bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Gothic Identity as Cultural Practice: Paleogenomic Evidence for Multi - Ethnic Assemblages Under Gothic Material Culture in Late Antique Bulgaria (4th - 6th centuries CE)

Svetoslav Stamov[1,2], Todor Chobanov[4,2], Tianyi Wang[3], Kremena Stoeva[5], Dimcho Momchilov[6,7], Andrey Aladzhov[8], Kaloyan Chobanov, Milen Nikolov[6], Desislava Nesheva[1,9], Peter Heather[10], Draga I. Toncheva[11], Milen Zamfirov[12], Iosif Lazaridis[3], David Reich[3,13]

[1]National Museum of History, Sofia, Bulgaria; Laboratory of Interdisciplinary Research and Paleogenomic Knowledge, Sofia, Bulgaria

[2]UNIBIT, Sofia, Bulgaria

[3]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA; Department of Genetics, Harvard Medical School, Boston, MA, USA

[4]Institute for Balkan Studies with Centre for Thracology, Bulgarian Academy of Sciences, Sofia, Bulgaria

[5]Archaeological Museum, Veliki Preslav, Bulgaria

[6]Regional Historical Museum, Burgas, Bulgaria

[7]Asen Zlatarov University, Burgas, Bulgaria

[8]National Archaeological Institute with Museum, Bulgarian Academy of Sciences, Sofia, Bulgaria

[9]Medical University of Sofia, Sofia, Bulgaria

[10]King's College London, London, United Kingdom

[11]Bulgarian Academy of Sciences, Sofia, Bulgaria

[12]Sofia University St. Kliment Ohridski, Sofia, Bulgaria

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

[13]Howard Hughes Medical Institute, Boston, MA, USA

*Correspondence: stamovsvetoslav@gmail.com

Abstract

Ethnonyms such as “Goth” in Late Antique sources capture political and cultural affiliations that may not map cleanly onto biological descent. Here we report genome - wide ancient DNA from 38 individuals associated with Gothic - period mortuary contexts at two sites in present - day Bulgaria: the Aquae Calidae necropolis (~320 - 375 CE) and the Aul of Khan Omurtag necropolis (~350 - 489 CE). Using PCA, f - statistics, qpAdm, uniparental markers, and IBD/kinship analyses, we find: (i) strong within - site heterogeneity, rejecting a single “Gothic” genetic profile; (ii) a reproducible north - south genetic contrast, with Aquae Calidae individuals shifted toward a Balkan/Anatolian - related ancestry axis and AKO individuals enriched in northern European - related ancestry consistent with Wielbark/Chernyakhiv proxies; and (iii) admixture dating with DATES placing the mixing between northern and southern ancestry poles at ~11 - 13 generations before burial (point estimates in the 1st century CE, depending on target grouping), based on 23 individuals with sufficient coverage. Together, these results support models in which Gothic material culture in the Balkans was practiced by multi - ethnic communities and illustrate how cultural “Gothic” identity could persist despite substantial genetic diversity. Full f3/qpAdm/DATES outputs, f4 validation, and kinship/IBD summaries are provided in Supplementary Tables S1-S6, Supplementary Notes S2-S4, and the Supplementary IBD Workbook.

1. Introduction

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1.1 Historical Context of Gothic Migrations to the Balkans

The Goths represent one of the most dynamic and transformative groups during the Migration Period (c. 300 - 700 CE), an era marked by the dissolution of the Western Roman Empire and the reconfiguration of European ethnic landscapes. Traditional historical narratives, drawing from sources such as Jordanes' *Getica* (c. 551 CE), trace Gothic origins to Scandinavia, positing a coherent migration of East Germanic - speaking peoples southward through Eastern Europe and into the Roman Balkans (Fig. 1). However, archaeological and genetic evidence reveals a far more complex process of ethnogenesis, involving progressive admixture with diverse populations and the formation of multi - ethnic confederations (Heather and Matthews, 1991; Stolarek et al., 2019).



Figure 1. Locations of Aquae Calidae and the Aul of Khan Omurtag (AKO) sites in Bulgaria, with inferred Gothic migration routes and timeline (320 - 490 CE) (Heather, 1986; Kulikowski, 2007).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

The Gothic presence in the Balkans spans five centuries (2nd - 7th CE), characterized by multiple waves of migration, raiding, and settlement that incorporated Iranian steppe nomads (Sarmatians and Alans), Hunnic elements (Maenchen-Helfen, 1973), and local Thracian/Dacian provincials (see Supplementary Note S4 for detailed chronology). Key phases include the Wielbark culture homeland in northern Poland (1st - 4th centuries CE), characterized by high Scandinavian - related ancestry (Speidel et al., 2025; Stolarek et al., 2023) and elevated frequencies of Y - haplogroup I1 (~41%; Stolarek et al., 2023); the multi - ethnic Chernyakhov culture (late 2nd - 5th centuries CE) in Ukraine/Moldova/Romania, where Gothic - associated populations merged with Sarmatians, Dacians, and Alans (Saag et al., 2025; Järve et al., 2019); 3rd - century raids on Anatolia capturing thousands (Thompson, 2008), including Central Anatolian Christians (as the ancestors of Bishop Ulfilas); 4th - century treaties establishing foederati settlements culminating in the 382 CE treaty under Theodosius I; post - Hunnic reconfigurations after 454 CE; and Theodoric's 488 CE departure for Italy (Wolfram, 1988, Heather, 1996), leaving residual populations in the region.

1.2 The Debate on Gothic Identity: Traditional vs. Ethnogenesis Models

Historical interpretations of Gothic identity have evolved from essentialist views, treating "Goths" as a biologically and culturally coherent tribe migrating en masse (e.g., Heather, 1996, emphasizing core Germanic continuity), to ethnogenesis models, which posit identities as dynamic cultural - political constructs. Wolfram (1988) and Pohl (2013) argue that Gothic ethnogenesis involved a "tradition - bearing core" (e.g., the Amal and Balt dynasties) around which diverse groups coalesced through shared institutions such as Arian Christianity, Gothic language, codified by Ulfilas' Bible translation (Streitberg, 1950), and federate status. This process incorporated non - Germanic elements - Alans, Sarmatians, Huns, Roman provincials - maintaining identity through social practices rather than strict descent.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Neither model is without limitations. Traditional frameworks struggle with the multi - ethnic evidence from the Chernyakhov culture, while ethnogenesis approaches risk overemphasizing fluidity and potentially underplaying biological continuity visible, for example, in the persistence of Scandinavian - associated Y - haplogroups across time and space. The implications extend to reevaluating Migration Period "barbarian" movements as complex integrations rather than wholesale demographic replacements, with broader relevance for understanding early medieval state formation, including the First Bulgarian State, founded in 681 CE.

1.3 The Role of Ancient DNA in Resolving Identity Questions

Ancient DNA provides a powerful tool to test ethnogenesis models by quantifying biological ancestry against cultural affiliations. Formal admixture methods such as qpAdm and f - statistics can partition ancestry into discrete streams and distinguish pre - from post - migration admixture, while identity - by - descent (IBD) analysis reveals social boundaries (e.g., endogamy, marriage practices). Admixture dating methods such as DATES (Narasimhan et al., 2019) can estimate the timing of mixture events by measuring the decay of ancestry - specific linkage disequilibrium, providing independent chronological constraints. For Gothic - associated populations, aDNA can assess the degree of northern European continuity - using Eastern European Hunter - Gatherer (EHG) ancestry as a diagnostic proxy, as Germanic speakers entering the Balkans would carry more EHG than Balkan natives (Olalde et al., 2023), though EHG is not a sufficient marker for Germanic attribution since other northern European groups such as Slavs (Lazaridis et al., 2022) and Finnic peoples also carry elevated levels - versus incorporation of Anatolian, steppe, and East Asian components. This study leverages aDNA from two Bulgarian sites to evaluate ethnogenesis at the molecular level.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1.4 Study Sites and Archaeological Context

Aquae Calidae (Thrace, ~320 - 375 CE, n=23) is a 4th - century necropolis adjacent to the thermal Roman baths at Aquae Calidae (Momchilov et al., 2025), renowned in antiquity as a healing center. The burials ignore prior Roman structures, and some individuals exhibit artificial cranial deformation, suggesting non - Roman upbringing. Extreme male bias (17:3 among sexed individuals) is consistent with a military garrison or monastic establishment. Diagnostic Gothic artifacts (beads, fibulae) and Arian burial orientations are present. Two historical sources document Gothic presence at the site: the first (Jordanes in *Getica* , 20.109) places Gothic arrival around 270 CE for a brief raiding episode; a second source, Procopius (*On Buildings* 3.7.18–23, ~550 CE), mentions that "barbarian villagers" still occupied the site, preventing Roman provincials from using the baths, until Justinian expelled them and fortified the bath buildings. Additional archaeological context, dating anchors, and site-specific chronology are summarized in Supplementary Note S4.

The Aul of Khan Omurtag (AKO, Moesia Secunda, ~350 - 489 CE, n=15) is an Arian episcopal center with four basilicas, potentially tied to Ulfilas' 348 CE settlement of Christian Goths in Moesia (Stoeva, 2014). The term *aul* (from Turkic) refers to a fortified royal encampment or settlement characteristic of steppe nomadic polities; the site was later reused as a Protobulgarian royal camp, from which its modern name derives. Sequential necropoleis (C1 - C5) span pre - Hunnic phases (C1/C2, with Chernyakhov - like material culture), Hunnic - period contexts (C3/C4, showing disruptions and East Asian admixture), and post - Nedao occupation (C5, a burial ground with Pannonian - influenced material (Balabanov, 2006), temporally aligned with Theodoric's departure from the region). Some burials display rich Germanic jewelry and cranial

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

deformation (a steppe elite marker), reflecting multi - wave occupation. Phase definitions, archaeological dating anchors, and site context are summarized in Supplementary Note S4.

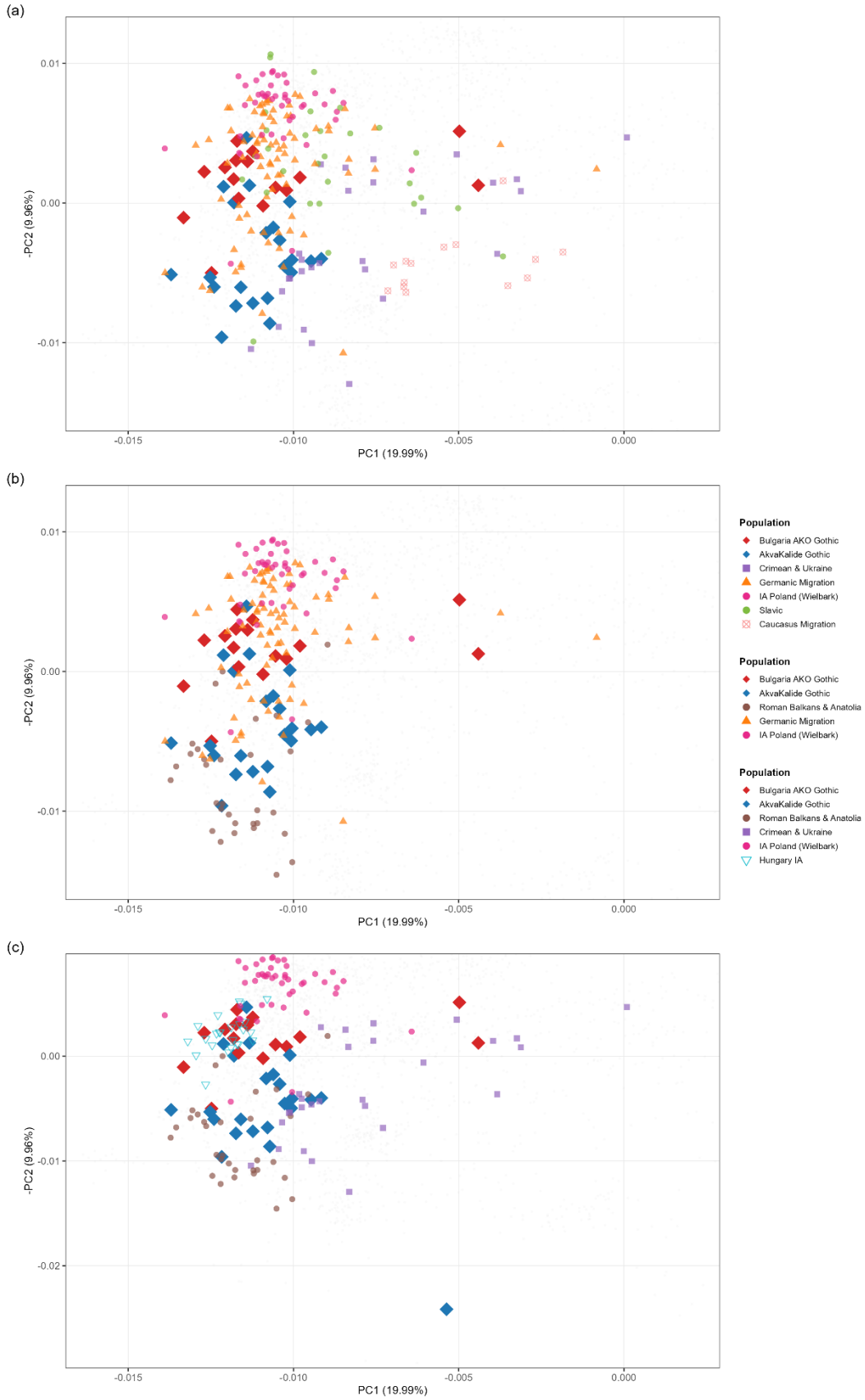
Together, these two sites encapsulate the timeline's complexity and enable ancient DNA to probe Gothic ethnogenesis across different phases of the Migration Period. Detailed archaeological chronology, site plans, and phase assignments are provided in Supplementary Note S4 (Figs. S1-S2).

2. Results

2.1 Population Structure and Affinities

Principal Component Analysis. PCA of genome - wide data, projected onto modern West Eurasian variation (PCs 1 - 2 capturing ~19.9% and ~9.9% of total variance, respectively; first 10 PCs capture ~79%), reveals pronounced genetic differentiation between and within the two Gothic - associated sites despite their shared material culture and Arian Christian burial practices

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Figure 2. Principal component analysis (PCA) of ancient individuals from Aquae Calidae and AKO projected onto modern West Eurasian variation. (A) Overview showing clustering of AKO individuals with Migration Period Germanic groups and Aquae Calidae individuals with Balkan Late Antiquity populations. (B) Detailed view highlighting intra - site subclusters at Aquae Calidae (PCA - South and PCA - North subclusters).

Individuals from AKO (n=15) form a relatively tight cluster in the upper portion of the PCA plot, aligning with other Migration Period Germanic - associated groups (e.g., Hungary_Langobard, Germanic Migration composites) and Pannonian references (Amorim et al., 2018). AKO individuals are positioned closer to Iron Age Poland/Wielbark (Stolarek et al., 2023) and Crimean/Ukrainian steppe references (Saag et al., 2025), than to Aquae Calidae, consistent with a genetic profile deriving primarily from northern European (Wielbark) and Pontic multi - ethnic (Chernyakhov) sources, with secondary steppe inputs.

In contrast, Aquae Calidae individuals (n=23) scatter toward the lower - left portion of the plot, clustering with Balkan Late Antiquity and Roman – period (Olade et al., 2023, Lazaridis et al., 2022) populations. The two sites are more distant from each other in PCA space than either is from its respective closest reference populations. Within Aquae Calidae, individuals separate into two subclusters along PCs 1 - 2: a southern subcluster (“PCA - South,” n=12, including one extreme outlier positioned near ancient Egyptian samples) shifted toward Anatolian Bronze Age proxies (Kalehöyük, ~2000 BCE), and a northern subcluster (“PCA - North,” n=10) aligned more closely with Byzantine - era Marmara/Iznik references and local Balkan Antiquity samples. Both subclusters share identical Gothic diagnostic artifacts and east - west burial orientations consistent with Arian Christianity.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Outgroup f3 statistics. f3 statistics of the form f3(Mbuti; Target_Gothic, Reference), where higher values indicate greater shared drift, corroborate and quantify the PCA patterns (Fig. 3; Supplementary Table S1).

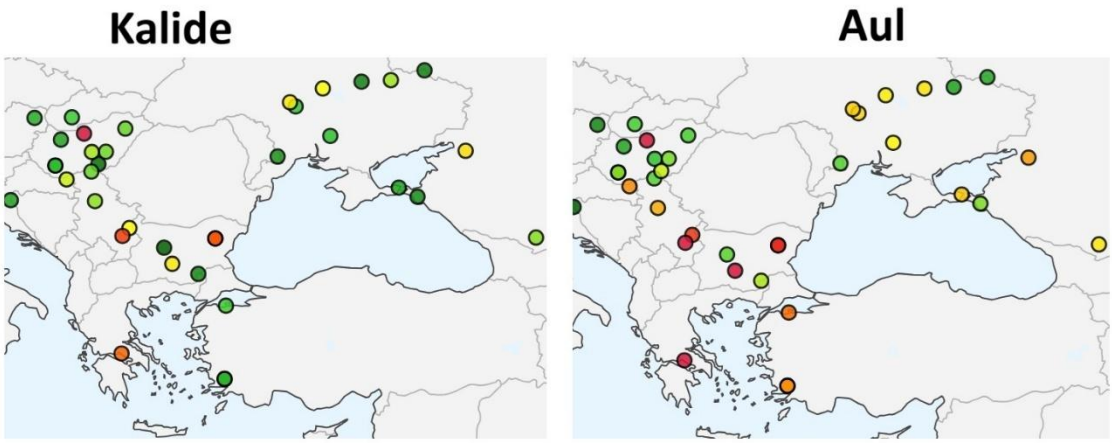


Figure 3. Outgroup f3 statistics heatmaps. (A) Aquae Calidae f3 outgroup statistics; (B) AKO f3 outgroup statistics. Color gradient from green (higher affinity) to red (lower affinity). The f3 form is f3(Mbuti; Target, Reference).

Both sites show strongest affinities to Central and Eastern European Migration Period populations. Top shared affinities ($f3 > 0.052$, $|Z| > 70$) include Hungary_Langobard (AKO: 0.0528; Aquae Calidae: 0.0523), Austria_Klosterneuburg_Roman (AKO: 0.0527; Aquae Calidae: 0.0522), and Poland_Wielbark_IA (AKO: 0.0527; Aquae Calidae: 0.0522), indicating common ancestry involving northern and central European migratory groups.

Site - specific patterns emerge in the ranking of affinities. Aquae Calidae ranks Bulgaria_LAntiquity as its highest overall affinity (f3=0.0524, rank #1), a connection 2 - 3× stronger than in AKO (rank #21), indicating greater local Balkan continuity or pre - arrival Mediterranean admixture. AKO shows relatively higher values with Wielbark and Chernyakhov

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

composites, consistent with more direct central or eastern European derivation. Moderate shared drift with Slavic Early Medieval (Stolarek et al., 2023), Pannonian Sarmatian (Gnecchi-Ruscone et al., 2022), and Pontic Scythian references (f3 ~0.045 - 0.050) is present at both sites, while affinities with eastern nomadic groups carrying substantial East Asian ancestry (e.g., Kazakhstan_Nomad_Hun_Sarmatian) are uniformly low (f3 < 0.040). The f3 patterns are robust to reference choice and temporal offsets, with all top results at $|Z| > 70$. However, f3 measures average population affinity and thus masks individual - level heterogeneity.

Demographic context. Aquae Calidae exhibits extreme male bias (17 males : 3 females : 3 indeterminate among sexed individuals; M:F ratio ~5.67:1), consistent with a military garrison or monastic setting. AKO shows near sex parity (consistent with a stable multi - generational community). No significant sex - biased ancestry patterns emerge in f4 tests, indicating bidirectional gene flow rather than unidirectional capture scenarios. Formal f4 test results supporting these statements are reported in Supplementary Table S6; limited sample sizes reduce power to detect subtle sex bias.

2.2 Ancestry Modeling

We employed qpAdm admixture modeling (admixtools2, Maier et al., 2023) to estimate ancestry proportions from specified source populations (right populations: Mbuti.DG, Israel_Natufian.SG, Russia_Yana_UP.SG, Russia_MA1_HG.SG, Russia_Kostenki14.SG, Italy_North_Villabruna_HG.SG, Morocco_Iberomaurusian.SG, Turkey_Boncuklu_PPN.SG, Iran_GanjDareh_N.SG, Karitiana.DG, Russia_DevilsCave_N.SG). We tested both proximal models (using temporally relevant sources: Wielbark, Chernyakhov, Anatolian references, Bulgaria_EIA) and distal models (EHG, CHG, WHG, Turkey_N, Yamnaya). Results are visualized in Figures 4 A, B, C and D; full model details are in Supplementary Tables S2–S4.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Models were evaluated for statistical fit ($p > 0.05$ considered feasible) and parsimony. To assess heterogeneity across individuals, we applied a uniform set of simpler one - , two - , and three - way models to all individuals, determining which models each individual fits versus rejects (Tables 1 - 2; Supplementary Tables S2 - S4).

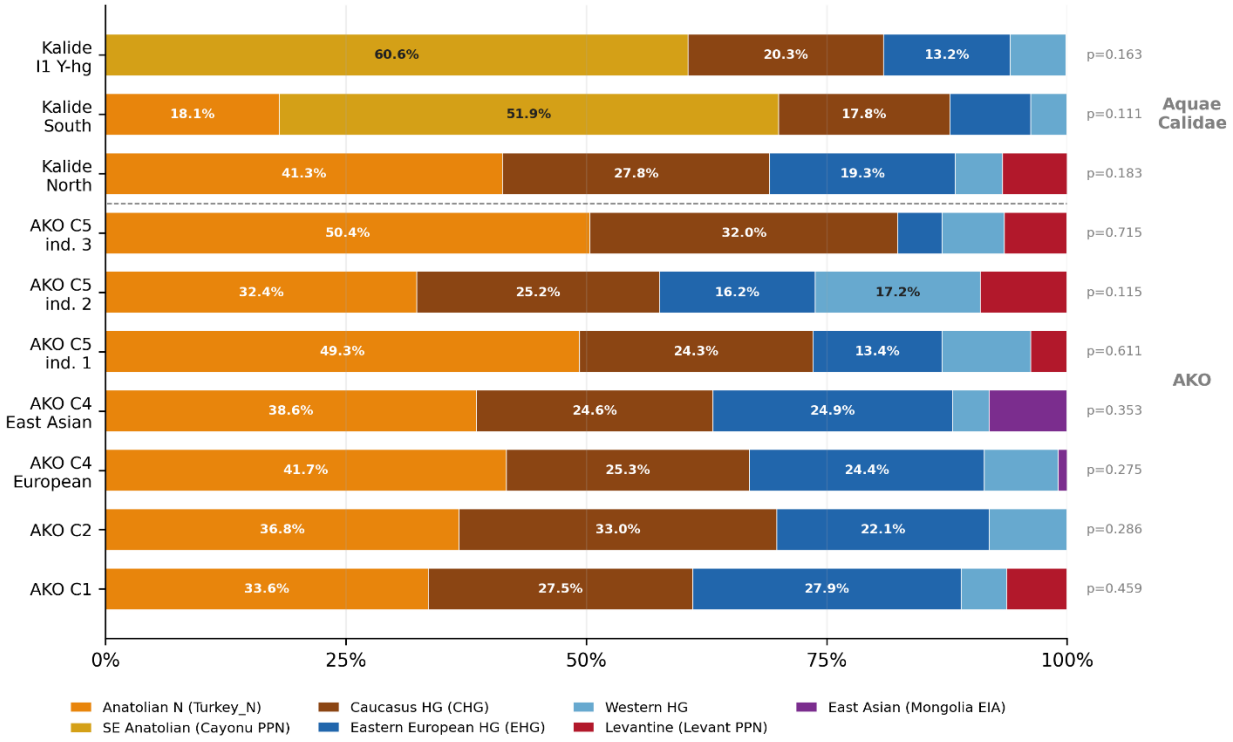
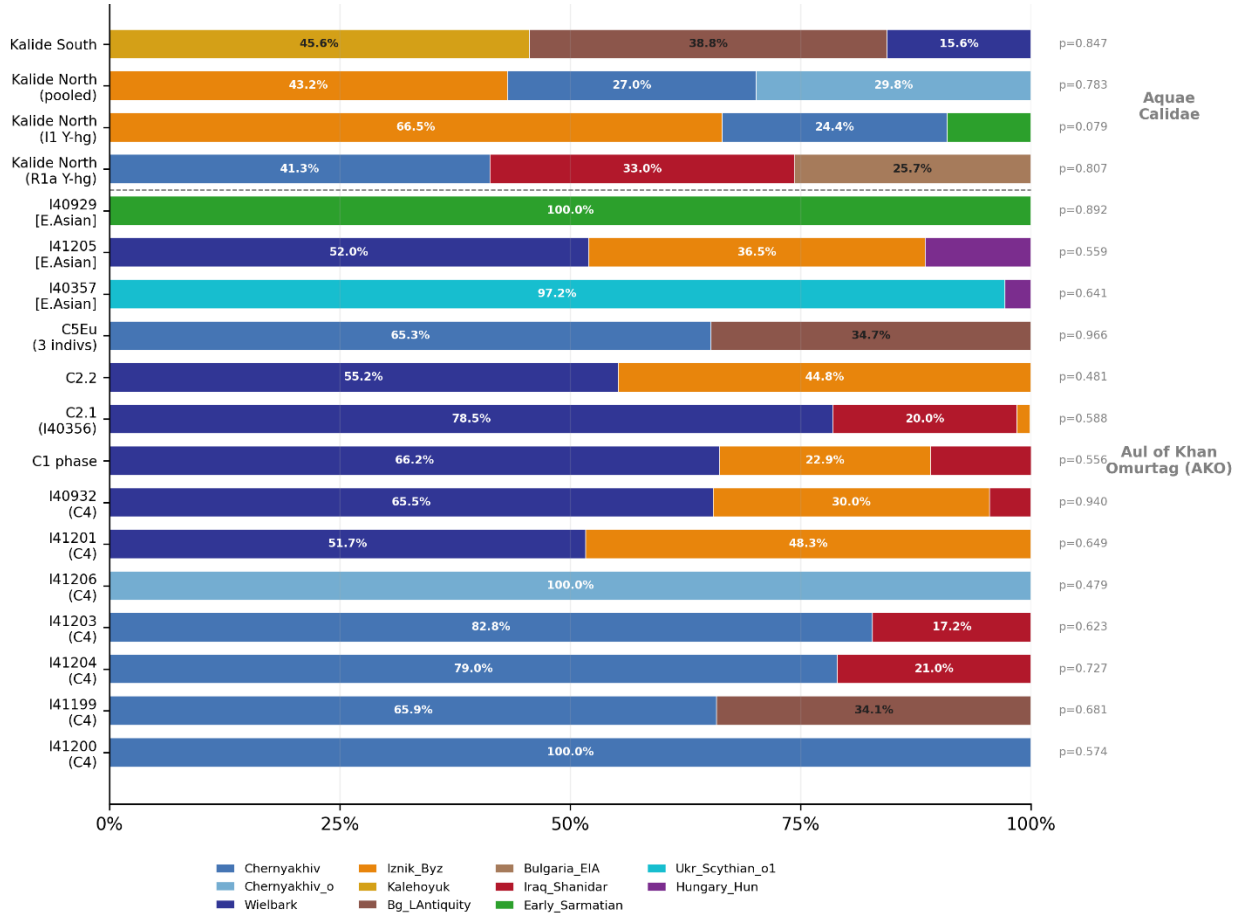


Figure 4A-C. qpAdm ancestry decomposition of Gothic populations. (A) Distal qpAdm deep-ancestry decomposition using extended source set (Turkey_N, Cayonu_PPN, CHG, EHG, WHG, Levant_PPN, Mongolia_EIA). Each bar shows the best-fitting model for a population or individual group. Aquae Calidae populations show dominant Anatolian Neolithic (Turkey_N) ancestry, while AKO populations display more balanced profiles with substantial Caucasus and Eastern European hunter-gatherer components. All models pass at $p > 0.05$.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



(B) Proximal qpAdm ancestry models using temporally relevant source populations (Tables S2–S3). Upper panel: Aquae Calidae subclusters showing Chernyakhov, Iznik_Byz, Kalehöyük, and Bulgaria_EIA sources. Lower panel: AKO individuals and phase groups, with color-coded sources including Chernyakhov, Wielbark, and steppe-related proxies. Dashed line separates the two sites. All displayed models have p ≥ 0.05.

2.2.1 Aquae Calidae Ancestry Profiles

Population - level proximal qpAdm indicates that Aquae Calidae individuals derive primarily from Anatolian - related sources, with modest contributions from northern/Pontic Gothic proxies and local Balkan substrates. The best - fitting feasible three - way model for the southern subcluster

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(PCA - South, n=12) is ~48% Central Anatolian (Kalehöyük ~2000 BCE) + ~34% Bulgaria Early Iron Age (EIA) + ~18% Chernyakhov (p=0.713, SE ±4 - 6%; Table 1). We acknowledge the substantial temporal gap (~2,300 years) between the Kalehöyük proxy and proposed Late Antique contacts; in the absence of published ancient DNA from Roman - era Cappadocia, this Bronze Age source represents the best available approximation among ~100 ancient Anatolian samples from more than 20 populations tested (spanning Roman, Byzantine, Iron Age, and Bronze Age periods). We note that comparable or greater Anatolian ancestry appears in non - Gothic contexts at Roman - period Balkan sites (e.g., Viminacium), indicating that this Anatolian signal - while robustly present - is not demonstrably Gothic - specific and likely reflects broader Hellenistic - Roman period population movements into the Balkans. Full source-rotation results and retesting are provided in Supplementary Tables S2-S3 and Supplementary Table S2b.

Table 1. qpAdm results for Aquae Calidae South (n=12). Best - fitting feasible models. SE = standard error from jackknife resampling.

Rank	Sources	Proportions (±SE)	p - value
1	Kalehöyük + Bulgaria_EIA + Wielbark	54.0±5.2% + 26.5±4.8% + 19.5±3.9%	0.865
2	Kalehöyük + Bulgaria_EIA	36.6±6.1% + 26.6±6.1%	0.863
3	Kalehöyük + Chernyakhov + Bulgaria_EIA	47.9±5.5% + 18.1±4.2% + 34.0±5.1%	0.713
FAIL	Wielbark only	100%	10^{-202}

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

The northern subcluster (PCA - North, n=10) shows a distinct pattern, preferring Byzantine Marmara (Iznik) over Bronze Age Kalehöyük as the Anatolian source (Table 2). This suggests temporally or geographically distinct incorporation pathways between the two subclusters.

Table 2. qpAdm results for Aquae Calidae North (n=10). Note: no feasible models with Kalehöyük source (distinct from South).

Rank	Sources	Proportions (±SE)	p - value
1	Iznik_Byzantine + Chernyakhov + Chernyakhov_o	43.4±5.8% + 26.9±4.5% + 29.8±5.3%	0.805
2	Iznik_Byzantine + Chernyakhov + Chernyakhov_o + Wielbark	43.4±5.9% + 24.2±4.7% + 29.5±5.4% + 2.9±3.1%	0.755
FAIL	Chernyakhov_o only	100%	0.008

Individual - level modeling (Figures 4C and 4D) reveal that no single model fits all Aquae Calidae individuals. Outliers include individuals modeled as nearly 100% Bulgaria_LAntiquity (p=0.681) - potentially representing local recruits - and others with elevated Near Eastern affinity (proxied by Ancient Egypt or Iraq_Shanidar Neolithic, ~11 - 21% in mixtures).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

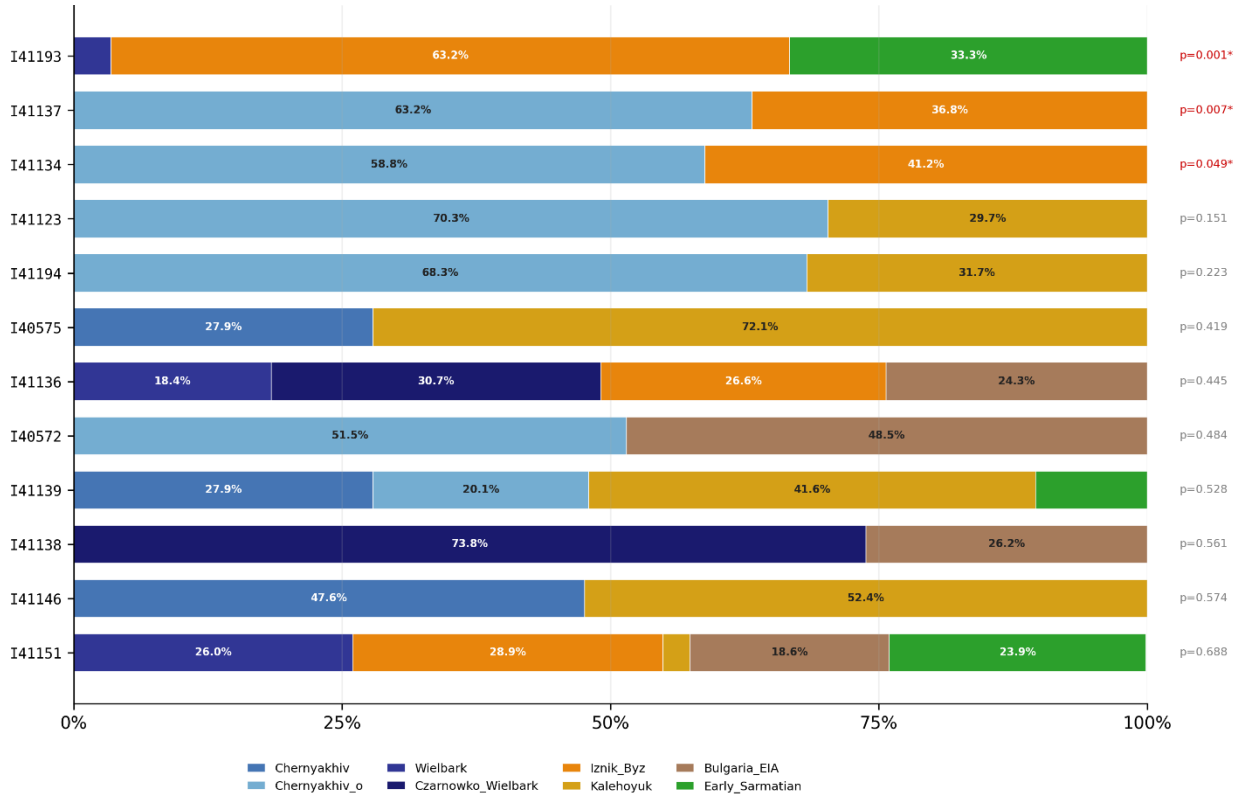
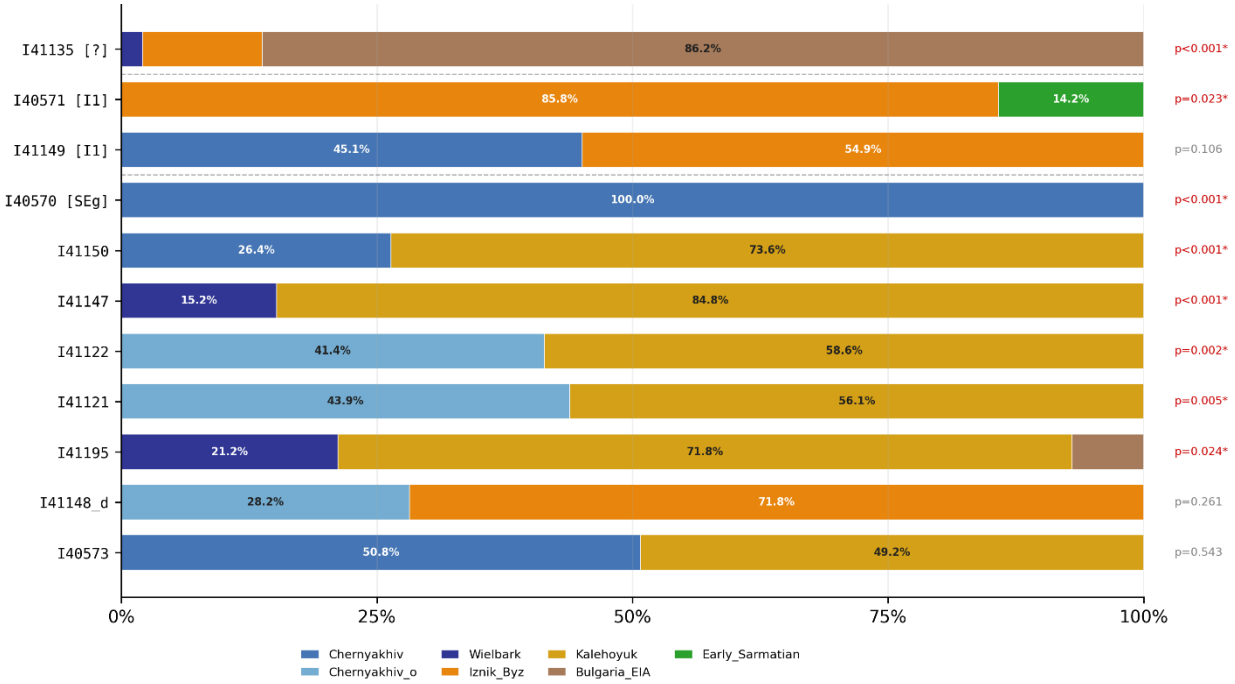


Figure 4C. Individual-level proximal qpAdm models for Aquae Calidae subclusters. (C) North subcluster (n=13). Each bar shows the best-fitting model per individual, sorted by p-value. Models rejected at $p < 0.05$ are marked with asterisks. Source populations include Chernyakhov, Wielbark, Czarnowko_Wielbark, Iznik_Byz, Kalehöyük, Bulgaria_EIA, and Early_Sarmatian. Note the diversity of source preferences even within this subcluster, with some individuals requiring Kalehöyük while others fit Chernyakhov-dominated models (Supplementary Table S3).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



(D) South subcluster, I1 Y-haplogroup carriers, and unassigned individuals. The South subcluster (below dashed line) shows predominantly Kalehöyük-derived ancestry, while I1 Y-haplogroup carriers display mixed Chernyakhov + Iznik_Byz profiles. Most South individuals have rejected models ($p < 0.05$), indicating that simplified proximal models do not fully capture their complex ancestry; these individuals require the Anatolian-enriched models detailed in Tables 1–3.

Model sensitivity analysis (Table 3) demonstrates that Anatolian ancestry is the essential component: models lacking Anatolian sources fail catastrophically ($p < 10^{-149}$), while removal of local Balkan substrates produces only moderate degradation and Serbian Roman - period sources prove redundant.

Table 3. Model sensitivity analysis for Aquae Calidae South. Models (left + right), proportions, standard errors, and p - values.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Model Configuration	**Sources and Proportions (±SE)**	**p - value**
Full model	Kalehöyük 47.9±5.5% + Chernyakhov 18.1±4.2% + Bulgaria_EIA 34.0±5.1%	0.713
Full + Serbia_Rit	Kalehöyük 46.8±5.6% + Chernyakhov 17.5±4.3% + Bulgaria_EIA 33.2±5.2% + Serbia 2.5±3.8%	0.713
Replace Kalehöyük with Iznik	Iznik 31.2±6.3% + Chernyakhov 22.4±4.8% + Bulgaria_EIA 46.4±5.9%	0.293
Remove Bulgaria_EIA	Kalehöyük 55.3±4.8% + Chernyakhov 44.7±4.8%	0.200
Remove all Anatolian sources	Chernyakhov + Bulgaria_EIA	10^{-149}
Germanic sources only	Wielbark 100%	10^{-216}

2.2.2 AKO Ancestry Profiles and Temporal Dynamics

At AKO, proximal qpAdm indicates a core deriving from Pontic multi - ethnic sources, with substantial heterogeneity. Population averages model as 60 - 78% Chernyakhov (or equivalent Wielbark rotations) + 15 - 30% Anatolian/Near Eastern + 5 - 15% steppe (Sarmatian/Scythian proxies; p=0.479 - 0.963). Distal modeling tracks temporal transformation across AKO phases (Figure 4A): EHG (used here as a diagnostic proxy for northern European heritage) declines from 27.9±3.2% in C1 (~350 CE, pre - Hunnic) to 24.4±2.8% in C4 (~376 - 450 CE, Hunnic hegemony) to 4.6 - 16.2±4.1% in C5 (~454 - 489 CE, post - Nedao). Concurrently, Anatolian Farmer ancestry (Turkey_N) rises from ~21% (C1/C2) to ~32 - 50% (C5), reflecting dilution via local admixture. Individual - level qpAdm applied uniformly across phases reveals substantial intra - phase diversity (Supplementary Table S3). Applying simplified one - , two - , and three - way models to each

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

individual shows that: C1 individuals (n=1) fit Wielbark + Anatolian two - way models (66% + 34%, p=0.556) but reject single - source Chernyakhov; C2 individuals (n=2) range from 79% Wielbark + 21% Near Eastern to 55% Wielbark + 45% Anatolian; C4 European individuals (n=6) show three distinct profiles - some fitting single - source Chernyakhov (100%, p=0.574), others requiring Wielbark + Anatolian two - way models, and still others fitting Chernyakhov + Zagros - related (Iraq_Shanidar) models; C4 individuals with East Asian ancestry (n=3) are heterogeneous, with two fitting predominantly steppe (Scythian) models and only one requiring direct Hunnic admixture (Hungary_Hun ~12%); and C5 individuals (n=3) require substantial Mediterranean and Pannonian components (Hungary_IA_LaTene ~16 - 52%), with reduced Gothic ancestry (11 - 30% Wielbark/Chernyakhov).

To estimate the proportion of Hunnic - related ancestry in the C5 sample (n=3), we forced Hungary_Hun as a source in qpAdm models. The estimated proportion was 1.2±3.4% (SE), consistent with zero (0.35 SE from zero; 95% CI: −5.5% to 7.9%). The small sample size precludes definitive conclusions.

f4 statistics and reference population validation (Supplementary Table S6). Systematic f4 tests of the form f4(Mbuti, Anatolian; Aquae Calidae, Chernyakhov) confirmed excess Anatolian affinity in PCA - South individuals beyond the Chernyakhov baseline (Gordion Z=−3.11, Kalehöyük Z=−2.09, Mardin Z=−2.61), while PCA - North showed weaker or non - significant signals (Z=−0.65 to −1.72). This North - South differentiation in f4 mirrors the qpAdm source preferences.

A potential contradiction between f4 and qpAdm was resolved by testing for substructure within the Chernyakhov reference itself. f4 tests reveal that Chernyakhov_main is significantly Balkan -

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

shifted relative to Central Anatolian sources (f4(Mbuti, Chernyakhov_main; Kalehöyük, Bulgaria_EIA) Z= 2.55), while Chernyakhov_outlier shows no significant preference (Z=−0.74 to 0.39). This internal substructure means that f4 tests using Chernyakhov_main as a reference detect its own Balkan affinity rather than a true source preference in Bulgarian Goths, whereas qpAdm, which explicitly partitions ancestry streams, correctly identifies the requirement for substantial Anatolian - proxied ancestry.

2.3 Admixture Dating (DATES)

To estimate the timing of admixture between northern European and southern (Balkan/Anatolian) ancestral components, we applied the DATES method (Narasimhan et al., 2019), which infers admixture dates from the decay of weighted ancestry covariance along the genome. DATES requires specification of two reference populations representing the mixing sources and a target admixed population; it does not require phased genomes and is applicable to low - coverage ancient DNA data, with sensitivity primarily to the sample size of the target population and the allele frequency contrast between the two reference sources.

We constructed reference panels designed to maximize allele frequency contrast: northern sources comprised Wielbark - associated individuals from three Polish sites (Pruszcz Gdański, Masłomęcz, Kowalewko; combined n=68, designated “Pol3”) or Chernyakhov (n=3); southern sources comprised either a Balkan Iron Age/Late Bronze Age pool (“BalkanCore”: Bulgaria_EIA, Greek IA/LBA, North Macedonian IA; n=44), an Anatolian pool (“AnatoliaMax”: Aegean, Central, and Eastern Anatolian populations; n=37), or Aegean Anatolian specifically (“Mugla3”: Roman/Byzantine - period Muğla region; n=23). Target populations were KalideN (the northern subcluster of Aquae Calidae, n=12), AKO (n=11), and the pooled Gothic - associated sample (KalideN+AKO, n=23). Results are summarized in Table 4 and Figure 5; the full set of tested

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

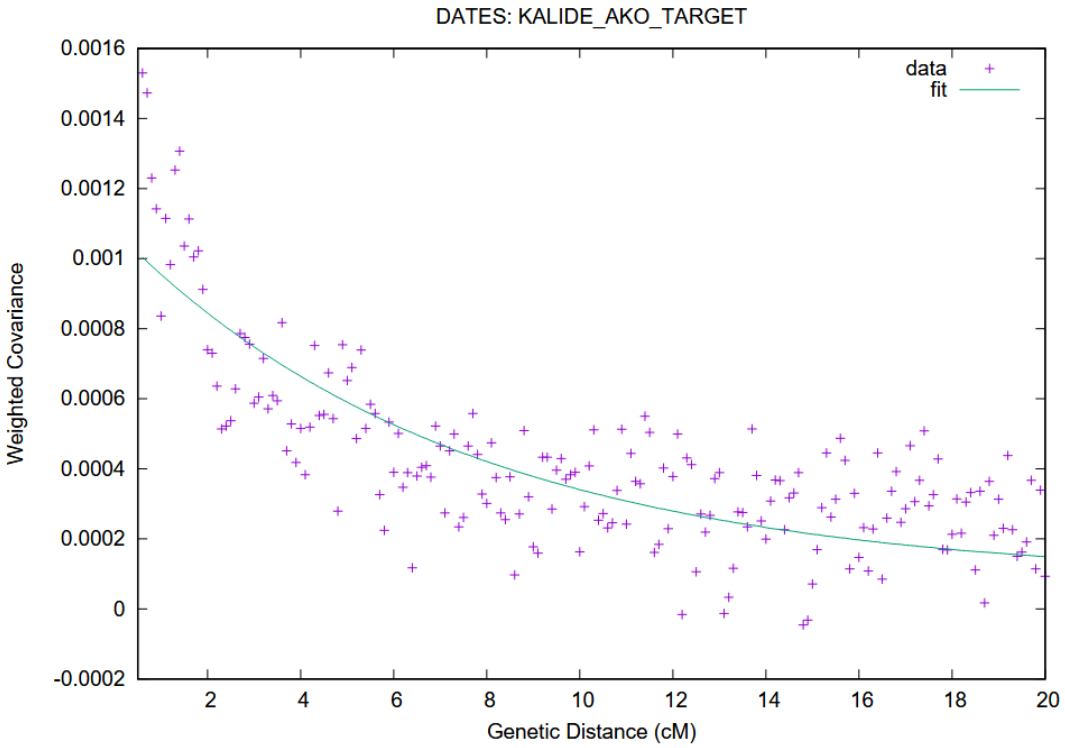
configurations, including sensitivity tests and control populations, is provided in Supplementary Table S5 and Supplementary Note S3.

Table 4. DATES admixture dating results. Top models ranked by Z-score. Generation time = 29 years. nrmsd: normalized root mean squared deviation (model fit quality; values <0.5 indicate acceptable fit).

Target	North	South	Mean (gen)	SE	Z	nrmsd	Point est.	95% CI
KalideN+AKO (n=23)	Pol3 (n=68)	AnatoliaMax (n=37)	12.44	2.36	**5.26**	0.186	~49 CE	85 BCE - 183 CE
KalideN (n=12)	Pol3 (n=68)	Mugla3 (n=23)	12.74	2.70	**4.73**	0.191	~4 CE	136 BCE - 144 CE
KalideN (n=12)	Pol3 (n=68)	BalkanCore (n=44)	12.09	2.61	**4.63**	0.318	~41 CE	97 BCE - 179 CE
KalideN (n=12)	Pol3 (n=68)	AnatoliaMax (n=37)	11.01	2.42	**4.55**	0.242	~41 CE	97 BCE - 179 CE
AKO (n=11)	Chernyakhov (n=3)	Mugla_Stratonikeia (n=10)	17.23	3.94	**4.38**	0.454	~40 BCE	294 BCE - 214 CE
KalideN+AKO (n=23)	Pol3 (n=68)	BalkanCore (n=44)	12.12	2.76	**4.39**	0.221	~49 CE	85 BCE - 183 CE

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Figure 5. DATES admixture dating results. (A) Weighted covariance decay curve for the best - fitting model (KalideN+AKO × Pol3 vs. AnatoliaMax; Z=5.26).



(B) Comparison of admixture date estimates across different southern source configurations, showing convergence at ~12 generations:

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Target × North	South (single)	Mean (gen)	SE	Z	Δ from BalkanCore
KalideN × Pol3	BalkanCore (n=44)	**12.09**	2.61	4.63	baseline
KalideN × Pol3	AnatoliaMax (n=37)	**11.01**	2.42	4.55	-1.08 gen (younger)
KalideN × Pol3	Mugla3 (n=23)	**12.74**	2.70	4.73	+0.65 gen (older)
KalideN+AKO × Pol3	BalkanCore (n=44)	**12.12**	2.76	4.39	baseline
KalideN+AKO × Pol3	AnatoliaMax (n=37)	**12.44**	2.36	5.26	+0.32 gen (same)
KalideN+AKO × Pol3	Mugla3 (n=23)	**14.65**	3.14	4.67	+2.53 gen (older)

BalkanCore and AnatoliaMax give **very close, nearly identical timing** (~12 gen, difference <1 gen, well within SE of ~2.5 gen). This seems to suggests **AGAINST two separate pulses**.

What Happens When South Is Pooled (Balkan + Anatolia)?

Target × North	South Pool	Mean	SE	Z	Effect
KalideN+AKO × Pol3	BalkanCore only	12.12	2.76	4.39	baseline
KalideN+AKO × Pol3	AnatoliaMax only	12.44	2.36	5.26	baseline
KalideN+AKO × Pol3	BalkanCore + Mugla3 UNION	14.24	4.83	2.95	+2 gen, SE +75%
KalideN+AKO × Pol3	BalkanCore + AnatoliaMax UNION	15.54	5.12	3.04	+3 gen, SE +85%

Full DATES decay-curve diagnostics, sensitivity tests, and alternative reference-panel configurations are documented in Supplementary Note S3 and Supplementary Table S5.

Three principal findings emerge from the DATES analysis:

Robust admixture signal. Across all well - fitting models with Wielbark northern sources, admixture dates converge on ~11 - 13 generations before burial, with Z - scores ranging from 4.39 to 5.26. The best overall model (pooled KalideN+AKO target, Pol3 northern source, AnatoliaMax southern source) yields 12.44±2.36 generations (Z=5.26). For burials dated to ~360 CE (KalideN) and ~460 CE (AKO), this places admixture in the 1st century BCE to early 2nd century CE range - predating the earliest historically recorded Gothic intrusions into the eastern Roman Empire (~170 CE) and suggesting that at least some of the admixture may have occurred outside Bulgaria, potentially during the Chernyakhov formation in the Pontic region or during trans - Danubian frontier mixing after the establishment of Roman Dacia (post - 106 CE).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Signal specificity. Adding Scandinavian and Danish Iron Age samples to the northern reference panel systematically shifted admixture dates older by 3 - 5 generations and inflated standard errors (e.g., Pol3+Scandinavia vs. BalkanCore: 14.63±4.59 gen, Z=3.19, compared with Pol3 alone: 12.09±2.61 gen, Z=4.63). This pattern is consistent with the detected admixture signal being specifically associated with Polish Wielbark - like ancestry rather than generic northern European Iron Age ancestry. Adding Polish Iron Age sites associated with non - Gothic cultures (notably Przeworsk, a contested Protoslavic or Vandal culture) also degraded inference, possibly because pre - existing southern admixture in those populations reduced the north - south allele frequency contrast. The complete sensitivity matrix (including Scandinavian-expanded northern panels and alternative southern proxies) is provided in Supplementary Note S3 and Supplementary Table S5.

Balkan and Anatolian sources yield indistinguishable timing. When tested individually as southern sources, BalkanCore (12.09 gen, Z=4.63), AnatoliaMax (11.01 gen, Z=4.55), and Mugla3 (12.74 gen, Z=4.73) all produce highly significant signals with timing estimates differing by less than two generations - well within the standard error of ~2.5 generations. Pooling Balkan and Anatolian sources into a single “union - south” reference slightly degraded inference (Z dropped to ~3, SE approximately doubled), consistent with violation of the two - way admixture model assumption. These results indicate that the Balkan and Anatolian components entered the Gothic - associated gene pool at similar times, consistent with admixture involving an already - blended Balkan - Anatolian substrate rather than two temporally separable pulses.

The two target populations show different optimal northern proxy preferences: KalideN favors Wielbark (Z=4.34 vs. 2.21 for Chernyakhov as northern source), while AKO favors Chernyakhov (Z=4.38 vs. 2.64 for Wielbark). This likely reflects their ancestry compositions: KalideN, with its larger southern base and minority northern component, creates stronger allele - frequency contrast

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

with Wielbark, producing crisper DATES signals, while AKO's more evenly distributed components (Chernyakhov itself being partially admixed) reduce contrast. When tested as a southern proxy for AKO, KalideN failed (Z<1), ruling out an ancestor - descendant relationship between the two groups.

The southern-shifted subcluster at Aquae Calidae (KalideS: the southern subcluster, n=5) produced weaker DATES signals (Z~2.7 - 3.2, poorer fit quality), consistent with these individuals having minimal northern European ancestry (~10 - 20% per qpAdm) and thus insufficient allele - frequency contrast for robust inference. This within - site heterogeneity, where individuals sharing identical Gothic material culture produce different admixture signals, itself constitutes evidence consistent with ethnogenesis.

Control tests on non - Gothic populations. To assess whether the ~12 - generation admixture clock is specific to Gothic - associated targets or represents a generic feature of Late Antique Balkan populations, we applied identical DATES contrasts to Roman - period populations from outside Gothic archaeological contexts. Serbia_Viminacium_Roman_Rit (n=3; Olalde et al., 2023), a published Roman - period ritual context, failed all DATES contrasts: Pol3 vs. BalkanCore, Pol3 vs. AnatoliaMax, and BalkanCore vs. AnatoliaMax all yielded Z<1.8 with implausible or negative mean estimates and poor fit quality (nrmsd>0.66), indicating that these individuals do not carry a detectable two - way admixture signal of the type observed in Gothic targets. We additionally tested unpublished Roman and Late Antique period individuals from Bulgarian sites (n=16 and n=3 respectively; data to be presented in a forthcoming study). These populations similarly failed to produce stable admixture signals under Gothic - relevant contrasts: the larger group (n=16) yielded ancient - looking dates (>100 generations) under substrate - style (BalkanCore vs. AnatoliaMax) contrasts consistent with deep population heterogeneity rather than a single recent

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

pulse, while Pol3 - based contrasts returned either negative estimates or weak signals ($Z<2.3$) in the wrong temporal range. The smaller group (n=3), despite plausible archaeological associations with Migration Period populations, proved too heterogeneous for pooled analysis: per - individual runs revealed sensitivity to northern proxy choice, with different individuals aligning better with Wielbark versus steppe/Scythian - like northern references, precluding stable pooled inference. Full control-population outputs and diagnostics are provided in Supplementary Table S5 and Supplementary Note S3.

The pattern across these controls is informative: the clean ~12 - generation admixture clock with high Z - scores (4.4 - 5.3) is not a generic property of Roman - period or Late Antique Balkan populations under DATES analysis. Rather, it appears specifically in Gothic - associated targets under Wielbark - like northern contrasts. This specificity strengthens the interpretation that the DATES signal in Gothic - associated populations reflects a discrete admixture event involving Wielbark - related ancestry, even if the geographic location and ethnic context of that event remain ambiguous.

2.4 Uniparental Markers and Kinship

2.4.1 Y - Chromosome Haplogroups

Y - haplogroup assignments were determined for 19 of 24 males (79%) using yHaplo (terminal SNPs with ≥3× coverage; ISOGG v15.73). The overall assemblage is cosmopolitan: only ~23% carry Germanic - associated lineages (I1, R - DF90), below expectations for a biologically homogeneous Gothic population (Fig. 6).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

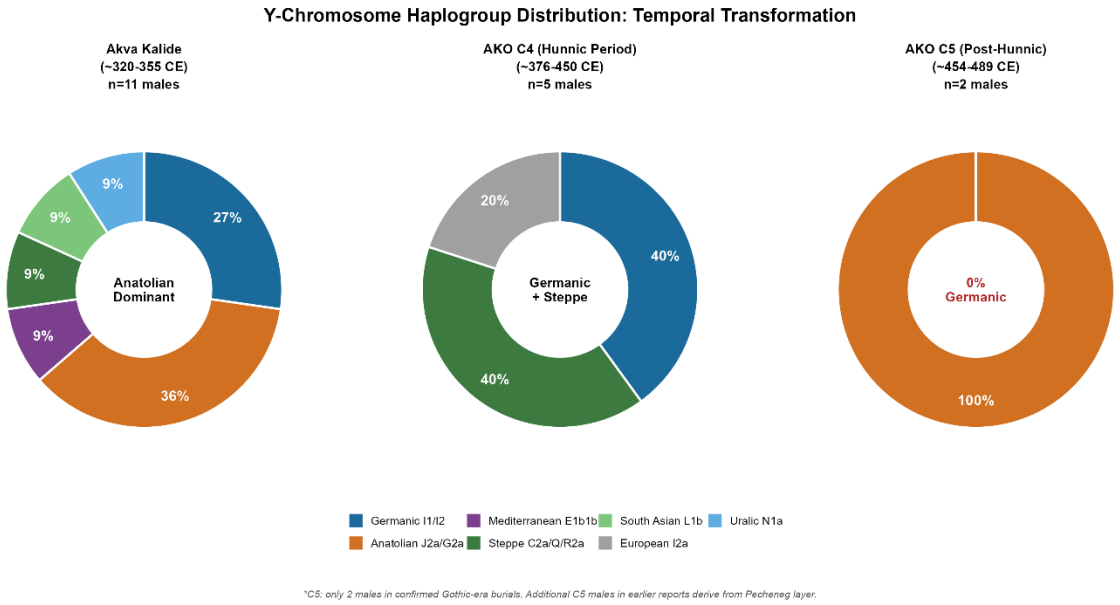


Figure 6. Y-chromosome haplogroup distribution across Gothic populations from Aquae Calidae and AKO by temporal phase.

The distribution differs between sites and phases. Aquae Calidae shows ~36% Anatolian/Near Eastern (J2a, G2a), ~29% Germanic - associated (I1 subclades, R - DF90), ~18% Balto - Slavic (R1a branches), and ~12% Steppe/East Asian (C2a, Q1b). AKO exhibits phase - specific patterns: C1/C2 is dominated by Northern European and Eastern European lineages (~50% I1/I2; ~30% R1a); C4 shows a mixed composition (~40% North European, ~30% Steppe/East Asian); and C5 shows Anatolian/Near Eastern lineages in both males with determined haplogroups (2/2 carrying G2a and J2a, respectively; given the small sample, the 95% CI for the Anatolian haplogroup frequency ranges from 16% to 100%).

Integration with nuclear ancestry profiles (per - individual comparison). To assess correspondence between uniparental and autosomal signals, we examined the nuclear ancestry profiles of individuals carrying specific Y - haplogroup categories. The two C5 males with Anatolian Y - haplogroups (G2a, J2a) both show nuclear profiles dominated by Mediterranean/Anatolian ancestry (qpAdm: 37 - 54% Turkey_N/Marmara, 16 - 52%

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Hungary_IA_LaTene, 11 - 30% Wielbark/Chernyakhov), consistent with concordance between paternal lineage and autosomal composition. The R1a carrier is best modeled as Chernyakhiv (41%) + Iraq_Shanidar (33%) + Bulgaria_EIA (26%), p = 0.807 suggesting Pontic steppe and eastern trajectory requiring no Wielbark or Anatolian Bronze Age source. Confirmed I1 carriers (n=3; two from Aquae Calidae, one from AKO; one additional AKO individual with uncertain I1 subclade assignment was excluded) demonstrate different model: Roman_Byzantine (28%) + Kalehöyük (26%) + Wielbark (22%) + Bulgaria_EIA (15%) + Chernyakhiv (9%), p = 0.695. The Wielbark component , not required in the R1a individual modeling , appeared alongside the Scandinavian I1 haplogroup, suggesting genuine Scandinavian - derived autosomal ancestry in these individuals, combined, however, with substantial Anatolian admixture.

Key lineages in both sites include: Germanic - associated I1 subclades (I - Y6229, I - Y22019) and R - DF90, which are recurrent in Wielbark contexts (~41% I1 in Iron Age Poland; Stolarek et al., 2023) and appear in early AKO phases but are absent from C5; Steppe/East Asian C2a and Q1b, confined to AKO C4 (Hunnic period); and Anatolian/Near Eastern J2a and G2a, dominant at Aquae Calidae (~36%) and exclusively present in AKO C5.

2.4.2 Mitochondrial DNA Haplogroups

mtDNA haplogroups were assigned for all 38 individuals using HaploGrep2 (PhyloTree Build 17). The assemblage is predominantly local Balkan/Aegean (H, J, T2, K, HV, V, X2, W, I; ~55 - 60%), with steppe/forager (U5, U4a2, U2e1f1; ~20 - 25%), Near Eastern/Caspian (U7b, N1a1a3, N1b1a; ~10 - 15%), rare East Eurasian (C5a1; ~3%), and African (L1b1a; ~2%) components.

Aquae Calidae displays an extraordinary degree of maternal diversity: the 20 successfully typed individuals carry 20 different mtDNA haplogroups, with no sharing of any maternal lineage between any pair of individuals. The haplogroup assemblage spans multiple continental

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

affiliations: European H - lineage subclades (H6a1a, H7, H14a), Near Eastern T - lineage variants (T2b4, T2b5, T2d2), the South Asian - associated U7b, Near Eastern N1b1a, the sub - Saharan African L1b1a (I41194) and rare West Eurasian lineages (HV15, I5a2, I5b1) whose modern distributions are confined to small geographic pockets. The complete absence of shared maternal lineages, combined with the absence of detected close kinship among any Aquae Calidae individuals, is inconsistent with a founding population and indicates diverse source populations. In contrast to the Y - chromosome composition, which shifts across AKO phases, mtDNA distributions remain relatively stable (~55% Balkan haplogroups throughout), suggesting possible male - biased migration patterns, though formal f4 tests lack statistical power to confirm sex - bias with current sample sizes.

2.4.3 IBD Kinship Networks

IBD analysis (READv2; ~220,000 pairs evaluated, 64 Gothic - involved; cutoff 0.04) reveals strict internal social structure: 9 intra - Gothic pairs spanning three generations (~80 - 100 years), including 3 first - degree (parent - sibling), 3 second - degree, and 3 third - degree relationships, all within sites. No first - or second - degree kinship was detected between the two Gothic sites or between Gothic individuals and external reference populations (55 external pairs, all 3rd - degree or more distant), indicating nuclear - family endogamy despite shared material culture. Full pairwise kinship calls, network statistics, enrichment calculations, and sensitivity analyses are provided in Supplementary Note S2 and the Supplementary IBD Workbook.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

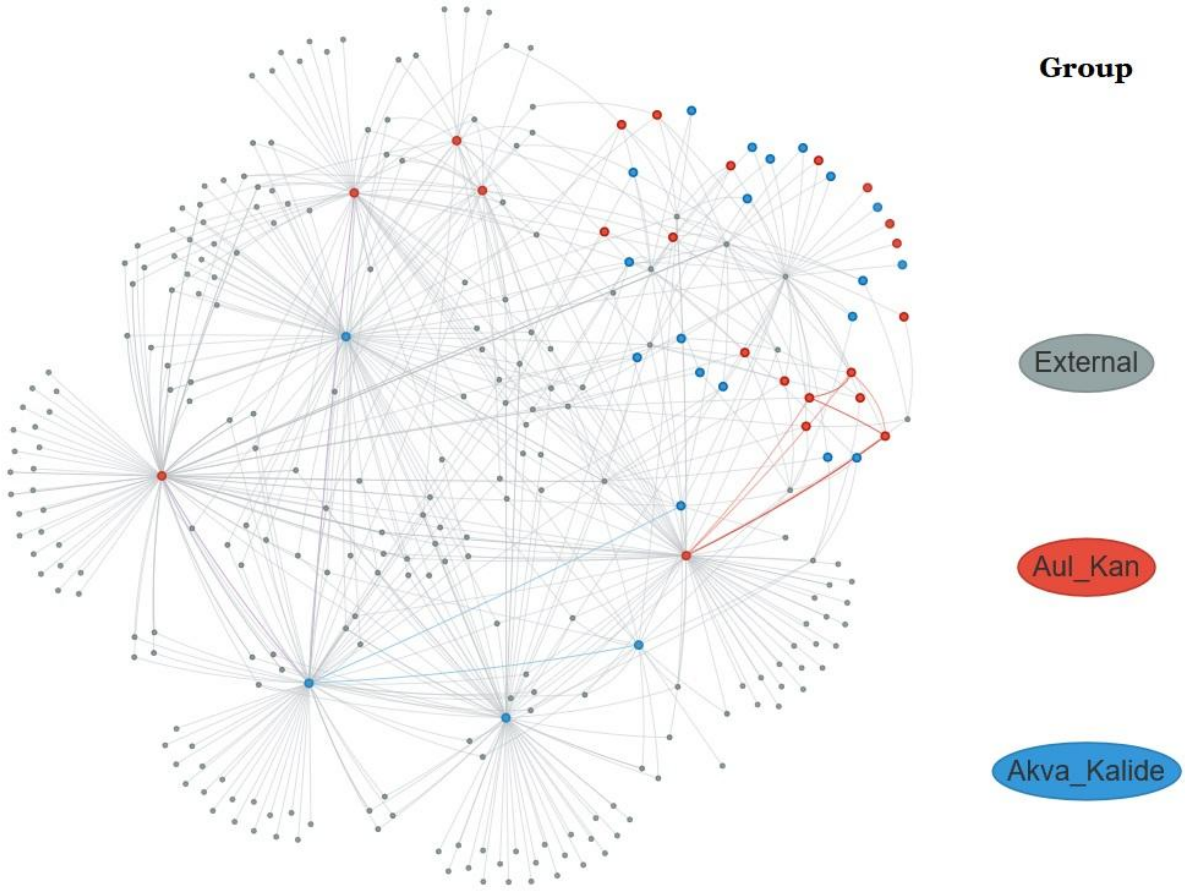
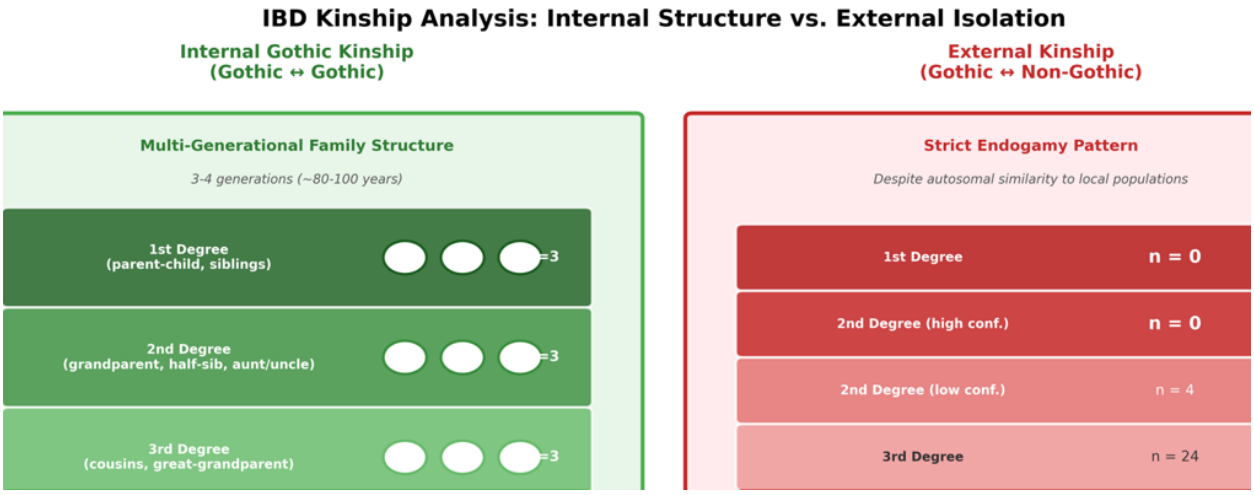


Figure 7. IBD kinship network analysis. (A) Network visualization showing IBD connections; node colors represent population groups, edge weights indicate kinship degree.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.



(B) Population - normalized enrichment values showing bimodal temporal connectivity.

Population - normalized enrichment analysis reveals bimodal temporal connectivity: strong enrichment with prehistoric Balkan populations (Chalcolithic: 20 - 30× enrichment) and later Medieval Balkan populations (4 - 11×), but weak connectivity with contemporary Late Antiquity Balkan populations. This pattern - elevated IBD sharing with temporally distant rather than contemporary Balkan populations - is consistent with admixture with local Balkan populations having occurred prior to settlement in Bulgaria (e.g., during 2nd - 4th century migrations through southeastern Europe), followed by relative social isolation. However, this interpretation assumes adequate Late Antique Balkan representation in reference panels; sampling gaps or population bottlenecks could produce similar patterns through reduced contemporary IBD detectability. Sensitivity analyses varying kinship cutoffs and software parameters confirm the robustness of both the endogamy finding and bimodal connectivity pattern (Supplementary Note S2).

3. Discussion

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

3.1 Genetic Heterogeneity of Cultural Goths

The most striking finding of this study is the extreme genetic heterogeneity among individuals buried with diagnostic Gothic material culture. At the population level, qpAdm models for Gothic - associated individuals range from ~100% Chernyakhov to ~85% Anatolian - related, encompassing northern Germanic (Wielbark - derived, ~5 - 28% in different individuals), Pontic multi - ethnic (Chernyakhov, ~18 - 100%), Anatolian (Kalehöyük/Marmara proxies, ~15 - 85%), steppe (Scythian/Sarmatian, ~2 - 67%), limited East Asian (Hunnic - related, ~3 - 12%), and local Balkan (Bulgaria_EIA/LAntiquity, ~20 - 40%) ancestries. Uniform model fitting across all individuals confirms that no single ancestry model fits the entire assemblage: individuals that accept single - source Chernyakhov models are found alongside others requiring substantial Anatolian or steppe components (Heather, 1986; Stolarek et al., 2019). The individual-level qpAdm accept/reject matrix and full model list underlying these summaries are provided in Supplementary Table S3.

This heterogeneity is consistent with historical accounts of Gothic ethnogenesis as a process of progressive accretion. Goths originating from Scandinavian - related populations in the Wielbark culture (1st - 4th centuries CE) merged with local Dacians, Sarmatians, and Alans during the formation of the Chernyakhov culture (late 2nd - 5th centuries CE), creating a multi - ethnic confederation. Our data suggest that this accretion continued - and perhaps intensified - during the Balkan settlement phases, incorporating individuals of predominantly Anatolian ancestry alongside those retaining substantial northern European ancestry. Three individuals illustrate the multi - ethnic character of these Gothic communities at the biographical level. I41205 (AKO, C4) carries Y - haplogroup C2a1a1b1b1 - East Asian lineage whose closest modern parallels occur among Buryats and Khalkha Mongols - combined with mtDNA J1c1b1a (European) and 6.5%

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

East Asian autosomal ancestry. I41194 (Aquae Calidae) combines a J2a Y - haplogroup - characteristic of Anatolian populations with mtDNA L1b1a, a sub-Saharan African lineage. His genome indicates the African diaspora's presence within Late Antique Mediterranean mobility networks and the capacity of Gothic identity to encompass individuals of diverse biological ancestry. I40570 (Aquae Calidae) shows approximately 50% Egyptian ancestry (modeled as Third Intermediate Period component) with J1 Y-haplogroup (Arabian Peninsula-associated) and falls entirely outside the main Aquae Calidae PCA cluster. This individual, whose genetic profile suggests Nile Valley or Levant rather than European population from the Migration Period, was nonetheless buried among Goths—the most extreme example in our current dataset of biological ancestry diverging from cultural identity.

The finding that individuals with dramatically different ancestry compositions share identical Gothic material culture, burial practices, and Arian Christian rites provides molecular evidence consistent with ethnogenesis theory as articulated by Wolfram (1988) and Pohl (2013): Gothic identity functioned as a permeable cultural - political category maintained by institutional frameworks rather than as a marker of common biological descent. This interpretation accords with the broader pattern observed in other Migration Period groups, where steppe - associated or Germanic - associated signals vary substantially by phase and site, as documented for Hunnic (Maenchen-Helfen, 1973) and Avar (Gnecchi-Ruscone et al., 2022) elites.

3.2 Distinct North/South Genetic Clusters

Despite shared material culture, the two sites form genetically distinct clusters. AKO individuals cluster with Migration Period Germanic and Pannonian groups in PCA, reflecting predominant Wielbark/Chernyakhov - derived ancestry (60 - 78% in qpAdm models), while Aquae Calidae individuals cluster with Balkan Late Antiquity populations and carry substantial Anatolian -

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

related ancestry (50 - 85%). Within Aquae Calidae, further substructure emerges: PCA - South individuals require Bronze Age Central Anatolian sources (Kalehöyük) while PCA - North individuals preferentially match Byzantine Marmara/Iznik references, suggesting temporally or geographically distinct incorporation pathways within a single site.

This north - south differentiation is corroborated by multiple independent lines of evidence. f3 - outgroup statistics show Aquae Calidae's top affinity with Bulgaria_LAntiquity (rank #1) compared to AKO's much weaker local Balkan signal (rank #21). f4 statistics confirm excess Anatolian affinity in PCA - South beyond the Chernyakhov baseline ($Z=-2.09$ to -3.11 for various Anatolian sources), while PCA - North shows weaker signals. The DATES analysis shows that the two sites have different optimal northern proxy preferences - KalideN favoring Wielbark, AKO favoring Chernyakhov - and that KalideN cannot serve as a proxy for AKO ancestry ($Z<1$), ruling out a simple ancestor - descendant relationship. Formal f4 results are in Supplementary Table S6; full DATES target/source combinations and diagnostics are in Supplementary Table S5 and Supplementary Note S3.

The Anatolian ancestry at Aquae Calidae is robustly established through extensive source rotation testing (~200 ancient samples from >20 populations): models without Anatolian sources fail ($p < 10^{-149}$). However, the specific historical pathway that introduced this ancestry remains ambiguous. Possible explanations include descent from Cappadocian Christians captured during 3rd - century Gothic raids on Anatolia (documented by Philostorgius), pre - existing Anatolian ancestry in local Roman - era Thracian provincials, incorporation of eastern Roman military or administrative personnel of Anatolian origin, or mixed origins combining several pathways. Anatolian - shifted ancestry is attested in other Roman - period Balkan populations (e.g., Viminacium, Serbia), indicating that Anatolian demographic contributions to the Balkans were not restricted to Gothic -

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

associated contexts. However, DATES control tests reveal an important distinction: while Anatolian ancestry may be broadly distributed in Roman - period Balkan populations, the specific admixture signature - a ~12 - generation - old mixture of Wielbark - related northern ancestry with a Balkan - Anatolian substrate - is not replicated in these non - Gothic groups (Section 2.3). This suggests that the pathway by which Anatolian ancestry entered Gothic - associated populations was distinct from the general Hellenistic - Roman era Balkan mixing process, even if the ultimate Anatolian source populations may have been similar. Strontium isotope analysis, not yet available for these samples, could further distinguish locally born individuals from first - generation migrants.

3.3 Temporal Dynamics at AKO

AKO's sequential necropoleis (~150 years) document progressive transformation. EHG declines from 27.9±3.2% in C1 (~350 CE) to 24.4±2.8% in C4 (~376 - 450 CE) to 4.6 - 16.2±4.1% in C5 (~454 - 489 CE), with corresponding increases in Anatolian Farmer ancestry (~21% → 32 - 50%). This ~70 - 80% dilution of the northern European signal over approximately five generations reflects continuous local admixture. Y - chromosome composition shifts from ~50% Germanic - associated lineages in early phases (C1/C2) to exclusively Anatolian/Near Eastern haplogroups in C5 (both confirmed males: G2a, J2a), while mtDNA distributions remain relatively stable (~55% Balkan haplogroups throughout), suggesting possible male - biased demographic processes.

We note potential circularity in linking C5 dating (based partly on Zeno coins, 474 - 491 CE) to Theodoric's departure and then interpreting C5's genetic shift as evidence for that departure. Without independent radiocarbon dating, this correlation remains suggestive. Two historical scenarios merit consideration. The first associates AKO's later phases with Theodoric the Amal, who resided in lower Moesia from 474 to 488. However, Heather (1991, 1996) argues that

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Theodoric's chronological window in the eastern Balkans was narrow (perhaps 477 - 478 and 483 - 487/8), offering limited opportunity for the progressive admixture documented in our data. The alternative associates the site with Thracian Gothic foederati established in the region since at least the 440s (Heather, 1996) under Theoderic Strabo (son of Triarius), whose communities numbered over 10,000 warriors (Heather, 1991) and were deeply integrated into eastern Roman society over 40 - 60 years. The genomic trajectory - progressive dilution rather than abrupt replacement - is more consistent with long - term local integration. The archaeological/numismatic basis of phase dating and its uncertainty bounds are summarized in Supplementary Note S4.

3.4 Admixture Timing, Geographic Constraints, and the Trans - Danubian Frontier

The DATES analysis provides an independent chronological constraint on the mixture of northern and southern ancestral components, and that constraint carries a geographic implication that has not, to our knowledge, been articulated for Migration Period populations. The robust signal at ~12 generations before burial (Z=4.39 - 5.26 across multiple configurations) places admixture in the 1st century BCE to early 2nd century CE - earlier than the earliest historically recorded Gothic intrusions into the eastern Roman Empire (~170 CE, during the Marcomannic Wars - Heather, 1986). If the admixture indeed occurred before Goths are attested south of the Danube, the event itself cannot have taken place on Balkan soil, and we must look to regions where Wielbark - related populations could have encountered Balkan - Anatolian - shifted groups prior to that date (Heather, 1986). Given the broad confidence intervals and the approximate pulse-admixture assumption of DATES, we treat this geographic implication as a working hypothesis; sensitivity analyses and assumptions are discussed in Supplementary Note S3, and chronological anchoring in Supplementary Note S4.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

The control tests reinforce this point. The ~12 - generation clock is not a generic feature of Late Antique Balkan populations: Serbia_Viminacium_Roman_Rit fails all DATES contrasts, and additional unpublished Roman - period Bulgarian populations (to be presented separately) similarly fail to replicate the signal. If the admixture signature reflected a general Hellenistic - Roman era Balkan mixing process - one that should have affected provincial populations broadly - we would expect to recover analogous signals in non - Gothic Roman - period groups. The failure to do so indicates that the specific combination of Wielbark - related northern ancestry with a Balkan - Anatolian substrate, dated to ~12 generations, is confined to (or at least strongly concentrated in) populations associated with Gothic material culture, even if the admixture event that produced it was not itself "Gothic" in an ethnic sense.

Several interpretive frameworks can accommodate these constraints, and we present them in order of decreasing compatibility with the temporal and geographic evidence.

The scenario most consistent with the DATES temporal constraint involves trans - Danubian frontier mixing in the territory of Roman Dacia. In 106 CE, Trajan's conquest of the Dacian Kingdom created a Roman province north of the Danube, extending deep into the Carpathian basin. The new province received a documented influx of settlers from Anatolia and the central Balkans - a colonization process attested epigraphically and archaeologically (Oltean, 2007) - creating precisely the kind of Balkan - Anatolian blended substrate that DATES detects as the southern component in Gothic - associated targets. This mixed provincial population coexisted in a long frontier zone with indigenous "barbarian" groups: Dacians (a Thracian - related people), Sarmatians, Carpi (whose ethnic classification remains debated), and Bastarnae, some of whom may have included early Germanic - speaking elements. The Wielbark expansion into the northern Pontic region during the 2nd century CE, culminating in the formation of the Chernyakhov culture,

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

brought populations carrying specifically Wielbark - related ancestry into proximity with these trans - Danubian frontier communities. The DATES point estimate of ~50 CE (95% CI: 85 BCE - 183 CE) is compatible with admixture occurring near this window - after the establishment of Roman Dacia and its associated population transfers, but before the major Gothic - Roman conflicts of the mid - 3rd century. When Romans abandoned Dacia under Aurelian (~271 - 275 CE), much of the provincial population was relocated south of the Danube into what became the province of Dacia Mediterranea (roughly corresponding to modern northern Bulgaria), while the Tervingi settled the evacuated territory. Populations already carrying the Wielbark × Balkan - Anatolian admixture signature could thus have entered the Balkans through either pathway: as part of the Aurelian relocation, or as members of the Gothic Tervingi polity that subsequently crossed the Danube during the crises of 376 - 382 CE.

An alternative interpretation places the admixture within the Chernyakhov formation (late 2nd - 3rd century CE) in the Pontic region, where Gothic - associated populations mixed with local groups that may have already carried Balkan - Anatolian - shifted ancestry through earlier steppe - Mediterranean interactions. This scenario is plausible but faces the difficulty that well - characterized Chernyakhov populations (from Ukrainian sites) do not themselves show the strong Anatolian signal that dominates Aquae Calidae, suggesting that the southern component was acquired through a pathway distinct from the main Chernyakhov synthesis. The fact that the DATES signal is specifically associated with Wielbark - like rather than generic northern European ancestry (Scandinavian sources shift dates older by 3 - 5 generations and inflate standard errors) nonetheless supports a connection to the specific Wielbark → Chernyakhov migration trajectory (Kulikowski, 2007; Stolarek et al., 2019).

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

A third possibility - that the admixture occurred after Gothic settlement south of the Danube, with the ~12 - generation estimate reflecting averaging across multiple admixture episodes - cannot be excluded given the broad confidence intervals. However, this scenario would require explaining why the resulting signal so closely tracks Wielbark - specific ancestry (rather than the broader Germanic ancestry pool available in the 4th - century Balkans) and why comparable signatures are absent from non - Gothic populations in the same region.

The finding that Balkan and Anatolian southern sources yield statistically indistinguishable admixture timing (BalkanCore: 12.09 gen; AnatoliaMax: 11.01 gen; difference < 1.1 gen, well within SE of ~2.5) is consistent with all three scenarios but most naturally explained by the Roman Dacia model, where Anatolian settlers and Balkan provincials were mixed together in Dacia before encountering Wielbark - related populations, producing a substrate that DATES cannot decompose into separate Balkan and Anatolian pulses because they were already blended prior to the admixture event. The slight degradation of inference when Balkan and Anatolian sources are pooled into a single southern reference (Z drops from ~5 to ~3) is consistent with some internal structure within this substrate, but the similar timing argues against two cleanly separable admixture events.

We acknowledge that DATES estimates may be affected by multiple layers of admixture, potentially producing composite signals that are difficult to decompose into individual events, and that the confidence intervals (95% CI spanning ~250 years for the best model) preclude precise historical attribution. The method dates when ancestry - specific linkage disequilibrium was created - that is, when segments from the northern and southern ancestral populations became associated in the same genomes - but cannot independently determine the geographic location of that event. Nonetheless, the analysis provides the first quantitative constraint on admixture timing

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

in Migration Period Gothic - associated populations and, together with the control tests demonstrating that the signal is not a generic Balkan feature, narrows the range of historically plausible scenarios.

3.5 Institutional Maintenance of Gothic Identity

How did ancestrally diverse populations - Aquae Calidae's predominantly Anatolian - ancestry individuals, AKO C5's individuals with Anatolian Y - haplogroups yet Gothic burial rites - maintain a shared cultural identity? The absence of a unifying genomic profile points toward superstructural frameworks: Arian Christianity (distinguished from Nicene orthodoxy, with episcopal centers like AKO's basilicas facilitating integration); Gothic language and material culture; and foederati political - legal status (tribal autonomy under the 382 CE treaty enabling incorporation without cultural dissolution).

3.6 Limitations

Several constraints affect interpretation. Sample sizes for individual phases, particularly C5 (n=3), limit statistical power. Reference panel gaps - notably the absence of Roman - era Central Anatolian ancient DNA and limited geographic coverage of published Chernyakhov samples (all from Ukrainian sites; Moldova - Romania / Roman Dacia substructure unassessed) - required the use of temporally distant proxies. Low genome coverage (~0.5 - 1.5× for 60% of samples) affects Y - chromosome subclade resolution and IBD sensitivity. The absence of direct radiocarbon dating introduces temporal uncertainty of ±25 - 50 years per phase. DATES admixture estimates may be complicated by multiple admixture layers and assume pulse admixture, which is an approximation for populations experiencing continuous gene flow. The DATES method constrains the timing of admixture but cannot independently determine its geographic location; the Roman Dacia

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

interpretation proposed in Section 3.4 is the scenario most consistent with the temporal evidence but remains a hypothesis pending confirmation from ancient DNA sampling in the trans - Danubian frontier region itself. Control tests on non - Gothic populations strengthen the specificity of the Gothic - associated DATES signal, but are limited by small sample sizes (Serbia_Viminacium_Roman_Rit, n=3) and the preliminary status of some Bulgarian Roman - period reference populations; expanded control sampling from published Roman - period Balkan datasets would further test the robustness of this specificity. Yet, genetic heterogeneity consistent with ethnogenesis theory does not definitively prove that model; similar patterns could arise from military conscription, serial migrations, or exogamous practices without shared ethnic identification. Complete validation requires integration with textual, epigraphic, and isotopic evidence demonstrating self - identification and institutional integration.

Future directions include higher - coverage sequencing; strontium isotope analysis to distinguish locally born from first - generation migrant individuals; direct radiocarbon dating of C5 burials; expanded Anatolian reference panels (particularly Roman - era Central Anatolia); ancient DNA sampling from Roman Dacia and trans - Danubian frontier contexts to test the proposed frontier mixing model directly; Hotelling's T^2 testing for formal statistical assessment of PCA group differences; and extended sampling at both sites and comparable Gothic - associated contexts elsewhere in the Balkans.

4. Conclusions

This paleogenomic study of 38 individuals from Aquae Calidae and the Aul of Khan Omurtag provides molecular evidence that populations archaeologically identified as "Gothic" in Late Antique Bulgaria were profoundly multi - ethnic assemblages rather than biologically coherent populations. Three principal findings emerge.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

First, extreme genetic heterogeneity exists within and between sites sharing identical Gothic material culture and Arian Christian burial practices. Individual ancestry profiles range from nearly 100% Chernyakhov to over 80% Anatolian - related, with no single qpAdm model fitting the full assemblage. This heterogeneity is consistent with historical accounts of Gothic ethnogenesis as a process of progressive accretion - from the Wielbark culture through the Chernyakhov synthesis to the Balkan settlement phases ((Heather and Matthews, 1991) - during which culturally identified "Goths" incorporated individuals of diverse biological.

Second, the two sites form genetically distinct clusters corresponding to different phases and pathways of incorporation. AKO preserves a predominantly northern European/Pontic genetic profile (Wielbark/Chernyakhov - derived, 60 - 78%), with progressive dilution of this ancestry across ~150 years of occupation. Aquae Calidae shows a predominantly Anatolian - shifted profile (50 - 85%), with intra - site substructure distinguishing Central Anatolian and Byzantine Marmara source preferences. The Anatolian ancestry is robustly documented through extensive source rotation testing, and qpAdm models without Anatolian sources fail ($p < 10^{-149}$).

Third, DATES admixture dating places the north - south mixture at ~12 generations before burial (point estimate ~50 CE; 95% CI: 85 BCE - 183 CE), with the signal specifically associated with Polish Wielbark - like ancestry. Control tests on non - Gothic Roman - period Balkan populations - including the published Serbia_Viminacium_Roman_Rit and additional unpublished Bulgarian Roman - period groups - fail to replicate this signal, indicating that the ~12 - generation Wielbark - Balkan/Anatolian admixture clock is specific to Gothic - associated targets rather than a generic feature of Late Antique Balkan populations. However, the temporal constraint - placing admixture before the earliest documented Gothic - Roman contacts (~170 CE) - indicates that the event itself likely began outside the Balkans proper, most plausibly in the trans - Danubian frontier zone

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

created by the Roman colonization of Dacia (post - 106 CE), where documented Anatolian and Balkan provincial settlers mixed with indigenous "barbarian" and early Germanic populations. Balkan and Anatolian southern sources yield indistinguishable timing, consistent with admixture involving a substrate already blended prior to its encounter with Wielbark - related populations. These findings are consistent with Gothic identity in the Late Antique Balkans functioning as a permeable cultural - political category sustained by Arian Christianity, Gothic language, material culture, and foederati political status - incorporating diverse ancestries around an initially northern European core that progressively diluted through local admixture. While the data support ethnogenesis models over essentialist interpretations, we emphasize that genetic heterogeneity alone does not constitute proof of ethnogenesis; integration with textual, archaeological, and isotopic evidence remains essential.

4. Methods

Sample Collection and Processing

Petrous bone samples from 38 individuals (Aquae Calidae: n=23; AKO: n=15) were collected from archaeological excavations conducted by the National Museum of History, Sofia, and the Regional Historical Museum, Burgas. All pre - PCR work was performed in dedicated clean room facilities at the Reich Laboratory, Harvard Medical School, following established ancient DNA protocols. DNA extraction followed Dabney et al. (2013), using a silica - based method optimized for low - input ancient samples. Double - stranded libraries were prepared following Meyer and Kircher (2010), with uracil - DNA glycosylase (UDG) treatment to reduce deamination artifacts while preserving authentic damage patterns. In - solution enrichment targeted ~1.24 million informative SNPs (1240K panel). Sequencing was performed on Illumina NovaSeq 6000 platforms to a median

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

depth of ~0.5 - 1.5× genome - wide coverage. Raw sequences were processed using the EAGER pipeline (Peltzer et al., 2016). MapDamage2.0 (Jónsson et al., 2013) assessed damage patterns; samples with <10% C - to - T transitions at read ends were excluded. Pseudo - haploid genotypes were called by randomly sampling one allele per site. Samples with >5% contamination or <30,000 SNPs covered were excluded; all retained samples passed these thresholds. Samples with <50,000 overlapping SNPs were excluded from formal ancestry modeling.

Principal Component Analysis

PCA was conducted using smartpca from EIGENSOFT v7.2.1 (Patterson et al., 2006), projecting ancient samples onto principal components computed from 1,240 modern West Eurasian individuals (shrinkmode: YES; lsqproject: YES). PC1 and PC2 captured approximately 11% and 19% of total variance, respectively (first 10 PCs: ~79%).

f3 - Outgroup and f4 Statistics

Shared drift and allele frequency correlations were quantified using admixtools2 in R (Maier et al., 2023). f3 statistics used the form f3(Mbuti; Target, Reference); f4 statistics used variable forms including f4(Mbuti, Anatolian; Gothic_Target, Chernyakhov) for Anatolian specificity tests and f4(Mbuti, Chernyakhov; Anatolian, Balkan) for reference validation. Block - jackknife standard errors were computed; $|Z| \geq 3$ was considered strong significance. Substructure within the Chernyakhov reference was resolved by testing Chernyakhov_main and Chernyakhov_outlier separately.

qpAdm Ancestry Modeling

qpAdm modeling was performed using admixtools2 (Patterson et al., 2012; Harney et al., 2021). The fixed right set comprised: Mbuti.DG, Israel_Natufian.SG, Russia_Yana_UP.SG,

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Russia_MA1_HG.SG, Russia_Kostenki14.SG, Italy_North_Villabruna_HG.SG, Morocco_Iberomaurusian.SG, Turkey_Boncuklu_PPN.SG, Iran_GanjDareh_N.SG, Karitiana.DG, Russia_DevilsCave_N.SG. Sources were rotated in combinations of 1 - 4 proxies, with model acceptance requiring $p > 0.05$ and feasible (non - negative) proportions. Nested F - tests assessed whether additional sources significantly improved fit ($p < 0.05$). Both population - level (pooled by site/phase) and individual - level analyses were conducted. To assess heterogeneity systematically, a uniform set of one - , two - , and three - way models was applied to all individuals, with results indicating which models each individual accepts or rejects.

DATES Admixture Dating

Admixture dates were estimated using the DATES method (Narasimhan et al., 2019), which infers the time since admixture from the decay of weighted ancestry covariance along the genome. The method fits an exponential decay curve to ancestry - weighted LD between pairs of sites, with the decay rate proportional to the number of generations since admixture. Two - way models were tested with northern sources (Wielbark - associated: Pol3 pool from Pruszcz Gdański, Masłomęcz, Kowalewko, n=68; or Chernyakhov, n=3) and southern sources (BalkanCore: Bulgaria_EIA + Greek/North Macedonian IA/LBA, n=44; AnatoliaMax: multi - regional Anatolian pool, n=37; or Mugla3: Aegean Anatolian, n=23). Target populations comprised KalideN (n=12), AKO (n=11), and pooled KalideN+AKO (n=23). As controls, identical contrasts were applied to Serbia_Viminacium_Roman_Rit (n=3; Olalde et al., 2023) and additional unpublished Roman and Late Antique period populations from Bulgarian sites. Z - scores > 3 and nrmsd < 0.5 were required for model acceptance. Calendar dates were estimated using 29 years per generation (following Moorjani et al., 2016) and burial midpoint estimates from archaeological and numismatic dating.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Full configuration grid, parameter settings, sensitivity analyses, and control tests are documented in Supplementary Note S3 and Supplementary Table S5.

Uniparental Markers

Y - chromosome haplogroups were assigned using yHaplo (Poznik, 2016) with terminal SNPs from ISOGG v15.73 (≥3× coverage required). Mitochondrial haplogroups were assigned using HaploGrep2 (Weissensteiner et al., 2016; PhyloTree Build 17). Haplogroup - ancestry associations are probabilistic and based on published ancient and modern distributions.

Identity - by - Descent and Kinship Analysis

Pairwise kinship was estimated using READv2 (Monroy Kuhn et al., 2018) on ~220,000 pairs (64 Gothic - involved), with a kinship cutoff of 0.04 (capturing ~4th - 5th degree relationships). Population - normalized enrichment ratios were calculated as observed/expected connections based on reference population sizes (~979 individuals). Sensitivity analyses varied cutoff thresholds, segment lengths, and software (pcangsd/NGSTools cross - validation; correlation r=0.92 with READv2). Full pairwise outputs, network statistics, enrichment calculations, and sensitivity analyses are provided in Supplementary Note S2 and the Supplementary IBD Workbook.

Data Availability

Raw sequencing data: European Nucleotide Archive (PRJXXXXXX). Genotypes: Allen Ancient DNA Resource (AADR). Interactive PCA, f3, and IBD visualizations: [GitHub repository URL].

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Acknowledgments

This publication has been developed within the framework of project BG05SFPR001-3.004-0005-C01 “Achieving Innovation and Growth through Doctoral Research”, implemented under Programme “Education” 2021–2027, co-financed by the European Union through the European Social Fund Plus (ESF+).







This research was conducted as part of the “Epochs of History” paleogenomic project.

This work was supported by the Bulgarian National Science Fund (grant/contract no. KP-06-N70/8; Contract № КП-06-Н70/8) as an interdisciplinary collaboration between the IBCT - Bulgarian Academy of Science, UNIBIT – Sofia, the National Institute for Archaeology with Museum, National Historical Museum, Bulgarian Regional Museums of History and Harvard University's David Reich Laboratory of Archaeogenetics.

We are deeply grateful to Prof. Stoyan Denchev (UNIBIT, Sofia) for his unwavering support as doctoral adviser to SS and for generously providing access to the computational server infrastructure on which many of the analyses in this study were performed.

We extend our heartfelt thanks to Prof. Boni Petrunova (National Museum of History, Sofia) for hosting the research at National Museum of History, for facilitating access to archaeological collections, and for her sustained support and encouragement throughout the project. This work would not have been possible without her institutional vision and personal commitment.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We thank Prof. Rumyana Preshlenova (Institute for Balkan Studies with Centre for Thracology, Bulgarian Academy of Sciences / IBCT-BAS) for her support of the interdisciplinary framework within which this research was conducted.

We gratefully acknowledge Aisling Kearns and the ancient DNA clean-room staff at the Reich Laboratory, Harvard Medical School, for expert sample processing and library preparation, and the Harvard sequencing facility for high-throughput sequencing support.

Author Contributions

S.S.: Conceptualization, Formal analysis, Data curation, Methodology, Software (bioinformatics pipeline implementation), Visualization, Writing – original draft, Writing – review & editing, Project administration. T.C.: Conceptualization, Funding acquisition, Resources (archaeological and chronological framework), Writing – original draft (Supplementary Note S4), Writing – review & editing. T.W.: Formal analysis, Methodology, Writing – review & editing. K.S.: Resources (archaeological collections, Veliki Preslav assemblage), Investigation. D.M.: Resources (archaeological collections, Regional Historical Museum Burgas), Investigation. A.A.: Resources (National Archaeological Institute collections and site documentation), Investigation. M.N.: Resources (Regional Historical Museum Burgas collections), Investigation. D.N.: Investigation (skeletal sampling and physical specimen collection), Resources. P.H.: Writing – review & editing (historical interpretation and ethnogenesis framework). D.I.T.: Writing – review & editing. M.Z.: Writing – review & editing. I.L.: Methodology, Formal analysis, Writing – review & editing, Supervision. D.R.: Conceptualization, Funding acquisition, Resources (sequencing and laboratory infrastructure), Supervision, Writing – review & editing.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

References

Allentoft, M. E., Sikora, M., Sjögren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., et al. (2015). Population genomics of Bronze Age Eurasia. Nature 522:167-172. doi: 10.1038/nature14507

Alt, K. W., Knipper, C., Peters, D., Müller, W., Maurer, A. F., Kollig, I., et al. (2014). Lombards on the move—An integrative study of the migration period cemetery at Szólád, Hungary. PLOS ONE 9:e110793. doi: 10.1371/journal.pone.0110793

Amorim, C. E. G., Vai, S., Posth, C., Modi, A., Koncz, I., Hakenbeck, S., et al. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. Nature Communications 9:3547. doi: 10.1038/s41467-018-06024-4

Antonio, M. L., Gao, Z., Moots, H. M., Lucci, M., Rickards, O., Jakobsson, M., et al. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science 366:708-714. doi: 10.1126/science.aay6826

Berndt, G. M., Steinacher, R., and eds. (2014). Arianism: Roman Heresy and Barbarian Creed. Farnham: Ashgate.

Bierbrauer, V. (1994). Archäologie und Geschichte der Goten vom 1.-7. Jahrhundert. Frühmittelalterliche Studien 28:51-171.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences 104:14616-14621. doi: 10.1073/pnas.0704665104

Burns, T. S. (1984). A History of the Ostrogoths. Bloomington: Indiana University Press.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., et al. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. American Journal of Human Genetics 93:852-864. doi: 10.1016/j.ajhg.2013.10.002

Childebayeva, A., Amorim, C. E. G., Günther, T., Prüfer, K., Malmström, H., Rohrlach, A. B., et al. (2024). High-resolution genomic history of early medieval Europe. Nature 637:118-126. doi: 10.1038/s41586-024-08275-2

Csáky, V., et al. (2020). Genetic insights into the social organisation of the Avar period elite in the 7th century AD Carpathian Basin. Scientific Reports 10:948. doi: 10.1038/s41598-019-57378-8

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences 110:15758-15763. doi: 10.1073/pnas.1314445110

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., et al. (2021). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. PeerJ 9:e10947. doi: 10.7717/peerj.10947

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. Nature 524:216-219. doi: 10.1038/nature14558

Gansauge, M. T. and Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8:737-748. doi: 10.1038/nprot.2013.038

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Gansauge, M. T., Aximu-Petri, A., Nagel, S., and Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nature Protocols 15:2279-2300. doi: 10.1038/s41596-020-0338-0

Geary, P. J. (2002). The Myth of Nations: The Medieval Origins of Europe. Princeton: Princeton University Press.

Gînguță, A., Rusu, I., Mirițoiu, N., Kis, L., Szabo, T. A., Nemeș, E., et al. (2022). Maternal Lineages of Gepids from Transylvania. Genes 13:563. doi: 10.3390/genes13040563

Gnecchi-Ruscone, G. A., Khussainova, E., Kahbatkyzy, N., Musralina, L., Spyrou, M. A., Bianco, R. A., Radzevičiūtė, R., Gomes Martins, N. F., Freund, C., Iksan, O., et al. (2021). Ancient genomic time transect from the Central Asian Steppe unravels the history of the Scythians. Science Advances 7:eabe4414. doi: 10.1126/sciadv.abe4414

Gnecchi-Ruscone, G. A., Szécsényi-Nagy, A., Koncz, I., Csiky, G., Rácz, Z., Rohrlach, A. B., et al. (2022). Ancient genomes reveal origin and rapid trans-Eurasian migration of 7th century Avar elites. Cell 185:1402-1413.e21. doi: 10.1016/j.cell.2022.03.007

Gnecchi-Ruscone, G. A., et al. (2024). Network of large pedigrees reveals social practices of Avar communities. Nature 625:312-320. doi: 10.1038/s41586-024-07312-4

Goffart, W. (1988). The Narrators of Barbarian History (A.D. 550-800): Jordanes, Gregory of Tours, Bede, and Paul the Deacon. Princeton: Princeton University Press.

Goffart, W. (2006). Barbarian Tides: The Migration Age and the Later Roman Empire. Philadelphia: University of Pennsylvania Press.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Gretzinger, J., Sayer, D., Justeau, P., Altena, E., Pala, M., Dulias, K., et al. (2022). The Anglo-Saxon migration and the formation of the early English gene pool. Nature 610:112-119. doi: 10.1038/s41586-022-05247-2

Hakenbeck, S. E. (2019). Present pasts in the archaeology of genetics, identity, and migration in Europe: a critical essay. World Archaeology 51:528-545. doi: 10.1080/00438243.2019.1627907

Halsall, G. (2007). Barbarian Migrations and the Roman West, 376-568. Cambridge Medieval Textbooks. Cambridge: Cambridge University Press.

Harney, E., Patterson, N., Reich, D., and Wakeley, J. (2021). Assessing the performance of qpAdm: a statistical tool for studying population admixture. Genetics 217(4):iyaa045. doi: 10.1093/genetics/iyaa045

Heather, P. (1986). The Crossing of the Danube and the Gothic Conversion. Greek, Roman, and Byzantine Studies 27:289-318.

Heather, P. J. (1991). Goths and Romans 332-489. Oxford Historical Monographs. Oxford: Clarendon Press. ISBN: 0-19-820234-2

Heather, P. and Matthews, J. (1991). The Goths in the Fourth Century. Translated Texts for Historians, Vol. 10. Liverpool: Liverpool University Press. ISBN: 0-85323-426-4

Heather, P. (1996). The Goths. The Peoples of Europe. Oxford: Blackwell Publishers. ISBN: 0-631-16536-3

Järve, M., Saag, L., Scheib, C. L., Pathak, A. K., Montinaro, F., Pagani, L., et al. (2019). Shifts in the genetic landscape of the western Eurasian steppe associated with the beginning and end of the Scythian dominance. Current Biology 29:2430-2441.e10. doi: 10.1016/j.cub.2019.06.019

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29:1682-1684. doi: 10.1093/bioinformatics/btt193

Juras, A., Dabert, M., Kushniarevich, A., Malmström, H., Raghavan, M., Kosicki, J. Z., et al. (2014). Ancient DNA reveals matrilineal continuity in present-day Poland over the last two millennia. PLOS ONE 9:e110839. doi: 10.1371/journal.pone.0110839

Juras, A., Krzewińska, M., Nikitin, A. G., Ehler, E., Chyleński, M., Łukasik, S., et al. (2017). Diverse origin of mitochondrial lineages in Iron Age Black Sea Scythians. Scientific Reports 7:43950. doi: 10.1038/srep43950

Kazanski, M. (1991). Les Goths (Ier-VIIe siècles ap. J.-C.). Paris: Editions Errance.

Kircher, M., Sawyer, S., and Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40:e3. doi: 10.1093/nar/gkr771

Korneliussen, T. S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics 15:356. doi: 10.1186/s12859-014-0356-4

Krzewińska, M., Kjellström, A., Günther, T., Hedenstierna-Jonson, C., Zachrisson, T., Omrak, A., et al. (2018). Genomic and strontium isotope variation reveal immigration patterns in a Viking Age town. Current Biology 28:2730-2738.e10. doi: 10.1016/j.cub.2018.06.053

Krzewińska, M., Kılınç, G. M., Juras, A., Koptekin, D., Chyleński, M., Nikitin, A. G., et al. (2018). Ancient genomes suggest the eastern Pontic-Caspian steppe as the source of western Iron Age nomads. Science Advances 4:eaat4457. doi: 10.1126/sciadv.aat4457

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Kulikowski, M. (2007). Rome's Gothic Wars: From the Third Century to Alaric. Key Conflicts of Classical Antiquity. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167277

Lazaridis, I., Alpaslan-Roodenberg, S., Acar, A., et al. (2022). A genetic probe into the ancient and medieval history of Southern Europe and West Asia. Science 377:940-951. doi: 10.1126/science.abq0755

Lazaridis, I., Alpaslan-Roodenberg, S., et al. (2022). The genetic history of the Southern Arc: A bridge between West Asia and Europe. Science 377:eabm4247. doi: 10.1126/science.abm4247

Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database Collaboration (2011). The Sequence Read Archive. Nucleic Acids Research 39:D19-D21. doi: 10.1093/nar/gkq1019

Maenchen-Helfen, O. J. (1973). The World of the Huns: Studies in Their History and Culture. Edited by Max Knight. Berkeley: University of California Press.

Maier, R., Flegontov, P., Isildak, U., Mallick, S., Reich, D., and Patterson, N. (2023). admixtools2: Inferring demographic history from genetic data. R package. Available at: https://github.com/uqrmaie1/admixtools

Mallick, S. and Reich, D. (2024). The Allen Ancient DNA Resource (AADR): A curated compendium of ancient human genomes [Dataset]. Harvard Dataverse. doi: 10.7910/DVN/FFIDCW

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., et al. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Scientific Data 11:182. doi: 10.1038/s41597-024-03031-7

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., et al. (2020). Population genomics of the Viking world. Nature 585:390-396. doi: 10.1038/s41586-020-2688-8

Maróti, Z., Neparáczki, E., Schütz, O., Maár, K., Varga, G. I. B., Kovács, B., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., Marcsik, A., Pálfi, G., Bernert, Z., Gallina, Z., Horváth, C., Varga, S., Költő, L., Raskó, I., Nagy, P. L., et al. (2022). The genetic origin of Huns, Avars, and conquering Hungarians. Current Biology 32:2858-2870.e7. doi: 10.1016/j.cub.2022.04.093

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Alpaslan-Roodenberg, S., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499-503. doi: 10.1038/nature16152

Meyer, M. and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010:pdb.prot5448. doi: 10.1101/pdb.prot5448

Monroy Kuhn, J. M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLOS ONE 13:e0195491. doi: 10.1371/journal.pone.0195491

Moorjani, P., Sankararaman, S., Fu, Q., Przeworski, M., Patterson, N., and Reich, D. (2016). A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proceedings of the National Academy of Sciences of the United States of America 113(20):5652–5657. doi: 10.1073/pnas.1514696113

Nakatsuka, N., Harney, E., Mallick, S., Mah, M., Patterson, N., and Reich, D. (2020). ContamLD: estimation of ancient nuclear DNA contamination using breakdown of linkage disequilibrium. Genome Biology 21:199. doi: 10.1186/s13059-020-02111-2

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Narasimhan, V. M., et al. (2019). The formation of human populations in South and Central Asia. Science 365(6457):eaat7487. doi: 10.1126/science.aat7487

Olalde, I., Carrión, P., Mikić, I., Rohrlach, A. B., Mallick, S., Lazaridis, I., et al. (2023). A genetic history of the Balkans from Roman frontier to Slavic migrations. Cell 186:5472-5485.e9. doi: 10.1016/j.cell.2023.10.018

Oltean, I. A. (2007). Dacia: Landscape, Colonisation and Romanisation. Routledge.

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. PLOS Genetics 2:e190. doi: 10.1371/journal.pgen.0020190

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. Genetics 192:1065-1093. doi: 10.1534/genetics.112.145037

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., et al. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology 17:60. doi: 10.1186/s13059-016-0918-z

Peyrégne, S. and Peter, B. M. (2020). AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. Genome Biology 21:246. doi: 10.1186/s13059-020-02123-y

Pohl, W., Reimitz, H., and eds. (1998). Strategies of Distinction: The Construction of Ethnic Communities, 300-800. The Transformation of the Roman World, Vol. 2. Leiden: Brill.

Pohl, W., Heydemann, G., and eds. (2013). Strategies of Identification: Ethnicity and Religion in Early Medieval Europe. Cultural Encounters in Late Antiquity and the Middle Ages, Vol. 13. Turnhout: Brepols.

Poznik, G. D., Henn, B. M., Yee, M. C., Sliwerska, E., Euskirchen, G. M., Lin, A. A., et al. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science 341:562-565. doi: 10.1126/science.1237619

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Poznik, G. D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped individuals. bioRxiv 088716. doi: 10.1101/088716

Reich, D. (2018). Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past. New York: Pantheon Books.

Renaud, G., Slon, V., Duggan, A. T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biology 16:224. doi: 10.1186/s13059-015-0776-0

Ricaut, F. X., Auriol, V., von Cramon-Taubadel, N., Keyser, C., Murail, P., Ludes, B., et al. (2011). Mitochondrial analysis of a Byzantine population reveals the differential impact of multiple historical events in South Anatolia. European Journal of Human Genetics 19:571-576. doi: 10.1038/ejhg.2010.230

Ringbauer, H., et al. (2024). Accurate detection of identity-by-descent segments in human ancient DNA. Nature Genetics 56:143-151. doi: 10.1038/s41588-023-01582-w

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society B 370:20130624. doi: 10.1098/rstb.2013.0624

Rohland, N., Glocke, I., Aximu-Petri, A., and Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. Nature Protocols 13:2447-2461. doi: 10.1038/s41596-018-0050-5

Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., and Reich, D. (2022). Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. Genome Research 32:2068-2078. doi: 10.1101/gr.276728.122

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Saag, L., et al. (2025). North Pontic crossroads: Mobility in Ukraine from the Bronze Age to the early modern period. Science Advances 11(2):eadr0695. doi: 10.1126/sciadv.adr0695

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Sherry, S. T., Yankie, L., et al. (2024). GenBank 2024 Update. Nucleic Acids Research 52:D134-D137. doi: 10.1093/nar/gkad903

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., et al. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. Nature Communications 7:10408. doi: 10.1038/ncomms10408

Schönherr, S., Weissensteiner, H., Kronenberg, F., and Forer, L. (2023). Haplogrep 3 - an interactive haplogroup classification and analysis platform. Nucleic Acids Research 51:W263-W268. doi: 10.1093/nar/gkad284

Skourtanioti, E., Erdal, Y. S., Frangipane, M., Balossi Restelli, F., Yener, K. A., Pinnock, F., et al. (2020). Genomic history of Neolithic to Bronze Age Anatolia, Northern Levant, and Southern Caucasus. Cell 181:1158-1175.e28. doi: 10.1016/j.cell.2020.04.044

Speidel, L., Silva, M., Booth, T., Raffield, B., Anastasiadou, K., Barrington, C., et al. (2025). High-resolution genomic history of early medieval Europe. Nature 637:118–126. doi: 10.1038/s41586-024-08275-2

Stolarek, I., Juras, A., Handschuh, L., Marcinkowska-Swojak, M., Piontek, J., Kozlowski, P., et al. (2018). A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age. Scientific Reports 8:2455. doi: 10.1038/s41598-018-20705-6

Stolarek, I., Juras, A., Handschuh, L., Marcinkowska-Swojak, M., Kukla-Bartoszek, M., Piontek, J., et al. (2019). Goth migration induced changes in the matrilineal genetic structure of the central-east European population. Scientific Reports 9:6737. doi: 10.1038/s41598-019-43183-w

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Stolarek, I., Handschuh, L., Juras, A., Nowaczewska, W., Piontek, J., Kozlowski, P., et al. (2023). Genetic history of East-Central Europe in the first millennium CE. Genome Biology 24:173. doi: 10.1186/s13059-023-03013-9

Streitberg, W. and ed. (1950). Die gotische Bibel. 3rd edition. Heidelberg: Carl Winter.

Thompson, E. A. (2008). The Visigoths in the Time of Ulfila. Second Edition. With a new Foreword by Michael Kulikowski. London: Duckworth. doi: 10.5040/9781472541000

Unterländer, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanová, Z., Groß, M., Sell, C., Blöcher, J., Kirsanow, K., Rohland, N., et al. (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nature Communications 8:14615. doi: 10.1038/ncomms14615

Vai, S., Ghirotto, S., Pilli, E., Migliori, L., Caramelli, D., and Lari, M. (2019). A genetic perspective on Longobard-Era migrations. European Journal of Human Genetics 27:647-656. doi: 10.1038/s41431-018-0319-8

Veeramah, K. R., Rott, A., Groß, M., van Dorp, L., López, S., Quach, H., et al. (2018). Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. Proceedings of the National Academy of Sciences 115:3494-3499. doi: 10.1073/pnas.1719880115

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., et al. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research 44:W58-W63. doi: 10.1093/nar/gkw233

Wenskus, R. (1961). Stammesbildung und Verfassung: Das Werden der frühmittelalterlichen gentes. Cologne: Böhlau.

bioRxiv preprint doi: https://doi.org/10.64898/2026.03.03.709317; this version posted March 5, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Wolfram, H. (1988). History of the Goths. Translated by Thomas J. Dunlap. Berkeley: University of California Press. doi: 10.1525/9780520310773

Wolfram, H. (1997). The Roman Empire and Its Germanic Peoples. Translated by Thomas Dunlap. Berkeley: University of California Press.

Yuan, D., Ahamed, A., Burgin, J., Cummins, C., Devraj, R., Gueye, K., et al. (2024). The European Nucleotide Archive in 2023. Nucleic Acids Research 52:D92-D97. doi: 10.1093/nar/gkad1067