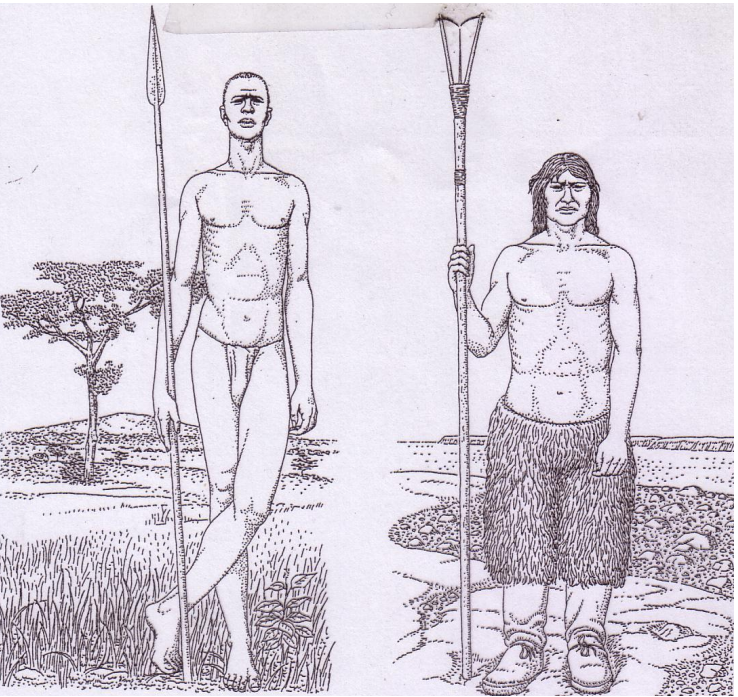


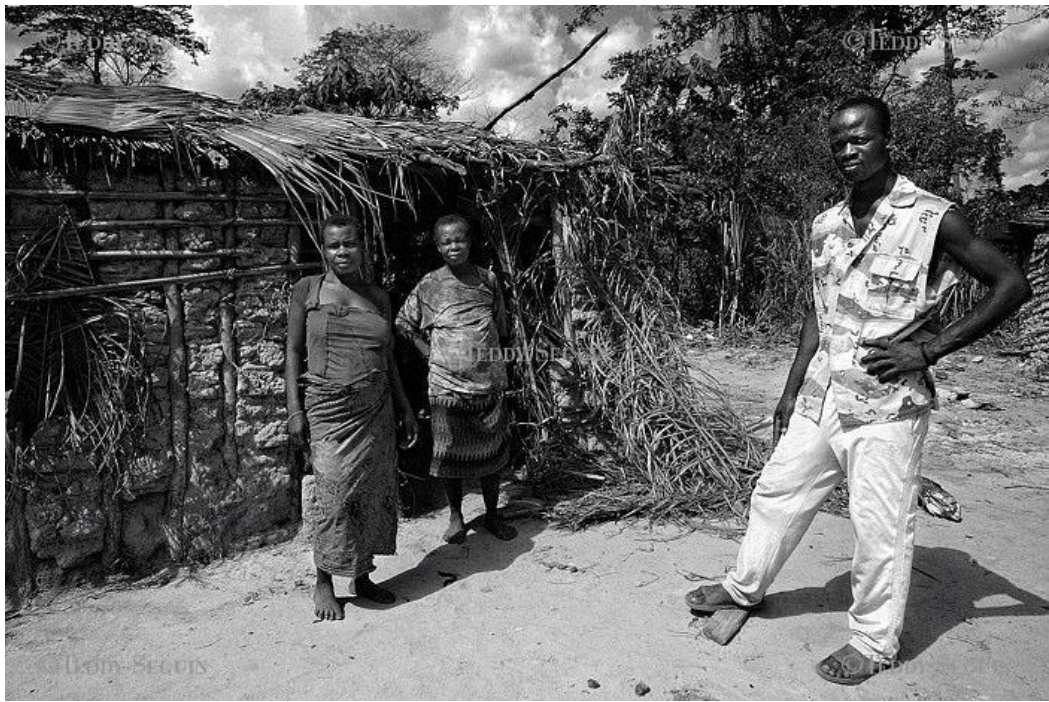
signatures in present day populations

inferences from extinct populations

Migration and human evolution

changes { biological / cultural } bio-cultural





…genetic inferences to be used with caution

Given the growing cachet associated with genetic inference in human evolution, it is perhaps wise to begin with the limitations of genetic approaches. The first point to emphasize is that genetic inference is all but helpless on its own. The fundamental reason for this is that present-day patterns of genetic variation could have been generated by many different demographic, mutational, and/or selective processes [e.g., (91)]. As a consequence, statements about past populations are inferential in nature and depend on information from other fields such as archaeology. Inference is moreover subject to considerable uncertainty stemming from the inherent variability of the evolutionary process and a variety of other complications such as modeling assumptions. On the other hand, information coming from archaeology, palaeontology, and linguistics often stems directly from ancient material.

For these reasons, the role of genetics in the study of human history has been and will continue to be largely limited to testing hypotheses formulated using data derived from other disciplines. Notable examples include testing models for the origin

Goldstein and Chikhi, Annu. Rev. Genomics Hum. Genet. 2002. 3:129–52

1. **Out of Africa II: signatures of our African origin**

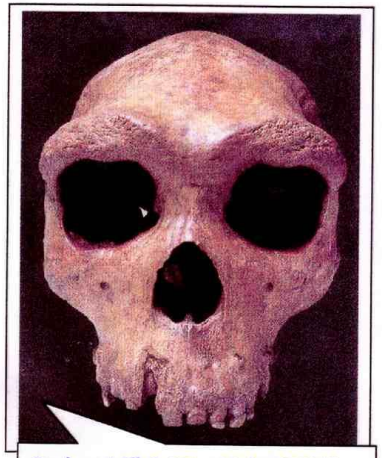
2. The Bantu expansion: how social structure shapes gene flow

3. “Out of Africa III”, a "forced" migration

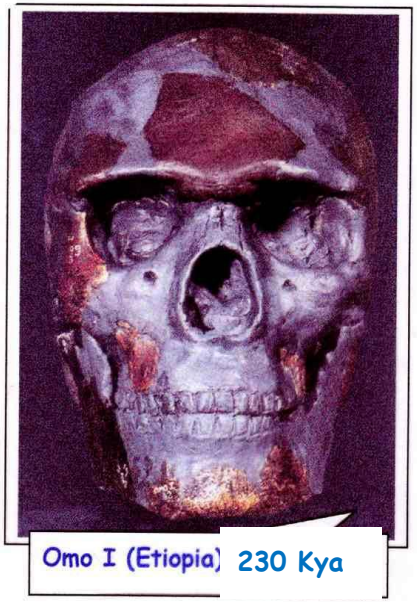
4. Neolithic in Europe: concept or gene dispersal ?

Origine di *Homo sapiens*: prove neontologiche

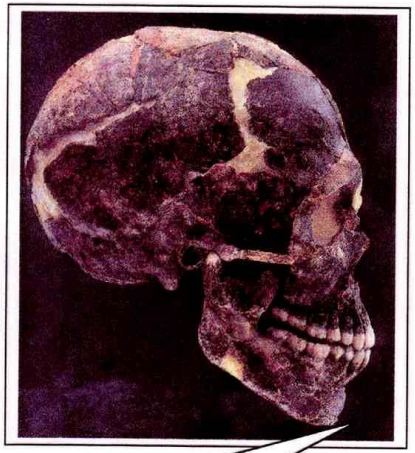
" Origine Africana recente"



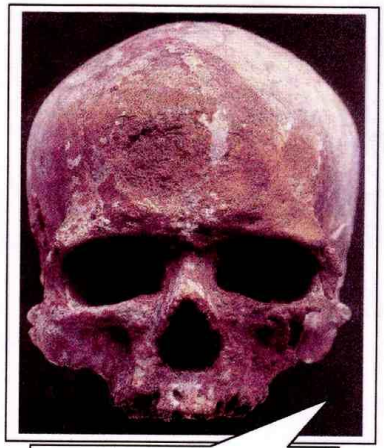
Broken Hill 1 (Zambia), 300 Kya
H. *rodhesiensis*



Omo I (Etiopia) 230 Kya



Qafzeh IX (Israele)
90-100 KYa

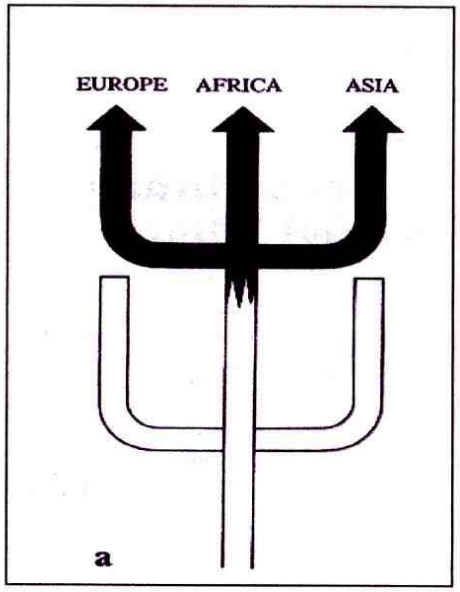


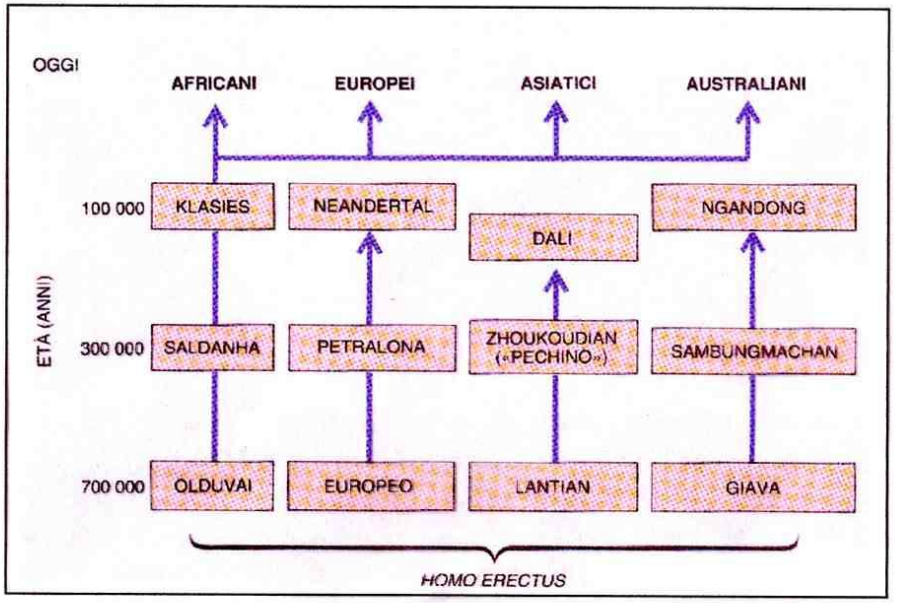
Cro-Magnon (Francia)
30-32 KYa

Origine di *Homo sapiens*: prove neontologiche

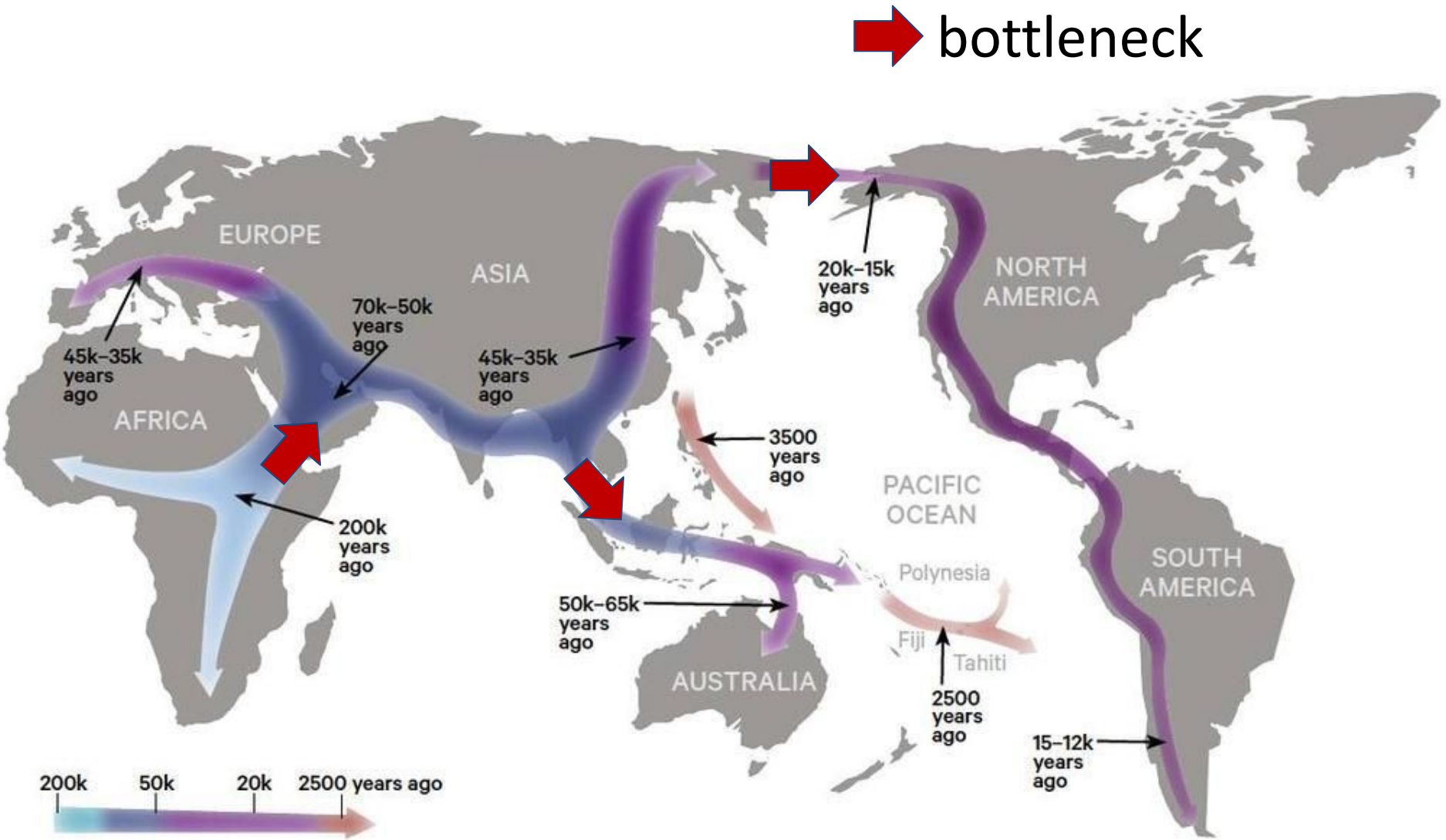
1." Origine Africana recente"

- comparsa di *Homo sapiens* in Africa tra 100 e 200 Kya
- migrazione in Asia S.Oc. e Europa tra 100 e 40 Kya
- evento di speciazione: collo di bottiglia e cladogenesi: equilibri punteggiati
  - ✓ rimpiazzamento delle popolazioni "locali"
  - ✓ cultura (linguaggio) e/o malattie?



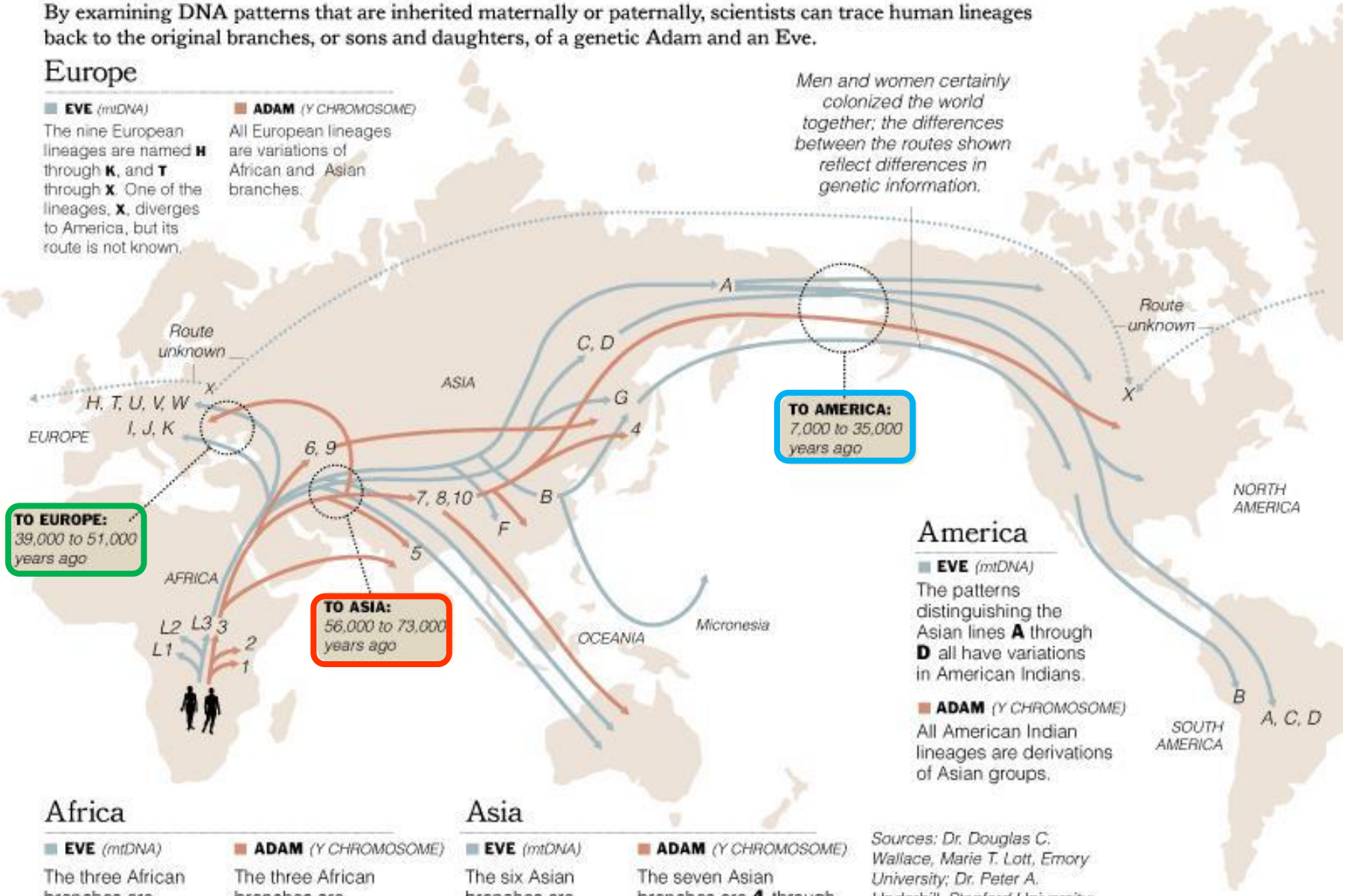


Le più antiche migrazioni umane



Tracing Human History Through Genetic Mutations

By examining DNA patterns that are inherited maternally or paternally, scientists can trace human lineages back to the original branches, or sons and daughters, of a genetic Adam and an Eve.



Europe

EVE *(mtDNA)*

The nine European lineages are named **H** through **K**, and **T** through **X**. One of the lineages, **X**, diverges to America, but its route is not known.

ADAM *(Y CHROMOSOME)*

All European lineages are variations of African and Asian branches.

America

EVE *(mtDNA)*

The patterns distinguishing the Asian lines **A** through **D** all have variations in American Indians.

ADAM *(Y CHROMOSOME)*

All American Indian lineages are derivations of Asian groups.

Africa

EVE *(mtDNA)*

The three African branches are named **L1** through **L3**, and **L3** separates into all the other branches.

ADAM *(Y CHROMOSOME)*

The three African branches are named **1**, **2** and **3**, and **3** separates into all the other branches.

Asia

EVE *(mtDNA)*

The six Asian branches are named **A** through **D** and **F** and **G**.

ADAM *(Y CHROMOSOME)*

The seven Asian branches are **4** through **10**, and these groups branch off into Oceania, Europe and America.

Sources: Dr. Douglas C. Wallace, Marie T. Lott, Emory University; Dr. Peter A. Underhill, Stanford University; "Genes, Peoples, and Languages," by Dr. Luca Cavalli-Sforza

Steve Duenes/The New York Times

migrazioni (umane)

Spostamento, d**efinitivo o temporaneo**, di gruppi da un territorio a un altro, da una ad altra sede, determinato da ragioni varie, ma essenzialmente da necessità di vita.
Le m. possono essere di grandi masse di uomini (sono allora per lo più permanenti) oppure per **piccoli contingenti**.

Colonizzazione, in ecologia e biogeografia, il processo di insediamento di una **comunità biologica** in un'area che ne è priva, in seguito a una forte perturbazione (per es. un incendio o un'eruzione) o in quanto **neoformata**;

grande spostamento di popolazione

come ?

quale migrazione?

molte piccole migrazioni “**stop and go**” in un lungo arco di tempo

lenta **diffusione** geografica (**dispersal**) con colonizzazione

Sotto la spinta di…

cambiamenti climatici (pulsati?)

migrazioni fauna

Con l'aiuto di…

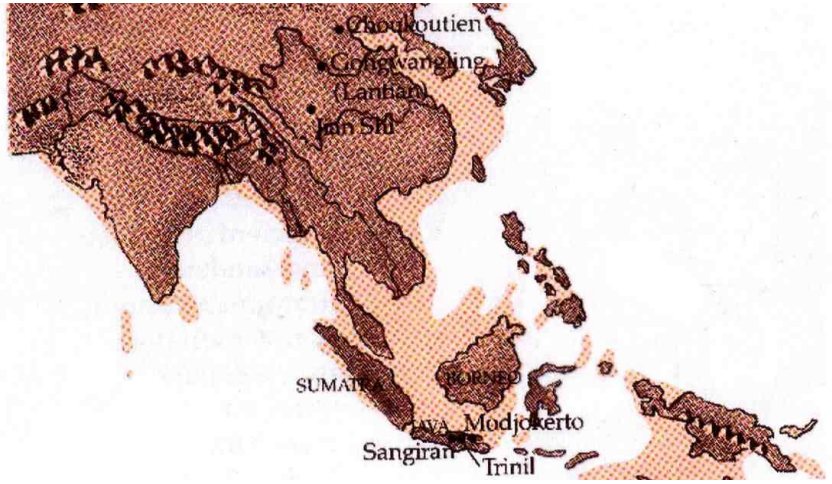
innovazioni tecnologiche (fuoco)

piattaforme continentali

Con l'effetto di…

adattamenti (biologici e culturali)

diversificazione (caso e necessità)

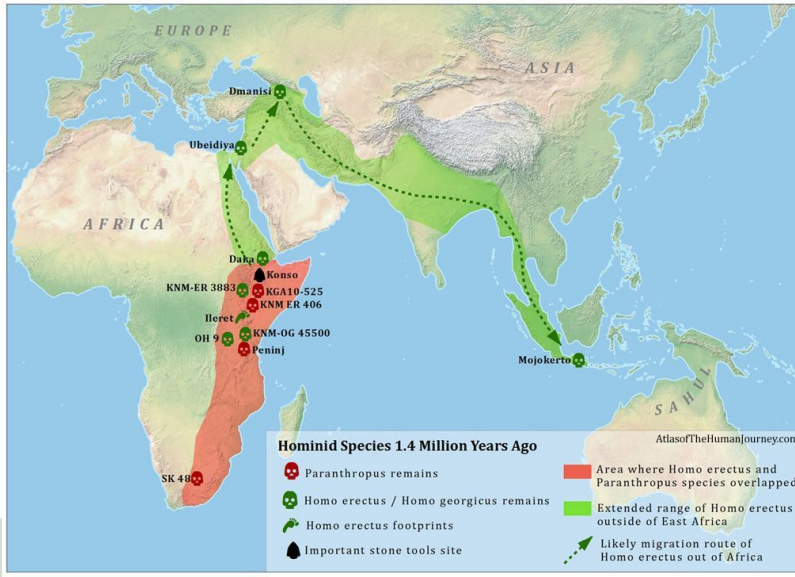


passato presente

perché migriamo?

effetti sociali delle migrazioni

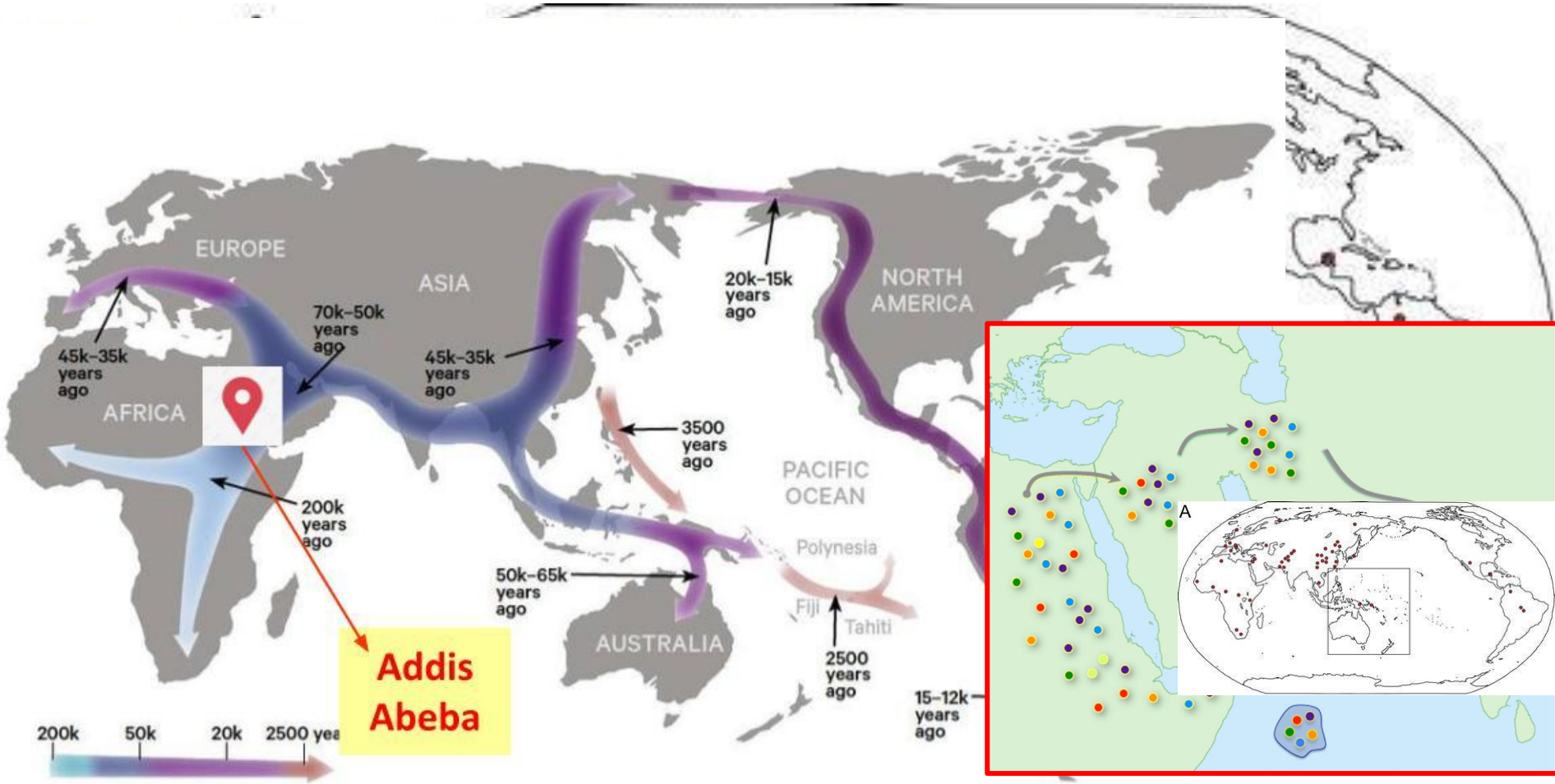
… di chi è la terra?



15942–15947 | PNAS | November 1, 2005

Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa

Sohini Ramachandran*†, Omkar Deshpande‡, Charles C. Roseman§, Noah A. Rosenberg¶, Marcus W. Feldman*, and L. Luca Cavalli-Sforza†‖



HGDP-CEPH population locations

783 autosomal microsatellite loci

DISTANZA GENETICA

Fst TRA DUE POPOLAZIONI AD UN LOCUS CON DUE ALLELI

Fst= Vp / P (1-P)

p = frequenza allelica

P = frequenza allelica media

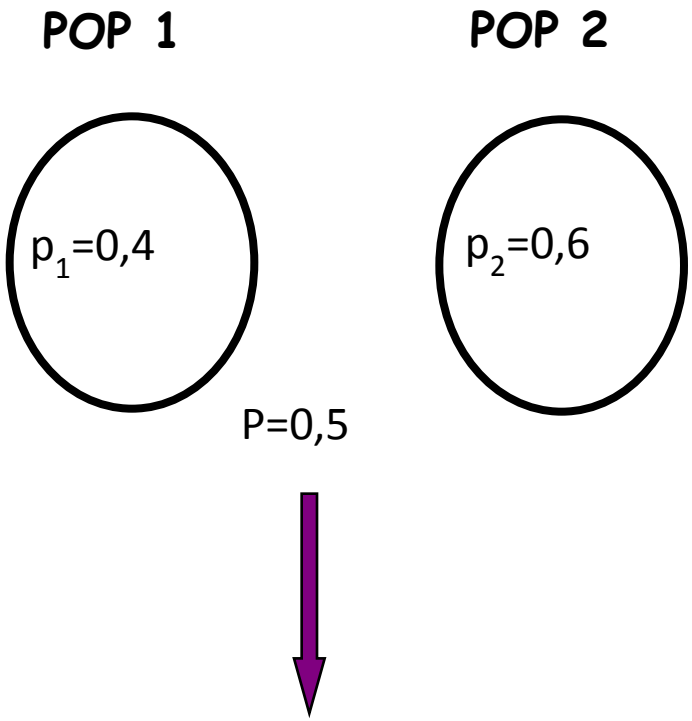
1 e 2 = popolazione 1 e 2

Devianza = $\sum(X- Xm)^2$
Somma degli scarti
al quadrato

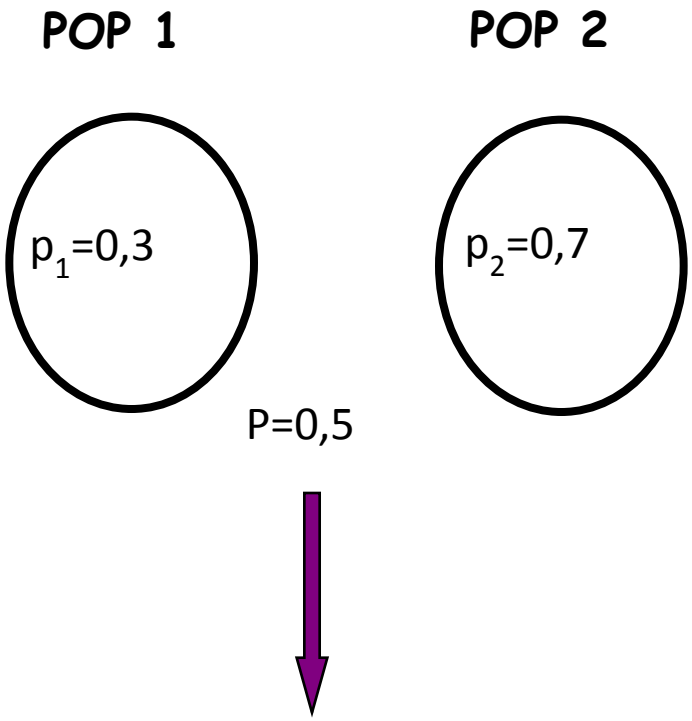
Varianza =$\sum(X-Xm)^2/N$
La devianza/N

$$Fst = \frac{(p_1-P)^2 + (p_2-P)^2}{2} \times \frac{1}{P(1-P)}$$

ESEMPIO DI CALCOLO DELLA DISTANZA GENETICA Fst



$$Fst = \frac{(0,4-0,5)^2 + (0,6-0,5)^2}{2 \times [0,5 \times (1-0,5)]} = \mathbf{0,04}$$

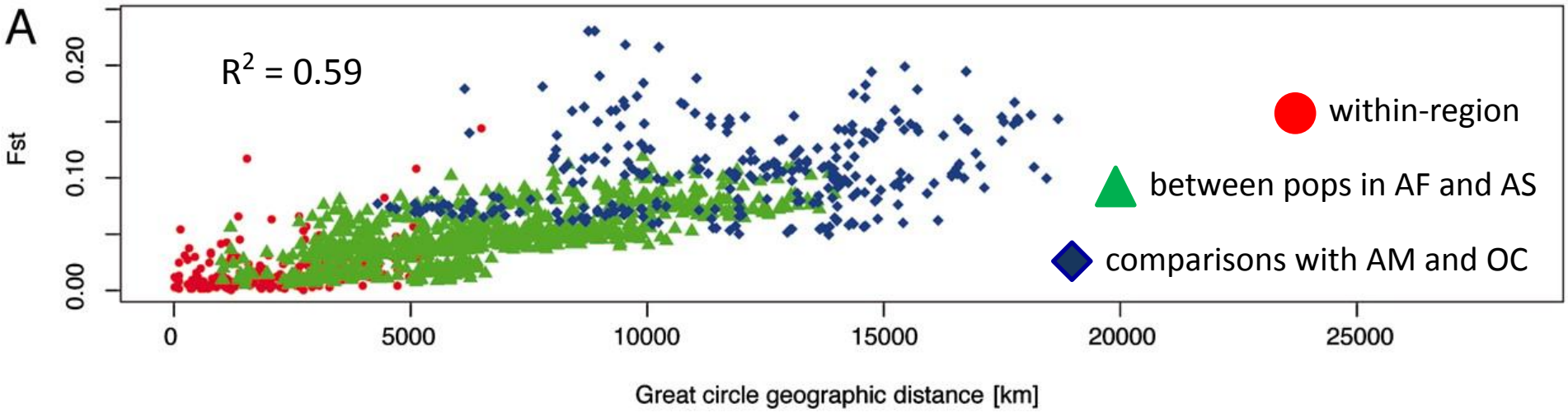


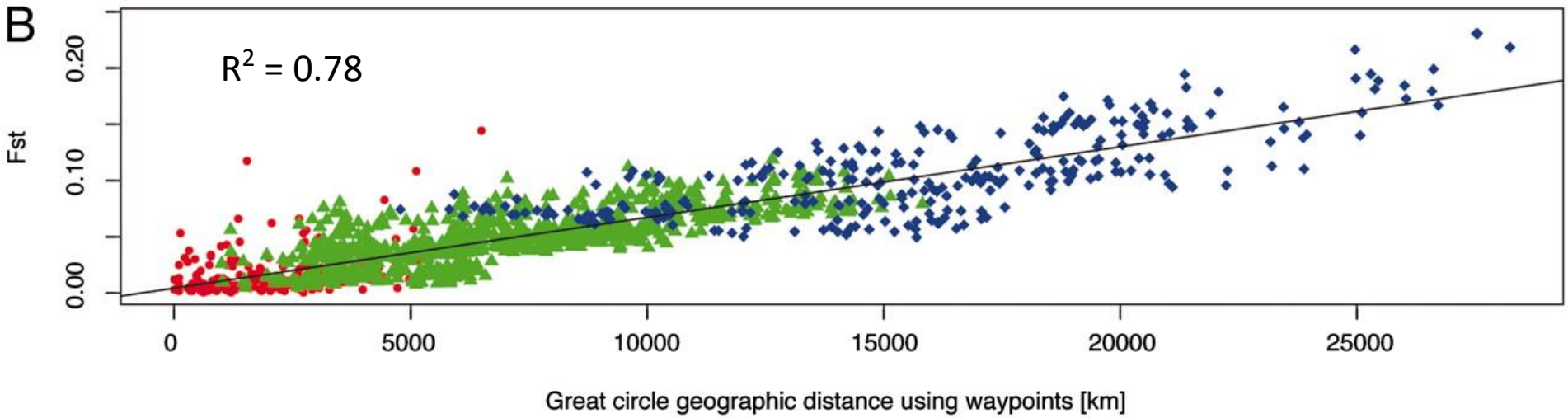
$$Fst = \frac{(0,3-0,5)^2 + (0,7-0,5)^2}{2 \times [0,5 \times (1-0,5)]} = \mathbf{0,16}$$

15942–15947 | PNAS | November 1, 2005

Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa

Sohini Ramachandran*†, Omkar Deshpande‡, Charles C. Roseman§, Noah A. Rosenberg¶, Marcus W. Feldman*, and L. Luca Cavalli-Sforza†||

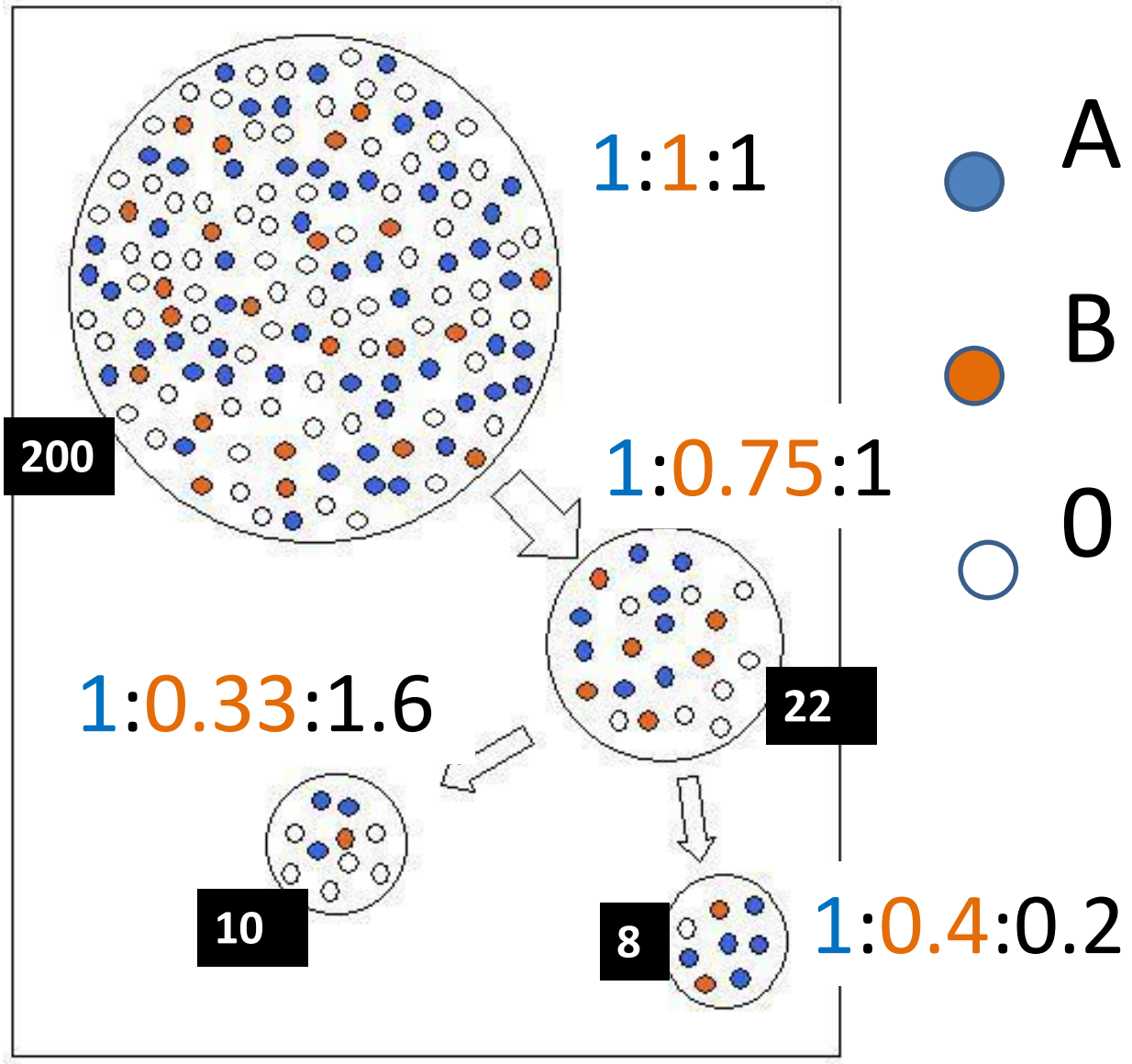




Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa

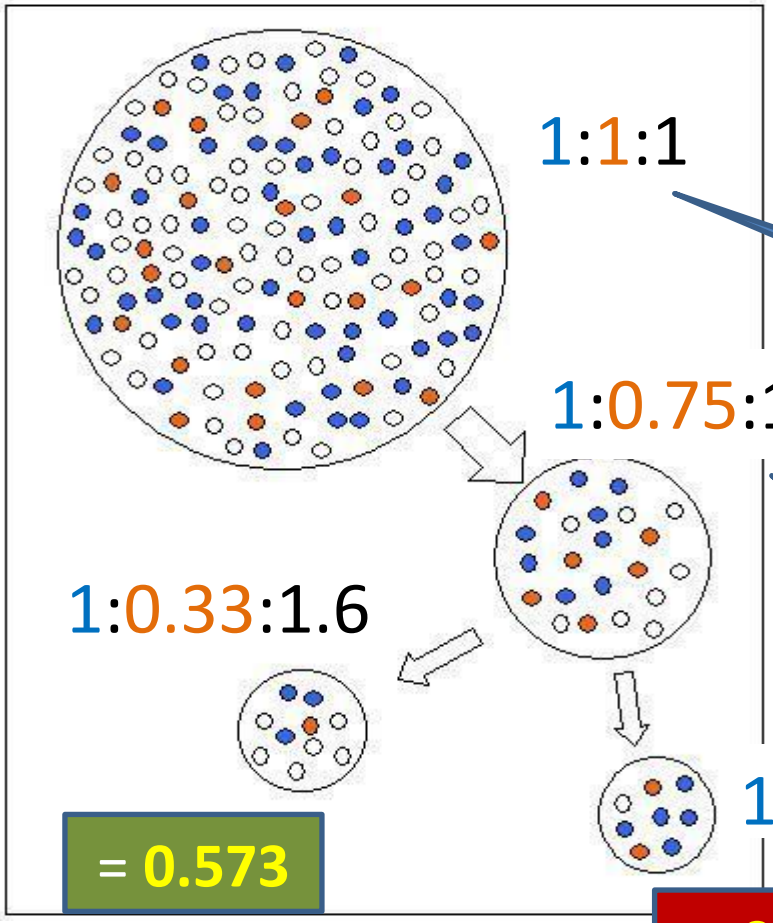
Sohini Ramachandran*†, Omkar Deshpande‡, Charles C. Roseman§, Noah A. Rosenberg¶, Marcus W. Feldman*, and L. Luca Cavalli-Sforza†‖

15942–15947 | PNAS | November 1, 2005



Eterozigosità attesa

$$H = (1 - \sum_i x_i^2)$$



1:1:1

1:0.75:1

1:0.33:1.6

1:0.4:0.2

= 0.573

= 0.531

= 1 – (.33² +.33² +.33²)

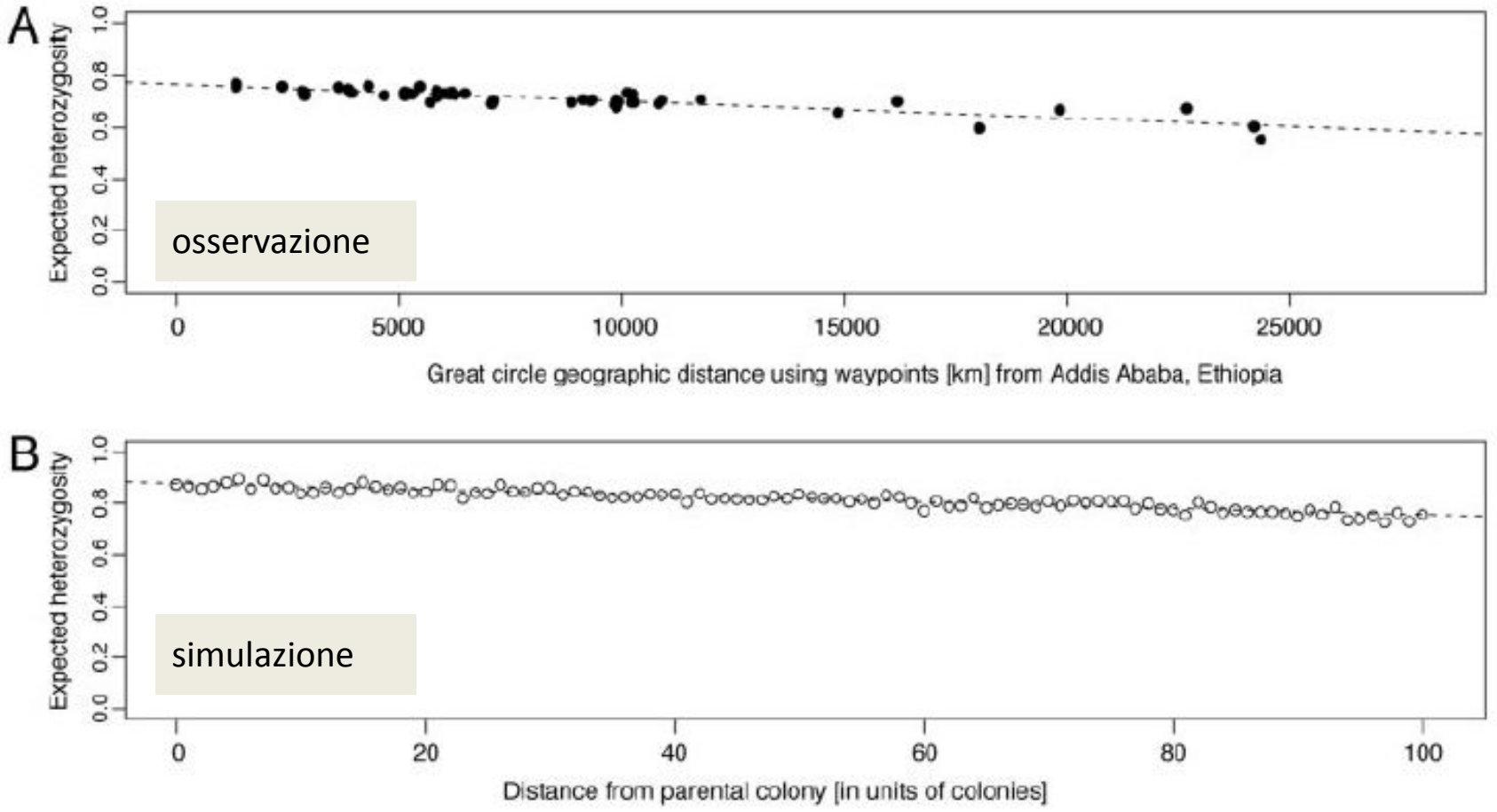
= 1 - .327 ≈ **0.673**

= 1 – (.36² +.27² +.36²)

= 1 - .338 ≈ **0.661**

Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa

Sohini Ramachandran*†, Omkar Deshpande‡, Charles C. Roseman§, Noah A. Rosenberg¶, Marcus W. Feldman*, and L. Luca Cavalli-Sforza†||



15 APRIL 2011 VOL 332 SCIENCE

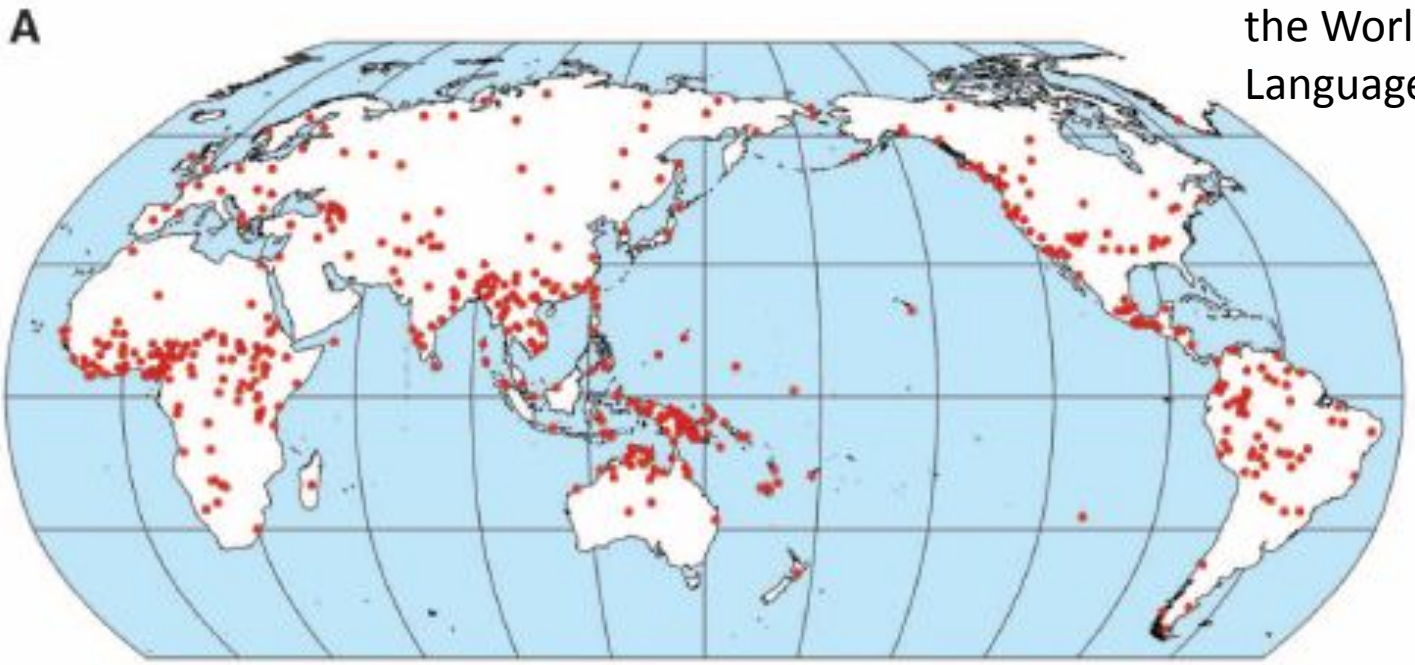
Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa

Quentin D. Atkinson[1,2]*

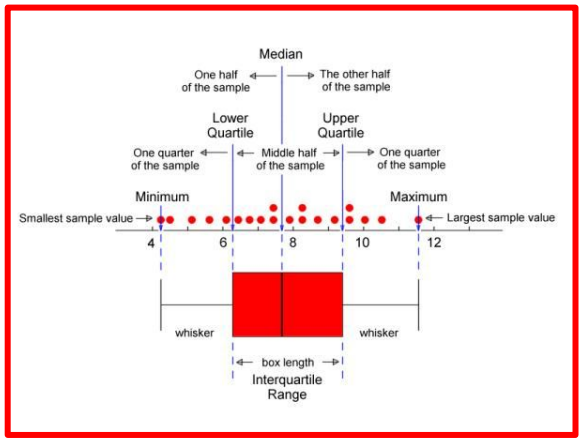
perceptually distinct units of sound that differentiate words: vowel, consonant, and tone inventories

Un fonema è un'unità minima differenziante, indivisibile ed astratta di un sistema linguistico. Più precisamente, esso è la rappresentazione astratta di un suono

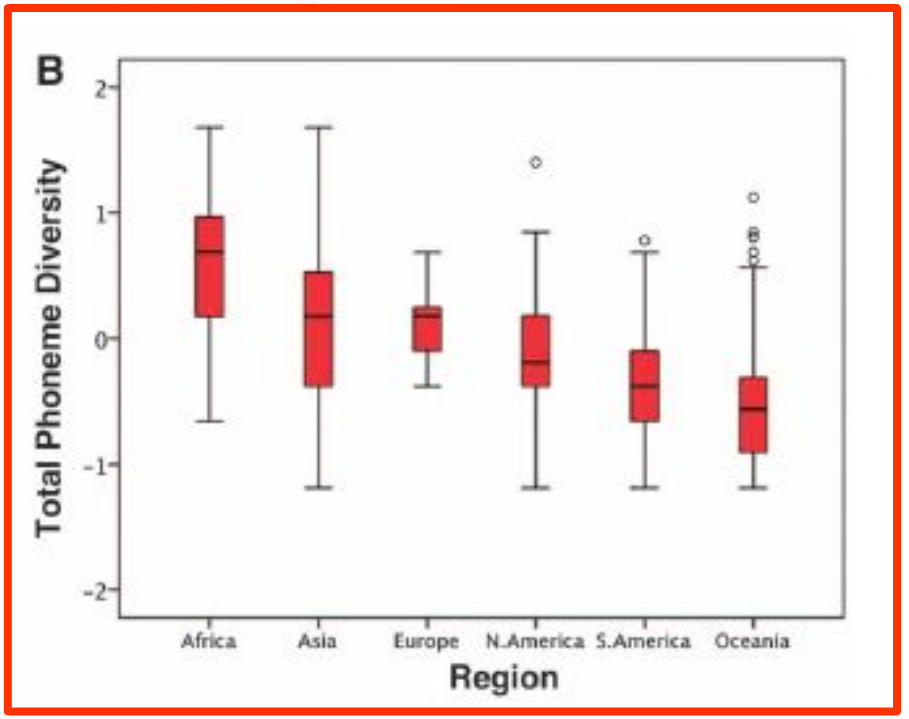
Human genetic and phenotypic diversity declines with distance from Africa, as predicted by a serial founder effect in which successive population bottlenecks during range expansion progressively reduce diversity, underpinning support for an African origin of modern humans. Recent work suggests that a similar founder effect may operate on human culture and language. Here I show that the number of phonemes used in a global sample of 504 languages is also clinal and fits a serial founder–effect model of expansion from an inferred origin in Africa. This result, which is not explained by more recent demographic history, local language diversity, or statistical non-independence within language families, points to parallel mechanisms shaping genetic and linguistic diversity and supports an African origin of modern human languages.



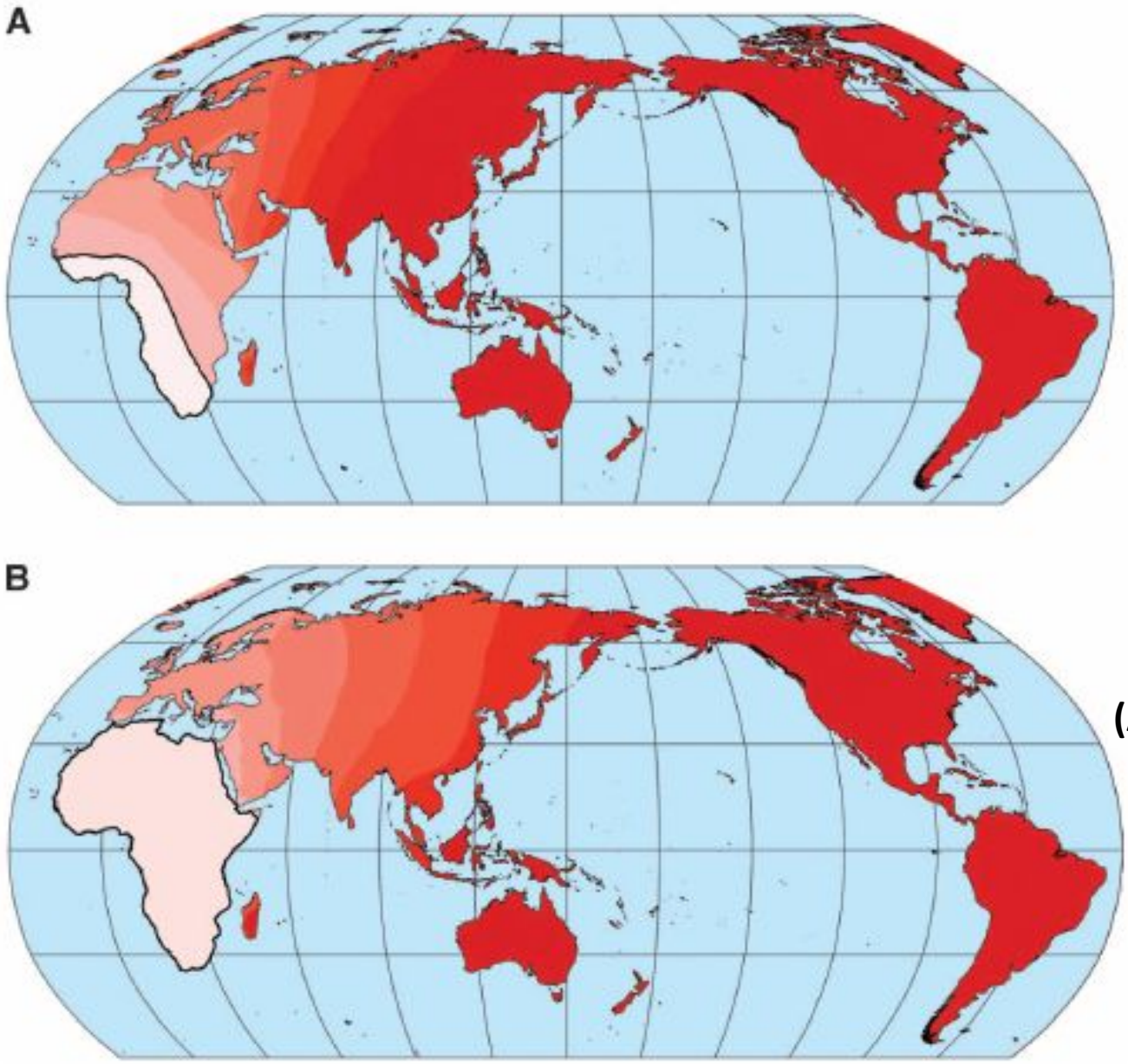
the World Atlas of Language Structures (WALS)



largest phoneme inventories in Africa and the smallest in South America and Oceania.



Lighter shading implies a stronger inverse relationship between phonemic diversity and distance from the origin and better fit of the model



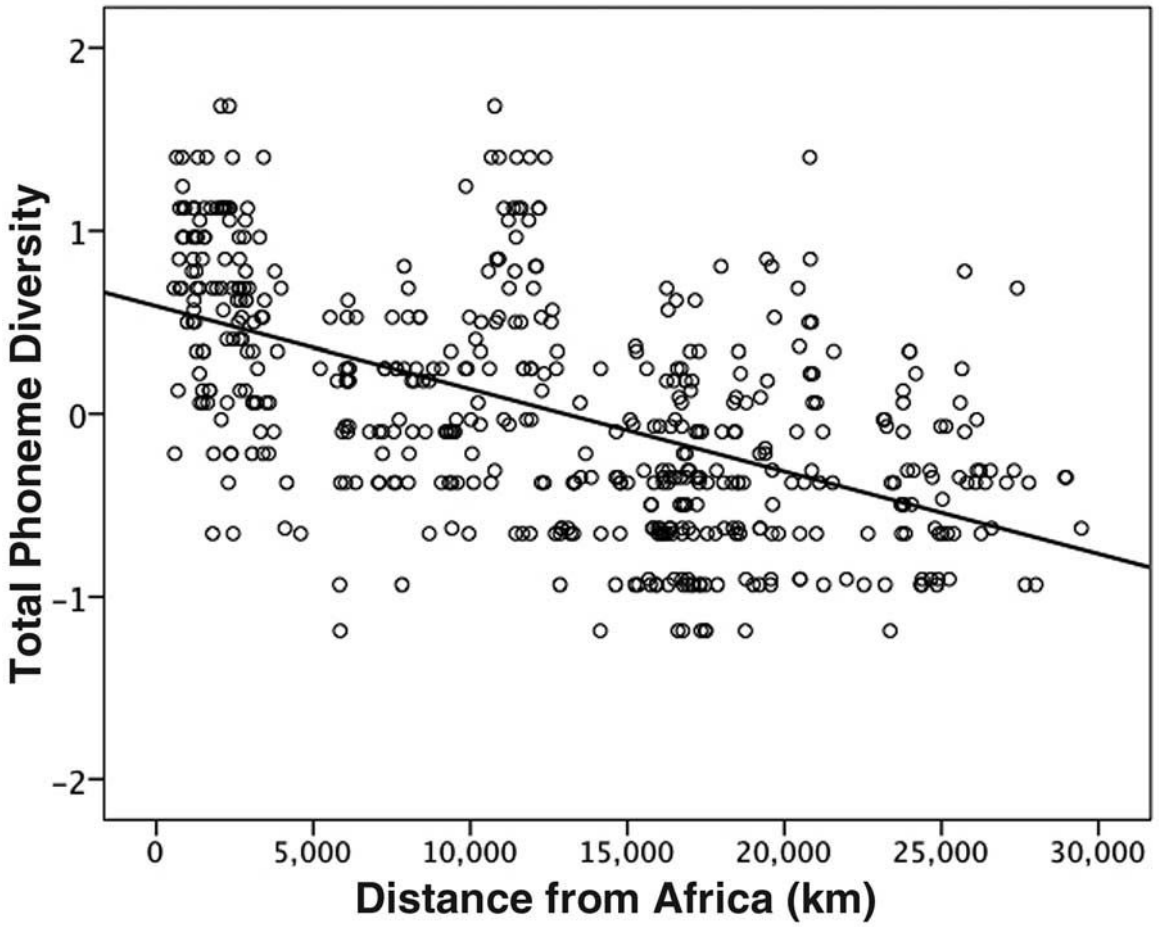
likely location of a single language origin under a founder effect model of phonemic diversity (controlling for population size) inferred from

(A) individual languages

(B) mean diversity across language families

The best-fit model incorporating population size and distance from the origin explains **31%** of the variance in phoneme inventory size

Controlling for population size, distance from origin accounts for **19%** of the variance in phonemic diversity.



neutral genetic markers: 80 to 85%
human mitochondrial DNA: 18%

Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa

Quentin D. Atkinson[1,2]*

cultural evolution more generally, these findings support the proposal that a cultural founder effect operated during our colonization of the globe, potentially limiting the size and cultural complexity of societies at the vanguard of the human expansion (*10*, *11*). An origin of modern languages predating the African exodus 50,000 to 70,000 years ago puts complex language alongside the earliest archaeological evidence of symbolic culture in Africa 80,000 to 160,000 years ago (*27*, *28*). Truly modern language, akin to languages spoken today, may thus have been the key cultural innovation that allowed the emergence of these and other hallmarks of behavioral modernity and ultimately led to our colonization of the globe (*29*).

1. Out of Africa II: signatures of our African origin

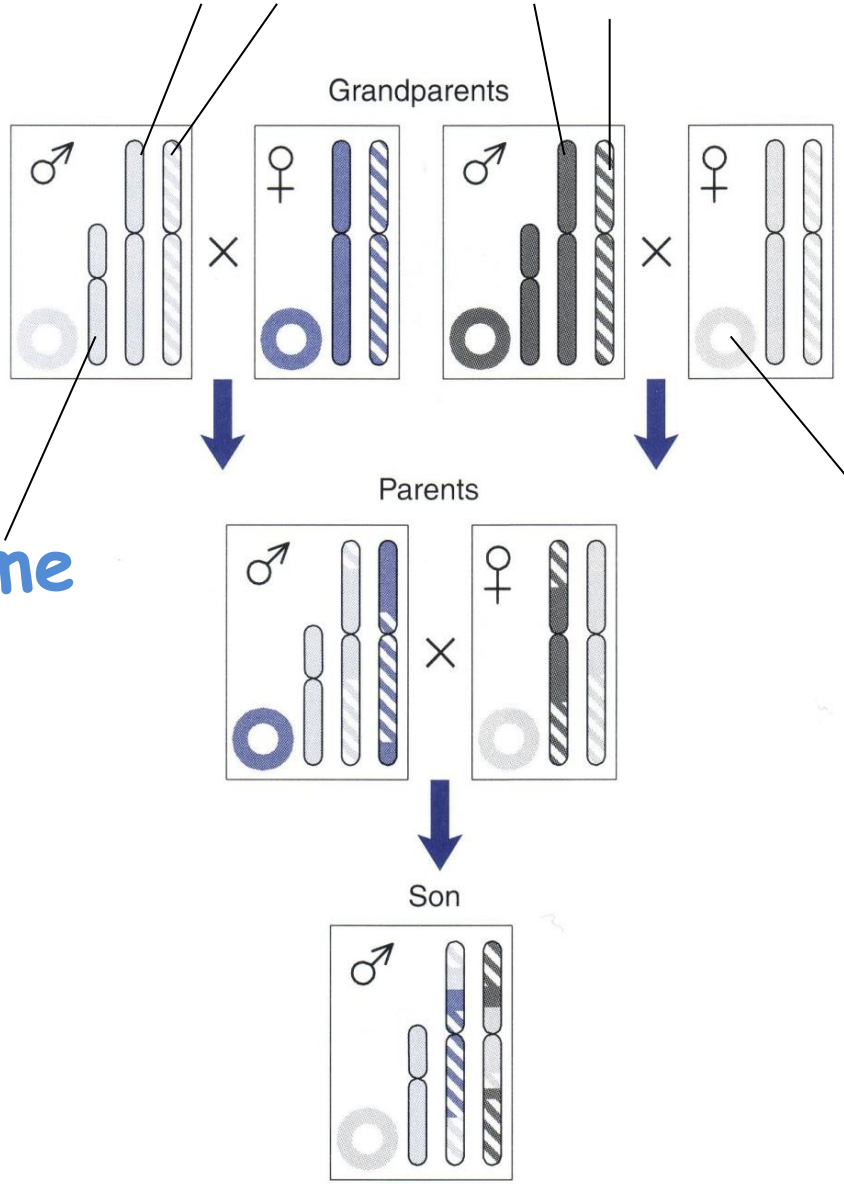
2. The Bantu expansion: how social structure shapes gene flow

3. “Out of Africa III”, a "forced" migration

4. Neolithic in Europe: concept or gene dispersal ?

3. The Bantu expansion:
how social structure shapes gene flow

Autosomal polymorphisms



Y-chromosome

mtDNA

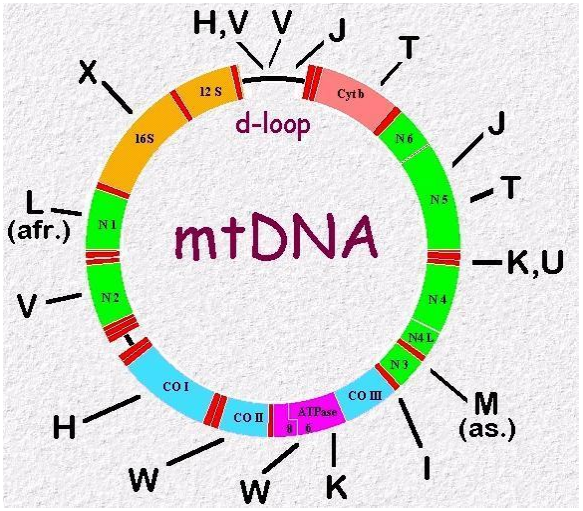
mtDNA and Y-chromosome

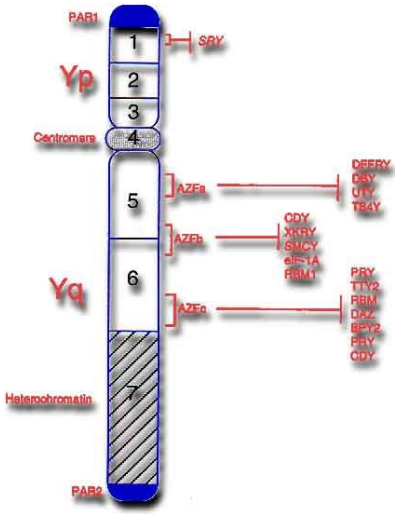
✔SEPARATE STUDY OF MALE AND FEMALE INHERITANCE

✔ LACK OF RECOMBINATION: PHYLOGEOGRAPHIC APPROACH

✔COMPARABLE MODE AND RATE OF EVOLUTION

(Fast evolving and slow-evolving polymorphisms)

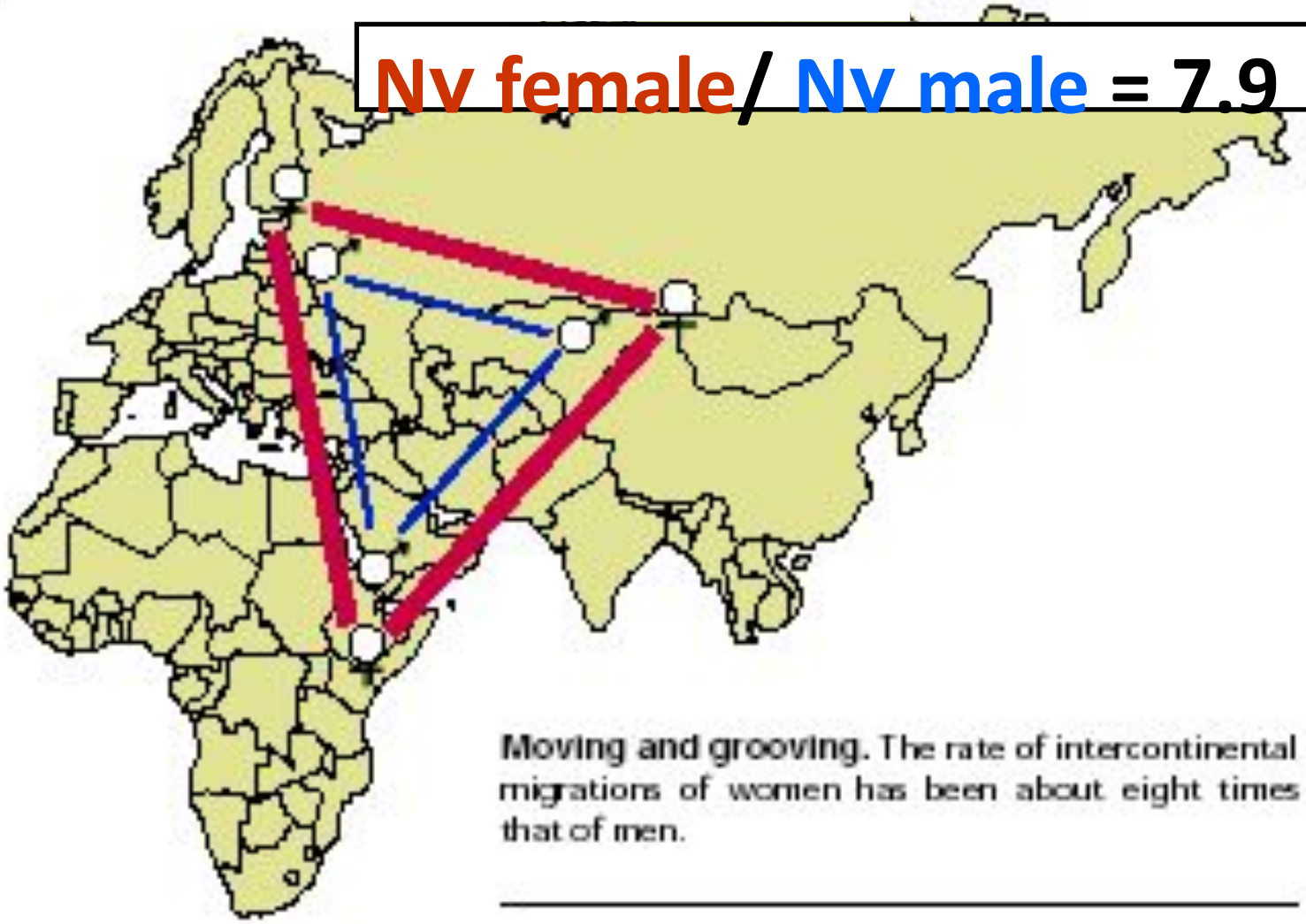




Genetic evidence for a higher female migration rate in humans

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

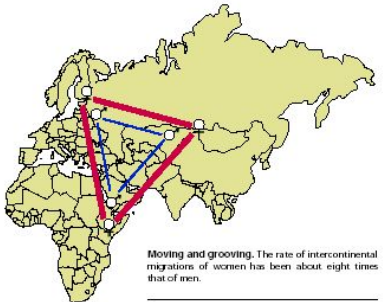
Nv female/ Nv male = 7.9



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

Genetic evidence for a higher female migration rate in humans

Mark T. Seielstad[1], Eric Minch[2] & L

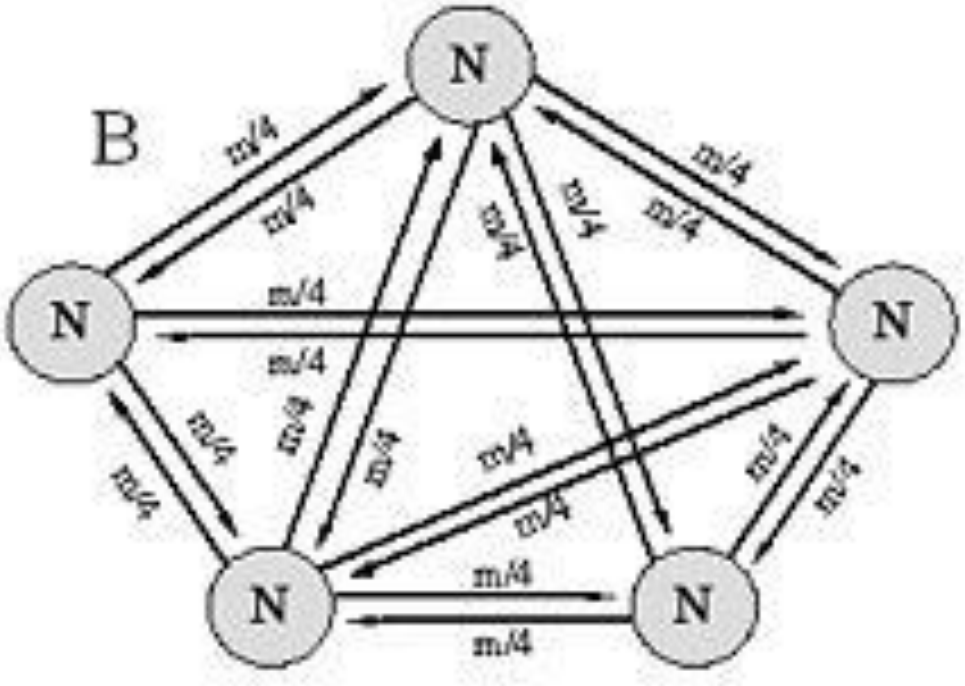


Moving and grooving. The rate of intercontinental migrations of women has been about eight times that of men.

Table 2 • F_{ST}, Nv and their ratios for SNP data

Genetic system	F_{ST}	Nv	Nv ratios
mtDNA	0.186	4.38	2.94 (mtDNA:autosomes) 7.96 (mtDNA:Y) 2.71 (autosomes:Y)
autosomes	0.144	1.49	
Y chromosome	0.645	0.550	

$$F_{st} = \frac{V_p}{p(1-p)}$$



$$Nv = (1/Fst)-1$$

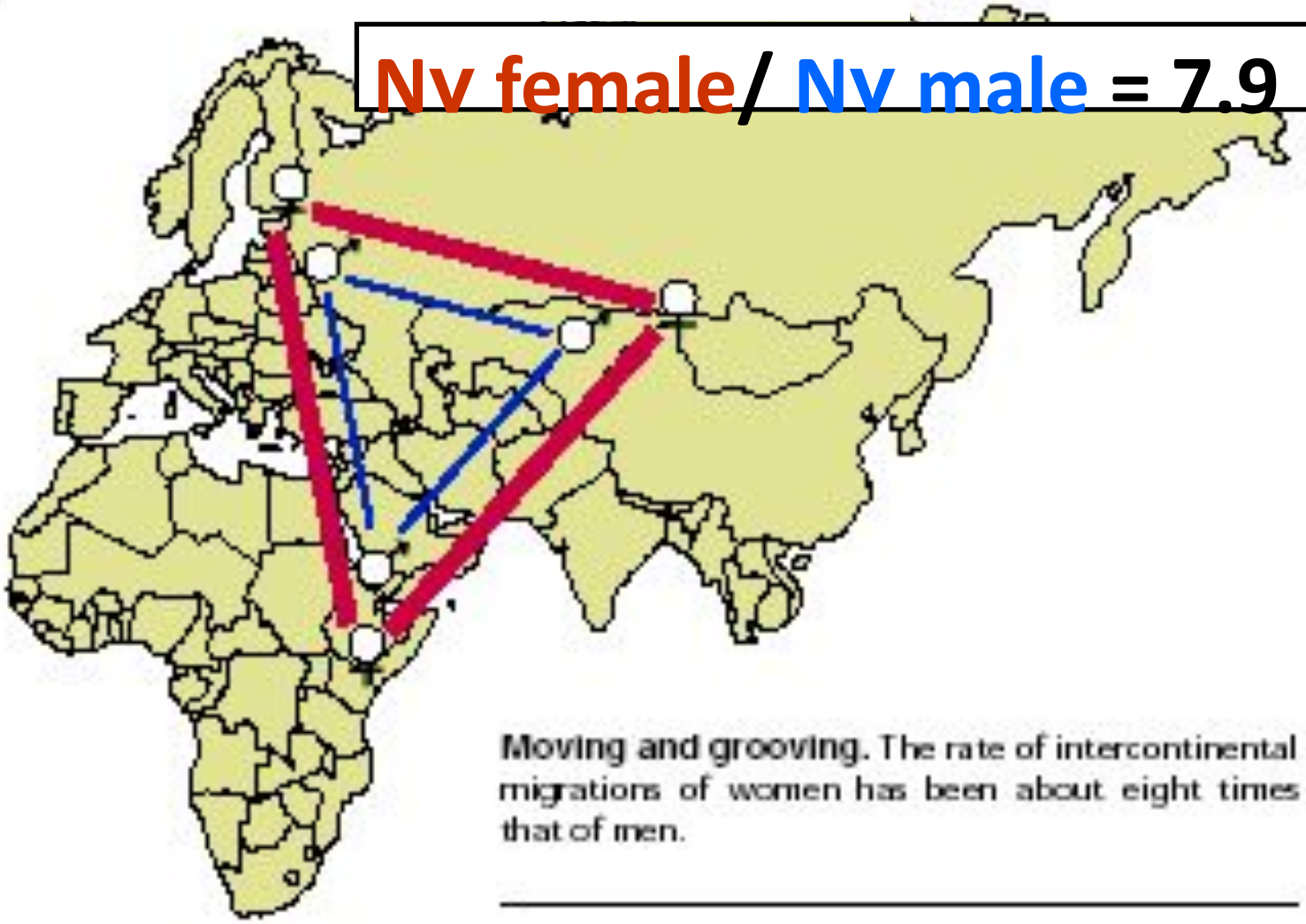
N = effective population size

v (ni)= the sum of migration (m) and mutation (mu)

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.

se Fst = 0.112

mtDNA: PICCOLA DISTANZA GENETICA

Nv = (1/0.112)-1

□ 8.9-1 □ 7.9

se Fst = 0.5

Y GRANDE DISTANZA GENETICA

Nv = (1/0.5)-1

□ 2-1 □ 1

Are there other cultural factors which influence the way in which genes arrange in the human populations?

Hierarchical Patterns of Global Human Y-Chromosome Diversity

Michael F. Hammer,† Tatiana M. Karafet,* Alan J. Redd,* Hamdi Jarjanazi,* Silvana Santachiara-Benerecetti,‡ Himla Soodyall,§ and Stephen L. Zegura†*

*Laboratory of Molecular Systematics and Evolution and †Department of Anthropology, University of Arizona; ‡Department of Genetics, Universita degli Studi di Pavia, Pavia, Italy; and §SAMIR, University of Witwatersrand, Johannesburg, South Africa

	Fst	
	mtDNA	Cromosoma Y
Africa	0,339	0,251
Asia	0,045	0,271
Europa	0,007	0,128

$$Nni = (1/Fst)-1$$

Fst	mtDNA	cromosoma Y
Africa	1,95	2,98

Nv female/ Nv male = 0.7

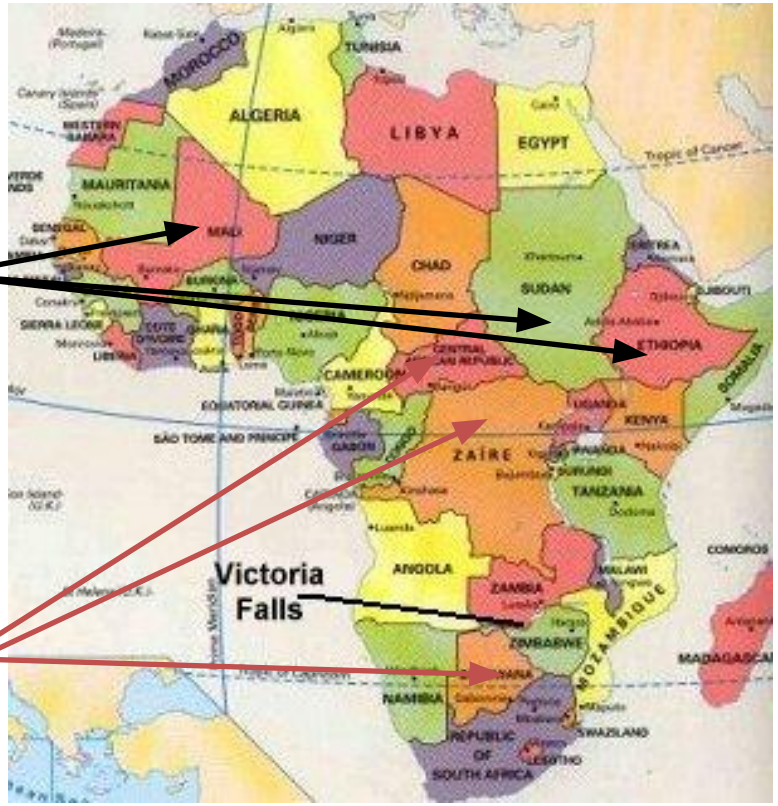
letter 1998 Nature America Inc. • http://genetics.nature.com

Genetic evidence for a higher female migration rate in humans

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

14 populations from Sudan, Ethiopia and Mali

Pygmies, Bantus from Central and Eastern Africa, and Bushmen



Hierarchical Patterns of Global Human Y-Chromosome Diversity

Michael F. Hammer,† Tatiana M. Karafet,* Alan J. Redd,* Hamdi Jarjanazi,* Silvana Santachiara-Benerecetti,‡ Himla Soodyall,§ and Stephen L. Zegura†*

*Laboratory of Molecular Systematics and Evolution and †Department of Anthropology, University of Arizona; ‡Department of Genetics, Universita degli Studi di Pavia, Pavia, Italy; and §SAMIR, University of Witwatersrand, Johannesburg, South Africa

Variation of Female and Male Lineages in Sub-Saharan Populations: the Importance of Sociocultural Factors

Mol. Biol. Evol. 21(9):1673–1682. 2004

Giovanni Destro-Bisol, Francesco Donati,* Valentina Coia,* Ilaria Boschi,†
Fabio Verginelli,‡ Alessandra Caglià,† Sergio Tofanelli,§ Gabriella Spedini,*
and Cristian Capelli†*

Food producers

Degree of polygyny

Cultural barriers to symmetric gene flow

Respect of patrilocality

Hunter-gatherers

The Bantu expansion

Africa's largest language family

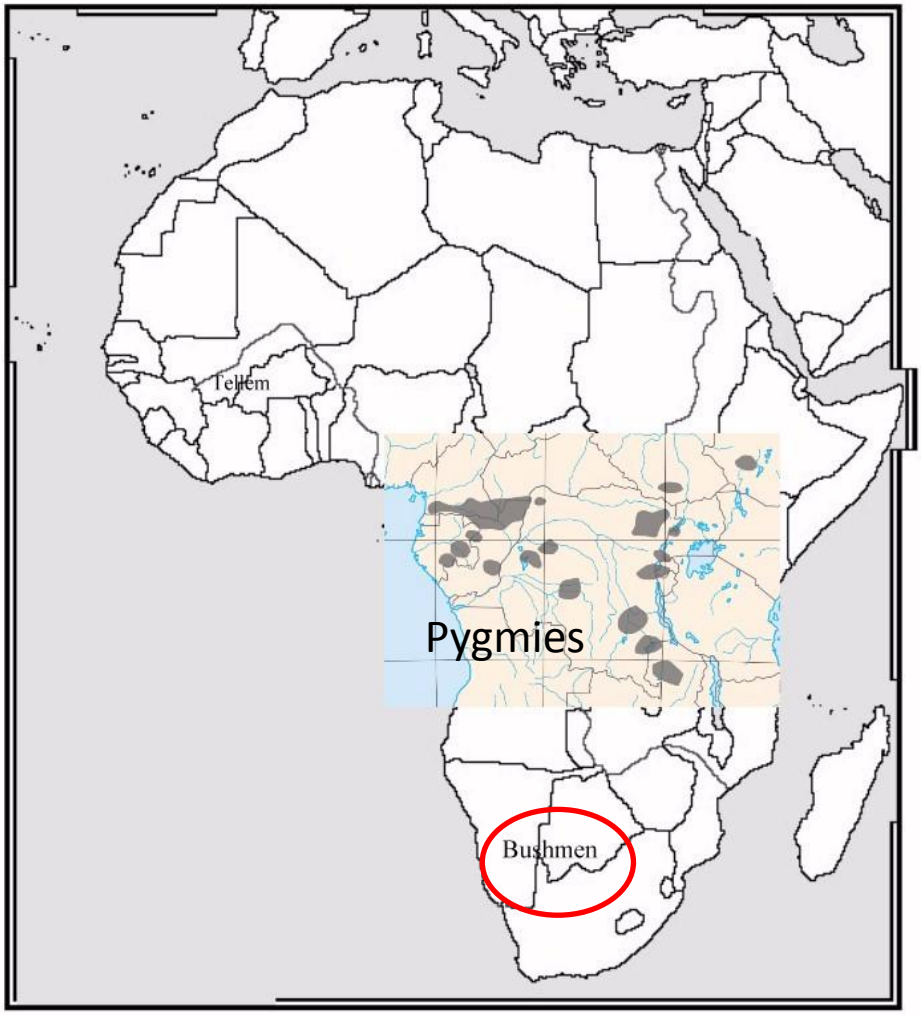
1 African out of 3 (300 millions)

Languages : 440 to 680

subgroup of the Niger-Congo phylum

time depth: 5,000–4,000 bp

African hunter gatherers



900,000 Pygmies of central Africa

100,000 eastern African HGs

100,000 Bushmen/San, southern Africa

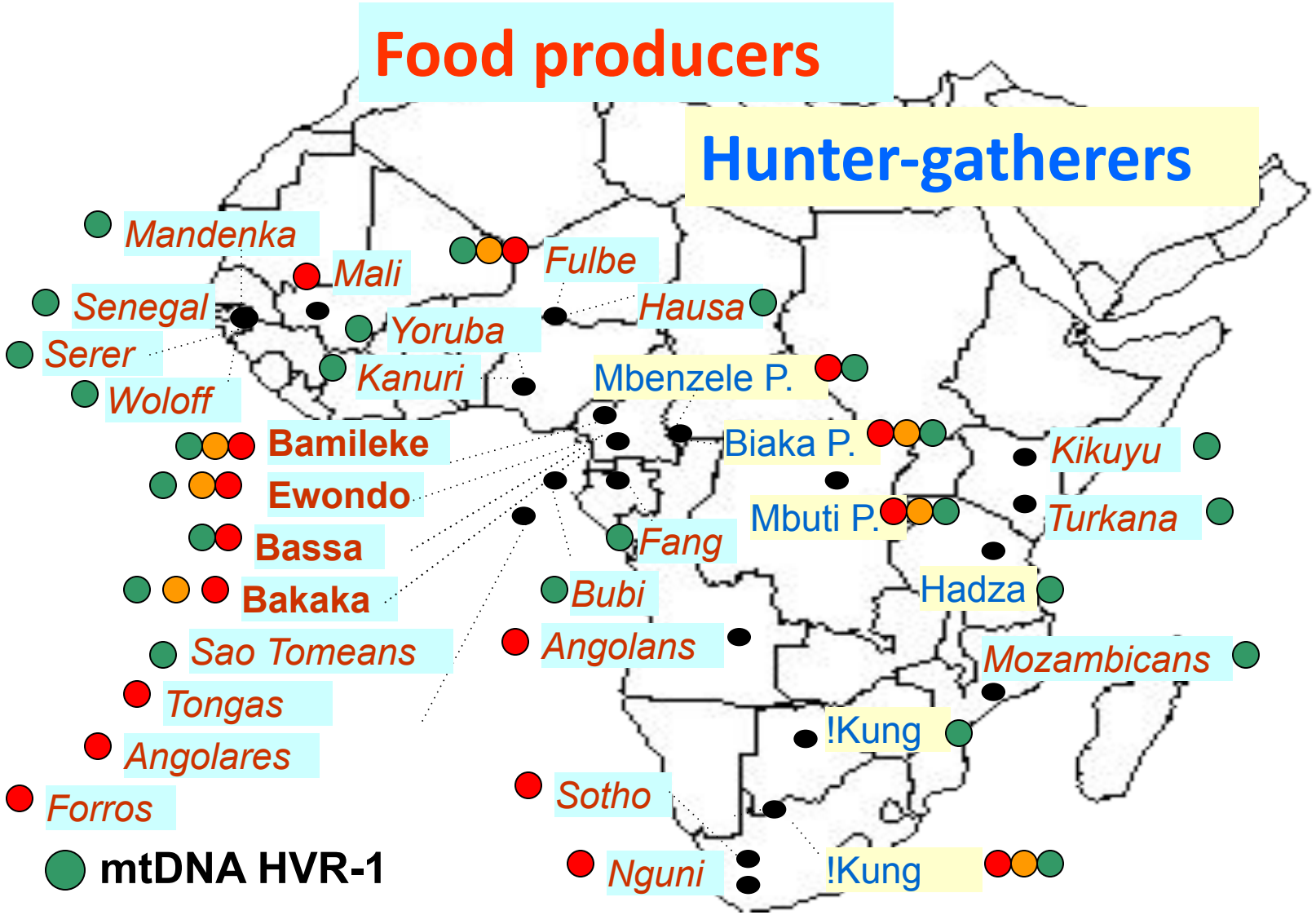
low population densities

Egalitarian societies; No property rights

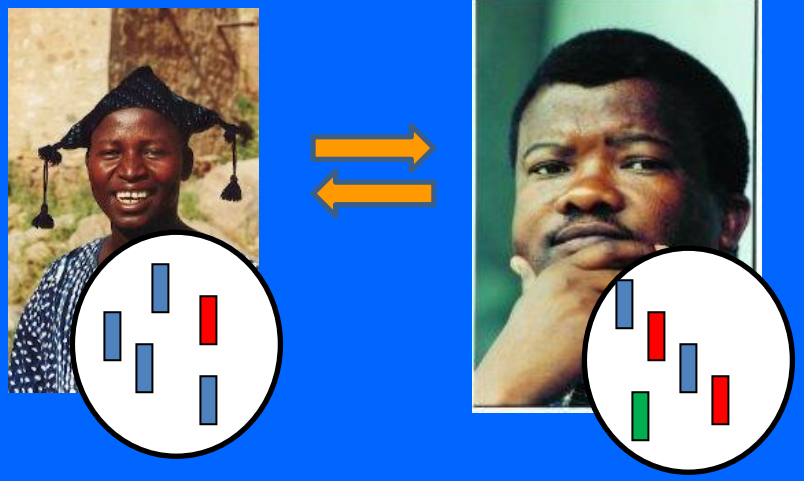
absence of food surplus

co-operation and sharing a general tenet

sustainable ecologically, needs are modest

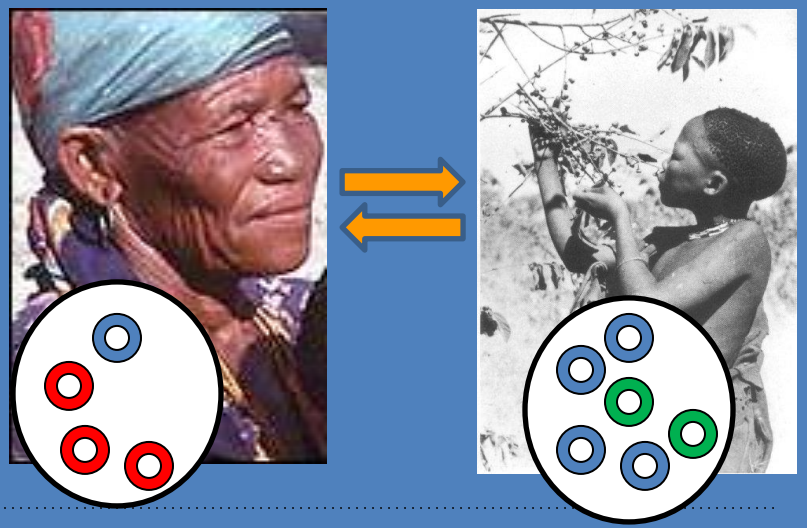

Food producers
Hunter-gatherers
Mandenka
Mali
Fulbe
Senegal
Yoruba
Hausa
Serer
Kanuri
Mbenzele P.
Wolloff
Bamileke
Biaka P.
Kikuyu
Ewondo
Mbuti P.
Turkana
Bassa
Fang
Bakaka
Bubi
Hadza
Sao Tomeans
Angolans
Mozambicans
Tongas
!Kung
Angolares
Sotho
Forros
Nguni
!Kung
mtDNA HVR-1
Y-chromosome SNPs
Y-chromosome microsatellites (DYS19, 389I, 390, 391, 392 and 393)

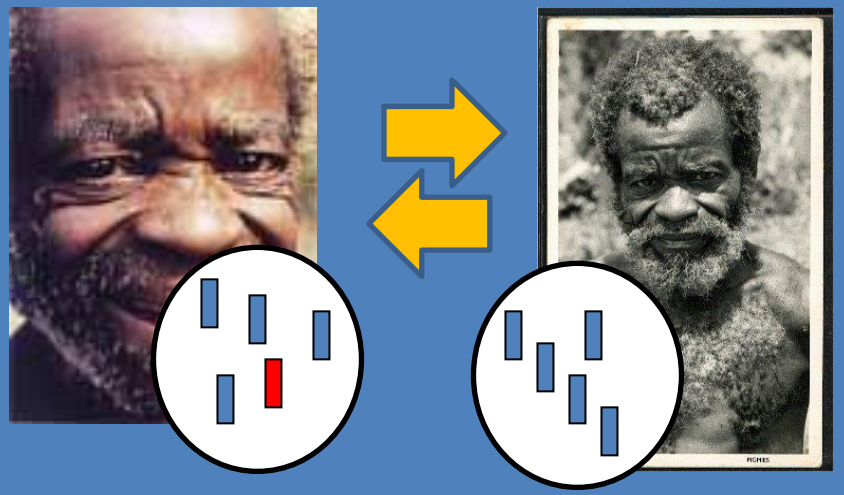
Food producers



Nv female/ Nv male = 1.97

Hunter-gatherers



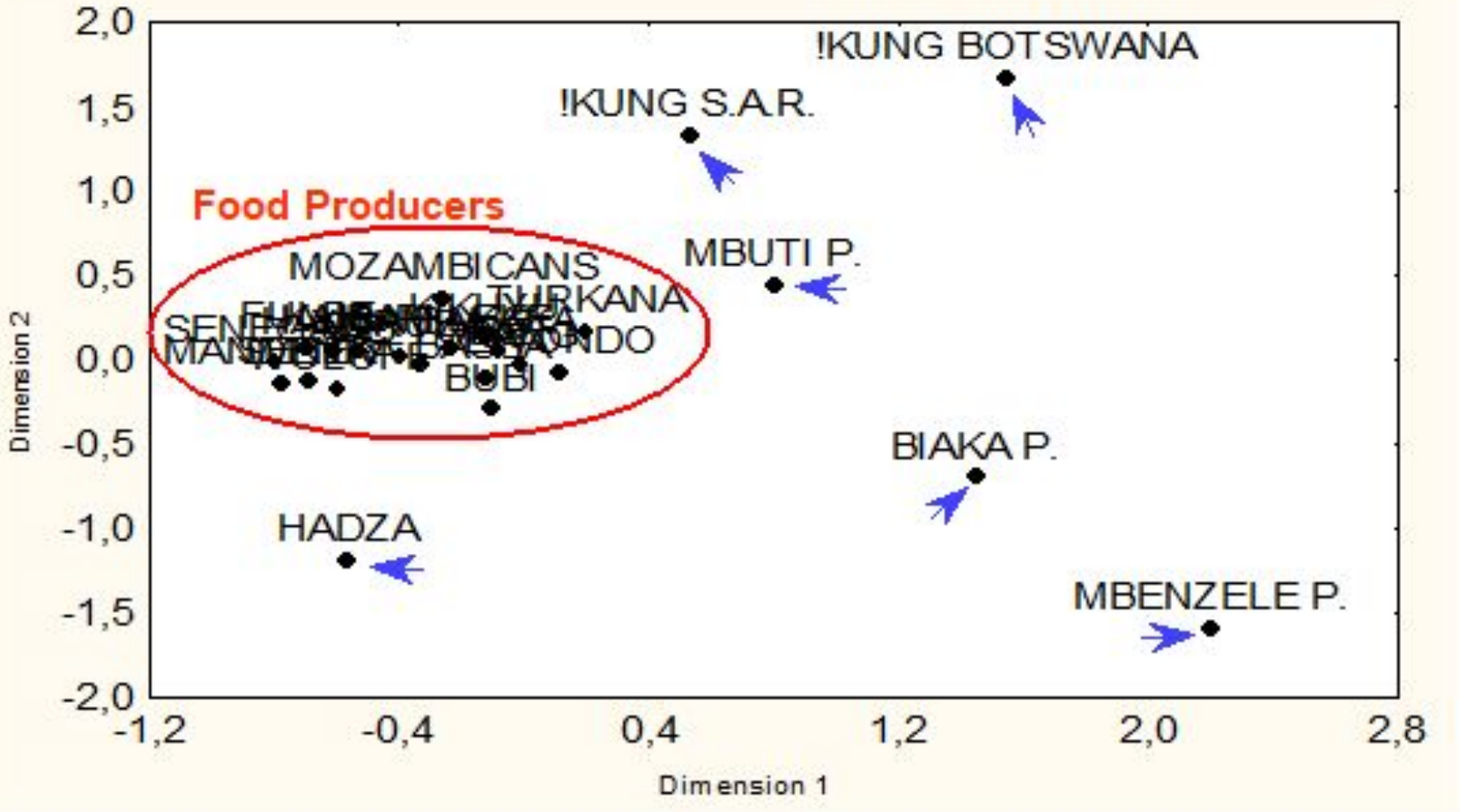


Nv female/ Nv male = 0.04

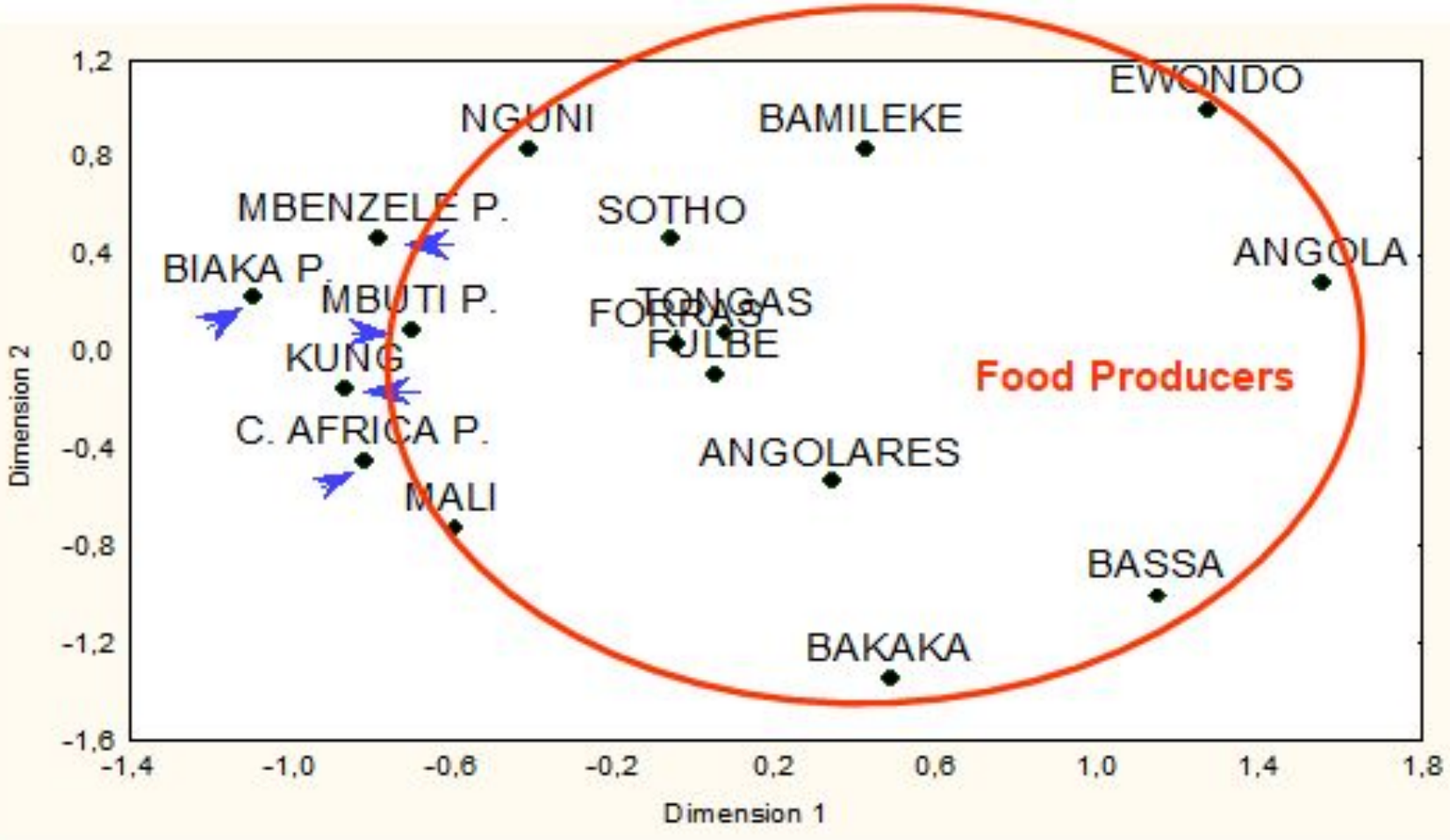
Complete dataset

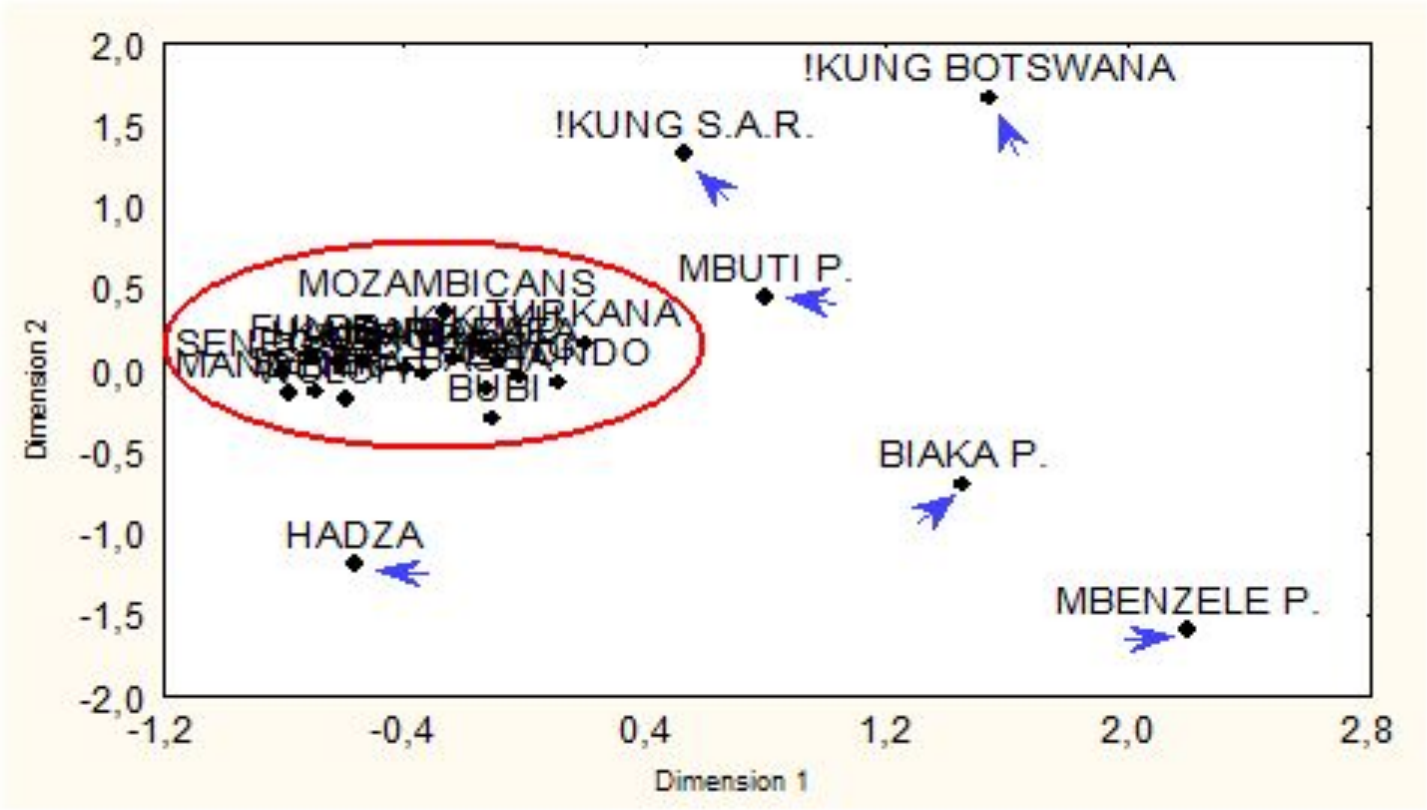
mtDNA HVR-1, Fst genetic distance

multidimensional scaling (stress 0.12)



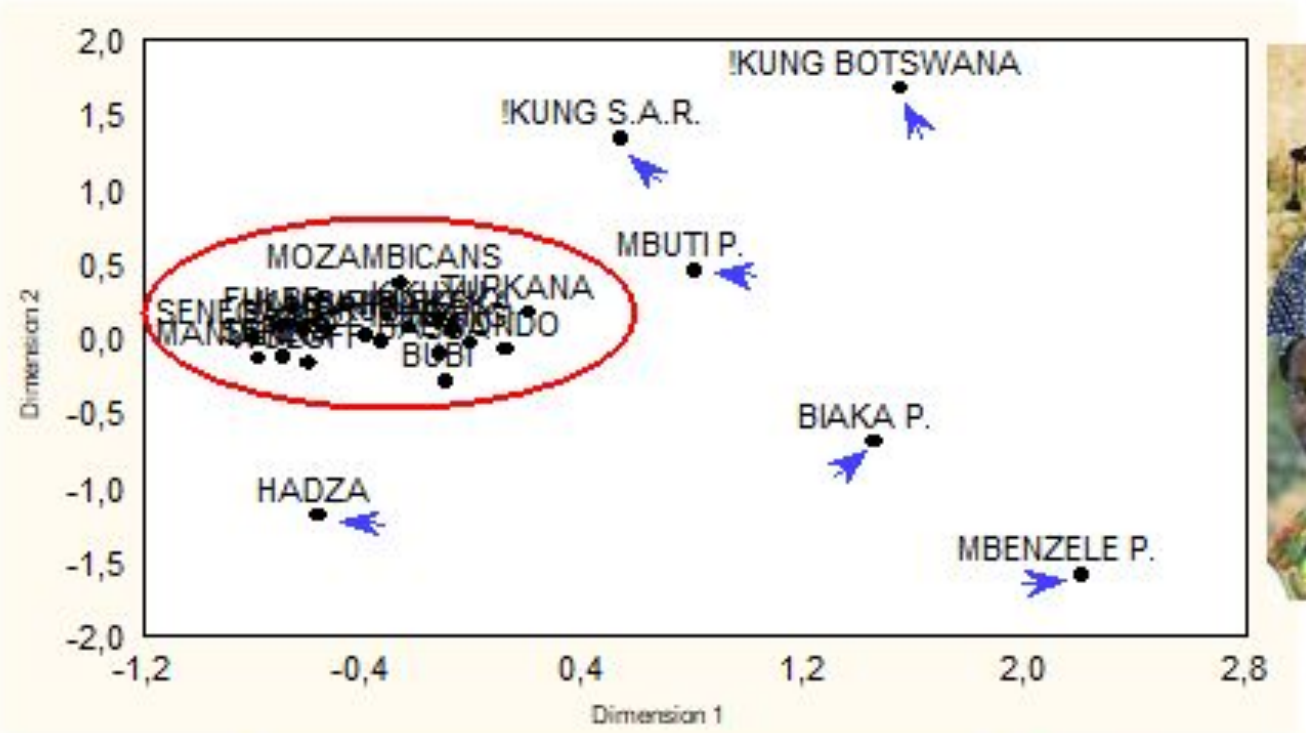
Complete dataset
Y-chromosome microsatellites
Fst genetic distance, MDS (stress 0.13)





$$F_{st} = \frac{V_p}{\bar{p}(1-\bar{p})}$$

$$Nv = (1/Fst)-1$$

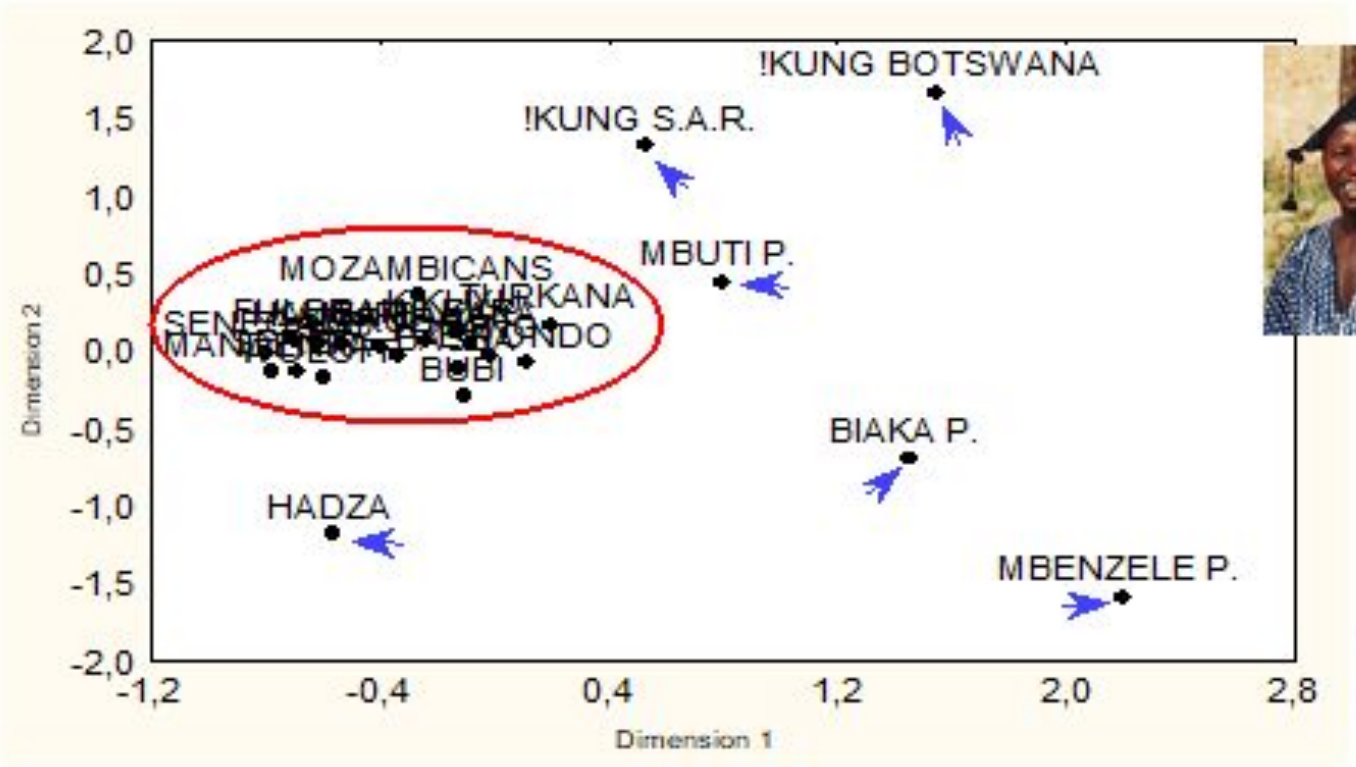


F_{st} = 0.167 intero dataset

Nv = (1/0.167)-1

Nv = 6,00-1

Nv = 5,00



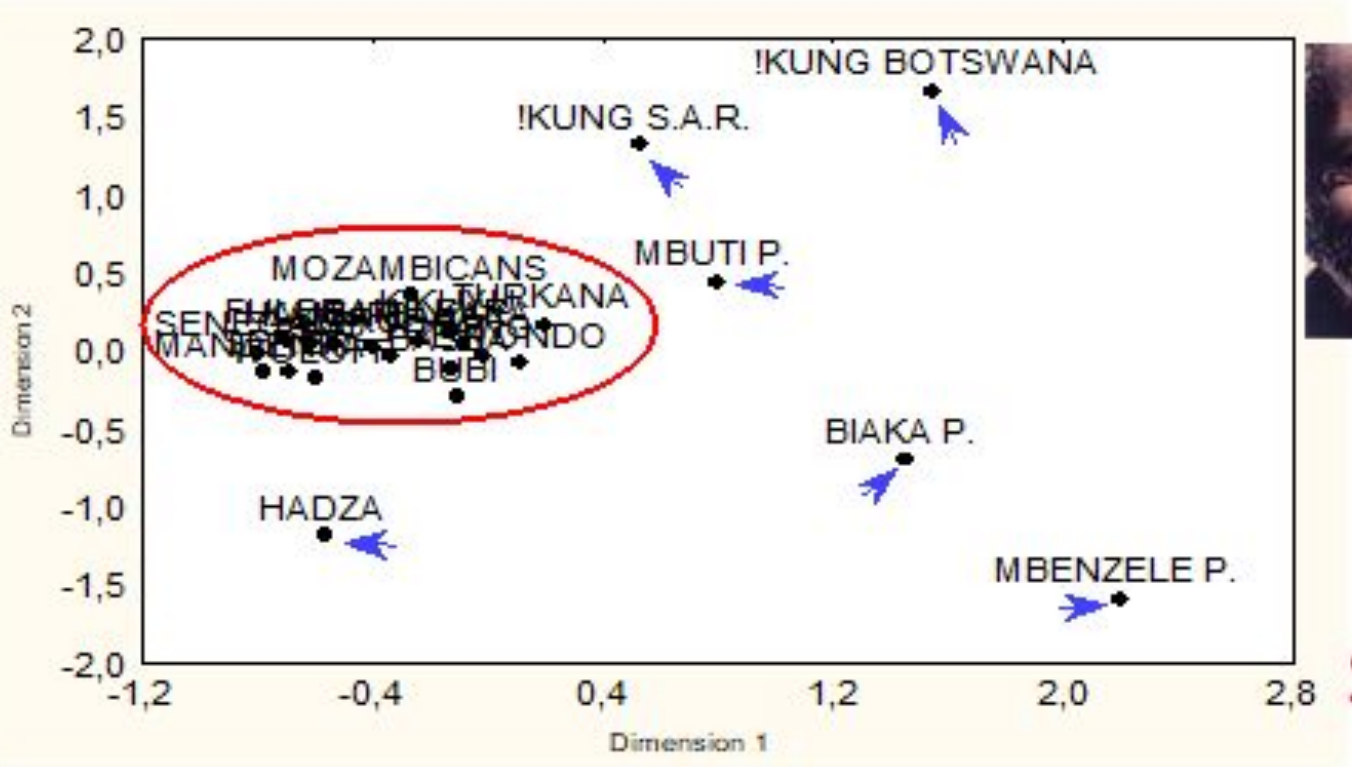
solo FP

F_{st} = 0.049

Nv = (1/0.049)-1

Nv = 20,02-1

Nv = 19,02



solo HGs

F_{st} = 0.461

Nv = (1/0.461)-1

Nv = 2.17-1

Nv = 2,17

Complete dataset
mtDNA HVR-1
Y-chromosome microsatellites

	Nni*		
	mtDNA	**Y**	**mtDNA/Y**
all	5.00	9.52	**0.52**
FPs	19.02	9.65	**1.97**
HGPs	1.17	32.54	**0.04**

*estimated according to the island model

as Nv = (1/Fst)-1

Paired population dataset
mtDNA HVR-1
Y-chromosome microsatellites

	Nni*	
	mtDNA/Y	jacknife
all populations	0:49	-
FPs (5 pops)	**6:38**	4.75-10.64
HGPs (4pops)	**0:03**	0.02-0.05

*estimated according to the island model

as Nv = (1/Fst)-1

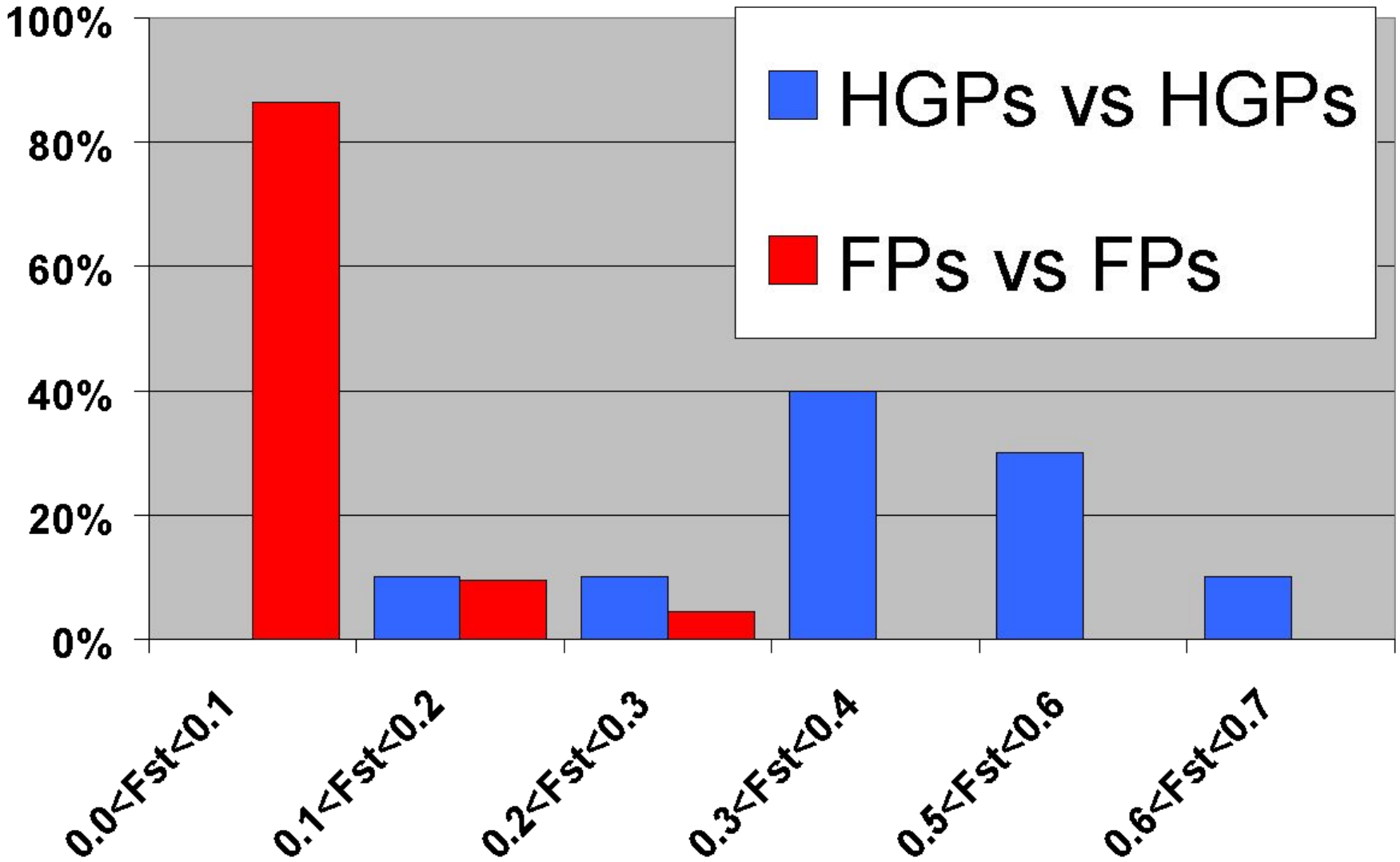
Paired population dataset
mtDNA and Y-chromosome haplogroups

	Nni*	
	mtDNA/Y	
all populations	8.08	3.99
FPs (8 pops)	6.24	[1.59-7.51]
HGPs (3 pops)	0.26	[0.12-0.35]

*estimated according to the island model

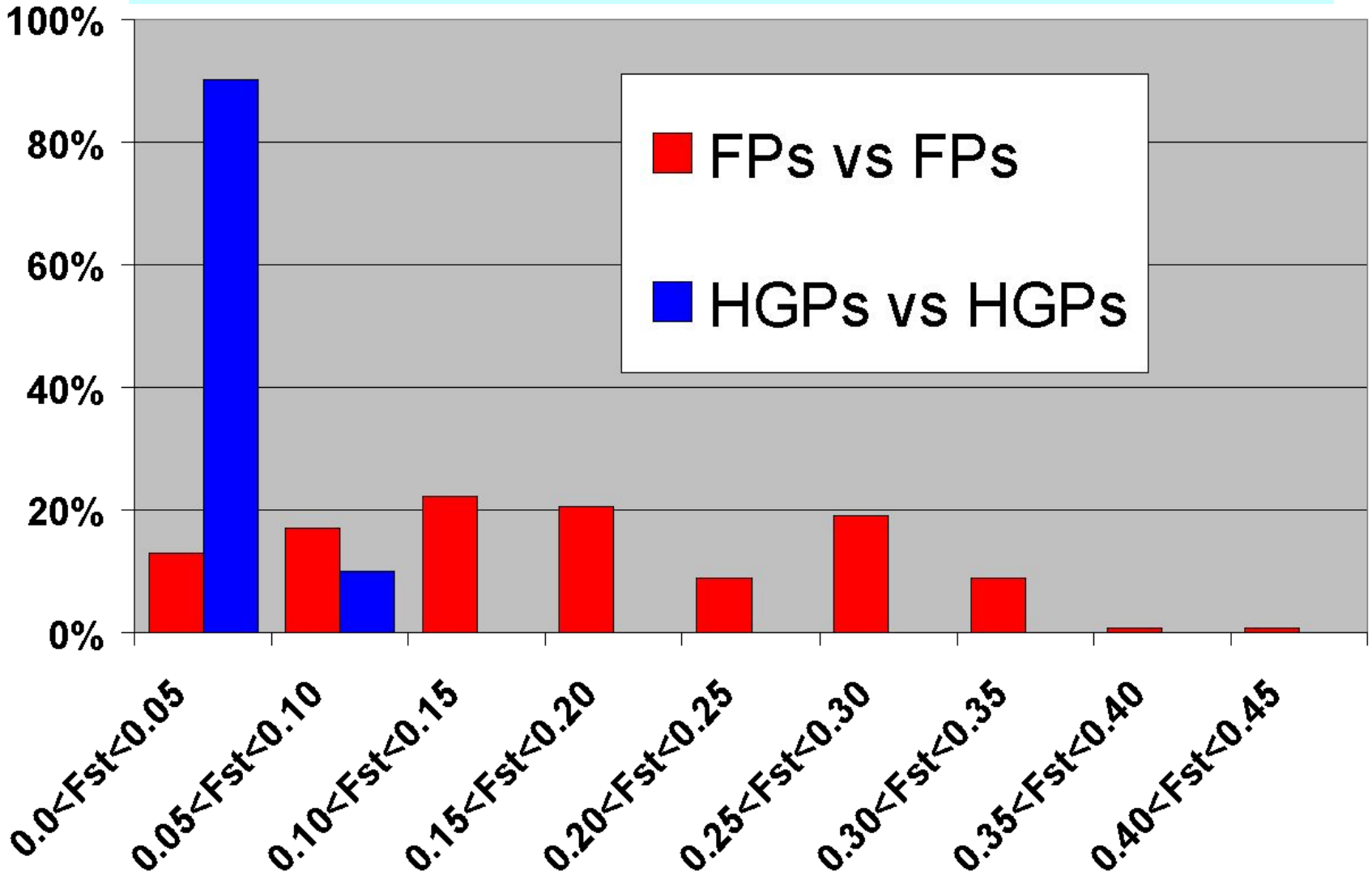
as Nv = (1/Fst)-1

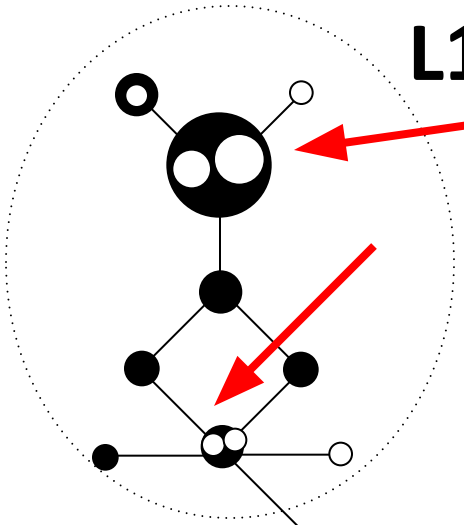
mtDNA, Fst genetic distances



Y-chromosomal microsatellites

Fst genetic distances





L1c1a1

female-biased gene flow from HGPs to FPs

- Mbenzele P
- Biaka P.
- Ewondo
- Other Africans

16214

16223

16234

16249

*

	Mbenzele P.	Biaka P.	Ewondo
frequency	0.596 (0.046)	0.294 (0.078)	0.132 (0.033)
haplotype diversity	0.642 (0.087)	0.700 (0.218)	0.523 (0.208)
Mean number P.C.	1.274 (0.821)	2.200 (1.450)	1.142 (0.835)

Destro-Bisol et al. 2004, American Naturalist

(*): 16129, 16187, 16189, 16223, 16274, 16278, 16293, 16294, 16311, 16360

Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes

Chiara Batini,†[1,2] Joao Lopes,[3] Doron M. Behar,[4] Francesc Calafell,[1,5] Lynn B. Jorde,[6] Lolke van der Veen,[7] Lluis Quintana-Murci,[8] Gabriella Spedini,[2,9] Giovanni Destro-Bisol,[2,9] and David Comas*[,1,5]

Mol. Biol. Evol. 28(2):1099–1110. 2011

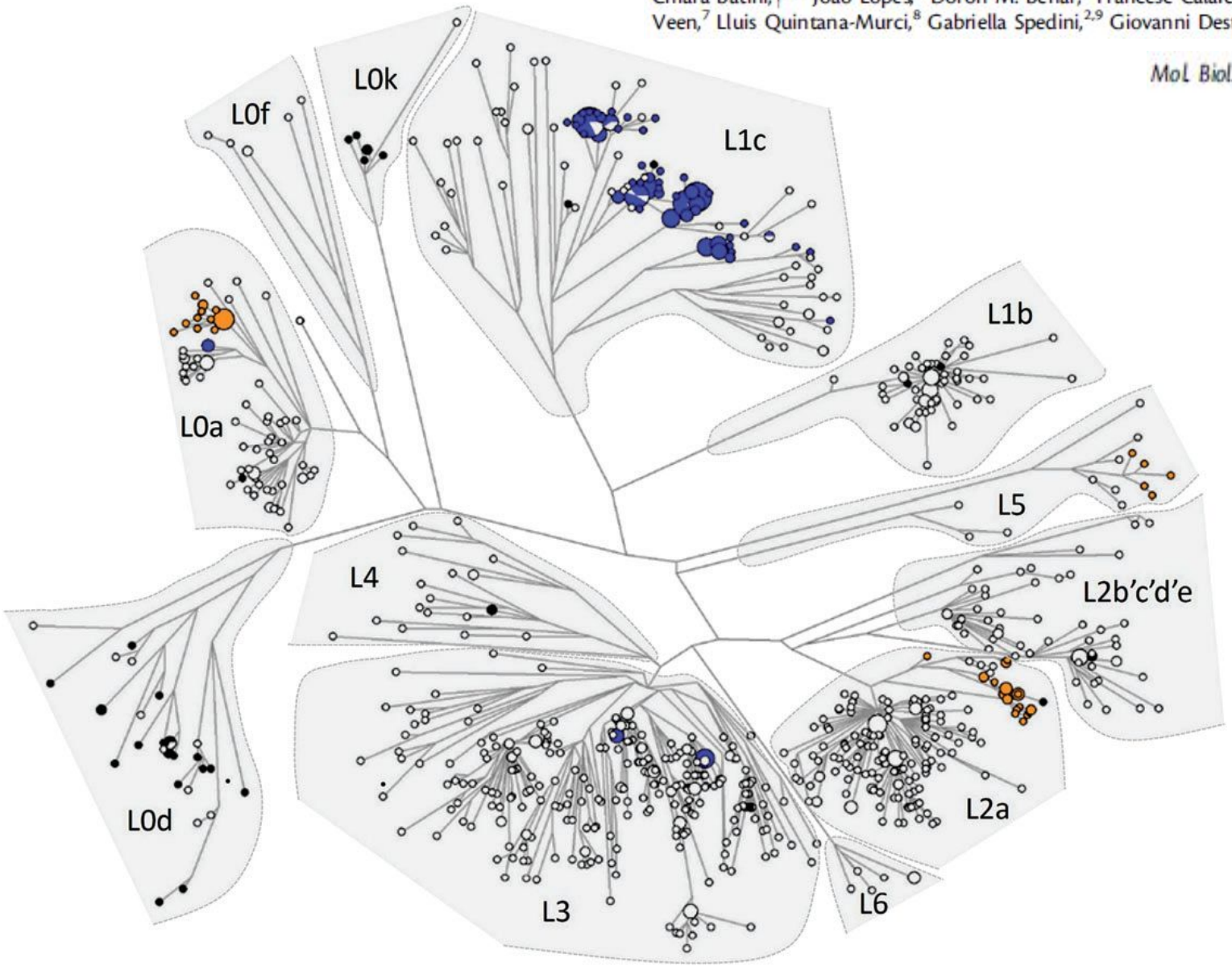
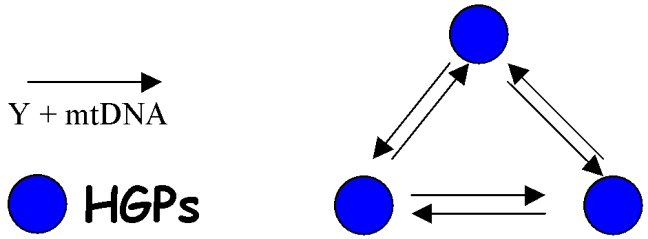


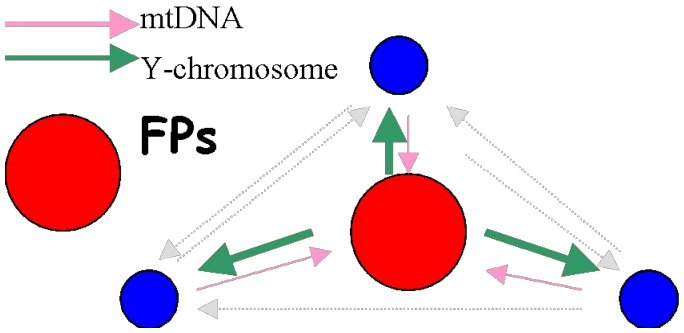
FIG. 3. Median-joining network of 973 mitochondrial-coding region sequences (corresponding to positions 435–16023). Blue, Western Pygmies; orange, Eastern Pygmies; black, Khoisan speakers.

1. before the Bantu expansion



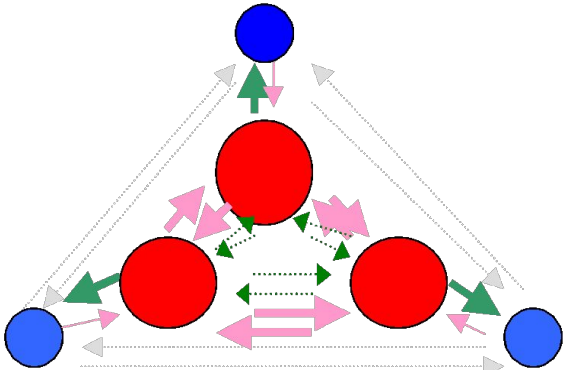
Demographic pattern	Effect on genetic structure
high migration rate among HGP subpopulations	maintenance of large effective size in HGPs

2. with the Bantu expansion

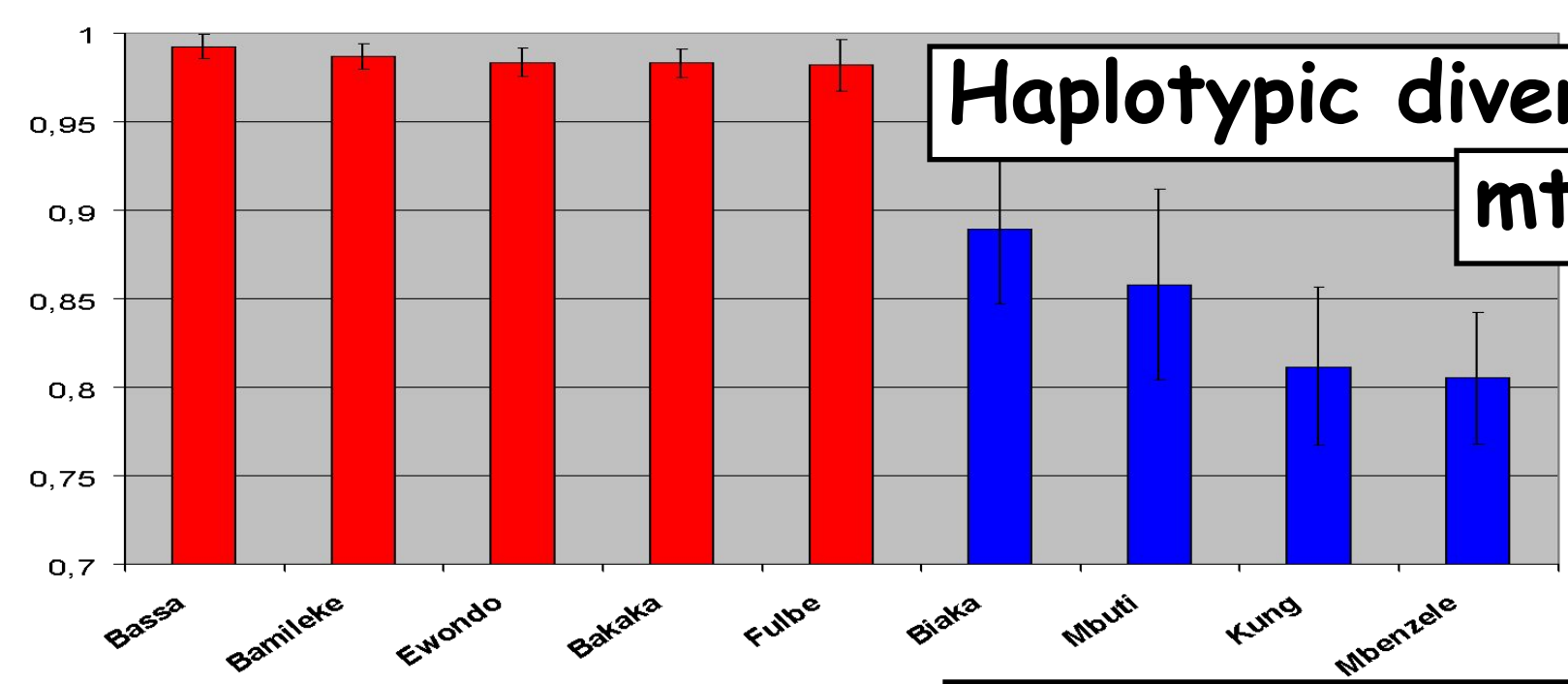


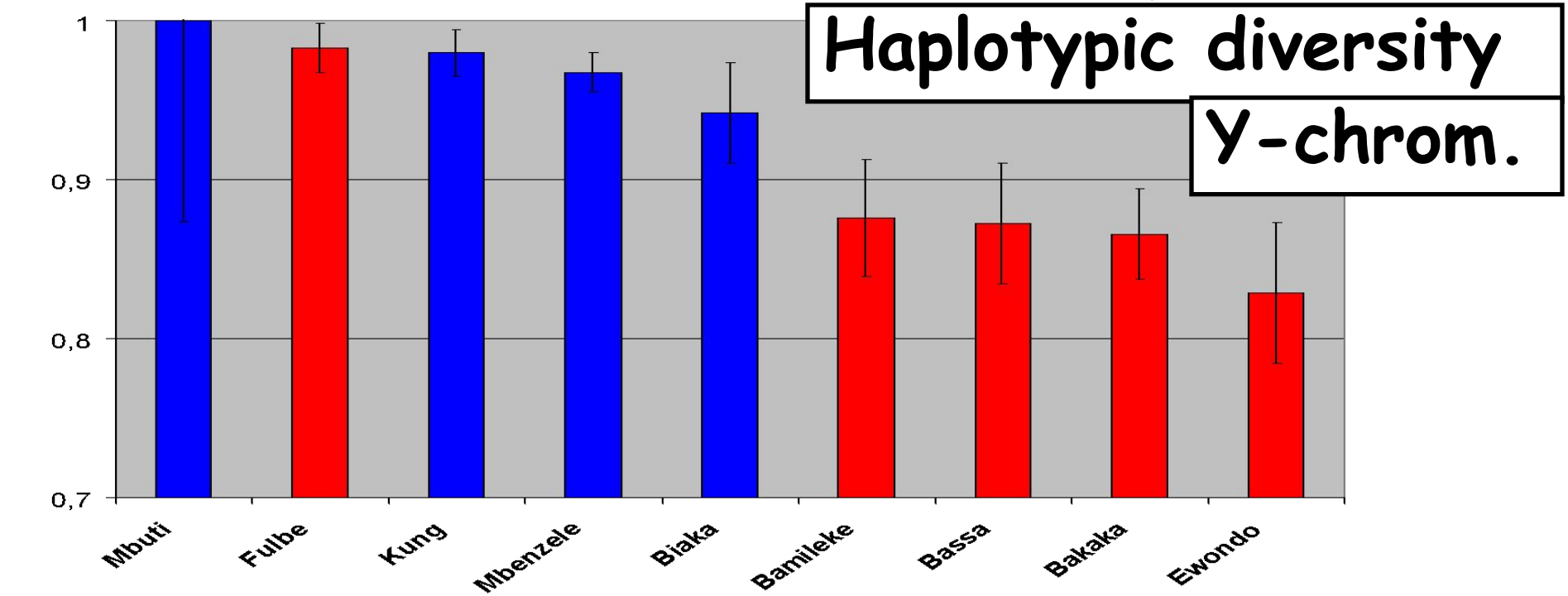
decreased migration rate among HG subpopulations	increased differentiation among HG subpopulations
male biased gene flow from FPs to HGPs	introgression of FP Y-chrom. into HGPs
female biased gene flow from HGPs to FPs	introgression of HGP mtDNAs into FPs

3. during the Bantu expansion



higher female than male migration rate in FPs *via* patrilocality; higher level of polygyny among FPs	increased ratio of Y-chromosome to mtDNA Fst among FPs


Haplotypic diversity
mtDNA
1
0,95
0,9
0,85
0,8
0,75
0,7
Bassa
Bamileke
Ewondo
Bakaka
Fulbe
Biaka
Mbuti
Kung
Mbenzele


Haplotypic diversity
Y-chrom.
1
0,9
0,8
0,7
Mbuti
Fulbe
Kung
Mbenzele
Biaka
Bamileke
Bassa
Bakaka
Ewondo

male-biased gene flow from FPs to HGPs

	Y-chromosome haplogroups		
	22	24	tot
Biaka P.	40%	25%	**65%**
Mbuti P.	33%	8%	**41%**
Kung	16%	23%	**39%**

data from Cruciani et al. 2002, AJHG 70:1197

Open problems

- Low number of population sampled and their heterogenetity
- Long distance among populations
- Environmental differences
- Simplistic model
- Fst and level of variation

Mol. Biol. Evol. 30(4):918–937 doi:10.1093/molbev/mss328 Advance Access publication January 7, 2013

Sociocultural Behavior, Sex-Biased Admixture, and Effective Population Sizes in Central African Pygmies and Non-Pygmies

Paul Verdu,*[1,2] Noémie S.A. Becker,[2] Alain Froment,[3] Myriam Georges,[2] Viola Grugni,[4] Lluis Quintana-Murci,[5] Jean-Marie Hombert,[6] Lolke Van der Veen,[6] Sylvie Le Bomin,[2] Serge Bahuchet,[2] Evelyne Heyer,[2] and Frédéric Austerlitz[2]

However, the study by Destro-Bisol et al. (2004) could not provide estimates of male and female effective population size. It could also not estimate levels of gender-related migrations within groups of hunter–gatherer and farmer populations, respectively, as well as the variable levels of sex-specific admixture from the farmers into the hunter–gatherers. The availability of large sets of nuclear markers, computer-intensive clustering algorithms and coalescent-based methods now allow obtaining such demographic parameter estimates (for a review see Nielsen and Beaumont 2009). Furthermore, Destro-Bisol et al. (2004)'s hunter–gatherer sample pooled six populations (Hadza, Hadzabe, !Kung, and three Central African Pygmy populations). Therefore, they could not study the variability of gender-specific processes across the various hunter–gatherer populations. Such variability is however expected, especially across the numerous populations categorized as Pygmies throughout Central Africa. Indeed, more than 20 hunter–gatherer Central African populations are historically designated as "Pygmies" and have been shown to be culturally very heterogeneous and genetically widely differentiated while likely sharing a common origin (Hewlett 1996; Verdu et al. 2009; Patin et al. 2009; Tishkoff et al. 2009; Batini, Lopes, et al. 2011; Bahuchet 2012; Jarvis et al. 2012; Lachance et al. 2012). Furthermore, these studies, mainly based on autosomal and mtDNA data, highlighted that Pygmy populations had variable levels of admixture from neighboring non-Pygmy farmer populations.

Re-evaluating a model of gender biased gene flow between sub-Saharan hunter gatherers and farmers

Paolo Anagnostou[1,2], Cinzia Battaggia[1], Marco Capocasa[2,3], Ilaria Boschi[4], Francesca Brisighelli[4], Gabriella Spedini and Giovanni Destro-Bisol*[1,2]

Human Biology, 2013

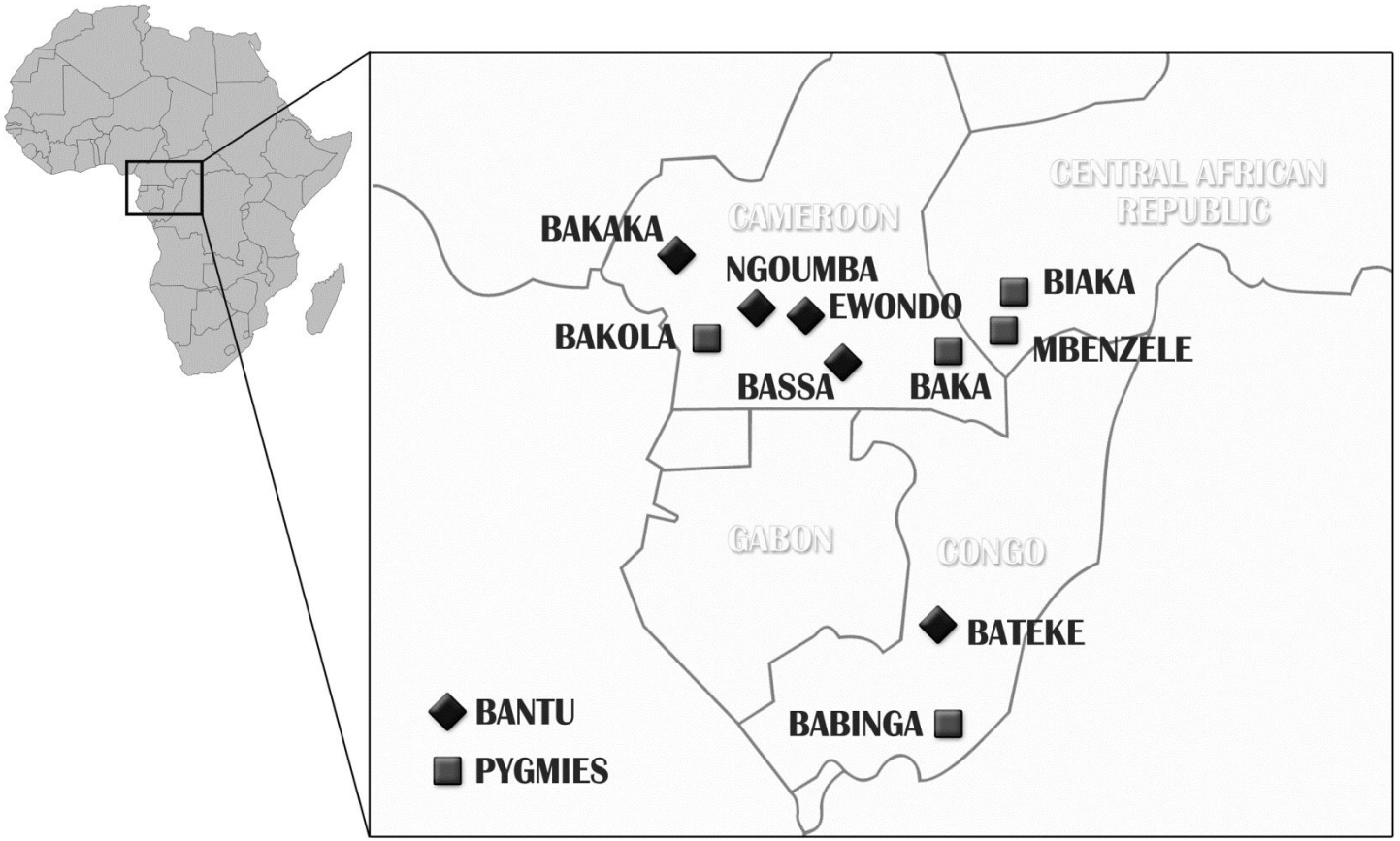


Table 2. Estimates of F_{st} and Nm of Pygmies and farmers from Central Western Africa.
Note - The Nm values were calculated by application of an island model of migration. Interval estimates of Nm parameter (in brackets) were obtained by use of the jacknife procedure.

	Fst **mtDNA**	Fst **Y chromosome**	$N_f m_f / N_m m_m$
Pygmies	0.269 (0.085 – 0.335)	0.054 (0.043 – 0.064)	0.154 (0.119 – 0.544)
Bantu farmers	0.019 (0.009 – 0.027)	0.116 (0.095 – 0.140)	6.759 (5.259 – 17.362)

Table 3. mtDNA and Y chromosome diversity in the populations analyzed.

	mtDNA		Y chromosome	
	HD (s.d.)	MNPD (s.d.)	HD (s.d.)	MNPD (s.d.)
Pygmies				
Babinga	0.693 (0.074)	4.615 (2.308)	0.884 (0.042)	3.463 (1.844)
Baka	0.871 (0.028)	6.260 (3.022)	0.953 (0.011)	3.259 (1.701)
Bakola	0.257 (0.048)	3.590 (1.854)	0.943 (0.022)	3.187 (1.690)
Biaka	0.889 (0.043)	8.117 (3.963)	0.938 (0.024)	3.617 (1.884)
Mbenzele	0.798 (0.035)	4.904 (2.425)	0.966 (0.015)	3.246 (1.704)
Bantu farmers				
Bakaka	0.983 (0.008)	9.781 (4.554)	0.866 (0.028)	1.899 (1.102)
Bassa	0.993 (0.006)	9.433 (4.408)	0.872 (0.038)	2.293 (1.279)
Bateke N.	0.944 (0.017)	6.621 (3.179)	0.909 (0.038)	2.409 (1.337)
Ewondo	0.983 (0.008)	10.165 (4.715)	0.829 (0.044)	1.742 (1.035)
Ngoumba	0.990 (0.007)	8.973 (4.213)	0.959 (0.016)	2.979 (1.594)

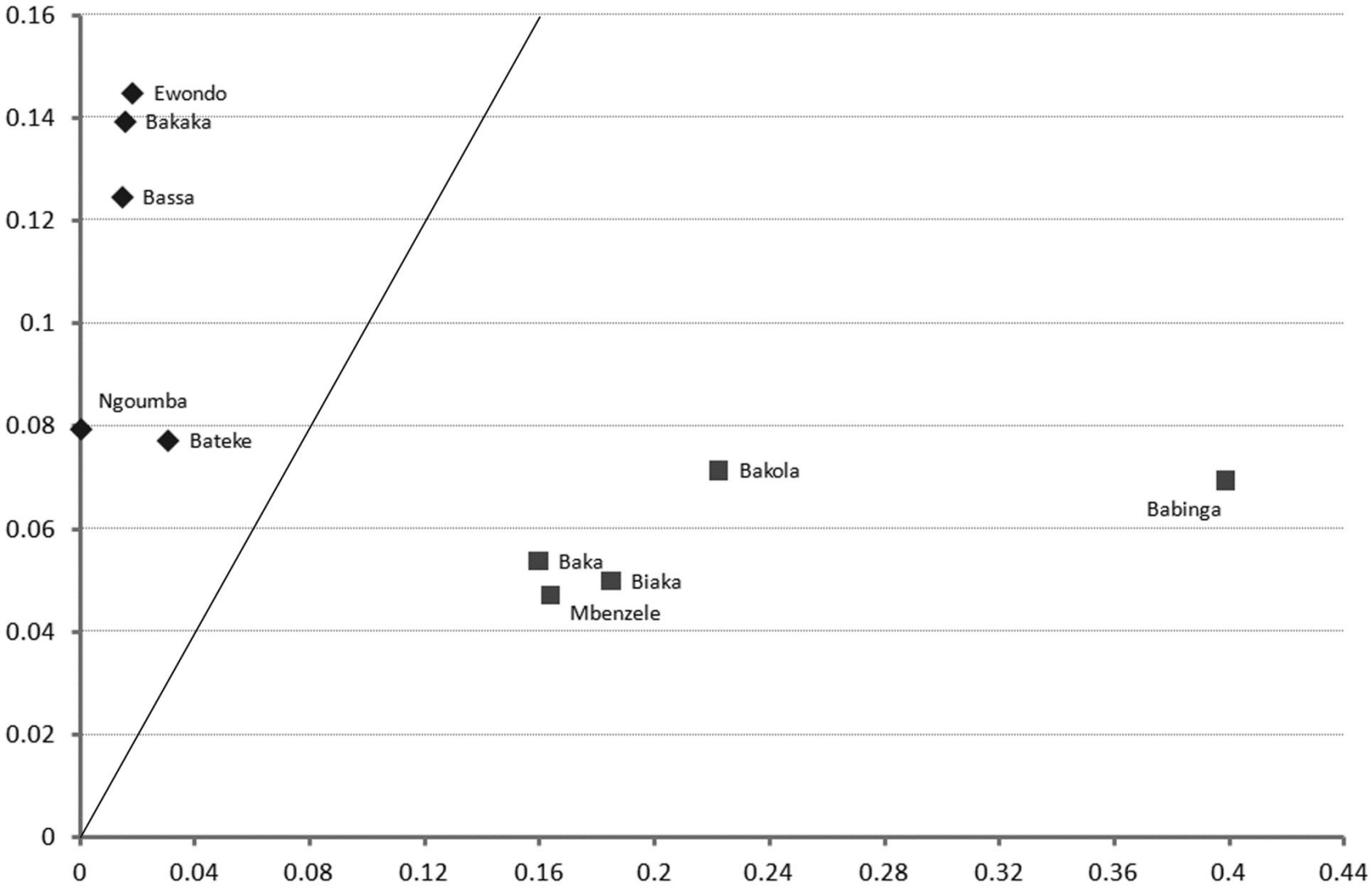


Figure 3. Average F_{st} for mtDNA (x axis) and Y chromosome (Y axis) between Pygmies (squares) and between farmers (diamonds). The black line indicates a 1:1 ratio.