

Genetic Distance

Introduction

Genetic distance is a core concept in biological anthropology and population genetics. It refers to the **degree of genetic divergence** between populations or species. Simply put, it quantifies how genetically “far apart” two populations are, offering insights into their **evolutionary history, migration patterns, and biological relatedness**.

This concept is central to understanding **microevolutionary changes** and has vast applications in anthropological genetics, including **forensic anthropology, medical genetics, and human population studies**. It also helps reconstruct the **tree of human evolution**, and its relevance has been highlighted repeatedly, particularly in questions related to **population differentiation, isolation, and genetic drift**.

I. What is Genetic Distance?

Genetic distance measures the **accumulated genetic differences** between populations. It is not simply a reflection of visible traits, but a mathematical estimate of how much the **allele frequencies**—i.e., the variations of genes—differ between groups.

- **Small genetic distance** implies **recent common ancestry or ongoing gene flow**.
- **Large genetic distance** suggests **ancient divergence, isolation, or long-term evolutionary change**.

II. Mechanisms Behind Genetic Distance

1. Mutation

Random **mutations** introduce new genetic variants. Over time, isolated populations accumulate **different mutations**, increasing genetic distance.

2. Genetic Drift

This refers to **random changes** in allele frequencies, especially in small populations. Drift causes populations to become genetically distinct even in the absence of selection.

Example: The small population of the Finnish Saami has unique mtDNA haplotypes due to drift, studied at the University of Helsinki.

3. Natural Selection

Environmental pressures lead to **population-specific adaptations**. These differences in adaptive alleles can contribute to measurable genetic distance.

Example: Tibetans exhibit specific EPAS1 gene adaptations to high-altitude life, not found in lowland Han Chinese. Research from the University of Chicago and Fudan University supports this divergence.

4. Geographic or Cultural Isolation

Physical barriers (mountains, oceans) or cultural practices (endogamy, caste) restrict **gene flow**, accelerating divergence.

5. Time Since Divergence

The longer two populations have been **genetically separated**, the more different they become. This concept underpins **molecular clock** models in evolutionary biology.

III. Measuring Genetic Distance

1. Allele Frequency-Based Measures

Populations are compared using **statistical estimates** of their allele frequencies.

Nei's Genetic Distance (1972)

Nei's Genetic Distance is one of the most widely used and influential measures of genetic divergence between populations. Introduced by **Masatoshi Nei** in 1972, it provides a **quantitative estimate of how genetically distinct two populations are**. This metric is rooted in population genetics and evolutionary biology and is essential for understanding patterns of divergence, migration, and evolutionary history.

Basic Concept

Nei's method is based on **allele frequencies**—that is, how common different versions of a gene are within each population. The idea is simple: **the more alleles two populations share, the more genetically similar they are**. If they share fewer alleles, they are more genetically distant.

Thus, **Nei's genetic distance** reflects the **accumulated genetic mutations** and **allele frequency changes** that have occurred since two populations diverged from a common ancestor.

Formula

The original formula is derived from comparing gene identity between two populations:

$$D = -\ln I$$

Where:

- D = Nei's genetic distance
- I = Genetic identity between the two populations (a value from 0 to 1)

The **genetic identity (I)** is calculated from the **average proportion of shared alleles** at multiple loci between two populations.

- If $I = 1$, then $D = 0$: the populations are genetically identical.
- If I is small, then D is larger: the populations are genetically divergent.

Mechanism: How Nei's Distance Works

Nei's method operates on the principle that **as two populations accumulate genetic differences**, their **allele frequencies shift** due to **mutation, genetic drift, and selection**. By mathematically analyzing these shifts, the method:

1. **Calculates average gene identity (I)** across many loci.
2. **Takes the natural logarithm of the inverse of identity** to measure genetic distance (D).
3. **Generates a value that reflects the evolutionary time and change** since the two populations split.

This approach assumes that differences in allele frequencies arise from **neutral mutations** accumulating over time—making it ideal for modeling **evolutionary distance**.

Interpretation of Nei's Distance

Genetic Distance (D)	Interpretation
0	Populations are genetically identical
< 0.05	Very closely related
0.05–0.15	Moderately related
0.15–0.30	Distantly related
> 0.30	Highly divergent populations (deep evolutionary split)

Examples

African vs. European Populations

When comparing sub-Saharan African populations to European populations, Nei's genetic distance values typically range from **0.07 to 0.12**, reflecting a **moderate genetic divergence**.

- **Explanation:** This supports the “**Out of Africa**” **model** of modern human origins. As humans migrated out of Africa and settled in new environments, genetic drift and selection led to changes in allele frequencies, increasing genetic distance from the ancestral African populations.

Cavalli-Sforza's Chord Distance (1967)

Cavalli-Sforza's Chord Distance, introduced by **Luigi Luca Cavalli-Sforza and A. Edwards** in 1967, is a statistical method to measure the **genetic distance between populations** based on **allele frequencies**. Unlike Nei's method, which is logarithmic, the chord distance is a **geometric model** — visualizing populations as points on the surface of a hypersphere in multidimensional genetic space.

This method is especially useful in constructing **phylogenetic trees** and **mapping evolutionary relationships** among human groups and other species.

Basic Concept

The Chord Distance interprets each population as a vector in multi-allelic space. The **shortest path (or chord)** connecting these points represents their **genetic difference**.

The method focuses on:

- **Angular differences** between population vectors.
- **Length of the chord** that connects them — hence the name.

It assumes that **allele frequency differences** arise mainly due to **genetic drift**, especially in the absence of natural selection.

Formula

Let p_{ij} and q_{ij} be the frequencies of allele j at locus i in two populations.

The **Cavalli-Sforza Chord Distance** (D_C) is:

$$D_C = \sqrt{2 \left(1 - \sum \sqrt{p_j \cdot q_j} \right)}$$

Where:

- p_j and q_j are the frequencies of allele j in population 1 and 2 respectively.
- The **sum** runs over all alleles and loci.

The value lies between 0 (identical populations) and a maximum (fully divergent).

Mechanism: How Chord Distance Works

1. **Each allele frequency set** is treated as a coordinate in multidimensional space.
2. The populations are positioned as vectors on the surface of a **unit hypersphere**.
3. The **chord (shortest straight-line distance)** between these points reflects the degree of genetic divergence.
4. **Genetic drift** is assumed to be the main force causing these differences.

Unlike Nei's distance, which can model evolutionary time, Chord Distance is **non-parametric**—focusing purely on observed differences, not on mutation models.

Interpretation

Chord Distance	Genetic Relationship
0.0	Populations are genetically identical
< 0.05	Closely related
0.05–0.15	Moderately related
> 0.15	Distantly related

This scale helps anthropologists build **genetic trees** and identify clusters of related populations.

Examples

European Populations

Cavalli-Sforza's original studies in the 1960s–80s used Chord Distance to study **European genetic structure**. For example:

- **Italians** and **Greeks** showed low Chord Distance (~ 0.02 – 0.04), indicating **close genetic ties**, consistent with Mediterranean gene flow and shared ancestry.
- **Basques** had higher distance values from surrounding groups, supporting their **status as a genetic isolate**.

Fixation Index (FST)

The **Fixation Index (FST)** is a key concept in population genetics used to measure the **genetic divergence** between subpopulations. It tells us **how much of the total genetic variation in a species is due to differences between populations**, rather than variation within each population.

Basic Concept

When a species is divided into different populations, these populations may become genetically different over time due to processes like **genetic drift, natural selection, mutation, or geographic isolation**. F_{ST} helps us **quantify** this difference using genetic data.

Formula

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Where:

- H_T (**Total heterozygosity**) is the expected genetic diversity if the entire species were one large, randomly mating population.
- H_S (**Subpopulation heterozygosity**) is the average genetic diversity **within** each subpopulation.

Mechanism: How F_{ST} Works

- If all subpopulations have **similar allele frequencies**, then $H_S \approx H_T$, and F_{ST} **will be close to 0**, indicating little or no genetic differentiation.
- If subpopulations have **very different allele frequencies**, then H_S is **much smaller than H_T** , and F_{ST} **increases**, showing greater genetic divergence.

Thus, F_{ST} is a ratio that reflects the **proportion of genetic diversity due to population structure**.

Interpretation of F_{ST} Values

F_{ST} Value	Interpretation
0	No genetic differentiation
0.0–0.05	Little differentiation

0.05–0.15	Moderate differentiation
0.15–0.25	Great differentiation
>0.25	Very great genetic differentiation

Examples

Genetic Differentiation between Europeans and East Asians

Studies using genome-wide data have found that **FST values between European and East Asian populations** typically range between **0.10 and 0.15**. This indicates a **moderate level of genetic differentiation**.

Explanation: The two populations diverged tens of thousands of years ago after the Out-of-Africa migration and adapted to different environments. Over time, genetic drift and selection caused allele frequencies to shift differently in each region.

2. Molecular Tools Used

1. Microsatellites (STRs – Short Tandem Repeats)

Microsatellites are small sections of DNA where a short sequence of bases (like "AG" or "CAT") repeats many times in a row. Different people or populations often have different numbers of these repeats.

Why are they useful?

- These repeats **change frequently**, which makes them good for studying **recent changes** in populations.
- They are inherited, so comparing them between populations shows how similar or different those populations are.

How they help with genetic distance:

- Scientists count the different versions (called **alleles**) of microsatellites in two populations.
- The more differences in allele patterns, the **greater the genetic distance**.

Example:

In some South American tribes, researchers used microsatellites to show that tribes living in isolated forest areas had developed **genetic differences** due to being separated for generations.

2. SNPs (Single Nucleotide Polymorphisms)

SNPs (pronounced "snips") are places in the DNA where a single base (A, T, G, or C) is different between people. For example, one person might have A, while another has G at the same spot.

Why are they useful?

- SNPs are **very common** in the human genome — they occur about every 300 bases.
- They are **very stable**, so they are useful for comparing both recent and ancient differences.

How they help with genetic distance:

- Scientists can study **thousands or even millions of SNPs** across the genome.
- They compare the frequency of these SNPs between populations.
- Differences in these frequencies show **how genetically distant** the populations are.

Example:

SNP studies have shown that people in northern India and southern India have different genetic patterns, showing historical **separation and limited gene flow** between the groups.

3. mtDNA (Mitochondrial DNA)

Mitochondrial DNA (mtDNA) is a special kind of DNA found in the mitochondria, which are the energy-producing parts of the cell. Unlike regular DNA, mtDNA is passed **only from mother to child**.

Why is it useful?

- It does not mix with the father's DNA, so it **traces maternal lineage**.
- It changes slowly and **accumulates mutations over time**, like a ticking genetic clock.

How it helps with genetic distance:

- Scientists compare the mtDNA sequences of different populations.
- If two groups have very different mtDNA, it means their **maternal lines separated long ago**.
- Smaller differences mean **more recent separation**.

Example:

mtDNA studies show that Europeans and East Asians have different maternal ancestry, and their **genetic distance** tells us they diverged over **40,000 years ago** during early human migrations.

4. Y-Chromosome Markers

The **Y chromosome** is passed only from father to son. It is used to trace **paternal lineages**, just like mtDNA is used for maternal lines.

Why is it useful?

- The Y chromosome does not recombine (mix) much, so it keeps a clear **record of male ancestry**.
- It has both **SNPs and STRs**, allowing for study of both ancient and recent changes.

How it helps with genetic distance:

- Populations are grouped into Y-DNA **haplogroups** (paternal line groups).
- Comparing the frequency of these haplogroups tells us how far apart the **male ancestry** of two populations is.
- Greater differences mean **greater genetic distance**.

Example:

In India, Y-chromosome studies show that some tribal groups have different paternal lineages compared to caste populations, pointing to **distinct historical origins and limited intermarriage**.

PYQ Insights

1. Definition and Conceptual Understanding

Gist: MCQs have asked to identify that genetic distance quantifies the genetic divergence between populations.

2. Common Measures of Genetic Distance

Gist: Questions commonly ask about **Nei's Genetic Distance (1972)** and **Cavalli-Sforza's Chord Distance (1967)** — including their developers or application scope.

3. Founder of the Concept

Gist: MCQs have included names like **Masatoshi Nei**, asking to match scientists to their models.

4. Molecular Tools Used in Genetic Distance Studies

Gist: Options include **SNPs, STRs (microsatellites), mtDNA, and Y-chromosome markers**, with one or more correct choices.

5. Applications in Human Populations

Gist: MCQs test if you can identify applications like **evolutionary history, population divergence, or biological relationships**.

6. Genetic Distance and Evolution

Gist: Questions test understanding that it increases with **isolation, lack of gene flow, or genetic drift over time**.

7. Match-the-Pair Questions

Gist: For example:

- Nei's Distance → Based on allele frequencies
- Cavalli-Sforza → Phylogenetic tree construction
- F_{ST} → Fixation index measure