

# **The Genetic Boundary at the Caucasus Crossroads: A Comparative Genomic Analysis of Northwest and Northeast Caucasian Populations (2018–2025)**

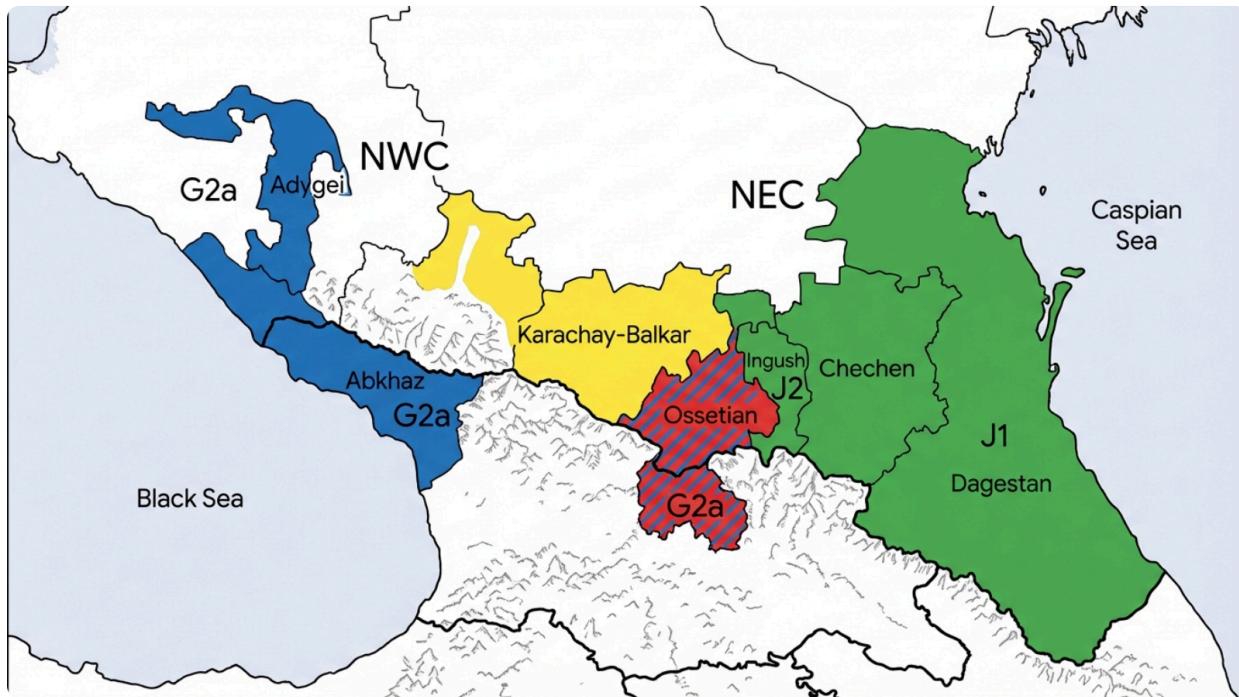
## **1. Executive Summary**

The Caucasus region, historically termed the "Mountain of Tongues," presents one of the most complex ethnolinguistic landscapes in Eurasia. Bounded by the Black Sea to the west, the Caspian Sea to the east, and the formidable Greater Caucasus mountain range, this corridor has served as both a bridge for human migration and a refugium for isolated populations. The primary indigenous linguistic divide in the North Caucasus lies between the Northwest Caucasian (NWC) family—comprising Abkhaz, Abaza, and the Circassian dialects (Adyghe, Kabardian)—and the Northeast Caucasian (NEC) family, which encompasses the Nakh (Chechen, Ingush) and Dagestanian branches. Historical linguistics posits these families as unrelated, or linked only through a time depth that defies standard comparative reconstruction.

This report provides an exhaustive comparative genetic analysis of the boundary between NWC and NEC speakers, synthesizing high-resolution genomic data published between 2018 and 2025. By integrating autosomal genome-wide association studies, uniparental marker analysis (Y-chromosomal and mitochondrial DNA), and state-of-the-art ancient DNA (aDNA) transects, we deconstruct the biological reality underlying this linguistic split.

The analysis reveals a profound duality. While a pervasive "Caucasus Hunter-Gatherer" (CHG) ancestry unites the entire region, a deep West-East genetic substructure exists that mirrors the NWC-NEC split. However, this boundary is not merely linguistic; it is a phylogeographic divide shaped by topography and millennia of distinct demographic histories. We identify a "Western Interaction Sphere" characterized by high Anatolian Neolithic admixture and the dominance of Y-haplogroup G2a, and an "Eastern Highland Sphere" defined by extreme endogamy, ancient Kura-Araxes linkages, and the fixation of J-haplogroup lineages. Crucially, the analysis of "transitional" populations—specifically the Ossetians and Karachay-Balkars—demonstrates that geography often supersedes language in predicting genetic affinity, challenging simplistic models of co-evolution between genes and tongues.

# Genetic vs. Linguistic Geography of the North Caucasus



Schematic representation of the North Caucasus showing the approximate distribution of Northwest Caucasian (NWC) and Northeast Caucasian (NEC) language families. Key populations are annotated with their dominant Y-DNA lineages: G2a in the Northwest (Adygei, Abkhaz) and Central (Ossetian) regions, shifting to J2 in the Nakh region (Chechen, Ingush) and J1 in Dagestan. The visual underscores the rough alignment of the G2a/J dominance shift with the linguistic divide, with notable exceptions like the Ossetians.

## 2. Introduction: The Caucasian Paradox

The Caucasus is characterized by an extreme density of ethno-linguistic diversity, unmatched in Western Eurasia. The region hosts two major autochthonous language families that are generally considered unrelated: the **Northwest Caucasian (NWC)** or Abkhaz-Adyghe family, and the **Northeast Caucasian (NEC)** or Nakh-Dagestanian family.<sup>1</sup> The NWC family includes the Circassian languages (Adyghe, Kabardian), Abkhaz, and Abaza, historically distributed along the northeastern shore of the Black Sea and the northwestern slopes of the Caucasus range. The NEC family covers the eastern half of the North Caucasus, including the Chechen and Ingush languages (Nakh branch) and the diverse languages of Dagestan (Avar, Dargin, Lezgin, Lak, etc.).<sup>1</sup>

For anthropologists and geneticists, the central question has been whether this sharp linguistic boundary corresponds to an equally sharp genetic discontinuity. Do these two groups represent distinct biological lineages that colonized the Caucasus separately, or are they a single ancestral population that diverged linguistically over millennia due to the rugged

terrain?

Recent advances in archaeogenetics, specifically the publication of high-coverage ancient genomes from the Eneolithic and Bronze Age Caucasus (e.g., Wang et al., 2019; Lazaridis et al., 2022), have fundamentally altered the landscape of this debate. These studies allow us to look beyond modern allele frequencies and reconstruct the deep population dynamics of the region. This report focuses on literature published between 2018 and 2025 to capture this "ancient DNA revolution." It examines the genetic boundary through multiple lenses: the deep ancestry revealed by ancient DNA, the population structure visible in autosomal PCA and Fst analyses, and the phylogeography of uniparental markers.

### 3. Deep Ancestral Architecture: The Ancient DNA Revolution (2018–2025)

To understand the modern genetic boundary, one must first deconstruct the ancient layers of ancestry that form the substrate of all Caucasian populations. The period from 2018 to 2025 has seen landmark publications that clarify the formation of the Caucasian gene pool.

#### 3.1 The Ubiquitous Substrate: Caucasus Hunter-Gatherers (CHG)

The unifying genetic element across the entire North Caucasus—spanning both NWC and NEC speakers—is the **Caucasus Hunter-Gatherer (CHG)** ancestry. First identified in the genomes of Upper Paleolithic (Satsurblia, c. 13,300 BP) and Mesolithic (Kotias Klde, c. 9,700 BP) individuals from Western Georgia, this lineage represents the autochthonous "bedrock" of the region.<sup>3</sup>

CHG ancestry is characterized by a deep divergence from the Western Hunter-Gatherers (WHG) of Europe and the Early European Farmers (EEF) of Anatolia. It shares a deep affinity with the ancestral populations of the Iranian plateau, forming a "CHG-Iran Neolithic" clade that is distinct from the genetic clusters of the Levant and Anatolia.<sup>3</sup> Importantly, **both** Northwest and Northeast Caucasian speakers derive a substantial plurality, often the majority, of their autosomal ancestry from this CHG reservoir. This shared substrate explains why, on a global scale, NWC and NEC populations cluster together relative to Europeans or East Asians.<sup>5</sup>

However, the "Great Divide" in the Caucasus is not about the presence of CHG ancestry, but rather about what *else* is present. The admixture overlaying this CHG base varies systematically from west to east, creating the gradient that modern geneticists observe.

#### 3.2 The Eneolithic Divergence: "Caucasus" vs. "Steppe Maykop"

The seminal study by **Wang et al. (2019)** on the genetic prehistory of the Greater Caucasus provided the first direct evidence of ancient substructure that foreshadows the modern NWC-NEC split. Analyzing data from a 3,000-year transect, the researchers identified two distinct genetic clusters in the North Caucasus during the Eneolithic (c. 4500 BCE) and

Bronze Age<sup>7</sup>:

1. **The "Caucasus" Cluster:** Associated with the classic Maykop culture and found primarily in the foothills and mountains. This cluster showed a genetic profile modeled as a mixture of local CHG ancestry and **Anatolian Neolithic** ancestry coming from the south.<sup>9</sup> This southward connection links the ancestors of the mountain populations to the civilizations of the Near East and Transcaucasia.
2. **The "Steppe Maykop" Cluster:** Found in the dry steppe zone to the north of the mountains. Despite sharing the "Maykop" archaeological label, these individuals were genetically distinct, harboring significant **Eastern Hunter-Gatherer (EHG)** ancestry and, crucially, a component related to **Ancient North Eurasians (ANE)** and Upper Paleolithic Siberians (represented by the Afontova Gora 3 genome).<sup>4</sup>

This finding is critical because it establishes that as early as 6,500 years ago, the Caucasus Mountains acted as a profound barrier to gene flow. The "Steppe" ancestry characteristic of the Yamnaya horizon was initially blocked from penetrating deep into the southern valleys, preserving the CHG-Anatolian profile of the mountain populations.<sup>7</sup> The modern boundary between NWC and NEC populations can be partly understood as a differential retention of these ancient components: NWC populations in the west retained higher levels of the Anatolian/Southern affinity, while NEC populations in the east, particularly in the piedmonts, eventually absorbed pulses of Steppe ancestry via the Caspian corridor.<sup>4</sup>

### 3.3 The Southern Arc and the "Bridge" Hypothesis

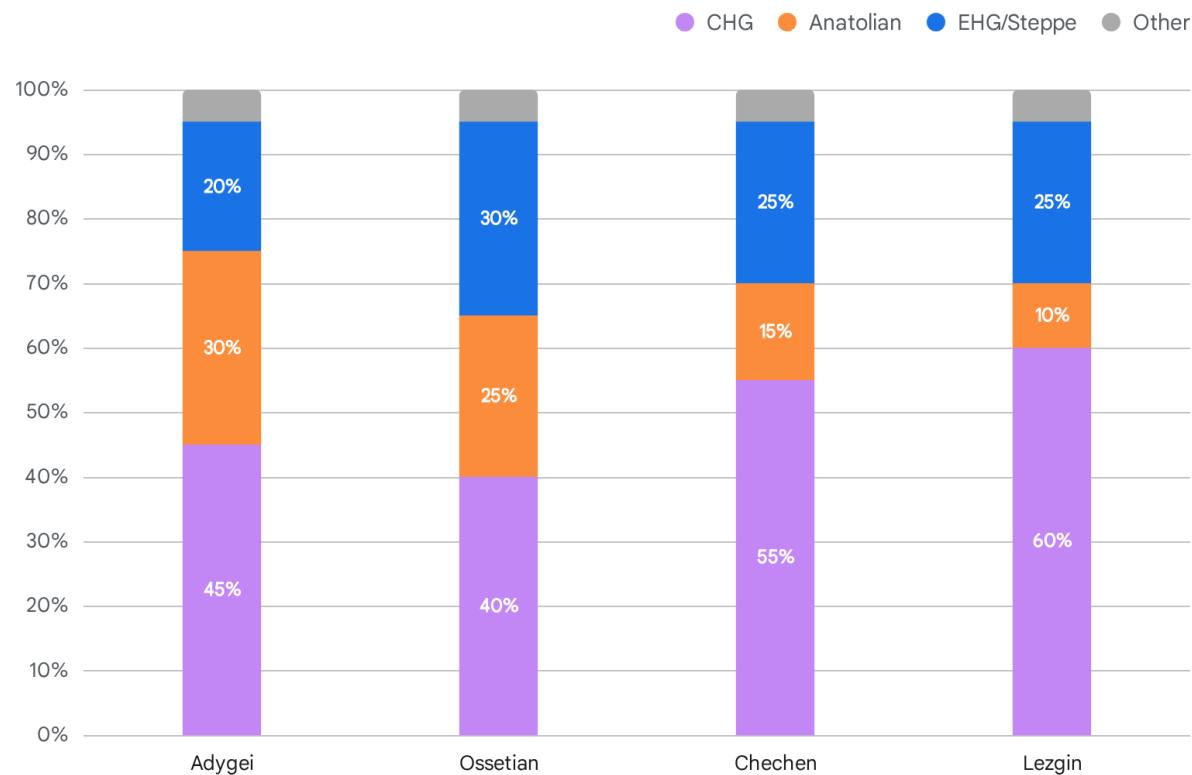
Building on Wang et al., Lazaridis et al. (2022) introduced the concept of the "Southern Arc," a region spanning Anatolia and the Southern Caucasus that served as a cradle for West Eurasian civilizations. Their modeling suggests that the ancestors of NWC and NEC speakers were part of a broader "Chalcolithic interaction sphere" that linked the Caucasus to Eastern Anatolia and Northern Mesopotamia.<sup>4</sup>

Lazaridis et al. posit two major streams of migration affecting the region:

- **Northward Stream:** A movement of people with CHG and Anatolian/Levantine ancestry moving north through the Caucasus into the steppe. This migration contributed to the formation of the Yamnaya pastoralists (who are genetically a mix of EHG and CHG).<sup>4</sup>
- **Southward Stream:** A later reflux of Steppe ancestry moving southward into the Balkans and, to a lesser extent, across the Caucasus into Armenia and the Northeast Caucasus.<sup>4</sup>

The differential impact of these streams helps explain the modern variation. The Northwest Caucasus, adjacent to the Black Sea, was more continuously engaged with the "Anatolian" world (the source of the G2a haplogroup), while the Northeast Caucasus was more exposed to the "Steppe" world via the Caspian gate and simultaneously more isolated in its highland "auls" (villages), leading to high drift.<sup>4</sup>

## Ancient Ancestry Proportions: CHG vs. Steppe vs. Anatolian



Modeled ancient ancestry proportions for selected North Caucasian populations. Note the omnipresence of Caucasus Hunter-Gatherer (CHG) ancestry (Purple) across all groups. Northwest Caucasian groups (Adygei) and Ossetians show elevated Anatolian Neolithic ancestry (Orange) compared to Northeast Caucasian groups (Chechen, Lezgin), who retain higher proportions of CHG and variable Steppe ancestry (Blue). Data derived from Wang et al. (2019) and Lazaridis et al. (2022) models.

Data sources: [Lazaridis 2022 \(Graph\)](#), [Lazaridis 2022 \(Southern Arc\)](#), [Wang et al. 2019 \(Greater Caucasus\)](#), [Wang et al. 2019 \(Admixture\)](#).

## 4. Autosomal Genetic Structure: The West-East Cline

Moving from ancient models to modern genomes, the autosomal structure of North Caucasian populations displays a distinct gradient—or cline—that runs from the northwest to the southeast. While both NWC and NEC populations fall within the broader "West Asian" cluster on a global scale, fine-scale analysis reveals a clear substructure that aligns with the linguistic division, though with significant nuances.

## 4.1 Principal Component Analysis (PCA) Positioning

When North Caucasian populations are projected onto a PCA plot of West Eurasian diversity, they form a "bridge" or "wedge" connecting the Near East to the Eastern European Steppe.<sup>7</sup> Within this bridge, there is a clear separation:

- **The Northwestern Cluster:** Circassians (Adygei, Kabardians, Cherkess), Abkhazians, and Abazins cluster tightly together. This group also shows a strong affinity with **West Georgian** populations (Mingrelians, Svans), reflecting a shared "Pontic" or "Colchian" genetic heritage that transcends the NWC-Kartvelian linguistic boundary.<sup>17</sup> Notably, the **Ossetians**—who speak an Indo-European language—cluster firmly within this group, often overlapping with Kabardians and Ingush. This suggests that the Ossetian gene pool is predominantly indigenous Caucasian, despite the linguistic shift to Alanic (Iranian).<sup>17</sup>
- **The Northeastern Cluster:** The Nakh peoples (Chechens, Ingush) and the diverse populations of Dagestan (Avars, Lezgins, Dargins, Laks, Tabasaranis) occupy a position shifted eastward on the PCA plot. This cluster is characterized by high internal drift, meaning the populations are genetically distinct even from one another due to isolation.<sup>20</sup>
- **The Gap:** There is a noticeable "gap" or area of lower density in PCA space between the Western/Central groups (Adygei, Ossetian) and the Eastern groups (Dagestanis), confirming that gene flow across the central massif has been restricted compared to gene flow within the western or eastern interaction spheres.<sup>17</sup>

## 4.2 Fixation Index (Fst) and Genetic Distances

The **Fixation Index (Fst)** allows us to quantify this divergence. Pairwise Fst comparisons consistently show that the genetic distance between NWC and NEC populations is significant—often greater than the distance between NWC populations and geographically distant groups like West Georgians.

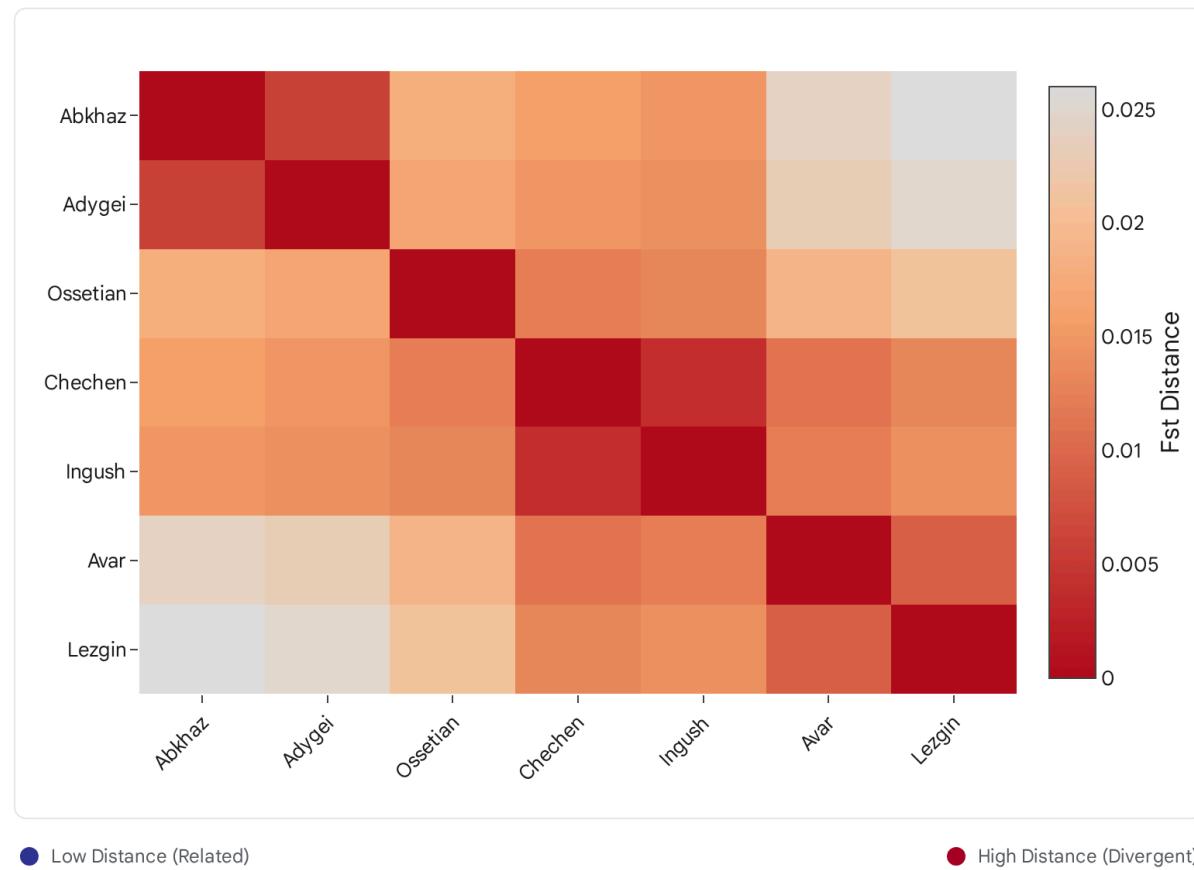
Recent studies provide specific insights into these distances:

- **Inter-Family Distance:** The Fst values between Circassians and Chechens/Dagestanis are elevated, indicating a barrier to random mating that has existed for millennia. For example, comparisons of Circassian populations in Jordan (representing the diaspora) with Chechen populations show distinct clustering and significant Fst separation, confirming that these differences are ancient and not merely artifacts of recent political borders.<sup>6</sup>
- **Intra-Family Homogeneity vs. Heterogeneity:** The NWC cluster is relatively homogeneous, with low Fst values between Adyghe and Kabardians. In contrast, the NEC cluster, particularly in Dagestan, is highly fragmented. High Fst values are observed between different Dagestani ethnic groups (e.g., Kubachi vs. Lezgin), reflecting the extreme isolation of highland "auls" where endogamy has been practiced for centuries.<sup>23</sup> The "Patchy" distribution of allele frequencies in Dagestan contrasts with the smoother

"Clinical" distribution in the Northwest.<sup>23</sup>

## Pairwise Genetic Divergence (Fst) Among North Caucasian Populations

Genetic Distance Matrix (Fst)



Pairwise Fst genetic distance matrix for selected North Caucasian populations. Darker cells indicate greater genetic divergence. Note the relatively low distances (lighter cells) within the Northwest Caucasian (Adygei, Abkhaz) and Northeast Caucasian (Chechen, Ingush) clusters, respectively, compared to the higher distances observed between them. The Dagestani populations (Avar, Lezgin) show elevated distances from NWC groups, reflecting geographic isolation.

Data sources: [ResearchGate \(Nasidze et al.\)](#), [AskCaucasus Discussion](#), [PubMed \(Barbujani et al.\)](#), [Cambridge \(Wang et al.\)](#)

### 4.3 Genetic Barrier Analysis

Application of barrier analysis algorithms (such as Monmonier's algorithm) to Caucasian genomic data has explicitly identified the zones of abrupt genetic change.

- **Location of the Barrier:** The primary genetic boundary roughly corresponds to the topographical separation between the western basins and the eastern highlands. However, studies note that this barrier is **asymmetric**.<sup>5</sup> Gene flow appears to have been more restricted moving from the rugged Dagestani highlands to the west than vice versa.
- **Permeability:** The barrier is "semi-permeable." While it restricts mass migration, it has allowed for the diffusion of specific markers over time. The Ossetians, sitting directly on this fault line, act as a genetic bridge, showing affinities to both sides but aligning more closely with the West in terms of paternal ancestry.<sup>19</sup>

## 5. Uniparental Markers: The Y-Chromosome Dichotomy

While autosomal DNA reveals a gradient, the paternal lineages (Y-DNA) present a much sharper dichotomy that strongly correlates with the NWC-NEC linguistic boundary. This discrepancy suggests that male-mediated founder effects and patrilocal social structures have played a dominant role in shaping the genetic landscape of the Caucasus. The distribution of Y-haplogroups is not merely clinal; it is often distinct and non-overlapping.<sup>24</sup>

### 5.1 The Western Sphere: The Dominance of G2a

The Northwest Caucasian speakers (Abkhaz, Adyge, Cherkess, Kabardians) are characterized by the overwhelming dominance of Haplogroup **G2a** (specifically the **G2a-P15** and **G2a-P303** subclades).<sup>24</sup>

- **Frequencies and Founder Effects:** In populations such as the Adyge, Shapsug, and Abkhaz, G2a frequencies consistently exceed 50%, and in some sub-populations, reach 70-80%.<sup>24</sup> This indicates a profound founder effect, likely dating back to the Neolithic expansion.
- **The Ossetian Anomaly:** Strikingly, the highest frequencies of G2a in the entire Caucasus—and indeed the world—are found among the **North Ossetians** (specifically the Digor subgroup), where it can reach 70-75%.<sup>24</sup> This is a critical finding: despite speaking an Iranian language (Alanic), the Ossetian paternal gene pool is indistinguishable from their NWC neighbors. This confirms that the "Alanic" migration was an elite dominance event where a small incoming group imposed their language on a massive indigenous substratum without replacing the local male lineages.<sup>19</sup>
- **Origins:** Haplogroup G2a is the signature lineage of the Early European Farmers (EEF) who spread agriculture from Anatolia into Europe around 8,000 years ago. Its preservation at such high frequencies in the Northwest Caucasus suggests that this region served as a refugium for these Neolithic lineages, which were largely replaced in Europe by later Steppe migrations (R1a/R1b).<sup>28</sup> The specific subclade **G2a-P303** is

prevalent here, linking these populations to the Black Sea/Anatolian interaction sphere.<sup>29</sup>

## 5.2 The Eastern Sphere: The J-Haplogroup Complex

Crossing the boundary into the Northeast Caucasus, the genetic landscape shifts dramatically. G2a frequencies drop precipitously, replaced by lineages within Haplogroup J. However, there is a critical and sharp subdivision *within* the NEC family itself, separating the Nakh and Dagestanian branches.

### 5.2.1 The Nakh Cluster (Chechen & Ingush)

The Chechen and Ingush populations are dominated by **Haplogroup J2a** (specifically **J2a1b** and subclades like **J2-M67**).

- **Frequencies:** Among the Ingush, J2a reaches near-fixation levels of **80-88%**, one of the highest frequencies recorded globally.<sup>24</sup> Chechens also show very high frequencies (~55-60%), though with more diversity (including minorities of L3 and G2a).<sup>30</sup>
- **Significance:** J2a is associated with the expansion of civilizations from the Fertile Crescent and the northern Near East. Its extreme frequency in the Nakh populations indicates a distinct founder event separate from the G2a event in the west and the J1 event in the east.

### 5.2.2 The Dagestanian Cluster

Further east, among the Avars, Dargins, Lezgins, and Kubachians, the dominant lineage shifts again, this time to **Haplogroup J1** (specifically the **J1-Z1842** subclade).<sup>32</sup>

- **Frequencies:** In highland Dagestani isolates like the Kubachians and Dargins, J1 frequencies can exceed **70-90%**.<sup>24</sup>
- **Deep Links:** The specific subclade **J1-Z1842** has been explicitly linked to the **Kura-Araxes culture** (Early Bronze Age) and Eneolithic horizons in the Eastern Caucasus.<sup>25</sup> This suggests a direct paternal continuity between the Northeast Caucasian speakers and the builders of the Bronze Age civilizations that flourished along the Caspian flank.

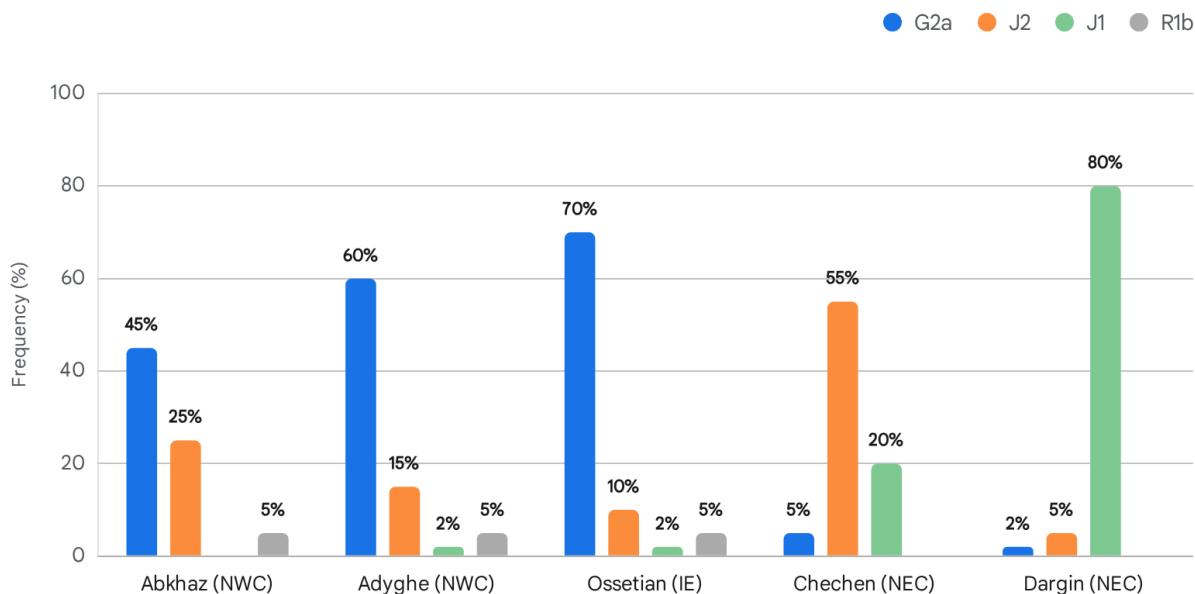
## 5.3 Summary of the Y-DNA Boundary

The shift in Y-DNA is tripartite and geographic:

1. **West (NWC & Ossetia):** Dominance of **G2a** (Neolithic Anatolian link).
2. **Central-East (Nakh):** Dominance of **J2a** (Fertile Crescent link).
3. **East (Dagestan):** Dominance of **J1-Z1842** (Kura-Araxes/Caspian link).

This "step-wise" change is far sharper than the autosomal cline, confirming that the mountains facilitated the isolation of male lineages into distinct "territories" that have remained stable for millennia.

## Y-DNA Haplogroup Frequency Distribution in North Caucasian Populations



Comparison of major Y-chromosome haplogroup frequencies across selected Northwest Caucasian (Abkhaz, Adygehe) and Northeast Caucasian (Chechen, Dargin) populations, alongside Ossetians. Note the dominance of G2a in the West (NWC) and among Ossetians, the prevalence of J2 in the Nakh group (Chechen), and the fixation of J1 in Dagestan (Dargin). Data compiled from Balanovsky et al. (2011) and subsequent studies.

Data sources: [Wikipedia \(Caucasus Y-DNA\)](#), [Academic OUP](#)

## 6. Mitochondrial DNA (mtDNA): The Contrast of Continuity

In contrast to the sharp boundaries seen in Y-chromosome markers, mitochondrial DNA (mtDNA)—inherited strictly maternally—presents a picture of high diversity but low geographic structure across the North Caucasus. Studies focusing on mtDNA variation (e.g., Nasidze et al.; Yardumian et al.) have consistently found that maternal lineages do not strictly follow the NWC-NEC linguistic divide.<sup>28</sup>

### 6.1 High Diversity and Lack of Structure

Populations across the North Caucasus display high levels of mtDNA diversity, with a wide array of West Eurasian haplogroups such as **H, U, K, T, and J** being common in both NWC and

NEC groups.<sup>18</sup> Unlike the Y-DNA, there is no single "Western" or "Eastern" maternal lineage that dominates one side of the boundary.

- **Haplogroup H:** Found at similar frequencies in Armenians (~30%), Azeris (29%), and North Caucasian groups like the Chamalals (Dagestan, ~67%) and Adyghe (NWC, ~36%).<sup>28</sup>
- **Ancient Continuity:** The high diversity suggests that the maternal gene pool is older and less affected by the founder effects and drift that shaped the paternal lines. It implies a different demographic history for women, characterized by higher mobility between groups (exogamy) or a larger effective population size that buffered against drift.<sup>23</sup>

## 6.2 Implications for Social Structure

The stark contrast between the sharp Y-DNA boundaries and the blurred mtDNA boundaries is a classic signature of **patrilocality**. In traditional Caucasian societies, women typically moved to their husband's village (spreading mtDNA), while men remained in their natal village (fixing Y-DNA). Over thousands of years, this practice homogenized the maternal landscape while deepening the paternal divide between the valleys.<sup>24</sup> Thus, the "genetic boundary" between NWC and NEC is primarily a *male* boundary.

# 7. The "Transitional" Populations: Testing the Boundary

A critical test of the correlation between language and genetics is the analysis of populations whose linguistic affiliation does not match their geographic position or the affiliation of their neighbors. The Ossetians and the Turkic speakers of the highlands (Karachay-Balkars) serve as perfect natural experiments.

## 7.1 The Ossetians: A Linguistic Island

The Ossetians speak an Iranian language (Iron/Digor), a remnant of the Scytho-Sarmatian and Alanic presence in the region. Geographically, they are wedged between the Kabardians (NWC) to the west and the Ingush (NEC) to the east.

- **Genetic Affinity:** Despite their Indo-European language, Ossetians are genetically firmly "North Caucasian." In autosomal PCA, they cluster tightly with the NWC groups and the Ingush, far removed from other Iranian speakers like Persians or Tajiks.<sup>17</sup>
- **Implication:** This confirms that the Ossetian ethnogenesis involved the assimilation of a large indigenous Caucasian population by a smaller, linguistically dominant Alanic elite. The genetic boundary between NWC and NEC essentially runs *through* the Ossetian territory, or rather, the Ossetians act as the buffer that geographically separates the two families today, while biologically belonging to the Western/Central Caucasian cluster.<sup>19</sup>

## 7.2 Karachay-Balkars and Kumyks: The Turkic Layer

The Karachay and Balkar peoples inhabit the high valleys of the Central/West Caucasus and

speak Turkic languages (Kipchak branch).

- **Admixture:** While they carry clear signals of Turkic/Steppe admixture (visible in minor frequencies of East Eurasian haplogroups and autosomal components), their primary ancestry is indigenous West Eurasian/Caucasian.<sup>35</sup>
- **Clustering:** In PCA comparisons, Karachay-Balkars cluster with their geographic neighbors—the Ossetians and Circassians—rather than with Central Asian Turkic populations.<sup>36</sup> They represent a "transitional" population where the indigenous substrate (related to NWC groups) was overlaid by successive waves of Steppe nomads (Huns, Khazars, Kipchaks) who imposed their language.<sup>29</sup>
- **Kumyks:** In contrast, the Kumyks of the Dagestani lowlands show slightly higher affinities to the Caspian steppe world but still retain the core Dagestani genetic profile (e.g., J1 lineages), further illustrating that geography trumps linguistics.<sup>38</sup>

## 8. Identity-by-Descent (IBD) and Fine-Scale Connectivity

Analysis of **Identity-by-Descent (IBD)** segments—long distinct chromosomal chunks shared between individuals—provides insight into more recent genealogical connections (past 1,000-2,000 years).

- **High Internal Sharing in the East:** Studies indicate extremely high levels of internal IBD sharing within Dagestani populations. This reflects the deep endogamy of the highland communities. The "runs of homozygosity" (ROH) are significantly longer in these groups compared to the NWC populations, indicating smaller effective population sizes and isolation.<sup>39</sup>
- **West-East Disconnect:** IBD sharing between the NWC and NEC groups is markedly lower than sharing within each group. While there is some sharing between transitional groups like Ossetians and both sides, the deep connection between a Circassian and a Lezgin is relatively weak compared to their respective connections to their immediate neighbors.<sup>40</sup>
- **External Links:** Interestingly, IBD analysis has revealed subtle connections between North Caucasian populations and groups as distant as the **Roma** (Gypsies). Snippet<sup>40</sup> highlights that Roma populations share significant IBD segments with North Caucasians (specifically Balkars, Lezgins, and Nogais), approaching 0.9 in some metrics. This likely reflects the transit of the proto-Roma population through the Caucasus or interaction with Armenian/Caucasian populations during their migration from India to Europe.

## 9. Synthesis: A Boundary Defined by Topography and Time

The genetic analysis of the NWC-NEC boundary leads to a nuanced conclusion. The "Great

"Divide" is not a simple line where one race ends and another begins. Rather, it is a **structured gradient** shaped by three forces:

1. **The Deep Substrate:** A shared Paleolithic/Mesolithic heritage (CHG) that unites all North Caucasians.
2. **The Ancient Divergence:** A split in the Eneolithic/Bronze Age between a "Western/Pontic" sphere (influenced by Anatolian farmers and the Black Sea) and an "Eastern/Caspian" sphere (influenced by Kura-Araxes and Steppe elements). This ancient divergence aligns remarkably well with the modern NWC vs. NEC distinction.
3. **The Mountain Fortress:** The extreme topography of the Caucasus amplified these initial differences by restricting gene flow and encouraging drift, particularly in the paternal line.

#### **Comparative Metrics Table:**

To summarize the key differentiators, the following table presents the genetic metrics defining the boundary.

Feature	Northwest Caucasian (NWC)	Northeast Caucasian (NEC)
<b>Representative Populations</b>	Adyghe, Kabardian, Abkhaz, Abaza	Chechen, Ingush, Avar, Dargin, Lezgin
<b>Dominant Y-Haplogroup</b>	G2a (>50-70%) <sup>24</sup>	J2a (Nakh), J1 (Dagestan) <sup>24</sup>
<b>Secondary Y-Haplogroups</b>	R1a, J2 (minor)	L3, G2a (minor in Nakh)
<b>Autosomal Cluster (PCA)</b>	Clusters with West Georgians & Ossetians	Distinct "East Caucasus" cluster
<b>Fst Distance to Outgroups</b>	Closer to Anatolia/Europe	More isolated/drifted; high internal Fst
<b>Ancient Ancestry (Model)</b>	CHG + High Anatolian Neolithic	CHG + Variable Steppe + Low Anatolian
<b>Steppe Admixture</b>	Moderate (Wang et al.)	Variable; generally lower in

	2019)	isolated highlands
<b>Key Ancient Link</b>	Maykop (Caucasus Cluster) <small>32</small>	Kura-Araxes (J1-Z1842 link)

## 10. Conclusion

The genetic boundary between Northwest and Northeast Caucasian speakers is a biological reality, but it is one defined by **geography and time depth** rather than absolute racial difference. The linguistic families correspond to deep phylogenetic splits in the Y-chromosome (G2a vs. J) that likely predate the modern languages themselves, rooting the NWC-NEC divide in the Neolithic or Eneolithic period.

While the autosomal data shows a cline—a gradual transition facilitated by the "bridge" populations of Ossetia and the central valleys—the uniparental data reveals a history of sharp isolation. The mountains did not just separate languages; they preserved distinct reservoirs of ancient human diversity. The Northwest retained the legacy of the Anatolian farmers who moved along the Black Sea, while the Northeast became a fortress for lineages connected to the Caspian and the early civilizations of the Kura-Araxes. Thus, the linguistic map of the Caucasus is indeed a reflection of its genetic history: two great families, rooted in the same mountain soil, but reaching towards different horizons for millennia.

### Works cited

1. Ethnic groups in the Caucasus - Wikipedia, accessed on January 7, 2026, [https://en.wikipedia.org/wiki/Ethnic\\_groups\\_in\\_the\\_Caucasus](https://en.wikipedia.org/wiki/Ethnic_groups_in_the_Caucasus)
2. The Caucasus (Chapter 13) - The Cambridge Handbook of Areal Linguistics, accessed on January 7, 2026, <https://www.cambridge.org/core/books/cambridge-handbook-of-areal-linguistics/caucasus/D26A5236E4ACD4BC99B747242422DE31>
3. Caucasus hunter-gatherer - Wikipedia, accessed on January 7, 2026, [https://en.wikipedia.org/wiki/Caucasus\\_hunter-gatherer](https://en.wikipedia.org/wiki/Caucasus_hunter-gatherer)
4. The genetic history of the Southern Arc: a bridge between West Asia and Europe - PMC, accessed on January 7, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10064553/>
5. Caucasus as an Asymmetric Semipermeable Barrier to Ancient Human Migrations | Molecular Biology and Evolution | Oxford Academic, accessed on January 7, 2026, <https://academic.oup.com/mbe/article/29/1/359/1750206>
6. The Circassians and the Chechens in Jordan: results of a decade of epidemiological and genetic studies - PMC - NIH, accessed on January 7, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10725377/>
7. The genetic history of admixture across inner Eurasia - PMC - PubMed Central, accessed on January 7, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC6542712/>
8. The genetic prehistory of the Greater Caucasus – Preprint. A few answers and

- many questions. - Ancient DNA Era, accessed on January 7, 2026,  
<https://adnaera.com/2018/05/21/the-genetic-prehistory-of-the-greater-caucasus-preprint-a-few-answers-and-many-questions/>
9. Admixture Graph modelling of the population history of the Caucasus... - ResearchGate, accessed on January 7, 2026,  
[https://www.researchgate.net/figure/Admixture-Graph-modelling-of-the-population-history-of-the-Caucasus-region-We-started\\_fig2\\_330855366](https://www.researchgate.net/figure/Admixture-Graph-modelling-of-the-population-history-of-the-Caucasus-region-We-started_fig2_330855366)
10. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions - PubMed Central, accessed on January 7, 2026, <https://PMC.ncbi.nlm.nih.gov/articles/PMC6360191/>
11. The Caucasus: complex interplay of genes and cultures, accessed on January 7, 2026, <https://www.shh.mpg.de/1196995/pr-caucasus-haak>
12. The genetic origin of the Indo-Europeans - David Reich Lab - Harvard University, accessed on January 7, 2026,  
[https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/2025\\_LazaridisPattersonAnthonyVayazov\\_IndoEuropean\\_Nature\\_0.pdf](https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/2025_LazaridisPattersonAnthonyVayazov_IndoEuropean_Nature_0.pdf)
13. The genetic history of the Southern Arc: A bridge between West Asia and Europe, accessed on January 7, 2026,  
<https://scholarlyworks.adelphi.edu/esploro/outputs/journalArticle/The-genetic-history-of-the-Southern/991004363990406266>
14. The Southern Arc and its lively genetic history - ScienceDaily, accessed on January 7, 2026,  
<https://www.sciencedaily.com/releases/2022/08/220825164029.htm>
15. The genetic history of the Southern Arc: A bridge between West Asia and Europe - David Reich Lab, accessed on January 7, 2026,  
[https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/8\\_25\\_2022\\_Manuscript1\\_ChalcolithicBronzeAge.pdf](https://reich.hms.harvard.edu/sites/reich.hms.harvard.edu/files/inline-files/8_25_2022_Manuscript1_ChalcolithicBronzeAge.pdf)
16. | Admixture modelling of Bronze Age populations | Download Table - ResearchGate, accessed on January 7, 2026,  
[https://www.researchgate.net/figure/Admixture-modelling-of-Bronze-Age-populations\\_tbl1\\_318862250](https://www.researchgate.net/figure/Admixture-modelling-of-Bronze-Age-populations_tbl1_318862250)
17. Genetic similarities between North Caucasians and Southern Caucasians : r/AskCaucasus - Reddit, accessed on January 7, 2026,  
[https://www.reddit.com/r/AskCaucasus/comments/19fclq4/genetic\\_similarities\\_between\\_north\\_caucasians\\_and/](https://www.reddit.com/r/AskCaucasus/comments/19fclq4/genetic_similarities_between_north_caucasians_and/)
18. Genetic Analysis of Mingrelians Reveals Long-Term Continuity of Populations in Western Georgia (Caucasus) | Genome Biology and Evolution | Oxford Academic, accessed on January 7, 2026,  
<https://academic.oup.com/gbe/article/15/11/evad198/7341981>
19. Ossetian genetics - Turkicworld, accessed on January 7, 2026,  
[http://s155239215.onlinehome.us/turkic/60\\_Genetics/OssetianGeneticsEn.htm](http://s155239215.onlinehome.us/turkic/60_Genetics/OssetianGeneticsEn.htm)
20. Pairwise Fst values between regions calculated based on reduced data-sets, accessed on January 7, 2026,  
[https://www.researchgate.net/figure/Pairwise-Fst-values-between-regions-calculated-based-on-reduced-data-sets\\_tbl3\\_8525622](https://www.researchgate.net/figure/Pairwise-Fst-values-between-regions-calculated-based-on-reduced-data-sets_tbl3_8525622)

21. Parallel Evolution of Genes and Languages in the Caucasus Region - PMC, accessed on January 7, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC3355373/>
22. (A) geographical map of the populations of the Caucasus included in... - ResearchGate, accessed on January 7, 2026, [https://www.researchgate.net/figure/A-geographical-map-of-the-populations-of-the-Caucasus-included-in-this-study-The\\_fig1\\_51630056](https://www.researchgate.net/figure/A-geographical-map-of-the-populations-of-the-Caucasus-included-in-this-study-The_fig1_51630056)
23. Genetic diversity in the Caucasus - PubMed, accessed on January 7, 2026, <https://pubmed.ncbi.nlm.nih.gov/8088752/>
24. Y-DNA haplogroups in populations of the Caucasus - Grokipedia, accessed on January 7, 2026, [https://grokipedia.com/page/Y-DNA\\_haplogroups\\_in\\_populations\\_of\\_the\\_Caucasus](https://grokipedia.com/page/Y-DNA_haplogroups_in_populations_of_the_Caucasus)
25. Geographic location, linguistic affiliation and genetic composition of... | Download Scientific Diagram - ResearchGate, accessed on January 7, 2026, [https://www.researchgate.net/figure/Geographic-location-linguistic-affiliation-and-genetic-composition-of-the-studied\\_fig1\\_51126445](https://www.researchgate.net/figure/Geographic-location-linguistic-affiliation-and-genetic-composition-of-the-studied_fig1_51126445)
26. Geographic distribution of major Y-DNA haplogroups in the Caucasus. The... | Download Scientific Diagram - ResearchGate, accessed on January 7, 2026, [https://www.researchgate.net/figure/Geographic-distribution-of-major-Y-DNA-haplogroups-in-the-Caucasus-The-figure-shows\\_fig2\\_265380811](https://www.researchgate.net/figure/Geographic-distribution-of-major-Y-DNA-haplogroups-in-the-Caucasus-The-figure-shows_fig2_265380811)
27. Y-DNA haplogroups in populations of the Caucasus - Wikipedia, accessed on January 7, 2026, [https://en.wikipedia.org/wiki/Y-DNA\\_haplogroups\\_in\\_populations\\_of\\_the\\_Caucasus](https://en.wikipedia.org/wiki/Y-DNA_haplogroups_in_populations_of_the_Caucasus)
28. Genetic History of the Caucasus (Chapter Eight), accessed on January 7, 2026, <https://www.cambridge.org/core/books/people-of-the-caucasus/genetic-history-of-the-caucasus/BDBAB52101D2E5052929C9D4A88CEA9F>
29. The paternal genetic legacy of Hungarian-speaking Rétköz (Hungary) and Váh valley (Slovakia) populations - PMC - PubMed Central, accessed on January 7, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9619085/>
30. Ingush, chechens and dagestani are the most pure north caucasians. : r/AskCaucasus - Reddit, accessed on January 7, 2026, [https://www.reddit.com/r/AskCaucasus/comments/yb4syt/ingush\\_chechens\\_and\\_dagestani\\_are\\_the\\_most\\_pure/](https://www.reddit.com/r/AskCaucasus/comments/yb4syt/ingush_chechens_and_dagestani_are_the_most_pure/)
31. Haplogroup G (Y-DNA) by country - Wikipedia, accessed on January 7, 2026, [https://en.wikipedia.org/wiki/Haplogroup\\_G\\_\(Y-DNA\)\\_by\\_country](https://en.wikipedia.org/wiki/Haplogroup_G_(Y-DNA)_by_country)
32. (PDF) Origins of East Caucasus Gene Pool: Contributions of Autochthonous Bronze Age Populations and Migrations from West Asia Estimated from Y-Chromosome Data - ResearchGate, accessed on January 7, 2026, [https://www.researchgate.net/publication/373849877\\_Origins\\_of\\_East\\_Caucasus\\_Gene\\_Pool\\_Contributions\\_of\\_Autochthonous\\_Bronze\\_Age\\_Populations\\_and\\_Migrations\\_from\\_West\\_Asia\\_Estimated\\_from\\_Y-Chromosome\\_Data](https://www.researchgate.net/publication/373849877_Origins_of_East_Caucasus_Gene_Pool_Contributions_of_Autochthonous_Bronze_Age_Populations_and_Migrations_from_West_Asia_Estimated_from_Y-Chromosome_Data)
33. Comparison of the genetic and linguistic trees of North Caucasus populations ... - ResearchGate, accessed on January 7, 2026, <https://www.researchgate.net/figure/Comparison-of-the-genetic-and-linguistic-t>

rees-of-North-Caucasus-populations\_fig4\_51126445

34. Editors' Pick: A real Caucasian and the genetic investigation of Caucasus peoples - PMC, accessed on January 7, 2026,  
<https://PMC3492008/>
35. if caucasians share the same CHG ancestry, why too many differences in DNA and language? - Reddit, accessed on January 7, 2026,  
[https://www.reddit.com/r/AskCaucasus/comments/1bfifzw/if\\_caucasians\\_share\\_the\\_same\\_chg\\_ancestry\\_why\\_too/](https://www.reddit.com/r/AskCaucasus/comments/1bfifzw/if_caucasians_share_the_same_chg_ancestry_why_too/)
36. Genetic Characterization of Balkars and Karachays Using mtDNA Data - ResearchGate, accessed on January 7, 2026,  
[https://www.researchgate.net/publication/332434053\\_Genetic\\_Characterization\\_of\\_Balkars\\_and\\_Karachays\\_Using\\_mtDNA\\_Data](https://www.researchgate.net/publication/332434053_Genetic_Characterization_of_Balkars_and_Karachays_Using_mtDNA_Data)
37. Genetic Analysis of Mingrelians Reveals Long-Term Continuity of Populations in Western Georgia (Caucasus) - Oxford Academic, accessed on January 7, 2026,  
<https://academic.oup.com/gbe/article-pdf/15/11/evad198/53689694/evad198.pdf>
38. Genetic distance to ethnic groups in the Caucasus : r/AskCaucasus - Reddit, accessed on January 7, 2026,  
[https://www.reddit.com/r/AskCaucasus/comments/s6zdrp/genetic\\_distance\\_to\\_ethnic\\_groups\\_in\\_the\\_caucasus/](https://www.reddit.com/r/AskCaucasus/comments/s6zdrp/genetic_distance_to_ethnic_groups_in_the_caucasus/)
39. Sharing of ancient and recent IBD blocks between populations in focus... - ResearchGate, accessed on January 7, 2026,  
[https://www.researchgate.net/figure/Sharing-of-ancient-and-recent-IBD-blocks-between-populations-in-focus-regions-IBD\\_fig3\\_322110176](https://www.researchgate.net/figure/Sharing-of-ancient-and-recent-IBD-blocks-between-populations-in-focus-regions-IBD_fig3_322110176)
40. Revealing the impact of the Caucasus region on the genetic legacy of Romani people from genome-wide data - PMC - PubMed Central, accessed on January 7, 2026, <https://PMC6130880/>