Pigmentation Genetics in Archaic Hominins: Evaluating DNA phenotyping in Neanderthal and Denisovan genomes

Lily Heald

Honors Thesis in the Department of Anthropology, University of Michigan

Advised by Dr. Tina Lasisi

20 April, 2026

List of figures

Figure 1: Distribution of pigmentation-associated SNPs across the genome..... 11
Figure 2: Skin pigmentation evolution in response to UV radiation. Figure adapted from Guermazi & Saliba, 2025b.....13
Figure 3: Map of discovery sites of archaic genomic samples with latitude. Available archaic genomic samples all come from similar latitudes.....15
Figure 4: Map of 15 sampled Simons Genome Diversity Project genomes with December 2010 average UV radiation (NASA AURA)..... 21
Figure 5: Heatmap of log-transformed sequencing depth across archaic samples with missing coverage (depth = 0) included as black..... 26
Figure 6: Probability consensus base call is correct is modeled as a function of sequencing depth, error profile, and position in read length..... 28
Figure 7: Principal component analysis of polymorphic SGDP loci.....29
Figure 8: Principal component analysis of 222 pigmentation-related SNPs..... 30

List of tables

Table 1: Key definitions.....5
Table 2: Table of genes involved in melanin production and skin pigmentation..... 12
Table 3: Age, species, and location of the five available high-coverage archaic genomes..... 15

Acknowledgements

I would like to thank my advisors and mentors, Dr. Tina Lasisi and Dr. Yemko Pryor, for advice and guidance throughout the research process. This project would not have been possible without resources made publicly available by the Max Planck Institute for Evolutionary Anthropology and the Simons Genome Diversity Project.

Abstract

The question of what our archaic hominin ancestors may have looked like has captured the attention of scientists and popular culture. Inferring phenotypes from DNA is a challenging statistical problem, further complicated by inference in unobservable archaic populations. Previous research has sequenced pigmentation related variants in ancient and archaic samples, but this work remains spread across single locus studies and phenotype prediction attempts rather than directly consolidated into a systematic archaic-focused resource. Here, I attempt to evaluate the feasibility of phenotype prediction in 14 archaic Neanderthal and Denisovan individuals through the analysis of depth and missingness in available genomes, and compare available pigmentation-related SNPs in these archaic populations to diverse modern human genomes obtained from the Simons Genome Diversity Project. PCAs were constructed using the modern human genomes, and archaic samples were then projected onto the PC space. Sequencing depth of the archaic samples was calculated using SAMtools. Analyses of depth and missingness were performed as a mixed effects logistic regression of missingness and two simulations of the confidence in the consensus base call conferred by sequence depth. There is significant variation in coverage and depth across archaic samples. Missingness, for which age and structural influences are significant predictors, is also non-random. Furthermore, untreated ancient DNA is significantly affected by low coverage. Archaic samples cluster near Eurasians in a genome-wide SNP panel, but with Africans in a pigmentation panel. This emphasizes the importance of caution and conservative statistical methods when attempting to make sweeping species-wide phenotypic predictions from archaic DNA.

Table of Contents

List of figures....1
List of tables....1
Acknowledgements....1
Abstract....2
1.0 Introduction....4
1.1 Background....6
1.2 Melanin pigmentation....7
1.3 Adaptive function of human skin pigmentation....9
1.3 Modern variation in pigmentation related SNPs....10
1.5 Archaic genome availability....14
1.6 Archaic introgression....19
2.0 Methods....20
2.1 Modern comparative dataset....20
2.2 SNP panels....21
2.3 Archaic Samples....22
2.5 Conversion to PLINK and LD pruning....22
2.6 Principal component analysis....23
2.7 Probability that a majority-rule consensus call recovers the true nucleotide....24
2.8 Mixed effects modeling....25
3.0 Results....26
3.1 There is a lot of variation in sample coverage....26
3.2 Sample age is a significant predictor of the probability of missingness....27
3.3 Untreated ancient DNA is significantly affected by coverage depth....28
3.4 Archaic samples cluster near Eurasians in whole-genome SNPs....29
3.5 Pigmentation SNPs lack clear geographic structure....30
4.0 Discussion....30
References....34

1.0 Introduction

Current scholarship on pigmentation in ancient and archaic humans is disproportionately centered on the study of introgressed archaic alleles in modern human populations rather than on cataloging the pigment-related variants directly in the genomes of archaic individuals (Dannemann et al., 2017a; Kang et al., 2023; McArthur et al., 2020). As a result, understanding of pigmentation genetics remains dispersed across partly overlapping literature including selection studies, phenotype inference, and selection-based analyses. This has resulted in a fragmented understanding of the genetic architecture underlying pigmentation in ancient and archaic hominins.

Existing work has sequenced and analyzed pigmentation related variants in ancient and archaic genomes (Brand et al., 2022; C. C. S. Cerqueira et al., 2012; Ferrando-Bernal et al., 2025; Lalueza-Fox et al., 2007; Perry et al., 2015), but the research remains spread across single locus studies and phenotype prediction attempts rather than directly consolidated into a systematic archaic focused resource. Discussion of pigmentation-related variation in archaic genomes has often centered around a small number of loci. In particular, the discussion of pigmentation-related alleles is disproportionately focused on *MC1R*, with studies focusing on the introgression of archaic haplotypes into modern human populations (Ding et al., 2014; Lalueza-Fox et al., 2007). Selection-based approaches in modern and ancient human populations have successfully identified key loci associated with pigmentation; however, these studies typically infer selection dynamics without comparing archaic genotypes to those of modern populations (Ju & Mathieson, 2021a). Genome-wide analyses of archaic introgression have identified regulatory and trait-associated effects relevant to skin and hair pigmentation, but these studies do not provide locus-level catalogs for variants associated with the pigmentation of

keratinized tissues in archaic genomes (Dannemann et al., 2017b; Kang et al., 2023; Rinker et al., 2019; Witt et al., 2022). In addition, earlier multi-locus approaches used smaller marker panels, whereas the present study examines a larger set of 222 pigmentation related single-nucleotide polymorphisms (SNPs), allowing broader locus coverage (C. C. S. Cerqueira et al., 2012).

Vocabulary

Term	Definition
Loci	Position of a gene or DNA sequence on a chromosome
Allele	One form of a gene
Haplotype	A set of closely linked genetic markers on a single chromosome that are inherited together from a single parent
Genotype	Two alleles present in an individual at a specific locus of the genome

Table 1: *Key definitions*

No study has yet curated a single, comprehensive locus-level inventory of pigmentation-associated variants in archaic genomes with explicit comparison to diverse modern human genomes. The primary aims of this study are to evaluate patterns of missingness and the implications of low sequence depth in archaic hominin genomic samples, and to provide a comprehensive locus catalog for variants associated with the skin pigmentation. Analyses of the distribution of pigmentation-related variants in archaic genomes will be conducted in comparison to a diverse sample of 15 modern human genomes to assess patterns of shared variation and divergence. By shifting focus from inference based on modern populations to direct characterization of archaic genomic data, this study seeks to refine current models of pigmentation evolution and its underlying genetic architecture.

1.1 Background

Since the publication of the first whole-genome sequence from an ancient sample in 2010, computational advances have revolutionized the availability of ancient DNA sequence data (Green et al., 2010). Genome-wide data from over 10,000 ancient humans and archaic hominins is now available (Mallick et al., 2024). Here, ancient humans are defined as ancient *Homo sapiens* samples, and archaic samples are defined as non-*sapiens Homo* specimens, such as *Homo neanderthalensis* and Denisovans. These datasets have enabled increasingly detailed analyses of population history and admixture.

A major area of interest in the study of ancient human and archaic hominin DNA is the prediction of phenotypes which are not able to be directly observed from skeletal remains (C. C. Cerqueira et al., 2012; Ferrando-Bernal et al., 2025; Pinson et al., 2022; Wang et al., 2023). Skin pigmentation is one polygenic trait that has been shown to have experienced directional selection in humans, and has garnered interest for phenotype prediction in ancient populations (Jablonski & Chaplin, 2018). However, DNA phenotyping is a challenging statistical problem, further complicated by the sparse nature of ancient DNA sequence availability, both in sequence depth and population coverage. The inference of genetic loci associated with phenotypes requires large cohorts, and even in well characterized populations, the predictive accuracy of polygenic models remains limited. For instance, when testing genetic predictions in anthropometric and blood panel traits in the UK Bio Bank, it was found that prediction accuracy was far lower in non-European populations, and up to 4.9-fold lower in Africans on average (Martin et al., 2019). These predictions are even less reliable in ancient samples, where there are issues of low coverage sequencing and small, unevenly distributed sample sizes.

In genome-wide association studies (GWAS), variants are associated with traits detected in large cohorts, requiring upwards of hundreds of cases for statistical significance when studying the effect of a single-nucleotide polymorphism (SNP) (Hong & Park, 2012). Models such as polygenic risk scores (PRS) predict variation in complex traits from the combination of effects of multiple variants (Ferrando-Bernal et al., 2025). These methods, though constructed for the analysis of living populations, have been applied to ancient and archaic populations. More complex methods have been developed for the prediction of hair and eye color. HIrisPlex, for example, is a single multiplex assay targeting 24 hair and eye color variants (Walsh et al., 2013). These tools were designed for low amounts of template DNA or degraded samples, and were used for forensic applications as well as to archaic and ancient genomes.

However, the application of these methods to ancient datasets assume that GWAS effect sizes from modern populations are applicable to ancient populations. The limitations of this method come from the nature of ancient DNA itself: we are using quantitative estimates of variant effects observed in modern groups and applying it to an unobservable group which is unlike that in which the PRS was constructed (especially for archaic groups which are a different species entirely). Furthermore, these reconstructions rely on the assumption that GWAS effect sizes calculated from modern populations are stable across environmental contexts. Understanding if the genome-phenotype map has evolved remains a critical question for the interpretation of ancient DNA.

1.2 Melanin pigmentation

Human skin color is determined by the amount, composition, and distribution of epidermal melanin, with smaller contributions from chromophores including carotene and

hemoglobin (Lin & Fisher, 2007; Miyamura et al., 2007). Human skin contains two kinds of melanin: brownish-black eumelanin and reddish-yellow pheomelanin (Guermazi & Saliba, 2025a). Both types of melanin are present in the epidermis, but eumelanin is the dominant component of epidermal melanin across skin types, and visible differences in constitutive skin color are driven by an increase in total melanin levels, but not differences in relative proportions of eumelanin and pheomelanin (Ito & Wakamatsu, 2003). Regardless of overall levels of pigmentation, pheomelanin comprises a relatively constant 26% of total melanin (Del Bino et al., 2015).

Melanin is synthesized through the process of melanogenesis, a biochemical pathway occurring within melanocytes (Lin & Fisher, 2007). Melanocytes are specialized dendritic cells located in the epidermis. Each melanocyte forms functional units with surrounding keratinocytes. Within melanocytes, melanin is synthesized and stored in specialized cytoplasmic organelles called melanosomes. Mature melanosomes are transported along dendritic processes and transferred to keratinocytes, where they contribute to visible pigmentation. Melanogenesis is initiated by the enzyme tyrosinase, which catalyzes the oxidation of the amino acid tyrosine into dihydroxyphenylalanine (DOPA) and then dopaquinone. Dopaquinone is critical in the early steps of melanin production, leading to the production of either eumelanin or pheomelanin. Melanocyte function is regulated by a broader signaling network that includes *MC1R* and *MITF*, linking environmental and hormonal cues to expression of melanogenic enzymes and melanosome maturation (Kondo & Hearing, 2011).

Variation in skin pigmentation is influenced by multiple factors. Biochemical determinants of skin coloration include melanosome transfer to keratinocytes, intracellular distribution, and persistence within the epidermis (Kondo & Hearing, 2011). In highly pigmented

skin, melanosomes in keratinocytes are predominantly larger and individually distributed throughout all epidermal layers, whereas in lightly pigmented skin they are smaller and more frequently grouped as membrane-bound clusters concentrated in the stratum basale (Hurbain et al., 2018).

Functionally, melanin plays a critical role in photoprotection. Melanin has the ability to absorb and scatter ultraviolet rays (UVR), providing protection from sunburning (Miyamura et al., 2007). This reduces DNA damage in skin cells, consequently lowering the risk of mutagenesis and skin cancer. Variation in pigmentation is closely linked with environmental UVR exposure. Pigmentation evolution is highly adaptively significant in human evolution, particularly in relation to geographic variation in UVR intensity.

1.3 Adaptive function of human skin pigmentation

Human skin pigmentation is a very well-characterized example of adaptive variation in response to environmental pressures. The primary evolutionary tradeoff driving the diversity we see in human skin pigmentation is the role of ultraviolet B rays (UVB) radiation in facilitating vitamin D synthesis and inducing cellular damage, including the degradation of folate (Jablonski & Chaplin, 2010).

Melanin pigments are effective at absorbing and scattering UVB radiation, the wavelength range that is responsible for catalyzing the conversion of 7-dehydrocholesterol to previtamin D3 in the skin. Increased melanin concentration in the skin results in a decrease in efficiency of cutaneous vitamin D3 synthesis. Individuals with high levels of constitutive pigmentation require longer UVB exposure to produce the same amount of vitamin D as someone with lower constitutive pigmentation. In environments with limited UVB, such as high

latitudes, this can predispose individuals to vitamin D deficiency. Conversely, reduced melanin content enhances the efficiency of vitamin D production under low-UVB conditions.

However, ultraviolet radiation can also degrade folate, a B vitamin essential for DNA synthesis and cell division (Borradale et al., 2014). Maternal folate deficiency is associated with increased congenital abnormalities, including neural tube defects and congenital heart defects (Czeizel et al., 2013). Higher melanin concentrations provide a productive advantage by limiting UV penetration and preserving folate levels in regions with high UV radiation.

The global distribution of human skin pigmentation is widely interpreted as the outcome to this evolutionary tradeoff shaped by spatial variation in UV intensity. Populations in equatorial regions where UV exposure is consistently high tend to exhibit darker pigmentation, conferring protection against photodamage and the preservation of folate. Conversely, populations at higher latitudes have evolved lighter pigmentation, facilitating vitamin D production in low-UVB conditions. This adaptive framework provides context for understanding pigmentation-associated genetic variation.

1.3 Modern variation in pigmentation related SNPs

The genetic architecture of human pigmentation is highly polygenic, involving many loci that contribute to variation in skin, hair, and eye color (E. Quillen et al., 2018). Genome-wide association studies have identified hundreds of single-nucleotide polymorphisms (SNPs) associated with pigmentation phenotypes. Many of these variants lie in genes involved in melanogenesis, melanosome biogenesis, melanosomal transport, or regulation of melanocyte signaling, including *SLC45A2* (chr5), *ADAMTS12* (chr5), *TYR* (chr11), *SLC24A5* (chr15), *OCA2*

(chr15), *HERC2* (chr15), and *MC1R* (chr16) (Sturm, 2009). The distribution of pigmentation associated SNPs across the genome is not uniform.

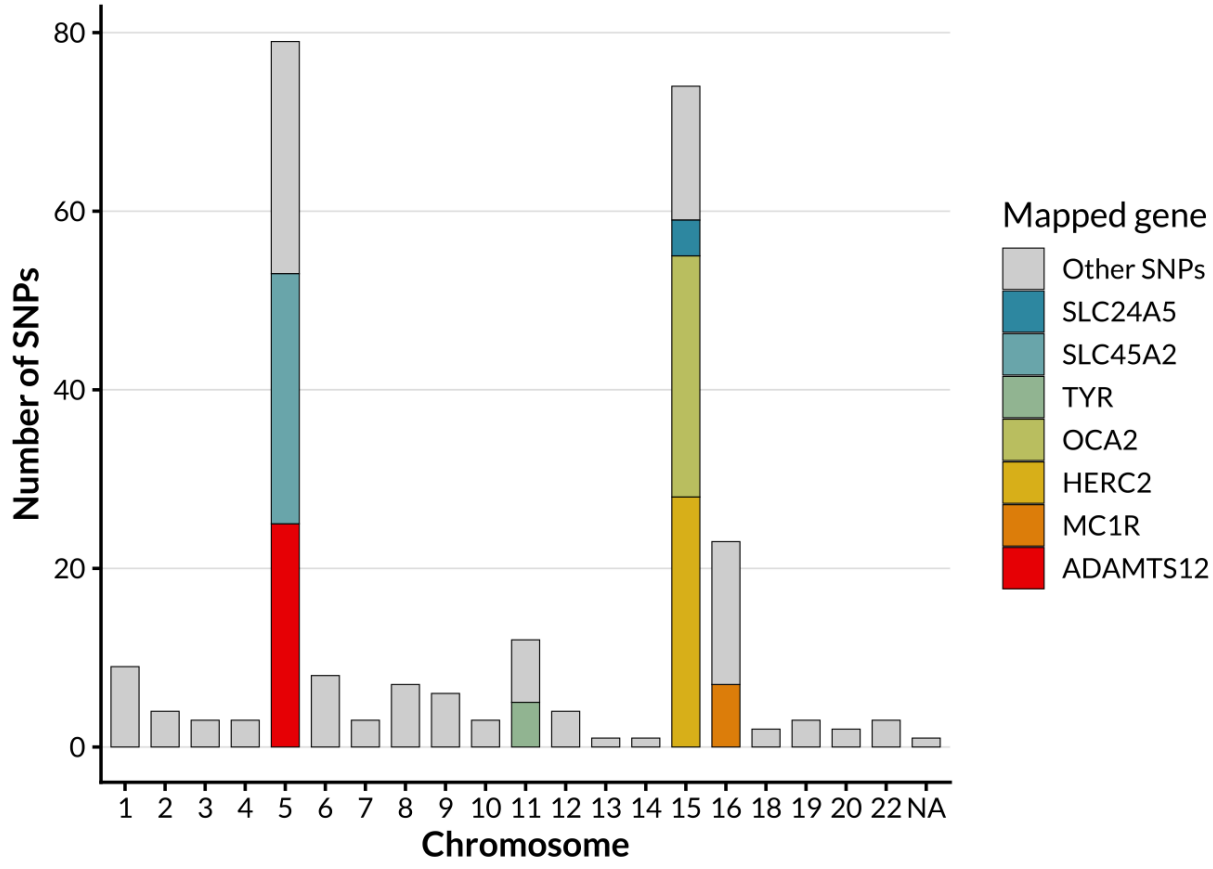


Figure 1: *Distribution of pigmentation-associated SNPs across the genome.*

Current research also shows that the architecture of skin pigmentation is broader and more population-specific than earlier Eurocentric models implied (Feng et al., 2021; E. E. Quillen et al., 2019). Studies in African and admixed populations have identified additional loci, including *MFSD12* and *DDB1*, that were not prominent in early European-focused analyses (Crawford et al., 2017; Feng et al., 2021; Martin et al., 2017). The number of known pigmentation SNPs is now large, effect sizes vary across populations, and older predictive panels

were built on samples that only capture a fraction of the global genetic landscape (E. E. Quillen et al., 2019).

Pigmentation-associated SNPs

Locus	Typical role in pigmentation biology	Evidence in modern humans
SLC24A5	Melanosome associated ion transport and melanogenesis	Major contributor to lighter skin in West Eurasia and beyond
SLC45A2	Melanosomal transport and pH related effects on melanogenesis	Repeatedly associated with skin color variation, especially in Europeans
TYR	Core melanogenic enzyme pathway	Associated with normal pigment variation and lighter pigmentation in some populations
OCA2/HERC2	Pigment regulation with strong effects on eye color and broader pigment traits	Important in pigmentation architecture, especially in Eurasian datasets
MC1R and ASIP	Regulation of melanocytic signaling and balance of pigment outputs	Strong effects on pigment-related traits and sun sensitivity, with marked population specificity

Table 2: *Table of genes involved in melanin production and skin pigmentation.*

Convergent evolution has played a significant role in shaping pigmentation diversity. Populations in Europe and East Asia have evolved lighter pigmentation through different genetic mechanisms, showing the flexibility of the underlying pathways (Guermazi & Saliba, 2025b). In European populations, lighter skin pigmentation evolved primarily through variants in *SLC24A5*, *SLC45A2*, *HERC2*/*OCA2*. In East Asian populations, this phenotype evolved due to variants in *DDB1* and *OPRM1*. Convergent evolution in lighter pigmentation emphasizes the importance of describing pigmentation biology as a shared pathway with multiple population-specific genetic architectures rather than as a simple, universal set of variants with uniform predictive power across populations (E. E. Quillen et al., 2019). Understanding pigmentation biology as multiple population-specific genetic architectures has implications for cross-population prediction models including those conducted in this present study, as SNPs identified in one population may not have the same effect in another.

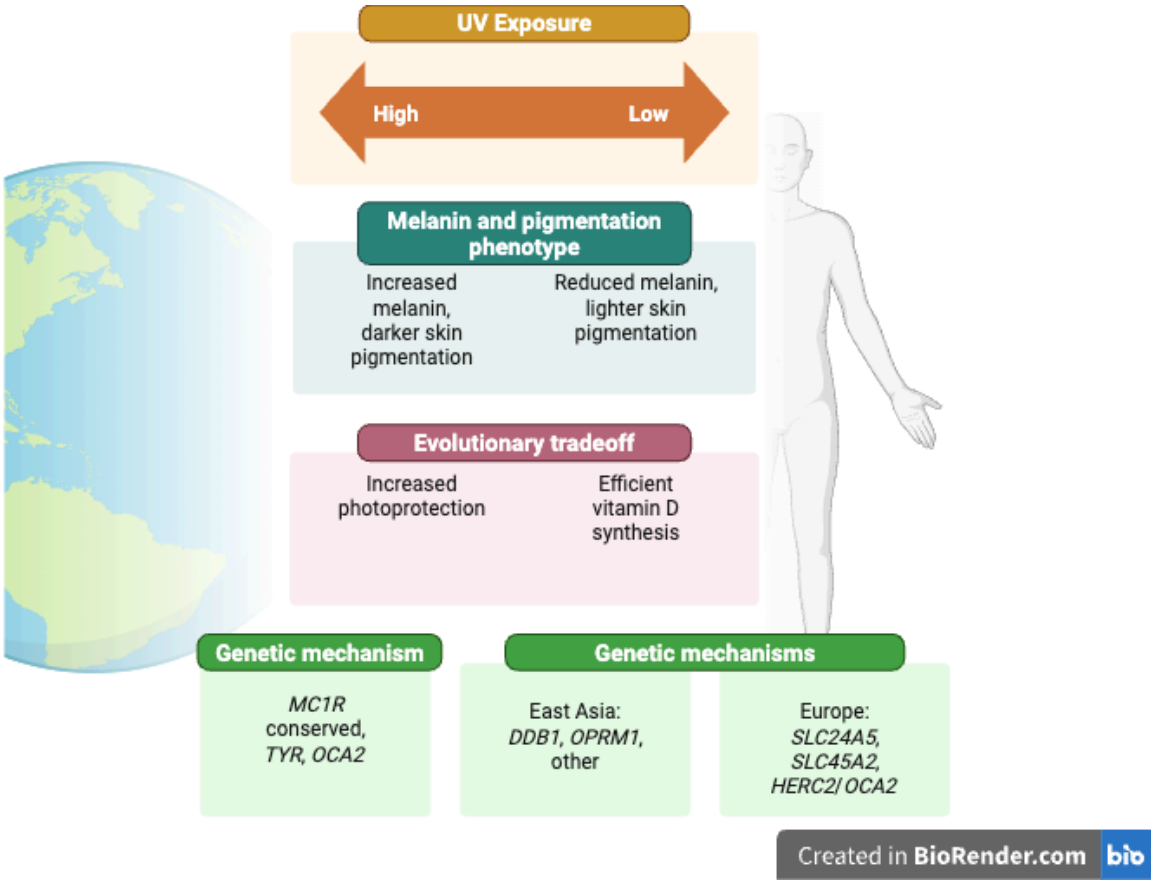


Figure 2: *Skin pigmentation evolution in response to UV radiation. Figure adapted from Guermazi & Saliba, 2025b.*

Looking to ancient DNA for questions about human skin pigmentation is an interesting opportunity to question whether the same selective pressures shaping modern human pigmentation variation acted similarly in archaic hominins. While studies of contemporary human variation suggest strong links between UV radiation, vitamin D synthesis, folate retention, and skin pigmentation, archaic genomes allow the interrogation of these evolutionary pressures in populations such as Neanderthals and Denisovans, or whether different environmental and genetic factors led to distinct pigmentation outcomes.

1.5 Archaic genome availability

Ancient DNA (aDNA) can persist for up to hundreds of thousands of years in skeletal remains, and can even be extracted from sediments (Orlando et al., 2021). However, extracting this kind of DNA poses a challenge. Ancient DNA is typically present in very small amounts, is highly fragmented, and contaminated with other endogenous organic molecules which inhibit enzymatic DNA manipulations which must be performed prior to sequencing, or exogenous DNA. Ancient DNA fragment length does not decrease uniformly over time; however, the frequency of cytosine to uracil deaminations near the 5' ends of the strands do increase over time (Sawyer et al., 2012). These damage patterns can introduce errors in sequencing, but can also serve as authentication signals and allow for the purification of ancient DNA with modern contaminants (Skoglund et al., 2014). There are also methods for sequencing DNA with this pattern of damage, such as the inclusion of uracil-DNA glycosylase (UDG), which catalyses the release of free uracil from uracil-containing DNA (Orlando et al., 2021).

High-coverage archaic nuclear genomes remain an exceptionally small record, currently including only five individuals. This narrow archaeological base limits locus-level inference, as the archaic dataset includes a handful of individuals separated by large gaps in time rather than from a representative archaic sample.

Available high-coverage archaic genomes

Specimen	Species	Location	Age
Denisova 5	Neanderthal	Denisova cave	122 ka
Chagyrskaya 8	Neanderthal	Chagyrskaya cave	60 ka
Vindija 33.19	Neanderthal	Vindija cave	52 ka
Denisova 25	Denisovan	Denisova cave	200 ka
Denisova 3	Denisovan	Denisova cave	72 ka

Table 3: *Age, species, and location of the five available high-coverage archaic genomes.*

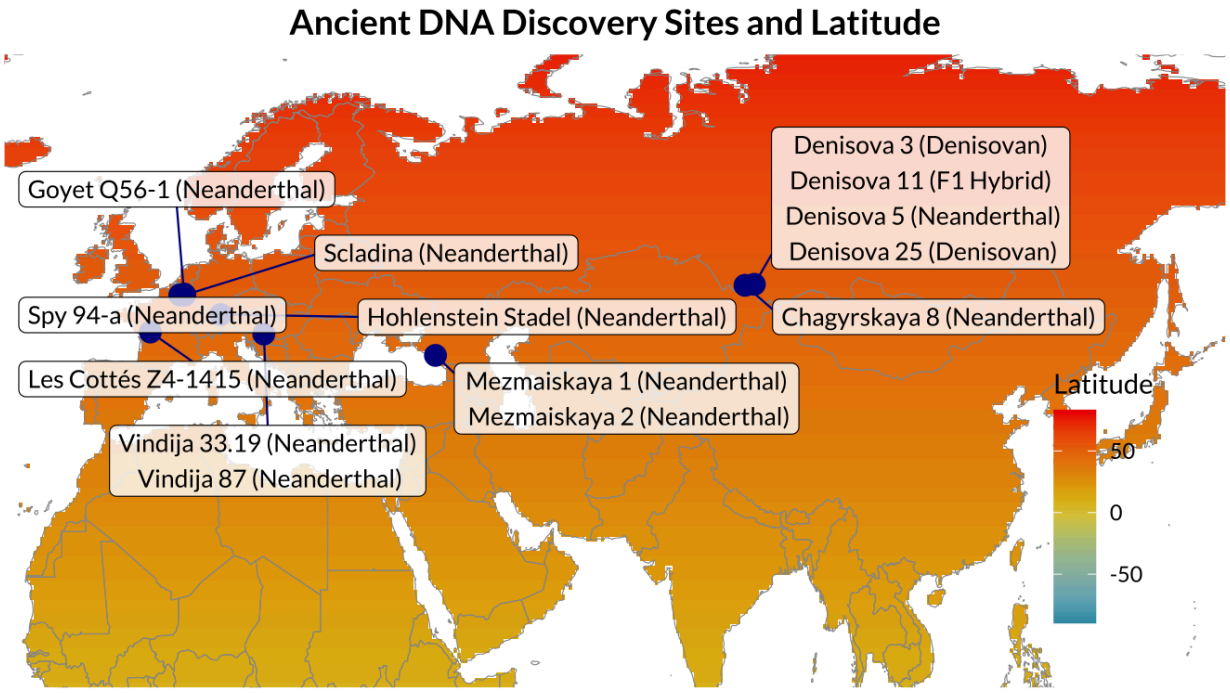


Figure 3: *Map of discovery sites of archaic genomic samples with latitude. Available archaic genomic samples all come from similar latitudes.*

1.6.1 Denisova 3 (Denisovan)

The first high coverage Denisovan genome was extracted from a finger phalanx from Denisova Cave in the Altai Mountains of Siberia and sequenced to ~30x genomic coverage. This phalanx from a juvenile female hominin was found in a layer which has been dated to 50-30ka. Initially, this phalanx was sequenced to ~1.9x coverage, which provided sufficient evidence for Denisovans to be established as a distinct archaic group (Reich et al., 2010). Later DNA library

preparation methods allowed the DNA from this phalanx to be reconstructed to 30x coverage (Meyer et al., 2012).

The archaeological setting helps explain the importance of this sample. Denisova Cave preserves a long depositional sequence with relatively intact microstratigraphy in parts of the site, with favorable sedimentary contexts for preservation (Morley et al., 2019). Denisova Cave also contains evidence of Denisovan, Neanderthal, and human occupations, as well as a F1 Neanderthal-Denisovan hybrid (Denisova 11) (Jacobs et al., 2019). The combination of cold continental conditions, the protection of a cave, and methodological advances allowed for the recovery of a high-coverage genome.

1.6.2 Denisova 25 (Denisovan)

The second high quality Denisovan genome was reconstructed from a molar found at Denisova cave, dated to ~200-170ka (Peyrégne et al., 2025). Denisova 25 is a male individual which was sequenced to 23.6x coverage. Archaeologically, this specimen pushes the Denisovan record deeper into the Middle Pleistocene. This drastic temporal separation between the two available high-coverage Denisovan genomes provides an exciting opportunity for comparison.

1.6.3 Denisova 5 (Altai Neanderthal)

In the same layer as the Denisova 3 finger phalanx was excavated, a proximal toe phalanx was recovered (Prüfer et al., 2014). Denisova 5, or the Altai Neanderthal, is from an adult female and predates the later high-coverage Neanderthals from Vindija and Chagyrskaya. Interestingly, this toe phalanx was found in the lowest sublayer, indicating that it is older than the Denisova 3

sample. Four DNA libraries were generated from this sample, providing a total of 52x coverage. This was the first high coverage Neanderthal genome sequenced.

1.6.4 Vindija 33.19 (Neanderthal)

The second high quality Neanderthal genome is a 30x coverage genome sequence from a Neanderthal found in Vindija, Croatia, dated to approximately 65-50ka (Prüfer et al., 2017). The initial analyses of this genome suggested that the Vindija 33.19 Neanderthal is more closely related to the Neanderthals which interbred with *Homo sapiens* than the Altai Neanderthal. This genome allowed more introgressed Neanderthal DNA to be identified in the modern human genome, including variants associated with low-density lipoprotein cholesterol concentrations and vitamin D, rheumatoid arthritis, eating disorders, and schizophrenia (Prüfer et al., 2017).

1.6.5 Chagyrskaya 8 (Neanderthal)

The third high quality neanderthal genome was obtained from the phalanx of a female in Chagyrskaya Cave in the Altai Mountains, Russia, which was sequenced to 27x coverage and dated to approximately 80ka (Mafessoni et al., 2020). Archaeological work at the site recovered a large Middle Paleolithic assemblage, numerous Neanderthal remains, and evidence for occupations dated roughly to 59-49ka (Kolobova et al., 2020). In the initial analyses of this genome, it was found that this individual was more closely related to Neanderthals in western Europe than those who lived in Denisova Cave earlier (Mafessoni et al., 2020).

1.6.6 The sparse record

Additional archaic genomes exist at low coverage, including the early Denisova 3 draft genome, the Denisova 11 ("Denny") admixed individual, and several later Neanderthals sequenced at ~1-2.7x coverage (Hajdinjak et al., 2018; Reich et al., 2010; Sawyer et al., 2015; Slon et al., 2018). These genomes are important for population history, but they are less reliable due to high levels of missingness and uncertainty. For that reason, the availability of high-coverage genomes limits phenotypic inference. The high-coverage record includes only five individuals, ranging from 200-50ka, and a narrow geographic range.

These genomes allow us to infer the pigmentation genotypes of those individuals based on loci of relevance discovered by GWAS of modern human populations. This limited dataset does not provide enough statistical power to make claims about the full pigmentation range of Neanderthals or Denisovans as groups. This motivates the aim of this study to provide a locus-level catalog of pigmentation-associated SNPs in available archaic genomes, and then compare those with variation in diverse modern humans.

Beyond DNA, dental proteomes, or the proteins found in oral tissues and fluids, have been sequenced and could provide additional interesting analyses. Dental proteomes extracted from a *Homo erectus* specimen in Dmanisi, Georgia, dated to ~1.77ma, and an *H. antecessor* specimen from Atapuerca, Spain, dated to 772-949ka, are the oldest protein sequence data available. Proteomic data does not provide direct information about pigmentation; however, it can inform reconstructions of diet and ecology. The availability of dental proteomes in conjugation with genomic data could provide interesting insight into the diets of our ancient ancestors, which would allow for a more thorough analysis of constitutive skin pigmentation as a function of vitamin D intake.

1.6 Archaic introgression

Archaic introgression has revealed pigmentation associated variants in the modern human genome which have originated from archaic haplotypes. Introgression, when DNA from an archaic group of interest remains in the modern genome due to past admixture, allows for an opportunity to infer phenotypes of the archaic group by observing the effect in the living population. Alleles in and around *BNC2*, which are related to skin pigmentation, have been shown to be introgressed from Neanderthals (Dannemann & Kelso, 2017; Racimo et al., 2015). However, the phenotypic effects of these introgressed variants are small compared to the large-effect alleles which dominate pigmentation differences in modern populations.

Ancient *Homo sapiens* genomes, as opposed to the genomes of our archaic relatives such as Neanderthals and Denisovans, provide time series data to analyze the evolution of pigmentation. In a landmark study, the genomes of 1,158 West Eurasians across 40,000 years were studied, showing that skin lightening was driven by concentrated sweeps at SLC24A5, SLC45A2, TYR, OCA2/HERC2, with most of the change occurring within the last 5-10ka (Ju & Mathieson, 2021b). When demographic history and stratification was properly modeled, polygenic selection tests on the full set of skin pigmentation GWAS variants failed to detect a signal. A significant signal was only detected when these tests were restricted to the top 5 largest GWAS effect size SNPs, suggesting that selection targeted a small subset of loci rather than diffuse selection (Ju & Mathieson, 2021b).

2.0 Methods

2.1 Modern comparative dataset

Modern human genomes were obtained from the Simons Genome Diversity Project (SGDP) (Mallick et al., 2016). A panel of 15 individuals available through the SGDP representing broad global diversity was constructed to provide a stable reference frame for principal component analysis (PCA). Variant call files (VCFs) were downloaded as whole-genome datasets aligned to the GRCh37/hg19 reference genome, and subsetted to a genome wide SNP panel and a panel of 222 pigmentation-related SNPs.

Individual SGDP BCFs were indexed using BCFtools and merged into a single multi-sample BCF using BCFtools merge (Li, 2011).

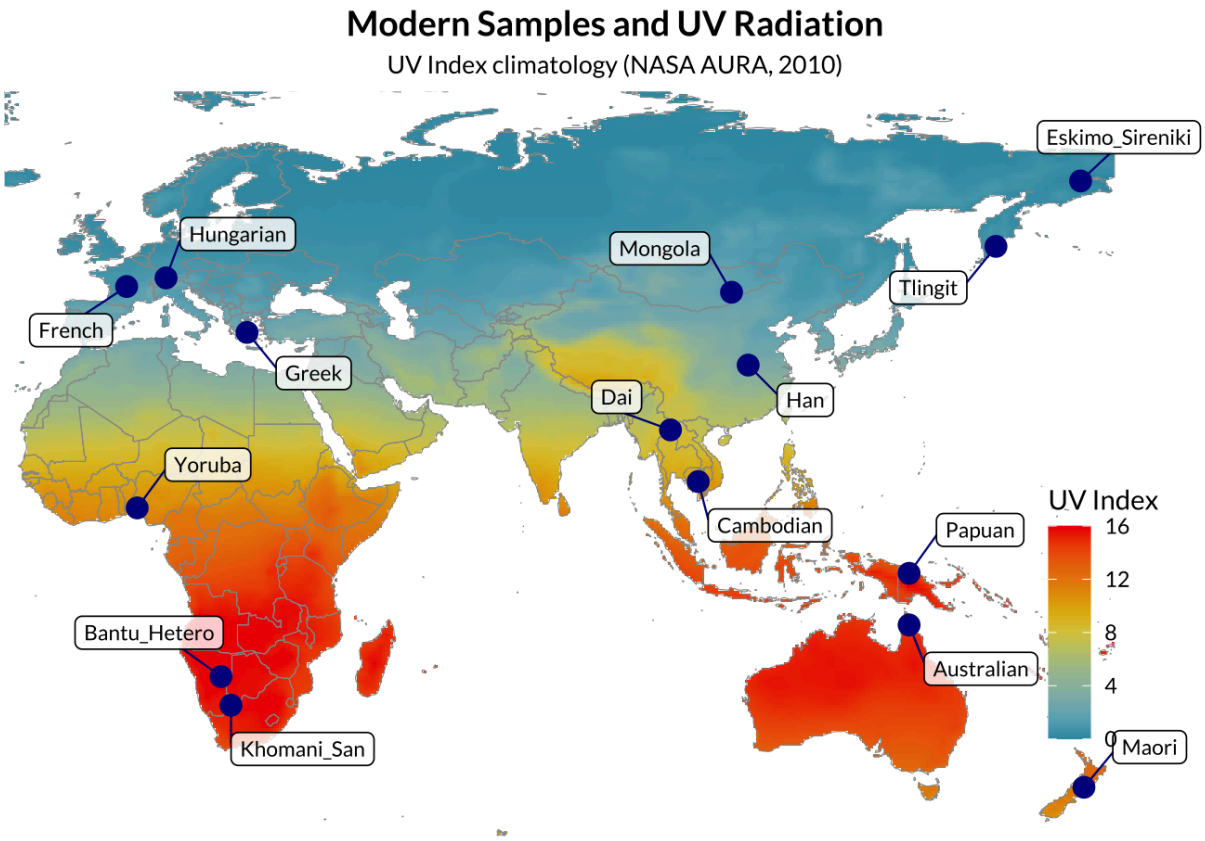


Figure 4: *Map of 15 sampled Simons Genome Diversity Project genomes with December 2010 average UV radiation (NASA AURA).*

2.2 SNP panels

Two SNP panels were used, in order to allow for comparison of genome-wide ancestry patterns with structure driven by a targeted functional SNP set. For pigmentation-specific analyses, a panel consisting of 222 SNPs previously associated with pigmentation phenotypes published on the NHGRI-EBI catalog of human genome-wide association studies was used (Sollis et al., 2023). A whole-genome panel of 34,418,131 SNPs found to be polymorphic in the

Simons Genome Diversity Project was used as a reference for genome-wide ancestry patterns (Fan et al., 2019).

2.3 Archaic Samples

Archaic genomes included archaic *Homo neanderthalensis* and Denisovan individuals. BAM files were previously mapped to the GRCh37/hg19 reference genome and sorted and indexed using SAMtools (Li et al., 2009).

Sequencing read alignments were processed using SAMtools. The input BAM file was first indexed, then coordinate-sorted and re-indexed to ensure compatibility with downstream analyses. Read depth was calculated at the specified pigmentation-related SNP panel using SAMtools depth including sites with zero coverage.

Reads were pre-filtered to the same set of pigmentation-associated loci prior to genotype calling. Genotypes were called using a pileup-based approach without depth filtering in order to maximize retention of loci due to the high degree of missingness in the archaic dataset. The resulting variant calls were converted to PLINK format.

2.5 Conversion to PLINK and LD pruning

Modern SGDP VCFs were converted to PLINK binary format using PLINK 2.0 (Chang et al., 2015). Variants were restricted to biallelic SNPs. Archaic genomes were processed separately using the same filtering criteria. In order to reduce the impact of linkage disequilibrium (LD), SNPs were pruned using a sliding window approach. Variants with minor allele frequency (MAF) less than 5% were excluded. Allele frequencies were computed from the

modern dataset and were subsequently used in both PCA computation and projection to ensure consistent scaling.

2.6 Principal component analysis

PCA was performed to characterize genetic variation in modern human populations and to project archaic genomes onto this reference space. Archaic individuals were not included in the eigenvector computation due to missingness. Instead, they were projected onto the modern PCA space using least squares projection, which is appropriate for the high level of missingness in the archaic dataset (Ausmees et al., 2022; Zabel et al., 2025).

PCA was performed on the LD-pruned modern dataset with 10 components and allele weight outputs. Archaic individuals were projected onto the PCA space defined by the modern samples with the allele weights derived from the PCA. Projection was performed in PLINK 2.0 with allele weights and frequency normalization. PCA results were visualized in R using the ggplot2 package.

Separate PCA analyses were conducted for the whole genome and pigmentation-associated SNP panels to allow for comparison of genome-wide ancestry patterns with structure driven by a targeted functional SNP set. Due to the low coverage of archaic genomes, projections were based on the subset of SNPs shared between each individual and the modern reference panel. As a result, the number of SNPs contributing to each projection varied across samples. Given the limited coverage of archaic genomes, no imputation was performed and projections only reflect observed genotypes.

2.7 Probability that a majority-rule consensus call recovers the true nucleotide

The probability that a consensus base call is correct was modeled as a function of sequencing depth and per-base error rate using a binomial distribution. For a given read depth (n), we assumed that each read represents an independent observation of the underlying true base with a constant per-base error rate (ϵ). The probability that an individual read reports the correct nucleotide was defined as $p = 1 - \epsilon$, assuming independence among reads and symmetric errors. Consensus accuracy was defined as a simple majority rule, meaning that a site was correctly called if more than half of the reads supported the true base $k = \lfloor n/2 \rfloor + 1$. For each combination of depth and error rate, the probability that the consensus call is correct is given by the cumulative binomial distribution:

$$P(correct) = 1 - \sum_{i=0}^{k-1} \binom{n}{i} p^{i} (1 - p)^{n-i}$$

This formulation models errors as independent Bernoulli trials; however, it is important to note that ancient DNA damage violates this assumption due to systematic position-dependent errors.

Four error profiles were defined to reflect both sequencing platform characteristics and post-mortem damage processes in ancient DNA. Modern Illumina short-read sequencing publishes low substitution error rates (~0.1%) (Pfeiffer et al., 2018), while Oxford Nanopore long-read Q20+ technology has slightly higher error rates (0.5%) (Nanopore Sequencing Accuracy, n.d.). Ancient DNA error profiles are higher due to damage patterns such as cytosine deamination, which increase both absolute and position-dependent error rates, particularly in untreated libraries where misincorporation rates can exceed 10-20% (Briggs et al., 2007).

UDG-treated libraries reduce these errors substantially, yielding error rates of ~1% (Sawyer et al., 2012).

To assess the impact of systematic position-dependent damage typical of ancient DNA, a position-dependent error model was implemented in which per-base error rates vary as a function of distance from fragment ends. Error probabilities were parametrized to reflect elevated misincorporation rates at strand ends, with untreated aDNA modeled as increasing from 5% in the interior to 20% at fragment ends, and UDG-treated aDNA libraries ranging from 1-2%. In this framework, each read was assigned an independent position along a fixed read length, and consensus accuracy was estimated using 1000 repetitions of a Monte Carlo simulation. All analyses were performed in R and visualized using ggplot2.

2.8 Mixed effects modeling

To understand what factors drive missingness in archaic genomic samples, a mixed effects logistic regression was conducted on the sample of pigmentation-related SNPs in the archaic individuals. Technical factors such as sample and overall coverage were accounted for as random effects. Coverage is included as a random effect, allowing the baseline probability of missingness to differ based on overall sequence quality. Sample was also included as a random effect, allowing for individual differences in preservation to vary. Mixed effects models were conducted in the lme4 package in R.

3.0 Results

3.1 There is a lot of variation in sample coverage

A heatmap of log-transformed sequencing depth including where depth = 0 shows substantial heterogeneity in sequencing coverage across both samples and positions. This heatmap shows non-random missingness and strong in-sample correlation.

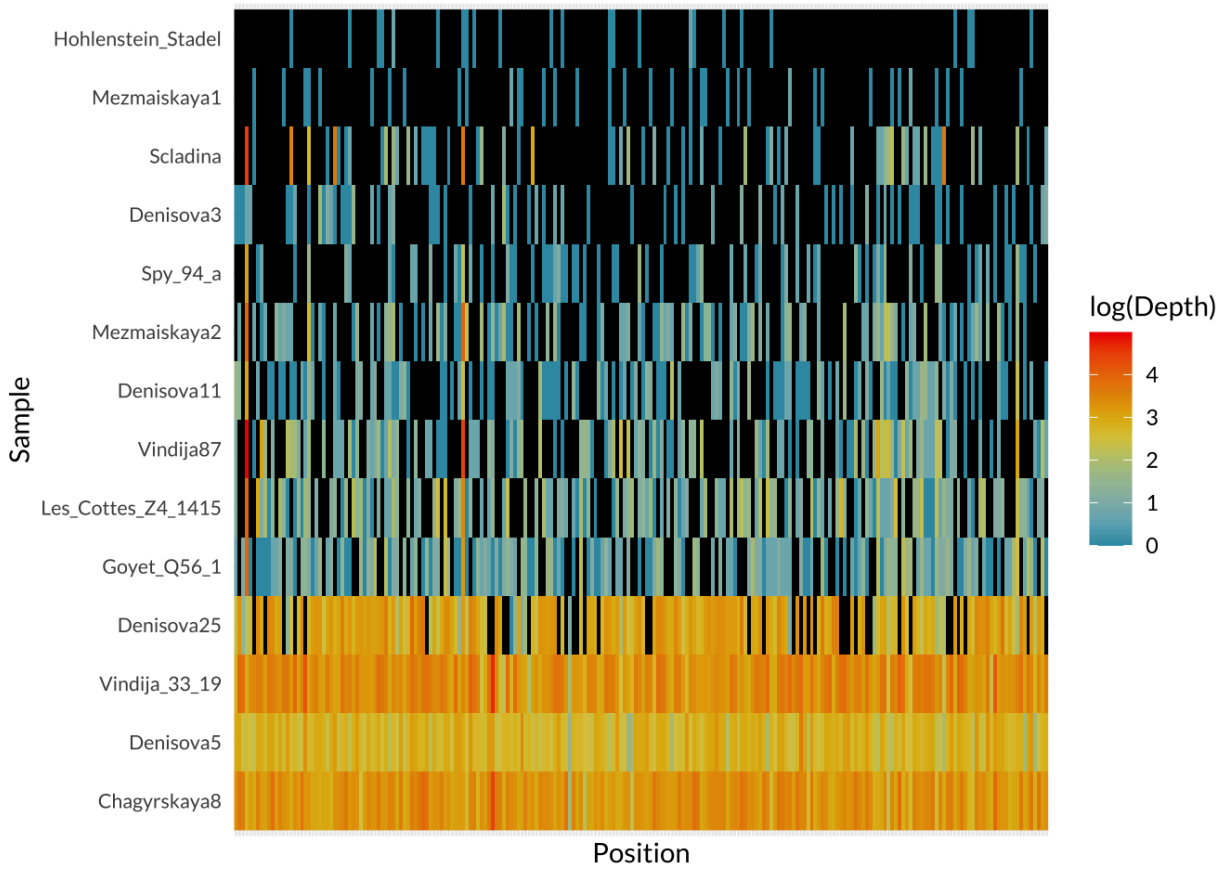


Figure 5: *Heatmap of log-transformed sequencing depth across archaic samples with missing coverage (depth = 0) included as black.*

There is strong in-sample correlation, as higher coverage samples show consistently high coverage shown as log-transformed depth is consistently above 3. Conversely, low coverage samples have high-levels of missingness and overall low sequence depth, shown as log-transformed depth consistently below 2. Within-sample variation shows that there is not uniform coverage and reveals some locus-level dropout. Interestingly, Denisova 25 (a high-coverage sample) shows some missingness.

Missingness also appears to be non-random, as missingness is not evenly distributed. Some positions show absence across many samples (min: 1; Q1: 5; Median: 6; Q3: 7; max 11). For example, position 50456592 on chromosome 22 is missing in 11 of the 14 samples. This indicates that some regions are systematically harder to recover.

Some individuals provide near-complete data whereas others are incredibly sparse, and even high-coverage samples contain missingness. This emphasizes the importance of accounting for structured missingness in downstream analyses.

3.2 Sample age is a significant predictor of the probability of missingness

A mixed effects logistic model shows that the probability of missingness is strongly influenced by age ($p < 0.001$), indicating that for each one standard deviation increase in age, the odds of missingness are 3.9 times higher, accounting for the chromosome and random effects. The effect of chromosome is significant for chromosomes 5, 7, 9, 13, 15, 16, and 22. Because of the small sample and non-uniform distribution of SNPs across the genome, this effect could be capturing structural influences, such as distance from a nucleosome or the end of a strand.

3.3 Untreated ancient DNA is significantly affected by coverage depth

Consensus accuracy increased with sequencing depth across all error profiles, but the rate of improvement varied substantially based on the underlying error rate. For modern DNA error profiles, near-perfect consensus accuracy was achieved at a sequence depth of 4. Ancient DNA error profiles converge to near-perfect consensus accuracy at a sequence depth of 10.

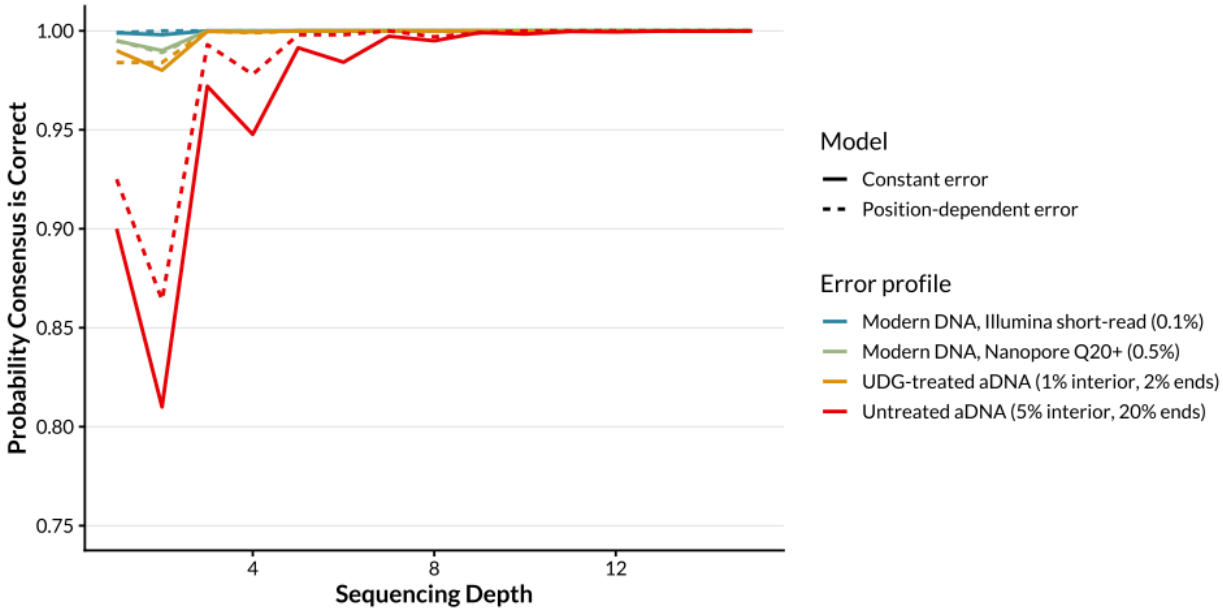


Figure 6: *Probability consensus base call is correct is modeled as a function of sequencing depth, error profile, and position in read length.*

Untreated ancient DNA is far more sensitive to low coverage than both modern DNA and UDG-treated aDNA. The constant error model assuming independent Bernoulli trials is conservative, as allowing for position-dependent error is less severe than assuming uniformly high error rates. The position-dependent shows slightly faster convergence to near-perfect consensus accuracy, especially for untreated aDNA.

3.4 Archaic samples cluster near Eurasians in whole-genome SNPs

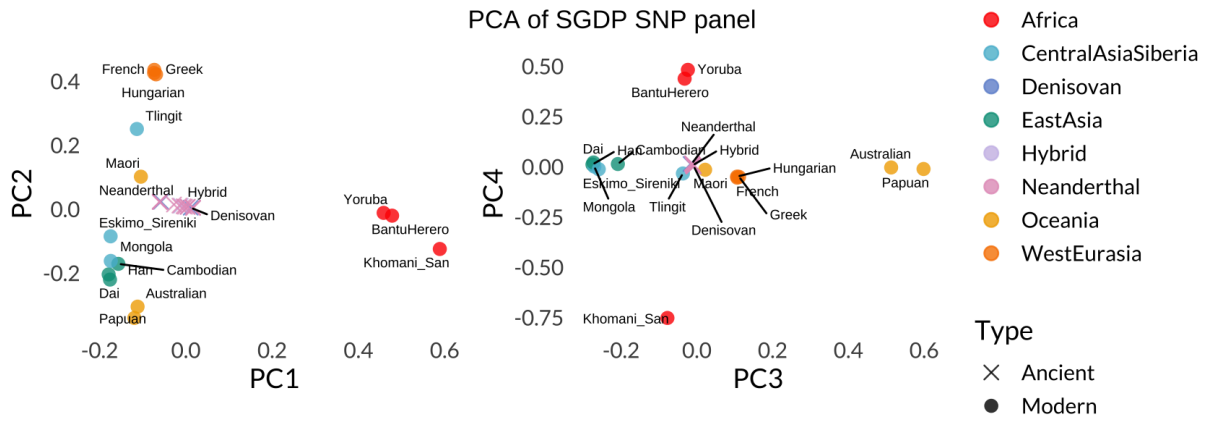


Figure 7: *Principal component analysis of polymorphic SGDP loci.*

The PCA of whole-genome SNPs shows clear population structure. In PC1-PC2, there is a strong separation between African and non-African populations along PC1. Within non-Africans, Europeans (WestEurasia), East Asians, and Oceanians form distinct clusters along PC2. Archaic individuals cluster near non-Africans on PC1 and form a distinct cluster on PC2. PC3 and PC4 show that Africans separate along PC4, with Yoruba and Bantu clustering together and Khomani San falling on the opposite end of the axis. Oceanians are clearly distinct along PC3. East Asians cluster tightly, distinct from both Oceanians and Europeans. Again, archaic individuals cluster near non-African populations. Whole-genome SNP data captures population history and demographic structure, producing clear clustering by continental ancestry and potentially reflecting introgression.

3.5 Pigmentation SNPs lack clear geographic structure

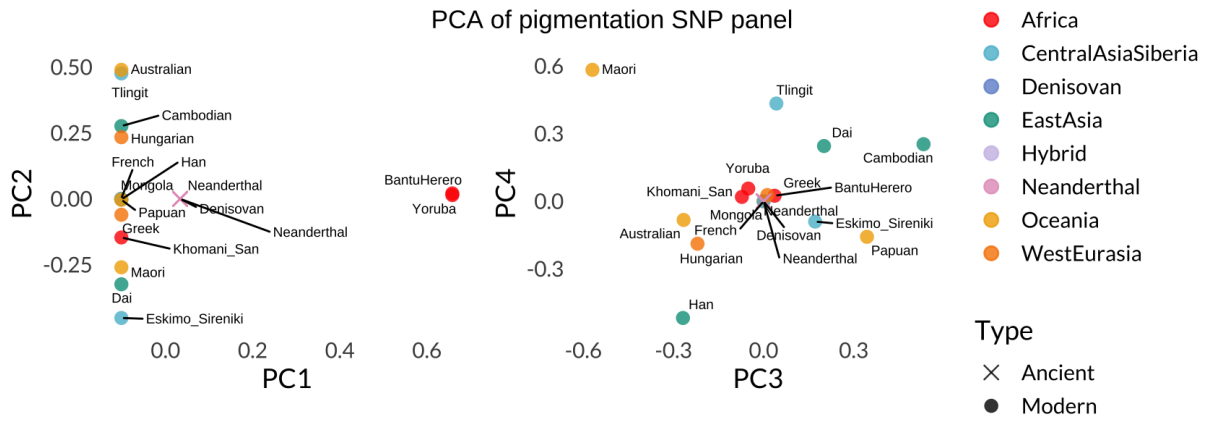


Figure 8: *Principal component analysis of 222 pigmentation-related SNPs.*

The PCA of pigmentation-associated SNPs shows minimal geographic structuring and greater overlap among populations compared to the genome-wide panel. In PCs 1 and 2, most populations fall closely near the origin, indicating limited differentiation captured by pigmentation loci alone. PC1 differentiates Bantu and Yoruba from the remaining samples.

In PC3 and PC4, there is some separation of Oceanians and East Asians; however, these groups do not form discrete clusters. Archaic individuals fall near the African populations. Additionally, the Khomani San sample is not clustering with other African populations.

4.0 Discussion

A major limitation of this project comes from the distribution of available archaic genomes. All currently sequenced archaic genomes come from samples from relatively high-latitude Eurasian contexts. This biased sample negates the ability to generalize

pigmentation variation across archaic hominins as the entire sample is derived from a geographically restricted sample, likely shaped by local environmental pressures. As a result, even high-confidence genotype calls cannot be assumed to represent species-wide phenotypic patterns and instead simply characterize the available sequences.

Beyond the limitation of specimen availability, the results demonstrate that the available genetic sequences complicate phenotypic inference. Sequencing depth varies dramatically both within and across samples. Importantly, missingness is not at random. Even high-coverage samples (e.g., Denisova 25) exhibit missingness. These patterns of missingness violate assumptions of uniform data quality and bias downstream analyses. The mixed-effects model further reinforces that missingness is structured. Sample age significantly increases the probability of missing data, even after accounting for technical factors and individual-level differences in preservation. Chromosome-level effects suggest that genomic context, such as the distance from a nucleosome or the end of a strand, may further shape what data is recoverable. Together, these results highlight that the absence of data at a locus cannot be treated as neutral.

These limitations are especially important when considering the relationship between sequence depth and confidence. Untreated ancient DNA requires substantially higher depth of coverage (~10x) to approach reliable consensus calls compared to modern DNA (~4x). At lower depths, confidence remains relatively low. This is important because many archaic genomes fall below these sequencing depth thresholds. Consequently, genotype calls at low coverage are inherently uncertain.

These findings provide important context for interpreting earlier studies that attempted to reconstruct archaic pigmentation from limited genetic data. Early attempts to infer pigmentation phenotypes in *Homo neanderthalensis*, such as Lalueza-Fox et al. (2007), focused on

sequencing the *MC1R* gene and interpreted some variants as consistent with lighter skin and red hair. However, there were several key issues with this study. First, one of the two purported Neanderthal samples examined in this study was later found to be from an anatomically modern human (Caramelli et al., 2006; Talamo et al., 2016). Second, these studies relied on low-coverage data from a single locus, which ignores the polygenic architecture of pigmentation.

A comparison of genome-wide and pigmentation-specific PCAs suggest an interesting difference between the genetic architecture of pigmentation and ancestry. At the whole-genome level, PCA captures patterns of human population structure shaped by ancestry and geography. The positioning of archaic individuals closer to non-Africans in this space aligns well with our understanding of introgression.

In contrast, the pigmentation-specific PCA does not reflect the expected clustering. This indicates that pigmentation may be influenced by a relatively small number of loci with complex evolutionary histories. Interestingly, archaic individuals fall nearer to African samples in this PC space. Additionally, the location of the Khomani San sample in the pigmentation-specific PC space complicates our current methods in pigmentation research. Khomani San individuals are often used as outgroups in genetic studies due to deep population divergence between Khomani San groups and all other human populations (Schlebusch et al., 2012, 2017; Veeramah et al., 2012). The separation of this sample from other African individuals in the pigmentation PCA may reflect unique evolutionary pressures or underrepresentation of African-specific variants in the SNP panel used.

Another question for pigmentation genetics, and population genetics broadly, in archaic DNA is how we choose to define populations. By defining populations for analyses of complex traits, we artificially discretize what is often a spatially-governed gradient of human diversity

(Bradburd & Ralph, 2019). Categorization, however, defines the study of population genetics (as you have to first define populations in order to study them). This is reflected in the varying resolutions with which we define our populations for analyses. Using ancient and archaic DNA as a sort of time-series data for population genetics adds an additional dimension to this choice. Increasingly, population geneticists are questioning these definitions in modern populations (Templeton, 2019). The currently available archaic genomes cover a vast temporal range, raising questions about how we should treat these samples.

The geographic restriction of available genomes, non-random missingness, and uneven sequencing depth means that the presence and absence of genetic variants are shaped by biases in preservation, recovery, and sampling. As a result, phenotypic inferences drawn from this data risk overinterpreting incomplete or skewed signals. Moving forward, interpretations of archaic pigmentation should be framed as provisional and context-dependent, with high levels of caution exercised when attempting to make species-wide generalizations.

References

Alaluf, S., Atkins, D., Barrett, K., Blount, M., Carter, N., & Heath, A. (2002). Ethnic Variation in Melanin Content and Composition in Photoexposed and Photoprotected Human Skin. *Pigment Cell Research*, *15*(2), 112–118. https://doi.org/10.1034/j.1600-0749.2002.1o071.x

Ausmees, K., Sanchez-Quinto, F., Jakobsson, M., & Nettelblad, C. (2022). An empirical evaluation of genotype imputation of ancient DNA. *G3: Genes|Genomes|Genetics*, *12*(6), jkac089. https://doi.org/10.1093/g3journal/jkac089

Borradale, D., Isenring, E., Hacker, E., & Kimlin, M. G. (2014). Exposure to solar ultraviolet radiation is associated with a decreased folate status in women of childbearing age. *Journal of Photochemistry and Photobiology B: Biology*, *131*, 90–95. https://doi.org/10.1016/j.jphotobiol.2014.01.002

Bradburd, G. S., & Ralph, P. L. (2019). Spatial Population Genetics: It's About Time. *Annual Review of Ecology, Evolution, and Systematics*, *50*(Volume 50, 2019), 427–449. https://doi.org/10.1146/annurev-ecolsys-110316-022659

Brand, C. M., Colbran, L. L., & Capra, J. A. (2022). Predicting Archaic Hominin Phenotypes from Genomic Data. *Annual Review of Genomics and Human Genetics*, *23*, 591–612. https://doi.org/10.1146/annurev-genom-111521-121903

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14616–14621. https://doi.org/10.1073/pnas.0704665104

Caramelli, D., Lalueza-Fox, C., Condemi, S., Longo, L., Milani, L., Manfredini, A., de Saint Pierre, M., Adoni, F., Lari, M., Giunti, P., Ricci, S., Casoli, A., Calafell, F., Mallegni, F., Bertranpetit, J., Stanyon, R., Bertorelle, G., & Barbujani, G. (2006). A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Current Biology*, *16*(16), R630–R632. https://doi.org/10.1016/j.cub.2006.07.043

Cerqueira, C. C., Paixão-Côrtes, V. R., Zambra, F. M. B., Salzano, F., Hünemeier, T., & Bortolini, M. (2012). Predicting homo pigmentation phenotype through genomic data: From neanderthal to James Watson. In *American Journal of Human Biology* (Vol. 24). https://doi.org/10.1002/ajhb.22263

Cerqueira, C. C. S., Paixão-Côrtes, V. R., Zambra, F. M. B., Salzano, F. M., Hünemeier, T., & Bortolini, M.-C. (2012). Predicting homo pigmentation phenotype through genomic data: From neanderthal to James Watson. *American Journal of Human Biology*, *24*(5), 705–709. https://doi.org/10.1002/ajhb.22263

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., Jewett, E., Ranciaro, A., Thompson, S., Lo, Y., Pfeifer, S. P., Jensen, J., Campbell, M., Beggs, W., Hormozdiari, F., Mpoloka, S., Mokone, G. G., Nyambo, T., Meskel, D. W., … Tishkoff, S. (2017). Loci associated with skin pigmentation identified in African populations. In *Science* (Vol. 358). https://doi.org/10.1126/science.aan8433

Czeizel, A. E., Dudás, I., Vereczkey, A., & Bánhidy, F. (2013). Folate Deficiency and Folic Acid Supplementation: The Prevention of Neural-Tube Defects and Congenital Heart Defects. *Nutrients*, *5*(11), 4760–4775. https://doi.org/10.3390/nu5114760

Dannemann, M., & Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *The American Journal of Human Genetics*, *101*(4), 578–589. https://doi.org/10.1016/j.ajhg.2017.09.010

Dannemann, M., Prüfer, K., & Kelso, J. (2017a). Functional implications of Neandertal introgression in modern humans. *Genome Biology*, *18*(1), 61. https://doi.org/10.1186/s13059-017-1181-7

Dannemann, M., Prüfer, K., & Kelso, J. (2017b). Functional implications of Neandertal introgression in modern humans. In *Genome Biology* (Vol. 18). https://doi.org/10.1186/s13059-017-1181-7

Del Bino, S., Ito, S., Sok, J., Nakanishi, Y., Bastien, P., Wakamatsu, K., & Bernerd, F. (2015). Chemical analysis of constitutive pigmentation of human epidermis reveals constant eumelanin to pheomelanin ratio. *Pigment Cell & Melanoma Research*, *28*(6), 707–717. https://doi.org/10.1111/pcmr.12410

Ding, Q., Hu, Y., Xu, S., Wang, C.-C., Li, H., Zhang, R., Yan, S., Wang, J., & Jin, L. (2014). Neanderthal origin of the haplotypes carrying the functional variant Val92Met in the MC1R in modern humans. In *Molecular biology and evolution: 31 8* (pp. 1994–2003). https://doi.org/10.1093/molbev/msu180

Fan, S., Kelly, D. E., Beltrame, M. H., Hansen, M. E. B., Mallick, S., Ranciaro, A., Hirbo, J., Thompson, S., Beggs, W., Nyambo, T., Omar, S. A., Meskel, D. W., Belay, G., Froment, A., Patterson, N., Reich, D., & Tishkoff, S. A. (2019). African evolutionary history inferred from whole genome sequence data of 44 indigenous African populations. *Genome Biology*, *20*(1), 82. https://doi.org/10.1186/s13059-019-1679-2

Feng, Y., McQuillan, M., & Tishkoff, S. (2021). Evolutionary genetics of skin pigmentation in African populations. In *Human molecular genetics*. https://doi.org/10.1093/hmg/ddab007

Ferrando-Bernal, M., Brand, C. M., & Capra, J. A. (2025). Inferring human phenotypes using ancient DNA: From molecules to populations. *Current Opinion in Genetics & Development*, *90*, 102283. https://doi.org/10.1016/j.gde.2024.102283

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., … Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, *328*(5979), 710–722. https://doi.org/10.1126/science.1188021

Guermazi, D., & Saliba, E. (2025a). The Genetics and Evolution of Human Pigmentation. *Biology*, *14*(8), 1026. https://doi.org/10.3390/biology14081026

Guermazi, D., & Saliba, E. (2025b). The Genetics and Evolution of Human Pigmentation. *Biology*, *14*(8), 1026. https://doi.org/10.3390/biology14081026

Hajdinjak, M., Fu, Q., Hübner, A., Petr, M., Mafessoni, F., Grote, S., Skoglund, P., Narasimham, V., Rougier, H., Crevecoeur, I., Semal, P., Soressi, M., Talamo, S., Hublin, J.-J., Gušić, I., Kućan, Ž., Rudan, P., Golovanova, L. V., Doronichev, V. B., … Kelso, J. (2018). Reconstructing the genetic history of late Neanderthals. *Nature*, *555*(7698), 652–656. https://doi.org/10.1038/nature26151

Hong, E. P., & Park, J. W. (2012). Sample Size and Statistical Power Calculation in Genetic Association Studies. *Genomics & Informatics*, *10*(2), 117–122. https://doi.org/10.5808/GI.2012.10.2.117

Hurbain, I., Romao, M., Sextius, P., Bourreau, E., Marchal, C., Bernerd, F., Duval, C., & Raposo, G. (2018). Melanosome Distribution in Keratinocytes in Different Skin Types: Melanosome Clusters Are Not Degradative Organelles. *Journal of Investigative Dermatology*, *138*(3), 647–656. https://doi.org/10.1016/j.jid.2017.09.039

Ito, S., & Wakamatsu, K. (2003). Quantitative Analysis of Eumelanin and Pheomelanin in Humans, Mice, and Other Animals: A Comparative Review. *Pigment Cell Research*, *16*(5), 523–531. https://doi.org/10.1034/j.1600-0749.2003.00072.x

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences*, *107*(supplement_2), 8962–8968. https://doi.org/10.1073/pnas.0914628107

Jablonski, N. G., & Chaplin, G. (2018). The roles of vitamin D and cutaneous vitamin D production in human evolution and health. *International Journal of Paleopathology, Vitamin D Deficiency in Bioarchaeology and beyond: The Study of Rickets and Osteomalacia in the Past*, *23*, 54–59. https://doi.org/10.1016/j.ijpp.2018.01.005

Jacobs, Z., Li, B., Shunkov, M. V., Kozlikin, M. B., Bolikhovskaya, N. S., Agadjanian, A. K., Uliyanov, V. A., Vasiliev, S. K., O'Gorman, K., Derevianko, A. P., & Roberts, R. G. (2019). Timing of archaic hominin occupation of Denisova Cave in southern Siberia. *Nature*, *565*(7741), 594–599. https://doi.org/10.1038/s41586-018-0843-2

Ju, D., & Mathieson, I. (2021a). The evolution of skin pigmentation-associated variation in West Eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(1), e2009227118. https://doi.org/10.1073/pnas.2009227118

Ju, D., & Mathieson, I. (2021b). The evolution of skin pigmentation-associated variation in West Eurasia. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(1), e2009227118. https://doi.org/10.1073/pnas.2009227118

Kang, J., Ramgolam, A., Vot, L. L., & Young, R. S. (2023). Functional archaic DNA regulates molecular variation and is associated with disease risk across global populations. In *bioRxiv*. https://doi.org/10.1101/2023.04.26.538367

Kolobova, K. A., Roberts, R. G., Chabai, V. P., Jacobs, Z., Krajcarz, M. T., Shalagina, A. V., Krivoshapkin, A. I., Li, B., Uthmeier, T., Markin, S. V., Morley, M. W., O'Gorman, K., Rudaya, N. A., Talamo, S., Viola, B., & Derevianko, A. P. (2020). Archaeological evidence for two separate dispersals of Neanderthals into southern Siberia. *Proceedings of the National Academy of Sciences*, *117*(6), 2879–2885. https://doi.org/10.1073/pnas.1918047117

Kondo, T., & Hearing, V. J. (2011). Update on the regulation of mammalian melanocyte function and skin pigmentation. *Expert Review of Dermatology*, *6*(1), 97–108. https://doi.org/10.1586/edm.10.70

Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J., & Hofreiter, M. (2007). A Melanocortin 1 Receptor Allele Suggests Varying Pigmentation Among Neanderthals. *Science*, *318*(5855), 1453–1455. https://doi.org/10.1126/science.1147417

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, J. Y., & Fisher, D. E. (2007). Melanocyte biology and skin pigmentation. *Nature*, *445*(7130), 843–850. https://doi.org/10.1038/nature05660

Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K. A., Viola, B., Markin, S. V., Chintalapati, M., Peyrégne, S., Skov, L., Skoglund, P., Krivoshapkin, A. I., Derevianko, A. P., Meyer, M., Kelso, J., Peter, B., Prüfer, K., & Pääbo, S. (2020). A high-coverage Neandertal genome from Chagyrskaya Cave. *Proceedings of the National Academy of Sciences*, *117*(26), 15132–15136. https://doi.org/10.1073/pnas.2004944117

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q., Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., … Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. https://doi.org/10.1038/nature18964

Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., & Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Scientific Data*, *11*(1), 182. https://doi.org/10.1038/s41597-024-03031-7

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Current clinical use of polygenic scores will risk exacerbating health disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Martin, A. R., Lin, M., Granka, J. M., Myrick, J. W., Liu, X., Sockell, A., Atkinson, E. G., Werely, C. J., Möller, M., Sandhu, M. S., Kingsley, D. M., Hoal, E. G., Liu, X., Daly, M. J., Feldman, M. W., Gignoux, C. R., Bustamante, C. D., & Henn, B. M. (2017). An Unexpectedly Complex Architecture for Skin Pigmentation in Africans. *Cell*, *171*(6), 1340-1353.e14. https://doi.org/10.1016/j.cell.2017.11.015

McArthur, E., Rinker, D., & Capra, J. A. (2020). *Quantifying the contribution of Neanderthal introgression to the heritability of complex traits* (p. 2020.06.08.140087). bioRxiv. https://doi.org/10.1101/2020.06.08.140087

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science*, *338*(6104), 222–226. https://doi.org/10.1126/science.1224344

Miyamura, Y., Coelho, S. G., Wolber, R., Miller, S. A., Wakamatsu, K., Zmudzka, B. Z., Ito, S., Smuda, C., Passeron, T., Choi, W., Batzer, J., Yamaguchi, Y., Beer, J. Z., & Hearing, V. J. (2007). Regulation of human skin pigmentation and responses to ultraviolet radiation. *Pigment Cell Research*, *20*(1), 2–13. https://doi.org/10.1111/j.1600-0749.2006.00358.x

Morley, M. W., Goldberg, P., Uliyanov, V. A., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Jacobs, Z., & Roberts, R. G. (2019). Hominin and animal activities in the microstratigraphic record from Denisova Cave (Altai Mountains, Russia). *Scientific Reports*, *9*(1), 13785. https://doi.org/10.1038/s41598-019-49930-3

Nanopore sequencing accuracy. (n.d.). Oxford Nanopore Technologies. Retrieved April 14, 2026, from https://nanoporetech.com/platform/accuracy

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*(1), 14. https://doi.org/10.1038/s43586-020-00011-0

Perry, G. H., Kistler, L., Kelaita, M. A., & Sams, A. J. (2015). Insights into hominin phenotypic and dietary evolution from ancient DNA sequence data. *Journal of Human Evolution, Special Issue: Ancient DNA and Human Evolution*, *79*, 55–63. https://doi.org/10.1016/j.jhevol.2014.10.018

Peyrégne, S., Massilani, D., Swiel, Y., Boyle, M. J., Iasi, L. N. M., Sümer, A. P., Mesa, A. B., Filippo, C. de, Viola, B., Essel, E., Nagel, S., Richter, J., Weihmann, A., Schellbach, B., Zeberg, H., Visagie, J., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., … Kelso, J. (2025). *A high-coverage genome from a 200,000-year-old Denisovan* (p. 2025.10.20.683404). bioRxiv. https://doi.org/10.1101/2025.10.20.683404

Pfeiffer, F., Gröber, C., Blank, M., Händler, K., Beyer, M., Schultze, J. L., & Mayer, G. (2018). Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports*, *8*(1), 10950. https://doi.org/10.1038/s41598-018-29325-6

Pinson, A., Xing, L., Namba, T., Kalebic, N., Peters, J., Oegema, C. E., Traikov, S., Reppe, K., Riesenberg, S., Maricic, T., Derihaci, R., Wimberger, P., Pääbo, S., & Huttner, W. B. (2022). Human TKTL1 implies greater neurogenesis in frontal neocortex of modern humans than Neanderthals. *Science*, *377*(6611), eabl6422. https://doi.org/10.1126/science.abl6422

Prüfer, K., de Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., … Pääbo, S. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, *358*(6363), 655–658. https://doi.org/10.1126/science.aao1887

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., De Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., … Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49. https://doi.org/10.1038/nature12886

Quillen, E. E., Norton, H. L., Parra, E. J., Lona-Durazo, F., Ang, K. C., Illiescu, F. M., Pearson, L. N., Shriver, M. D., Lasisi, T., Gokcumen, O., Starr, I., Lin, Y.-L., Martin, A. R., & Jablonski, N. G. (2019). Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. *American Journal of Physical Anthropology*, *168*(S67), 4–26. https://doi.org/10.1002/ajpa.23737

Quillen, E., Norton, H., Parra, E., Lona-Durazo, F., Ang, K. C., Illiescu, F. M., Pearson, L., Shriver, M., Lasisi, T., Gokcumen, O., Starr, I., Lin, Y.-L., Martin, A. R., & Jablonski, N. (2018). Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. In *American journal of physical anthropology: 168 Suppl 67* (pp. 4–26). https://doi.org/10.1002/ajpa.23737

Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, *16*(6), 359–371. https://doi.org/10.1038/nrg3936

Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., … Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*, *468*(7327), 1053–1060. https://doi.org/10.1038/nature09710

Rinker, D., Simonti, C. N., McArthur, E., Shaw, D., Hodges, E., & Capra, J. (2019). Neanderthal introgression reintroduced functional ancestral alleles lost in Eurasian populations. In *bioRxiv*. https://doi.org/10.1101/533257

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS ONE*, *7*(3), e34131. https://doi.org/10.1371/journal.pone.0034131

Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M. V., Derevianko, A. P., Prüfer, K., Kelso, J., & Pääbo, S. (2015). Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proceedings of the National Academy of Sciences*, *112*(51), 15696–15700. https://doi.org/10.1073/pnas.1519905112

Schlebusch, C. M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A. R., Vicente, M., Steyn, M., Soodyall, H., Lombard, M., & Jakobsson, M. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*, *358*(6363), 652–655. https://doi.org/10.1126/science.aao6266

Schlebusch, C. M., Skoglund, P., Sjödin, P., Gattepaille, L. M., Hernandez, D., Jay, F., Li, S., De Jongh, M., Singleton, A., Blum, M. G. B., Soodyall, H., & Jakobsson, M. (2012). Genomic Variation in Seven Khoe-San Groups Reveals Adaptation and Complex African History. *Science*, *338*(6105), 374–379. https://doi.org/10.1126/science.1227721

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences*, *111*(6), 2229–2234. (world). https://doi.org/10.1073/pnas.1318934111

Slon, V., Mafessoni, F., Vernot, B., de Filippo, C., Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Douka, K., Higham, T., Kozlikin, M. B., Shunkov, M. V., Derevianko, A. P., Kelso, J., Meyer, M., Prüfer, K., & Pääbo, S. (2018). The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature*, *561*(7721), 113–116. https://doi.org/10.1038/s41586-018-0455-x

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J. A. L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., … Harris, L. W. (2023). The NHGRI-EBI GWAS Catalog: Knowledgebase and deposition resource. *Nucleic Acids Research*, *51*(D1), D977–D985. https://doi.org/10.1093/nar/gkac1010

Sturm, R. A. (2009). Molecular genetics of human pigmentation diversity. *Human Molecular Genetics*, *18*(R1), R9–R17. https://doi.org/10.1093/hmg/ddp003

Talamo, S., Hajdinjak, M., Mannino, M. A., Fasani, L., Welker, F., Martini, F., Romagnoli, F., Zorzin, R., Meyer, M., & Hublin, J.-J. (2016). Direct radiocarbon dating and genetic analyses on the purported Neanderthal mandible from the Monti Lessini (Italy). *Scientific Reports*, *6*(1), 29144. https://doi.org/10.1038/srep29144

Templeton, A. R. (2019). Chapter 6—Gene Flow and Subdivided Populations. In *Human Population Genetics and Genomics* (pp. 155–193). Academic Press. https://doi.org/10.1016/B978-0-12-386025-5.00006-3

Veeramah, K. R., Wegmann, D., Woerner, A., Mendez, F. L., Watkins, J. C., Destro-Bisol, G., Soodyall, H., Louie, L., & Hammer, M. F. (2012). An Early Divergence of KhoeSan Ancestors from Those of Other Modern Humans Is Supported by an ABC-Based Analysis of Autosomal Resequencing Data. *Molecular Biology and Evolution*, *29*(2), 617–630. https://doi.org/10.1093/molbev/msr212

Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., & Kayser, M. (2013). The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Science International: Genetics*, *7*(1), 98–115. https://doi.org/10.1016/j.fsigen.2012.07.005

Wang, K., Prüfer, K., Krause-Kyora, B., Childebayeva, A., Schuenemann, V. J., Coia, V., Maixner, F., Zink, A., Schiffels, S., & Krause, J. (2023). High-coverage genome of the Tyrolean Iceman reveals unusually high Anatolian farmer ancestry. *Cell Genomics*, *3*(9). https://doi.org/10.1016/j.xgen.2023.100377

Witt, K., Villanea, F. A., Loughran, E., Zhang, X., & Huerta-Sánchez, E. (2022). Apportioning archaic variants among modern populations. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 377). https://doi.org/10.1098/rstb.2020.0411

Zabel, S., Breitling, S., Posth, C., & Nieselt, K. (2025). A probabilistic approach to visualize the effect of missing data on PCA in ancient human genomics. *BMC Genomics*, *26*(1), 537. https://doi.org/10.1186/s12864-025-11728-1