UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"*Biologia Evoluzionistica ed Ecologia*"

CICLO XXXVI

Robust inference of phenotypic traits from low-coverage ancient genomes

Coordinatore:

Chiar.mo Prof. Pierluigi Viaroli

Tutore:

Chiar.ma Prof.ssa Silvia Ghirotto

Co-tutore:

Chiar.mo Prof. Guido Barbujani

Dottoranda: Dott.ssa Silvia Perretti

Anni Accademici 2020/2021 – 2022/2023

Il dottorato in Biologia Evoluzionistica ed Ecologia – XXXVI ciclo è in convenzione tra le Università di Ferrara, Firenze e Parma (sede amministrativa).

ABSTRACT

Forensic DNA phenotyping (FDP) enables the prediction of specific phenotypic traits of an individual, such as physical appearance, biogeographical ancestry, and age, from minimal amounts of DNA obtained from biological samples. FDP is particularly useful in identifying unknown perpetrators of crimes, especially when standard forensic profiling does not provide sufficient information due to the absence of known suspects or profiles in national DNA databases.

A validated forensic phenotyping tool is the HIrisPlex-S system, designed for the simultaneous prediction of eye, hair, and skin colour. Based on the analysis of 41 genetic polymorphisms associated with pigmentation, this system allows for the estimation of individual probabilities for three categories of eye colour, four categories of hair colour, and five categories of skin colour, using exclusively genotypic data.

Originally developed to support specific criminal investigations, this tool has also been extensively used in recent years to predict the phenotype of ancient human skeletal remains.

To date, phenotypic prediction based on ancient, degraded, and low-coverage data has been carried out by assuming knowledge of the allelic and genotypic states of the samples at the 41 loci of interest. However, it is well known that directly calling genotypic variants from such data presents numerous challenges due to genetic material fragmentation, contamination risks, and degradation. Over time, degradation causes the DNA to break into very small fragments, making it difficult to obtain complete and accurate sequences. This results in low sequencing depth, which drastically affects the ability to reliably identify genotypes, thereby contributing to errors in phenotypic inference.

In the first part of this thesis, the robustness of the HIrisPlex-S system in phenotypic inference applied to ancient and low-coverage data was evaluated by testing it at various coverage levels. The required coverage level for ensuring robust inference was determined, and guidelines were established for when to apply the standard HIrisPlex-S protocol versus when to use alternative methods that account for uncertainty in genotypic calls. The evaluation was conducted by analysing the results of phenotypic predictions for eye, hair, and skin colour obtained through three prediction models: 1) the standard HIrisPlex-S protocol, based on direct variant calling; 2) a model specifically developed in this work, which integrates the classic system with genotype likelihoods to account for

the uncertainty associated with low-coverage data in genotype calling; 3) imputation, used to handle potentially missing data in low-coverage genomes.

To this scope, three available high-coverage samples were analysed: one from Palaeolithic Siberia, one from Mesolithic Sweden, and one from Bronze Age Germany. Using a downsampling procedure, phenotypic predictions for eye, hair, and skin colour were made by applying the three different prediction models. The results of phenotypic predictions for each approach and coverage level were compared with the "*true phenotype*" inferred from the original high-coverage genome. Subsequently, the frequency of correct predictions by each method at each coverage level was evaluated to determine the limits and the most suitable approach for phenotypic prediction from ancient data, thereby minimizing errors in the estimations.

The conclusions of this first part indicate that, regardless of the historical period of the analysed sample, it is advisable to apply the standard HIrisPlex-S protocol if minimum coverage levels of at least 8x at each of the 41 loci of interest are achieved. For coverages below 8x, the method developed and tested in this study, which uses genotype likelihoods for genotype estimation, is recommended instead.

The second part of this thesis project involved applying the protocol to a dataset of 348 Eurasian individuals ranging from the Upper Palaeolithic to the Iron Age, with the goal of examining the variation in eye, hair, and skin colour over the past 45,000 years. The resulting picture reveals a complex evolution of these traits: while dark phenotypes predominated for much of the studied period, light phenotypes began to appear from the Mesolithic onward in regions requiring environmental adaptation to high latitudes. Subsequently, these light phenotypes further spread in later periods due to a complex interaction between environmental adaptation and demographic events.

RIASSUNTO

La fenotipizzazione forense del DNA (FDP) consente di prevedere determinati tratti fenotipici di un individuo, quali l'aspetto fisico, l'ascendenza biogeografica e l'età, a partire da minime quantità di DNA provenienti da campioni biologici. Tale tecnica risulta particolarmente utile nell'identificazione di autori sconosciuti di reati, nei casi in cui la profilazione forense standard non fornisca informazioni sufficienti a causa della mancanza di sospetti noti o profili presenti in database nazionali di DNA.

Uno strumento di fenotipizzazione forense già sviluppato e validato è il sistema HIrisPlex-S, progettato per la previsione simultanea del colore degli occhi, dei capelli e della pelle. Basato sull'analisi di 41 polimorfismi genetici associati alla pigmentazione, questo sistema permette di stimare le probabilità individuali per tre categorie di colore degli occhi, quattro di colore dei capelli e cinque di colore della pelle, utilizzando esclusivamente i dati genotipici.

Questo strumento, originariamente sviluppato per supportare indagini criminali specifiche, è stato ampiamente utilizzato negli ultimi anni anche per prevedere il fenotipo di resti scheletrici umani antichi.

Fino ad oggi, la predizione fenotipica basata su dati antichi, degradati e a bassa copertura, è stata effettuata assumendo la conoscenza dello stato allelico e genotipico dei campioni nelle 41 posizioni di interesse. Tuttavia, è noto che la chiamata diretta delle varianti genotipiche da tali dati presenta numerose sfide, a causa della frammentazione del materiale genetico, del rischio di contaminazione e della degradazione. Quest'ultima provoca, nel tempo, la rottura del DNA in frammenti molto piccoli, rendendo difficile l'ottenimento di sequenze complete e accurate. Ciò comporta una bassa profondità di sequenziamento, che influisce drasticamente sulla capacità di identificare in modo affidabile i genotipi, contribuendo così ad errori nell'inferenza fenotipica.

Nella prima parte di questo lavoro di tesi è stata valutata la robustezza del sistema HIrisPlex-S nella procedura inferenziale fenotipica applicata a dati antichi e a bassa copertura, testandolo a diversi livelli di copertura. È stato determinato il livello di copertura necessario per garantire un'inferenza robusta e valutato quando applicare il protocollo classico di HIrisPlex-S e quando invece ricorrere a metodi alternativi che considerino l'incertezza nella chiamata genotipica. La valutazione è stata condotta analizzando i risultati delle predizioni fenotipiche per il colore degli occhi, dei capelli e della pelle, ottenuti tramite tre modelli di predizione: 1) il protocollo standard di HIrisPlex-S, basato sulla chiamata diretta delle varianti; 2) un modello specificamente sviluppato in questo lavoro, che integra

il sistema classico con le genotype likelihoods, per considerare l'incertezza associata ai dati a bassa copertura nella chiamata dei genotipi; 3) l'imputazione, utilizzata per gestire i dati potenzialmente mancanti in genomi a bassa copertura.

A tale scopo, sono stati analizzati tre campioni presenti in letteratura ad alta copertura: uno proveniente dalla Siberia del Paleolitico, uno dalla Svezia del Mesolitico e uno dalla Germania dell'Età del Bronzo. Utilizzando una procedura di abbassamento artificiale della copertura, sono state effettuate previsioni fenotipiche sui colori degli occhi, dei capelli e della pelle, testando i tre diversi modelli di predizione. I risultati delle previsioni fenotipiche per ciascun approccio e livello di copertura sono stati confrontati con il "*fenotipo reale*", inferito dal genoma originale ad alta copertura. È stato quindi valutato quante volte ciascun metodo, a ogni livello di copertura, ha prodotto previsioni corrette, al fine di determinare i limiti e l'approccio più idoneo per la predizione fenotipica da dati antichi, riducendo al minimo gli errori nelle stime.

Le conclusioni di questa prima parte indicano che, indipendentemente dal periodo storico del campione analizzato, è consigliabile applicare il protocollo standard di HIrisPlex-S se si raggiungono livelli minimi di copertura di almeno 8x su ciascuna delle 41 posizioni di interesse. Per coperture inferiori a 8x, si raccomanda invece l'uso del metodo ideato e testato in questo studio, che utilizza le genotype likelihoods per la stima dei genotipi.

La seconda parte di questo progetto di tesi ha previsto l'applicazione del protocollo a un dataset di 348 individui eurasiatici, che spaziano dal Paleolitico superiore all'Età del Ferro, con l'obiettivo di osservare la variazione del colore degli occhi, dei capelli e della pelle negli ultimi 45.000 anni. La fotografia che emerge rivela un'evoluzione complessa di questi tratti: sebbene i fenotipi scuri siano stati predominanti per gran parte del periodo esaminato, i fenotipi chiari sono comparsi a partire dal Mesolitico, nelle regioni che richiedevano un adattamento ambientale alle alte latitudini. In seguito, tali fenotipi chiari si sono diffusi ulteriormente nei periodi successivi grazie a una complessa interazione tra adattamento ambientale ed eventi demografici.

Table of Contents

1 INTRODUCTION 1

1.1 Ancient DNA 1

1.1.1 Sequencing of Ancient Data 3

1.1.2 Challenges in Variant Calling 6

1.2 Genetic Origins of Eurasia 11

1.3 Evolutionary History of Eurasian Pigmentation 15

1.4 Reconstructing Phenotypes from Genomic Data: Insights into Ancient Humans 22

1.5 Forensic DNA Phenotyping of Pigmentation Traits 24

1.5.1 Eye Colour 25

1.5.2 Hair Colour 27

1.5.3 Skin Colour 28

1.6 The HIrisPlex-S System 30

1.7 The HIrisPlex-S System Applied to Ancient DNA 32

2 AIM 35

3 METHODS 36

3.1 Conceptual framework 36

3.2 Genomic dataset 37

3.3 Test sample selection and preparation 39

3.3.1 *SAMtools* Mpileup file 40

3.4 Genomic downsampling 41

3.4.1 Pointwise progressive downsampling 43

3.4.2 Mean progressive downsampling 43

3.5 HIrisPlex-S standard protocol 44

3.5.1 Indel check 44

3.5.2 Variant calling 45
3.5.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes 45
3.6 HIrisPlex-S Protocol implemented with a Genotype likelihoods Model 50
3.6.1 Indel check 51
3.6.2 Variant calling 51
3.6.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes 53
3.7 Imputation workflow 54
3.7.1 Indel check 54
3.7.2 Variant calling 54
3.7.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes 55
3.8 User Manual and Dataset Application 55
4 RESULTS E DISCUSSION 57
4.1 Ust'-Ishim sample 58
4.1.1 Eye colour prediction 58
4.1.2 Hair colour prediction 63
4.1.3 Skin colour prediction 68
4.2 SF12 sample 73
4.2.1 Eye colour prediction 73
4.2.2 Hair colour prediction 78
4.2.3 Skin colour prediction 83
4.2.4 Analysis of Red and Blond Hair Phenotypic Expression 89
4.3 I5832 sample 93
4.3.1 Eye colour prediction 93
4.3.2 Hair colour prediction 98
4.3.3 Skin colour prediction 103
4.4 General considerations 108

4.5 User Manual and Dataset Application 110

4.5.1 Palaeolithic inference of phenotypic traits 112

4.5.2 Mesolithic inference of phenotypic traits 114

4.5.3 Neolithic inference of phenotypic traits 117

4.5.4 Copper Age inference of phenotypic traits 120

4.5.5 Bronze Age inference of phenotypic traits 122

4.5.6 Iron Age inference of phenotypic traits 125

4.6 General considerations 127

5 CONCLUSION 130

6 BIBLIOGRAPHY 132

1 INTRODUCTION

1.1 Ancient DNA

The study of ancient DNA (hereafter aDNA) has given rise to the emerging field of paleogenomics. This discipline involves analysing genetic material extracted from organisms preserved over extended periods, typically exceeding 100 years. By examining ancient genomes, paleogenomics enables insightful assessments of evolutionary processes, population dynamics, and genetic adaptations of past organisms, including non-human species (Brunson and Reich, 2019; Kerner *et al.*, 2023).

The origins of paleogenomics can be traced back to the 1980s. In 1984, researchers from the Department of Biochemistry at the University of California, led by Allan Wilson, conducted pioneering research on aDNA. Their work involved extracting and sequencing 229 base pairs of mitochondrial DNA fragments from the dried muscle of a museum specimen of the quagga (*Equus quagga*), a zebra-like species that went extinct in 1883 (Higuchi *et al.*, 1984). A year later, Swedish geneticist Svante Pääbo, who was awarded the Nobel Prize in Medicine in 2022, focused on amplifying short sequences from biological material found in a 2,400-year-old mummy of a child (Pääbo, 1985). However, these early studies faced significant technical limitations due to the inefficiency of bacterial cloning as an amplification method for investigating aDNA samples (Willerslev and Cooper, 2005). These studies revealed that preserved genetic material in ancient specimens was predominantly microbial or fungal, with endogenous DNA typically present in low concentrations as short, damaged fragments of multi-copy loci, such as mitochondrial DNA (mtDNA) (Figure 1.1).

This fragmentation and degradation of DNA can be attributed to its susceptibility to external environmental factors. During DNA replication, random errors may occur, which, if not corrected, become permanently incorporated into the genome. External chemical and physical agents can also induce mutations. While living cells possess numerous repair mechanisms to detect and repair these errors, these mechanisms become inactive after cell death, leading to diagenetic processes in DNA, primarily through the action of nucleases (autolytic degradation) (Lindahl, 1993). DNA degradation is accelerated in warm, humid environments with acidic or basic pH, whereas degradation processes are considerably slowed in colder regions with neutral pH and in the absence of oxygen and water (Kirsanow and Burger, 2012; Hofreiter *et al.*, 2015). Consequently, this geographical distribution significantly affects the availability of aDNA samples.

The invention of PCR (Polymerase Chain Reaction), conceptualized by Kary Mullis in 1986, revolutionized DNA analysis by simplifying and standardizing the amplification process (Pääbo and Wilson, 1988; Thomas *et al.*, 1989). However, PCR faced significant constraints due to the requirement for specific primers (20-30 bp) designed to amplify sequences typically longer (~150 bp), posing challenges for aDNA fragments. Designing these primers required prior knowledge of the sequence to be amplified, which was a limitation when dealing with ancient samples that might contain genetic variants absent in contemporary genomes. Furthermore, aDNA molecules often suffer from molecular damage, favouring the amplification of longer and less damaged modern DNA molecules (Willerslev and Cooper, 2005; Knapp and Hofreiter, 2010).

The advent of Next Generation Sequencing (hereafter NGS) has played a crucial role in recent advancements in aDNA research. NGS allows simultaneous sequencing and analysis of ultra-short DNA fragments (35-300 bp) without the constraints of primer-based methods, addressing challenges such as contamination and DNA damage. NGS has also improved the authenticity of aDNA data through comprehensive genome-wide analysis and targeted enrichment techniques, making it cost-effective to recover sequence data at various genomic scales (Orlando *et al.*, 2021).

Over the past four decades of aDNA research, improvements in extraction technologies have enabled the retrieval of genetic material from diverse external sources such as soil (Willerslev *et al.*, 2003), coprolites (Gilbert *et al.*, 2008), or resins (Jensen *et al.*, 2019). The chronological range suitable for DNA analysis has expanded significantly, now encompassing periods exceeding half a million years ago (560,000–780,000 years ago) (Orlando *et al.*, 2013). Moreover, complete genomes of numerous extinct species, including woolly mammoths (Palkopoulou *et al.*, 2015) and cave bears (Barlow *et al.*, 2018), as well as various human populations from Vikings (Margaryan *et al.*, 2020) to Paleo-Inuit (Rasmussen *et al.*, 2010) to Neanderthals (Green *et al.*, 2010; Prüfer *et al.*, 2014; Mafessoni *et al.*, 2020), have been successfully sequenced.

In addition to these achievements, ancient genomics also holds significant promise in medical research by enabling the reconstruction of human health history, elucidating human adaptation to historical epidemics (Bos *et al.*, 2014; Kerner *et al.*, 2021), and identifying genetic factors influencing the immune systems of modern human populations (Kerner *et al.*, 2023; Mathieson *et al.*, 2015). These insights provide opportunities for advancements in drug development (Kerner *et al.*, 2023).

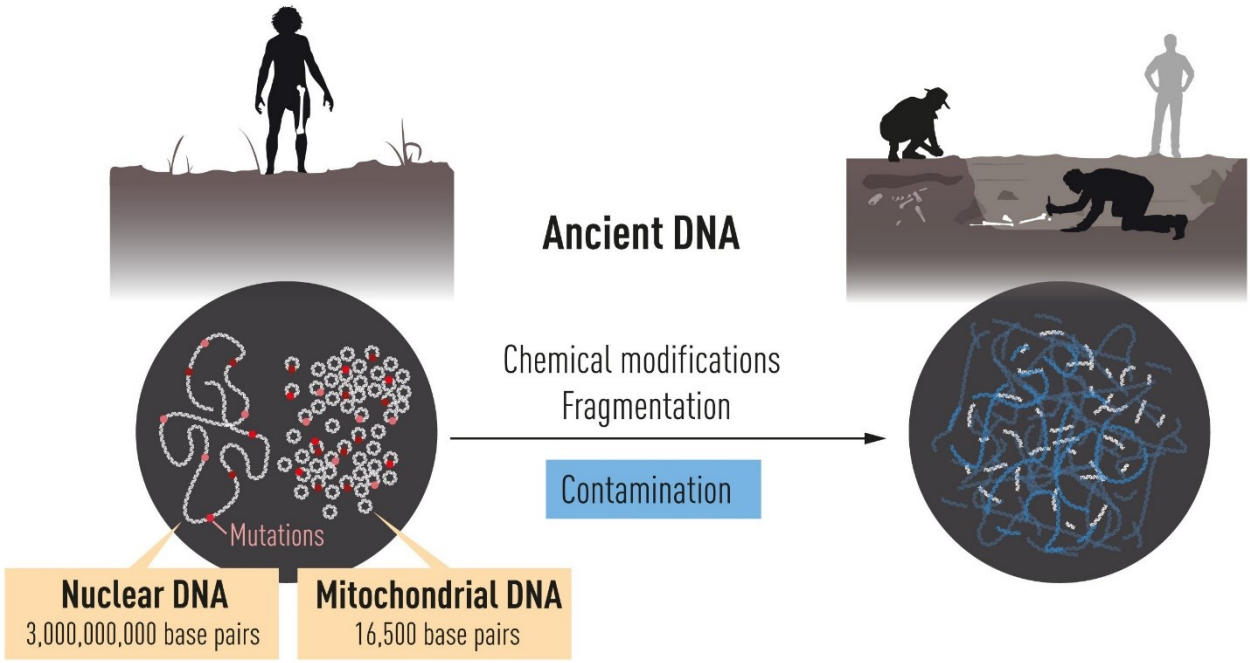


Figure 1.1 - The cellular DNA exists in two distinct compartments: nuclear DNA containing the majority of genetic information, and the smaller mitochondrial genome present in thousands of copies. Post-mortem, DNA undergoes gradual degradation, resulting in minimal remnants, often contaminated with DNA from bacteria and contemporary humans (image taken from: https://www.nobelprize.org/prizes/medicine/2022/press-release/).

1.1.1 Sequencing of Ancient Data

In the field of aDNA analysis, a wide variety of materials can be utilized to extract genetic information, with a particular emphasis on molecular preservation. These materials include bone fragments (Hagelberg *et al.*, 1989; Fu *et al.*, 2014), teeth, which also enable the retrieval of ancient blood-borne pathogens (Bos *et al.*, 2011; Margaryan *et al.*, 2018), petrous bone (Gamba *et al.*, 2014), hair (Gilbert *et al.*, 2006; Rasmussen *et al.*, 2010), and soft tissues obtained from mummified remains or artificially conserved specimens (Higuchi *et al.*, 1984; Pääbo, 1985).

Dedicated clean laboratory facilities specifically designed for aDNA research are essential to minimize the risk of contamination during DNA extraction and manipulation processes (Fulton and Shapiro, 2019). The preparation and analysis of aDNA involve several crucial steps to ensure the integrity and accuracy of the genetic material under study. Initially, DNA extraction is performed, which includes recovering DNA from the original matrix through demineralization and digestion (Sirak *et al.*, 2017; Patzold *et al.*, 2020). Techniques such as silica-based DNA isolation (Höss and Pääbo, 1993), utilizing spin columns (Damgaard *et al.*, 2015) or silica-coated magnetic beads (Glocke and Meyer, 2017) are employed. These methods are crucial for efficiently retrieving short DNA

fragments, essential for the analysis of degraded samples and help to separate aDNA from potential contaminants, thus maintaining the purity of the extracted sample.

Following extraction, DNA can be treated with the USER reagent, a commercial enzymatic mix containing uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII), aimed at reducing sequencing errors caused by post-mortem damage. This treatment eliminates damaged nucleotides but also removes patterns necessary for authenticating aDNA. In mammalian DNA, these patterns can still be identified in CpG dinucleotides (Briggs *et al.*, 2010), which are heavily methylated in vertebrate nuclear genomes and protected from USER degradation. A modified USER protocol known as UDG-half has been developed to preserve some damage markers while minimizing errors (Rohland *et al.*, 2015).

In palaeoenvironmental aDNA research, an optional step can be meta-barcoding, aimed at retrieving information about past species communities using barcode PCR, able to amplify specific sites from a diverse array of organisms, and NGS (Crump *et al.*, 2019).

Subsequently, aDNA fragments prepared for sequencing are converted into DNA library. Two primary methods are employed: double-stranded and single-stranded library preparations. In double-stranded preparation, aDNA molecules are end-repaired and ligated to double-stranded adapters (Gansauge and Meyer, 2013; Gansauge *et al.*, 2020), while single-stranded preparation involves ligating heat-denatured aDNA templates to adapters as single-stranded molecules. Single-stranded preparation is particularly advantageous for samples with low biomass, allowing incorporation of DNA with post-mortem damage and enhancing the fraction of endogenous DNA in sequencing libraries (Gansauge and Meyer, 2013; Gansauge *et al.*, 2020). During library construction, unique identifying sequences (indexes) are integrated within the adapters. These indexes facilitate the pooling of multiple samples and minimize contamination risks by ensuring that only sequences from samples with the correct adapter-index combinations are analysed (Kircher *et al.*, 2012; Rohland *et al.*, 2015; Gansauge *et al.*, 2020).

Additionally, target-enrichment methods can be utilized before sequencing to optimize efforts on specific regions of interest within aDNA libraries, given the often limited endogenous DNA content (Burbano *et al.*, 2010).

The resulting library can be amplified via PCR to increase the quantity of genetic material available for sequencing, thereby improving sensitivity and precision in genetic analyses. PCR amplification

uses specific enzymes capable of amplifying damaged templates (Gansauge and Meyer, 2013; Seguin-Orlando *et al.*, 2015).

Subsequently, aDNA libraries are subjected to sequencing using technologies such as Illumina, which utilize "sequencing by synthesis". This method involves sequencing tens-to-hundreds of millions of clusters containing small ssDNA templates simultaneously. Initially, DNA fragments are prepared with specific adapters, followed by cluster generation on a surface and sequential addition of fluorescent nucleotides (Rasmussen *et al.*, 2010). To each base call is assigned a measure of uncertainty, known as a quality score. The resulting sequencing data consists of relatively short segments, typically around 50-100 base pairs, referred to as "reads" (Fumagalli *et al.*, 2013).

These meticulous procedures in DNA extraction, library preparation, and amplification are essential for generating precise and reliable sequencing data from aDNA. These steps form the basis for subsequent computational analyses; however, before the data can be analysed, it must first undergo cleaning and processing.

Various tools and pipelines exist for the analysis of aDNA sequences, and the choice of methods depends on the specific objectives of the research project, such as paleoenvironmental reconstruction, microbial profiling, or population history modelling. Despite diverse aims, several common steps apply to most analytical workflows. These steps include processing raw sequencing data, aligning sequences against reference genomes, and evaluating the authenticity and error rates of the sequences, particularly addressing miscoding lesions caused by post-mortem damage.

The initial processing steps involve index demultiplexing to identify and assign reads to their respective libraries based on index combinations. This is followed by read trimming to remove adapter sequences and low-quality terminal regions. For paired-end reads, overlapping pairs are collapsed into consensus sequences (Renaud *et al.*, 2014; Schubert *et al.*, 2016). The performance of mapping tools can vary depending on the experimental procedures followed during sample preparation, including the type of DNA library constructed and whether the extracts underwent USER treatment (Poullet and Orlando, 2020).

Processed reads are usually aligned to a reference genome, or assembled into scaffolds in a *de novo* assembly when a reference genome is unavailable, using tools like *BWA* (Li and Durbin, 2010) or *Bowtie2* (Langmead and Salzberg, 2012) and the chosen parameters are optimized to maximise mapping specificity and sensitivity, influenced by factors such as DNA fragment size, and the extent of post-mortem damage (Green *et al.*, 2009; Schubert *et al.*, 2012). It is important to highlight that

reads are mapped against the reference genomes of the target species, such as the human genome, which contains only the reference allele being present at polymorphic sites. Short reads carrying alternate alleles are less likely to map compared to those with the reference allele, leading to a potential bias against non-reference variation. This bias can significantly impact population genetic inference (Günther and Nettelblad, 2019; Martiniano *et al.*, 2020). To address this, post-mapping read filtering procedures can be employed. Once the alignment phase is complete, any duplicates resulting from PCR are checked for, as they can lead to the overrepresentation of some reads.

Following alignment, the authentication of genomic DNA from individual samples requires the observation of post-mortem nucleotide misincorporation due to cytosine deamination (Hoss *et al.*, 1996; Jónsson *et al.*, 2013) and the estimation of contaminating DNA based on heterozygosity levels in haploid chromosomes (Krause *et al.*, 2010). Post-mortem DNA damage patterns can be quantified in non-USER-treated or partially USER-treated libraries using the tool *mapDamage2* (Jónsson *et al.*, 2013). Cytosine deamination can also be estimated with *PMDtools* (Skoglund *et al.*, 2014) on fully USER-treated data, focusing on CpG dinucleotides (Briggs *et al.*, 2010). Estimating contamination based on ploidy can be challenging, as it generally necessitates access to a polymorphism database to account for the probability that contaminating molecules may go undetected due to shared variants between the contaminant and the endogenous genome. However, various tools exist for assessing contamination levels in haploid chromosomes, such as *Schmutzi* (Renaud *et al.*, 2015) for mitochondrial DNA data and *ANGSD* (Korneliussen *et al.*, 2014) for the X chromosome in males.

1.1.2 Challenges in Variant Calling

The process of transforming base calls into genotypes is commonly split into two distinct stages: Single Nucleotide Polymorphisms (hereafter SNPs) calling and genotype calling. SNP calling seeks to identify positions where polymorphisms occur or where at least one base differs from the reference genome; the latter, termed as “variant calling”, involves assigning genotypes to individuals, typically restricted to positions where a SNP or variant has previously been identified (Nielsen *et al.*, 2011).

Considering the aforementioned challenges associated with aDNA, data typically exhibit short fragment lengths, low coverage, and a high intrinsic error rate. This limitation complicates the accurate of variants calling, thereby introducing considerable uncertainty into the results. Quantifying and addressing this uncertainty is crucial as it profoundly impacts downstream analyses reliant on inferred SNPs and genotypes.

Sequencing depth, also known as coverage, refers to the number of unique sequencing reads that align to specific regions in a reference genome, indicating how frequently each nucleotide is sequenced at those positions. For instance, a sequencing depth of 30x for a human genome means that, on average, each base in the genome is sequenced 30 times from different reads (Figure 1.2). In cases of low coverage, the chance of identifying heterozygous sites decreases if genotype calling relies solely on the presence or absence of an allele, as only one chromosome of a diploid individual might be sampled at specific sites (Nielsen *et al.*, 2012).

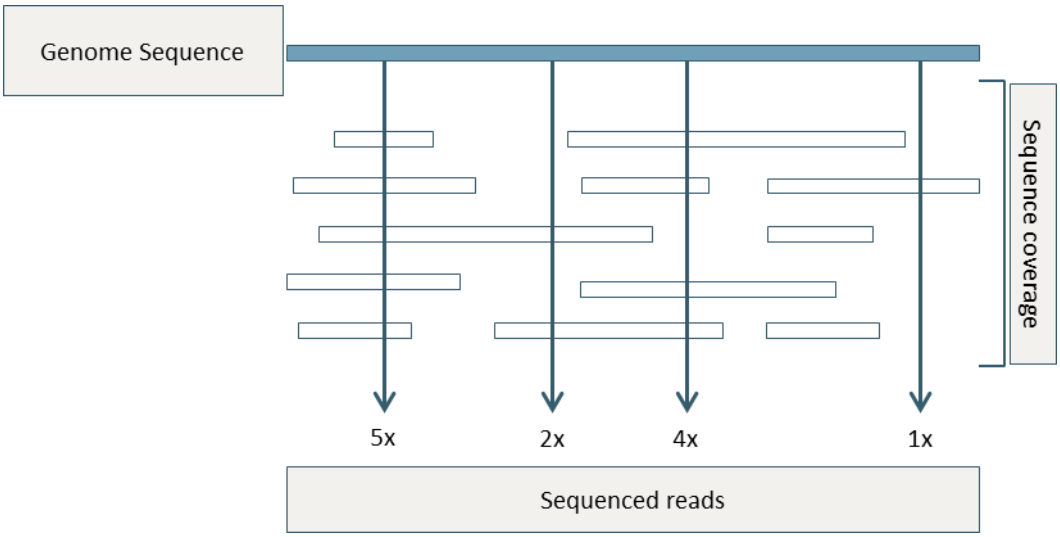


Figure 1.2 - Example of sequence coverage, ranging from 1x to 5x. Blue band stands for genome sequence, white bands stand for genomic reads.

The error rate partly arises from the inherent nature of ancient data, compounded by features of NGS platforms. This error, in the context of Illumina platform sequencing, is influenced by various factors, including mixed clusters, potential cross-contamination of indexed oligonucleotides, and the formation of chimeras during PCR (Nielsen *et al.*, 2011; Kircher *et al.*, 2012; Korneliussen *et al.*, 2014).

1.1.2.1 Genotype Likelihoods

Developing a strategy that utilizes available sequencing reads while addressing potential biases in variant calling would be advantageous. An efficient method involves employing a Bayesian framework to compute genotype posterior probabilities based on genotype priors and genotype likelihoods (Nielsen *et al.*, 2011). In essence, Bayes' theorem facilitates the inference of conclusions (*posteriors*) by integrating prior expectations (*priors*) before conducting the analysis, while also considering the likelihood of the data under a chosen model. These probabilities are conditional, representing the probability of event A given that event B has occurred.

The distribution chosen to represent knowledge about the parameter of interest (*priors*) is determined based on available information regarding the parameter before analysis. In the context of genotype posterior probabilities, prior probabilities are determined independently of information obtained from NGS data. For instance, when sequencing a single individual, the prior probability for genotypes may be uniformly distributed or informed by external sources such as the reference sequence or SNP databases. When dealing with multiple samples, prior probabilities can be computed by considering allele frequencies or genotype frequencies. If allele frequencies are known, genotype probabilities can be calculated using assumptions like the Hardy-Weinberg equilibrium (HWE) or other models relating allele frequencies to genotype frequencies. Uncertainty in allele frequency estimation can be addressed by assigning a prior to the allele frequency itself, rather than estimating it directly. This prior can be derived from observed data or established population genetic theory (Nielsen *et al.*, 2011).

The probability used to indicate the likelihood of observing data within a specific model can be selected based on available data. One example is the method utilized in the original *GATK* (McKenna *et al.*, 2010), and also implemented in the software *ANGSD* (Korneliussen *et al.*, 2014).

Given a genotype *G* of a biallelic locus consisting of the A_1 and A_2 alleles, for a position *i* covered by *M* reads, the genotype likelihoods are defined as the probability of observing base *b* corresponding to the A_1 and A_2 alleles, respectively:

$$Pr(D|G = \{A_1, A_2\}) = \prod_{i=1}^{M} Pr(b_i|G = \{A_1, A_2\}) = \prod_{i=1}^{M} \left(\frac{1}{2}Pr(b_i|A_1) + \frac{1}{2}Pr(b_i|A_2)\right)$$

The probability of the base *b* to be equal to an *A* allele is calculated using the probability of sequencing error *e* of base *b* obtained from the Phred-scaled base quality score *q* by the sequencer $e = 10^{-q/10}$.

$$Pr(b|A) = \begin{cases} \frac{e}{3} : b \neq A \\ 1 - e : b = A \end{cases}$$

This assumes independence between the reads.

This method prioritizes the likelihood of observing a specific set of sequencing reads rather than making absolute genotype calls at each position. As a result, it produces detailed measurements that account for their uncertainties, offering results that can be interpreted statistically with clarity (Nielsen *et al.*, 2011).

1.1.2.2 Genotype Imputation

When direct genotype determination encounters obstacles such as low coverage or missing data, genotype imputation becomes essential. This technique facilitates genotype inference through statistical methods that rely on the knowledge of haplotype distributions within the population (Cassidy *et al.*, 2020; Sousa da Mota *et al.*, 2023).

Genotype imputation is widely employed in contemporary data studies to reduce sequencing costs. Moreover, it has proven effective in aDNA research, achieving robust concordance (>99%) between imputed genotypes and high-quality counterparts in human ancient genomes, even at low coverage levels such as 0.5x (Erven *et al.*, 2022; Sousa da Mota *et al.*, 2023).

The process of genotype imputation leverages comprehensive genetic datasets known as reference panels (Li *et al.*, 2009). These panels are derived from genotypes collected across large sample sizes and undergo a phasing process, wherein alleles are assigned to chromosomes to construct ordered haplotype sequences. Once established, these reference haplotype panels enable the phasing of genotypic data from target samples, facilitating comparison and inference (Naj, 2019).

Inference of missing data relies on models of linkage disequilibrium (hereafter LD) and haplotype sharing, notably the Li and Stephens model (Li & Stephens, 2003). This model calculates the likelihood of encountering a chromosome (or haplotype) based on previously sampled haplotypes from a population, treating the new haplotype as a replication of segments from the sampled haplotypes while allowing for mutations. The extension of LD provides a measure of haplotype diversity within the same panel and allows for the detection of SNPs in high LD with other groups of SNPs present in a genomic region, thereby contributing to haplotype capture (Li and Stephens, 2003; Sousa da Mota *et al.*, 2023; Torkamaneh and Belzile, 2021). These data are compared with the samples under study to find the best possible match. Once the best match between the sample and the reference panel is determined, the absent variants in the sequencing data of the sample are inferred, and their corresponding probability of matching is assigned.

The choice of reference panel is critical and ideally should consider populations with similar LD patterns, haplotype distributions, and demographic histories of the study samples. However, challenges arise from the presence of rare variants and differing population dynamics, including migration and admixture (Wang and Chatterjee, 2017; Naj, 2019; Torkamaneh and Belzile, 2021; Deng *et al.*, 2022). Therefore, careful consideration is necessary when selecting the panel to ensure alignment with the population structure of the samples under study. Additionally, a thorough assessment of the inferred

data is required, considering the limitations of imputation, such as the inference of incorrect genotypes (Torkamaneh and Belzile, 2021; Orlando *et al.*, 2021).

aDNA imputation introduces extra challenges that may affect the accurate capture of ancient individuals' ancestries using present-day reference panels. In 2020, Hui and colleagues released a two-step imputation pipeline for ancient genomes (Hui *et al.*, 2020). Initially, genotype likelihoods were imputed using *Beagle* v4.1 (Browning and Browning, 2016), followed by site removal based on maximum genotype probability (hereafter GP). Subsequent imputation with *Beagle* v5 (Browning *et al.*, 2018), coupled with final GP filtering, increased the proportion of heterozygous sites meeting GP thresholds compared to single-step methods. Validation using a single downsampled ancient European genome supported these findings. In 2022, a group led by Iain Mathieson compared this pipeline with *GLIMPSE* (Rubinacci *et al.*, 2021), finding higher precision with *GLIMPSE* but greater recall with the Hui *et al.* pipeline (Cox *et al.*, 2022). Another study published in 2022 by Ausmees and colleagues evaluated imputation performance across various coverage levels and reference panel configurations, suggesting improved accuracy with broader population inclusion in the 1000 Genomes reference panel and simultaneous imputation of ancient genomes (Ausmees *et al.*, 2022).

In 2023, a study entitled "*Imputation of ancient human genomes*", published in Nature Communications by Bárbara Sousa da Mota, addressed the limitations of previous studies, such as limited sample sizes of ancient genomes and their predominantly European descent (Sousa da Mota *et al.*, 2023). In this study, an ancient trio (comprising a mother, father, and son) was re-sequenced, and 43 ancient genomes were downsampled and imputed, including 42 high-coverage genomes (exceeding 10x coverage). The accuracy of imputation was evaluated across various factors such as ancestry, time, coverage depth, and sequencing technology, revealing that error rates are influenced by the ancestry of the ancient genomes. The representation of a given population in the reference panel significantly affects imputation accuracy, with genotyping errors at alternative allele sites ranging from above 5% to 25% among African genomes at 1x coverage. Additionally, it was found that the age of samples can negatively affect the imputation accuracy of rare variants.

These findings highlight significant implications for the imputation of ancient genomes, particularly when a direct descendant population without admixture is not available in the present day. The results indicate that using admixed reference populations with recent shared ancestry with ancient genomes can achieve accurate imputation, even for rare variants. Furthermore, common variants

can be accurately imputed in very old non-African genomes or those poorly represented in the reference panel, likely due to their shared historical background.

1.2 Genetic Origins of Eurasia

The time period spanning approximately from 50,000 to 5,000 years ago witnessed significant demographic changes among human populations in Eurasia. This era marked the expansion of anatomically modern humans throughout Eurasia, alongside the disappearance of archaic human groups that had previously inhabited these regions up until 40,000 years ago (Skoglund and Mathieson, 2018).

Initially, early Eurasian populations of the Upper Palaeolithic were distinct and contributed minimally to later Eurasian genetic diversity (Fu *et al.*, 2014; Fu *et al.*, 2015).

Between 35,000 and 15,000 years ago, genomic analysis indicates that ancient *Homo sapiens* in western and eastern Eurasia began to resemble present-day Europeans or Asians (Seguin-Orlando *et al.*, 2014; Allentoft, Sikora, Refoyo-Martínez, *et al.*, 2024). Following the Last Glacial Maximum around 20,000 years ago, Eurasia saw a repopulation likely originating from refuges in southern and central Europe, resulting in a genetic amalgamation of European and Asian ancestries in northern Eurasia (Fu *et al.*, 2016; Yang and Fu, 2018).

Subsequent to the Palaeolithic era, early recolonizations were replaced by populations contributing to the formation of European Mesolithic hunter-gatherer populations (hereafter HGs). Their genetic legacy in modern European populations dates back approximately to 37,000 years (Seguin-Orlando *et al.*, 2014), yet their overall genetic impact was limited due to reduced variability and the influx of new populations (Marchi *et al.*, 2022).

Initially viewed as a homogeneous group, HGs exhibit genetic diversity reflective of their regional distribution: West hunter-gatherers (hereafter WHGs), Scandinavian hunter-gatherers (hereafter SHGs), East hunter-gatherers (hereafter EHGs), and Caucasian hunter-gatherers (hereafter CHGs) (Skoglund, Malmström, *et al.*, 2014; Lazaridis *et al.*, 2014; Haak *et al.*, 2015; Jones *et al.*, 2015; Allentoft, Sikora, Refoyo-Martínez, *et al.*, 2024) (Figure 1.3).

WHGs ancestry is believed to originate directly from groups associated with the Epigravettian, Azilian, and Epipalaeolithic cultures, specifically those within the Villabruna cluster (Fu *et al.*, 2016). In contrast, EHGs ancestry exhibits additional admixture with a source from the Upper Paleolithic

Siberian region, known as Ancient North Eurasian (hereafter ANE) (Raghavan *et al.*, 2014). The composition of WHGs ancestry was regionally diverse among Mesolithic populations. Evidence indicates ongoing local admixture among Iberian HGs (Villalba-Mouco *et al.*, 2019), in contrast to the more homogenous WHGs ancestry observed in Britain and northwestern continental Europe (Brace *et al.*, 2019). The admixture event that gave rise to EHGs ancestry is estimated to have occurred between 13,000 and 15,000 years ago, with a composition that aligns broadly with geographical gradients. Baltic and Ukrainian HGs demonstrate a greater affinity to the Villabruna Upper Paleolithic cluster compared to Russian HGs, who show a stronger ANE influence (Mathieson *et al.*, 2018; de Barros Damgaard *et al.*, 2018; Saag *et al.*, 2021; Posth *et al.*, 2023). SHGs exposes a varied mix of WHGs and EHGs ancestries among the later Mesolithic populations, revealing an east-west genetic gradient, distinct from patterns observed in other parts of Mesolithic Europe (Günther *et al.*, 2018). CHGs diverged from WHGs populations around 45,000 years ago and subsequently contributing to the ancestry of nomadic pastoralist populations of the Pontic-Caspian steppes. Another branch emerged from the ancestral population of CHGs, becoming the ancestors of the first Neolithic farmers in the Levant (Jones *et al.*, 2015).

Two major cultural transitions significantly reshaped the genetic landscape of Europe: the introduction of agriculture during the Neolithic period and the Bronze Age's arrival of the Yamnaya culture, marking distinct contributions to the genetic makeup of modern European populations (Günther and Jakobsson, 2016) (Figure 1.3).

The Neolithic revolution in Europe marked a shift from nomadic hunter-gatherer lifestyles to settled agricultural communities. This transition was primarily driven by population movements rather than mere cultural diffusion, as migrating farmers from the Middle East introduced agricultural practices (Ammerman and Cavalli-Sforza, 1984).

Unlike HGs, Neolithic farmers were genetically closer to populations in southwestern Europe than those in the Near East (Skoglund *et al.*, 2012), with genetic evidence tracing their origins back to Anatolia and the Levant approximately 12,000-11,500 years ago (Mathieson *et al.*, 2015; Omrak *et al.*, 2016; Lazaridis *et al.*, 2016; Kılınç *et al.*, 2016).

The spread of agriculture in Europe occurred through multiple routes, including the Danube and Mediterranean coast, resulting in diverse genetic and cultural impacts across the continent (Özdoğan, 2011; González-Fortes *et al.*, 2017)

Being more isolated from mainland Europe, Britain experienced a delayed adoption of agricultural and husbandry techniques, occurring around 6,000 years ago, approximately a millennium after these practices were established on the continent. This introduction came through the migration of peoples descended from Neolithic Iberian farmers who reached Britain via the Mediterranean route (Brace *et al.*, 2019). A similar delay in transitioning to the Neolithic lifestyle is observed among Danish Mesolithic individuals, lagging by over a millennium compared to Central Europe. This transition was rapid and led to a significant turnover in population dynamics, with minimal genetic input from local HGs (Allentoft, Sikora, Fischer, *et al.*, 2024). The impact of ancient Neolithic farmers was most pronounced in Iberia and Italy, whereas central, eastern, and Scandinavian Europe experienced distinct influences (Allentoft, Sikora, Refoyo-Martínez, *et al.*, 2024). Although initial gene flow between farmer populations and indigenous HGs was limited or absent (Jones *et al.*, 2017), some regions, particularly in the late Neolithic, saw the "resurgence" of HGs ancestry in genomes, possibly due to reduced gene flow from Anatolia (Figure 1.3). The interaction between these populations, especially in regions like France, where migration routes converged, was complex, resulting in a mosaic of cultures. Genetic studies in this area highlight a higher prevalence of local HGs ancestry, contrasting with more diluted traces found in central-eastern European regions (Rivollat *et al.*, 2020).

HGs populations east of the genetic barrier stretching from the Black Sea to the Baltic maintained genetic continuity until about 5,000 years ago, while western regions underwent nearly complete replacement of local HGs populations following the arrival of Neolithic farmers (Allentoft, Sikora, Refoyo-Martínez, *et al.*, 2024).

By the end of the Neolithic period, the genetic composition of Europeans reflected a combination of Anatolian farming community profiles and, to a lesser extent, HGs populations. The arrival of Mesolithic HGs and Neolithic farmer-herders was not the only significant episode of population movement that shaped the genetic pool of Europeans.

During the Bronze Age, a third ancestral component arrived from the steppes of present-day Ukraine, further diversifying the genetic profile of European populations throughout history. This component had a more substantial impact in northern Europe than in the southern regions of the continent (Haak *et al.*, 2015). Archaeological findings indicate that the semi-nomadic pastoralists of the Yamnaya culture migrated from the Pontic-Caspian steppes around 6,500 years ago, contributing to the emergence of the Corded Ware culture (Allentoft *et al.*, 2015; Haak *et al.*, 2015; Narasimhan *et al.*, 2019). Besides their westward expansion, evidence suggests eastward migration as well, as

seen in discoveries from the Altai and Sayan Mountains that align with the Afanasievo culture. The genetic profile of the Afanasievo population closely resembles that of the Yamnaya culture (Allentoft *et al.*, 2015). The spread of these groups into Europe was likely facilitated by advancements such as two-wheeled chariots and the domestication of horses.

Upon entering Europe, they introduced a new ancestral genetic component known as the Steppe component, which significantly shaped the genetic landscape of modern Europeans towards the end of the Bronze Age (Allentoft *et al.*, 2015; Haak *et al.*, 2015). Individuals associated with the Yamnaya culture carried structured genetic components linked to CHGs and EHGs in their genomes (Jones *et al.*, 2015; de Barros Damgaard *et al.*, 2018) (Figure 1.3). This Steppe genetic component in Europe is found alongside migrations from the steppes. EHGs show genetic affinities with Mal'ta (MA1) (Figure 1.3), a Siberian child from 24,000 years ago who represents ANE, contributing to the genetic makeup of present-day Europeans and Native Americans, though less so to Asian populations (Raghavan *et al.*, 2014; Skoglund *et al.*, 2014; Haak *et al.*, 2015).

Simultaneously, with the spread of the Corded Ware culture in north-central and northeastern Europe around 5,750 years year ago, the Bell Beaker complex expanded from the Iberian Peninsula in western Europe. The dissemination of this culture across the continent involved both cultural diffusion and genetic exchanges among neighbouring populations. In Iberia, individuals associated with the Bell Beaker culture lacked Steppe ancestry, contrasting with Bell Beaker populations in central Europe where Steppe ancestry was prevalent, suggesting primarily cultural diffusion. Conversely, in Great Britain, the arrival of the Bell Beaker culture introduced significant Steppe ancestry previously absent, leading to substantial genetic replacement, affecting approximately 90% of the population's genetic makeup (Olalde *et al.*, 2018).

Modern European populations have thus received genetic contributions primarily from 3 distinct ancestral populations, each contributing to varying degrees depending on the geographical region analysed (Lazaridis *et al.*, 2014).

From the Iron Age onward, the genetic makeup of European populations began to resemble that of contemporary populations in the same regions today (Günther and Jakobsson, 2016; Moots *et al.*, 2023; Antonio *et al.*, 2024). These populations exhibited genetic structures distinct from those of Mesolithic HGs or Neolithic farmers (Günther and Jakobsson, 2016).

Indeed, from around 3,000 years ago, genomic analysis reveals high genetic diversity among continental populations attributed to increased mobility (Olalde *et al.*, 2019; Antonio *et al.*, 2019).

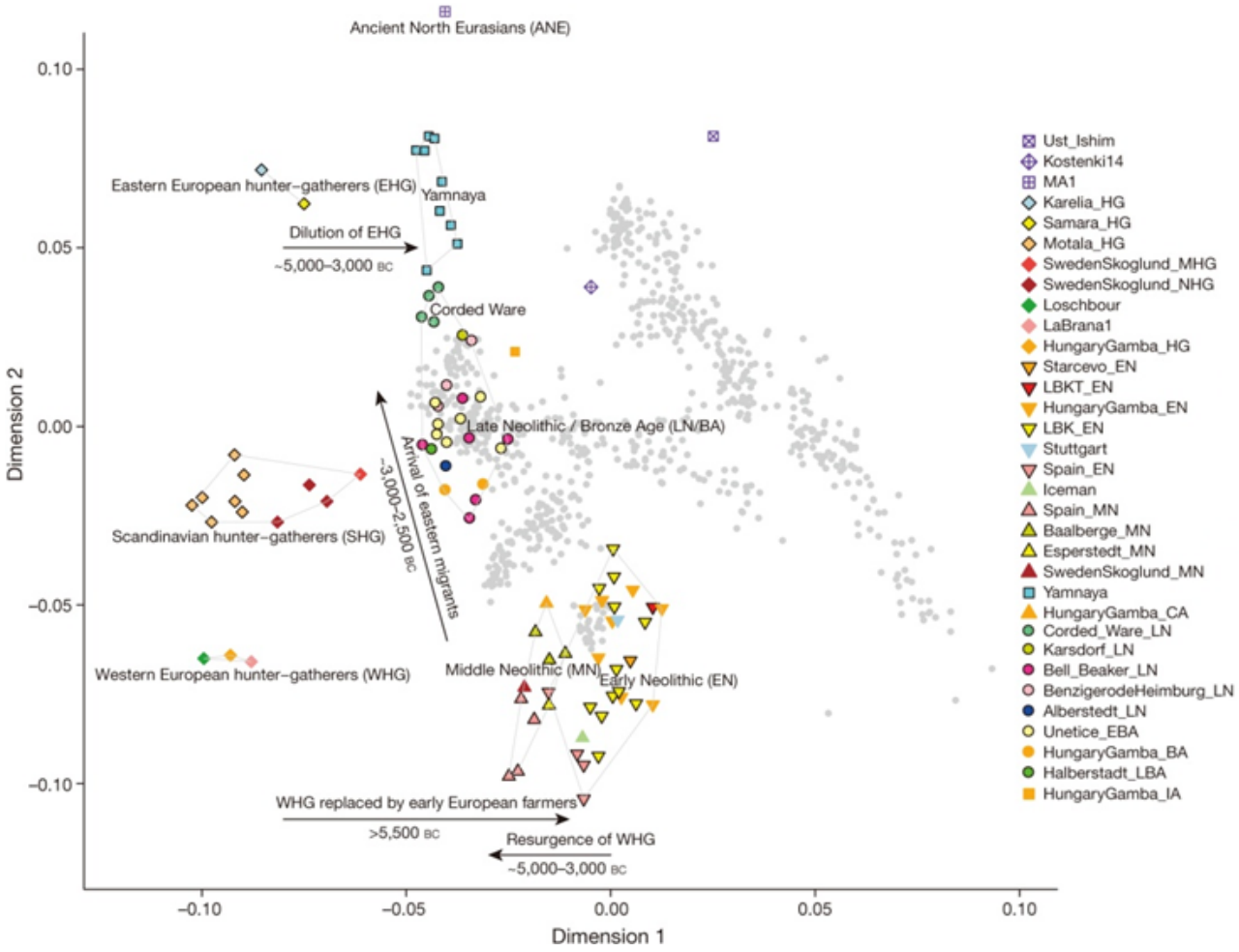


Figure 1.3 - Principal components analysis (PCA) of all ancient individuals, along with present-day West Eurasians, replicates the positioning of present-day Europeans between the Near East and HGs and the clustering of early farmers from across Europe. EHG from Russia are closer to MA1 and ANE and form a cline with WHG from Luxembourg, Spain, and Hungary. SHG are intermediate. Middle Neolithic Europeans from Germany, Spain, Hungary, and Sweden are intermediate between the earlier farmers and WHG, suggesting an increase in WHG ancestry throughout much of Europe. By contrast, in Russia, the later Yamnaya steppe herders plot between the EHG and the present-day Near East/Caucasus, indicating a decrease in EHG ancestry during the same period. Late Neolithic and Bronze Age samples from Germany and Hungary are distinct from the preceding Middle Neolithic samples, plotting between them and the Yamnaya (image taken from Haak et al., 2015).

1.3 Evolutionary History of Eurasian Pigmentation

Human coloration has evolved distinctively from that of most non-human primates due to different evolutionary trajectories and adaptive pressures. Non-human primates typically have coats consisting of brown or grey agouti hairs, though there is significant variation. They commonly exhibit darker dorsal and lighter ventral pelage on their trunk and limbs, likely serving purposes such as concealment and body temperature regulation (Walsberg, 1988; Caro, 2005; Bradley and Mundy, 2008; Caro, 2013). Certain primate lineages, particularly forest-dwelling diurnal species, display notable

coat colour patterns (Kingdon, 1980; Chaplin and Jablonski, 1998; Caro, 2013). In contrast, nocturnal non-human primates often have cryptic body coloration with distinctive mask-like patterns aiding in species and mate recognition under low-light conditions (Caro, 2013).

Humans, however, lack extensive fur coverage and exhibit minimal regional hair coloration patterns except for facial hair in some populations (Held, 2010). Most human skin is sparsely covered by nearly invisible vellus hairs, leading to the characterization of humans as "functionally naked". The remaining terminal hairs on the scalp and different types found in axillary and pubic regions reflect residual primate hair (Jablonski and Chaplin, 2017).

Evolutionary perspectives on human hair loss explore comparative anatomy, physiology, paleoecology, and climatology, despite challenges in hair and skin preservation in the fossil record (Wheeler, 1985; Zihlman and Cohn, 1988; Bramble and Lieberman, 2004). This evidence supports natural selection favouring enhanced thermoregulation during physical activity in warm environments, coinciding with reductions in body hair and the development of dense eccrine sweat glands facilitating heat dissipation (Jablonski and Chaplin, 2000; Lieberman, 2014). However, reduced body hair in humans has drawbacks such as decreased protection against abrasion and ultraviolet radiation (hereafter UVR) (Branda and Eaton, 1978).

Compensatory adaptations in Hominin skin evolved rapidly, evidenced by genomic comparisons with chimpanzees showing accelerated evolution in keratinization and epidermal differentiation genes (The Chimpanzee Sequencing and Analysis Consortium, 2005; Toulza *et al.*, 2007; Gautam *et al.*, 2015). Simultaneously, constitutive pigmentation changes occurred with the emergence of permanently darkly pigmented skin, inferred from anatomical and physiological evidence (Walter, 1971). Genomic studies of the human melanocortin 1 receptor (*MC1R*) locus indicate early evolution of dark eumelanin-rich skin pigmentation in *Homo* history, around 1,2 million years ago or earlier, correlating with functional hairlessness and increased eccrine sweat gland density (Rana *et al.*, 1999; Rogers *et al.*, 2004). This adaptive evolution towards sun-protective *MC1R* alleles persisted through purifying selection in early *Homo sapiens* within sub-Saharan Africa, crucial for survival in regions with high UVR (Harding *et al.*, 2000; Hudjashov *et al.*, 2013). In addition, *MFSD12* and *DDB1* potential genetic loci, may have been present in early *Homo sapiens*, given their prevalence at high frequencies in modern African populations and in some non-European populations today (Crawford *et al.*, 2017; Feng *et al.*, 2021) (Figure 1.4).

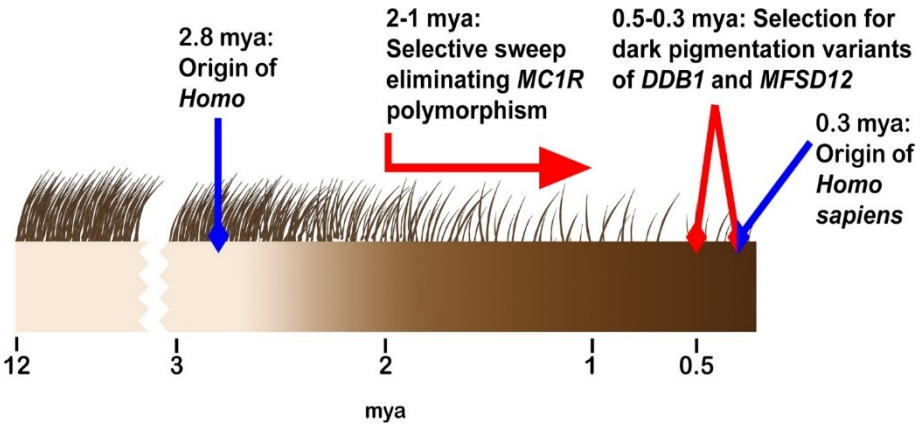


Figure 1.4 - The "hairy timeline of human evolution", schematically depicting the development of body hair and skin colour in pre-Homo sapiens members of the human lineage, indicates that prior to the emergence of early Homo species around 2.8 million years ago, hominin skin was lightly pigmented and covered with dark hair. The transition to predominantly hairless bodies starting approximately 2 million years ago marked a selective process that phased out the MC1R polymorphism. Subsequently, selection for dark pigmentation variants of DDB1 and MFSD12 genes continued through the emergence of H. sapiens around 0.3 million years ago (image taken from Jablonski, 2021).

Nevertheless, *Homo sapiens* has globally expanded over the past 100,000 years, with dispersal patterns influenced by geographic, demographic, technological, and cultural factors. Variation in UVR intensity across time of day, season, latitude, and altitude affects individual UVR exposure, body coverage, and outdoor activities, likely shaping localized adaptations of pigmentation phenotypes (Jablonski, 2021).

Hair and eye coloration in humans are primarily determined by the quantity and type of melanin produced and stored in melanosomes (Edwards *et al.*, 2016; Norton *et al.*, 2016). Unlike skin coloration, there is currently no consistent geographical pattern observed in hair and eye coloration based on the intensity and seasonality of UVR (Mackey *et al.*, 2011). Indigenous populations outside of Europe, the Middle East, Central Asia, and South Asia typically exhibit limited variation in hair coloration (Westgate *et al.*, 2013; Edwards *et al.*, 2016; Norton *et al.*, 2016), with scalp hair predominantly very dark brown in non-European populations (Shriver and Parra, 2000). A relatively high prevalence of blond hair is noted in specific geographic areas, such as Northern Island Melanesia (Kenny *et al.*, 2012; Norton *et al.*, 2014) and Europe, where red hair also occurs (Liu *et al.*, 2013). Iris coloration is mainly brown but varies significantly in shade (Edwards *et al.*, 2016), with blue, green, and hazel eyes prevalent in Europe and some regions of the Middle East, Central Asia, and South Asia, where Europeans typically display the lightest eye colours (Norton *et al.*, 2016).

Hair and eye coloration appear to have undergone weaker natural selection, with genetic bottleneck events potentially influencing observed phenotypic patterns (Lazaridis *et al.*, 2014).

The variation in human skin coloration concerning geography and levels of UVR has been extensively studied. Populations residing nearer to the Equator, where UVR is intense, tend to have darker skin pigmentation, while those in regions farther from the Equator, with lower UVR levels, typically exhibit lighter skin pigmentation (Figure 1.5). This geographical stratification in skin pigmentation is believed to stem from the evolutionary and adaptive history that has shaped human populations at various latitudes and altitudes (Sulem *et al.*, 2007).

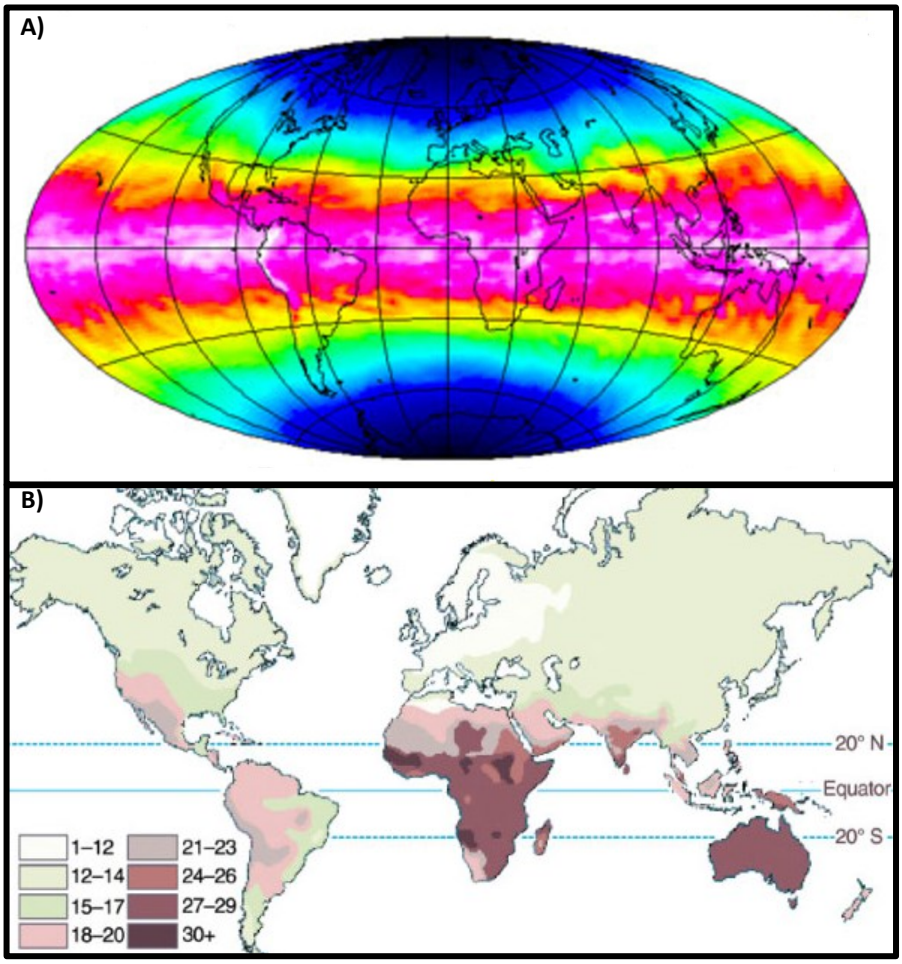


Figure 1.5 - The two maps illustrate how, with increased UVR exposure, skin tends to exhibit darker tones. A) Global clear-sky UV exposure map depicting varying levels of UV incidence from minimal (blue) to maximum (magenta) (image taken from ESA. Credit: KNMI/ESA). B) Global map illustrating distribution of skin pigmentation levels, derived from the research of Italian geographer R. Biasutti. Darker skin tones are represented by higher numerical values (image taken from Parra et al., 2007).

The prevailing hypothesis regarding the evolution of dark skin coloration in early *Homo* species suggests its role as a natural sunscreen, protecting against the photodegradation of cutaneous and systemic folate under high UVR conditions (Branda and Eaton, 1978). Folate is essential at the cellular level for DNA synthesis, the cysteine cycle, and regulation of homocysteine. Ultraviolet radiation can degrade folate, posing significant health risks such as neural tube defects in newborns and cardiovascular diseases in adults (Bower and Stanley, 1989; Branda and Blickensderfer, 1993; Fleming and

Copp, 1998; Lucock *et al.*, 2001). Therefore, the evolution of eumelanin-rich pigmentation served to mitigate fertility issues caused by folate depletion, ensuring reproductive success and early survival. Additionally, dark skin coloration plays a crucial role in protecting epidermal DNA (*DDB1* gene) against strand breaks (Jablonski, 2021).

As Hominins migrated out of equatorial Africa, they encountered lower and variable UVR levels. Darkly pigmented skin, which contains eumelanin acting as an effective sunscreen, needed to adapt to facilitate the cutaneous biosynthesis of vitamin D due to reduced UVB penetration (Clemens *et al.*, 1982; Webb *et al.*, 1988; Chen *et al.*, 2007). Vitamin D production is crucial for bone health and immune system function, making it essential for populations occupying non-tropical latitudes over the long term. This necessity led to the evolutionary reduction in eumelanin pigmentation, resulting in lighter skin, which allows for greater penetration of sunlight and increased vitamin D production (Murray, 1934; Loomis, 1967), while still maintaining folate stability.

The delicate balance between preserving folate and synthesizing vitamin D underscores the adaptive significance of skin colour intensity in human populations (Rocha, 2020; Jablonski, 2021).

The evolutionary development of depigmented skin in high-latitude Eurasian populations involved multiple genetic factors influencing both skin pigmentation and vitamin D metabolism. This adaptation was not uniform across Eurasia but was part of a broader genetic and cultural complex aimed at optimizing vitamin D availability through diet and skin synthesis (Norton *et al.*, 2007; Martínez-Cadenas *et al.*, 2013; Yang *et al.*, 2016; Yang *et al.*, 2018; Del Bino *et al.*, 2018).

In the extreme northwestern regions of Europe, such as northernmost Britain, geographic isolation during the Pleistocene and early Holocene periods favoured maximal depigmentation of the skin. This adaptation was facilitated by positive selection for genetic variants like *TYRP1, SLC24A5*, and *SLC45A2*, at 11,000-19,000 years ago respectively (Lamason *et al.*, 2005; Norton *et al.*, 2007; Sturm and Duffy, 2012). Extreme depigmentation in northwestern Europeans involved multiple skin pigmentation genes, including *KITLG*, inherited from the common ancestor of western European and east Asian populations (Hanel and Carlberg, 2020; Lao *et al.*, 2007; Sulem *et al.*, 2007), and likely accelerated by the advent of agriculture (Brace *et al.*, 2019). Alongside genetic factors, adaptations related to genes affecting vitamin D production and metabolism also played a crucial role in this biocultural compromise, sometimes counterbalancing extreme depigmentation in high-latitude Eurasian populations (Hanel and Carlberg, 2020). A third component of this adaptation involves a diet rich in vitamin D, typically centred around oily fish and, in certain locations, including marine

mammals or wild/domestic reindeer (Ross *et al.*, 2006; Chaplin and Jablonski, 2013; de Barros Damgaard *et al.*, 2018) (Figure 1.6).

The evolution of skin lightening in northeastern Asia occurred under conditions of generally low and seasonal UVB radiation, similar to those in northwestern Europe (Jablonski and Chaplin, 2000; Chaplin and Jablonski, 2013). However, distinct genetic pathways influenced depigmentation in northern Europeans and East Asians. Research from 2018 highlights that selective pressure on the upstream region of *KITLG*, a gene shared by all Eurasians' common ancestor, was more pronounced in East Asians (Yang *et al.*, 2018). This region likely conferred depigmentation and enhanced cold resistance, complemented by selection on variants of *OCA2* (Yang *et al.*, 2016) and *MFSD12* (Adhikari *et al.*, 2019) that are absent in Europeans. These adaptations facilitated comparable levels of depigmentation beneficial for cutaneous vitamin D production.

Thus, depigmentation in ancient western Europeans and East Asians exemplifies convergent evolution (Norton *et al.*, 2007; Yang *et al.*, 2016), where natural selection acted on different genes and variants in response to similar environmental pressures, achieving comparable physiological adaptations crucial for health and reproductive success (Jablonski and Chaplin, 2000; Chaplin and Jablonski, 2013; Yang *et al.*, 2018) (Figure 1.6).

East Asians and western Europeans share visibly similar light skin, but their responses to UVR are different. The skin of East Asians has a higher density of melanosomes and produces more eumelanin and pheomelanin in response to UVR exposure than the skin of western Europeans (Hennessy *et al.*, 2005). Consequently, most East Asians can tan, whereas many Europeans, particularly northern Europeans, can tan only slightly or not at all. The moderate tanning abilities observed among some Scandinavians and northern Europeans are likely due to recent genetic admixture across northern Eurasia (Del Bino *et al.*, 2018).

Following the end of the Pleistocene, improved climatic conditions led to the development of extensive steppe grasslands across much of hinterland Eurasia (Allentoft *et al.*, 2015; de Barros Damgaard *et al.*, 2018). This environment facilitated rapid human population growth associated with animal husbandry and relatively rapid, bidirectional east-west movements of people. The rise and expansion of agriculture from Anatolia and Iran led to the incursion of agriculturalists into southern and central Europe, gradually displacing earlier HGs populations (Skoglund and Mathieson, 2018). The evolution of skin pigmentation genes and phenotypes in the region reflects these significant movements and the influence of relatively few major-effect genes transmitted over long distances through

admixture (Brace *et al.*, 2019; Ju and Mathieson, 2021). Among these, the classic depigmentation variant of *SLC24A5* has the greatest impact, with evidence showing it reached western Europe via admixture with Anatolian agriculturalists, followed by continued positive selection (Ju and Mathieson, 2021) (Figure 1.6).

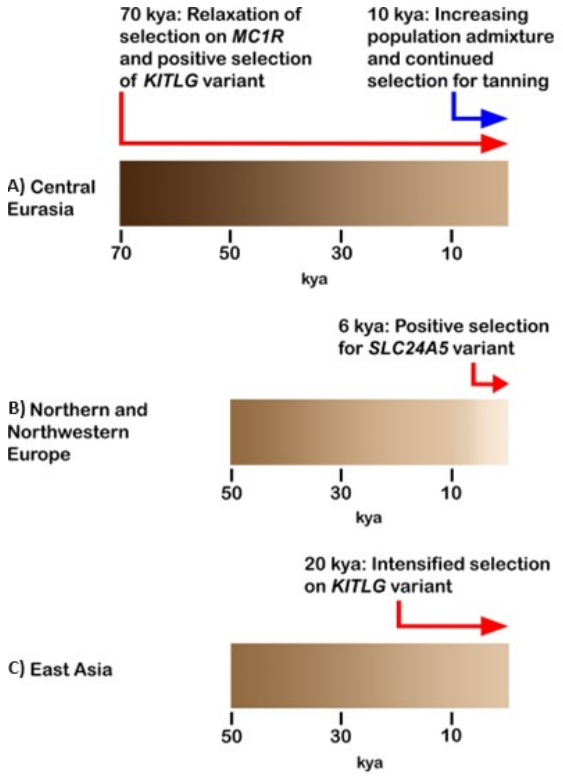


Figure 1.6 - Generalized timeline of key events in the evolution of skin pigmentation across hinterland Eurasia: A) Around 70,000 years ago in central Eurasia, there was a selection on MC1R and a KITLG variant promoting lighter skin. By 10,000 years ago, the advent of agriculture and animal husbandry led to increased population admixture and diverse gene combinations favouring moderately pigmented, tannable skin. B) About 6,000 years ago in northern and northwestern Europe, the SLC24A5 variant underwent strong positive selection following its introduction by Anatolian agriculturalists. C) In East Asia, intensified selection pressures on the KITLG variant and other genetic loci contributed to further depigmentation over time (image taken from Jablonski, 2021).

In summary, human skin, hair, and eye coloration varies widely, influenced by different evolutionary forces. Hair and eye coloration have been more affected by genetic bottlenecks and sexual selection. In contrast, the evolution of skin pigmentation in Eurasian populations reflects a complex interplay of genetic adaptation to environmental factors, including UVR exposure and dietary practices, over millennia. This evolutionary journey has shaped regional diversity in skin pigmentation, highlighting the multifaceted nature of human adaptation to diverse ecological niches.

1.4 Reconstructing Phenotypes from Genomic Data: Insights into Ancient Humans

The exploration of the human genome has posed significant challenges in the field of biology since Watson and Crick revealed the double helix structure. Initially, researchers concentrated on decoding the connections among DNA, proteins and molecular mechanism (Raben *et al.*, 2022). Central to understanding the inheritance of traits was Wilhelm Johannsen's distinction between genotype and phenotype, where the genotype refers to the inherited genetic composition of an organism and the phenotype represents the observable characteristics resulting from the interaction of the genotype with the environment (Johannsen, 1911). Building on this foundational understanding, a crucial question in genetics has been the extent to which genetic variation influences phenotypic diversity. In the early 20th century, a vigorous debate occurred between Mendelians, who emphasized discrete monogenic traits inspired by Mendel's work, and biometricians, who investigated continuous traits such as height. Biometricians contended that Mendelian genetics could not account for the continuous variation observed in numerous traits. This controversy was addressed by R.A. Fisher in a paper in 1919 (Fisher, 1919), which elucidated that if multiple genes impact a trait, the random assortment of alleles produces a continuous, normally distributed phenotype within the population. Fisher illustrated that trait values and their (co)variances can be divided into components, indicating that the phenotypic observation of consistent within-family variance corresponds to a large number of Mendelian factors with additive effects (Barton *et al.*, 2017; Boyle *et al.*, 2017).

Despite the effectiveness of Fisher's infinitesimal model in explaining inheritance patterns, particularly in agricultural breeding, uncertainties persisted regarding the number of key genes influencing complex traits throughout the 20th century. Fisher's model proposed that complex traits could result from the combined effects of multiple genes with small additive contributions (Boyle *et al.*, 2017). This theoretical framework laid foundational concepts in quantitative genetics, exploring how genetic variation could contribute to phenotypic diversity through additive genetic effects. Subsequent technological advancements, such as Genome-Wide Association Studies (hereafter GWAS), have further clarified the polygenic nature of traits by revealing associations between specific genetic variants across the genome and complex phenotypes (Boyle *et al.*, 2017). These studies have provided crucial insights, indicating that phenotypic traits are often influenced by a multitude of genes distributed across the genome.

Nevertheless, studying the relationship between phenotype and genotype remains challenging. This task arises not only due to the polygenic nature of traits but also because phenotypic variability

results from a combination of environment and additive and epistatic effects involving multiple alleles across different genes. The reconstruction of phenotypes from genomic data is additionally complicated by inadequate phenotype descriptions, incomplete or insufficient genotype data and the intricate network of interactions governing cellular functions (Frazier *et al.*, 2003; Schulze and McMahon, 2004).

As a consequence of these interactions, a minority of individuals demonstrate extreme phenotypic traits, while the majority exhibit average characteristics. Additionally, phenotypic changes typically manifest as quantitative variations in intensity rather than the mere presence or absence of traits, as noted by Frazier *et al.* (2003). It is crucial to recognize the disparity between genotype stability and the dynamic and responsive nature of phenotypes throughout an individual's lifespan, as discussed by Schulze and McMahon (2004).

To advance understanding in this research field, GWAS can disclose significant genetic variants linked to specific traits. These studies identify molecular markers such as SNPs associated with traits of interest (Sulem *et al.*, 2007; Liu *et al.*, 2009; Branicki *et al.*, 2011; Fagertun *et al.*, 2015; Adhikari *et al.*, 2016; Walsh *et al.*, 2017; Böhringer and de Jong, 2019; Liu *et al.*, 2021; Dorado *et al.*, 2021), facilitating correlations between molecular markers and quantitative phenotypic traits. Furthermore, mathematical models based on genotypic data allow for phenotype prediction. Recent advancements in computing, particularly in bioinformatics, are invaluable for phenotyping research, enabling the integration of environmental, phenotypic and omics data to comprehensively understand genetic influences on complex traits and their interactions with external factors (Dorado *et al.*, 2021).

Predicting phenotypes from modern data remains a formidable challenge, despite technological advancements that offer assistance. However, this challenge becomes even greater when attempting to make predictions using ancient data. Reconstructing phenotypes from aDNA presents significant challenges, primarily arising from the quality and diversity of available genomic data. aDNA samples are frequently degraded and contaminated, posing obstacles to obtaining sequences of high quality. Moreover, the environmental conditions in which these samples are discovered can further impact the preservation of genetic material. Unlike modern humans, who benefit from extensive datasets and controlled environments for phenotype prediction, studying ancient Hominins presents more challenging circumstances due to the necessity of inferring phenotypes from genomic and environmental contexts that are no longer accessible. This requires an emphasis on molecular phenotypes

at lower levels, such as gene expression and protein function, which are comparatively better characterized and more stable across diverse populations and environments. As genomic technologies and computational methods progress, integrating comprehensive and high-quality genomic data from both modern and ancient sources will be indispensable for surmounting these challenges (Dorado *et al.*, 2021; Brand *et al.*, 2022).

Understanding the physical characteristics and traits of ancient humans is crucial for unravelling the complexities of human evolution and gaining insights into the shared heritage of humanity. By examining and reconstructing the phenotypic traits of ancient populations, valuable information can be gathered about the adaptations, behaviours and physical characteristics that defined ancestors. This knowledge enhances comprehension of historical human societies and offers essential context for interpreting contemporary genetic diversity and phenotypic variability. Additionally, exploring ancient phenotypes facilitates the investigation of how factors such as environmental influences, genetic adaptations and cultural practices have shaped the trajectory of human development over time. Through the study of ancient human phenotypes, connections between the past and present can be established, promoting a deeper understanding of the diversity and flexibility exhibited by the human species throughout history.

1.5 Forensic DNA Phenotyping of Pigmentation Traits

Forensic DNA Phenotyping (hereafter FDP) differs from traditional DNA profiling, which relies on matching DNA to existing profiles in databases and often fails to identify individuals without prior records. FDP predicts physical traits directly from DNA extracted from human biological samples (Santos *et al.*, 2013; Kayser, 2015; Kayser *et al.*, 2023), generating investigative leads. Recent advancements in FDP have expanded the scope and accuracy of predicted traits, encompassing a wide range of externally visible characteristics (hereafter EVCs) such as eyebrow colour, freckles, hair structure, male pattern baldness, and height (Kayser *et al.*, 2023), sub-continental ancestry inferences (Xavier *et al.*, 2022) and age estimations from various tissue types (blood, saliva, bones, and semen) (Heidegger *et al.*, 2020; Woźniak *et al.*, 2021; Pisarek *et al.*, 2021).

Technological progress, particularly in targeted massively parallel sequencing (MPS), has significantly increased the capacity for simultaneous analysis of numerous DNA predictors. Recent developments include innovative tools like the FORCE panel (Forensic Capture Enrichment) with thousands of SNPs for appearance and ancestry prediction (Tillmar *et al.*, 2021), the VISAGE Enhanced Tool for

appearance and ancestry inference (Heidegger *et al.*, 2020), HIrisPlex-S known for accurate eye, hair, and skin colour prediction (Chaitanya *et al.*, 2018), and age estimation tools based on DNA methylation analysis (Freire-Aradas *et al.*, 2017).

However, studies on DNA-based prediction of appearance, ancestry, and age for forensic purposes highlight common challenges. These include small sample sizes affecting marker predictability, and the use of limited, non-independent datasets during model development stages. Complex EVCs, similarly to common diseases, pose challenges in identifying responsible genes and predictive DNA markers due to the combined effects of multiple genetic factors. The genetic basis often involves minor contributions from individual genes, necessitating large-scale studies for statistical significance, particularly in the presence of significant environmental influences that can obscure genetic signals (Wray *et al.*, 2013; Kayser *et al.*, 2023).

Contrariwise, human pigmentation traits are among the least genetically complex of all EVCs, primarily influenced by a small number of genes. This relative simplicity facilitates advanced genetic understanding and DNA-based prediction compared to other EVCs, which typically involve numerous genes contributing to their complexity (Kayser, 2015; Kayser *et al.*, 2023). For this reason, in the following sections are summarized the recent advances in DNA prediction of eye, hair, and skin colour appearance.

1.5.1 Eye Colour

In 2007, two pioneering studies utilizing DNA for eye colour prediction were published. Frudakis and colleagues used 33 SNPs from the *OCA2* gene, achieving an 8% accuracy rate among over 1,000 samples (Frudakis *et al.*, 2007). Concurrently, Sulem *et al.* conducted the first GWAS on human pigmentation traits among thousands of Europeans, identifying 9 SNPs across 6 genomic regions significantly associated with eye colour (*SLC24A4*, *KITLG*, *6p25.3*, *TYR*, *OCA2-HERC2*, and *MC1R*). The model yielded a prediction accuracy of approximately 90% for blue eyes and 60% for brown eyes based on specific probability thresholds (Sulem *et al.*, 2007).

Subsequent research in 2008 highlighted the *HERC2* gene as crucial for eye colour prediction, particularly the rs12913832 SNP as a major predictor (R^2 from multiple linear regression modelling of categorical eye colour of 0.68%). The rs916977 SNP was found to be present in 97% of individuals with blue eyes (Kayser *et al.*, 2008; Sturm *et al.*, 2008; Eiberg *et al.*, 2008).

In 2009, Liu and colleagues conducted the first comprehensive study with 37 SNPs from 8 pigmentation genes among more than 6,100 Dutch Europeans. A predictive model for eye colour was developed utilizing 24 SNPs from 8 genes in a dataset of over 3,800 samples. Its accuracy was subsequently validated in an additional set of more than 2,300 independent samples, achieving AUC values of 0.93 for brown eyes, 0.91 for blue eyes, and 0.72 for intermediate eye colour. This study identified a minimal set of 6 SNPs, *HERC2* rs12913832, *OCA2* rs1800407, *SLC24A4* rs12896399, *SLC45A2* rs16891982, *TYR* rs1393350, and *IRF4* rs12203592, as primary predictors of eye colour (Liu *et al.*, 2009).

In 2010, Valenzuela *et al.* evaluated 75 SNPs across 24 genes and found that 3 SNPs (*HERC2* rs12913832, *SLC45A2* rs16891982, and *SLC24A5* rs1426654) could explain 76.45% of eye colour variation, with the vast majority attributed to *HERC2* rs12913832 alone. However, the research, which included both European and non-European populations, faced challenges in distinguishing the effects of ancestry from those of eye colour (Valenzuela *et al.*, 2010), complicating the interpretation of the prediction outcomes.

Between 2011 and 2013, various studies tested numerous SNPs and attempted to develop eye colour prediction tests. Notable examples include the IrisPlex system by Walsh *et al.* (Walsh *et al.*, 2012), the Spichenok prediction approach (Spichenok *et al.*, 2011), the work by Allwood and Harbison (Allwood and Harbison, 2013), and the study by Ruiz *et al.* with the *Snipper App suite* (http://mathgene.usc.es/snipper/) (Ruiz *et al.*, 2013). These studies demonstrated that incorporating *HERC2* and *OCA2* SNPs, which are in linkage disequilibrium, enhances prediction accuracies, particularly for intermediate eye colour.

Debate surrounding gender differences in eye colour prediction persists, with studies like those by F. Liu *et al.* (2014), Martinez-Cadenas *et al.* (2013), and Pietroni *et al.* (2014) suggesting varying effects, often attributed to sample size limitations (Martinez-Cadenas *et al.*, 2013; Liu *et al.*, 2014; Pietroni *et al.*, 2014). Emerging research on epistasis and SNP interactions in pigmentation traits, as explored by Branicki *et al.* (2009), F. Liu *et al.* (2010), and Pośpiech *et al.* (2011, 2014), underscores the complex genetic influences on eye colour and the potential for enhanced predictive models (Branicki *et al.*, 2009; F. Liu *et al.*, 2010; Pośpiech *et al.*, 2014; Pośpiech *et al.*, 2011).

Recent advancements in DNA-based eye colour prediction have significantly enhanced both predictive accuracy and genetic insight. Notably, refinements in models such as IrisPlex and HIrisPlex (Walsh *et al.*, 2011; Walsh *et al.*, 2013) and the introduction of VISAGE Software (Heidegger *et al.*,

2020) have enabled predictions even from incomplete DNA profiles commonly encountered in forensic settings. Furthermore, recent GWAS have identified numerous genetic loci associated with eye colour variation. For instance, a 2021 study by the VisiGen Consortium revealed 61 new genetic loci linked to eye colour, collectively explaining a substantial proportion of observed variation in European populations (Simcoe *et al.*, 2021).

1.5.2 Hair Colour

The first DNA test capable of predicting hair colour, specifically red hair, was published in 2001 by Grimes and colleagues. This study demonstrated that, using a DNA test based on 12 MC1R variants, 96% of individuals with two mutations associated with red hair were indeed confirmed to have red hair. The two individuals in the study who did not have red hair (one with blond hair and the other with light brown hair) reported having red hair in their youth (Grimes *et al.*, 2001).

In 2007, Branicki *et al.* sequenced the entire *MC1R* gene in over 180 individuals with various hair colours, including 40 with red hair and 36 with blond-red hair. They developed a DNA test for predicting red hair colour based on 5 *MC1R* variants (Branicki *et al.*, 2007). The same year, the first attempt to predict all categorical hair colours via DNA was published as part of the Sulem *et al.* pigmentation GWAS study. Using 2 *MC1R* SNPs, rs1805008 and rs1805007, they initially predicted red hair with about 70% accuracy. For other hair colours, 9 associated SNPs from 6 genes were used, but these predictions were less accurate (Sulem *et al.*, 2007).

In 2010, Valenzuela *et al.* reported that 3 SNPs (*SLC45A2* rs16891982, *SLC24A5* rs1426654, and *HERC2* rs12913832) could explain 76.3% of the total hair melanin variation (Valenzuela *et al.*, 2010).

In 2011, Branicki and colleagues conducted a systematic study using 46 SNPs from 13 genes previously associated with hair colour. They tested these for their predictive value in 385 Europeans from Poland and presented a model with 22 SNPs from 11 genes, achieving AUC values of 0.93 for red hair, 0.87 for black, 0.82 for brown, and 0.81 for blond. They could not validate the *SLC24A5* rs1426654 variant, as Valenzuela did in 2010, because in their Polish sample, 98.7% of the individuals were homozygous for the derived allele, with the few heterozygous individuals displaying different hair colours (Branicki *et al.*, 2011). Based on previous findings and the SNP rankings observed by Branicki's 2011 study, the first DNA test system for predicting all categorical hair colours, combined with eye colour prediction, was developed and published in 2013 as the HIrisPlex (Walsh *et al.*, 2013).

In 2013, 18 of the 22 SNPs utilized in the HIrisPlex system for hair colour prediction were incorporated into the commercial Identitas V1 Forensic Chip (http://identitascorp.com/). This tool, designed to provide forensic information including hair colour prediction, employs a modified HIrisPlex model that excludes 4 *MC1R* SNPs due to technical constraints, resulting in slightly diminished prediction accuracies (Keating *et al.*, 2013).

Recent advancements in DNA-based hair colour prediction have also demonstrated significant progress in predictive accuracy and genetic comprehension. Noteworthy updates include the refinement of predictive models such as HIrisPlex (Walsh *et al.*, 2013), now bolstered by a larger dataset encompassing nearly 1,900 samples and the development of VISAGE Software (Heidegger *et al.*, 2020), facilitating predictions from incomplete DNA profiles commonly encountered in forensic investigations. Additionally, recent GWAS studies have made substantial strides in identifying genetic loci linked to hair colour variation. For example, a comprehensive 2018 study by the VisiGen Consortium identified 124 significant genetic loci associated with hair colour (Hysi *et al.*, 2018).

1.5.3 Skin Colour

The current understanding of the genetics underlying skin colour variation is significantly less developed compared to that for eye and hair colour. This disparity arises from the global diversity of skin tone, in contrast to the predominantly European distribution of eye and hair colours, which allows for more straightforward genetic mapping in homogeneous populations. Traditional gene mapping methods, such as GWAS, are less effective for heterogeneous populations, resulting in skin colour research being conducted primarily within European or Asian groups. However, these studies are limited by the relatively small variation in skin tones within these groups compared to the broader intercontinental differences (Kayser, 2015).

For instance, the 2010 study by Valenzuela *et al.* demonstrated high predictive accuracy for eye and hair colour (76.4% and 76.3%, respectively) but only 45.7% for skin reflectance using 3 specific SNPs (Valenzuela *et al.*, 2010). Other studies by Spichenok *et al.*, Pneuman *et al.*, and Hart *et al.* using similar SNP sets showed low error rates but a significant proportion of inconclusive results, highlighting the limitations of these approaches and the necessity for careful interpretation of findings, especially given the lack of detailed skin colour variation data within European populations (Spichenok *et al.*, 2011; Pneuman *et al.*, 2012; Hart *et al.*, 2013).

In 2014, Maroñas *et al.* conducted a comprehensive study involving 59 SNPs linked to skin, eye, and hair colour in a diverse set of European and non-European individuals. They identified 29 SNPs significantly associated with skin tone variation. A principal component analysis (PCA) using these SNPs effectively separated individuals with white skin from those with intermediate or black skin, though there was considerable overlap between intermediate and black skin colours. The authors suggested 6 SNPs for skin colour prediction, achieving high classification success in Bayes analysis, and an expanded 10-SNP set that showed strong predictive accuracy in a small validation set. Nonetheless, the study limited sample size indicates the need for further data to validate these findings (Maroñas *et al.*, 2014).

In 2017, a group led by Manfred Kayser developed a prediction model for 5 skin tone categories for the first time both within and between continental regions (Walsh *et al.*, 2017). This study evaluated the predictive value of 77 SNPs from 37 genetic loci previously associated with human pigmentation in 2,025 individuals from 31 global populations. A minimal set of 36 highly informative SNPs was identified, leading to the development of a statistical prediction model for global skin colour prediction. The model demonstrated average cross-validated prediction accuracies, of approximately 0.74% for light skin tone, 0.73% for intermediate skin tone, and 0.90% for dark skin tone. Comparative analysis with 194 independent samples from 17 populations showed that this model outperformed the previously proposed 10-SNP classifier (Maroñas *et al.*, 2014), with AUCs increasing for white and black skin tones and comparable levels for intermediate. In 2018 these 36 SNPs were subsequently integrated into the HIrisPlex system, resulting in the development of the new HIrisPlex-S system (Chaitanya *et al.*, 2018).

In 2018, Visconti *et al.* conducted a GWAS on sun sensitivity, a proxy for skin colour, involving over 175,000 Europeans. This study identified 20 significant genetic loci, 14 of which were previously unknown to be associated with skin colour (Visconti *et al.*, 2018). Similar GWAS have been carried out in non-European populations, including Africans (Crawford *et al.*, 2017) and South Asians (Jonnalagadda *et al.*, 2019), revealing both previously known and novel genetic loci. However, these newly discovered SNPs require further independent validation to confirm their predictive value for skin colour.

Looking forward, ongoing research aims to integrate these genetic insights into forensic genetics and anthropological studies, potentially advancing eye, hair and skin colour prediction tools from

categorical to continuous models. These developments underscore the evolving capabilities in DNA-based colour prediction, with implications for both scientific exploration and practical forensic applications (Kayser *et al.*, 2023).

1.6 The HIrisPlex-S System

In the field of FDP, one of the most widely used tools is HIrisPlex-S (Chaitanya *et al.*, 2018), designed to predict eye, hair, and skin colour from DNA samples. It employs specific genetic markers linked to pigmentation traits. By analysing these markers, HIrisPlex-S yields probabilistic estimates of phenotypic traits.

The system originates from the earlier IrisPlex assay (Walsh *et al.*, 2011), initially focused on eye colour prediction, which later expanded into HIrisPlex (Walsh *et al.*, 2013) to encompass hair colour prediction. Finally, HIrisPlex-S was developed to include skin colour prediction as well (Chaitanya *et al.*, 2018).

HIrisPlex-S integrates data from 41 genetic loci, each associated with different pigmentation traits, enhancing the accuracy of phenotype predictions. Among these, 6 SNPs predict eye colour (Liu *et al.*, 2009), 18 predict hair colour (Branicki *et al.*, 2011), and 17 predict skin pigmentation, with 19 SNPs informative for both eye and hair colour also contributing to skin colour prediction. Therefore, a total of 36 SNPs are informative for skin coloration prediction (Walsh *et al.*, 2017).

For the purpose of accessibility, developers have instituted an online platform (https://hirisplex.erasmusmc.nl/) that facilitates genotype uploads and phenotype predictions derived from 41 SNPs.

The system predicts three potential eye colours (Blue, Intermediate, and Brown), four hair colours (Blond, Brown, Red, and Black), and five skin colours (Very Pale, Pale, Intermediate, Dark, and Dark-to-Black). This is achieved through two multiplex genotyping assays: a 24-plex assay from HIrisPlex for eye and hair pigmentation (Walsh *et al.*, 2013; Walsh *et al.*, 2014), and a novel 17-plex assay for skin pigmentation (Chaitanya *et al.*, 2018). Three distinct statistical prediction models were tailored for each phenotype category.

The laboratory protocol for HIrisPlex-S involves precise steps to ensure accurate genotyping results. Initially, DNA extraction and quantification are performed, requiring a minimum input of 63 picograms of DNA. This is followed by multiplex PCR under specific thermal conditions to amplify target SNPs. After purification, single-base extension (SBE) reactions detect SNPs using fluorescently

labelled nucleotides. Post-SBE, products undergo purification and analysis via an ABI 3130xl Genetic Analyzer (Liu *et al.*, 2009; Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018).

Statistical prediction models utilize three multinomial logistic regression models, one for each phenotype category (Liu *et al.*, 2009; Walsh *et al.*, 2013; Walsh *et al.*, 2017; Chaitanya *et al.*, 2018). Multinomial logistic regression extends the binary logistic regression framework by accommodating more than two possible categories in its predictions. This model estimates the likelihood of each category based on input variables, utilizing a logit function to transform the linear scale of independent variables into a probability scale ranging from 0 to 1. This transformation establishes the relationship between genetic markers (independent variables) and the likelihood of each outcome (Liu *et al.*, 2009).

Accurate prediction depends on specific SNP conditions: *HERC2* rs12913832 for eye colour, 11 *MC1R* SNPs for hair colour, and a combination of *HERC2-SLC45A2-IRF4* SNPs for both. For skin colour, including predictive SNPs like *SLC24A5* rs1426654, *OCA2* rs1800414, and *MC1R* rs3212355 is recommended to distinguish between intermediate and darker tones. Adhering to these conditions is crucial for maintaining prediction precision (Walsh *et al.*, 2014; Walsh *et al.*, 2017; Chaitanya *et al.*, 2018).

HIrisPlex-S accuracy varies across eye, hair, and skin colour categories, as indicated by Area Under the Curve (hereafter AUC) values derived from informative SNPs.

For eye colour, the AUC values indicate varying levels of prediction confidence (Liu *et al.*, 2009):

- AUC of 0.939 +/- 0.005 for the "Blue" category;
- AUC of 0.735 +/- 0.005 for the "Intermediate" category;
- AUC of 0.946 +/- 0.005 for the "Brown" category.

Regarding hair colour prediction accuracy (Branicki *et al.*, 2011):

- AUC of 0.800 +/- 0.020 for the "Blond" category;
- AUC of 0.719 +/- 0.024 for the "Brown" category;
- AUC of 0.916 +/- 0.029 for the "Red" category;
- AUC of 0.831 +/- 0.031 for the "Black" category.

In terms of skin colour prediction accuracy (Walsh *et al.*, 2017):

- AUC of 0.744 +/- 0.050 for the "Very Pale" category;

- AUC of 0.719 +/- 0.025 for the "Pale" category:
- AUC of 0.731 +/- 0.026 for the "Intermediate" category;
- AUC of 0.875 +/- 0.084 for the "Dark" category;
- AUC of 0.958 +/- 0.036 for the "Dark-to-Black" category.

Development included testing phases with diverse sample sets specific to each predictive model. For eye colour prediction, 3,804 Dutch individuals were used and this database was later enriched with studies on samples from different geographical origins (Walsh *et al.*, 2014). Meanwhile, for hair colouration, 1,243 samples of Polish, Irish and Greek origin were used, as well as 945 other samples representing 51 world populations (Walsh *et al.*, 2013). Skin colour prediction necessitated a global sample set due to its extensive variation across populations. Therefore, the testing phase for skin pigmentation predictions involved 1,159 individuals from southern Poland, 347 Irish samples, 119 Greeks and 329 individuals of non-U.S. parental origin (Nigeria, Mexico, Argentina, Colombia, India, China, Russia, Saudi Arabia, and so forth) as well as subjects from Senegal, Kenya, Nigeria and Papua New Guinea (Walsh *et al.*, 2017).

The two multiplex genotyping assays demonstrated resilience to UV-induced DNA degradation and the HIrisPlex-S system reliably genotyped and phenotyped single-source samples in forensic simulations, albeit with limitations in accurately predicting phenotypes from DNA mixtures (Chaitanya *et al.*, 2018).

The HIrisPlex-S system adheres to SWGDAM (Scientific Working Group on DNA Analysis Methods) guidelines for forensic validation, ensuring sensitivity, stability, and reliability in simulated casework samples.

1.7 The HIrisPlex-S System Applied to Ancient DNA

aDNA and forensic DNA analysis share several common challenges, necessitating sophisticated approaches for their study. Both fields must contend with DNA degradation, resulting in fragmented and low-quality DNA samples. Additionally, they face issues with contemporary contamination and the presence of inhibitors, such as humic acids, which can obstruct enzymatic reactions critical for DNA analysis. Various types of DNA damage, including deamination-induced C-T substitutions, may be present in both ancient and forensic DNA. Meticulous optimization of every step, from sample

selection to DNA extraction and sequencing, is required in both disciplines to ensure maximum yield and quality of DNA (Hofreiter *et al.*, 2021).

Another shared feature between ancient and forensic DNA analysis is the employment of the HIrisPlex-S kit, which has been applied not only in forensic contexts or with modern samples but also with ancient remains.

The application of the classic HIrisPlex-S system protocol, along with earlier versions like IrisPlex or HIrisPlex, has been conducted on a diverse array of specimens.

This research spans multiple studies (Olalde *et al.*, 2014; Lazaridis *et al.*, 2014; Gamba *et al.*, 2014; Jones *et al.*, 2015; Cassidy *et al.*, 2016; González-Fortes *et al.*, 2017; Günther *et al.*, 2018; Olalde *et al.*, 2019; Brace *et al.*, 2019; Jensen *et al.*, 2019; Seguin-Orlando *et al.*, 2021; Clemente *et al.*, 2021; Allentoft, Sikora, Fischer, *et al.*, 2024), showcasing the widespread application of these protocols in aDNA research. These include Palaeolithic samples such as Romanian PM1 (Svensson *et al.*, 2021), as well as medieval ones (Draus-Barini *et al.*, 2013). Among the notable cases are remains dating to around 1485 attributed to King Richard III (King *et al.*, 2014), the Mesolithic Spanish individuals “La Braña” (Olalde *et al.*, 2018), and “Chan” (González-Fortes *et al.*, 2017), and the English sample “Cheddar Man” (Brace *et al.*, 2019). Additionally, there is the Neolithic Danish individual “Lola”, whose DNA was sequenced from a chewing gum dating back approximately 5700 years (Jensen *et al.*, 2019) (Figure 1.7).

Figure 1.7 - A) The model of Cheddar Man rendered by Kennis and *Kennis Reconstructions. Credit: CHANNEL 4. B) Reconstruction of La Braña. Credit: PELOPANTON/CSIC. C) Reconstruction of Lola. Credit: Tom Björklund*

It is worth noting that, when dealing with aDNA data as discussed in paragraph 1.1.2, most samples exhibit low coverage, resulting in potential uncertainty in genotypes. Additionally, not all 41 positions of the HIrisPlex-S are uniformly covered across these samples. In cases of missing data, such as with "Cheddar Man", predictions were made with various missing positions. Similarly, for "Chan",

missing genotypes have been imputed using pipelines that have not yet been validated for ancient DNA analysis.

Moreover, while the laboratory protocols of the HIrisPlex-S system have been validated for forensic procedures, enabling the effective handling of degraded and low-coverage data, bioinformatic protocols necessary for robust management of this specific type of data have not yet been fully implemented, indicating a need for further development in this area.

2 AIM

The primary aim of this study is twofold: first, to give users a sense of the operational effectiveness and robustness of the HIrisPlex-S system when applied to low-coverage ancient data, thereby assessing the reliability of phenotypic inference procedures; and second, to propose a novel methodology based on genotype likelihoods that explicitly accounts for genotype calling uncertainty in low-coverage data.

Within this framework, this thesis aims to determine the optimal coverage threshold for accurate inference using the standard HIrisPlex-S protocol and to assess the efficacy of common imputation methods for managing missing data in low-coverage genomes, particularly in the context of predicting phenotypic traits in ancient samples.

To evaluate the accuracy of predictions regarding physical traits, a downsampling procedure was applied to high-coverage data from three test samples: one from the Palaeolithic period in Siberia, one from the Mesolithic period in Sweden and one from the Bronze Age in Germany.

Predictions of eye, hair, and skin colour were subsequently generated using three models: the classical HIrisPlex-S workflow, the genotype likelihood approach here proposed and the imputation method, each applied across varying data coverage levels achieved through downsampling.

Following the analysis of prediction accuracy, this study revised and implemented guidelines for inferring phenotypic traits using the HIrisPlex-S system.

These guidelines were then applied to a cohort of 348 Eurasian individuals spanning from the Upper Palaeolithic to the Iron Age, providing insights into the physical appearances of ancient populations.

3 METHODS

3.1 Conceptual framework

The HIrisPlex-S system can predict eye, hair, and skin colour by analysing 41 DNA genetic variants. While the standard protocol does not specifically address caution regarding sample coverage or age, genotyping aDNA presents a primary challenge due to the limited number of unique aDNA fragments available for sequencing. This often results in low coverage or missing data.

The approach followed in this study to quantify the impact of coverage is schematically described in Figure 3.1. Three different prediction workflows for phenotypic traits have been tested on three downsampled high-coverage ancient genomes. The first protocol follows the classical procedure of HIrisPlex-S, based on direct genotype calling (Chaitanya *et al.*, 2018). The second prediction is based on the new method here presented, which uses a direct genotype likelihood approach to define genotype posterior probabilities. The third prediction involves recovering missing positions by imputing missing variants following a pipeline specifically developed for ancient DNA data (Sousa da Mota *et al.*, 2023).

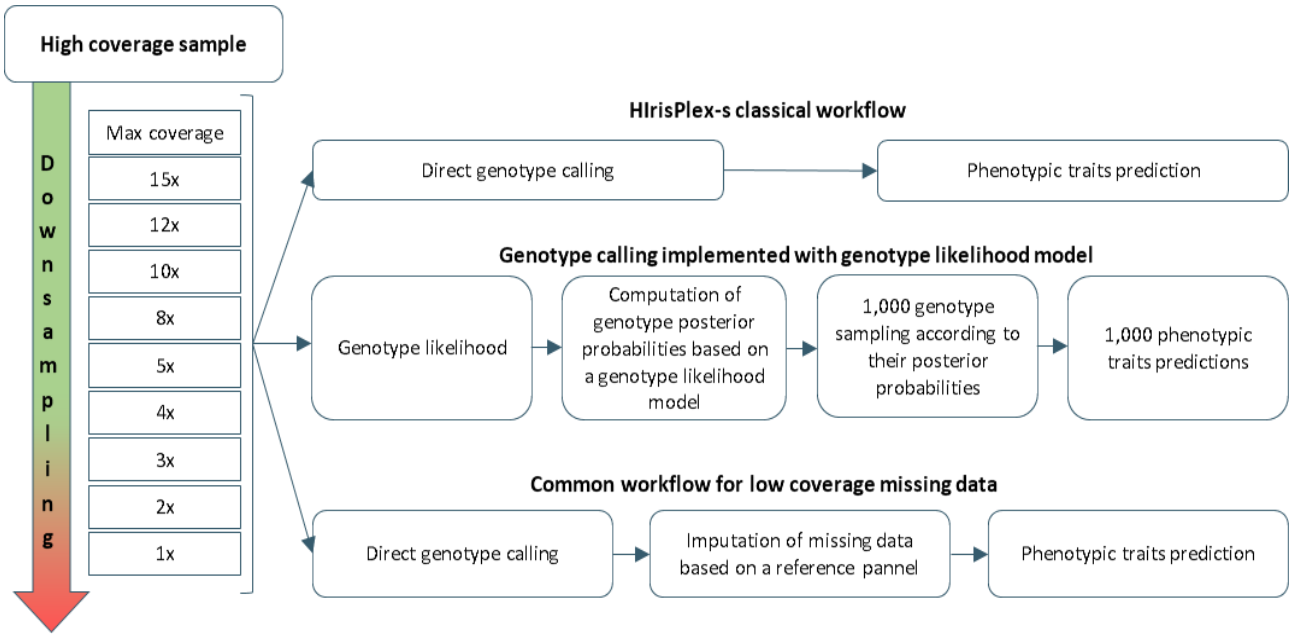


Figure 3.1 - Diagram illustrating the conceptual framework: a sequence of steps starting with the original data at the top, followed by a progressive reduction of coverage through multiple stages of downsampling. The process leads to three branches, each representing the different prediction workflow to be tested, explained step by step.

3.2 Genomic dataset

The dataset built in this thesis comprises samples from the Upper Palaeolithic to the Iron Age, encompassing the geographic region of Eurasia (Figure 3.2, Table 3.1).

The samples were gathered from previous studies and obtained from the European Nucleotide Archive (*ENA*) and from the Allen Ancient Genome Diversity Project (*AAGDP*) (https://reich.hms.harvard.edu/ancient-genome-diversity-project), required to have an average coverage of at least 1x and undergo whole-genome sequencing. In total, 348 samples were collected.

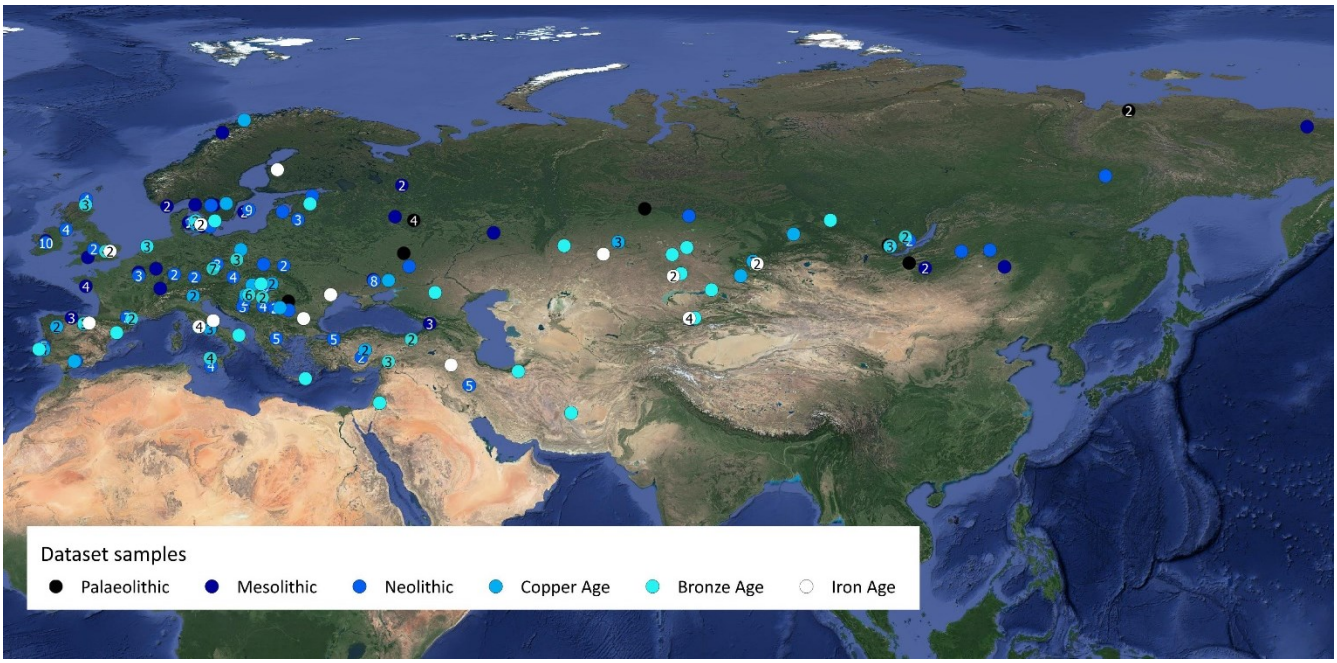


Figure 3.2 - Geographical distribution of selected samples. Black dots denote Palaeolithic period samples, white dots denote Iron Age samples and intermediate colours represent transitional periods. Numbers indicate multiple samples at a site.

PERIOD	COUNTRY	SAMPLE SIZE
Palaeolithic	Czech Republic	1
Palaeolithic	Romania	1
Palaeolithic	Russia	10
Mesolithic	Denmark	10
Mesolithic	France	5
Mesolithic	Georgia	3
Mesolithic	Ireland	2
Mesolithic	Italy	2
Mesolithic	Latvia	5
Mesolithic	Luxemburg	1
Mesolithic	Norway	3
Mesolithic	Romania	3
Mesolithic	Russia	9
Mesolithic	Serbia	7
Mesolithic	Spain	5

Mesolithic	Sweden	6
Mesolithic	Switzerland	1
Mesolithic	UK	1
Mesolithic	Ukraine	3
Neolithic	Austria	4
Neolithic	Bulgaria	1
Neolithic	Croatia	5
Neolithic	Czech Republic	3
Neolithic	Denmark	9
Neolithic	Estonia	1
Neolithic	France	5
Neolithic	Germany	4
Neolithic	Great Britain	8
Neolithic	Greece	5
Neolithic	Hungary	6
Neolithic	Iran	5
Neolithic	Ireland	12
Neolithic	Italy	6
Neolithic	Latvia	4
Neolithic	Malta	2
Neolithic	Poland	3
Neolithic	Portugal	5
Neolithic	Romania	1
Neolithic	Russia	9
Neolithic	Serbia	6
Neolithic	Spain	1
Neolithic	Sweden	12
Neolithic	Turkey	7
Neolithic	Ukraine	8
Copper Age	Bulgaria	1
Copper Age	Croatia	1
Copper Age	Denmark	2
Copper Age	Great Britain	2
Copper Age	Hungary	3
Copper Age	Italy	5
Copper Age	Kazakhstan	3
Copper Age	Norway	1
Copper Age	Poland	1
Copper Age	Portugal	6
Copper Age	Romania	2
Copper Age	Russia	5
Copper Age	Spain	5
Copper Age	Sweden	2
Copper Age	Turkey	2
Copper Age	Ukraine	1
Bronze Age	Bulgaria	2
Bronze Age	Croatia	3
Bronze Age	Czech Republic	5
Bronze Age	Denmark	2
Bronze Age	Estonia	1
Bronze Age	France	2
Bronze Age	Germany	2

Bronze Age	Great Britain	7
Bronze Age	Greece	6
Bronze Age	Hungary	4
Bronze Age	Iran	2
Bronze Age	Italy	6
Bronze Age	Jordan	1
Bronze Age	Kazakhstan	5
Bronze Age	Netherlands	3
Bronze Age	Poland	3
Bronze Age	Portugal	1
Bronze Age	Russia	8
Bronze Age	Serbia	2
Bronze Age	Spain	4
Bronze Age	Sweden	1
Iron Age	Bulgaria	1
Iron Age	Denmark	2
Iron Age	Finland	1
Iron Age	Great Britain	2
Iron Age	Iran	1
Iron Age	Italy	5
Iron Age	Kazakhstan	7
Iron Age	Pakistan	1
Iron Age	Russia	4
Iron Age	Spain	1
Iron Age	Ukraine	1

Table 3.1 - Summary table of whole-genome samples collected from ENA for phenotypic inference.

Details of all samples included in the dataset are available in the supplementary materials folder at the provided link: *Eurasian_Dataset.xlsx*.

3.3 Test sample selection and preparation

After creating the dataset, the next step was to identify samples suitable for downsampling. The main criteria were to ensure the representation of different historical periods and to select those with the highest coverage at the 41 HIrisPlex-S positions. Based on these conditions, test samples could only be identified from the Palaeolithic, Mesolithic, and Bronze Age periods:

- Ust'-Ishim, 45,045 calBP from western Siberia (Fu *et al.*, 2014);
- SF12, 9,033±8,757 calBP from Sweden (Günther *et al.*, 2018);
- I5835, 4,520–4,020 BP from Germany (Olalde *et al.*, 2018).

All the samples were processed following guidelines specifically designed for ancient genome reconstruction. For Ust'-Ishim and SF12 samples, the fastq files were downloaded from *ENA* and analysed

with *FastQC* v0.12.1 (Andrews S. (2010)) for quality control of the DNA fragments and removal of adapter sequences. The filtered reads were then aligned with *BWA* software v0.7.17 246 with the "*aln*" algorithm. The minimum read length *(--minlength 30*), base quality (*--minquality 30*) and mapping quality (*-q 30*) filters were applied, and seeding was disabled (*-l 1000*) to ensure alignment of all reads against the GRCh37.p13/hg19 human reference genome (Church *et al.*, 2011).

In contrast, I5835 sample underwent similar bioinformatic processing, as described in the literature, with the addition of a duplicate removal phase using *Picard* v2.24.1 (https://broadinstitute.github.io/picard/), for added reliability, alongside the other samples.

The final average coverage obtained, sex chromosomes excluded, was calculated using *SAMtools* software v1.11 (H. Li *et al.*, 2009) with the option "*-a*", to include in the calculus all positions (including 0 depth): 28.4822x for the Ust'-Ishim sample, 43.9357x for the SF12 sample and 22.5315x for the I5835.

3.3.1 *SAMtools* Mpileup file

The final step entailed generating a mpileup file for each sample using *SAMtools* software v1.11 248. As reported in the *SAMtools* manual page (https://www.htslib.org/doc/samtools-mpileup.html), the *SAMtools* mpileup command produces a "*pileup*" text format from alignment data, summarizing mapped reads derived from BAM alignment files. Within this context, several columns contain numeric quality values represented as individual ASCII characters, ranging from "*!*" to "~", with decoding achieved by subtracting 33 from the ASCII value (Figure 3.3).

The initial 3 columns provide essential positional and reference details:

1. Chromosome name.
2. 1-based position on the chromosome.
3. Reference base at this position ("N" for lines where -f/--fasta-ref has not been used).

Subsequent columns present the pileup data for each specified input BAM file:

4. Number of reads covering this position.
5. Read bases, encoding information on matches, mismatches, indels, strand, mapping quality, and read boundaries. For each read covering:
    - If this is the first position covered by the read, a "^" character followed by the alignment's mapping quality encoded as an ASCII character.

- If this is the last position covered by the read, a "$" character.
- A single character indicating the read base and the strand to which the read has been mapped:
  - "." dot or "," comma for base matches the reference base for forward or reverse strand, respectively.
  - Uppercase "ACGTN" or lowercase "acgtn" alphabetic characters for base in mismatch to the reference base for forward or reverse strand, respectively.
- ">" or "<" for reference skip (due to CIGAR "N").
- "*" (forward-del) or "/#" (reverse-del) for deletion of the reference base.

6. Base qualities, encoded as ASCII characters.
7. Additional comma-separated read field columns, as selected via "*--output-extra*". The fields selected appear in the same order as in SAM: QNAME, FLAG, RNAME, POS, MAPQ (displayed numerically) and so forth.

```
chr16 89986546 G 20 ..$,,..a.,.,,,,,.,.,,, ]]]]II]]]]]]]]]]]]]]
chr16 90024206 A 23 ..,..,,,,,,..,,,,.,.,,,^F.^F,>]]]]]]]]]G]E]?]]Q]?E]]
chr20 32665748 A 29 gGggGGgGggggggGGg,gGGGgGgGGG^F,^Fg ]]]I]]]]]EF]]I]]]]]]II]]]]]]]
chr20 32785212 C 18 .,.,.....,,..,.,^F, ]]]I]I]][]]]II]]]I
chr20 33218090 G 29 A,,aAa,AAa,,A.Aaaaaa.,Aa,A.,a ]I]E]]I]]]]]]]G]]G]]I]]]]D]]]
```

Figure 3.3 - Section of the Ust'-Ishim's mpileup file. The first column represents the chromosome, the second column indicates the position, the third column displays the reference base, the fourth column shows the number of reads covering the position, the fifth column details the read bases, and the sixth column presents base qualities encoded in ASCII.

3.4 Genomic downsampling

The principle underlying the downsampling procedure is to accurately evaluate the amount of information that might be lost at lower coverage levels by artificially reducing the sequencing depth (i.e., the number of times a single nucleotide is sequenced). Essentially, it enables the observation of what remains unseen with reduced coverage and facilitates the analysis of how nucleotide call precision and reliability vary across different depths of coverage.

In this thesis, two different types of downsampling strategies were chosen, starting from a maximum achievable coverage: 1) targeted downsampling aimed at achieving specific coverage at the 41 HIrisPlex-S positions to evaluate the performance of the classical HIrisPlex-S workflow and the here proposed genotype likelihood approach method (from here on pointwise progressive

downsampling, Figure 3.4); 2) downsampling aimed at achieving an average coverage to assess the imputation method (from here on mean progressive downsampling Figure 3.5).

The maximum achievable coverage was determined by the minimum coverage observed across all 41 HIrisPlex-S positions, as detailed in the SAMtools mpileup file column 4. Specifically:

- Ust'-Ishim sample:
  - Maximum coverage achieved: 17x.
  - Specific coverage levels tasted: 17x, 15x 12x, 10x, 8x, 5x, 4x, 3x, 2x, 1x.
- SF12 sample:
  - Maximum coverage achieved: 33x.
  - Specific coverage levels tasted: 33x, 20x, 15x 12x, 10x, 8x, 5x, 4x, 3x, 2x, 1x.
- I5835 sample:
  - Maximum coverage achieved: 8x.
  - Specific coverage levels tasted: 8x, 5x, 4x, 3x, 2x, 1x.

For both methods, pointwise progressive downsampling and mean progressive downsampling, 10 different downsamplings were conducted for each sample at each coverage level, starting from maximum coverage and progressively reducing from the same set of reads to achieve lower coverages. A total of 540 new BAM files were generated.

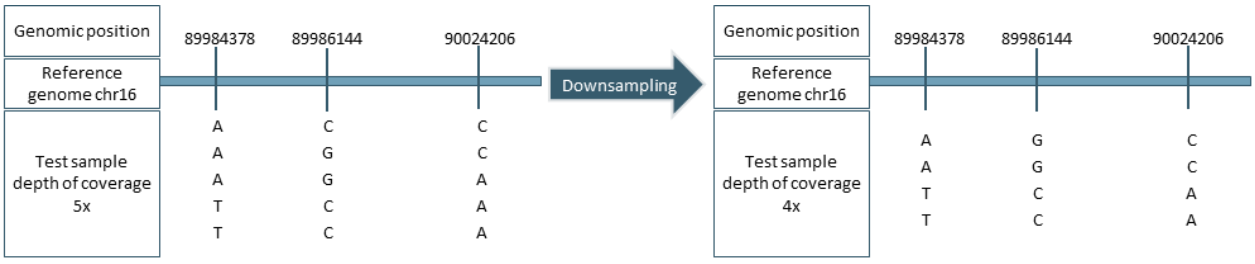


Figure 3.4 - Example of the conceptual scheme of the pointwise progressive downsampling. Letters stand for nucleotides.

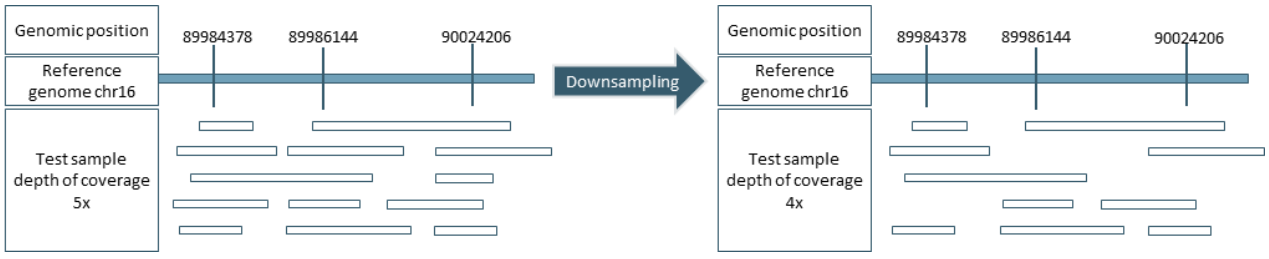


Figure 3.5 - Example of the conceptual scheme of mean progressive downsampling. Bands stand for genomic reads.

3.4.1 Pointwise progressive downsampling

To achieve incremental downsampling at each stage, the *SAMtools* mpileup file was employed to detect read identifiers ("*--output-extra QNAME*" flag) for sampling within the *R* environment (R Core Team, 2022). Specifically, non-replacement sampling was conducted, ensuring random selection of unique elements to prevent read and nucleotide duplication (using the option "*replace = false*").

Subsequently, reads from chromosomes 5, 6, 9, 11, 12, 14, 15, and 20 were collectively extracted using their identifiers directly from the original high-coverage BAM files via the *SAMtools* v1.11 "*view*" function (H. Li *et al.*, 2009), resulting in a single BAM file per chromosome. Chromosome 16 underwent a similar process but required 14 separate BAM files, one for each genomic position, to manage read overlap and potential coverage increases due to the proximity of the positions themselves.

For each BAM file, a new *SAMtools* mpileup file was generated to confirm the accurate extraction of reads and attainment of the target coverage level.

Once validated, read extraction continued from the new BAM files, progressively reducing coverage down to 1x. A total of 270 new BAM files were generated.

3.4.2 Mean progressive downsampling

The specific average coverage was determined using the "*view*" option of *SAMtools* v1.11 248. This process involved filtering reads with the "*-s*" flag to output a subset of the input alignments, followed by a fractional value representing the desired coverage ratio compared to the initial coverage. As described in previous sections, the process began with the high-coverage genome, progressively reducing it to 1x coverage. To ensure randomness in read selection, distinct seed values were employed for each downsampling operation, conducted on separate computing nodes.

To validate the accuracy of the achieved coverage, the *SAMtools* "*depth -a*" command was executed for each new BAM file to check for coverage. A total of 270 new BAM files were generated. Details of all mean coverage reached, fraction values and computing nodes are available in the supplementary materials folder at the provided link: Mean_progressive_downsampling.xlsx

3.5 HIrisPlex-S standard protocol

The standard HIrisPlex-S protocol comprises three main steps: indel verification, genotype calling, and phenotypic prediction via the website "https://hirisplex.erasmusmc.nl/" (Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018).

All procedures were conducted using pointwise progressive downsampled genomes. Specifically, for sample I5835, phenotypic predictions were performed using 10 iterations starting from 8x coverage, the maximum obtainable, and decreasing to 1x, the minimum coverage level to be achieved. For samples Ust'-Ishim and SF12, phenotypic predictions were also performed using 10 iterations. The first 5 iterations began from the maximum coverage achievable down to 1x and, due to the stability of phenotypic predictions and to standardize data with sample I5835, the last 5 iterations were carried out beginning from 8x coverage and progressed down to 1x. In total, 225 phenotypic predictions were generated.

3.5.1 Indel check

The indel at site rs312262906 (insertion of an "A" nucleotide at chromosome 16 position 89985753) was assessed using *VarScan* v2.3.9 (Koboldt *et al.*, 2012). Initially, *SAMtools* software v1.11 248 was employed to generate a mpileup file used by *VarScan* software to call indels with the algorithm "*mpileup2indel*". The resulting output file was in VCF format and included data such as the chromosome name, position, reference allele at the position, observed variant allele, total depth of coverage, number of reads supporting the reference allele, number of reads supporting the variant allele, variant allele frequency by read count, and p-value comparing observed reads versus expected non-variant reads. Additionally, it provided information to assess strand bias using all reads: forward strand support for reference allele, reverse strand support for reference allele, forward strand support for variant allele, reverse strand support for variant allele, and corresponding strand distribution p-values.

The presence of the indel was confirmed if "=+A" was reported at chromosome 16 positions 89985750, 89985751, 89985752, 89985753 or 89985754 (Figure 3.6).

chr16	89985750	C	=+A	*/+A:1080:583:496:45.93%:4.3379E-180	Pass:0.9858870967741935:582:1:489:7:2.0375E-2	0	1	0	0 */+A:1080:583:496:45.93%:4.3379E-180

Figure 3.6 - The image displays a sample showing the indel at position 89985750. Extracted from the HIrisPlex-S DNA Phenotyping Webtool User Manual Version 2.0 (2018), supplemental material.

Upon detection of an indel in a sample, critical examination focused on the read counts for the reference allele (e.g., "*" or Reads1) versus the alternative allele (e.g., "+A" or Reads2) at that locus. Based on these counts, determination was made whether the sample exhibited a heterozygous genotype (if reads were significant for both alleles) or a homozygous genotype (if reads predominantly favoured one allele). For instance, with a coverage (COV) of 1080, where READS1 numbered 583 and READS2 numbered 496 (Fig 3.7), the sample was classified as heterozygous, with the reference allele being "C" and the alternate allele being "+A". A predominance of "C" alleles indicated a homozygous genotype "CC", whereas a predominance of "+A" alleles indicated a homozygous genotype "AA".

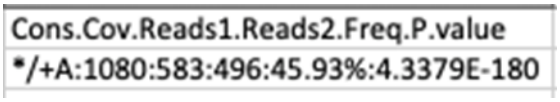


Figure 3.7 - Extracted from the HIrisPlex-S DNA Phenotyping Webtool User Manual Version 2.0 (2018), supplemental material, showing an example with COV of 1080, where READS1 numbered 583 and READS2 numbered 496.

3.5.2 Variant calling

Variants were called on BAM files using the *GATK* (Genome Analysis Toolkit) *UnifiedGenotyper* algorithm v3.5 (DePristo *et al.*, 2011). Key parameters included a minimum confidence threshold for calling variants (*-stand_call_conf* = 30), a minimum confidence threshold for emitting variants (*-stand_emit_conf* = 20) and a minimum base quality threshold (*-mbq* = 30). The analysis was configured to output all sites (*-out_mode EMIT_ALL_SITES*), regardless of whether they contain variant or invariant alleles compared to the reference genome (GRCh37.p13). Additionally, the "*-glm BOTH*" option was employed to simultaneously compute genotype likelihoods using both SNP and indel models, which is advantageous for accurately genotyping and calling variants of both types from the same dataset. The option "*-L*" was used to compute calling only on HIrisPlex-S sites. The final output was a VCF file reporting the 41 called genotypes.

3.5.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes

The HIrisPlex-S website "*https://hirisplex.erasmusmc.nl*" is designed to facilitate phenotypic predictions based on genotyped data. Users input genotype information corresponding to specific single nucleotide polymorphisms (SNPs) associated with eye, hair, and skin pigmentation (Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018).

Prior to conducting phenotypic predictions, the data were meticulously reviewed for accuracy. As recommended in the HIrisPlex-S Webtool User Manual Version 2.0 (2018) (Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018), it was necessary to adjust the strand orientation before entering them into this prediction model.

Several SNPs required conversion from NCBI's forward orientation, while all other G/C and A/T SNPs were correctly oriented for input (https://www.ncbi.nlm.nih.gov/snp/).

The correct orientation is reported in the table below (Table 3.2):

Gene	**SNP**	**Website Allele**	**Checked Allele**
MC1R	rs312262906	A	A
MC1R	rs11547464	A	A
MC1R	rs885479	T	A
MC1R	rs1805008	T	T
MC1R	rs1805005	T	T
MC1R	rs1805006	A	A
MC1R	rs1805007	T	T
TUBB3	rs1805009	C	C
MC1R	rs201326893	A	A
MC1R	rs2228479	A	A
MC1R	rs1110400	C	C
SLC45A2	rs28777	C	C
SLC45A2	rs16891982	C	C
KITLG	rs12821256	G	C
LOC105374875	rs4959270	A	A
IRF4	rs12203592	T	T
TYR	rs1042602	T	A
OCA2	rs1800407	A	T
SLC24A4	rs2402130	G	G
HERC2	rs12913832	T	A
PIGU	rs2378249	C	G
LOC105370627	rs12896399	T	T
TYR	rs1393350	T	A
TYRP1	rs683	G	C
ANKRD11	rs3114908	T	T
OCA2	rs1800414	C	C
BNC2	rs10756819	G	G
HERC2	rs2238289	C	G
SLC24A4	rs17128291	C	G
HERC2	rs6497292	C	G
HERC2	rs1129038	G	C
HERC2	rs1667394	C	C
TYR	rs1126809	A	A
OCA2	rs1470608	A	T
SLC24A5	rs1426654	G	G
ASIP	rs6119471	C	G
OCA2	rs1545397	T	T

RALY	rs6059655	T	A
OCA2	rs12441727	A	A
MC1R	rs3212355	A	T
DEF8	rs8051733	C	G

Table 3.2 - Correct strand orientation of the 41 HIrisPlex-S SNPs. Red SNPs indicate those that have been corrected.

Once these checks were completed, the data were fed into the designated section of the website (Figure 3.8) as here outlined:

- In the absence of the input allele, “0” should be entered into the “No. of alleles” box.
- For a heterozygous condition (one allele present), “1” should be entered.
- For a homozygous condition (both alleles identical), “2” should be entered.
- If any SNP is missing, “NA” should be entered.

Working with pointwise progressively downsampled genomes, no positions were missing and at 1x coverage genomes lost the heterozygous condition.

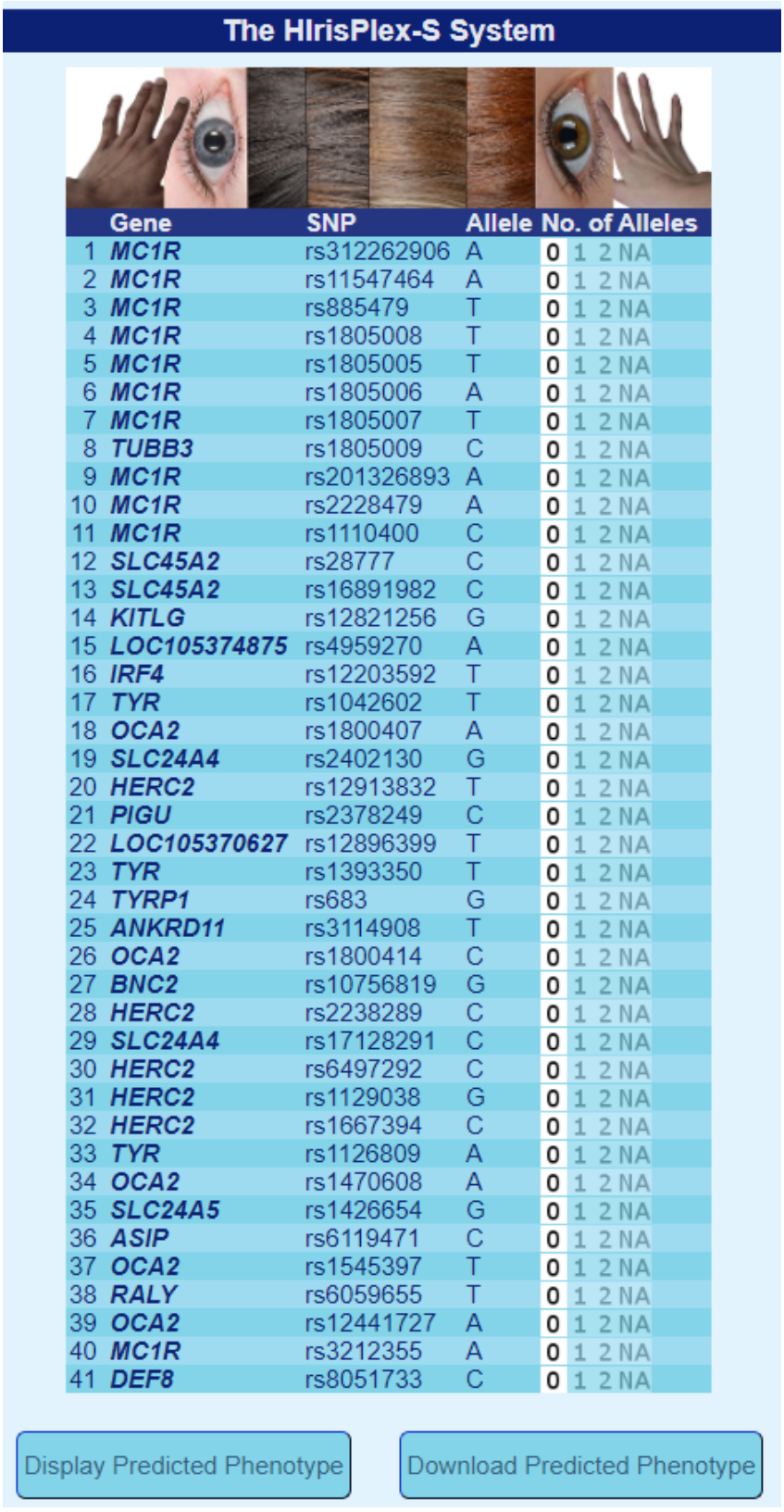


Figure 3.8 - Image extracted from the HIrisPlex-S website https://hirisplex.erasmusmc.nl/, illustrating the section dedicated to entering the number of informative alleles for each SNP.

The prediction results were downloaded as a CSV file ("Results.csv") and analysed in *RStudio* v2024.4.2.764 (RStudio Team, 2020), following the guidelines outlined in the HIrisPlex-S user Webtool User Manual Version 2.0 (2018) (Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018). The predictive models were then translated into *R* language, according to the decision models presented below, and prediction results rounded to three significant figures.

In terms of eye colour, the highest p-value is considered the predicted phenotype, with threshold accuracy levels set to 0.7, as indicated in Walsh *et al.* (2012).

In terms of hair colour, the methodology employed utilizes the highest p-value approach in conjunction with a stepwise model (Figure 3.9) (Walsh *et al.*, 2014). For instance, with a Blond p-value of 0.742 and a light value of 0.93, the prediction process proceeds as follows: the "Blond" category is selected due to its highest p-value among the categories, surpassing the threshold of 0.7. Subsequently, the evaluation checks if the light value exceeds 0.95p; as it measures 0.93p, the final prediction concludes with "Blond/Dark Blond."

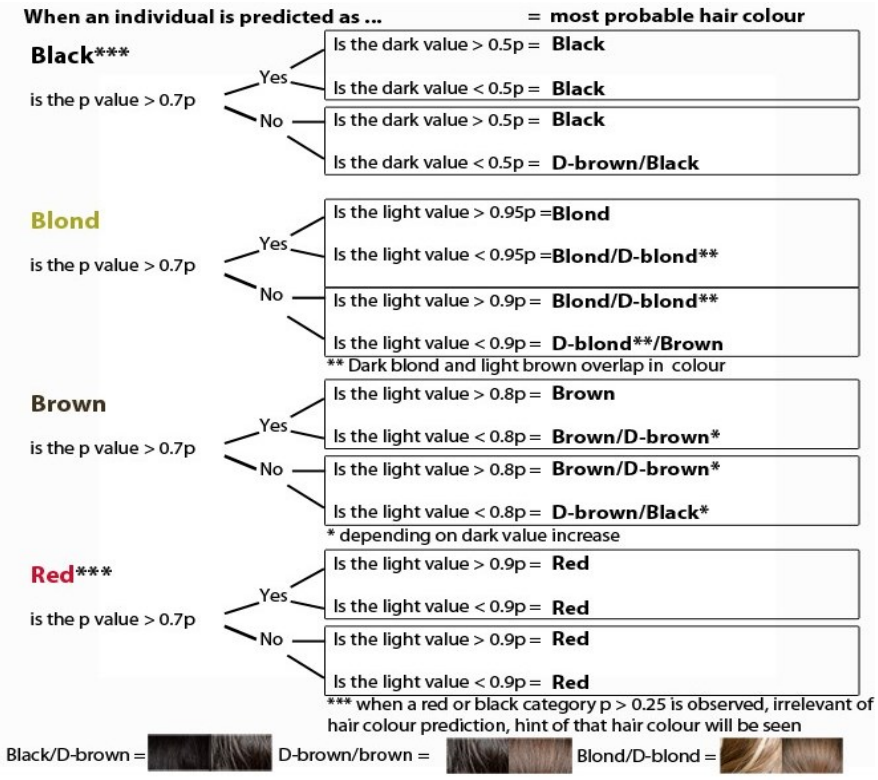
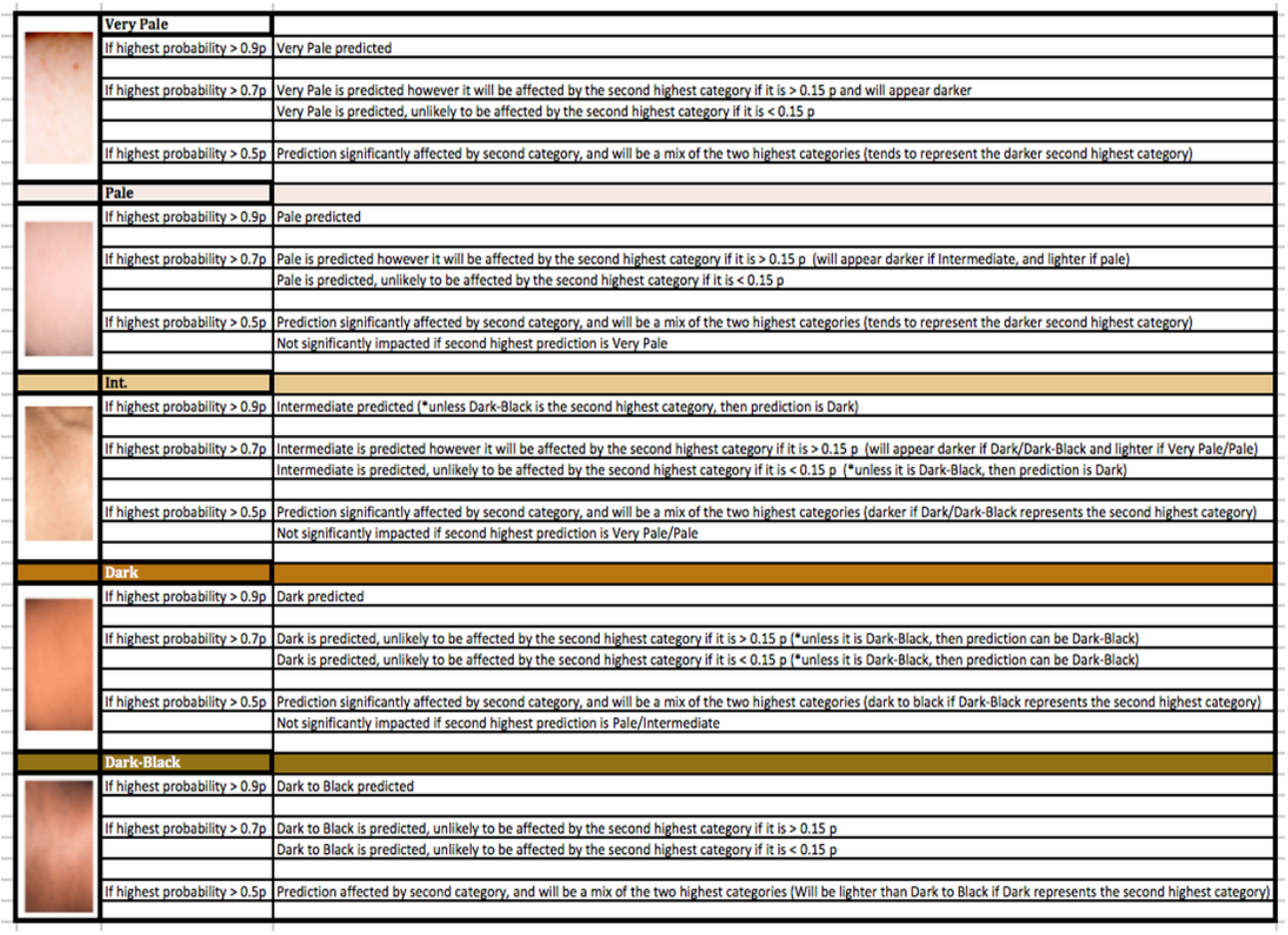


Figure 3.9 - Image extracted from the HIrisPlex-S DNA Phenotyping Webtool User Manual Version 2.0 (2018) illustrating the decision model for hair colour inference.

In terms of skin colour, the methodology employed utilizes the highest p-value approach in conjunction with the second-highest probability value, as described in 2018 by Chaitanya and colleagues (Chaitanya *et al.*, 2018) (Figure 3.10).



	Very Pale	
	If highest probability > 0.9p	Very Pale predicted
	If highest probability > 0.7p	Very Pale is predicted however it will be affected by the second highest category if it is > 0.15 p and will appear darker
		Very Pale is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (tends to represent the darker second highest category)
	Pale	
	If highest probability > 0.9p	Pale predicted
	If highest probability > 0.7p	Pale is predicted however it will be affected by the second highest category if it is > 0.15 p (will appear darker if Intermediate, and lighter if pale)
		Pale is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (tends to represent the darker second highest category)
		Not significantly impacted if second highest prediction is Very Pale
	Int.	
	If highest probability > 0.9p	Intermediate predicted (*unless Dark-Black is the second highest category, then prediction is Dark)
	If highest probability > 0.7p	Intermediate is predicted however it will be affected by the second highest category if it is > 0.15 p (will appear darker if Dark/Dark-Black and lighter if Very Pale/Pale)
		Intermediate is predicted, unlikely to be affected by the second highest category if it is < 0.15 p (*unless it is Dark-Black, then prediction is Dark)
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (darker if Dark/Dark-Black represents the second highest category)
		Not significantly impacted if second highest prediction is Very Pale/Pale
	Dark	
	If highest probability > 0.9p	Dark predicted
	If highest probability > 0.7p	Dark is predicted, unlikely to be affected by the second highest category if it is > 0.15 p (*unless it is Dark-Black, then prediction can be Dark-Black)
		Dark is predicted, unlikely to be affected by the second highest category if it is < 0.15 p (*unless it is Dark-Black, then prediction can be Dark-Black)
	If highest probability > 0.5p	Prediction significantly affected by second category, and will be a mix of the two highest categories (dark to black if Dark-Black represents the second highest category)
		Not significantly impacted if second highest prediction is Pale/Intermediate
	Dark-Black	
	If highest probability > 0.9p	Dark to Black predicted
	If highest probability > 0.7p	Dark to Black is predicted, unlikely to be affected by the second highest category if it is > 0.15 p
		Dark to Black is predicted, unlikely to be affected by the second highest category if it is < 0.15 p
	If highest probability > 0.5p	Prediction affected by second category, and will be a mix of the two highest categories (Will be lighter than Dark to Black if Dark represents the second highest category)

Figure 3.10 - Image extracted from the HIrisPlex-S DNA Phenotyping Webtool User Manual Version 2.0 (2018) illustrating the decision model for skin colour inference.

3.6 HIrisPlex-S Protocol implemented with a Genotype likelihoods Model

As described, in the classical protocol, the HIrisPlex-S website allows a single phenotypic prediction per sample by taking as input an informative allele from a 1 genotype call for each of the 41 positions.

Two substantial modifications have been implemented in the method proposed in this thesis. The first affects the variant calling phase, where instead of exclusively adopting the most probable genotype proposed by GATK (despite using genotype likelihoods, *UnifiedGenotyper* selects a single genotype), all 10 potential genotypes are considered based on their respective probabilities. It is noteworthy that selecting the most probable genotype in high-coverage scenarios is generally believed a confident procedure, as the position coverage ensures a robust representation of nucleotides. However, in low coverage contexts, where positions may be covered only scarcely if at all, this approach carries higher risk (calling a genotype even with a posterior probability as low as 51%). This

approach ensures that multiple genotypes may be evaluated for the same position due to their genotype likelihood values, facilitating a more comprehensive analysis.

The second modification affects the number of predictions made for each sample. This transitions from 1 prediction per sample to 1000 predictions per sample, encompassing all possible genotype combinations across the 41 positions.

All the steps were performed on samples subjected to pointwise progressive downsampling. As previously detailed in paragraph 3.5, the 10 iterations were conducted from 8x coverage downwards, with the exception of the 5 initial iterations for Ust'-Ishim and SF12, which included higher coverage levels. This approach ensured consistency and accuracy in data handling across all samples involved in the study. The total number of phenotypic predictions made was 225,000.

3.6.1 Indel check

The indel discovery phase followed the classical protocol aforementioned in subparagraph 3.5.1.

3.6.2 Variant calling

Genotype likelihoods were calculated following the formula of the first version of *GATK* (dragon) (McKenna *et al.*, 2010):

$$Pr(D|G = \{A_1, A_2\}) = \prod_{i=1}^{M} Pr(b_i|G = \{A_1, A_2\}) = \prod_{i=1}^{M} \left(\frac{1}{2}Pr(b_i|A_1) + \frac{1}{2}Pr(b_i|A_2)\right)$$

$$Pr(b|A) = \begin{cases} \frac{e}{3} : b \neq A \\ 1 - e : b = A \end{cases}$$

This formula facilitates the comparison of the base read during sequencing (b_i), its respective depth of coverage (M) and the probability of sequencing error, calculated from the Phred scaled quality score $e = 10^{-q/10}$ 63. Given the assumption of diploid individuals, this entails consideration of 10 distinct genotypes, i.e. AA, AC, AG, AT, CC, CG, CT, GG, GT and TT.

In addition to computing the genotype likelihood, a prior probability for each genotype must be assumed to produce posterior probabilities for the genotypes (Nielsen *et al.*, 2011).

The prior genotype probability chosen was the same used by *GATK UnifiedGenotyper* v3.5, i.e. the expected probability that an individual is heterozygous at a given locus. For humans, a default heterozygosity value (hets) of 0.001 is provided (DePristo *et al.*, 2011).

The calculation of genotype likelihoods and genotype posteriors was executed by developing a pipeline in *RStudio* v2024.4.2.764 (RStudio Team, 2020), managing the data using the libraries *matrixStats* (Bengtsson *et al.*, 2017), *tibble* (Müller and Wickham, 2023), *tidyr* (Wickham *et al.*, 2024) and *dplyr* (Wickham *et al.*, 2023).

All the steps were performed starting from the *SAMtools* mpileup file imported in *RStudio* environment and all the calculations were achieved using logarithmic values to avoid machine underflow.

To summarize the pipeline:

1. **Structuring of the SAMtools mpileup file**:
   - Data cleaning: non-coding symbols were removed (starts "^" and ends "$" of reads) and forward (.) and reverse (,) symbols were replaced with nucleotides.
   - Quality base filter: quality scores of bases were converted from ASCII to numeric values and bases with quality scores below 30 were filtered out.
   - Depth recalculation
   - Error calculation: ASCII quality scores were converted to numeric values and the logarithmic probability of sequencing error was calculated as $e = 10^{-q/10}$.
2. **Computation of Genotype Likelihoods**:
   - Function definitions: functions for calculating genotype likelihoods in logarithmic values was defined following the formula of the first version of *GATK* (dragon).
   - Genotype Likelihoods Calculation: a vector of possible genotypes was created. For each position in the mpileup file, logarithmic likelihoods for each possible genotype was calculated and stored in a list.
   - Logarithmic Genotype likelihoods were normalized and checked by summing to 0.
3. **Computation of Genotype Posteriors:**
   - Prior definitions: logarithmic prior values for homozygous (0.999) and heterozygous (0.001) genotypes were set.
   - Genotype Posterior Probabilities calculation: logarithmic genotype likelihoods were combined with the defined priors.
   - Logarithmic Genotype Posterior Probabilities were normalized and transformed to exponential values. Normalisation was verified by summing to 1.

At the end of the pipeline, 10 possible genotypes were maintained for each of the 41 positions within the HIrisPlex-S.

These genotypes must adhere to the requirements stipulated by the HIrisPlex-S website, which specify that each input should consist of a single allele derived from a genotype.

To meet this requirement, genotypes were sampled using the "*sample*" function, considering their genotype posterior probabilities. The likelihood of selecting a particular genotype for sampling thus depended on the strength of support it received from the sequencing data, with genotypes possessing higher posterior probabilities being more likely to be sampled.

Each sampling event covered all 41 positions and was repeated 1000 times.

3.6.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes

Following the sampling process, an additional step was introduced in *RStudio* v2024.4.2.764 (RStudio Team, 2020) to expedite the recognition of informative alleles. This step involved the incorporation of a CSV file containing HIrisPlex-S allele inputs adjusted for strand orientation. Subsequently, a comparison was made between the sampled genotype and the presence of 0, 1 or 2 informative alleles at each position. The final result consists of a table containing 1000 rows, each representing a set of 41 allelic combinations for the website (Table 3.3).

This table was saved in a CSV file, formatted for compatibility with loading into the HIrisPlex website as specified in the HIrisPlex-S Webtool User Manual Version 2.0 (2018) (Walsh *et al.*, 2013; Walsh *et al.*, 2014; Chaitanya *et al.*, 2018).

sam- pleid,rs312262906_A,rs11547464_A,rs885479_T,rs1805008_T,rs1805005_T,rs1805006_A,rs1805007_T,rs1805009_C,rs201326893_A,rs2228479_A,rs1110400_C,rs28777_C,rs16891982_C,rs12821256_G,rs4959270_A,rs12203592_T,rs1042602_T,rs1800407_A,rs2402130_G,rs12913832_T,rs2378249_C,rs12896399_T,rs1393350_T,rs683_G,rs3114908_T,rs1800414_C,rs10756819_G,rs2238289_C,rs17128291_C,rs6497292_C,rs1129038_G,rs1667394_C,rs1126809_A,rs1470608_A,rs1426654_G,rs6119471_C,rs1545397_T,rs6059655_T,rs12441727_A,rs3212355_A,rs8051733_C
UstIshim_1x_Sampling1,0,0,0,0,0,0,0,0,0,0,0,2,2,0,2,0,0,0,0,2,0,2,0,2,0,0,2,2,0,2,2,2,0,2,2,0,2,0,0,0,0
UstIshim_1x_ Sampling2,0,0,0,0,0,0,0,0,0,0,0,2,2,0,2,0,0,0,0,2,0,2,0,2,0,0,2,2,0,2,2,2,0,2,2,0,2,0,0,0,0
UstIshim_1x_ Sampling3,0,0,0,0,0,0,0,0,0,0,0,2,2,0,2,0,0,0,0,2,0,2,0,2,0,0,2,2,0,2,2,2,0,2,2,0,2,0,0,0,0
…
UstIshim_1x_ Sampling1000,0,0,0,0,0,0,0,0,0,0,0,2,2,0,2,0,0,0,0,2,0,2,0,2,0,0,2,2,0,2,2,2,0,2,2,0,2,0,0,0,0

Table 3.3 - Example extracted from the uploaded CSV of the Ust'-Ishim sample, downsampling 1 at coverage 1x. The first row includes the CSV file header containing the sample ID and Rs IDs of informative alleles for HIrisPlex-S. Rows 2 to end present a set of 41 sampled alleles.

Phenotypic prediction followed the classical protocol aforementioned in subparagraph 3.5.3.

3.7 Imputation workflow

Low-coverage imputation is increasingly prevalent in aDNA studies. However, the literature shows that imputation pipelines typically utilized for contemporary genomes may be inaccurate when applied to older samples, particularly concerning rare variants (Hui *et al.*, 2020; Sousa da Mota *et al.*, 2023; Garrido Marques *et al.*, 2024), such as the A allele at rs312262906 (MAF = 0,01) or rs11547464 (MAF < 0.01) associated with the red hair colour phenotype (Branicki *et al.*, 2011; https://www.ncbi.nlm.nih.gov/snp/).

In the light of this, the golden standard imputation pipeline for aDNA released in June 2023 by Bárbara Sousa da Mota and colleagues and published in Nature Communications was tested to observe how phenotypic predictions vary with changes in average coverage and the presence of missing variants.

All the steps were performed on samples subjected to mean progressive downsampling. As explained earlier in paragraph 3.5, the 10 repetitions were carried out starting from 8x coverage downward, with the 5 initial repetitions for Ust'-Ishim and SF12 including higher coverage levels. This methodological approach was implemented to maintain robustness and consistency throughout the analysis of the samples. The total number of phenotypic predictions made was 225.

3.7.1 Indel check

The indel discovery phase followed the classical protocol aforementioned in subparagraph 3.5.1.

3.7.2 Variant calling

Following the guidelines of Sousa da Mota, the reference panel selected was the 1000 Genomes phase 3 v5, comprising 3202 samples re-sequenced at 30x (Byrska-Bishop *et al.*, 2022; Sousa da Mota *et al.*, 2023). Data were downloaded in VCF format from NCBI (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/) and checked for HIrisPlex-S. Two positions in chromosome 16 were missing: 89985753 (rs312262906) and 89986122 (rs201326893). The reference panel was lifted over from build GRCh38 to GRCh37 reference genome assembly using *Picard liftoverVCF* v2.24.1 (https://broadinstitute.github.io/picard/) with the hg38ToHg19 chain from the University of California, Santa Cruz liftOver tool (http://hgdownload.cse.ucsc.edu/goldenpath/hg38/liftOver/). *BCFtools* v1.11 software (Li, 2011) was used with the

options "*-S*" and "*-m 2 -M 2 -v snps*" to remove related individuals and multiallelic records, resulting in approximately 90 million SNPs and 2504 samples remaining.

The first phase of imputation involved generating genotype likelihoods and calls, which were calculated using *BCFtools* v1.11 261 with the command *bcftools mpileup* and the parameters "*-I*" (output variants in the form FORMAT/GT), "*-E*" (uniform error probability), "*-a*" *(*print all sites in *FORMAT/DP*) "*-ignore-RG*" (treat all reads in one alignment file as one sample). Subsequently, *bcftools call* was executed with the options "*-Aim*" (ignore reference genome, output only sites with indels, merge multiple VCF/BCF files into a single output file) and "*-C alleles*" (call genotypes given alleles). Mapping (*-Q*) and base quality (*-q*) filters were set to 30.

Before proceeding, HIrisPlex-S positions were removed from all the mean progressive downsampled genomes.

The imputation of the downsampled genomes was performed using *GLIMPSE* v1.1.1 (Rubinacci *et al.*, 2021). Chromosomes were split into chunks of sizes in the range 1–2 Mb using *GLIMPSE_chunk*, with a 200-kb buffer region at each side of a chunk. Imputation was then carried out with *GLIMPSE_phase* using parameters "*--burn 10*" (number of burn-in iterations for sampling a new pair of haplotypes per target individual by conditioning on haplotypes from the reference panel), "--main 15" (number of main iterations that contributes to output genotypes) and "*--pbwt-depth2*" (model state space to store reference and target haplotypes as a positional Burrows–Wheeler transform). Finally, the imputed chunks were ligated using *GLIMPSE_ligate* and the most likely haplotype pair for each sample was determined using the "*--solve*" option in *GLIMPSE_sample*.

The final output was a VCF file reporting all the called and phased genotypes in one chromosome. The HIrisPlex-S genotypes were then extracted for the phenotypic predictions.

3.7.3 Phenotypic Prediction and Interpretation of the Prediction Outcomes

Phenotypic prediction followed the classical protocol aforementioned in subparagraph 3.5.3.

3.8 User Manual and Dataset Application

Upon conclusion of the testing phase on the three approaches, the most robust method for performing phenotypic inference on ancient and low-coverage data was identified.

The proposed guidelines recommend applying the classical HIrisPlex-S protocol exclusively when a minimum coverage of 8x is guaranteed, while the newly proposed protocol utilizing the Genotype Likelihoods Model is advised for lower coverage. Following these guidelines, phenotypic inferences for eye colour, hair colour and skin tone were conducted on the samples in the dataset described in paragraph 3.2.

To optimize the prediction protocol, the pipeline was implemented with Bash scripts designed to monitor the sample coverage status. For each sample, a *SAMtools* mpileup file was generated for the 41 HIrisPlex-S positions. Checks were conducted using the mpileup for each sample at these positions, ensuring minimum coverage, identifying missing data, assessing the absence of critical positions impacting predictions (e.g., *HERC2*, *SLC45A2*, *IRF4* and *MC1R* positions) and detecting in-dels. After verifying the sample status, if the guaranteed minimum coverage at each position was at least 8x, direct calling was performed; otherwise, the new prediction protocol utilizing genotype likelihood support was applied.

4 RESULTS E DISCUSSION

The initial phase of this project involved the evaluation of three phenotypic prediction methods to provide an optimized protocol of inference of phenotypic traits based on the specific coverage of the HIrisPlex-S positions. The methods assessed were: the HIrisPlex-S Standard Protocol, the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data. Each method was tested on three different samples (Ust'-Ishim, SF12, and I5835) from three distinct historical periods (Palaeolithic, Mesolithic, and Bronze Age), with each sample subjected to 10 distinct downsampling procedures.

Given the limitation of lacking observed information on phenotypic traits of these samples, it was considered "true" phenotypic traits those inferred by applying the Standard HIrisPlex-S Protocol to high-coverage data. Consequently, it was determined that the Ust'-Ishim sample exhibited brown eyes, black hair, and dark to black skin; the SF12 sample had blue eyes, brown to dark brown hair, and dark skin; and the I5835 sample displayed brown eyes, dark brown to black hair, and an intermediate skin tone.

Throughout this thesis, eye, hair, and skin colour phenotypic inference results are presented using bar graphs. Each graph displays the variation in coverage on the X-axis and the percentage of prediction for each phenotypic class on the Y-axis, with percentages calculated over 10 repetitions of each downsampling. Each graph represents a predictive model, and for the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model, two graphs are presented: the first illustrates all obtained predictions, while the second applies a 90% threshold. Specifically, if a phenotypic class is predicted in at least 900 out of 1000 cases, it is considered the sole predicted phenotype, and the values for this class are reported as if they were predicted in all 1000 cases. If the threshold is not met, predictions for all classes are reported. Following these, graphs depict the percentages of correct and incorrect predictions, based on the phenotype considered as "true" from high-coverage data.

At the end of this part of the project, it should be possible to identify which procedure is more robust to predict phenotypic traits for each coverage level.

In the second part of the project these guidelines of phenotypic inference were applied to a dataset comprising 348 samples, including the three previously tested, spanning from the Palaeolithic period to the Iron Age. Results are presented using maps which illustrate the geographical distribution of

samples and pie charts collocated on sample collection sites. Each pie chart represents the percentage of predicted phenotypes and their size corresponds to the number of samples used for the phenotypic inference.

The results, along with the corresponding discussions, are organized according to these two main points: first the testing phase and then the application of the guidelines on the dataset. For clarity, results are further divided by historical period (Palaeolithic, Mesolithic, Neolithic, Copper Age, Bronze Age, and Iron Age) and predicted phenotypic trait (eye colour, hair colour, and skin colour).

4.1 Ust'-Ishim sample

For the Palaeolithic sample Ust'-Ishim (with an average coverage of 28.5x), the maximum coverage at the HIrisPlex-S positions, from which downsampling initiated, was 17x.

The results of each specific downsampling are provided in the folder at the following link: Ust'-Ishim.

4.1.1 Eye colour prediction

The results for eye colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.1, 4.2, 4.3, and 4.4). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered "true", namely brown eye (Figures 4.5, 4.6, 4.7, and 4.8).

- HIrisPlex-S Standard Protocol

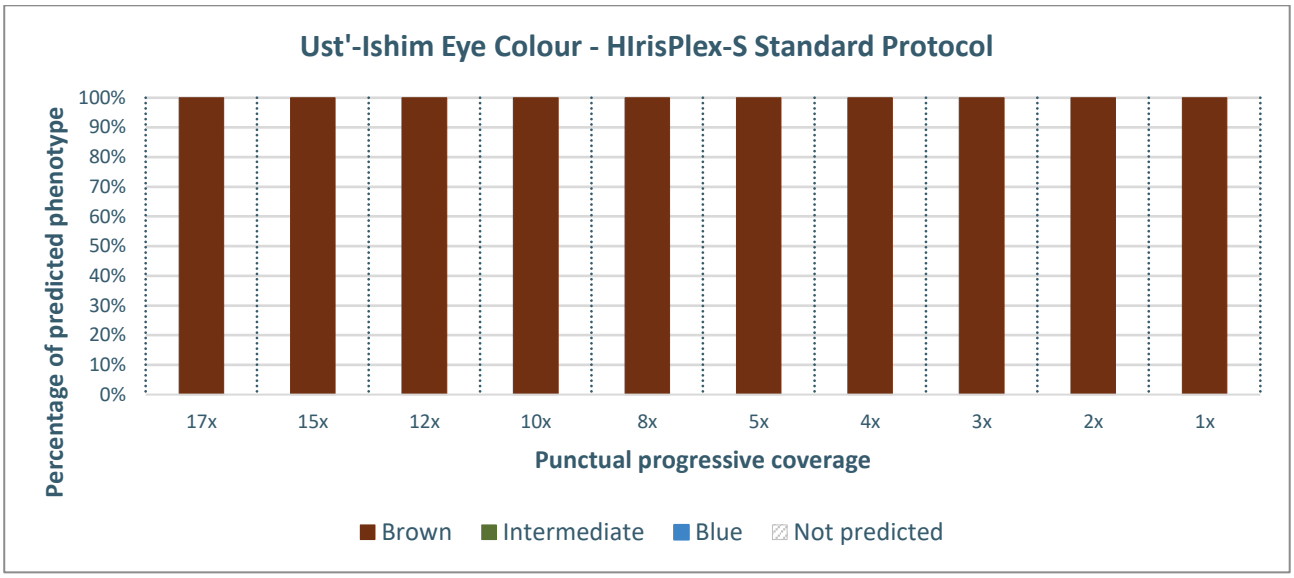


Figure 4.1 - Percentage distribution of predicted eye colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

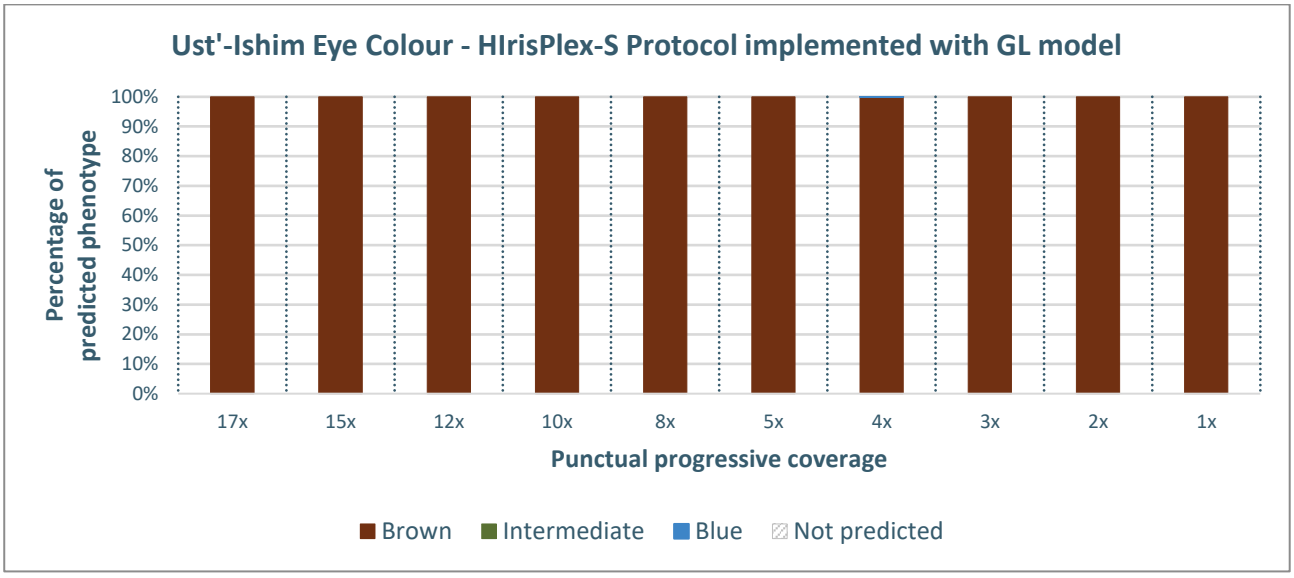


Figure 4.2 - Percentage distribution of predicted eye colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

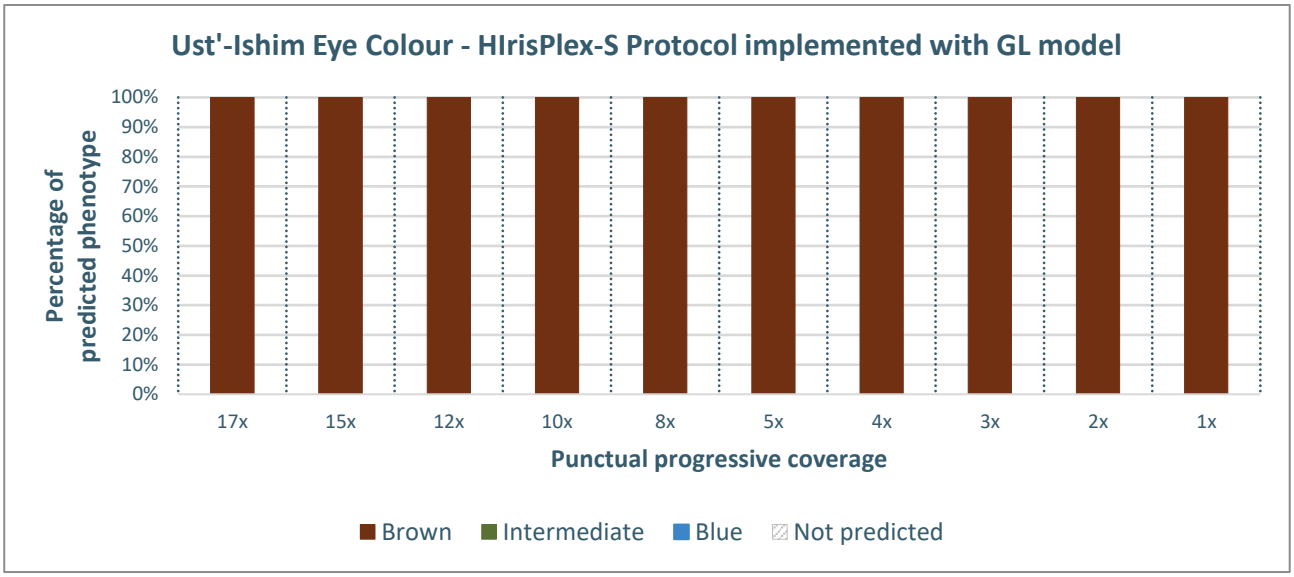


Figure 4.3 - Percentage distribution of predicted eye colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

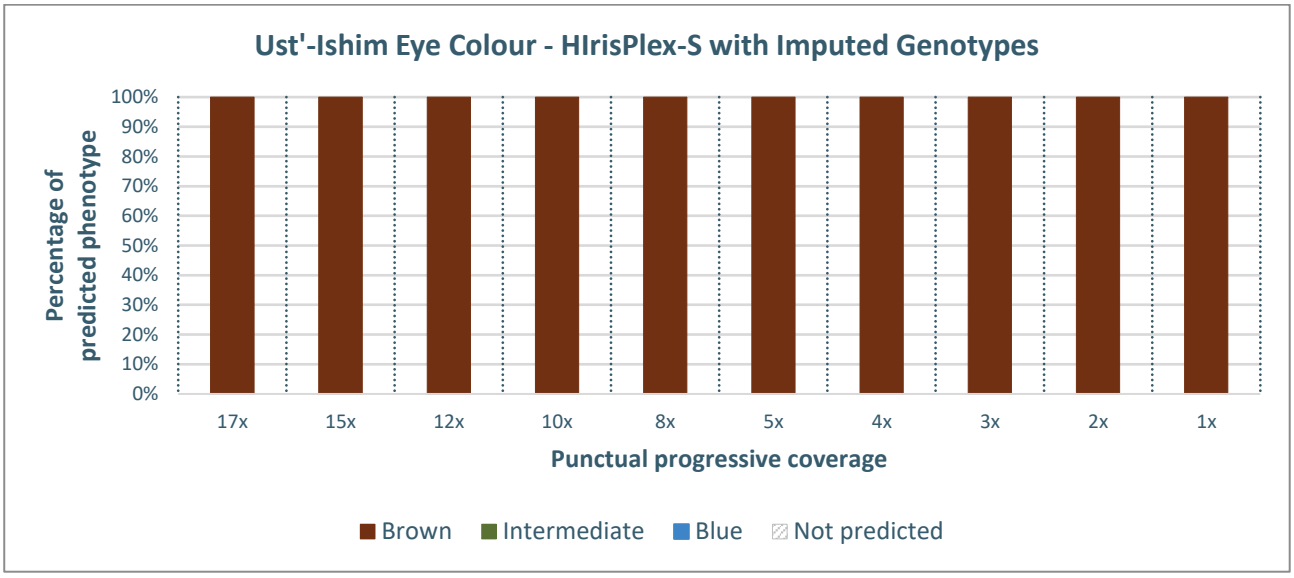


Figure 4.4 - Percentage distribution of predicted eye colour phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figures 4.5, 4.6, 4.7, and 4.8). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

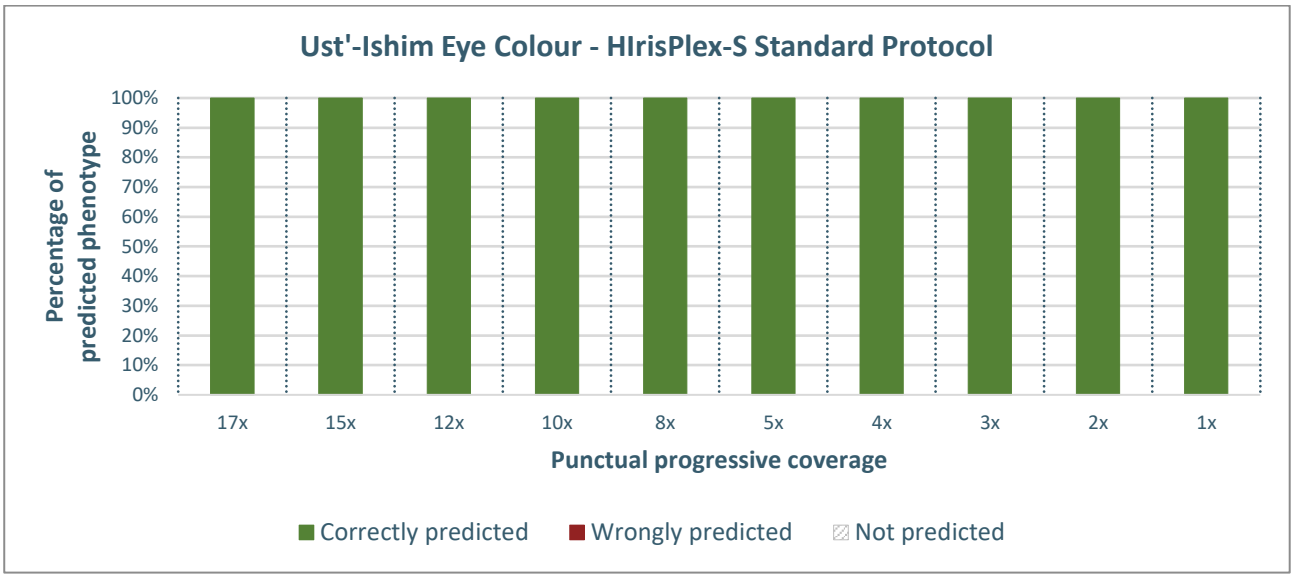


Figure 4.5 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

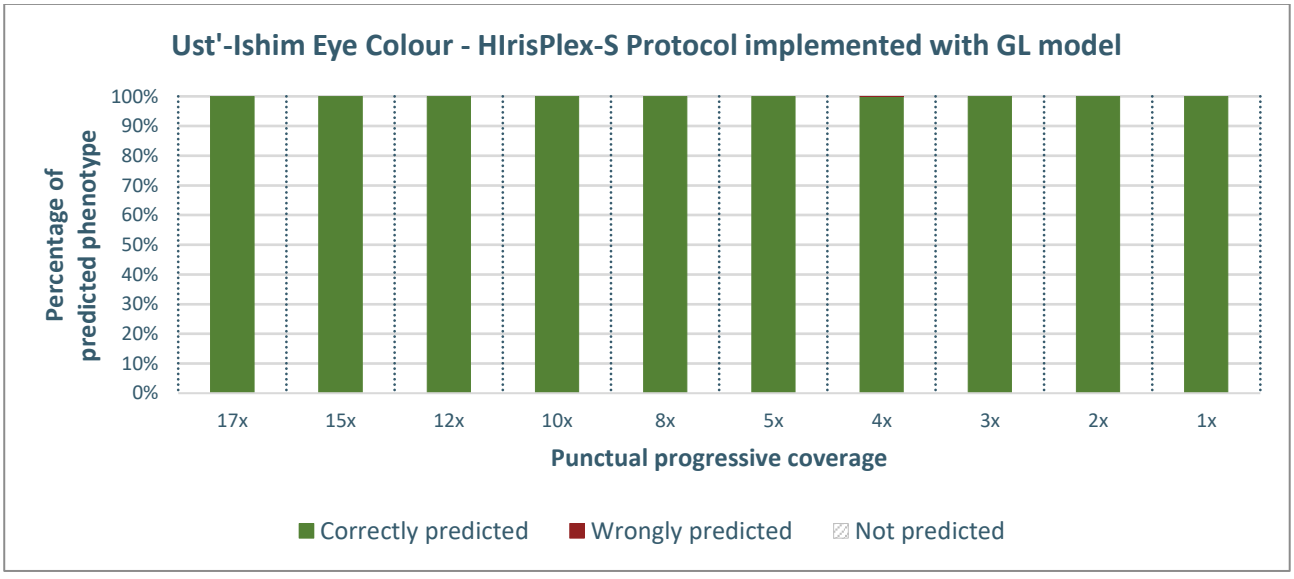


Figure 4.6 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

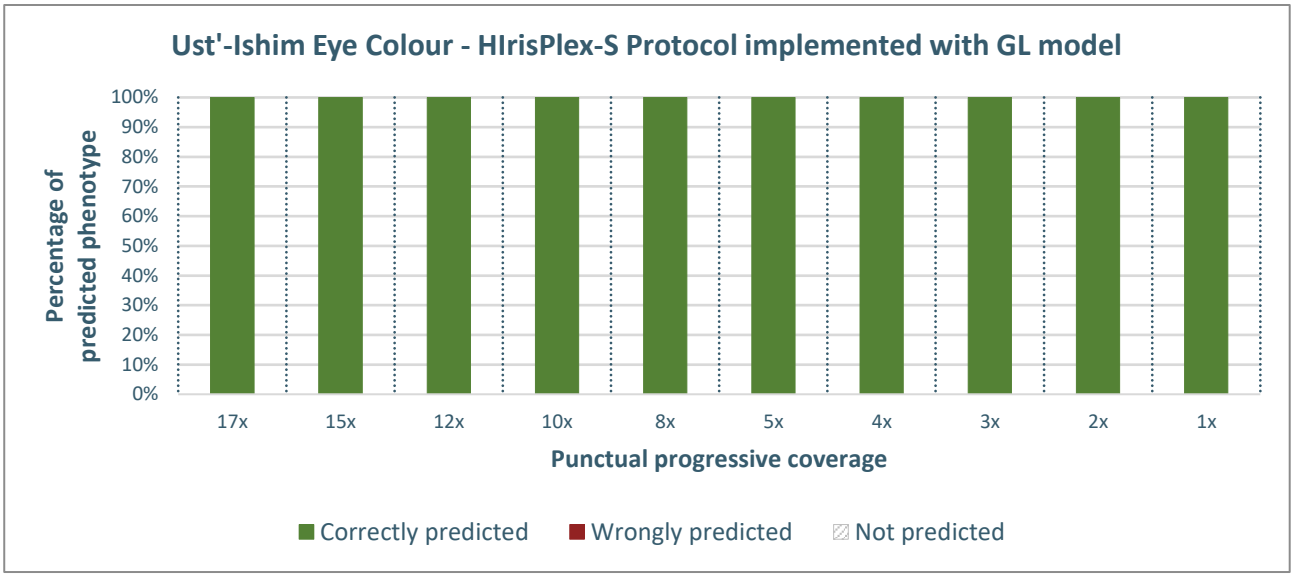


Figure 4.7 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Geno-type Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

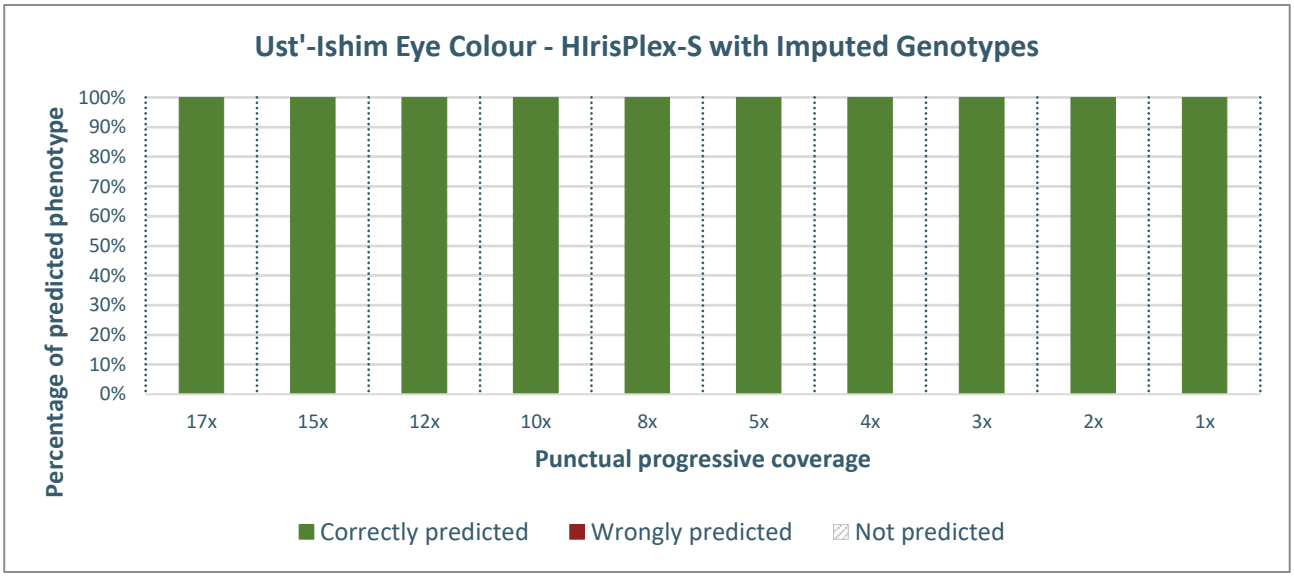


Figure 4.8 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.1.2 Hair colour prediction

The results for hair colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.9, 4.10, 4.11, and 4.12). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely black hair (Figures 4.13, 4.14, 4.15, and 4.16).

- HIrisPlex-S Standard Protocol

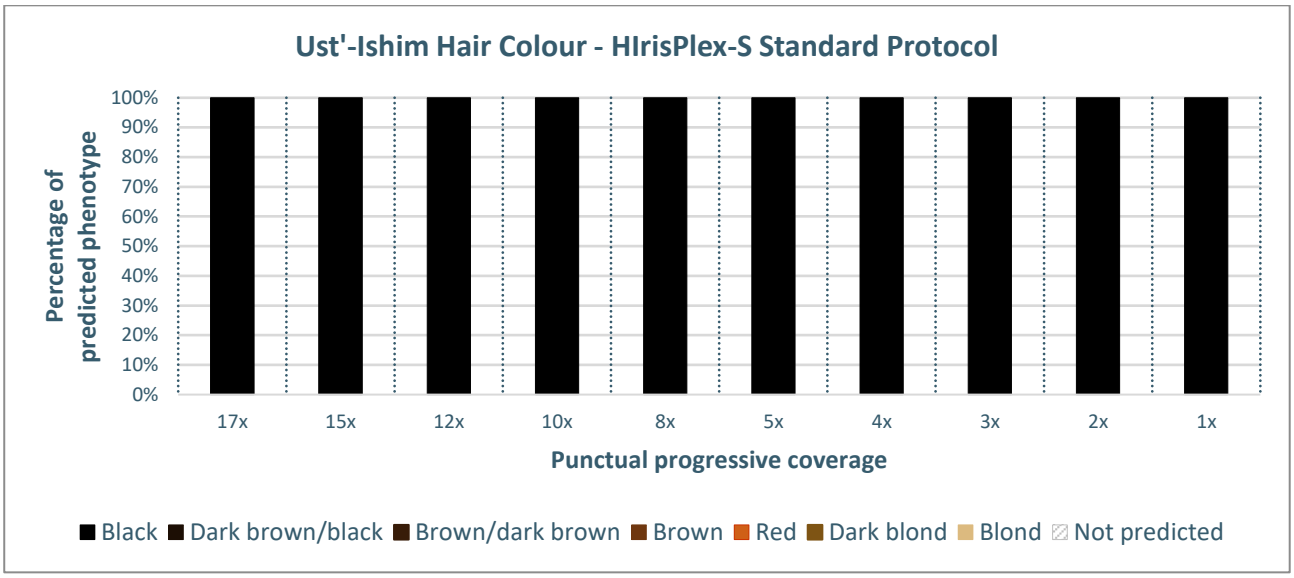


Figure 4.9 - Percentage distribution of predicted hair colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

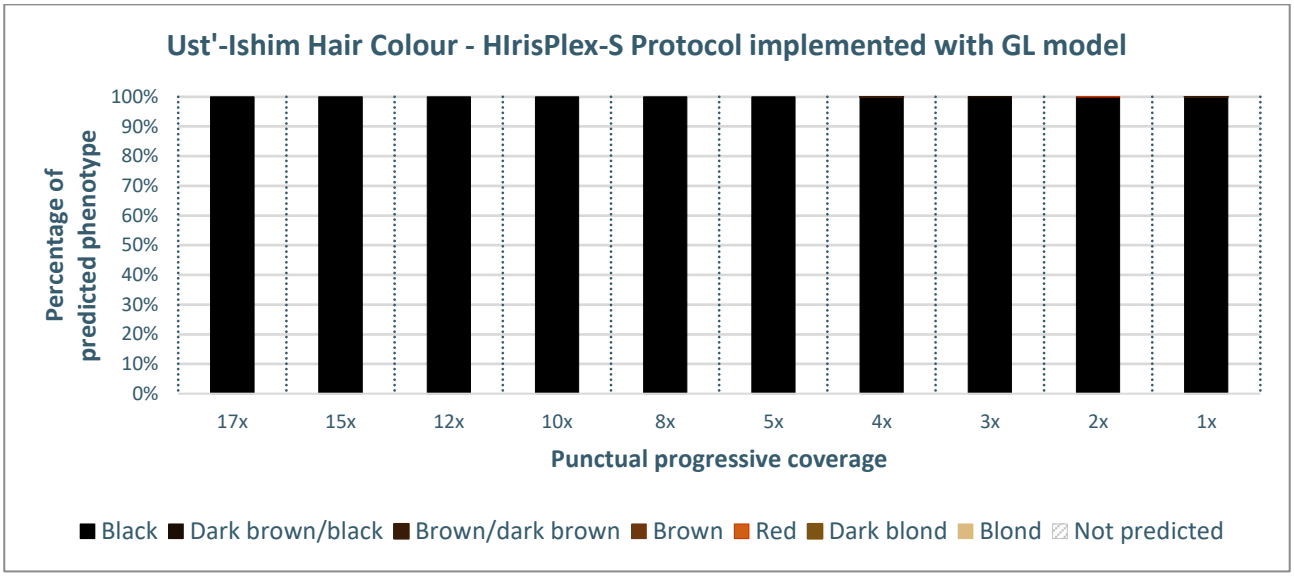


Figure 4.10 - Percentage distribution of predicted hair colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

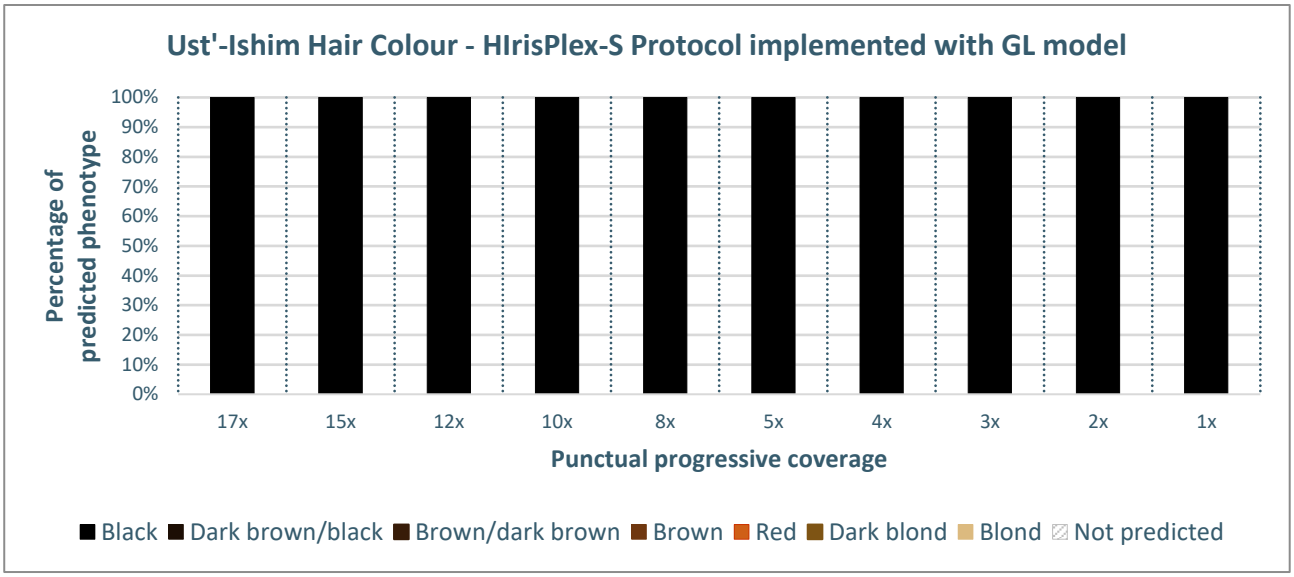


Figure 4.11 - Percentage distribution of predicted hair colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

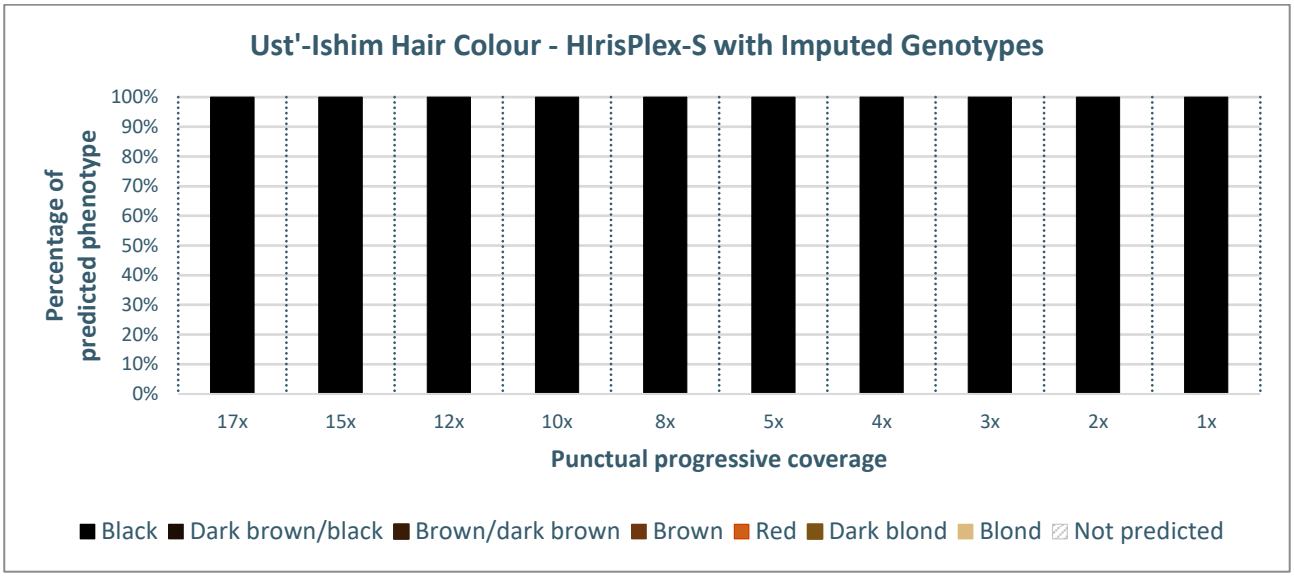


Figure 4.12 - Percentage distribution of predicted hair colour phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions, based on the presumed true phenotype, derived from high-coverage data (Figures 4.13, 4.14, 4.15, and 4.16). The graphs are presented in the order: HIrisPlex-S Standard Protocol, HIrisPlex-S Protocol

Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

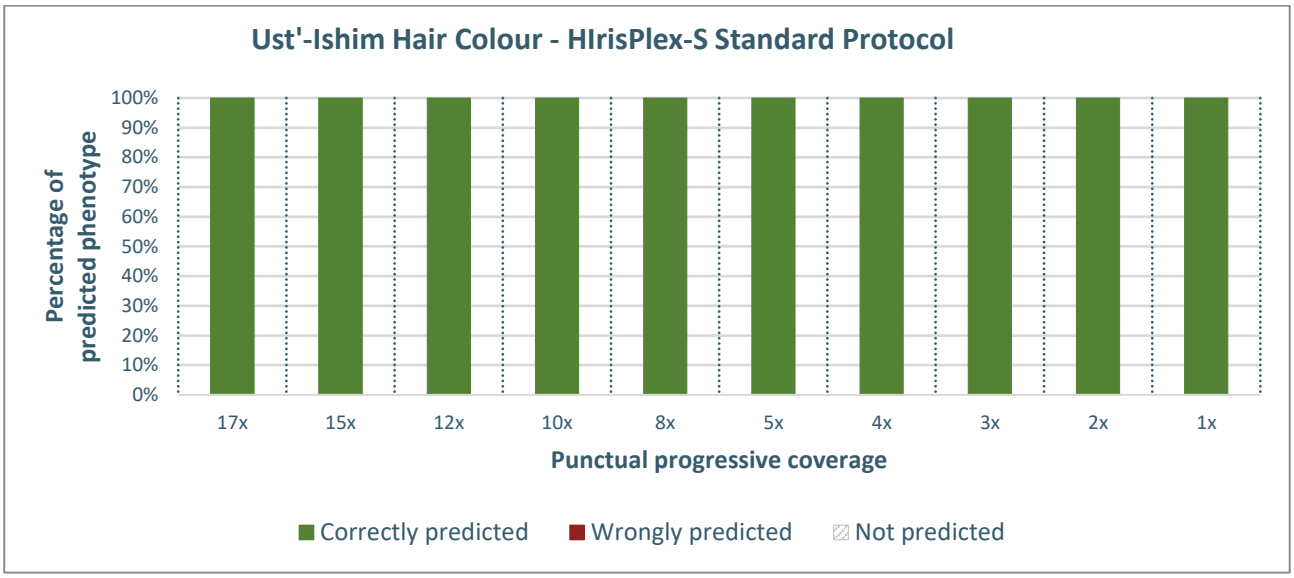


Figure 4.13 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

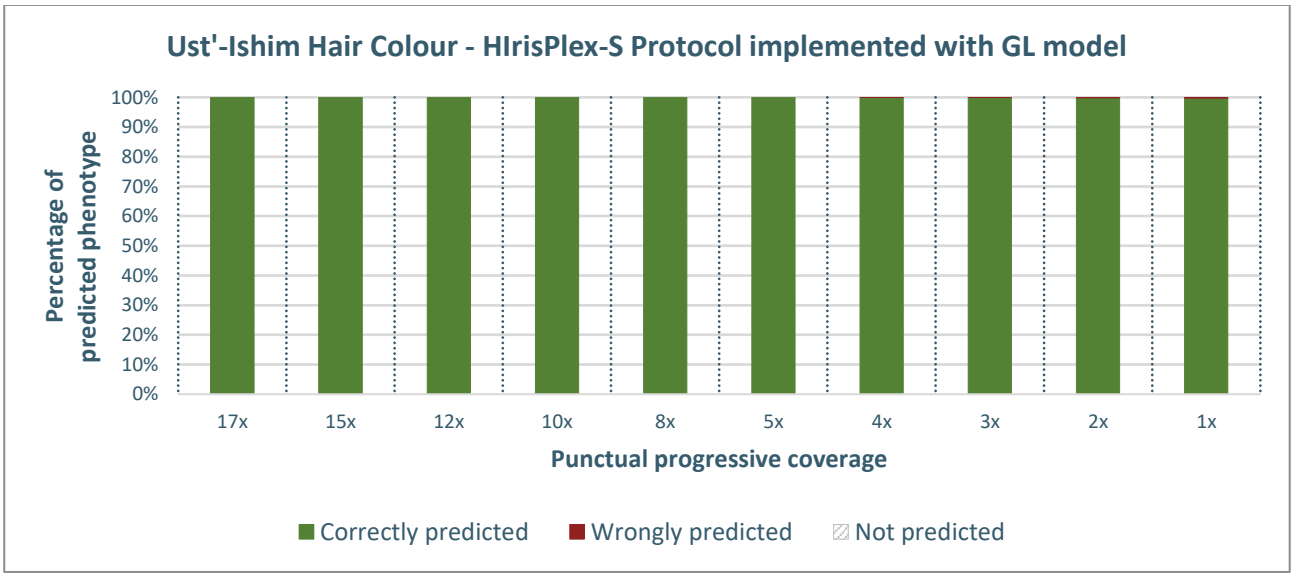


Figure 4.14 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

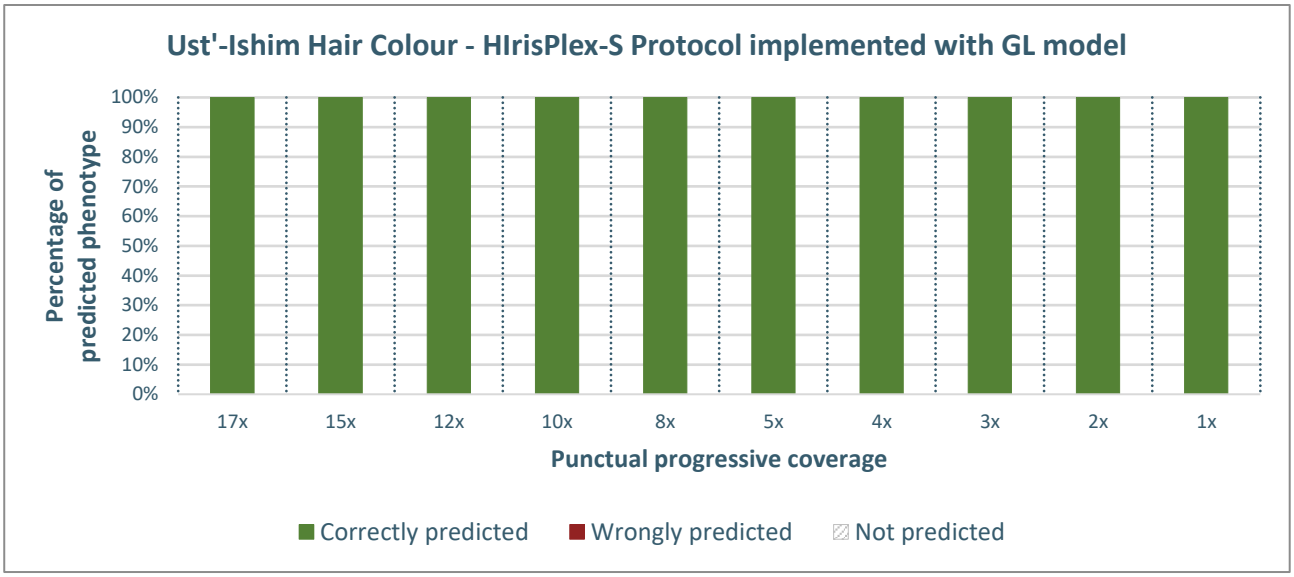


Figure 4.15 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

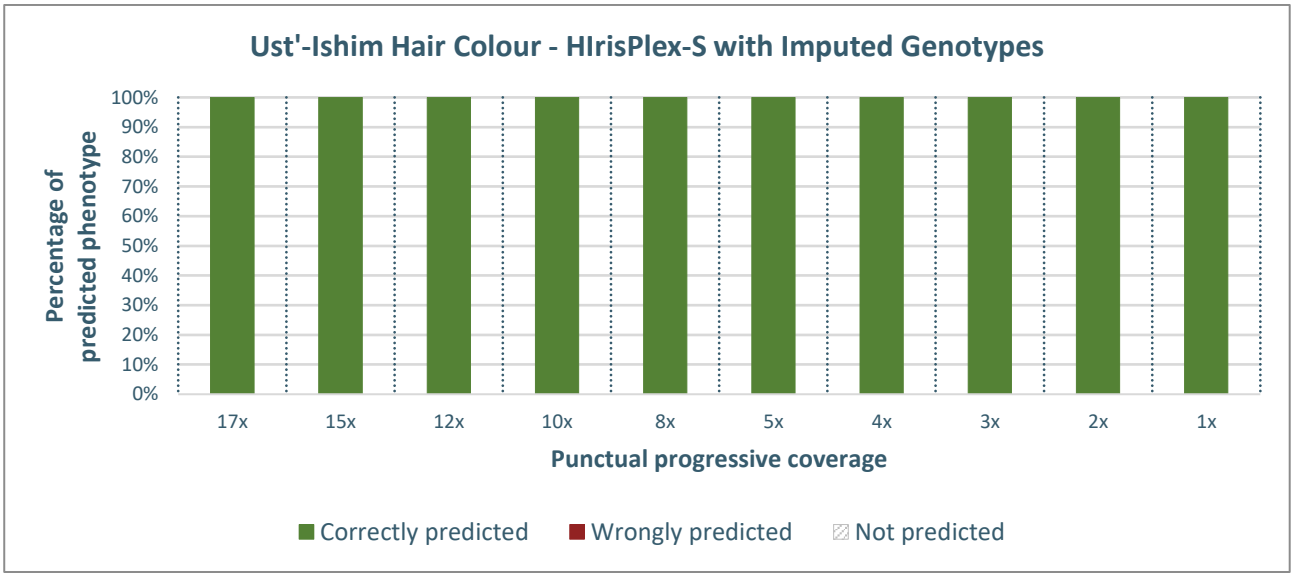


Figure 4.16 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.1.3 Skin colour prediction

The results for skin colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.17, 4.18, 4.19, and 4.20). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely dark to black skin (Figures 4.21, 4.22, 4.23, and 4.24).

- HIrisPlex-S Standard Protocol

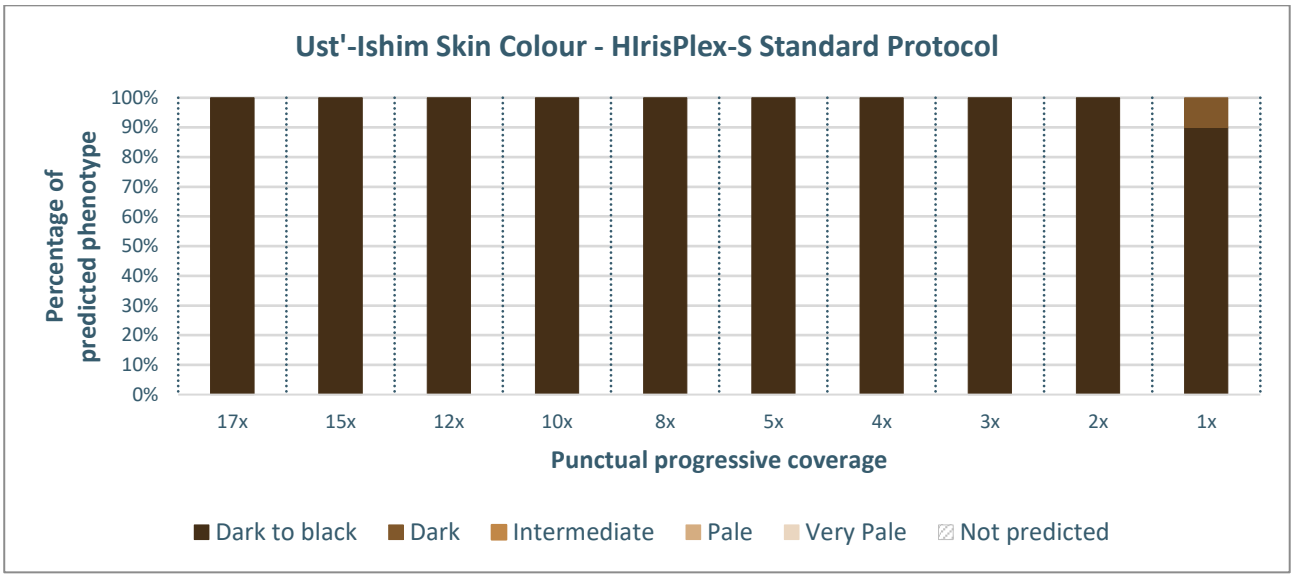


Figure 4.17 - Percentage distribution of predicted skin colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

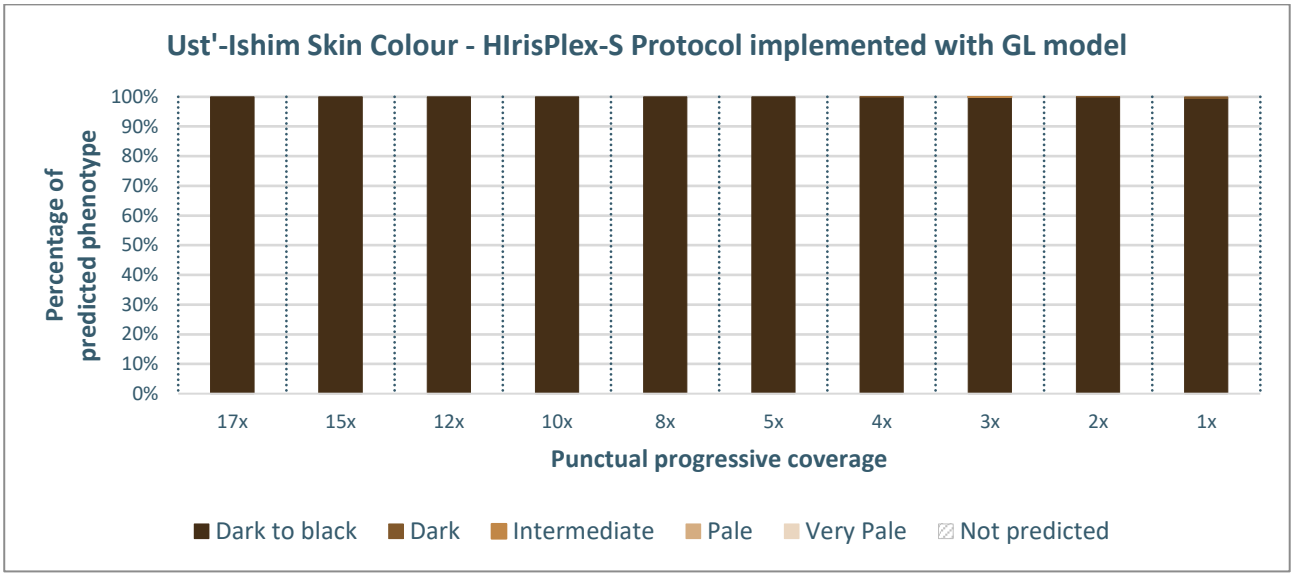


Figure 4.18 - Percentage distribution of predicted skin colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

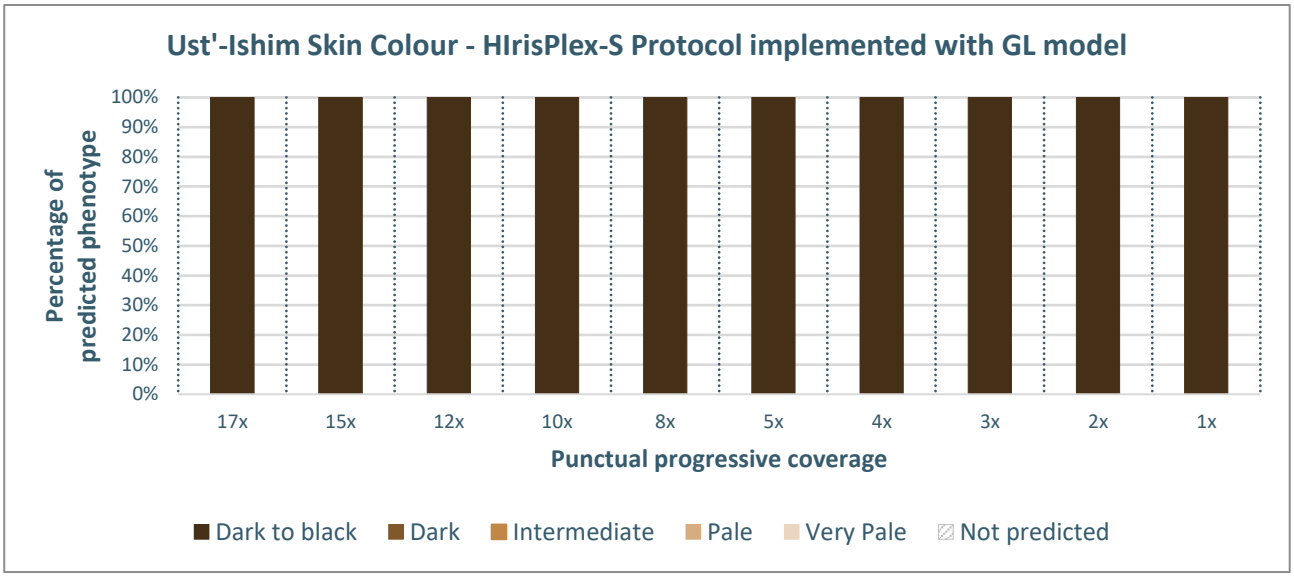


Figure 4.19 - Percentage distribution of predicted skin colour phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

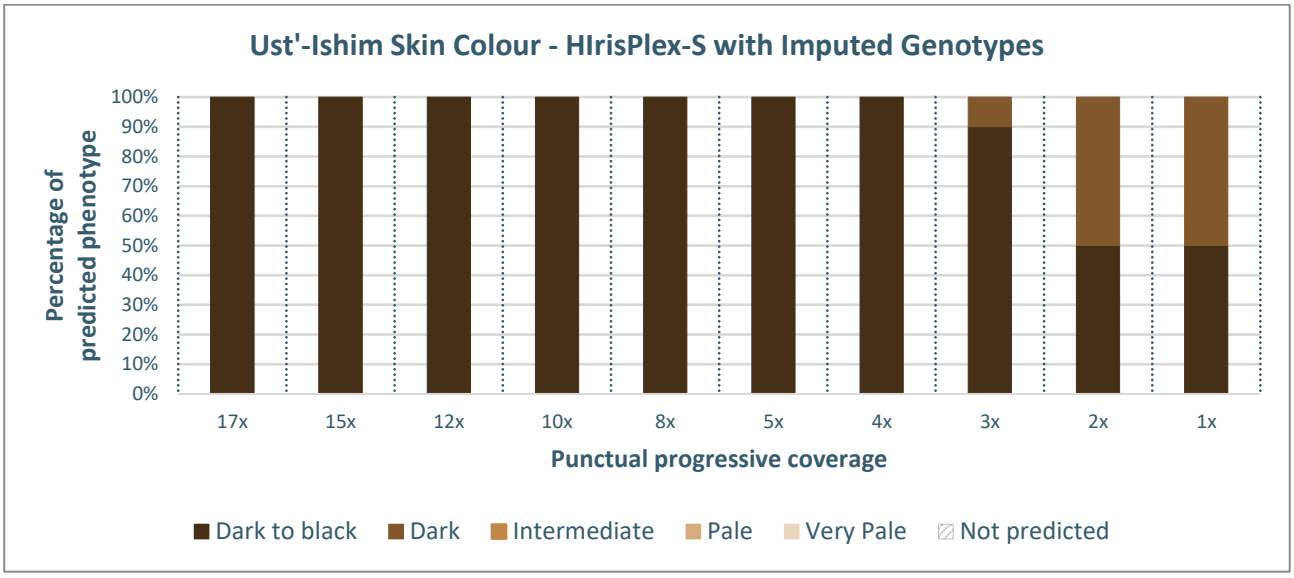


Figure 4.20 - Percentage distribution of predicted skin colour phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions, based on the presumed true phenotype, derived from high-coverage data (Figures 4.21, 4.22, 4.23, and 4.24). The graphs are presented in the order: HIrisPlex-S Standard Protocol, HIrisPlex-S Protocol

Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

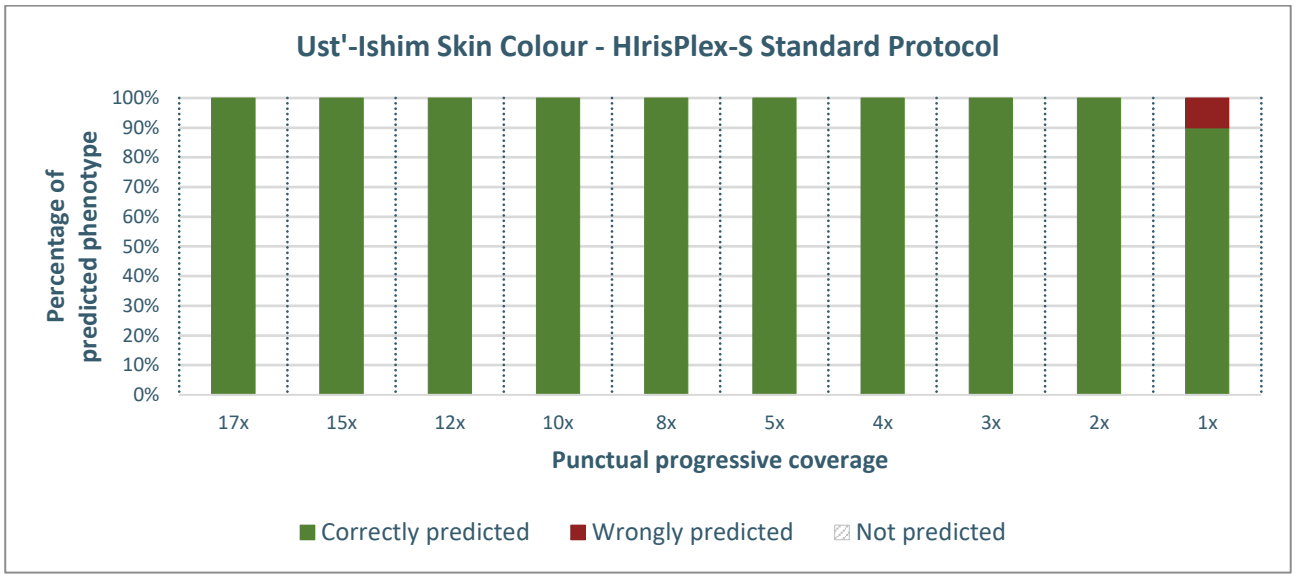


Figure 4.21 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

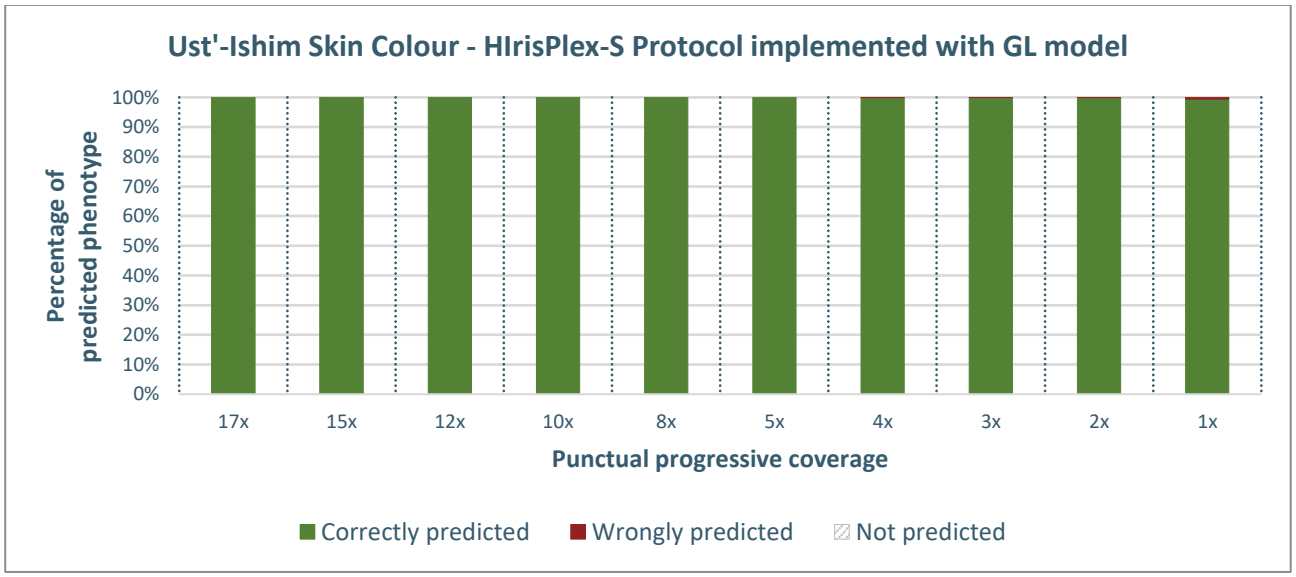


Figure 4.22 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

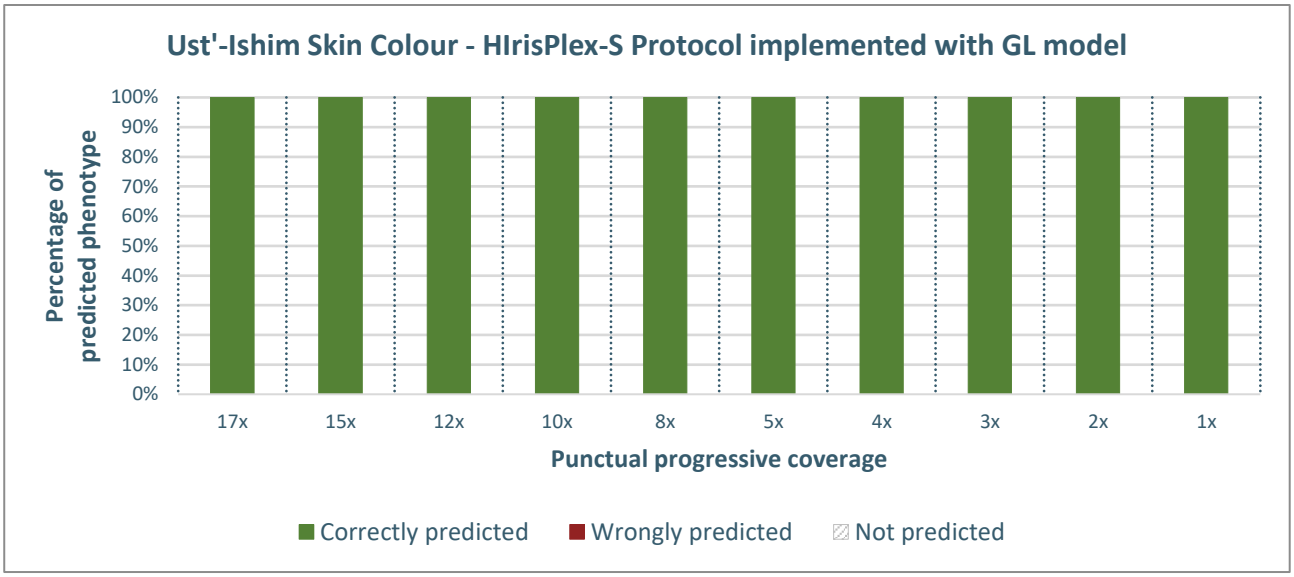


Figure 4.23 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

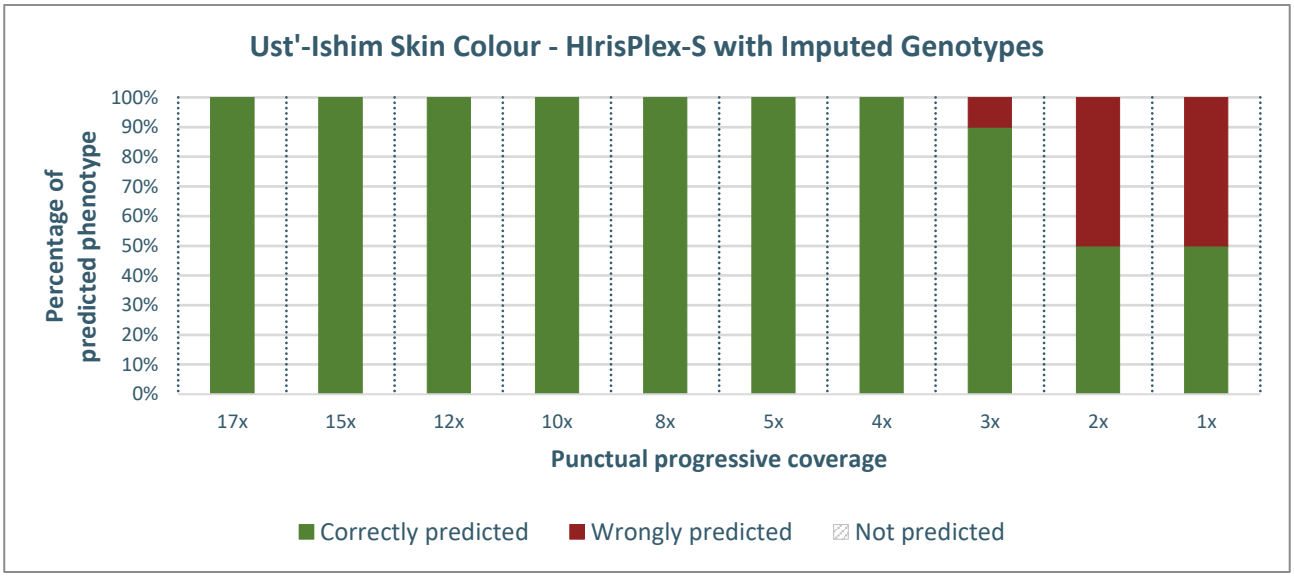


Figure 4.24 - Percentage distribution of predicted phenotypes for the Ust'-Ishim sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

The results obtained for the Palaeolithic sample Ust'-Ishim demonstrate robust phenotypic prediction for both eye colour (Figures 4.1, and 4.4) and hair colour (Figures 4.9, and 4.12) using both the HIrisPlex-S Standard Protocol and imputed genotypes. Regardless of the coverage level, predictions for brown eye colour and black hair colour remain consistent.

For the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model, the error rate for eye colour prediction is less than 0.2% at 4x coverage (Figure 4.2). For hair colour inference, the error rate is less than 0.3% across all coverage levels from 4x to 1x (Figure 4.10). Applying the 90% threshold (Figure 4.3), stochastic associations among genotypes with lower posterior probabilities, which could otherwise lead to erroneous inferences, are filtered out.

For skin colour inference, the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model demonstrates greater robustness compared to other methods, with an error rate for phenotypic predictions below 0.55% (Figures 4.18). This error rate vanishes once the 90% threshold is applied (Figures 4.19). Specifically, the HIrisPlex-S Standard Protocol, at 1x coverage, infers an incorrect phenotype in 10% of cases (Figure 4.17). In contrast, the imputation-based method fails in 50% of cases at 1x and 2x coverage, and in 10% of cases at 3x coverage (Figure 4.20).

4.2 SF12 sample

For the Mesolithic sample SF12 (with an average coverage of 43.9x), the maximum coverage at the HIrisPlex-S positions, from which downsampling initiated, was 33x.

The results of each specific downsampling are provided in the folder at the following link: SF12.

4.2.1 Eye colour prediction

The results for eye colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of predictions for each phenotypic class across 10 repetitions of downsampling (Figures 4.25, 4.26, 4.27, and 4.28). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered "true", namely blue eye (Figures 4.29, 4.30, 4.31, and 4.32).

- HIrisPlex-S Standard Protocol

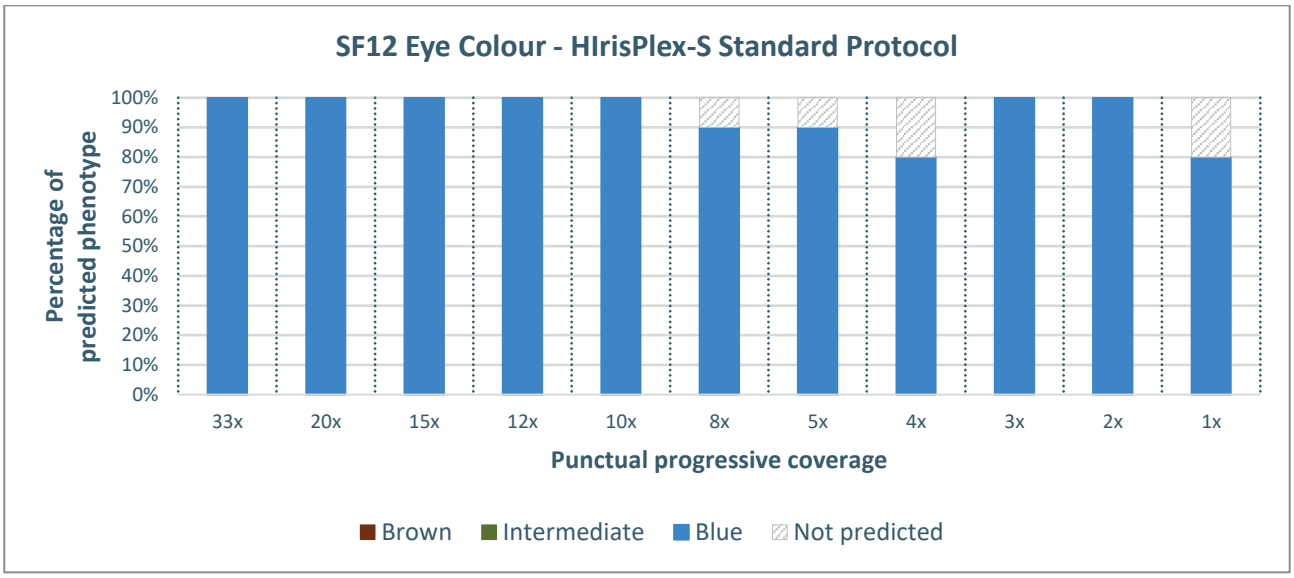


Figure 4.25 - Percentage distribution of predicted eye colour phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

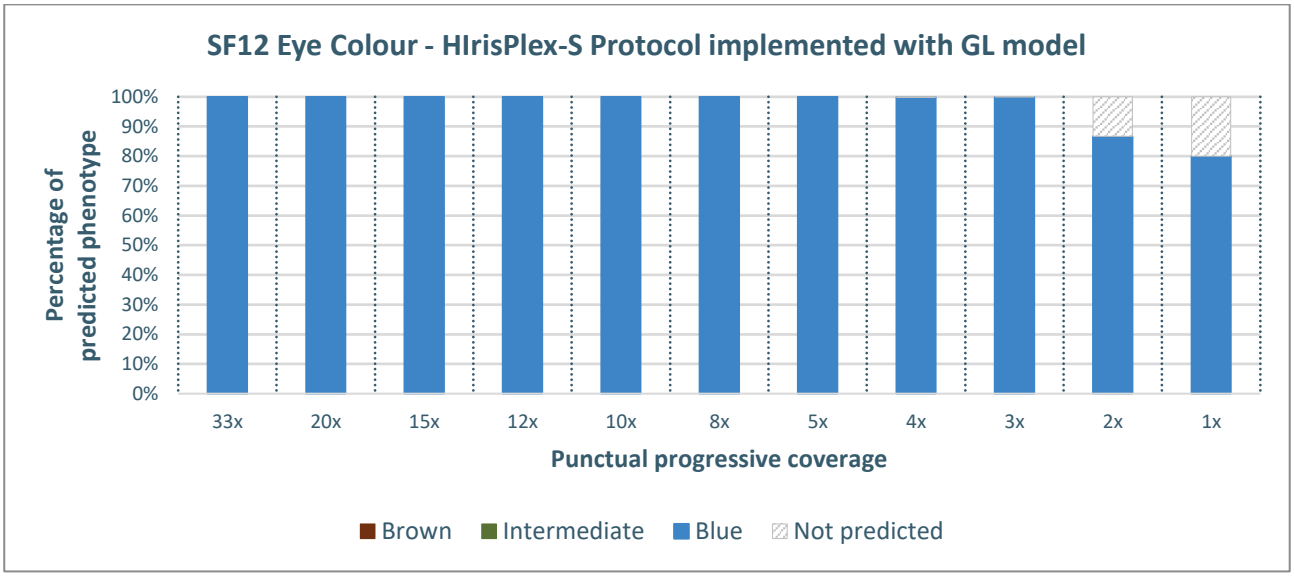


Figure 4.26 - Percentage distribution of predicted eye colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

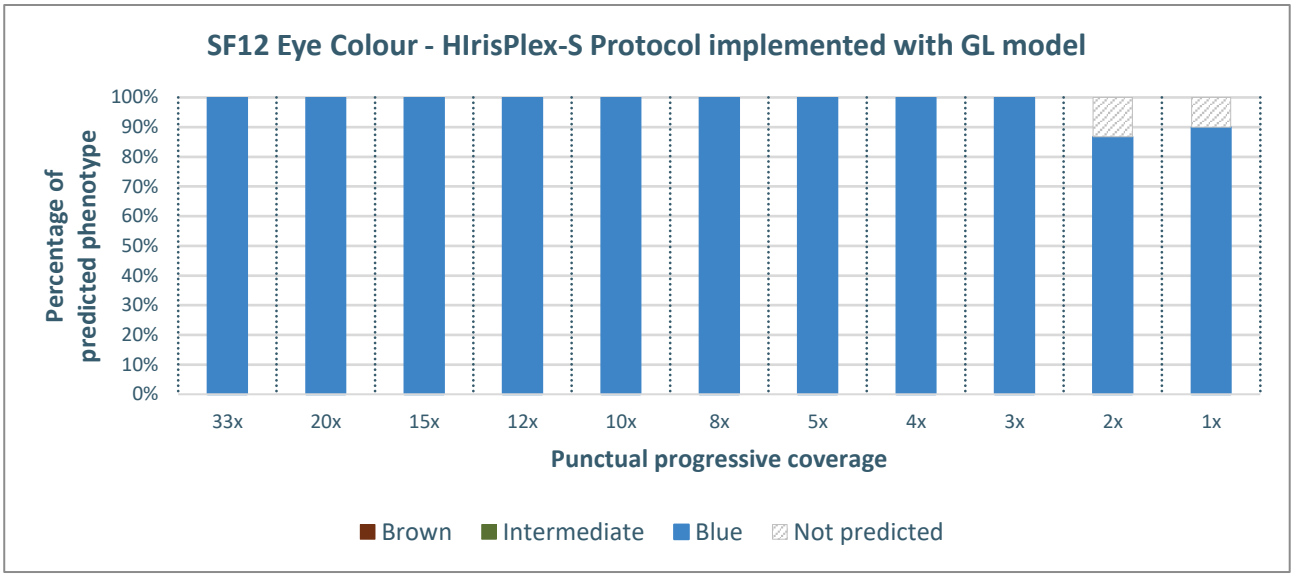


Figure 4.27 - Percentage distribution of predicted eye colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

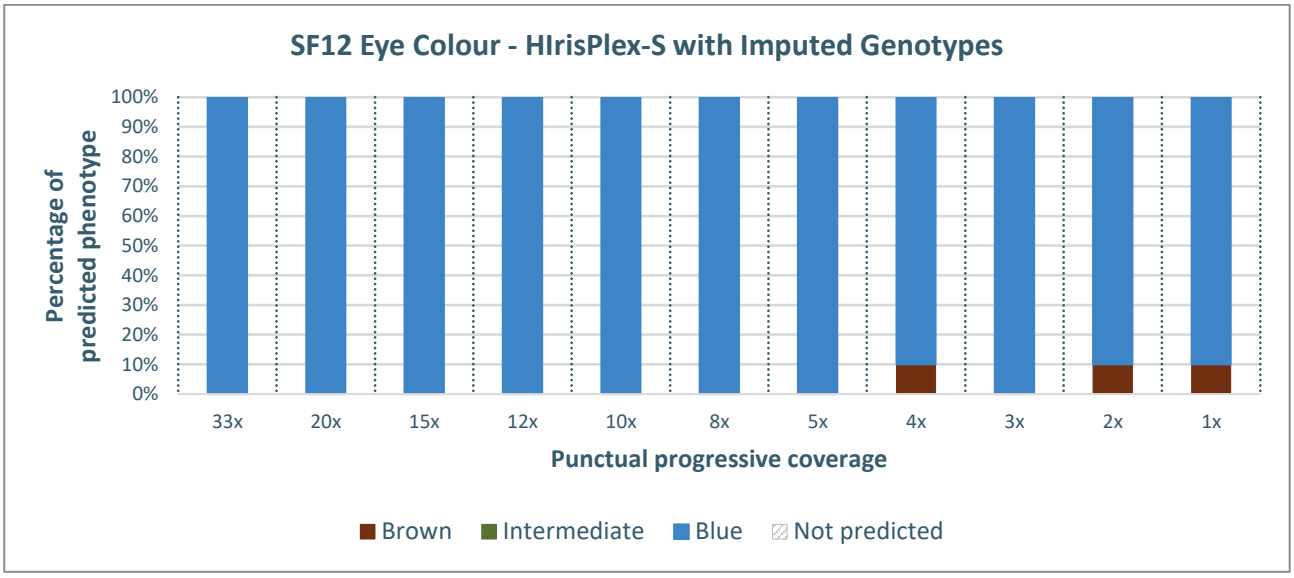


Figure 4.28 - Percentage distribution of predicted eye colour phenotypes for the SF12 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figure 4.29, 4.30, 4.31, and 4.32). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

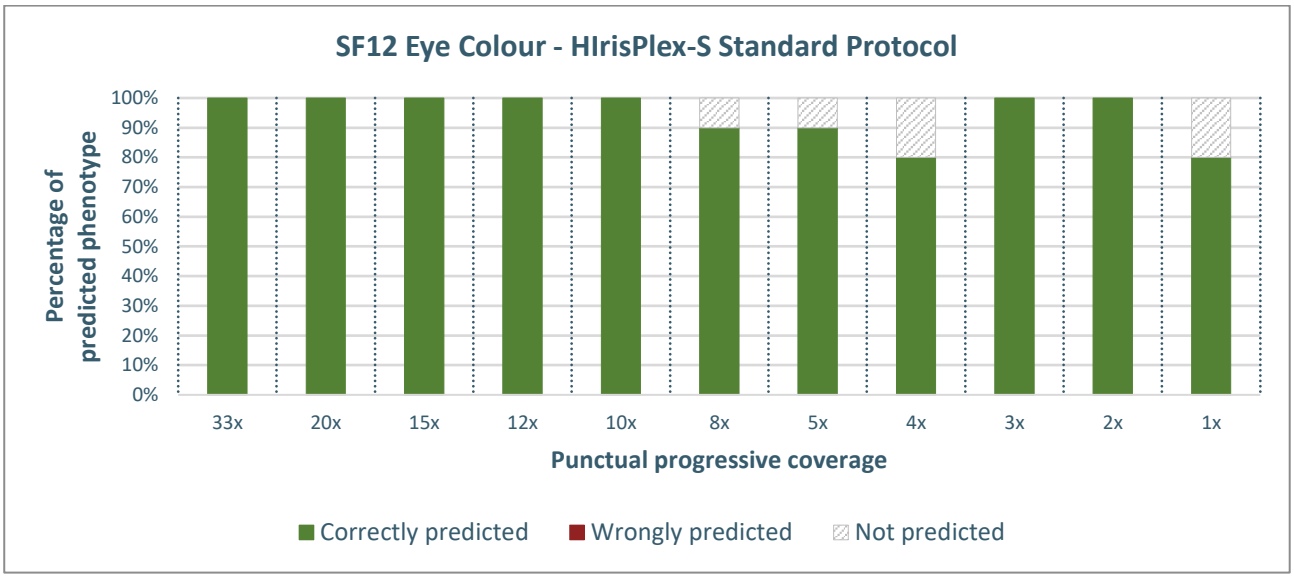


Figure 4.29 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

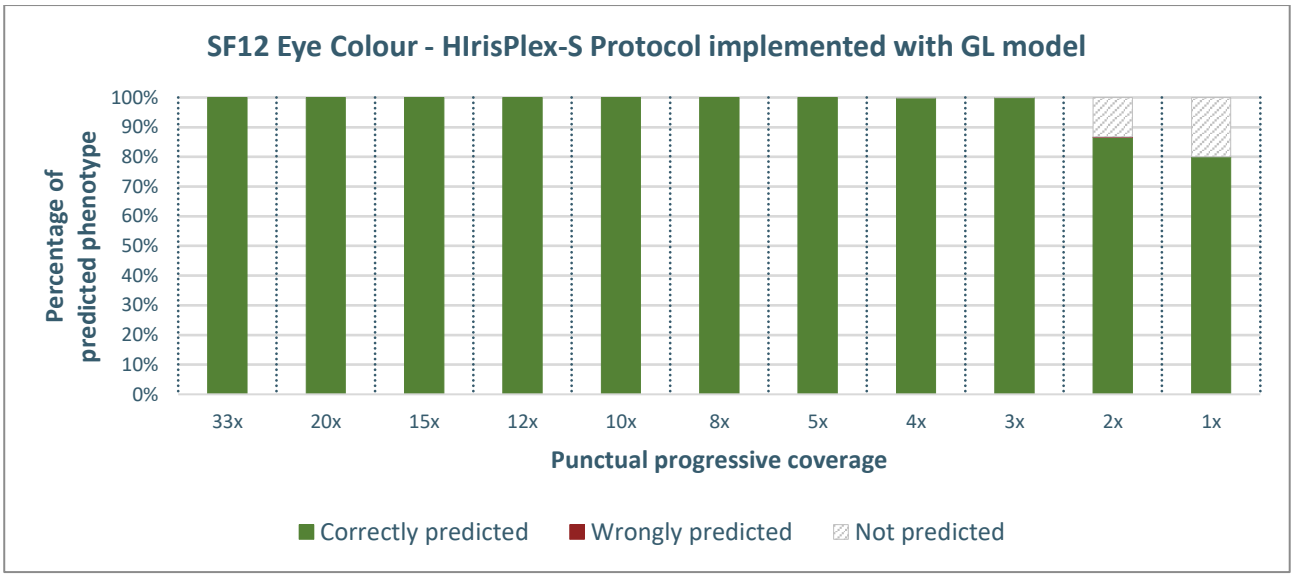


Figure 4.30 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

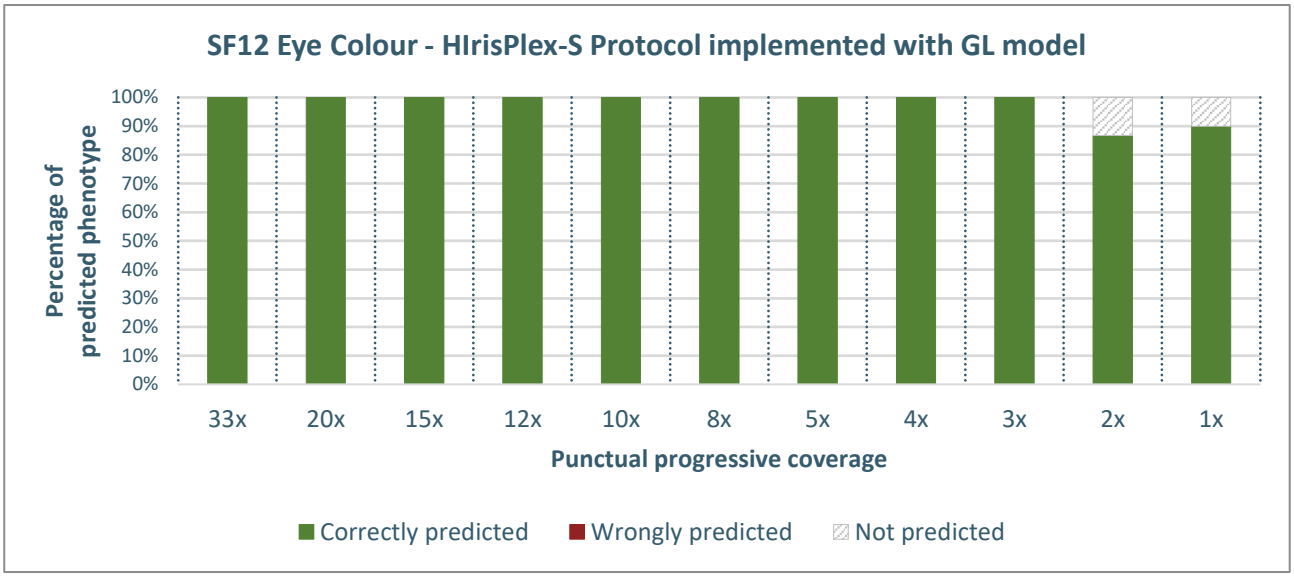


Figure 4.31 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Geno-type Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

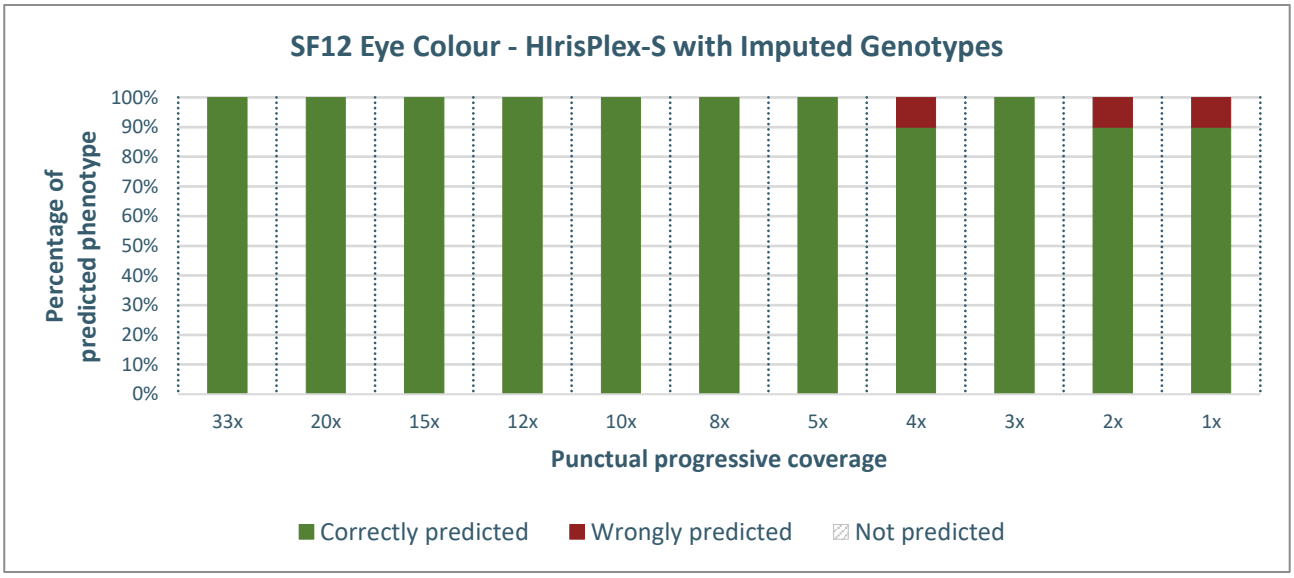


Figure 4.32 - Percentage distribution of predicted phenotypes for the SF12 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.2.2 Hair colour prediction

The results for hair colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.33, 4.34, 4.35, and 4.36). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely brown to dark brown hair (Figures 4.37, 4.38, 4.39, and 4.40).

- HIrisPlex-S Standard Protocol

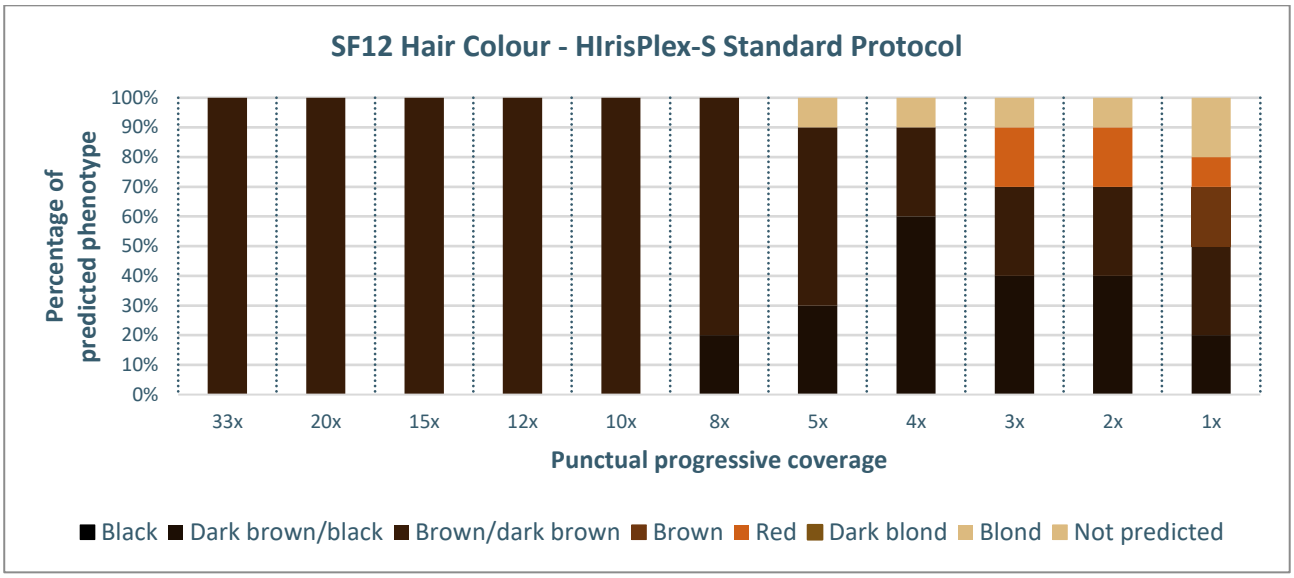


Figure 4.33 - Percentage distribution of predicted hair colour phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

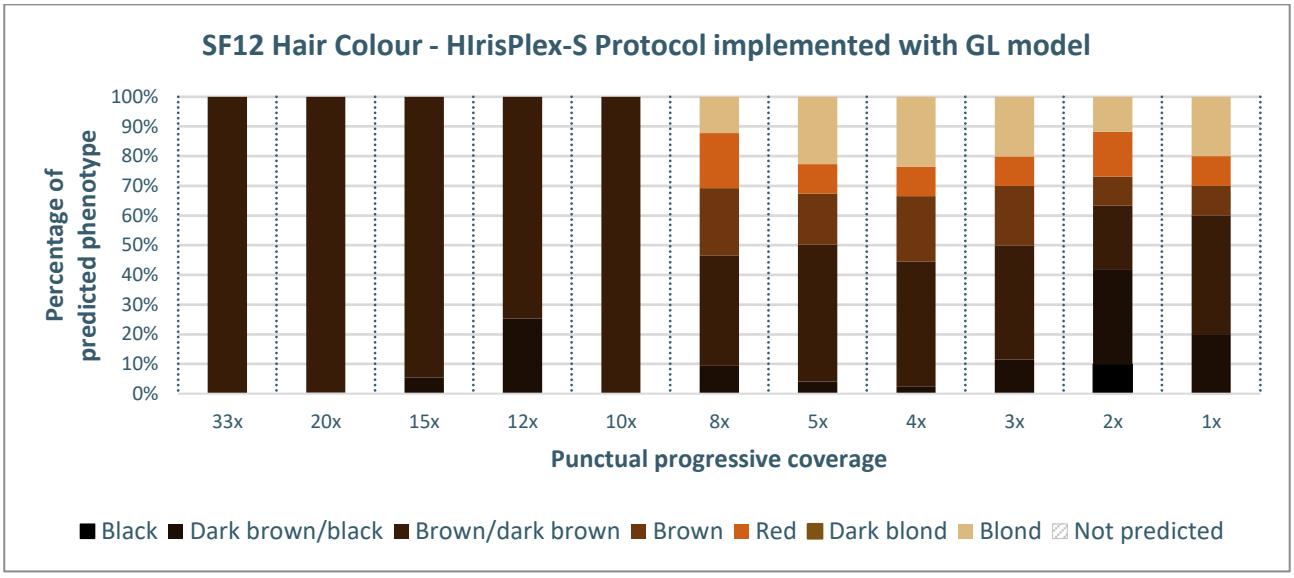


Figure 4.34 - Percentage distribution of predicted hair colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

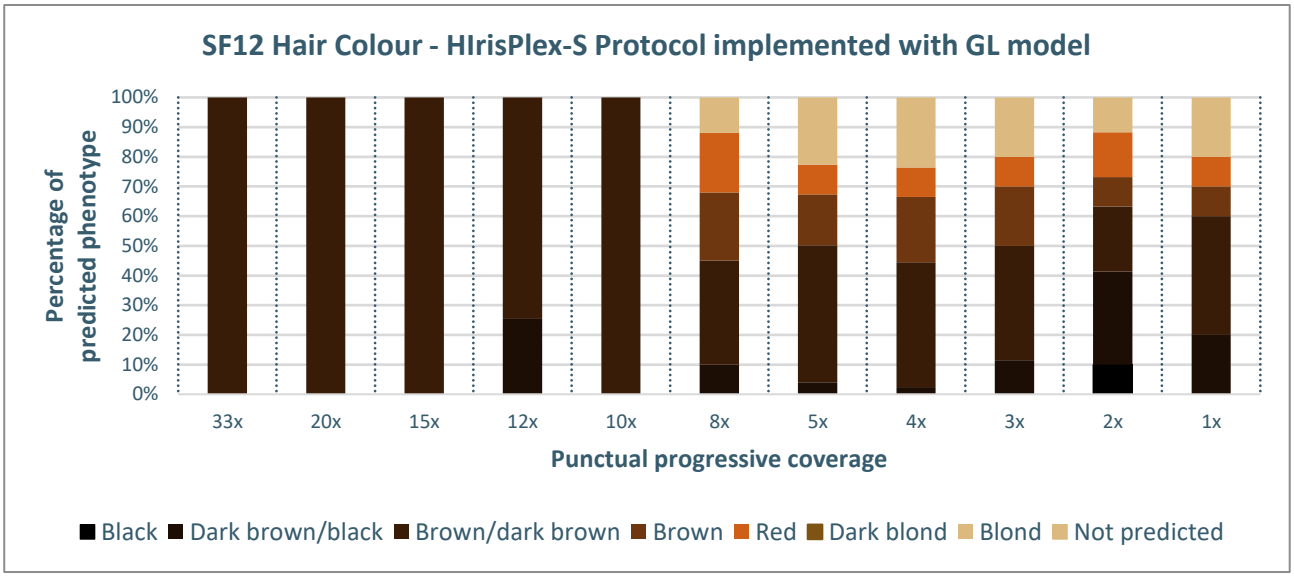


Figure 4.35 - Percentage distribution of predicted hair colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

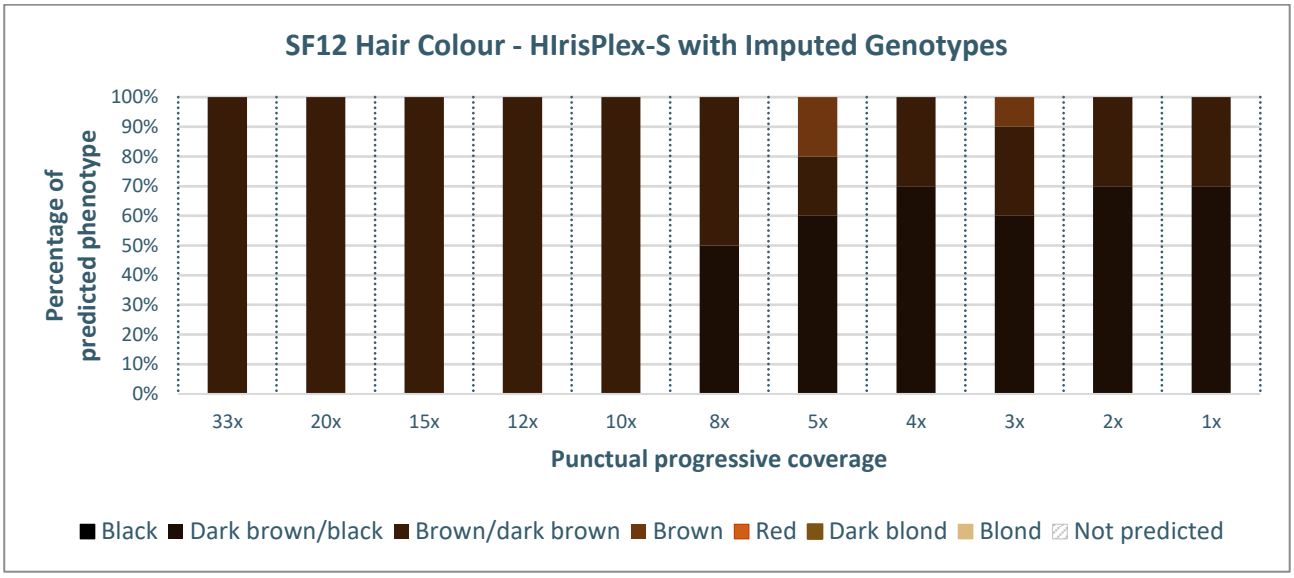


Figure 4.36 - Percentage distribution of predicted hair colour phenotypes for the SF12 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figures 4.37, 4.38, 4.39, and 4.40). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

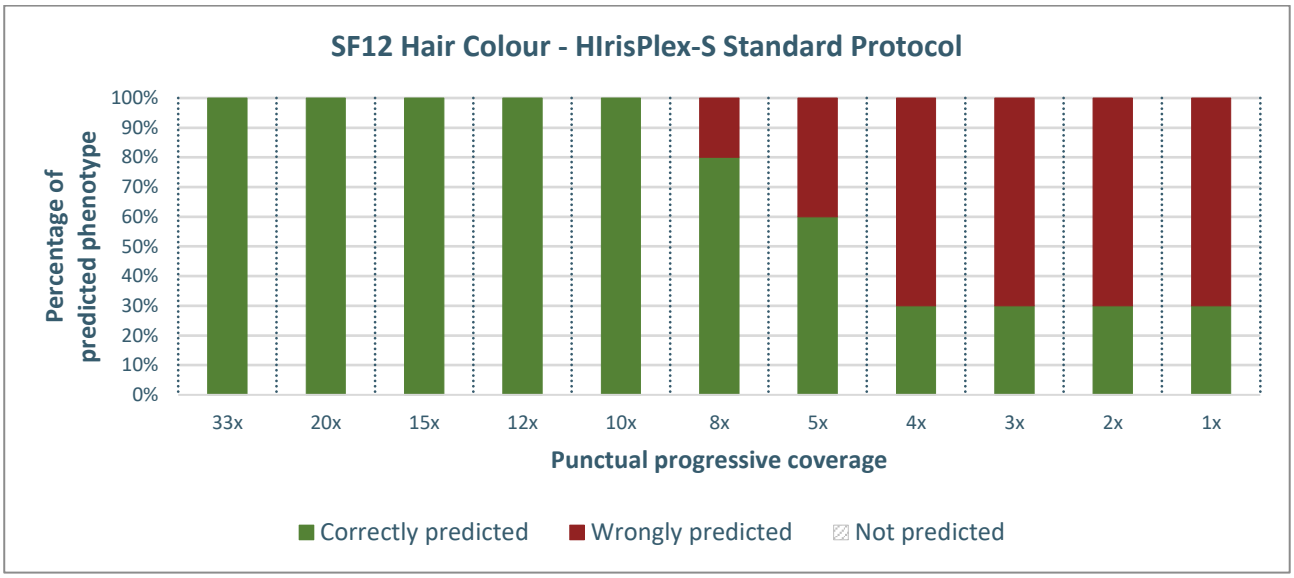


Figure 4.37 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

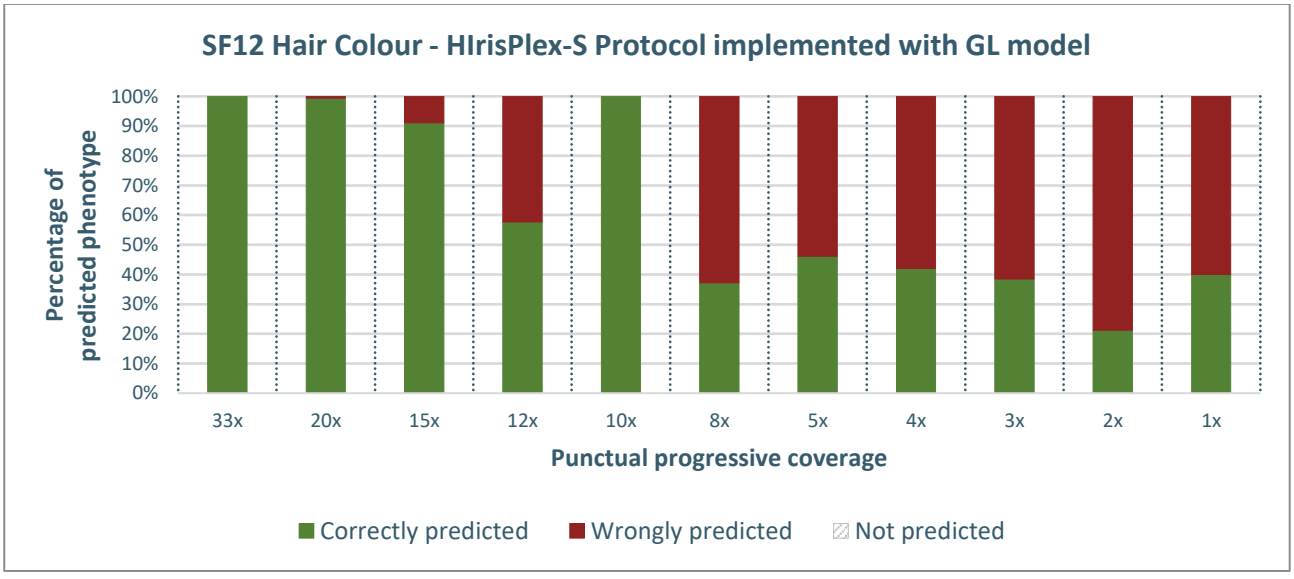


Figure 4.38 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

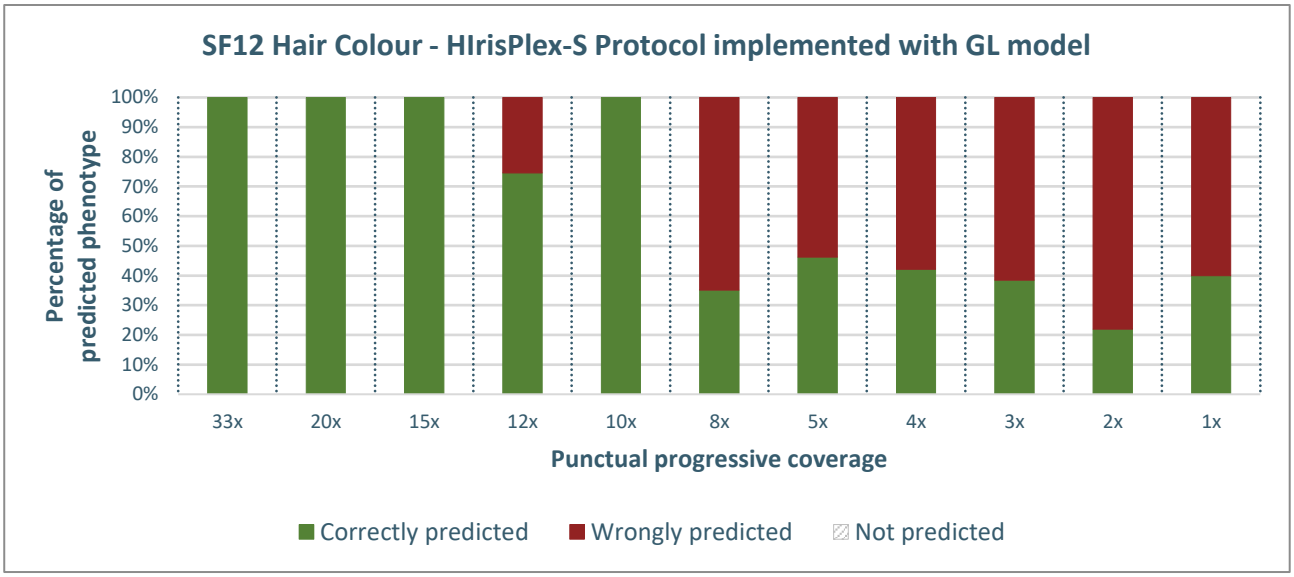


Figure 4.39 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

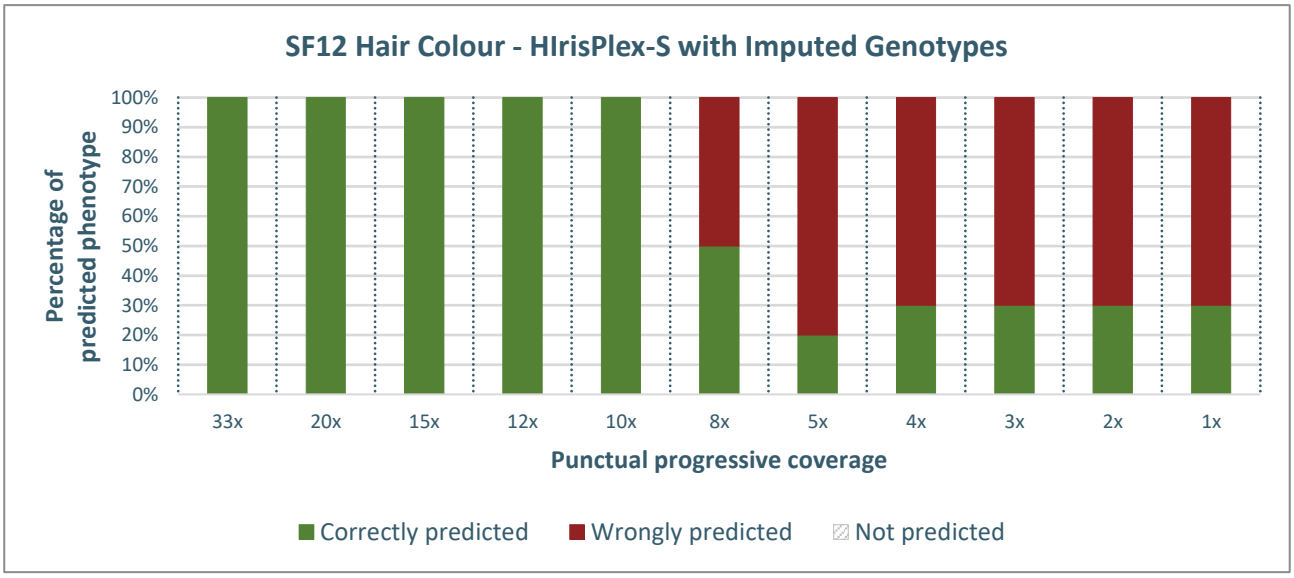


Figure 4.40 - Percentage distribution of predicted phenotypes for the SF12 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.2.3 Skin colour prediction

The results for skin colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.41, 4.42, 4.43, and 4.44). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely dark skin (Figures 4.45, 4.46, 4.47, and 4.48).

- HIrisPlex-S Standard Protocol

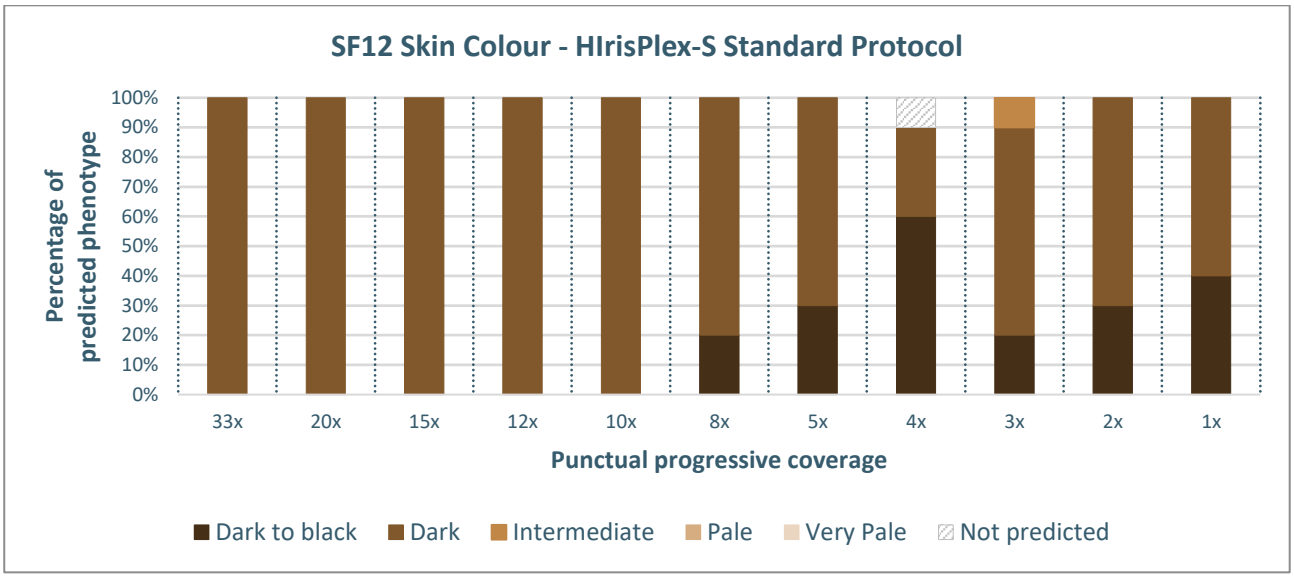


Figure 4.41 - Percentage distribution of predicted skin colour phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

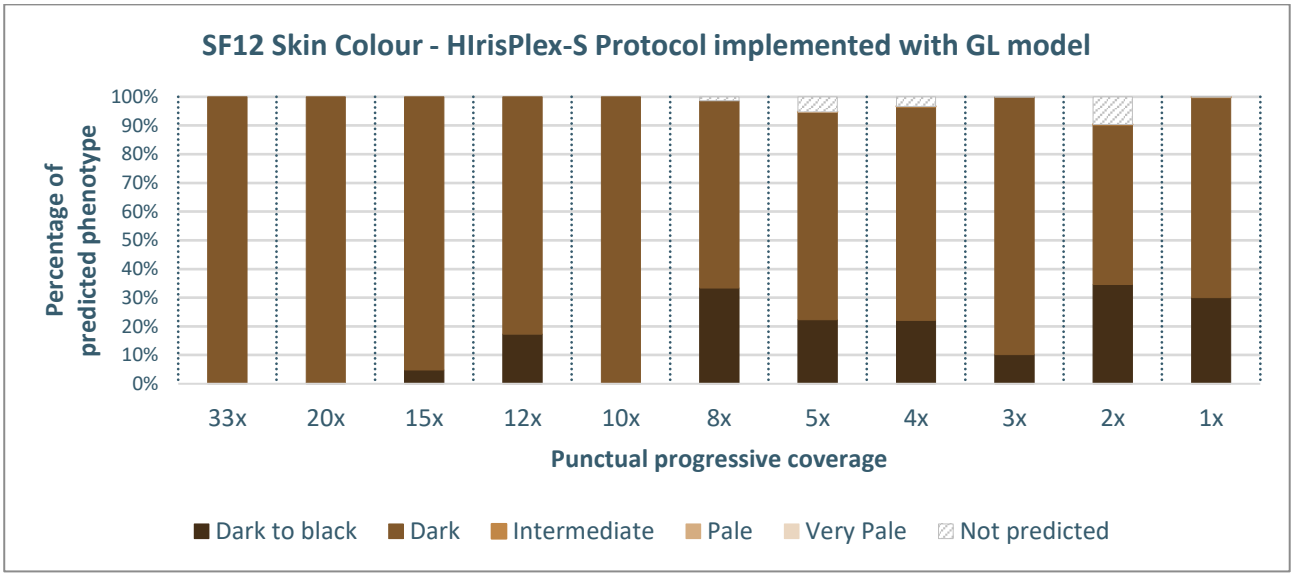


Figure 4.42 - Percentage distribution of predicted skin colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

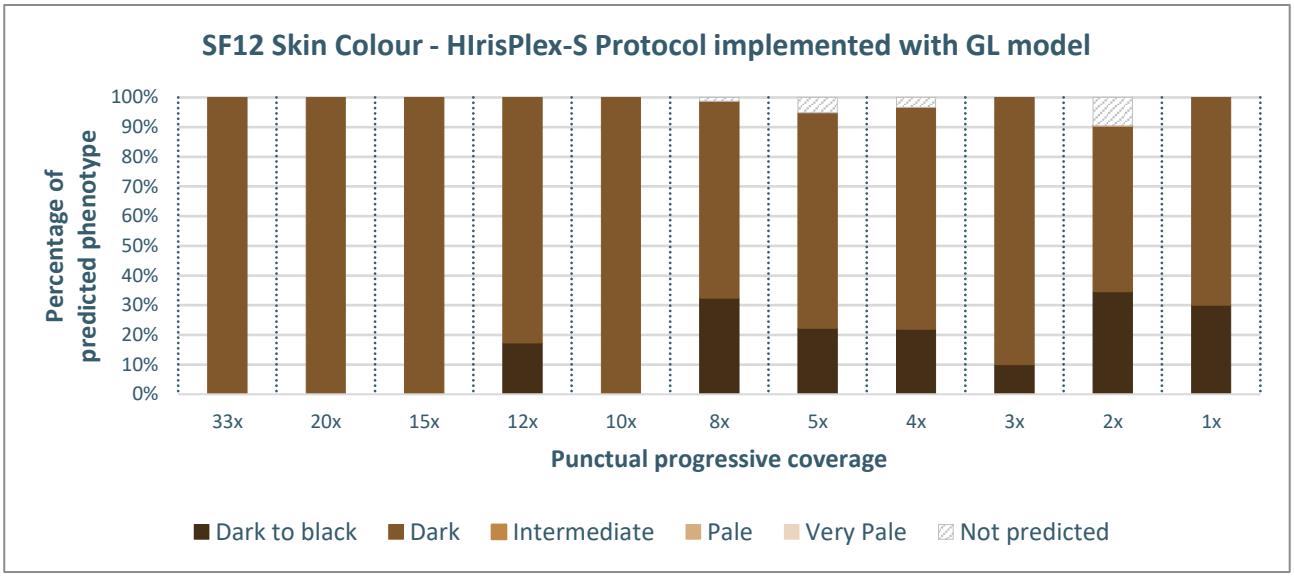


Figure 4.43 - Percentage distribution of predicted skin colour phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

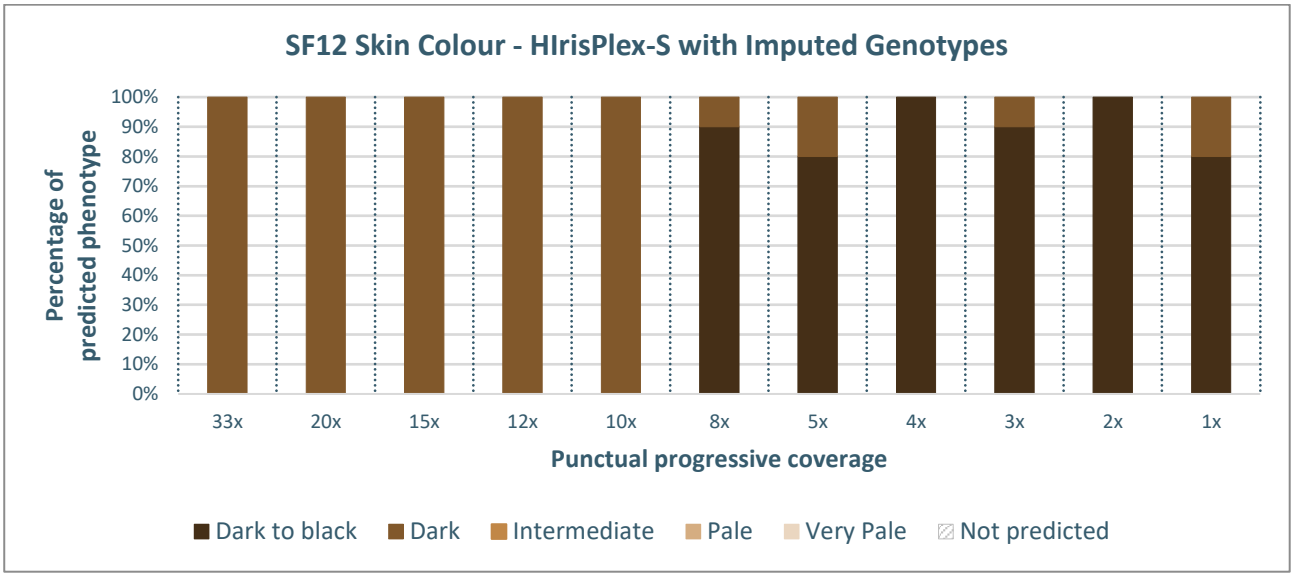


Figure 4.44 - Percentage distribution of predicted skin colour phenotypes for the SF12 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figures 4.45, 4.46, 4.47, and 4.48). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

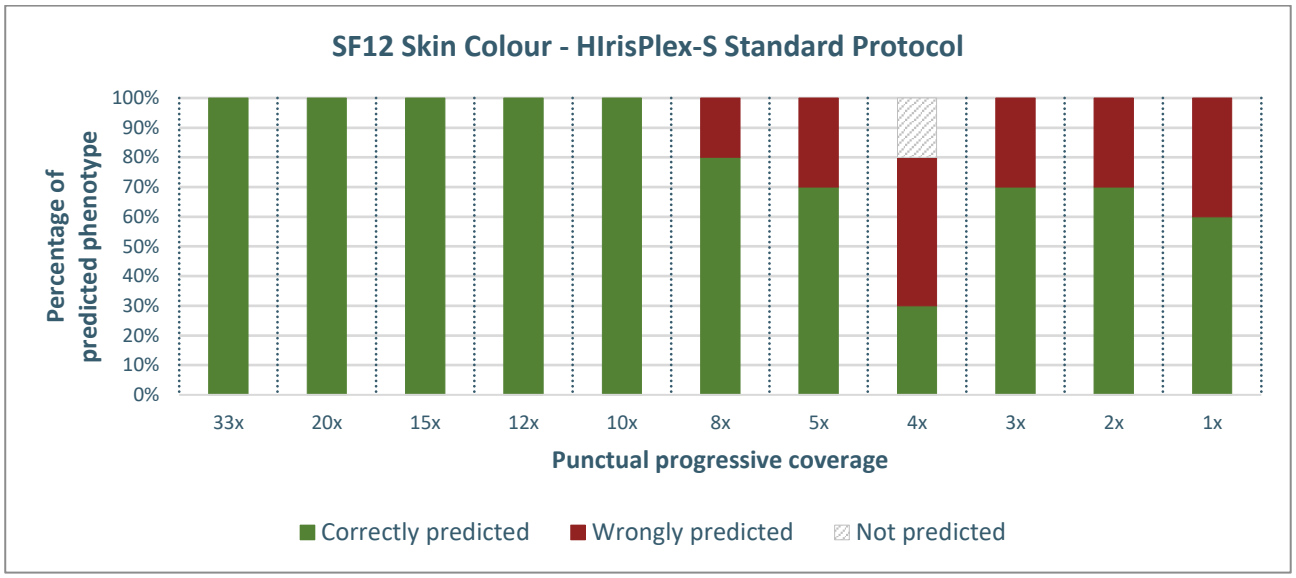


Figure 4.45 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

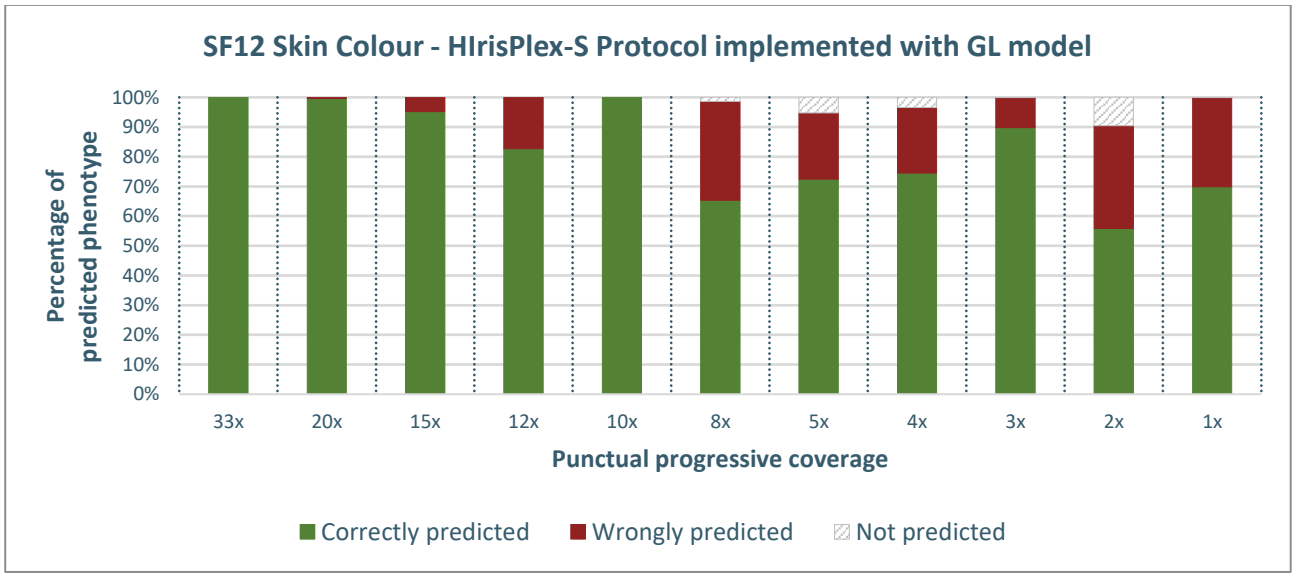


Figure 4.46 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

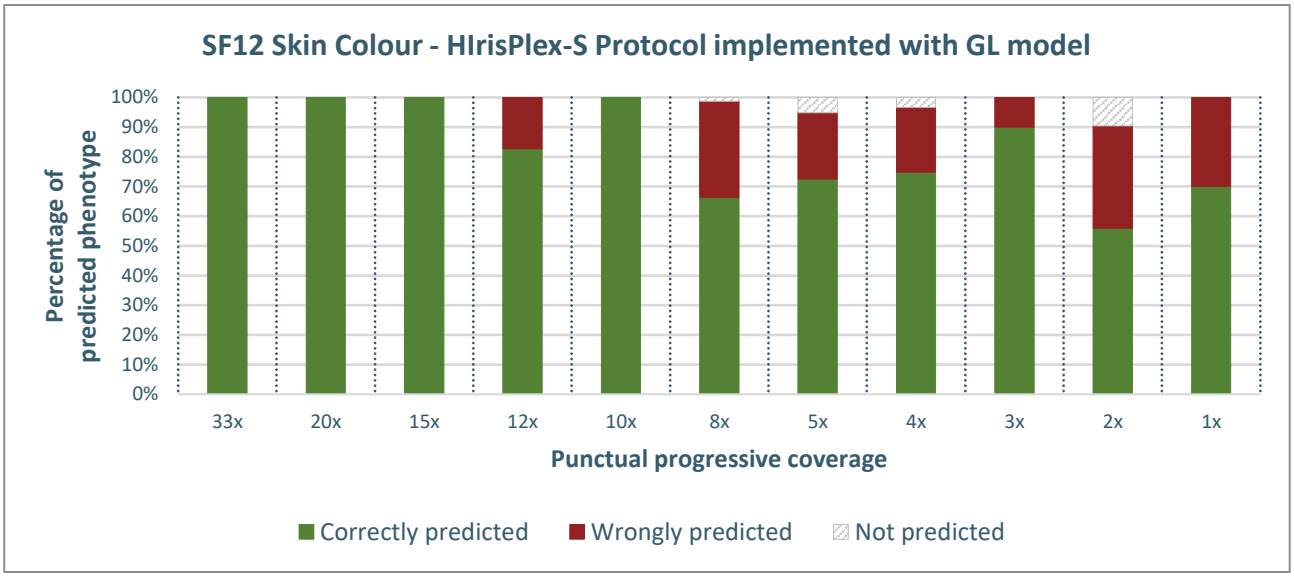


Figure 4.47 - Percentage distribution of predicted phenotypes for the SF12 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

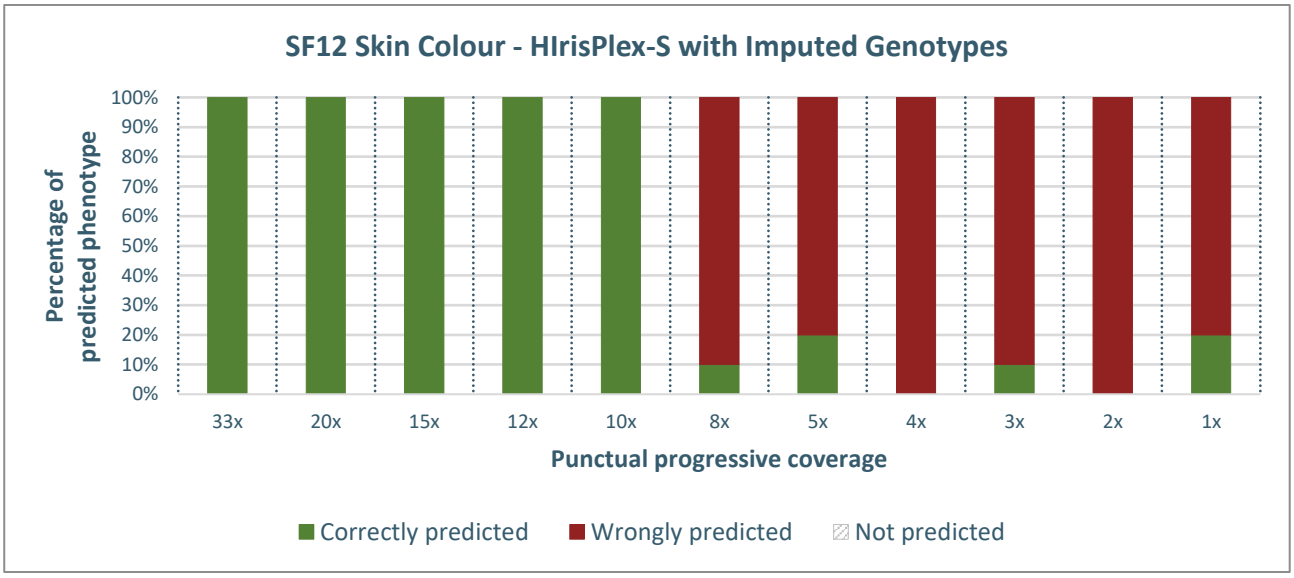


Figure 4.48 - Percentage distribution of predicted phenotypes for the SF12 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

In the context of eye colour phenotypic inference for the Mesolithic sample SF12, the HIrisPlex-S Protocol, when implemented with a Genotype Likelihoods Model, demonstrates better performance compared to both the HIrisPlex-S Standard Protocol and imputed genotypes. As illustrated in Figures 4.25, 4.26, and 4.28, the HIrisPlex-S Standard Protocol consistently underperforms in phenotypic prediction at intermediate to low coverage levels (e.g., 8x, 5x, 4x), with failure rates ranging from 10% to 20% of cases, and reaching up to 20% at 1x coverage. Predictions based on imputed data are inaccurate in 10% of cases at 5x, 2x, and 1x coverage levels. In contrast, the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model exhibits a high predictive capacity, with phenotype prediction failures occurring in only 20% of cases at 1x coverage and 13% at 2x coverage. Furthermore, it shows improved performance in phenotypic prediction at 8x, 5x, and 4x coverage levels, where the HIrisPlex-S Standard Protocol fails.

In the context of hair colour predictions, making phenotypic inferences proves challenging across all three models. The HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model demonstrates inaccuracies of less than 0.2% at 20x coverage, less than 0.09% at 15x coverage, 42% at 12x coverage, 62% at 8x coverage, an average of 50% at 5x, 4x, and 3x coverage, and 80% at 2x coverage. Notably, at 1x coverage, this model outperforms the HIrisPlex-S classical workflow, achieving 40% correct predictions compared to 30%. When applying the 90% threshold (Figure 4.35), incorrect phenotypic predictions at 20x and 15x coverage are eliminated, the rate at 12x coverage decreases from 42% to 25%, while for lower coverage values, the rates remain similar. HIrisPlex-S Standard Protocol exhibits a 20% of incorrect phenotypic predictions at 8x coverage, 40% at 5x coverage and 70% from 4x downward. The conventional workflow for low coverage missing data shows 50% incorrect predictions at 8x coverage, 80% at 5x, and 70% of incorrect phenotypic predictions from 5x coverage downward.

Of particular interest is the phenomenon of overexpression of the red and blond hair phenotype observed in the initial two prediction models, a subject that will be explored in further detail in section 4.2.4.

For skin colour predictions in Mesolithic samples, the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model demonstrates superior performance compared to both the HIrisPlex-S Standard Protocol and the common workflow for low coverage missing data. As depicted in Figures 4.41, 4.42, and 4.44, the HIrisPlex-S Standard Protocol shows significant inaccuracies, with incorrect

prediction rates of 20% at 8x coverage, 40% at 5x coverage, and 40% at 4x coverage. Additionally, it fails to make predictions in 30% of cases at 3x and 2x coverage, and 40% of cases at 1x coverage.

In contrast, the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model produces incorrect predictions in less than 0.05% of cases at 20x and 15x coverage, approximately 20% of cases at 12x coverage, nearly 30% of cases at 8x coverage, about 20% of cases at 5x and 4x coverage, 10% of cases at 3x coverage, 34% of cases at 2x coverage, and 30% of cases at 1x coverage. When applying the 90% threshold (Figure 4.43), incorrect phenotypic predictions are eliminated at coverages higher than 12x. Predictions derived from imputed data also display inaccuracies, starting from 8x coverage and decreasing further. Error rates are 100% at 4x and 2x coverage, 90% at 3x coverage, and 80% at 5x and 1x coverage.

4.2.4 Analysis of Red and Blond Hair Phenotypic Expression

As illustrated by the data presented in Figures 4.33 and 4.34, predictions of hair colour using both the HIrisPlex-S Standard Protocol and the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model consistently show an increased prevalence of the red and blond hair phenotypes.

To elucidate this observation, an example derived from SF12 downsampling 3 is provided for red hair. In this case, the red hair phenotype is absent when using the HIrisPlex-S Standard Protocol but becomes evident with the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model. Specifically, at 8x coverage, the HIrisPlex-S Standard Protocol predicts a brown to dark brown hair colour, reflecting the "true" phenotype (Figure 4.49), while the model based on genotype likelihoods predicts a red hair colour in over 90% of cases (Figure 4.50).

- HIrisPlex-S Standard Protocol

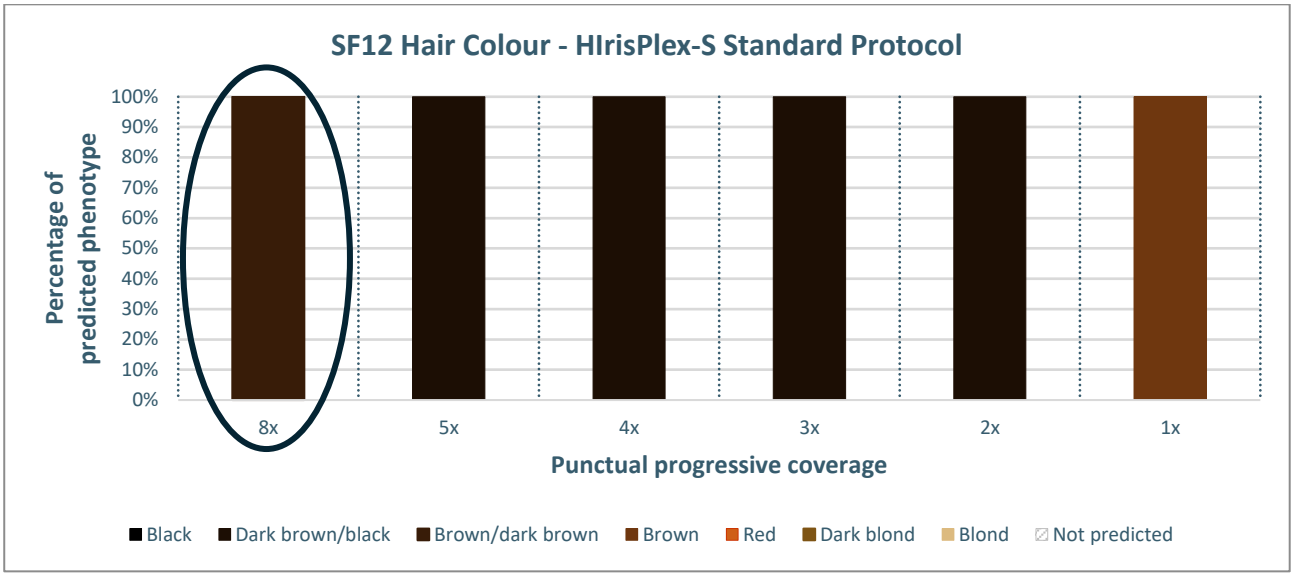


Figure 4.49 - Results of SF12 downsampling number 3 hair phenotypic prediction applying the HIrisPlex-S Standard Protocol. Brown to dark brown phenotype at 8x coverage is circled.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

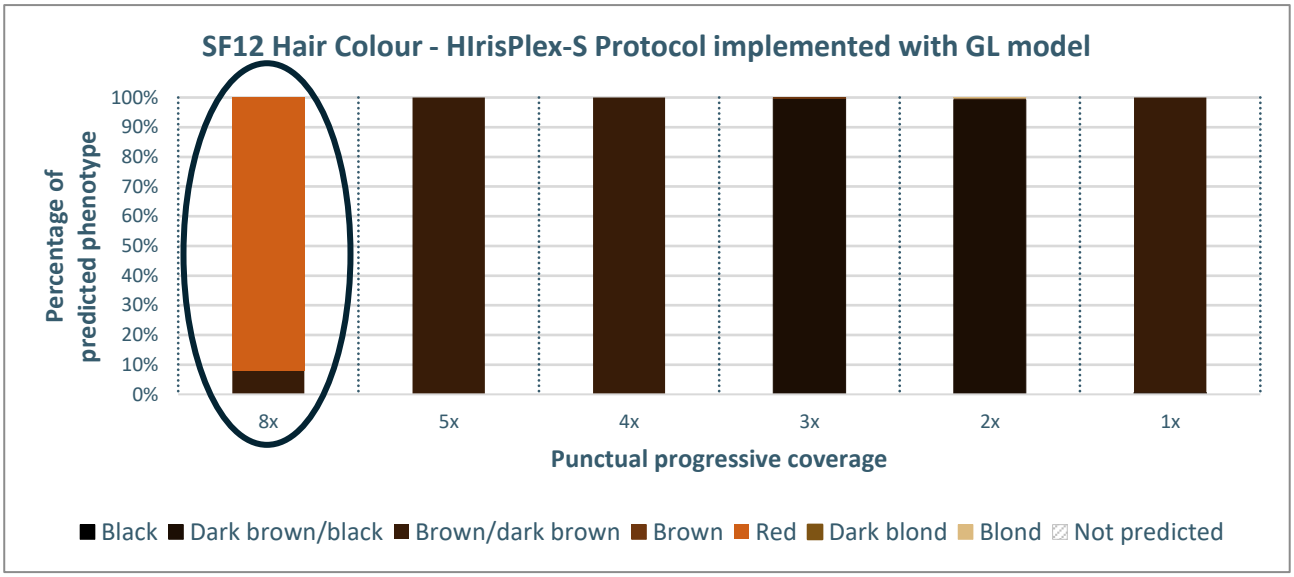


Figure 4.50 - Results of SF12 downsampling number 3 hair phenotypic prediction applying the HIrisPlex-S Protocol implemented with GL model. Red hair phenotype at 8x coverage is circled.

Upon examining the input alleles, it becomes evident that only one allele at rs201326893 leads to a distinct phenotypic prediction, as demonstrated in Table 4.1.

					SF12 8x HIrisPlex-S Standard Protocol		SF12 8x HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model	
HIrisPlex-S position	**Gene**	**Genomic position**	**Rs allele**	**Input allele**	**Genotype**	**Allele number**	**Genotype**	**Allele number**
9	MC1R	chr16:89986122	rs201326893	A	CC	0	CA	1

Table 4.1 - Comparison of different input alleles at position 89986122 on chromosome 16 for sample SF12, as determined by the HIrisPlex-S Standard Protocol and the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model.

Following the inspection of the alignment at this position, it is evident that there are 8 bases (8x punctual coverage), consisting of 7 C bases and 1 A base. The base qualities for all bases remain consistently high, with each base recording a score of 60 (Table 4.2).

Alignment	C	C	C	A	C	C	C	C
Base quality	60	60	60	60	60	60	60	60

Table 4.2 - Alignment and base quality at position 89986122 on chromosome 16 for sample SF12 at downsampling 3.

These observations lead to the consideration of two potential explanations for the phenotypic inference: one based on biological factors and the other on statistical analysis.

The biological explanation is derived from the research conducted by Walsh and colleagues in 2013, the same year the HIrisPlex system was released. This research explored the impact of *MC1R* variant alleles at N29insA (now rs312262906) and Y152OCH (now rs201326893), which significantly influence the probabilities of red hair. Specifically, the presence of an A allele at either N29insA or Y152OCH results in a red hair prediction probability of 1. This prediction is unique to these specific *MC1R* variants, indicating that these exceptionally rare alleles (in both heterozygote and homozygote states) are associated with individuals having a red hair phenotype (Walsh *et al.*, 2013).

Figure 4.51 illustrates the contribution of the SNP rs201326893.

Number	DNA variant	Gene	Blond(beta)	Blond(p)	Brown(beta)	Brown(p)	Black(beta)	Black(p)	Red(beta)	Red(p)
1	N29insA	MC1R	–	–	–	–	–	–	-21.9731	0.994026
2	rs11547464	MC1R	–0.947299	0.081175	–0.4007191	0.441688	–16.782634	0.995907	–2.8866	4.42E–08
3	rs885479	MC1R	0.272536	3.36E–01	0.1938828	0.460717	2.29E–01	0.575679	0.315529	0.707292
4	rs1805008	MC1R	–0.57034	0.003874	–0.3058868	0.097798	–5.66E–01	0.084668	–3.02472	2.20E–16
5	rs1805005	MC1R	0.20689	2.28E–01	0.2382036	0.128146	–1.57E–01	0.539306	–0.86742	0.025064
6	rs1805006	MC1R	1.718508	0.045418	2.1268136	0.009857	–1.70E+01	0.996356	–2.43626	0.001714
7	rs1805007	MC1R	–0.53542	0.030279	–0.1503278	0.508278	–1.32E+00	0.009567	–3.59956	2.20E–16
8	rs1805009	MC1R	0.550547	5.60E–01	0.5309897	0.49513	–4.70E–01	0.693758	–4.25774	4.14E–08
9	Y152OCH	MC1R	–	–	–	–	–	–	–19.3501	0.992969
10	rs2228479	MC1R	0.025643	8.83E–01	–0.1128742	0.483857	1.98E–01	0.413966	–0.61967	0.110936
11	rs1110400	MC1R	–0.366071	0.338334	–0.5920858	0.123046	6.63E–01	0.21252	–1.67775	0.009302
12	rs28777	SLC45A2	0.566568	0.414238	0.3138274	0.561428	4.85E–01	0.468883	–0.41607	0.743869
13	rs16891982	SLC45A2	0.863795	0.194837	0.2562763	0.618846	6.29E–01	0.326034	0.891013	0.522114
14	rs12821256	KITLG	–0.434962	0.020898	–0.1743193	0.32142	–6.87E–01	0.056556	0.406751	0.312582
15	rs4959270	EXOC2	–0.251437	0.019073	–0.1555227	0.120958	–2.71E–01	0.104087	–0.34639	0.107774
16	rs12203592	IRF4	1.741377	2.20E–16	1.0810914	2.22E–16	8.80E–01	2.35E 06	0.071132	0.773323
17	rs1042602	TYR	0.125113	0.24551	0.141479	0.155781	–4.52E–02	0.779493	–0.3842	0.071464
18	rs1800407	OCA2	–0.204189	0.331948	–0.0048133	0.97935	–3.53E–01	0.202517	0.223931	0.580501
19	rs2402130	SLC24A4	0.354085	0.00797	0.2752735	0.023746	4.36E–02	0.820086	–0.08861	0.724429
20	rs12913832	HERC2	1.372353	2.20E–16	0.6797949	6.83E–10	1.19E+00	6.65E 13	0.754729	0.004319
21	rs2378249	PIGU/ASIP	0.088319	0.526489	0.1828612	0.154928	–1.64E–01	0.449722	–0.72184	0.002302
22	rs683	TYRP1	0.197865	0.066913	0.168184	0.08995	1.58E–03	0.992081	0.129235	0.540918

Figure 4.51 - Assessment of the contribution of each HIrisPlex-S DNA variant for hair colour prediction within the model in terms of betas and probability (p) values. rs201326893 is circled. Extracted from Walsh et al. 2013.

The statistical reasoning is based on the genotype likelihood approach. When analysing 8 bases of uniformly high quality, this method indicates a 92% probability of a heterozygous site among all 10 genotypes (Table 4.3).

Genotype	AA	AC	AG	AT	CC	CG	CT	GG
Posterior probability	1.05E-40	0.92347223	5.27E-44	5.27E-44	0.07652653	6.13E-07	6.13E-07	3.50E-47

Table 4.3 - Genotype posterior probabilities for rs201326893 for sample SF12 at downsampling 3. In red AC genotype posterior probabilities.

Direct calling and imputation, although using a genotype likelihood model during genotype calling, involve a secondary recalibration step (McKenna *et al.*, 2010; Rubinacci *et al.*, 2021). This step adjusts genotypes based on allele frequencies derived from a reference panel of contemporary human genotypes. Given the rarity of this variant in modern humans (MAF < 0.01) (Branicki *et al.*, 2011; https://www.ncbi.nlm.nih.gov/snp/), the likelihood of observing this genotype decreases, causing the genotype calling process to favour an alternative genotype.

A similar discussion can be applied to the blond phenotype. For instance, in SF12 downsampling number 5, number 6, and number 9, the blond phenotype is associated with the GG genotype at

position 33951693 on chromosome 5 (rs16891982) (whereas the SF12 high-coverage sample exhibits the CG genotype). Branicki's 2011 study demonstrates that the C allele is associated with the black phenotype, whereas the G allele (with a minor allele frequency of 0.2) is linked to a lighter phenotype (Branicki *et al.*, 2011).

4.3 I5832 sample

For the Bronze Age sample I5832 (with an average coverage of 22.5x), the maximum coverage at the HIrisPlex-S positions, from which downsampling initiated, was 8x.

The results of each specific downsampling are provided in the folder at the following link: I5835.

4.3.1 Eye colour prediction

The results for eye colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.52, 4.53, 4.54, and 4.55). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely brown eye (Figures 4.56, 4.57, 4.58, and 4.59).

- HIrisPlex-S Standard Protocol

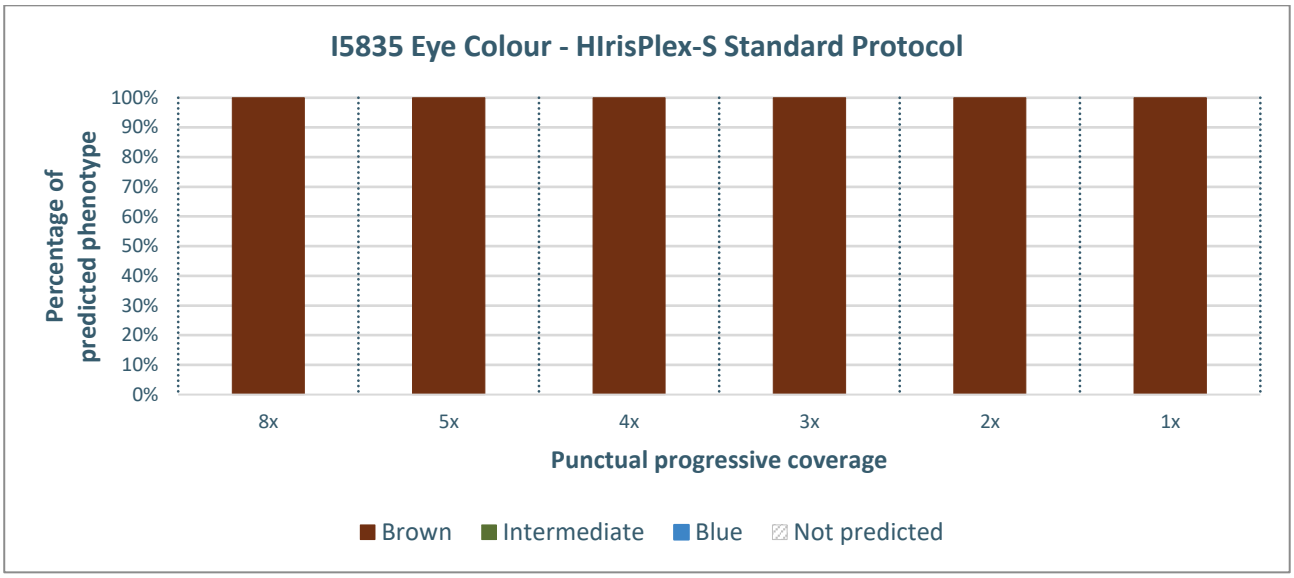


Figure 4.52 - Percentage distribution of predicted eye colour phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

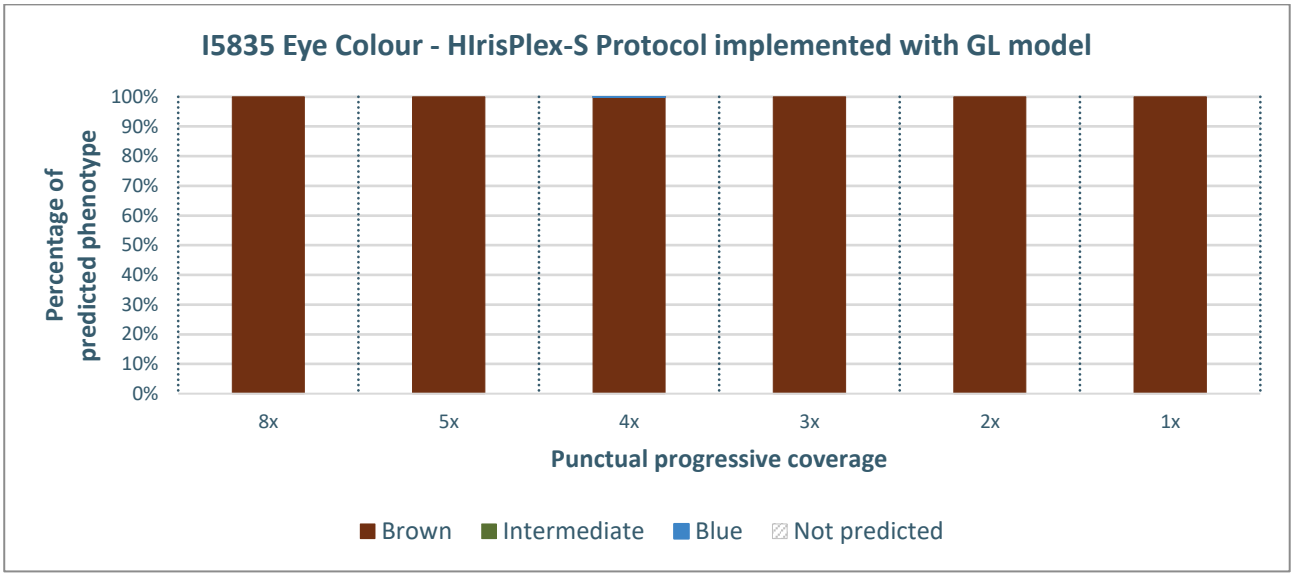


Figure 4.53 - Percentage distribution of predicted eye colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

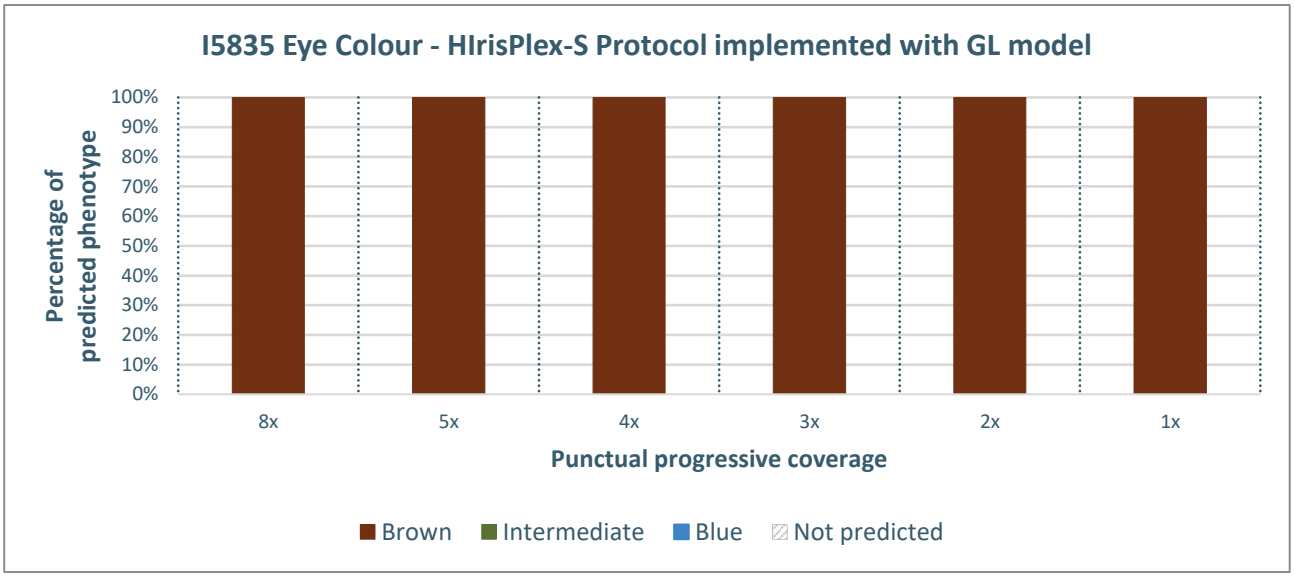


Figure 4.54 - Percentage distribution of predicted eye colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

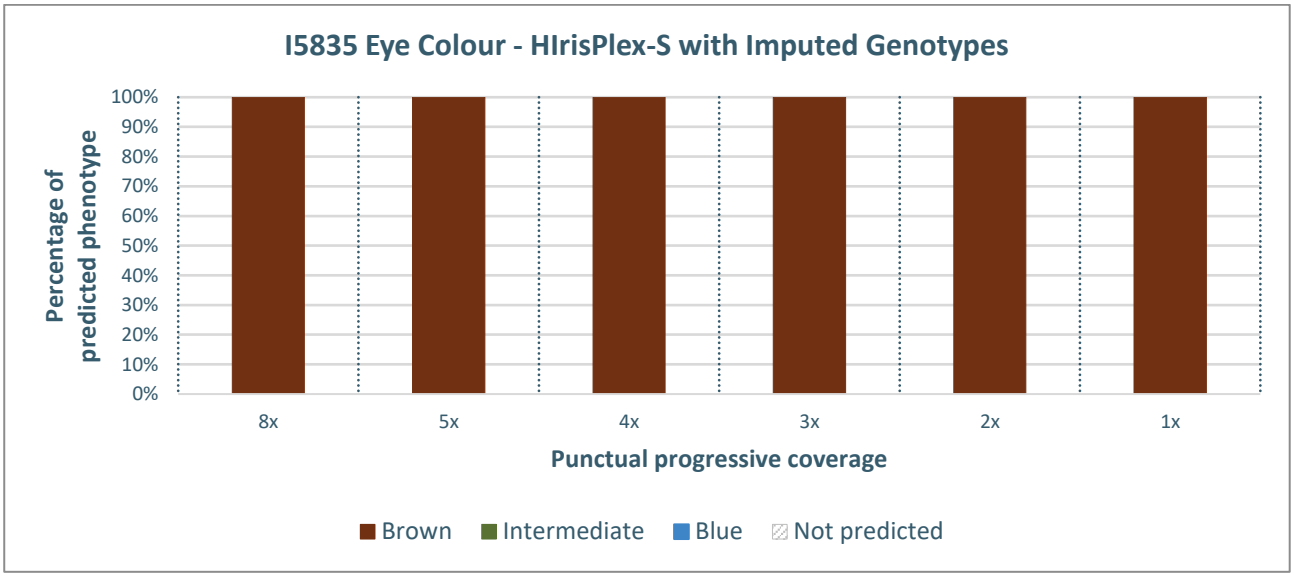


Figure 4.55 - Percentage distribution of predicted eye colour phenotypes for the I5832 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figure 4.56, 4.57, 4,58, and 4.59). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

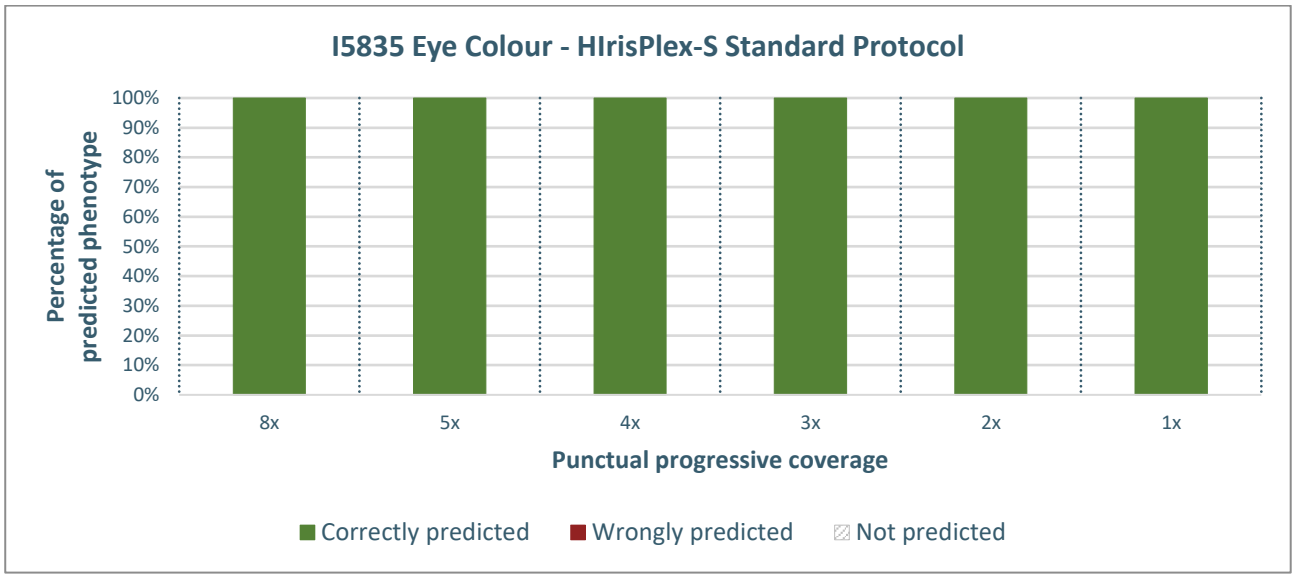


Figure 4.56 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

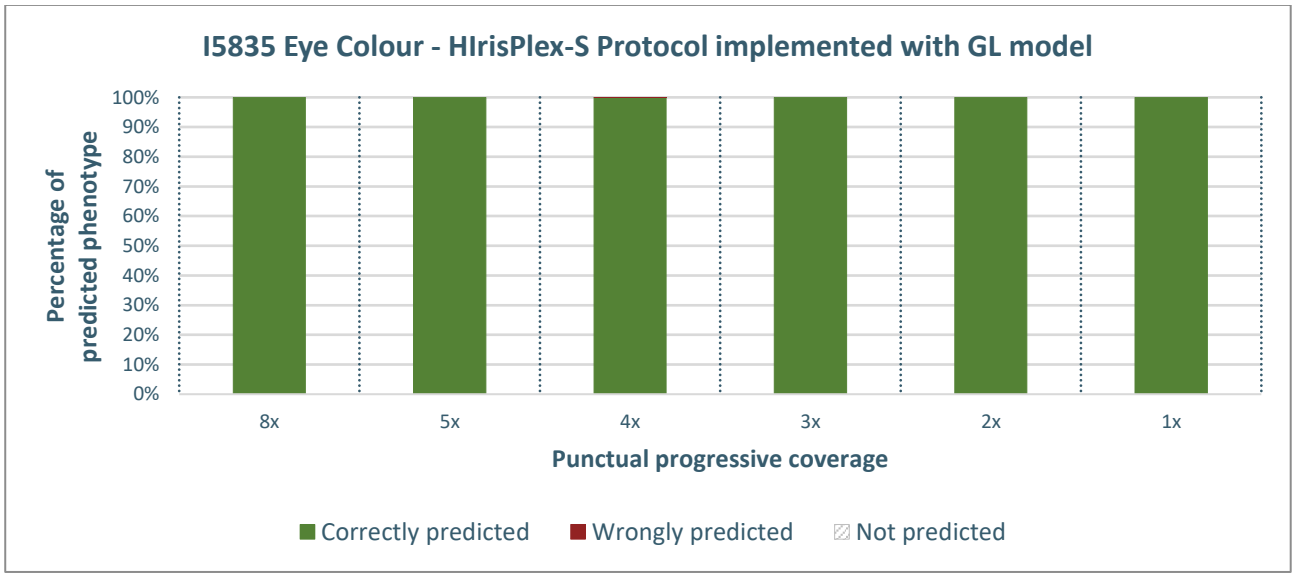


Figure 4.57 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

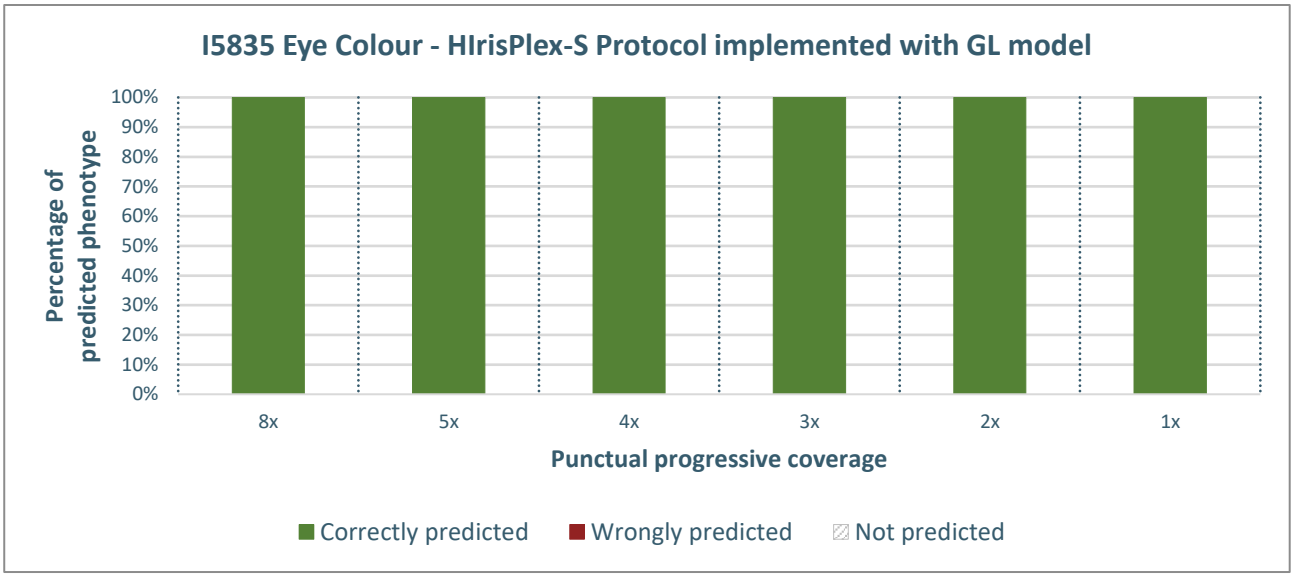


Figure 4.58 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Geno-type Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

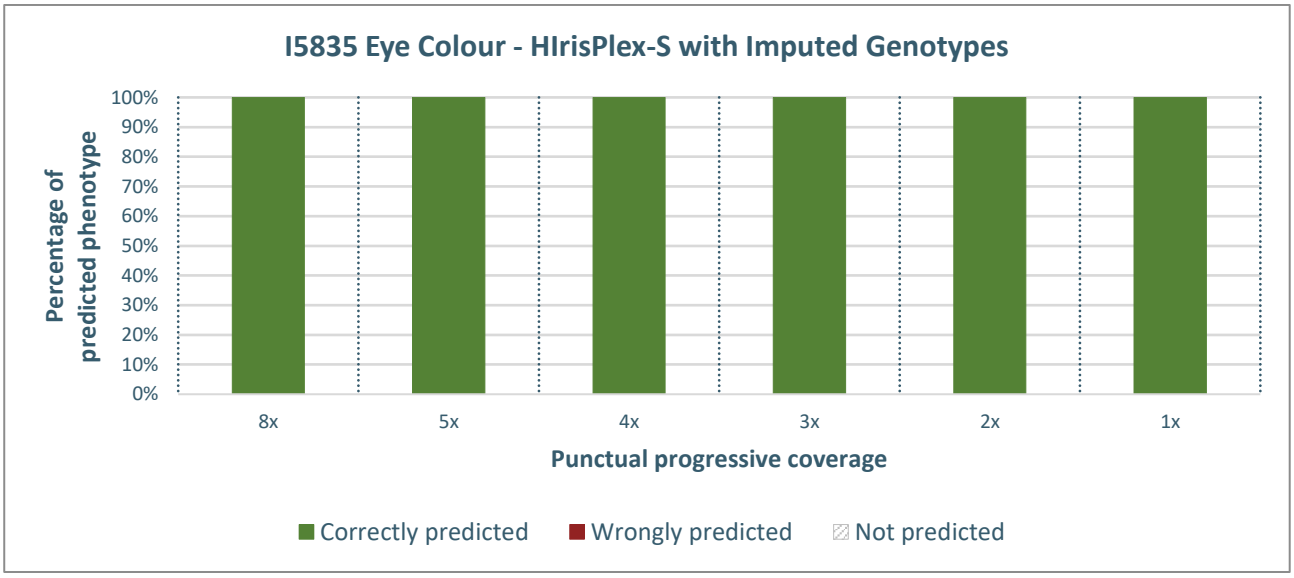


Figure 4.59 - Percentage distribution of predicted phenotypes for the I5832 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.3.2 Hair colour prediction

The results for hair colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.60, 4.61, 4.62, and 4.63). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely dark brown to black hair (Figures 4.64, 4.65, 4.66, and 4.67).

- HIrisPlex-S Standard Protocol

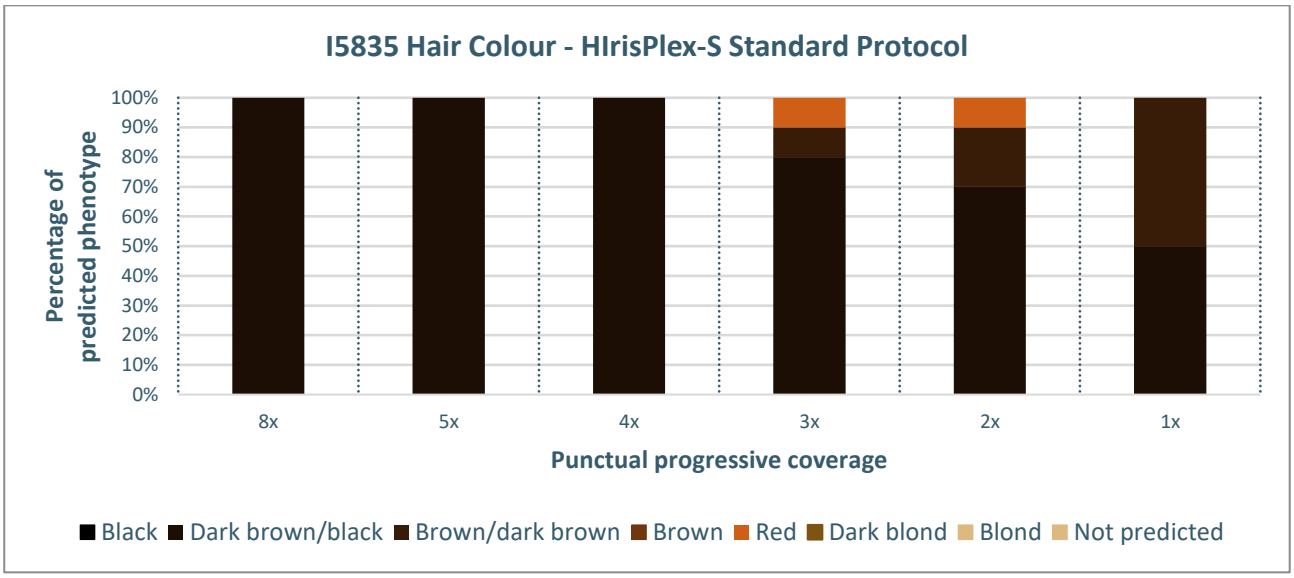


Figure 4.60 - Percentage distribution of predicted hair colour phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

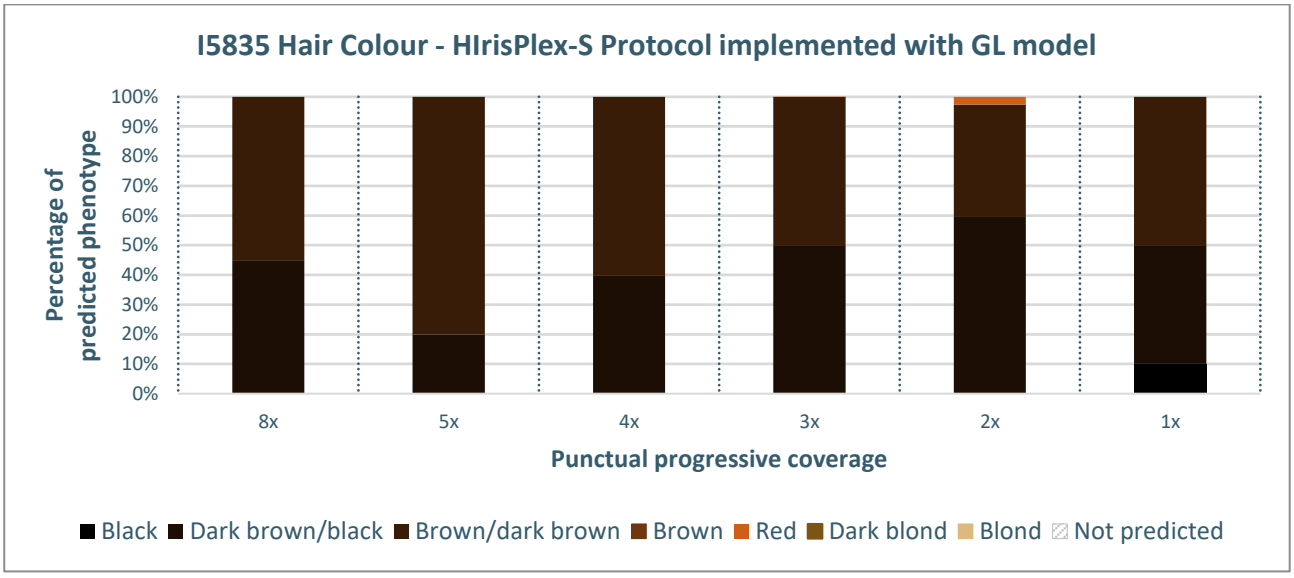


Figure 4.61 - Percentage distribution of predicted hair colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

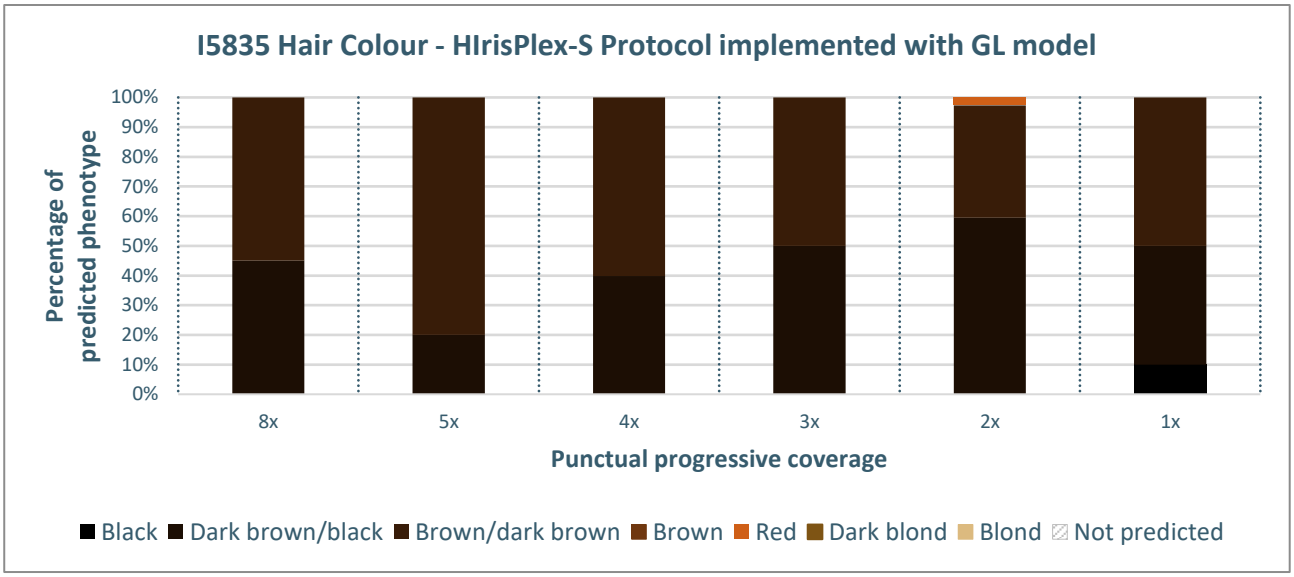


Figure 4.62 - Percentage distribution of predicted hair colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

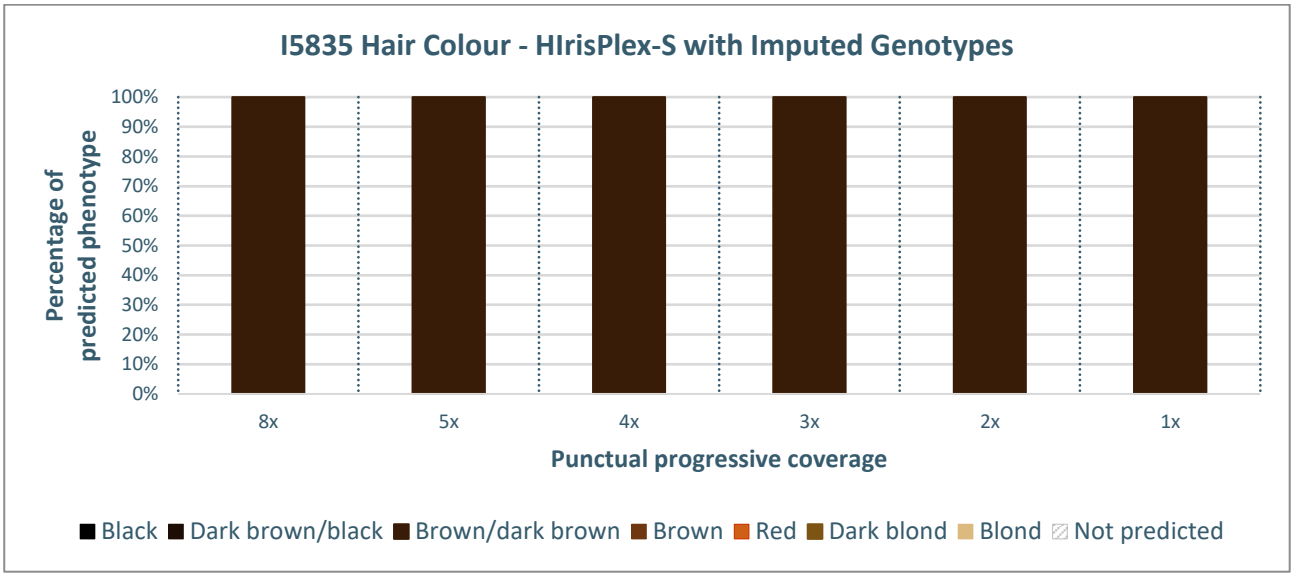


Figure 4.63 - Percentage distribution of predicted hair colour phenotypes for the I5832 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figures 4.64, 4.65, 4.66, and 4.67). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

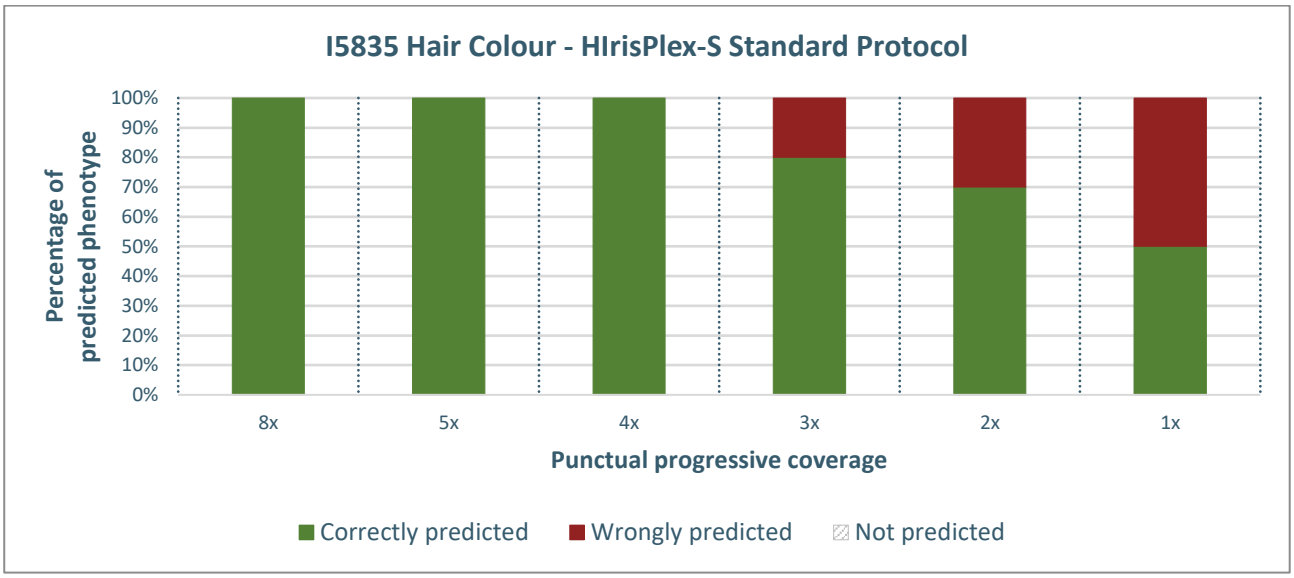


Figure 4.64 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

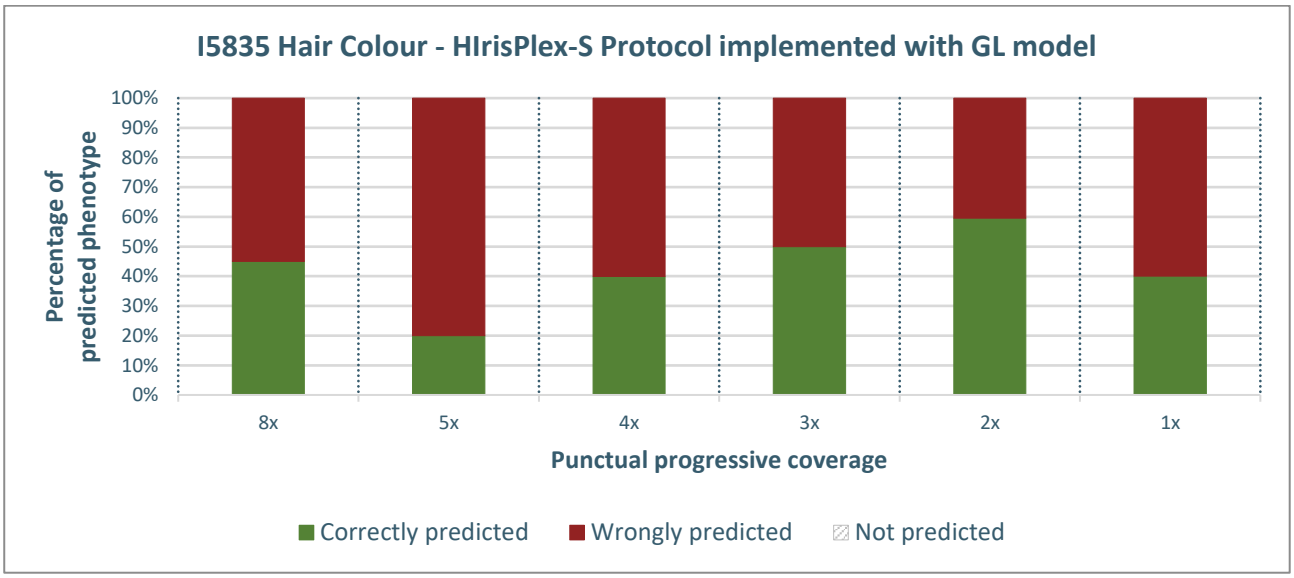


Figure 4.65 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

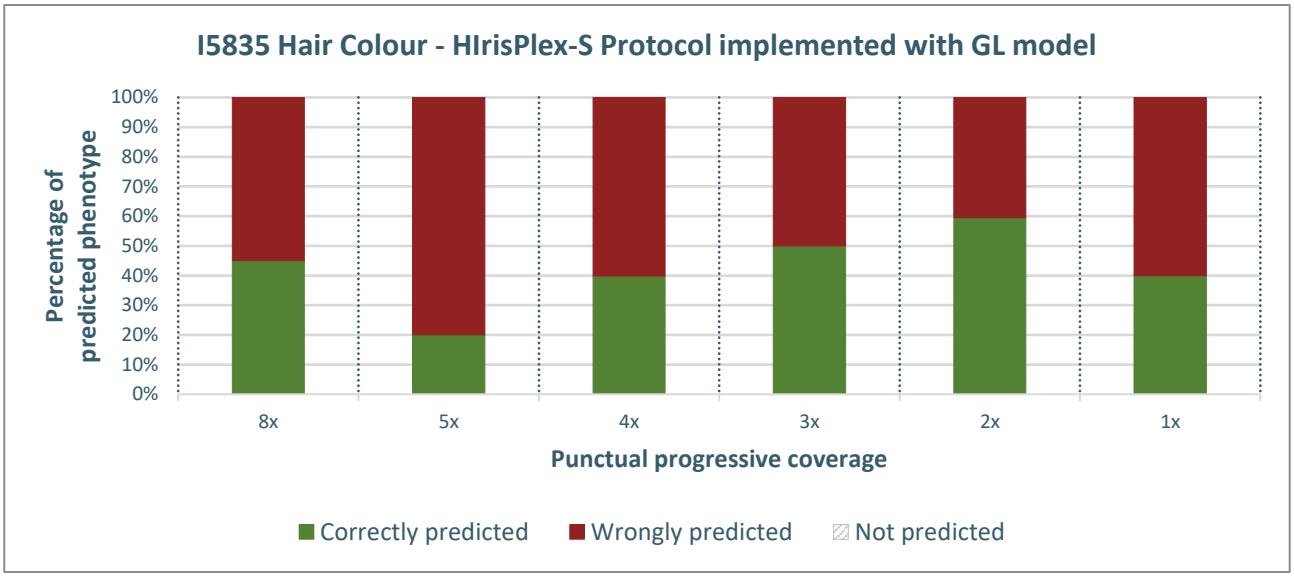


Figure 4.66 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

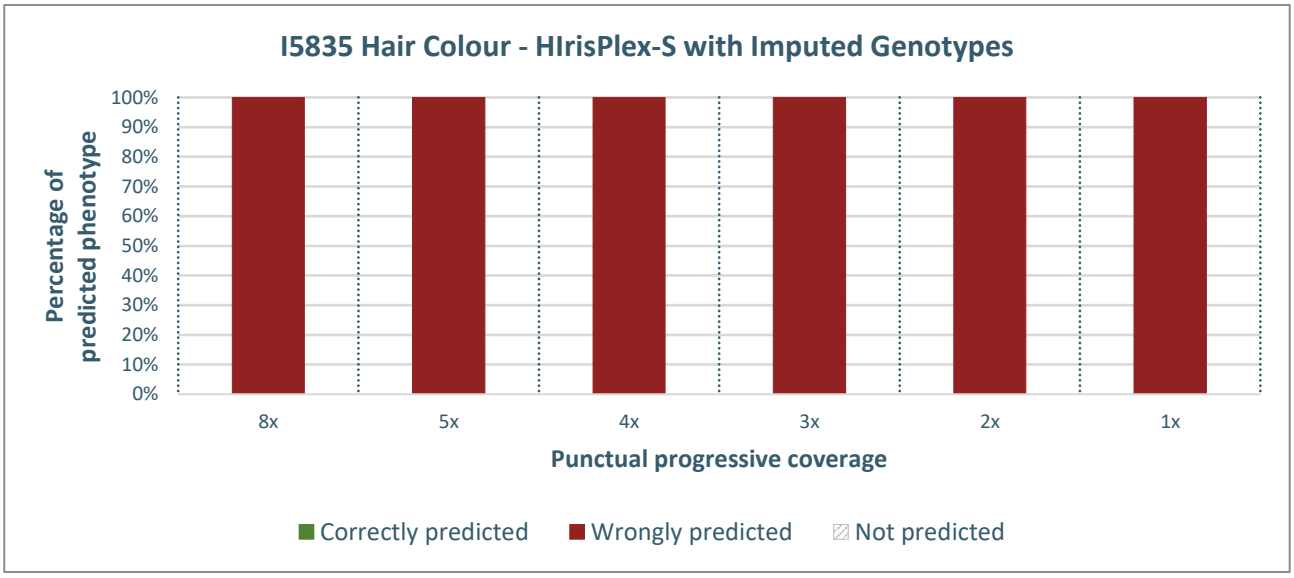


Figure 4.67 - Percentage distribution of predicted phenotypes for the I5832 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

4.3.3 Skin colour prediction

The results for skin colour phenotypic inference follow the order and criteria explained in the beginning of chapter 4.

Initially, the graphs display the percentages of prediction for each phenotypic class across 10 repetitions of downsampling (Figures 4.68, 4.69, 4.70, and 4.71). Subsequently, graphs illustrate the percentages of correct and incorrect predictions based on the phenotype considered “true”, namely intermediate skin (Figures 4.72, 4.73, 4.74, and 4.75).

- HIrisPlex-S Standard Protocol

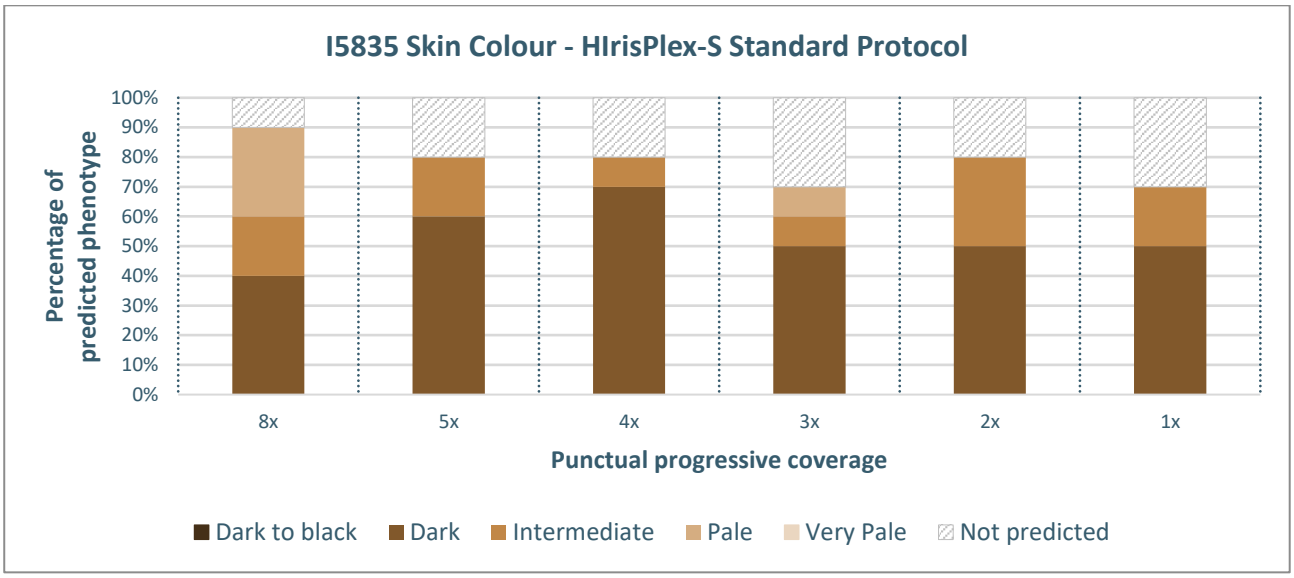


Figure 4.68 - Percentage distribution of predicted skin colour phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

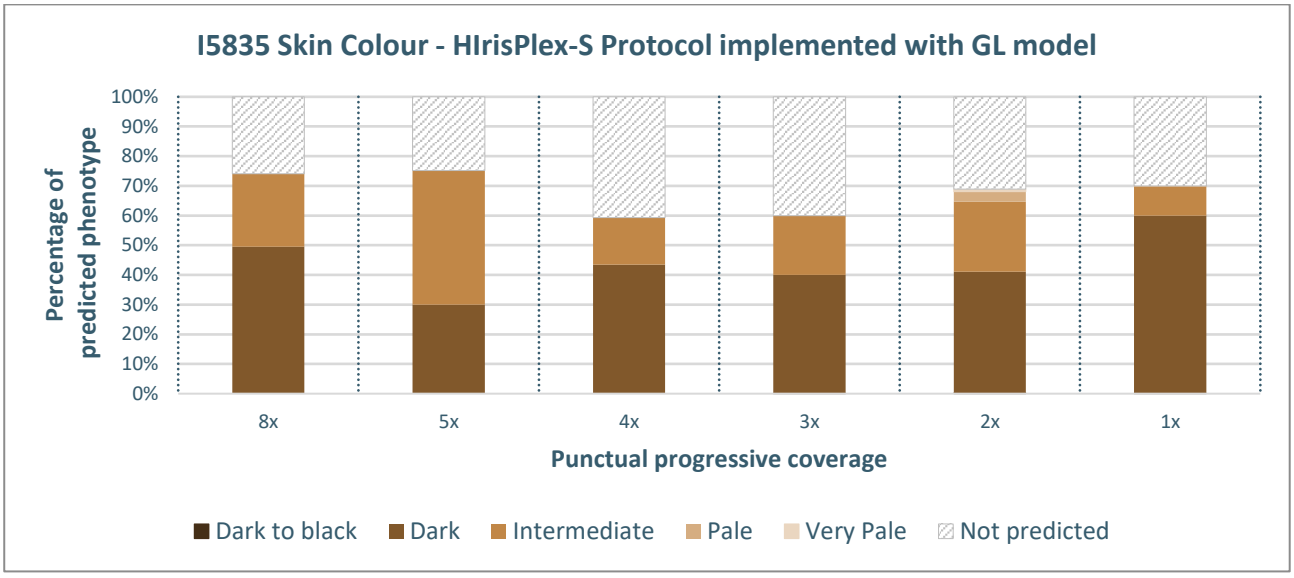


Figure 4.69 - Percentage distribution of predicted skin colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations. This graph depicts all obtained predictions.

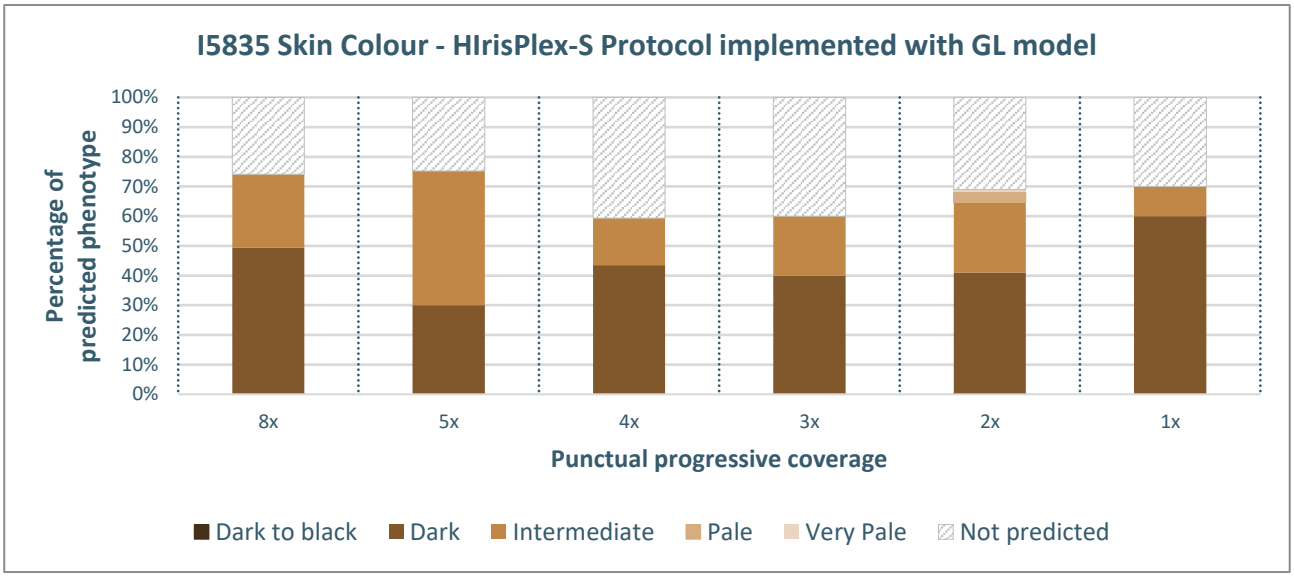


Figure 4.70 - Percentage distribution of predicted skin colour phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different sequencing coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

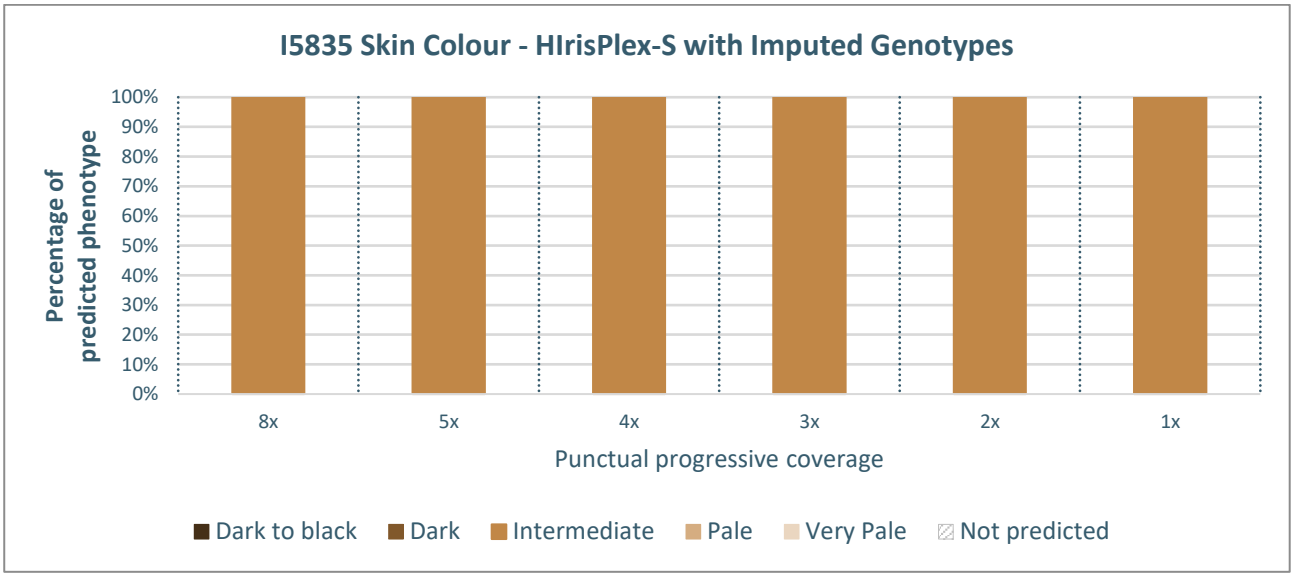


Figure 4.71 - Percentage distribution of predicted skin colour phenotypes for the I5832 sample using imputed genotypes. The x-axis represents the coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over 10 iterations.

The subsequent bar graphs display the percentages of correct and incorrect phenotypic predictions based on the presumed true phenotype, derived from high-coverage data (Figure 4.72, 4.73, 4.74, and 4.75). The graphs are arranged in the following order: HIrisPlex-S Standard Protocol, HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model, and phenotypic predictions derived from imputed data.

- HIrisPlex-S Standard Protocol

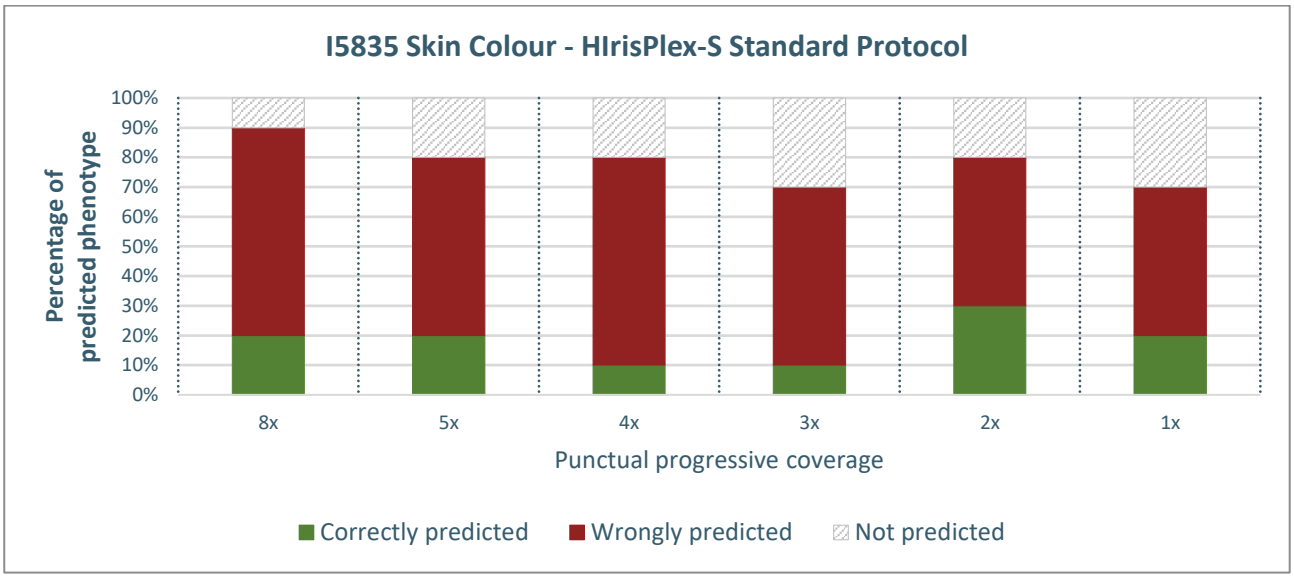


Figure 4.72 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Standard Protocol across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

- HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model

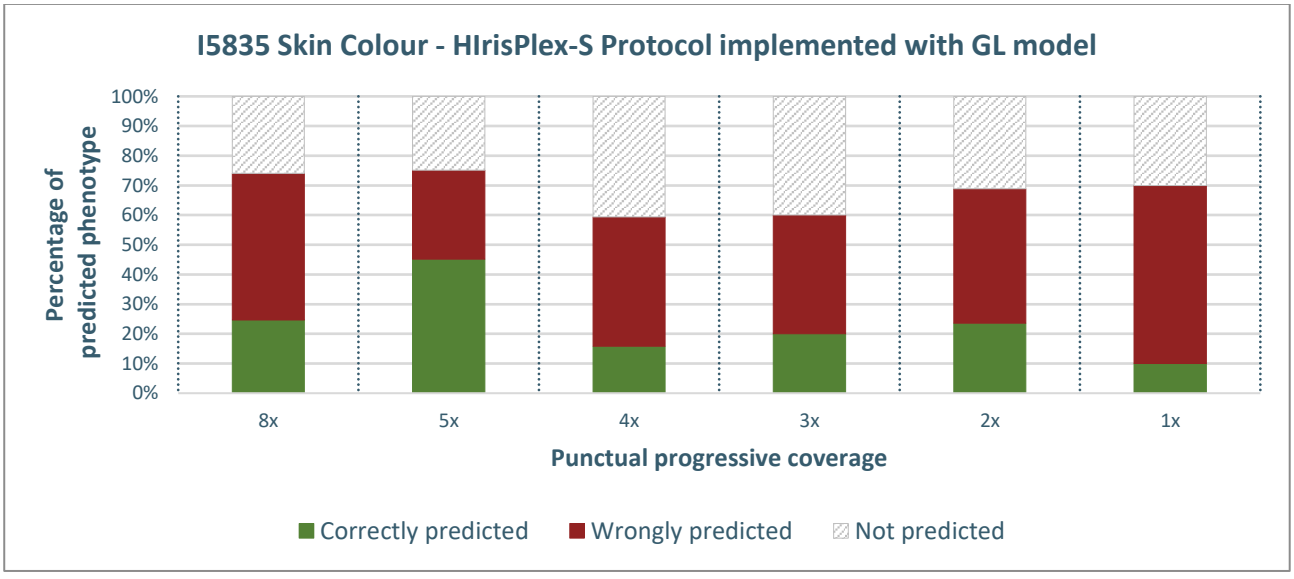


Figure 4.73 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. This graph depicts all obtained predictions.

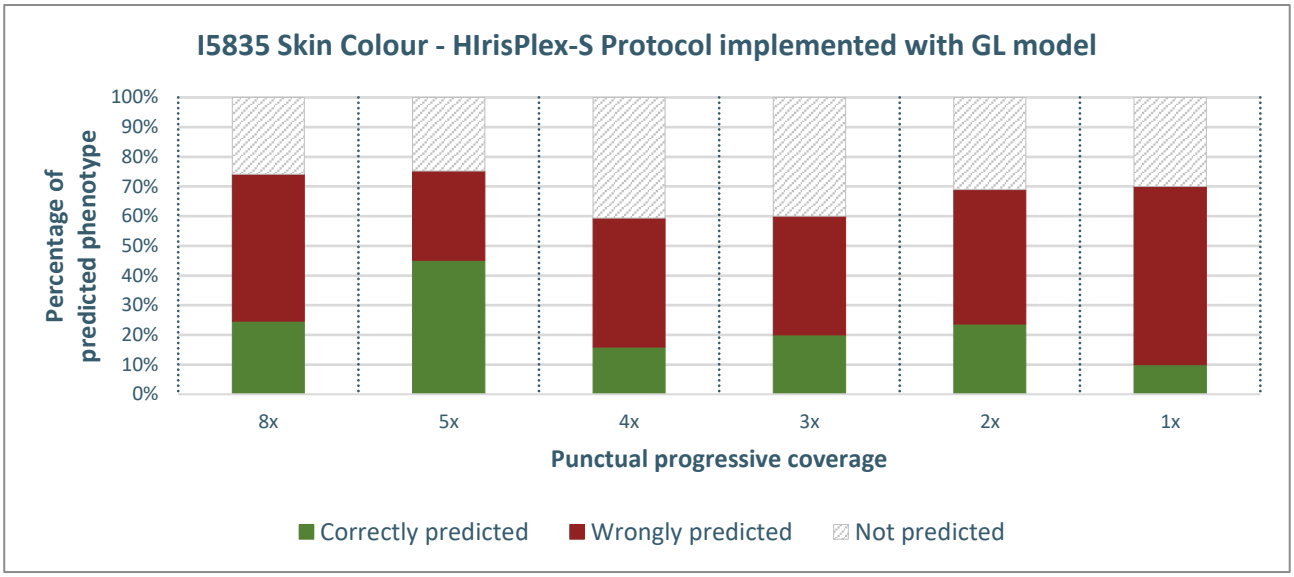


Figure 4.74 - Percentage distribution of predicted phenotypes for the I5832 sample using the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model across different coverage levels. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance. A 90% threshold is applied: if a phenotype reaches at least 900 predictions out of 1000, it is considered the sole predicted phenotype, with only its values retained. If the threshold is not met, values for all phenotypes are reported.

- Phenotypic predictions from imputed data

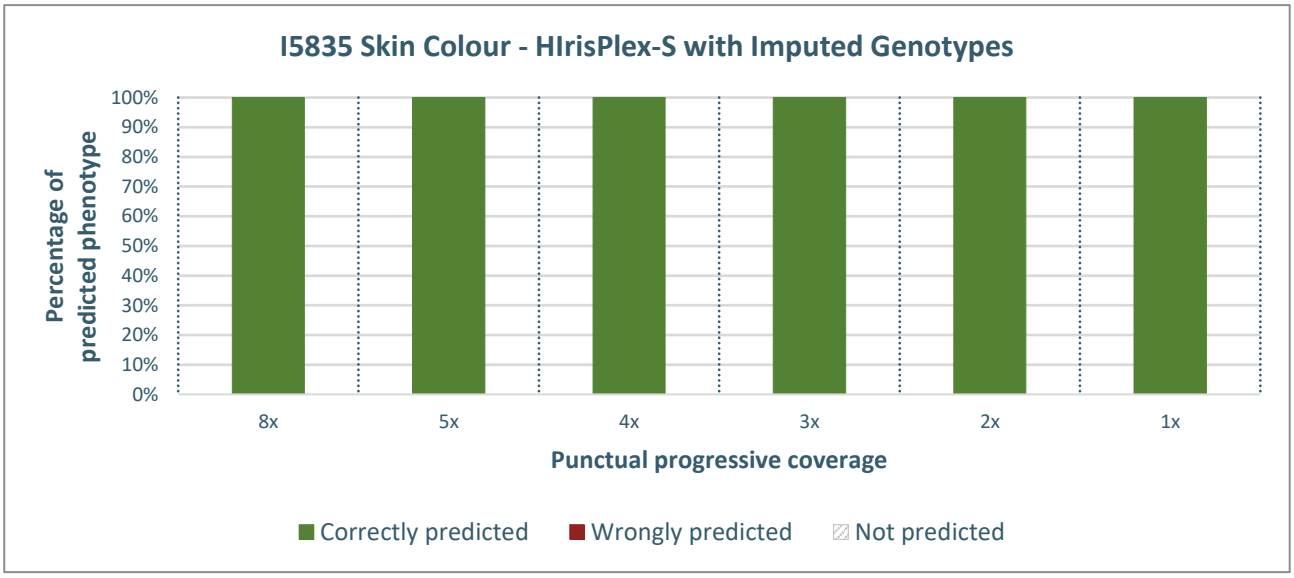


Figure 4.75 - Percentage distribution of predicted phenotypes for the I5832 sample using imputed genotypes. The x-axis represents coverage levels, while the y-axis shows the percentage of times each phenotype was predicted over ten iterations. Correct predictions are shown in green, and incorrect predictions are shown in red, highlighting the model's performance.

The results obtained for the Bronze Age sample I5832 demonstrate robust phenotypic prediction for eye colour across all 3 models (Figures 4.52, 4.53, and 4.55). In the context of hair colour predictions (Figures 4.60, 4.61, and 4.63), the HIrisPlex-S Standard Protocol outperforms both the HIrisPlex-S

Protocol Implemented with a Genotype Likelihoods Model and the imputation workflow. Specifically, the HIrisPlex-S Standard Protocol exhibits an inaccuracy of 20% at 3x coverage, 30% at 2x coverage, and 50% at 1x coverage. In contrast, the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model consistently shows inaccuracies of 55% at 8x coverage, 80% at 5x coverage, and an average of 50% from 3x coverage to 1x coverage. The application of a 90% threshold slightly improves these results. The conventional workflow for low coverage missing data shows 100% incorrect phenotypic predictions at all coverage levels. It is important to note that a major error in both the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model and the imputation workflow is the confusion between the dark/brown black and brown/dark brown phenotypes, which are both very dark phenotypic classes.

For skin colour predictions (Figures 4.68, 4.69, and 4.71), the HIrisPlex-S Protocol with Imputed Genotypes outperforms both the HIrisPlex-S Standard Protocol and the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model, achieving 100% correct phenotypic inferences. In contrast, the HIrisPlex-S Standard Protocol shows a higher error rate, with inaccuracies peaking at 70% at 8x coverage, decreasing to 60% at 5x and 3x, 70% at 4x, and 50% at 2x and 1x. The rate of "*not predicted*" values ranges from 10% at 8x coverage, 20% at 5x, 4x, and 2x coverage, up to 30% at 3x and 1x coverage. The HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model demonstrates improved performance relative to the HIrisPlex-S Standard Protocol, with incorrect phenotypic predictions of 49% at 8x coverage, 30% at 5x coverage, 43% at 4x coverage, 40% at 3x coverage, 45% at 2x coverage, and 60% at 1x coverage. The "*Not Predicted*" values are higher compared to the HIrisPlex-S Standard Protocol because some of the incorrect predictions made by the HIrisPlex-S Standard Protocol are classified as "*Not Predicted*" by the HIrisPlex-S Protocol implemented with a Genotype Likelihoods Model.

4.4 General considerations

The evaluation of phenotypic prediction methods for eye colour, hair colour, and skin colour across Palaeolithic, Mesolithic, and Bronze Age samples reveals several key findings.

Generally, reconstructing eye colour phenotypes is more straightforward than reconstructing hair and skin colour phenotypes. This is partly because eye colour reconstruction relies on significantly fewer SNPs (only 6), and with just one SNP (rs12913832), it is possible to achieve a phenotypic prediction with 68% probability. The CC genotype corresponds to a blue eye colour phenotype, whereas

the TT and CT genotypes are strongly associated with a brown iris coloration(Kayser *et al.*, 2008; Eiberg *et al.*, 2008; Walsh *et al.*, 2011).

In contrast, the phenotypic prediction of hair and skin colour is much more sensitive to allelic variation, with hair colour inference being particularly challenging. This is especially true for distinguishing between dark hair colour phenotypes, such as dark/brown black, brown/dark brown black, and black, as well as due to the significant impact that only a few SNPs can have on predicting red or blonde hair phenotypes (Branicki *et al.*, 2011).

Regarding the three methods tested, the HIrisPlex-S Standard Protocol generally demonstrates higher accuracy in hair colour predictions, particularly for the Bronze Age sample, but exhibits significant errors at lower and intermediate coverage levels for eye and skin colour predictions. Specifically, the method begins to lose robustness in phenotypic inference at coverage levels below 8x.

The HIrisPlex-S Protocol with Imputed Genotypes demonstrates exceptional performance in skin colour predictions for the Bronze Age sample, achieving 100% accuracy; however, it consistently underperforms in hair and eye colour predictions, especially at lower coverage levels, where it shows high error rates. The stability of the imputation-based method is influenced not only by sample coverage, where lower coverage complicates haplotype matching between the sample and the reference panel, but also by the age of the sample. When applied to samples with ancestry traceable to reference panels composed of modern human genomes dating from the Neolithic period onward (Hui *et al.*, 2020; Sousa da Mota *et al.*, 2023), this method proves to be relatively robust, even though it is computationally and time-intensive. For samples predating the Neolithic period, the method exhibits reduced predictive stability.

The model developed in this thesis utilizes a probabilistic approach, incorporating all 10 possible genotypes along with their respective probabilities and increasing the number of phenotypic predictions to 1000, as opposed to a single prediction. This model demonstrates enhanced robustness, particularly in predicting eye and skin colour, exhibiting greater accuracy and lower error rates compared to other methods. By implementing a 90% confidence threshold, the model significantly reduces errors arising from stochastic associations among genotypes with lower posterior probabilities and minimises the risk of erroneous predictions. Only predictions with a confidence level of 90% or higher are considered definitive, while those below this threshold are flagged as insufficiently certain, thus avoiding reliance on less reliable or stochastic data. This approach improves overall accuracy by reducing errors from less reliable predictions and still reporting lower-probability predictions

with indications of uncertainty. Additionally, it allows for the consideration of all possible phenotypes that may characterize a given sample. Unlike other methods that may fail or produce incorrect phenotypes, the HIrisPlex-S Protocol, when implemented with a Genotype Likelihoods Model, can present the phenotype of interest, even at a low percentage, alongside other possible phenotypes, thereby providing an assessment of the reliability of the phenotypic prediction.

Consequently, this method provides both a more robust phenotypic inference, derived from a genotyping approach that accounts for call uncertainty and a greater number of phenotypic predictions supporting the final inference, and a warning when the phenotype cannot be determined with certainty.

4.5 User Manual and Dataset Application

The results obtained from the testing phase demonstrate that sample coverage significantly affects phenotypic predictions. The HIrisPlex-S Standard Protocol is relatively robust at medium-to-high coverage levels but shows diminished predictive consistency at coverage levels of 8x or lower. In contrast, the phenotypic inference model developed in this thesis offers more reliable predictions at lower coverage levels and identifies cases where predictions may be inaccurate.

This has led to the establishment of optimized guidelines based on the analysis of coverage across the 41 positions of the HIrisPlex-S system. The recommended protocol involves using the HIrisPlex-S Standard Protocol when a coverage of at least 8x is achieved across all HIrisPlex-S positions. Otherwise should be used the HIrisPlex-S Protocol implemented with the Genotype Likelihoods Model developed in this study.

The following results present the application of these guidelines to a dataset comprising Eurasian human samples selected from the Palaeolithic period through to the Iron Age. Details of all samples included in the dataset are available in the supplementary materials folder at the provided link: *Eurasian_Dataset.xlsx*.

To provide a statistically supported view of the current phenotypes during these historical periods, the results include only those samples that have exceeded the 90% threshold as supported by the new predictive method. Furthermore, given the difficulty in distinguishing dark colours for the phenotypic trait of hair, once the 1000 phenotypic predictions were obtained, phenotypic macro-groups

were identified (Table 4.5). To standardize the data, this approach was also applied to the phenotypes of eyes (Table 4.4) and skin (Table 4.6).

Eye phenotype colour	**Eye phenotypic group**
Brown	Dark
Intermediate	Intermediate
Blue	Light

Table 4.4 – Classification of eye colour phenotypes into standardized groups.

Hair phenotype colour	**Hair phenotypic group**
Black	Dark
Dark brown/black	
Brown/dark brown	
Brown	Intermediate
Dark blond	
Blond	Light
Red	Red

Table 4.5 - Classification of hair colour phenotypes into standardized groups.

Skin phenotype colour	**Skin phenotypic group**
Dark to black	Dark
Dark	
Intermediate	Intermediate
Pale	Light
Very Pale	

Table 4.6 - Classification of skin colour phenotypes into standardized groups.

The results obtained from applying the user manual are presented chronologically, from the earliest to the most recent period, and are categorized according to predicted phenotypic traits. These results are illustrated with pie charts that display the percentages of each predicted phenotype. Additionally, the data are overlaid on a geographical map representing the sampling locations of the samples used.

4.5.1 Palaeolithic inference of phenotypic traits

The total number of samples from the Palaeolithic period is 12, of which one is Ust'-Ishim sample. Of these 12 samples, 11 provide results for eye colour phenotypic inference (Figure 4.76), 10 for hair colour (Figure 4.77), and 12 for skin colour (Figure 4.78).

- Eye colour phenotypes

Figure 4.76 - Geographical Distribution of Predicted Eye Phenotypes for Palaeolithic period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

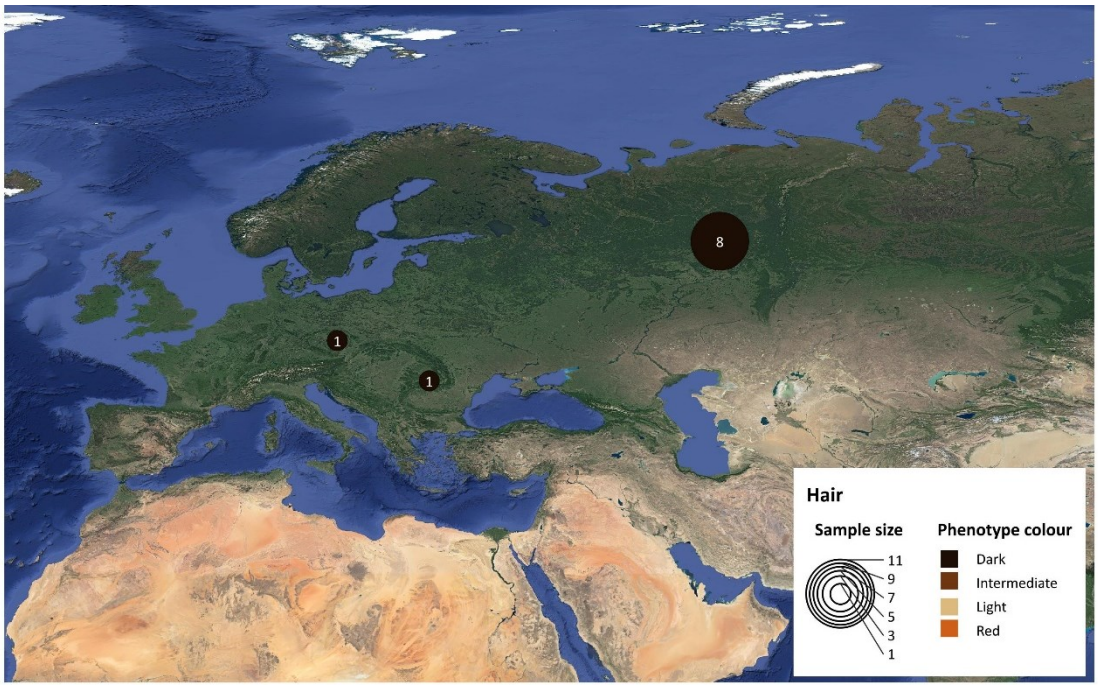


Figure 4.77 - Geographical Distribution of Predicted Hair Phenotypes for Palaeolithic period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

Figure 4.78 - Geographical Distribution of Predicted Skin Phenotypes for Palaeolithic period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The analysed samples from the Palaeolithic period reveal a dark phenotypic profile for all 3 traits under study. The sole exception is a Russian sample, Kostenki 14, dated to between 38,700 and 36,200 years ago, which exhibits an intermediate skin colour (Seguin-Orlando *et al.*, 2014).

Among the analysed samples is included PM1, characterized by brown eyes, black hair, and dark to black skin. This prediction contrasts with the one made in the original article by Svensson in 2021, where PM1 was found to have an intermediate skin colour. This discrepancy is due to the predictive method used; the original article employed the HIrisPlex-S Standard Protocol, whereas this project applied the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model. Although the sample is covered at the 41 HIrisPlex-S positions, the minimum coverage for the sample is 5x at 2 positions (Svensson *et al.*, 2021).

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Palaeolithic*.

4.5.2 Mesolithic inference of phenotypic traits

The total number of samples from the Mesolithic period is 66, of which one is SF12 sample. Of these 66 samples, 35 provide results for eye colour phenotypic inference (Figure 4.79), 63 for hair colour (Figure 4.80), and 53 for skin colour (Figure 4.81).

- Eye colour phenotypes

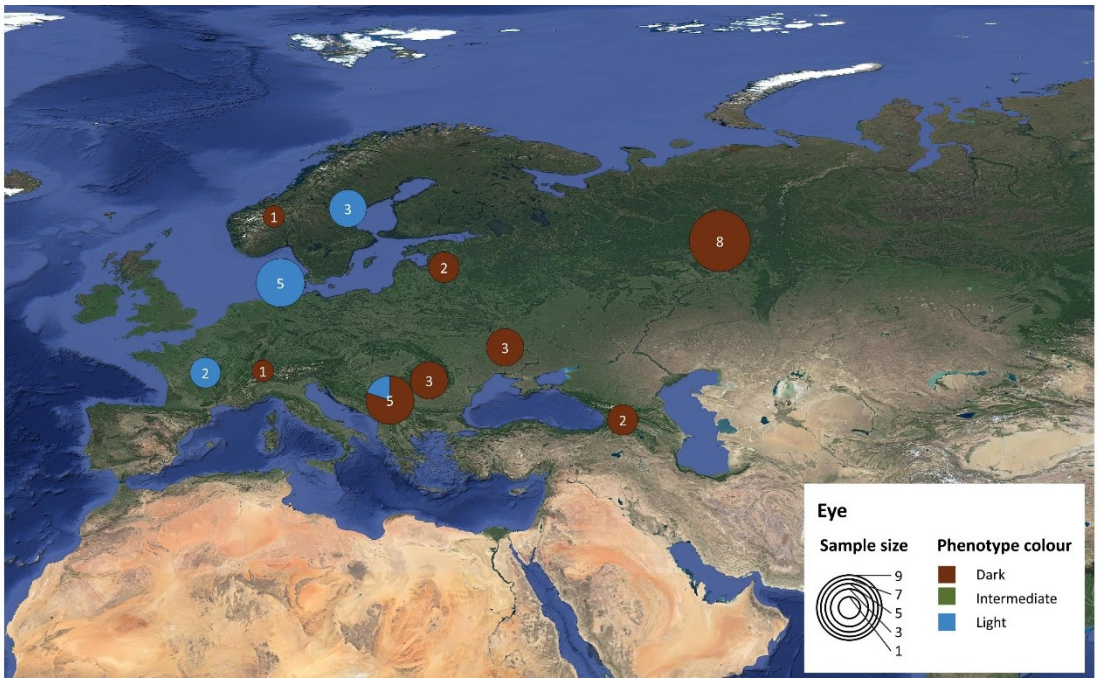


Figure 4.79 - Geographical Distribution of Predicted Eye Phenotypes for Mesolithic period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

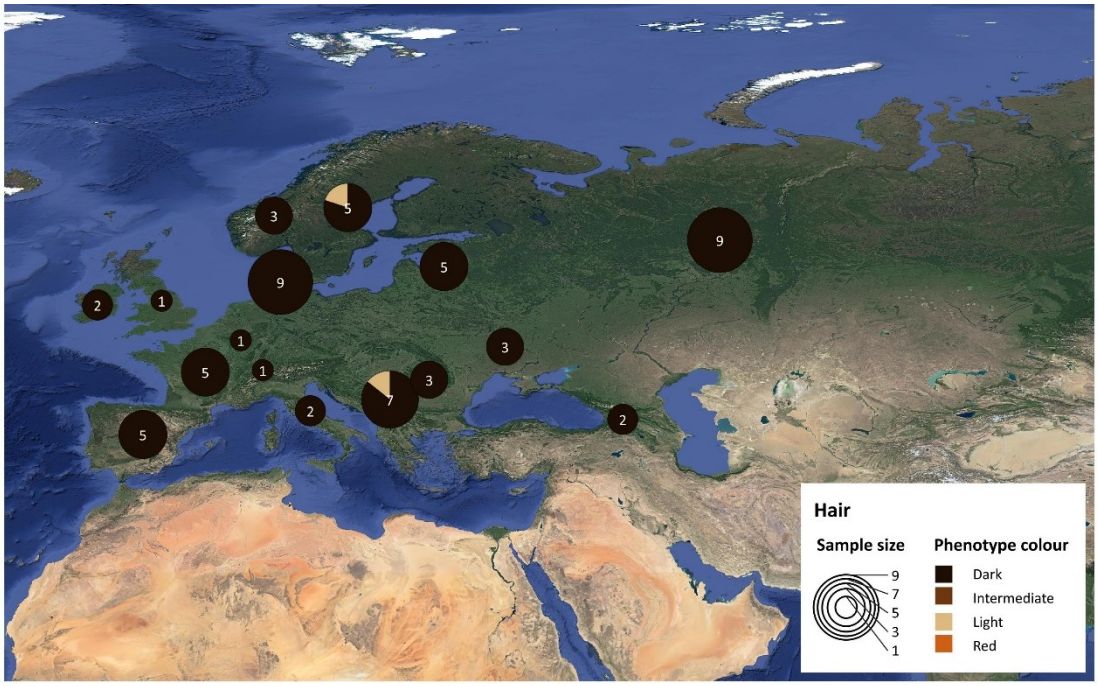


Figure 4.80 - Geographical Distribution of Predicted Hair Phenotypes for Mesolithic period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

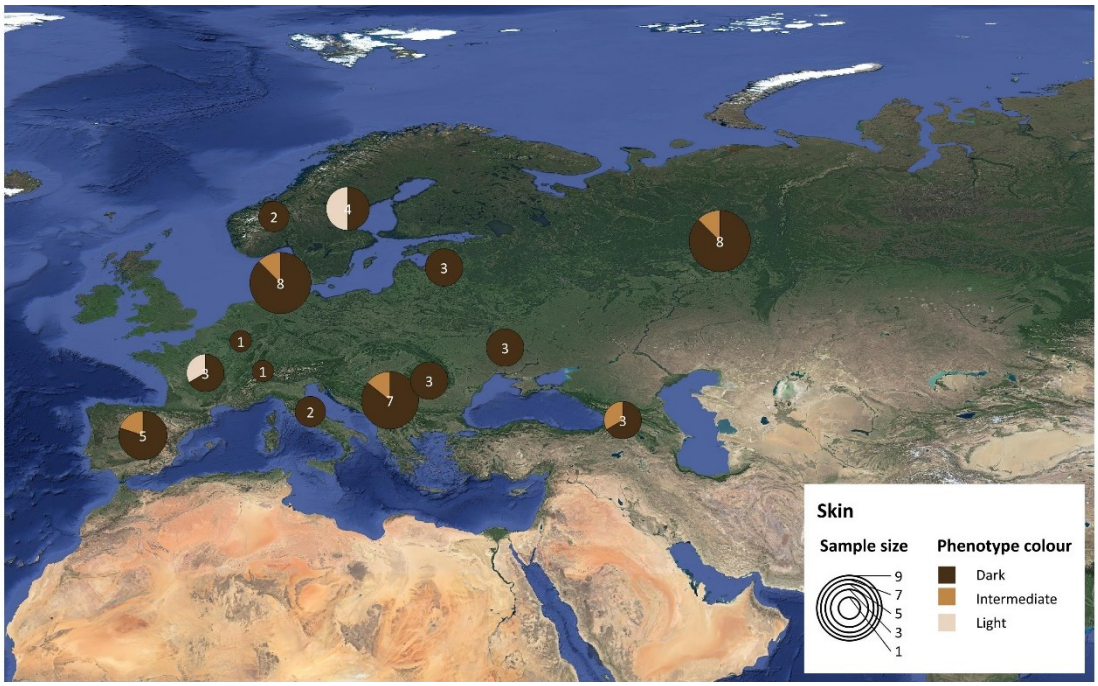


Figure 4.81 - Geographical Distribution of Predicted Skin Phenotypes for Mesolithic period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The phenotypic analysis of Mesolithic samples exhibits the emergence of light phenotypes. Regarding eye colour, the samples from Northern Europe, Denmark, Sweden, and Central-Western Europe, France, exclusively exhibit the light phenotype (11 samples), in contrast to the more eastern regions which instead display the dark phenotype (24 samples). It is noteworthy that Serbia, in Southeastern Europe, exhibits both phenotypes, with 20% blue eyes (1 sample) and 80% brown eyes (4 samples). Hair colour predominantly remains dark (56 samples), with the exception of 1 Swedish and 1 Serbian sample that show the light phenotype. Skin colour, however, displays a broader range of phenotypes: predominantly dark (43 samples), with regions in Europe also showing intermediate phenotypes (7 samples from Denmark, France, Georgia, Russia, Serbia, and Spain) and light phenotypes (3 samples from France and Sweden).

Among the analysed samples is included the Cheddar Man, with the only predicted phenotype being dark brown/black hair colour. The values for eye and skin colour did not exceed the thresholds of 0.7 and 0.5 imposed by HIrisPlex-S. In the original publication by Brace in 2019, the sample is reported to have most probably blue/green eyes, dark brown (possibly black) hair, and dark or dark to black skin (Brace *et al.*, 2019). The difference here lies in the predictive method, as the original study

used the HIrisPlex-S Standard Protocol, whereas this project applied the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model. The Cheddar Man sample has an average coverage of 2.14x, with 38 out of 41 HIrisPlex-S positions covered showing a minimum coverage 1x.

The La Braña sample, with an average coverage of 1.67x, has 33 out of 41 positions covered and a minimum coverage of 1x. In this thesis, applying the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model, it is found to have dark brown/black hair and intermediate skin. The phenotype for eye colour does not exceed the HIrisPlex-S threshold of 0.7. In the original study by Olalde in 2014, following the HIrisPlex-S Standard Protocol, the sample is reported to have dark skin, non-brown eyes, and dark or brown hair (Olalde *et al.*, 2014).

Dating back to this period is the first sample exhibiting blue eyes, blonde hair, and light skin, NEO27, a Scandinavian HG from Sweden who lived approximately 12,000 years ago (Allentoft, Sikora, Fischer, *et al.*, 2024).

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Mesolithic*.

4.5.3 Neolithic inference of phenotypic traits

The total number of samples from the Neolithic period is 132. Of these 132 samples, 93 provide results for eye colour phenotypic inference (Figure 4.82), 120 for hair colour (Figure 4.83), and 93 for skin colour (Figure 4.84).

- Eye colour phenotypes

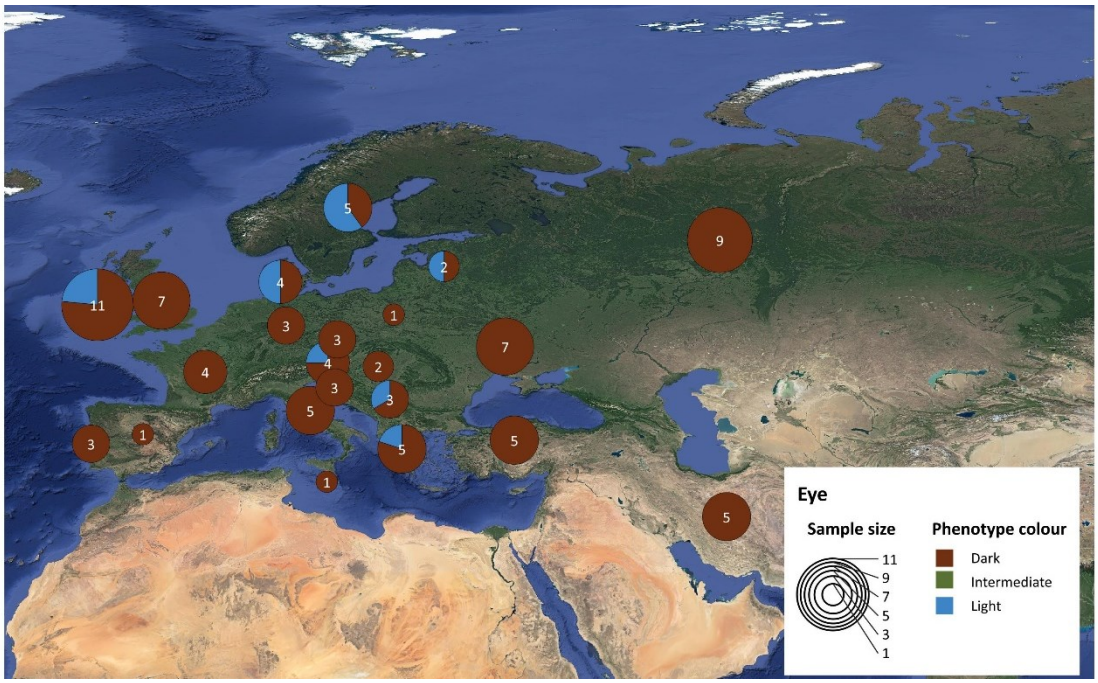


Figure 4.82 - Geographical Distribution of Predicted Eye Phenotypes for Neolithic period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

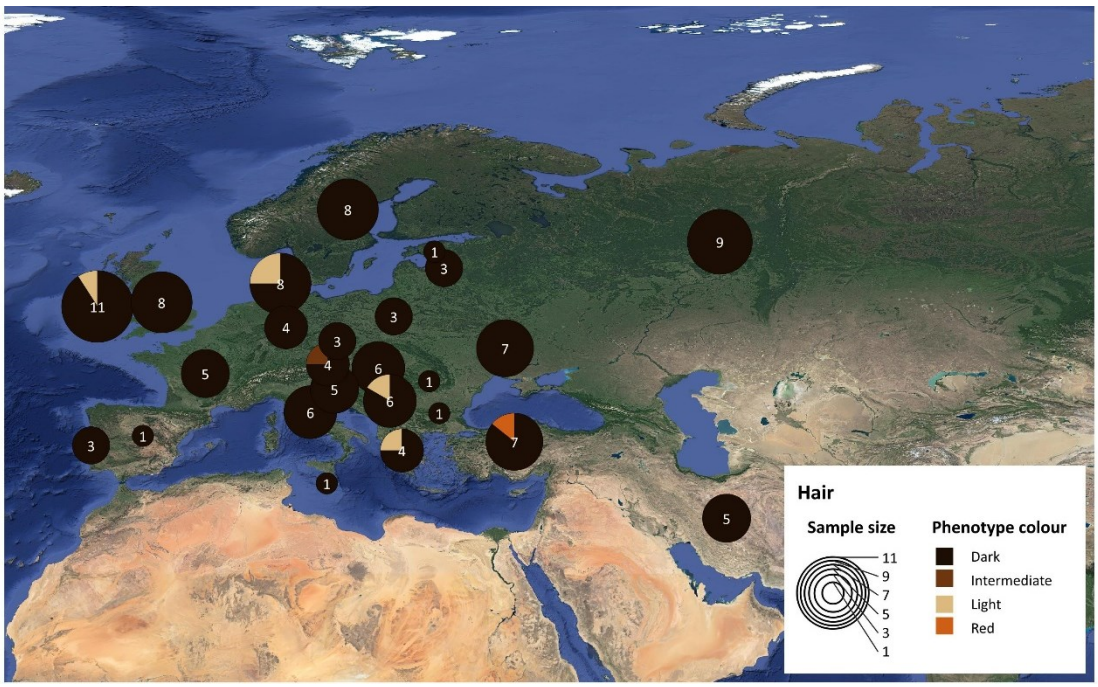


Figure 4.83 - Geographical Distribution of Predicted Hair Phenotypes for Neolithic period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

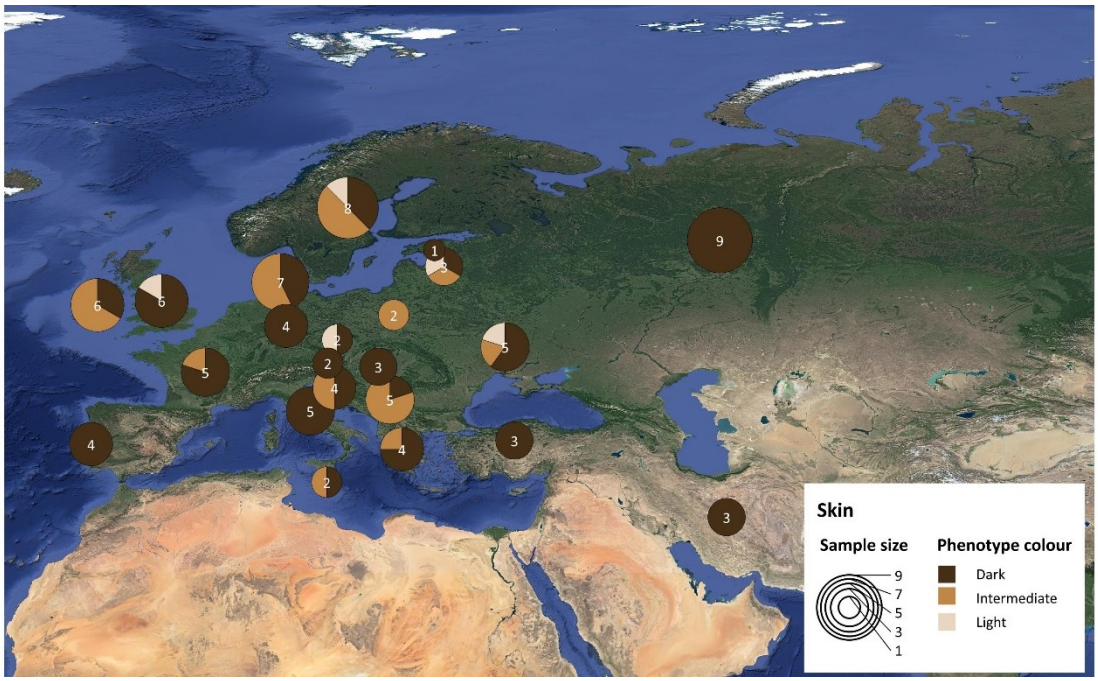


Figure 4.84 - Geographical Distribution of Predicted Skin Phenotypes for Neolithic period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The analysed samples from the Neolithic period indicate, regarding eye colour, a continued exclusive presence of the dark phenotype (81 samples) in the regions of Eastern Europe. This is now also observed in Western Europe, including France, which previously exhibited only the light phenotype, and in Western Asia. There are no longer regions with an exclusively light phenotype; however, both phenotypes can be observed in Central-Northern and Central-Eastern Europe (the light phenotype is present in 12 samples from Austria, Denmark, Greece, Ireland, Latvia, Serbia, and Sweden). Hair colour predominantly remains dark (113 samples), with the exception of 1 sample from Austria showing an intermediate phenotype and 5 samples from Central-Northern and Central-Eastern Europe (Denmark, Greece, Ireland, and Serbia). Additionally, 1 sample from Turkey displays red hair. Skin colour phenotype remains more variable, with regions in Europe (Portugal, Italy, Germany, Estonia, and Russia) and Western Asia (Iran and Turkey) exhibiting an exclusively dark phenotype, and regions showing either both dark and intermediate phenotypes (in 25 samples from Malta, Poland, Serbia, Sweden, and Ukraine) or light phenotypes (in 5 samples from Great Britain, Czech Republic, Sweden, and Ukraine).

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Neolithic*.

4.5.4 Copper Age inference of phenotypic traits

The total number of samples from the Copper Age period is 43. Of these 43 samples, 32 provide results for eye colour phenotypic inference (Figure 4.85), 34 for hair colour (Figure 4.86), and 29 for skin colour (Figure 4.87).

- Eye colour phenotypes

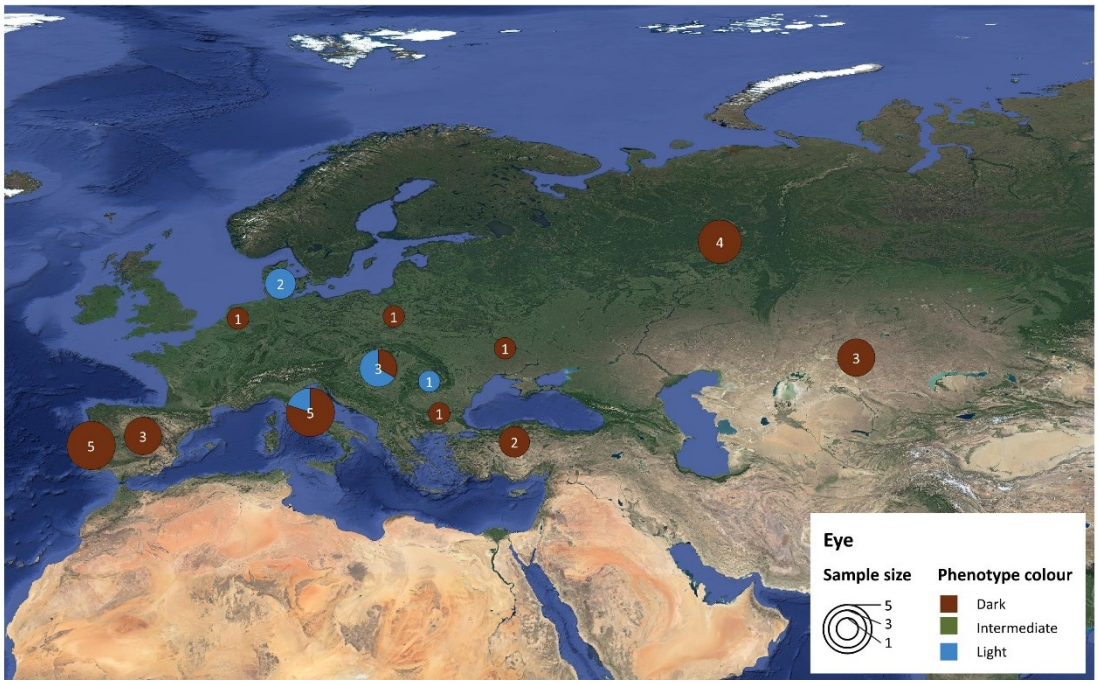


Figure 4.85 - Geographical Distribution of Predicted Eye Phenotypes for Copper Age period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

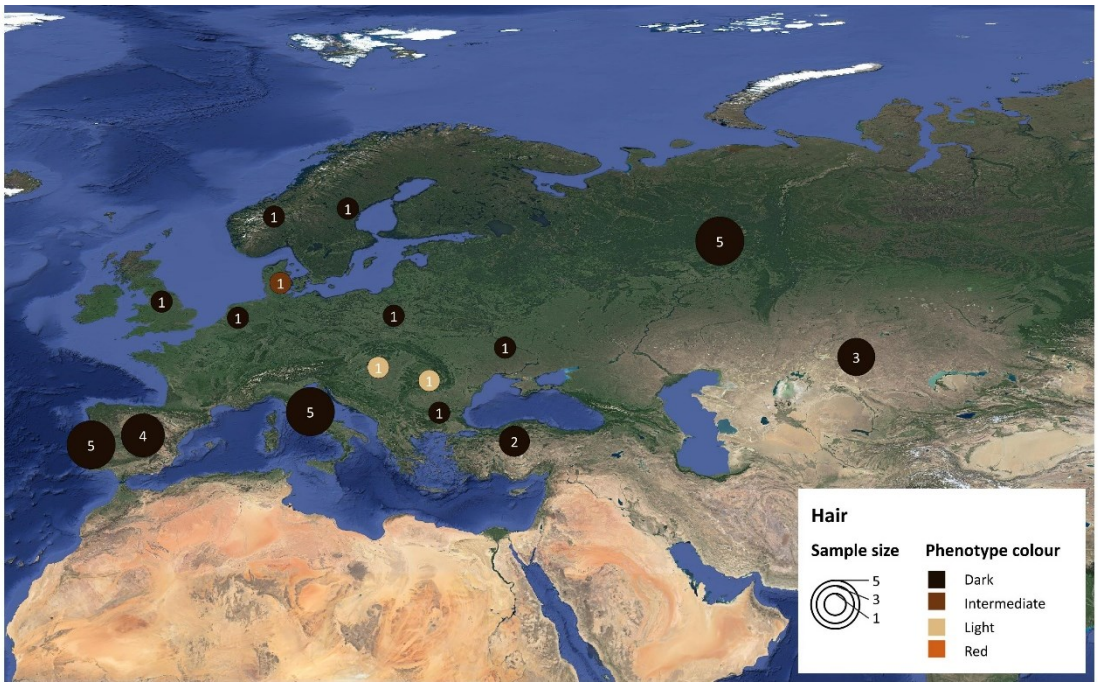


Figure 4.86 - Geographical Distribution of Predicted Hair Phenotypes for Copper Age period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

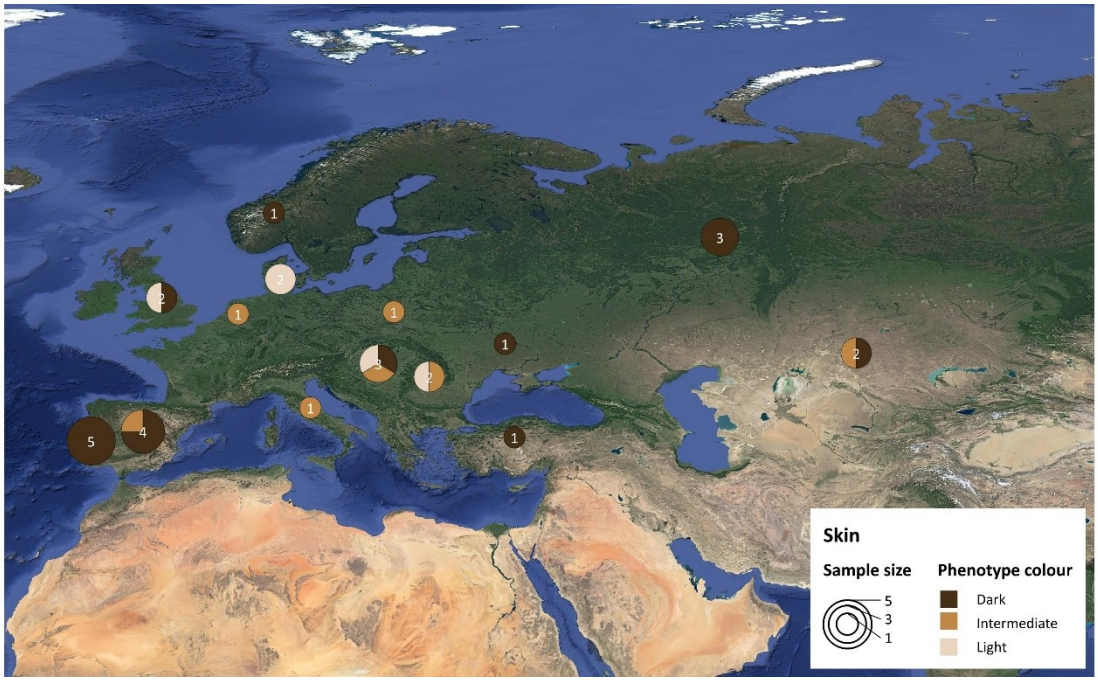


Figure 4.87 - Geographical Distribution of Predicted Skin Phenotypes for Copper Age period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The phenotypic analysis of Copper Age samples once again reveals a prevalence of dark eyes (26 samples), with the light phenotype present in Denmark, Hungary, Italy, and Romania (6 samples). Hair phenotype remains mostly dark, with only 1 sample showing intermediate hair colour (Denmark) and 2 samples exhibiting light hair colour (Hungary and Romania). Skin colour is also predominantly dark (17 samples) precisely in Eastern Europe, Iberian Peninsula, and Northern Europe, with intermediate skin tones observed in Spain, Kazakhstan, and Central Europe (7 samples from Hungary, Italy, the Netherlands, Poland, and Romania), while light skin colour is present in Denmark, Great Britain, Hungary, and Romania (5 samples). Notably, the Hungarian sample SZ1, dating back approximately 3000 years, displays all three light phenotypes (Amorim *et al.*, 2018).

Among the samples from this period is the Tyrolean Iceman, also known as "Ötzi", a man from approximately 5,300 years ago discovered in Val Senales (South Tyrol). The initial publication of his genome, dating back to 2012, reported an average coverage of 7.6x and inferred a phenotype of light skin and brown eyes (Keller *et al.*, 2012). In 2023, his genome was sequenced with an average coverage of 13.4x, resulting in inferred phenotypes of black hair and dark skin (Wang *et al.*, 2023). To infer these phenotypes, in both studies were analysed several pigmentation-associated SNPs from the UK Biobank GWAS. In this thesis, with a minimum coverage of 3x on the HIrisPlex-S positions, was applied the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model. The predicted phenotypes included dark eyes and hair, with the skin predicted to be 60% intermediate and 40% "*Not Predicted*".

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Copper Age*.

4.5.5 Bronze Age inference of phenotypic traits

The total number of samples from the Bronze Age period is 70, of which one is I5835 sample. Of these 54 samples, 63 provide results for eye colour phenotypic inference (Figure 4.88), 42 for hair colour (Figure 4.89), and 29 for skin colour (Figure 4.90).

- Eye colour phenotypes

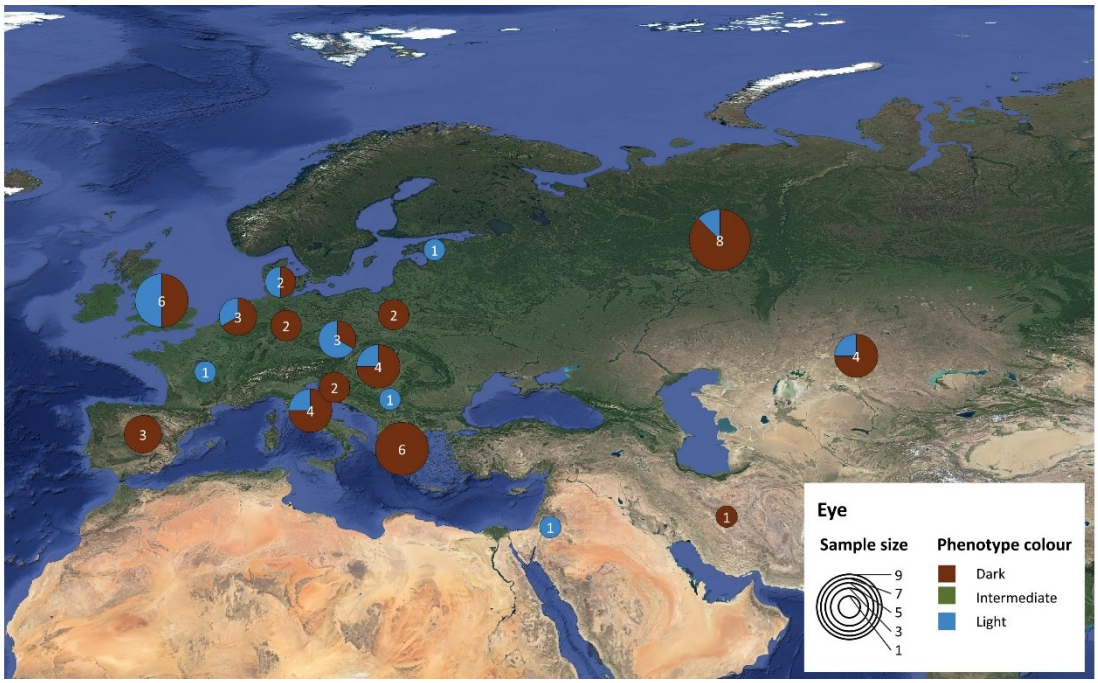


Figure 4.88 - Geographical Distribution of Predicted Eye Phenotypes for Bronze Age period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

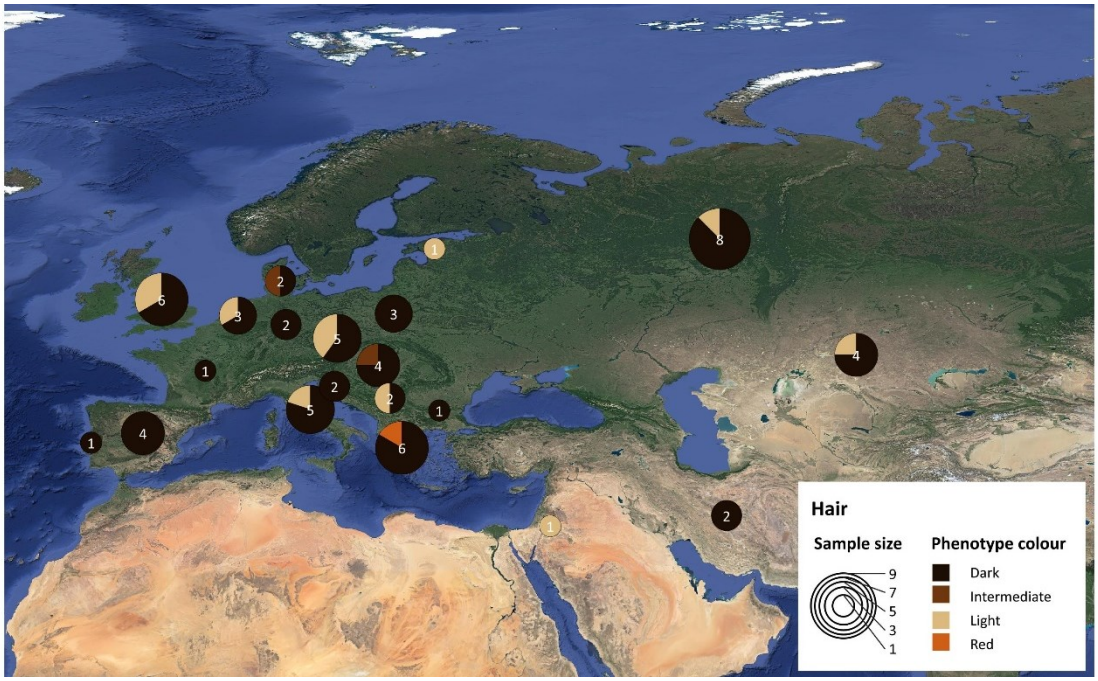


Figure 4.89 - Geographical Distribution of Predicted Hair Phenotypes for Bronze Age period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

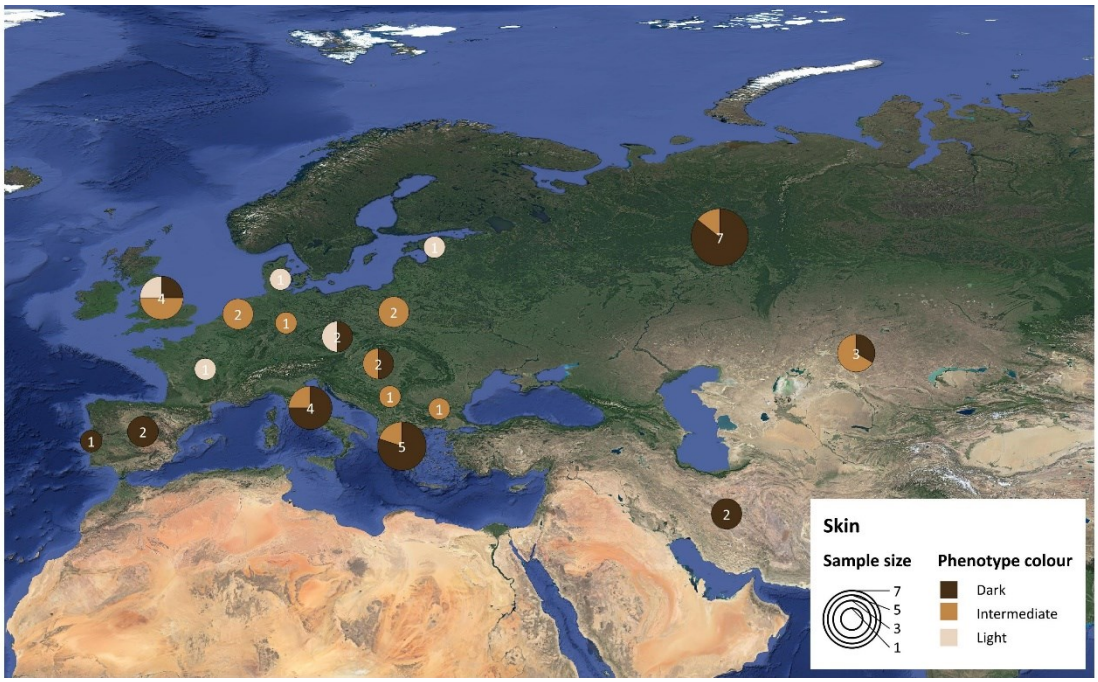


Figure 4.90 - Geographical Distribution of Predicted Skin Phenotypes for Bronze Age period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The analysed samples from the Bronze Age period indicate, with respect to the eyes colour, a continued prevalence of dark phenotypes throughout most of Europe and Asia (39 samples). Light phenotypes (15 samples), although increasing, are still primarily found in Northern and Central-Eastern Europe, but are also emerging in new geographical regions such as Russia, Jordan, and Kazakhstan. Concerning hair colour phenotype, dark phenotypes remain predominant in most of Europe and Asia (49 samples), with intermediate phenotypes present in 2 samples from Denmark and Hungary. However, there is a noted increase in light phenotypes (11 samples) in Northern Europe, Central-Eastern Europe, and their reappearance in Italy, Russia, Jordan, and Kazakhstan. One sample from Greece exhibits red hair. The analysis of skin colour phenotype also shows a continued prevalence of dark phenotypes (22 samples), especially in Western Europe, Southern Europe, and Southern Asia, with an increase in intermediate phenotypes (15 samples) in Central Europe and Central-Eastern Europe, as well as their appearance in Russia and Kazakhstan. The light phenotype, less common, is present in 5 samples from the Czech Republic, Denmark, Estonia, France, and Great Britain.

During this period, there is also an increase in the co-occurrence of blue eyes, blonde hair, and light skin. This trend is observed in a progression from a Swedish Mesolithic sample and a Hungarian Copper Age sample to 4 samples: I7198 from the Czech Republic (Olalde *et al.*, 2018), EKA1 from

Estonia (Malmström *et al.*, 2019), I2445 from England (Olalde *et al.*, 2018), and R11105 from Italy (Moots *et al.*, 2023).

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Bronze Age*.

4.5.6 Iron Age inference of phenotypic traits

The total number of samples from the Iron Age period is 25. Of these 25 samples, 15 provide results for eye colour phenotypic inference (Figure 4.91), 19 for hair colour (Figure 4.92), and 11 for skin colour (Figure 4.93).

- Eye colour phenotypes

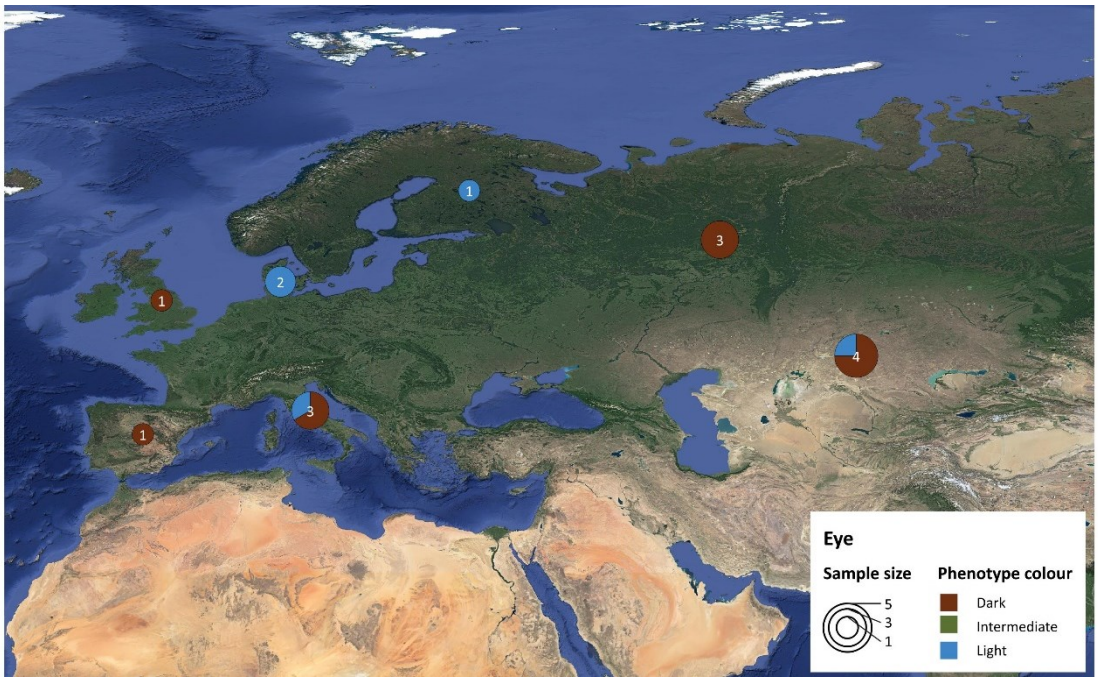


Figure 4.91 - Geographical Distribution of Predicted Eye Phenotypes for Iron Age period. The map illustrates the geographical distribution of predicted eye phenotypes across various regions. Pie charts represent the percentage distribution of eye phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Hair colour phenotypes

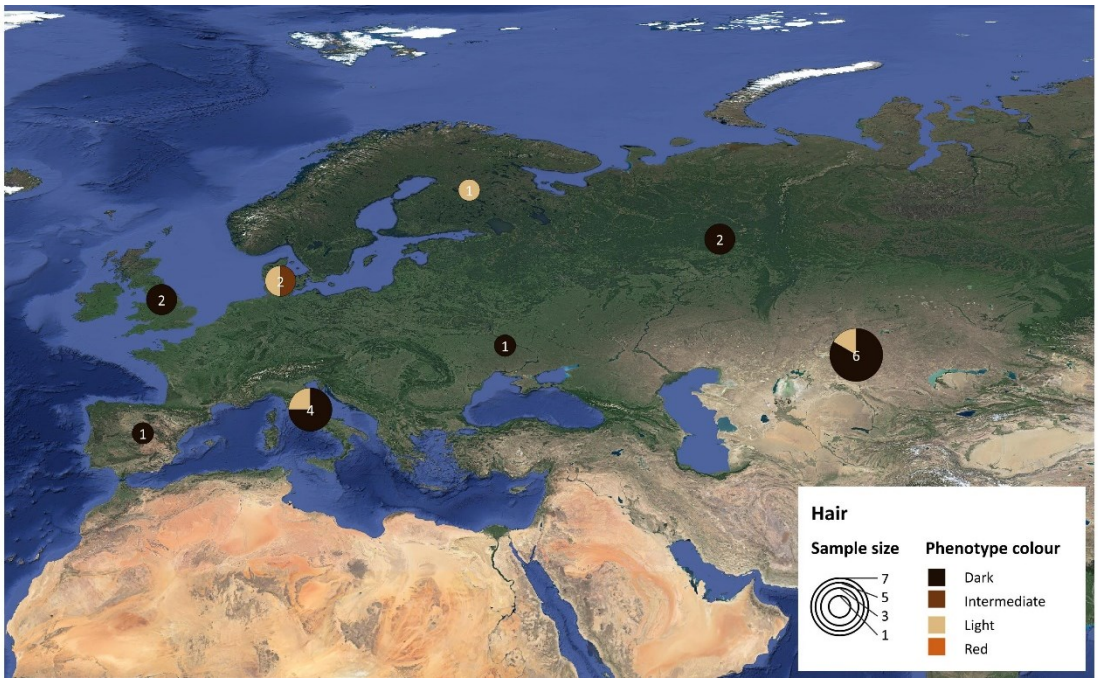


Figure 4.92 - Geographical Distribution of Predicted Hair Phenotypes for Iron Age period. The map illustrates the geographical distribution of predicted hair phenotypes across various regions. Pie charts represent the percentage distribution of hair phenotypes, with each segment coloured according to the phenotype (dark, intermediate, light, and red). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

- Skin colour phenotypes

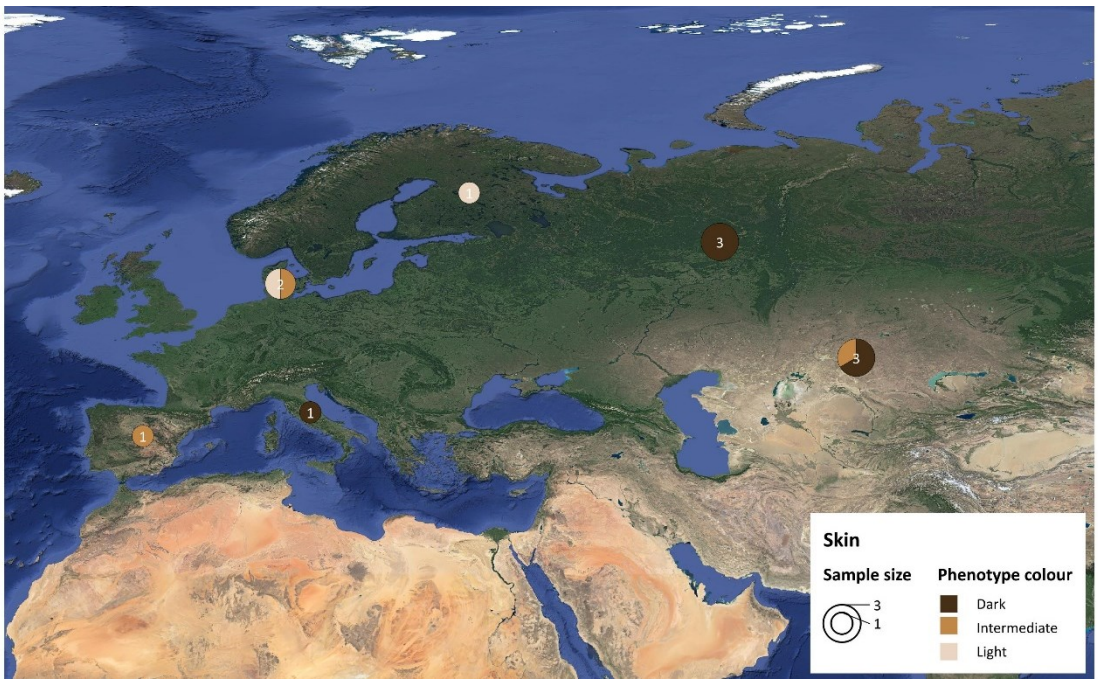


Figure 4.93 - Geographical Distribution of Predicted Skin Phenotypes for Iron Age period. The map illustrates the geographical distribution of predicted skin phenotypes across various regions. Pie charts represent the percentage distribution of skin phenotypes, with each segment coloured according to the phenotype (dark, intermediate, and light). The size of the pie charts corresponds to the number of samples collected at each site, as indicated by the legend in the lower right corner.

The phenotypic analysis of Iron Age samples indicates that the dark eye phenotype (10 samples), so far predominant, is present in Western and Eastern Europe, while the light eye phenotype (5 samples) is found in Northern Europe, Italy, and Kazakhstan. Hair remains predominantly dark throughout Europe and Asia (14 samples), with 1 intermediate phenotype sample in Denmark and 4 light phenotype samples in Denmark, Finland, Italy, and Kazakhstan. Skin colour analysis shows the dark phenotype (6 samples) in Eastern Europe, Asia, and Italy, the intermediate phenotype (3 samples) in Kazakhstan, Denmark, and for the first time in Spain, and the light phenotype (2 samples) still present in Northern Europe.

In Northern Europe, there are 2 samples with blue eyes, blonde hair, and light skin: VK521 from Denmark (Margaryan *et al.*, 2020) and DA236 from Finland (Sikora *et al.*, 2019).

Details of all samples, including information on genomic coverage at 41 HIrisPlex-S positions, the prediction model utilized for phenotypic inference, and the resulting phenotypic predictions, both those meeting and those falling below the threshold, are available in the supplementary materials folder at the provided link: *Iron Age*.

4.6 General considerations

This phase of phenotypic prediction of the Eurasian dataset primarily facilitated an additional comparison between the HIrisPlex-S Standard Protocol and the new method developed in this thesis project. For instance, with regard to the Mesolithic samples La Braña and Cheddar Man, both of which have an average coverage around 2, the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model indicates that coverage is not sufficient to predict all phenotypes. In the case of La Braña, eye colour cannot be predicted, while for Cheddar Man, information can only be retrieved concerning hair colour. An even more intriguing case is that of the Copper Age sample, Ötzi. Initial inferences suggested a light skin phenotype; however, recent studies report a much darker phenotype. The HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model developed in this thesis, however, issues a warning, stating that there is a 60% probability that the skin phenotype is intermediate, with a 40% probability of it being "*Not Predicted*". This leads to the conclusion that it is not possible to predict Ötzi's skin colour with certainty, which could account for the discrepancy in skin colour between the two publications (Keller *et al.*, 2012; Olalde *et al.*, 2014; Brace *et al.*, 2019; Wang *et al.*, 2023).

Furthermore, the phenotypic analysis of Eurasian populations from the Palaeolithic to the Iron Age reveals a complex evolution and diversification of physical traits. While dark phenotypes were predominant for most of the examined period, light phenotypes began to emerge and spread gradually from the Mesolithic onward.

Specifically, during the Palaeolithic period, samples predominantly exhibit dark phenotypic traits for eyes, hair, and skin, which may suggest that these populations were not yet well adapted to high-latitude conditions, having recently migrated from lower latitudes. An exception is the Russian sample Kostenki 14, which displays intermediate skin colour. This sample shares close ancestry with European Mesolithic HGs but not with East Asians (Seguin-Orlando *et al.*, 2014). In the Mesolithic period, light phenotypes, particularly in eye and skin colour, began to appear among HGs in Northern and Central Europe. These regions, associated with cold and low light conditions, suggest adaptation to reduced light levels, particularly among Scandinavian HGs. Supporting this, Gunther's 2018 study highlighted that Scandinavian populations exhibited high frequencies of low pigmentation variants, reinforcing the notion of adaptation to the environment (Günther *et al.*, 2018). Notably, the first sample to exhibit blue eyes, blond hair, and pale skin is a Mesolithic individual from Sweden.

In the Neolithic period, dark phenotypes for eyes, hair, and skin remain prevalent in Eastern and Western Europe, while Central and Northern Europe exhibit both light and dark phenotypes, with a decrease in the frequency of light phenotypes. This period is marked by increased phenotypic variability, reflecting the spread of agriculture and admixture between local HGs and migrating farmers (Olalde *et al.*, 2014; Lazaridis *et al.*, 2014; Gamba *et al.*, 2014).

During the Copper Age, dark phenotypes continue to dominate, but there is an increase in intermediate and light variants, particularly in Central and Northern Europe.

In the Bronze Age, although dark phenotypes remain prevalent, there is a notable increase in light phenotypes in Northern and Central Europe. Light phenotypes also begin to appear in new regions such as Russia, Kazakhstan, and Jordan, likely associated with the eastward expansion of the Yamnaya culture (Allentoft *et al.*, 2015). This period is characterized by an expansion in phenotypic diversity, with the co-occurrence of blue eyes, blonde hair, and light skin observed in multiple individuals.

In the Iron Age, dark phenotypes for eyes and hair remained dominant, but light phenotypes are present in Northern Europe, Italy, and Kazakhstan. Analysis of skin colour indicates a predominance of dark phenotypes, with intermediate and light phenotypes emerging in Western Europe.

What emerges from the scientific literature is that studies of ancient human populations attempting to reconstruct pigmentation phenotypes indicate that Mesolithic HGs had dark skin. These reconstructions have primarily focused on ancient individuals from Western Eurasia (Olalde *et al.*, 2014; González-Fortes *et al.*, 2017; Brace *et al.*, 2019; Jensen *et al.*, 2019). Populations with lighter skin may have arrived in Europe through Neolithic migration from the Middle East around 8,000 years ago (Mathieson *et al.*, 2018), after which a significant decrease in skin pigmentation is observed. In contrast, Günther *et al.* (2018) found high frequencies of alleles for light skin in Mesolithic Scandinavian individuals, suggesting an early environmental adaptation to high latitudes and Ju and Mathieson have noted that alleles for dark pigmentation have significantly decreased in the European population over the past 40,000 years (Günther *et al.*, 2018; Ju and Mathieson, 2021).

What emerges from this thesis project, due to the expanded geographical regions under study, is that light phenotypes were already present during the Mesolithic period, with even a Swedish HGs sample exhibiting blue eyes, blonde hair, and light skin. During the Neolithic period, there is no evidence of an increase in these traits; rather, there appears to be a decrease, as illustrated by France, which shows light traits in the Mesolithic but darker traits in the Neolithic. It cannot be excluded that Neolithic migration contributed to the presence of light skin in Europe, but it can be stated with certainty that light-skinned individuals were already present during the Mesolithic period. Thus, light skin in Europe might be the result of the interaction of various migratory and adaptive mechanisms, and a detailed molecular analysis of an increasing number of ancient samples will help clarify this point.

In conclusion, the phenotypic evolution observed across Eurasian populations from the Palaeolithic to the Iron Age illustrates a dynamic interplay between environmental adaptation, genetic admixtures, and migratory movements.

5 CONCLUSION

In recent years, methodological and theoretical advances in GWAS studies have led to a much deeper understanding of the genetics underlying many complex phenotypic traits of polygenic nature, such as eye, hair, and skin colour. These technological advancements have facilitated the consolidation of Forensic DNA Phenotyping (FDP), which allows for the prediction of phenotypic traits directly from DNA samples, without prior knowledge of the physical characteristics. In the field of FDP, one of the most widely used tools is HIrisPlex-S, designed to simultaneously predict eye, hair, and skin colour from DNA samples from 41 genetic loci.

The prediction of these traits from DNA initially had implications in the field of forensic genetics, but in recent years it has also provided phenotypic information through the analysis of ancient human remains, which often present degraded and low-coverage DNA. Although the laboratory protocols of the HIrisPlex-S system have been validated for forensic procedures, enabling the effective handling of degraded and low-coverage data, what is still lacking today is a bioinformatic validation of the genotype calling protocols.

The first part of this thesis aimed to test the robustness of the HIrisPlex-S system when applied to low-coverage ancient data, assessing the effectiveness of common imputation methods for managing missing data in low-coverage genomes, and particularly proposing a new protocol, defined as the HIrisPlex-S Protocol Implemented with a Genotype Likelihoods Model, which incorporates genotype likelihoods to address uncertainties in genotype calling associated with low-coverage data. What emerges is that this method, which employs a probabilistic approach to consider all 10 possible genotypes with their respective probabilities and increases the number of phenotypic predictions to 1000 rather than a single prediction, proves to be the most robust overall when dealing with low-coverage data.

In addition, it enables the identification of all potential phenotypes that could characterize a given sample. In contrast to other methods that may fail or generate incorrect phenotypes, the protocol developed in this thesis project can display the phenotype of interest, even at low probabilities, along with other possible phenotypes, thereby informing the user about the reliability of the phenotypic prediction.

The user manual defined for making phenotypic predictions thus requires the application of the standard HIrisPlex-S protocol up to a minimum coverage of 8x calculated for each of the 41 positions of interest. From a coverage of 8x and below, it is advisable to use the method tested here.

A future perspective for the testing phase involves replicating the entire procedure of downsampling and predictions on a real case to obtain direct information on the phenotypic traits of the case study, rather than inferring them. Additionally, it would also be of interest to test, starting from samples belonging to the Bronze Age onwards, a combined approach that utilizes the method developed here alongside an imputation approach for any missing positions.

The second part of this thesis project involved applying this protocol to a dataset of 348 Eurasian individuals spanning from the Upper Palaeolithic to the Iron Age, to provide insights into the physical appearances of ancient populations. The emerging picture shows a complex evolution of physical traits, with dark phenotypes being predominant initially. Light phenotypes started emerging in Europe during the Mesolithic, especially in northern and central regions, due to adaptation to lower light levels. The Neolithic period saw increased phenotypic variability due to the spread of agriculture and admixture between local HGs and migrating farmers. The Copper and Bronze Ages continued to exhibit dark phenotypes predominantly, but there was a gradual increase in intermediate and light variants, particularly in Northern and Central Europe. By the Iron Age, dark phenotypes remained dominant, but light phenotypes were present in Northern Europe, Italy, and Kazakhstan, with a noticeable emergence of intermediate and light skin tones in Western Europe.

aDNA studies have already attempted to reconstruct pigmentation phenotypes in ancient human populations. However, these reconstructions have primarily focused on ancient individuals from Western Eurasia, due to the relatively higher abundance of samples from European-centric studies. The accuracy of these predictions remains uncertain.

In this thesis project, for the first time, a protocol is tested and then applied to perform robust phenotypic inference. This phenotypic inference is applied to a dataset of samples, which not only span a time window from approximately 45,000 to approximately 1,000 years ago (defined by the age of the oldest and most recent samples), but also originate from regions ranging from Western Europe to Asia, encompassing a total of 34 countries.

Further molecular analyses and an increased availability of ancient data will enable a clearer identification of the complex adaptive and demographic mechanisms that have shaped Eurasian phenotypic diversity over the past 70,000 years, that is, the dispersal of *Homo sapiens* into Eurasia.

6 BIBLIOGRAPHY

Adhikari, K. *et al.* (2016) 'A Genome-Wide Association Scan Implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in Human Facial Variation'. *Nature Communications*, 7(1), p. 11616. DOI: 10.1038/ncomms11616.

Adhikari, K. *et al.* (2019) 'A GWAS in Latin Americans Highlights the Convergent Evolution of Lighter Skin Pigmentation in Eurasia'. *Nature Communications*, 10(1), p. 358. DOI: 10.1038/s41467-018-08147-0.

Allentoft, M.E., Sikora, M., Fischer, A., *et al.* (2024) '100 Ancient Genomes Show Repeated Population Turnovers in Neolithic Denmark'. *Nature*, 625(7994), pp. 329–337. DOI: 10.1038/s41586-023-06862-3.

Allentoft, M.E. *et al.* (2015) 'Population Genomics of Bronze Age Eurasia'. *Nature*, 522(7555), pp. 167–172. DOI: 10.1038/nature14507.

Allentoft, M.E., Sikora, M., Refoyo-Martínez, A., *et al.* (2024) 'Population Genomics of Post-Glacial Western Eurasia'. *Nature*, 625(7994), pp. 301–311. DOI: 10.1038/s41586-023-06865-0.

Allwood, J.S. and Harbison, S. (2013) 'SNP Model Development for the Prediction of Eye Colour in New Zealand'. *Forensic Science International: Genetics*, 7(4), pp. 444–452. DOI: 10.1016/j.fsigen.2013.03.005.

Ammerman, A.J. and Cavalli-Sforza, L.L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press Available at: http://www.jstor.org/stable/j.ctt7zvqz7.

Amorim, C.E.G. *et al.* (2018) 'Understanding 6th-Century Barbarian Social Organization and Migration through Paleogenomics'. *Nature Communications*, 9(1), p. 3547. DOI: 10.1038/s41467-018-06024-4.

Andrews S. (2010). *Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Available Online at: Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc*.

Antonio, M.L. *et al.* (2019) 'Ancient Rome: A Genetic Crossroads of Europe and the Mediterranean'. *Science*, 366(6466), pp. 708–714. DOI: 10.1126/science.aay6826.

Antonio, M.L. *et al.* (2024) 'Stable Population Structure in Europe since the Iron Age, despite High Mobility'. *ELife*, 13. DOI: 10.7554/eLife.79714.

Ausmees, K. *et al.* (2022) 'An Empirical Evaluation of Genotype Imputation of Ancient DNA'. *G3 Genes|Genomes|Genetics*, 12(6). DOI: 10.1093/g3journal/jkac089.

Barlow, A. *et al.* (2018) 'Partial Genomic Survival of Cave Bears in Living Brown Bears'. *Nature Ecology & Evolution*, 2(10), pp. 1563–1570. DOI: 10.1038/s41559-018-0654-8.

de Barros Damgaard, P. *et al.* (2018) 'The First Horse Herders and the Impact of Early Bronze Age Steppe Expansions into Asia'. *Science*, 360(6396). DOI: 10.1126/science.aar7711.

Barton, N., Etheridge, A. and Véber, A. (2016) 'The Infinitesimal Model'. *BioRxiv*.

Bengtsson, H. *et al.* (2017) 'MatrixStats: Functions That Apply to Rows and Columns of Matrices (and to Vectors). R Package Version 0.52.2.'

Del Bino, S., Duval, C. and Bernerd, F. (2018) 'Clinical and Biological Characterization of Skin Pigmentation Diversity and Its Consequences on UV Impact'. *International Journal of Molecular Sciences*, 19(9), p. 2668. DOI: 10.3390/ijms19092668.

Böhringer, S. and de Jong, M.A. (2019) 'Quantification of Facial Traits'. *Frontiers in Genetics*, 10. DOI: 10.3389/fgene.2019.00397.

Bos, K.I. *et al.* (2011) 'A Draft Genome of Yersinia Pestis from Victims of the Black Death'. *Nature*, 478(7370), pp. 506–510. DOI: 10.1038/nature10549.

Bos, K.I. *et al.* (2014) 'Pre-Columbian Mycobacterial Genomes Reveal Seals as a Source of New World Human Tuberculosis'. *Nature*, 514(7523), pp. 494–497. DOI: 10.1038/nature13591.

Bower, C. and Stanley, F.J. (1989) 'Dietary Folate as a Risk Factor for Neural-tube Defects: Evidence from a Case-control Study in Western Australia'. *Medical Journal of Australia*, 150(11), pp. 613–619. DOI: 10.5694/j.1326-5377.1989.tb136723.x.

Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) 'An Expanded View of Complex Traits: From Polygenic to Omnigenic'. *Cell*, 169(7), pp. 1177–1186. DOI: 10.1016/j.cell.2017.05.038.

Brace, S. *et al.* (2019) 'Ancient Genomes Indicate Population Replacement in Early Neolithic Britain'. *Nature Ecology & Evolution*, 3(5), pp. 765–771. DOI: 10.1038/s41559-019-0871-9.

Bradley, B.J. and Mundy, N.I. (2008) 'The Primate Palette: The Evolution of Primate Coloration'. *Evolutionary Anthropology: Issues, News, and Reviews*, 17(2), pp. 97–111. DOI: 10.1002/evan.20164.

Bramble, D.M. and Lieberman, D.E. (2004) 'Endurance Running and the Evolution of Homo'. *Nature*, 432(7015), pp. 345–352. DOI: 10.1038/nature03052.

Brand, C.M., Colbran, L.L. and Capra, J.A. (2022) 'Predicting Archaic Hominin Phenotypes from Genomic Data'. *Annual Review of Genomics and Human Genetics*, 23(1), pp. 591–612. DOI: 10.1146/annurev-genom-111521-121903.

Branda, R.F. and Blickensderfer, D.B. (1993) 'Folate Deficiency Increases Genetic Damage Caused by Alkylating Agents and Gamma-Irradiation in Chinese Hamster Ovary Cells.' *Cancer Research*, 53(22), pp. 5401–8.

Branda, R.F. and Eaton, J.W. (1978) 'Skin Color and Nutrient Photolysis: An Evolutionary Hypothesis'. *Science*, 201(4356), pp. 625–626. DOI: 10.1126/science.675247.

Branicki, W. *et al.* (2007) 'Determination of Phenotype Associated SNPs in the MC1R Gene'. *Journal of Forensic Sciences*, 52(2), pp. 349–354. DOI: 10.1111/j.1556-4029.2006.00361.x.

Branicki, W. *et al.* (2011) 'Model-Based Prediction of Human Hair Color Using DNA Variants'. *Human Genetics*, 129(4), pp. 443–454. DOI: 10.1007/s00439-010-0939-8.

Branicki, W., Brudnik, U. and Wojas-Pelc, A. (2009) 'Interactions Between HERC2, OCA2 and MC1R May Influence Human Pigmentation Phenotype'. *Annals of Human Genetics*, 73(2), pp. 160–170. DOI: 10.1111/j.1469-1809.2009.00504.x.

Briggs, A.W. *et al.* (2010) 'Removal of Deaminated Cytosines and Detection of in Vivo Methylation in Ancient DNA'. *Nucleic Acids Research*, 38(6), pp. e87–e87. DOI: 10.1093/nar/gkp1163.

Browning, B.L. and Browning, S.R. (2016) 'Genotype Imputation with Millions of Reference Samples'. *The American Journal of Human Genetics*, 98(1), pp. 116–126. DOI: 10.1016/j.ajhg.2015.11.020.

Browning, B.L., Zhou, Y. and Browning, S.R. (2018) 'A One-Penny Imputed Genome from Next-Generation Reference Panels'. *The American Journal of Human Genetics*, 103(3), pp. 338–348. DOI: 10.1016/j.ajhg.2018.07.015.

Brunson, K. and Reich, D. (2019) 'The Promise of Paleogenomics Beyond Our Own Species'. *Trends in Genetics*, 35(5), pp. 319 – 329. DOI: 10.1016/j.tig.2019.02.006.

Burbano, H.A. *et al.* (2010) 'Targeted Investigation of the Neandertal Genome by Array-Based Sequence Capture'. *Science*, 328(5979), pp. 723–725. DOI: 10.1126/science.1188046.

Byrska-Bishop, M. *et al.* (2022) 'High-Coverage Whole-Genome Sequencing of the Expanded 1000 Genomes Project Cohort Including 602 Trios'. *Cell*, 185(18), pp. 3426-3440.e19. DOI: 10.1016/j.cell.2022.08.004.

Caro, T. (2005) 'The Adaptive Significance of Coloration in Mammals'. *BioScience*, 55(2), pp. 125–136. DOI: 10.1641/0006-3568(2005)055[0125:TASOCI]2.0.CO;2.

Caro, T. (2013) 'The Colours of Extant Mammals'. *Seminars in Cell & Developmental Biology*, 24(6–7), pp. 542–552. DOI: 10.1016/j.semcdb.2013.03.016.

Cassidy, L.M. *et al.* (2020) 'A Dynastic Elite in Monumental Neolithic Society'. *Nature*, 582(7812), pp. 384–388. DOI: 10.1038/s41586-020-2378-6.

Cassidy, L.M. *et al.* (2016) 'Neolithic and Bronze Age Migration to Ireland and Establishment of the Insular Atlantic Genome'. *Proceedings of the National Academy of Sciences*, 113(2), pp. 368–373. DOI: 10.1073/pnas.1518445113.

Chaitanya, L. *et al.* (2018) 'The HIrisPlex-S System for Eye, Hair and Skin Colour Prediction from DNA: Introduction and Forensic Developmental Validation'. *Forensic Science International: Genetics*, 35, pp. 123–135. DOI: 10.1016/j.fsigen.2018.04.004.

Chaplin, G. and Jablonski, N.G. (2013) 'The Human Environment and the Vitamin D Compromise: Scotland as a Case Study in Human Biocultural Adaptation and Disease Susceptibility'. *Human Biology*, 85(4), pp. 529–552. DOI: 10.3378/027.085.0402.

Chaplin, G. and Jablonski, N.G. (1998) 'The Integument of the "Odd-Nosed" Colobines'. In *The Natural History of the Doucs and Snub-Nosed Monkeys*. WORLD SCIENTIFIC, pp. 79–104. DOI: 10.1142/9789812817020_0005.

Chen, T.C. *et al.* (2007) 'Factors That Influence the Cutaneous Synthesis and Dietary Sources of Vitamin D'. *Archives of Biochemistry and Biophysics*, 460(2), pp. 213–217. DOI: 10.1016/j.abb.2006.12.017.

Clemens, T.L. *et al.* (1982) 'Increased Skin Pigment Reduces the Capacity of Skin to Synthesise Vitamin D3'. *The Lancet*, 319(8263), pp. 74–76. DOI: 10.1016/S0140-6736(82)90214-8.

Clemente, F. *et al.* (2021) 'The Genomic History of the Aegean Palatial Civilizations'. *Cell*, 184(10), pp. 2565-2586.e21. DOI: 10.1016/j.cell.2021.03.039.

Cox, S.L. *et al.* (2022) 'Predicting Skeletal Stature Using Ancient DNA'. *American Journal of Biological Anthropology*, 177(1), pp. 162–174. DOI: 10.1002/ajpa.24426.

Crawford, N.G. *et al.* (2017) 'Loci Associated with Skin Pigmentation Identified in African Populations'. *Science*, 358(6365). DOI: 10.1126/science.aan8433.

Crump, S.E. *et al.* (2019) 'Arctic Shrub Colonization Lagged Peak Postglacial Warmth: Molecular Evidence in Lake Sediment from Arctic Canada'. *Global Change Biology*, 25(12), pp. 4244–4256. DOI: 10.1111/gcb.14836.

Damgaard, P.B. *et al.* (2015) 'Improving Access to Endogenous DNA in Ancient Bones and Teeth'. *Scientific Reports*, 5(1), p. 11184. DOI: 10.1038/srep11184.

Deng, T. *et al.* (2022) 'Comparison of Genotype Imputation for SNP Array and Low-Coverage Whole-Genome Sequencing Data'. *Frontiers in Genetics*, 12. DOI: 10.3389/fgene.2021.704118.

DePristo, M.A. *et al.* (2011) 'A Framework for Variation Discovery and Genotyping Using Next-Generation DNA Sequencing Data'. *Nature Genetics*, 43(5), pp. 491–498. DOI: 10.1038/ng.806.

Dorado, G. *et al.* (2021) 'Molecular Biology to Infer Phenotypes of Forensic and Ancient Remains in Bioarchaeology - Review'. *ARCHAEOBIOS* , 1(15). Available at: www.arqueobios.orgwww.arqueobios.org.

Draus-Barini, J. *et al.* (2013) 'Bona Fide Colour: DNA Prediction of Human Eye and Hair Colour from Ancient and Contemporary Skeletal Remains'. *Investigative Genetics*, 4(1), p. 3. DOI: 10.1186/2041-2223-4-3.

Edwards, M. *et al.* (2016) 'Iris Pigmentation as a Quantitative Trait: Variation in Populations of European, East Asian and South Asian Ancestry and Association with Candidate Gene Polymorphisms'. *Pigment Cell & Melanoma Research*, 29(2), pp. 141–162. DOI: 10.1111/pcmr.12435.

Eiberg, H. *et al.* (2008) 'Blue Eye Color in Humans May Be Caused by a Perfectly Associated Founder Mutation in a Regulatory Element Located within the HERC2 Gene Inhibiting OCA2 Expression'. *Human Genetics*, 123(2), pp. 177–187. DOI: 10.1007/s00439-007-0460-x.

Erven, J.A.M. *et al.* (2022) 'Imputation of Ancient Whole Genome Sus Scrofa DNA Introduces Biases Toward Main Population Components in the Reference Panel'. *Frontiers in Genetics*, 13. DOI: 10.3389/fgene.2022.872486.

Fagertun, J. *et al.* (2015) 'Predicting Facial Characteristics from Complex Polygenic Variations'. *Forensic Science International: Genetics*, 19, pp. 263–268. DOI: 10.1016/j.fsigen.2015.08.004.

Feng, Y., McQuillan, M.A. and Tishkoff, S.A. (2021) 'Evolutionary Genetics of Skin Pigmentation in African Populations'. *Human Molecular Genetics*, 30(R1), pp. R88–R97. DOI: 10.1093/hmg/ddab007.

Fisher, R.A. (1919) 'XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.' *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2), pp. 399–433.

Fleming, A. and Copp, A.J. (1998) 'Embryonic Folate Metabolism and Mouse Neural Tube Defects'. *Science*, 280(5372), pp. 2107–2109. DOI: 10.1126/science.280.5372.2107.

Frazier, M.E. *et al.* (2003) 'Realizing the Potential of the Genome Revolution: The Genomes to Life Program'. *Science*, 300(5617), pp. 290–293. DOI: 10.1126/science.1084566.

Freire-Aradas, A., Phillips, C. and Lareu, M. V. (2017) 'Forensic Individual Age Estimation with DNA: From Initial Approaches to Methylation Tests.' *Forensic Science Review*, 29(2), pp. 121–144.

Frudakis, T., Terravainen, T. and Thomas, M. (2007) 'Multilocus OCA2 Genotypes Specify Human Iris Colors'. *Human Genetics*, 122(3–4), pp. 311–326. DOI: 10.1007/s00439-007-0401-8.

Fu, Q. *et al.* (2015) 'An Early Modern Human from Romania with a Recent Neanderthal Ancestor'. *Nature*, 524(7564), pp. 216–219. DOI: 10.1038/nature14558.

Fu, Q. *et al.* (2014) 'Genome Sequence of a 45,000-Year-Old Modern Human from Western Siberia'. *Nature*, 514(7523), pp. 445–449. DOI: 10.1038/nature13810.

Fu, Q. *et al.* (2016) 'The Genetic History of Ice Age Europe'. *Nature*, 534(7606), pp. 200–205. DOI: 10.1038/nature17993.

Fulton, T.L. and Shapiro, B. (2019) 'Setting Up an Ancient DNA Laboratory'. In pp. 1–13. DOI: 10.1007/978-1-4939-9176-1_1.

Fumagalli, M. *et al.* (2013) 'Quantifying Population Genetic Differentiation from Next-Generation Sequencing Data'. *Genetics*, 195(3), pp. 979–992. DOI: 10.1534/genetics.113.154740.

Gamba, C. *et al.* (2014) 'Genome Flux and Stasis in a Five Millennium Transect of European Prehistory'. *Nature Communications*, 5(1), p. 5257. DOI: 10.1038/ncomms6257.

Gansauge, M.-T. *et al.* (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(8), pp. 2279–2300. DOI: 10.1038/s41596-020-0338-0.

Gansauge, M.-T. and Meyer, M. (2013) 'Single-Stranded DNA Library Preparation for the Sequencing of Ancient or Damaged DNA'. *Nature Protocols*, 8(4), pp. 737–748. DOI: 10.1038/nprot.2013.038.

Garrido Marques, A. *et al.* (2024) 'Assessing the Impact of Post-Mortem Damage and Contamination on Imputation Performance in Ancient DNA'. *Scientific Reports*, 14(1), p. 6227. DOI: 10.1038/s41598-024-56584-3.

Gautam, P. *et al.* (2015) 'Population Diversity and Adaptive Evolution in Keratinization Genes: Impact of Environment in Shaping Skin Phenotypes'. *Molecular Biology and Evolution*, 32(3), pp. 555–573. DOI: 10.1093/molbev/msu342.

Gilbert, M.T.P. *et al.* (2008) 'DNA from Pre-Clovis Human Coprolites in Oregon, North America'. *Science*, 320(5877), pp. 786–789. DOI: 10.1126/science.1154116.

Gilbert, M.T.P. *et al.* (2006) 'Resistance of Degraded Hair Shafts to Contaminant DNA'. *Forensic Science International*, 156(2–3), pp. 208–212. DOI: 10.1016/j.forsciint.2005.02.021.

Glocke, I. and Meyer, M. (2017) 'Extending the Spectrum of DNA Sequences Retrieved from Ancient Bones and Teeth'. *Genome Research*, 27(7), pp. 1230–1237. DOI: 10.1101/gr.219675.116.

González-Fortes, G. *et al.* (2017) 'Paleogenomic Evidence for Multi-Generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin'. *Current Biology*, 27(12), pp. 1801-1810.e10. DOI: 10.1016/j.cub.2017.05.023.

Green, R.E. *et al.* (2010) 'A Draft Sequence of the Neandertal Genome'. *Science*, 328(5979), pp. 710–722. DOI: 10.1126/science.1188021.

Green, R.E. *et al.* (2009) 'The Neandertal Genome and Ancient DNA Authenticity'. *The EMBO Journal*, 28(17), pp. 2494–2502. DOI: 10.1038/emboj.2009.222.

Grimes, E.A. *et al.* (2001) 'Sequence Polymorphism in the Human Melanocortin 1 Receptor Gene as an Indicator of the Red Hair Phenotype'. *Forensic Science International*, 122(2–3), pp. 124–129. DOI: 10.1016/S0379-0738(01)00480-7.

Günther, T. *et al.* (2018) 'Population Genomics of Mesolithic Scandinavia: Investigating Early Postglacial Migration Routes and High-Latitude Adaptation'. *PLOS Biology*, 16(1), p. e2003703. DOI: 10.1371/journal.pbio.2003703.

Günther, T. and Jakobsson, M. (2016) 'Genes Mirror Migrations and Cultures in Prehistoric Europe — a Population Genomic Perspective'. *Current Opinion in Genetics & Development*, 41, pp. 115–123. DOI: 10.1016/j.gde.2016.09.004.

Günther, T. and Nettelblad, C. (2019) 'The Presence and Impact of Reference Bias on Population Genomic Studies of Prehistoric Human Populations'. *PLOS Genetics*, 15(7), p. e1008302. DOI: 10.1371/journal.pgen.1008302.

Haak, W. *et al.* (2015) 'Massive Migration from the Steppe Was a Source for Indo-European Languages in Europe'. *Nature*, 522(7555), pp. 207–211. DOI: 10.1038/nature14317.

Hagelberg, E., Sykes, B. and Hedges, R. (1989) 'Ancient Bone DNA Amplified'. *Nature*, 342(6249), pp. 485–485. DOI: 10.1038/342485a0.

Hanel, A. and Carlberg, C. (2020) 'Skin Colour and Vitamin D: An Update'. *Experimental Dermatology*, 29(9), pp. 864–875. DOI: 10.1111/exd.14142.

Harding, R.M. *et al.* (2000) 'Evidence for Variable Selective Pressures at MC1R'. *The American Journal of Human Genetics*, 66(4), pp. 1351–1361. DOI: 10.1086/302863.

Hart, K.L. *et al.* (2013) 'Improved Eye- and Skin-Color Prediction Based on 8 SNPs'. *Croatian Medical Journal*, 54(3), pp. 248–256. DOI: 10.3325/cmj.2013.54.248.

Heidegger, A. *et al.* (2020) 'Development and Optimization of the VISAGE Basic Prototype Tool for Forensic Age Estimation'. *Forensic Science International: Genetics*, 48, p. 102322. DOI: 10.1016/j.fsigen.2020.102322.

Held, L.I. (2010) 'The Evo-Devo Puzzle of Human Hair Patterning'. *Evolutionary Biology*, 37(2–3), pp. 113–122. DOI: 10.1007/s11692-010-9085-4.

Hennessy, A. *et al.* (2005) 'The Photoadaptive Response to Ultraviolet Exposure in Human Skin Using Ultraviolet Spectrophotometry'. *Photodermatology, Photoimmunology & Photomedicine*, 21(5), pp. 229–233. DOI: 10.1111/j.1600-0781.2005.00170.x.

Higuchi, R. *et al.* (1984) 'DNA Sequences from the Quagga, an Extinct Member of the Horse Family'. *Nature*, 312(5991), pp. 282–284. DOI: 10.1038/312282a0.

Hofreiter, M. *et al.* (2021) 'Progress in Forensic Bone DNA Analysis: Lessons Learned from Ancient DNA'. *Forensic Science International: Genetics*, 54, p. 102538. DOI: 10.1016/j.fsigen.2021.102538.

Hofreiter, M. *et al.* (2015) 'The Future of Ancient DNA: Technical Advances and Conceptual Shifts'. *BioEssays*, 37(3), pp. 284–293. DOI: 10.1002/bies.201400160.

Hoss, M. *et al.* (1996) 'DNA Damage and DNA Sequence Retrieval from Ancient Tissues'. *Nucleic Acids Research*, 24(7), pp. 1304–1307. DOI: 10.1093/nar/24.7.1304.

Höss, M. and Pääbo, S. (1993) 'DNA Extraction from Pleistocene Bones by a Silica-Based Purification Method'. *Nucleic Acids Research*, 21(16), pp. 3913–3914. DOI: 10.1093/nar/21.16.3913.

Http://Hgdownload.Cse.Ucsc.Edu/Goldenpath/Hg38/LiftOver/.

Https://Broadinstitute.Github.Io/Picard/.

Https://www.ncbi.nlm.nih.gov/snp/. DbSNP [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004 – [Cited 2024/07]. Available from: Https://Www.Ncbi.Nlm.Nih.Gov/Snp/. Available at: https://www.ncbi.nlm.nih.gov/snp/ (Accessed: 1 July 2024).

Hudjashov, G., Villems, R. and Kivisild, T. (2013) 'Global Patterns of Diversity and Selection in Human Tyrosinase Gene'. *PLoS ONE*, 8(9), p. e74307. DOI: 10.1371/journal.pone.0074307.

Hui, R. *et al.* (2020) 'Evaluating Genotype Imputation Pipeline for Ultra-Low Coverage Ancient Genomes'. *Scientific Reports*, 10(1), p. 18542. DOI: 10.1038/s41598-020-75387-w.

Hysi, P.G. *et al.* (2018) 'Genome-Wide Association Meta-Analysis of Individuals of European Ancestry Identifies New Loci Explaining a Substantial Fraction of Hair Color Variation and Heritability'. *Nature Genetics*, 50(5), pp. 652–656. DOI: 10.1038/s41588-018-0100-5.

Jablonski N. G. (2021) 'The evolution of human skin pigmentation involved the interactions of genetic, environmental, and cultural variables'. *Pigment cell & melanoma research*, 34(4), pp. 707–729. DOI: 10.1111/pcmr.12976.

Jablonski, N.G. and Chaplin, G. (2017) 'The Colours of Humanity: The Evolution of Pigmentation in the Human Lineage'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1724), p. 20160349. DOI: 10.1098/rstb.2016.0349.

Jablonski, N.G. and Chaplin, G. (2000) 'The Evolution of Human Skin Coloration'. *Journal of Human Evolution*, 39(1), pp. 57–106. DOI: 10.1006/jhev.2000.0403.

Jensen, T.Z.T. *et al.* (2019) 'A 5700 Year-Old Human Genome and Oral Microbiome from Chewed Birch Pitch'. *Nature Communications*, 10(1), p. 5520. DOI: 10.1038/s41467-019-13549-9.

Johannsen, W. (1911) 'The Genotype Conception of Heredity'. *American Naturalist*, 45(531), pp. 129–159.

Jones, E.R. *et al.* (2017) 'The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers'. *Current Biology*, 27(4), pp. 576–582. DOI: 10.1016/j.cub.2016.12.060.

Jones, E.R. *et al.* (2015) 'Upper Palaeolithic Genomes Reveal Deep Roots of Modern Eurasians'. *Nature Communications*, 6(1), p. 8912. DOI: 10.1038/ncomms9912.

Jonnalagadda, M. *et al.* (2019) 'A Genome-Wide Association Study of Skin and Iris Pigmentation among Individuals of South Asian Ancestry'. *Genome Biology and Evolution*, 11(4), pp. 1066–1076. DOI: 10.1093/gbe/evz057.

Jónsson, H. *et al.* (2013) 'MapDamage2.0: Fast Approximate Bayesian Estimates of Ancient DNA Damage Parameters'. *Bioinformatics*, 29(13), pp. 1682–1684. DOI: 10.1093/bioinformatics/btt193.

Ju, D. and Mathieson, I. (2021) 'The Evolution of Skin Pigmentation-Associated Variation in West Eurasia'. *Proceedings of the National Academy of Sciences*, 118(1). DOI: 10.1073/pnas.2009227118.

Kayser, M. (2015) 'Forensic DNA Phenotyping: Predicting Human Appearance from Crime Scene Material for Investigative Purposes'. *Forensic Science International: Genetics*, 18, pp. 33–48. DOI: 10.1016/J.FSIGEN.2015.02.003.

Kayser, M. *et al.* (2023) 'Recent Advances in Forensic DNA Phenotyping of Appearance, Ancestry and Age'. *Forensic Science International: Genetics*, 65, p. 102870. DOI: 10.1016/J.FSIGEN.2023.102870.

Kayser, M. *et al.* (2008) 'Three Genome-Wide Association Studies and a Linkage Analysis Identify HERC2 as a Human Iris Color Gene'. *The American Journal of Human Genetics*, 82(2), pp. 411–423. DOI: 10.1016/j.ajhg.2007.10.003.

Keating, B. *et al.* (2013) 'First All-in-One Diagnostic Tool for DNA Intelligence: Genome-Wide Inference of Biogeographic Ancestry, Appearance, Relatedness, and Sex with the Identitas v1 Forensic Chip'. *International Journal of Legal Medicine*, 127(3), pp. 559–572. DOI: 10.1007/s00414-012-0788-1.

Keller, A. *et al.* (2012) 'New Insights into the Tyrolean Iceman's Origin and Phenotype as Inferred by Whole-Genome Sequencing'. *Nature Communications*, 3(1), p. 698. DOI: 10.1038/ncomms1701.

Kenny, E.E. *et al.* (2012) 'Melanesian Blond Hair Is Caused by an Amino Acid Change in TYRP1'. *Science*, 336(6081), pp. 554–554. DOI: 10.1126/science.1217849.

Kerner, G., Neehus, A.-L., *et al.* (2023) 'Genetic Adaptation to Pathogens and Increased Risk of Inflammatory Disorders in Post-Neolithic Europe'. *Cell Genomics*, 3(2), p. 100248. DOI: 10.1016/j.xgen.2022.100248.

Kerner, G. *et al.* (2021) 'Human Ancient DNA Analyses Reveal the High Burden of Tuberculosis in Europeans over the Last 2,000 Years'. *The American Journal of Human Genetics*, 108(3), pp. 517–524. DOI: 10.1016/j.ajhg.2021.02.009.

Kerner, G., Choin, J. & Quintana-Murci, L. (2023) 'Ancient DNA as a tool for medical research'. *Nat Med* 29, pp. 1048–1051. DOI: 10.1038/s41591-023-02244-4.

King, T.E. *et al.* (2014) 'Identification of the Remains of King Richard III'. *Nature Communications*, 5(1), p. 5631. DOI: 10.1038/ncomms6631.

Kingdon, J.S. (1980) 'The Role of Visual Signals and Face Patterns in African Forest Monkeys (Guenons) of the Genus Cercopithecus'. *The Transactions of the Zoological Society of London*, 35(4), pp. 425–475. DOI: 10.1111/j.1096-3642.1980.tb00062.x.

Kircher, M., Sawyer, S. and Meyer, M. (2012) 'Double Indexing Overcomes Inaccuracies in Multiplex Sequencing on the Illumina Platform'. *Nucleic Acids Research*, 40(1), pp. e3–e3. DOI: 10.1093/nar/gkr771.

Kirsanow, K. and Burger, J. (2012) 'Ancient Human DNA'. *Annals of Anatomy - Anatomischer Anzeiger*, 194(1), pp. 121–132. DOI: 10.1016/j.aanat.2011.11.002.

Kılınç, G.M. *et al.* (2016) 'The Demographic Development of the First Farmers in Anatolia'. *Current Biology*, 26(19), pp. 2659–2666. DOI: 10.1016/j.cub.2016.07.057.

Knapp, M. and Hofreiter, M. (2010) 'Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives'. *Genes*, 1(2), pp. 227–243. DOI: 10.3390/genes1020227.

Koboldt, D.C. *et al.* (2012) 'VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing'. *Genome Research*, 22(3), pp. 568–576. DOI: 10.1101/gr.129684.111.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R. (2014) 'ANGSD: Analysis of Next Generation Sequencing Data'. *BMC Bioinformatics*, 15(1), p. 356. DOI: 10.1186/s12859-014-0356-4.

Krause, J. *et al.* (2010) 'A Complete MtDNA Genome of an Early Modern Human from Kostenki, Russia'. *Current Biology*, 20(3), pp. 231–236. DOI: 10.1016/j.cub.2009.11.068.

Lamason, R.L. *et al.* (2005) 'SLC24A5, a Putative Cation Exchanger, Affects Pigmentation in Zebrafish and Humans'. *Science*, 310(5755), pp. 1782–1786. DOI: 10.1126/science.1116238.

Langmead, B. and Salzberg, S.L. (2012) 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods*, 9(4), pp. 357–359. DOI: 10.1038/nmeth.1923.

Lao, O. *et al.* (2007) 'Signatures of Positive Selection in Genes Associated with Human Skin Pigmentation as Revealed from Analyses of Single Nucleotide Polymorphisms'. *Annals of Human Genetics*, 71(3), pp. 354–369. DOI: 10.1111/j.1469-1809.2006.00341.x.

Lazaridis, I. *et al.* (2014) 'Ancient Human Genomes Suggest Three Ancestral Populations for Present-Day Europeans'. *Nature*, 513(7518), pp. 409–413. DOI: 10.1038/nature13673.

Lazaridis, I. *et al.* (2016) 'Genomic Insights into the Origin of Farming in the Ancient Near East'. *Nature*, 536(7617), pp. 419–424. DOI: 10.1038/nature19310.

Li, H. (2011) 'A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data'. *Bioinformatics*, 27(21), pp. 2987–2993. DOI: 10.1093/bioinformatics/btr509.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics*, 25(16), pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, H. and Durbin, R. (2010) 'Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform'. *Bioinformatics*, 26(5), pp. 589–595. DOI: 10.1093/bioinformatics/btp698.

Li, H. and Durbin, R. (2009) 'Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform'. *Bioinformatics*, 25(14), pp. 1754–1760. DOI: 10.1093/bioinformatics/btp324.

Li, N. and Stephens, M. (2003) 'Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data'. *Genetics*, 165(4), pp. 2213–2233. DOI: 10.1093/genetics/165.4.2213.

Li, Y. *et al.* (2009) 'Genotype Imputation'. *Annual Review of Genomics and Human Genetics*, 10(1), pp. 387–406. DOI: 10.1146/annurev.genom.9.081307.164242.

Lieberman, D.E. (2014) 'Human Locomotion and Heat Loss: An Evolutionary Perspective'. In *Comprehensive Physiology*. Wiley, pp. 99–117. DOI: 10.1002/cphy.c140011.

Lindahl, T. (1993) 'Instability and Decay of the Primary Structure of DNA'. *Nature*, 362(6422), pp. 709–715. DOI: 10.1038/362709a0.

Liu, D. *et al.* (2021) 'Impact of Low-Frequency Coding Variants on Human Facial Shape'. *Scientific Reports*, 11(1), p. 748. DOI: 10.1038/s41598-020-80661-y.

Liu, F. *et al.* (2010) 'Digital Quantification of Human Eye Color Highlights Genetic Association of Three New Loci'. *PLoS Genetics*, 6(5), p. e1000934. DOI: 10.1371/journal.pgen.1000934.

Liu, F. *et al.* (2009) 'Eye Color and the Prediction of Complex Phenotypes from Genotypes'. *Current Biology*, 19(5), pp. R192–R193. DOI: 10.1016/j.cub.2009.01.027.

Liu, F., Walsh, S. and Kayser, M. (2014) 'Of Sex and IrisPlex Eye Colour Prediction: A Reply to Martinez-Cadenas *et al.*' *Forensic Science International: Genetics*, 9, pp. e5–e6. DOI: 10.1016/j.fsigen.2013.06.006.

Liu, F., Wen, B. and Kayser, M. (2013) 'Colorful DNA Polymorphisms in Humans'. *Seminars in Cell & Developmental Biology*, 24(6–7), pp. 562–575. DOI: 10.1016/j.semcdb.2013.03.013.

Loomis, W.F. (1967) 'Skin-Pigment Regulation of Vitamin-D Biosynthesis in Man'. *Science*, 157(3788), pp. 501–506. DOI: 10.1126/science.157.3788.501.

Lucock, M. *et al.* (2001) 'An Examination of Polymorphic Genes and Folate Metabolism in Mothers Affected by a Spina Bifida Pregnancy'. *Molecular Genetics and Metabolism*, 73(4), pp. 322–332. DOI: 10.1006/mgme.2001.3205.

Mackey, D.A. *et al.* (2011) 'Classification of Iris Colour: Review and Refinement of a Classification Schema'. *Clinical & Experimental Ophthalmology*, 39(5), pp. 462–471. DOI: 10.1111/j.1442-9071.2010.02487.x.

Mafessoni, F. *et al.* (2020) 'A High-Coverage Neandertal Genome from Chagyrskaya Cave'. *Proceedings of the National Academy of Sciences*, 117(26), pp. 15132–15136. DOI: 10.1073/pnas.2004944117.

Malmström, H. *et al.* (2019) 'The Genomic Ancestry of the Scandinavian Battle Axe Culture People and Their Relation to the Broader Corded Ware Horizon'. *Proceedings of the Royal Society B: Biological Sciences*, 286(1912), p. 20191528. DOI: 10.1098/rspb.2019.1528.

Marchi, N. *et al.* (2022) 'The Genomic Origins of the World's First Farmers'. *Cell*, 185(11), pp. 1842-1859.e18. DOI: 10.1016/j.cell.2022.04.008.

Margaryan, A. *et al.* (2018) 'Ancient Pathogen DNA in Human Teeth and Petrous Bones'. *Ecology and Evolution*, 8(6), pp. 3534–3542. DOI: 10.1002/ece3.3924.

Margaryan, A. *et al.* (2020) 'Population Genomics of the Viking World'. *Nature*, 585(7825), pp. 390–396. DOI: 10.1038/s41586-020-2688-8.

Maroñas, O. *et al.* (2014) 'Development of a Forensic Skin Colour Predictive Test'. *Forensic Science International: Genetics*, 13, pp. 34–44. DOI: 10.1016/j.fsigen.2014.06.017.

Martinez-Cadenas, C. *et al.* (2013) 'Gender Is a Major Factor Explaining Discrepancies in Eye Colour Prediction Based on HERC2/OCA2 Genotype and the IrisPlex Model'. *Forensic Science International: Genetics*, 7(4), pp. 453–460. DOI: 10.1016/j.fsigen.2013.03.007.

Martínez-Cadenas, C. *et al.* (2013) 'Simultaneous Purifying Selection on the Ancestral MC1R Allele and Positive Selection on the Melanoma-Risk Allele V60L in South Europeans'. *Molecular Biology and Evolution*, 30(12), pp. 2654–2665. DOI: 10.1093/molbev/mst158.

Martiniano, R. *et al.* (2020) 'Removing Reference Bias and Improving Indel Calling in Ancient DNA Data Analysis by Mapping to a Sequence Variation Graph'. *Genome Biology*, 21(1), p. 250. DOI: 10.1186/s13059-020-02160-7.

Mathieson, I. *et al.* (2015) 'Genome-Wide Patterns of Selection in 230 Ancient Eurasians'. *Nature*, 528(7583), pp. 499–503. DOI: 10.1038/nature16152.

Mathieson, I. *et al.* (2018) 'The Genomic History of Southeastern Europe'. *Nature*, 555(7695), pp. 197–203. DOI: 10.1038/nature25778.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data'. *Genome Research*, 20(9), pp. 1297–1303. DOI: 10.1101/gr.107524.110.

Moots, H.M. *et al.* (2023) 'A Genetic History of Continuity and Mobility in the Iron Age Central Mediterranean'. *Nature Ecology & Evolution*, 7(9), pp. 1515–1524. DOI: 10.1038/s41559-023-02143-4.

Murray, F.G. (1934) 'Pigmentation, Sunlight, and Nutritional Disease'. *American Anthropologist*, 36(3), pp. 438–445.

Naj, A.C. (2019) 'Genotype Imputation in Genome-Wide Association Studies'. *Current Protocols in Human Genetics*, 102(1). DOI: 10.1002/cphg.84.

Narasimhan, V.M. *et al.* (2019) 'The Formation of Human Populations in South and Central Asia'. *Science*, 365(6457). DOI: 10.1126/science.aat7487.

Nielsen, R. *et al.* (2011) 'Genotype and SNP Calling from Next-Generation Sequencing Data'. *Nature Reviews Genetics*, 12(6), pp. 443–451. DOI: 10.1038/nrg2986.

Nielsen, R. *et al.* (2012) 'SNP Calling, Genotype Calling, and Sample Allele Frequency Estimation from New-Generation Sequencing Data'. *PLoS ONE*, 7(7), p. e37558. DOI: 10.1371/journal.pone.0037558.

Norton, H.L. *et al.* (2014) 'Distribution of an Allele Associated with Blond Hair Color across Northern Island Melanesia'. *American Journal of Physical Anthropology*, 153(4), pp. 653–662. DOI: 10.1002/ajpa.22466.

Norton, H.L. *et al.* (2007) 'Genetic Evidence for the Convergent Evolution of Light Skin in Europeans and East Asians'. *Molecular Biology and Evolution*, 24(3), pp. 710–722. DOI: 10.1093/molbev/msl203.

Norton, H.L. *et al.* (2016) 'Quantitative Assessment of Skin, Hair, and Iris Variation in a Diverse Sample of Individuals and Associated Genetic Variation'. *American Journal of Physical Anthropology*, 160(4), pp. 570–581. DOI: 10.1002/ajpa.22861.

Olalde, I. *et al.* (2014) 'Derived Immune and Ancestral Pigmentation Alleles in a 7,000-Year-Old Mesolithic European'. *Nature*, 507(7491), pp. 225–228. DOI: 10.1038/nature12960.

Olalde, I. *et al.* (2018) 'The Beaker Phenomenon and the Genomic Transformation of Northwest Europe'. *Nature*, 555(7695), pp. 190–196. DOI: 10.1038/nature25738.

Olalde, I. *et al.* (2019) 'The Genomic History of the Iberian Peninsula over the Past 8000 Years'. *Science*, 363(6432), pp. 1230–1234. DOI: 10.1126/science.aav4040.

Omrak, A. *et al.* (2016) 'Genomic Evidence Establishes Anatolia as the Source of the European Neolithic Gene Pool'. *Current Biology*, 26(2), pp. 270–275. DOI: 10.1016/j.cub.2015.12.019.

Orlando, L. *et al.* (2021) 'Ancient DNA Analysis'. *Nature Reviews Methods Primers*, 1(1), p. 14. DOI: 10.1038/s43586-020-00011-0.

Orlando, L. *et al.* (2013) 'Recalibrating Equus Evolution Using the Genome Sequence of an Early Middle Pleistocene Horse'. *Nature*, 499(7456), pp. 74–78. DOI: 10.1038/nature12323.

Özdoğan, M. (2011) 'Archaeological Evidence on the Westward Expansion of Farming Communities from Eastern Anatolia to the Aegean and the Balkans'. *Current Anthropology*, 52(S4), pp. S415–S430. DOI: 10.1086/658895.

Pääbo, S. (1985) 'Molecular Cloning of Ancient Egyptian Mummy DNA'. *Nature*, 314(6012), pp. 644–645. DOI: 10.1038/314644a0.

Pääbo, S. and Wilson, A.C. (1988) 'Polymerase Chain Reaction Reveals Cloning Artefacts'. *Nature*, 334(6181), pp. 387–388. DOI: 10.1038/334387b0.

Palkopoulou, E. *et al.* (2015) 'Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth'. *Current Biology*, 25(10), pp. 1395–1400. DOI: 10.1016/j.cub.2015.04.007.

Patzold, F., Zilli, A. and Hundsdoerfer, A.K. (2020) 'Advantages of an Easy-to-Use DNA Extraction Method for Minimal-Destructive Analysis of Collection Specimens'. *PLOS ONE*, 15(7), p. e0235222. DOI: 10.1371/journal.pone.0235222.

Pietroni, C. *et al.* (2014) 'The Effect of Gender on Eye Colour Variation in European Populations and an Evaluation of the IrisPlex Prediction Model'. *Forensic Science International: Genetics*, 11, pp. 1–6. DOI: 10.1016/j.fsigen.2014.02.002.

Pisarek, A. *et al.* (2021) 'Epigenetic Age Prediction in Semen – Marker Selection and Model Development'. *Aging*, 13(15), pp. 19145–19164. DOI: 10.18632/aging.203399.

Pneuman, A. *et al.* (2012) 'Verification of Eye and Skin Color Predictors in Various Populations'. *Legal Medicine*, 14(2), pp. 78–83. DOI: 10.1016/j.legalmed.2011.12.005.

Pośpiech, E. *et al.* (2011) 'Gene–Gene Interactions Contribute to Eye Colour Variation in Humans'. *Journal of Human Genetics*, 56(6), pp. 447–455. DOI: 10.1038/jhg.2011.38.

Pośpiech, E. *et al.* (2014) 'The Common Occurrence of Epistasis in the Determination of Human Pigmentation and Its Impact on DNA-Based Pigmentation Phenotype Prediction'. *Forensic Science International: Genetics*, 11, pp. 64–72. DOI: 10.1016/j.fsigen.2014.01.012.

Posth, C. *et al.* (2023) 'Palaeogenomics of Upper Palaeolithic to Neolithic European Hunter-Gatherers'. *Nature*, 615(7950), pp. 117–126. DOI: 10.1038/s41586-023-05726-0.

Poullet, M. and Orlando, L. (2020) 'Assessing DNA Sequence Alignment Methods for Characterizing Ancient Genomes and Methylomes'. *Frontiers in Ecology and Evolution*, 8. DOI: 10.3389/fevo.2020.00105.

Prüfer, K. *et al.* (2014) 'The Complete Genome Sequence of a Neanderthal from the Altai Mountains'. *Nature*, 505(7481), pp. 43–49. DOI: 10.1038/nature12886.

R Core Team. (2022) *R: A Language and Environment for Statistical Computing*. Available at: http://www.r-project.org/index.html (Accessed: 28 June 2022).

Raben, T.G. *et al.* (2022) 'From Genotype to Phenotype: Polygenic Prediction of Complex Human Traits'. In pp. 421–446. DOI: 10.1007/978-1-0716-2205-6_15.

Raghavan, M. *et al.* (2014) 'Upper Palaeolithic Siberian Genome Reveals Dual Ancestry of Native Americans'. *Nature*, 505(7481), pp. 87–91. DOI: 10.1038/nature12736.

Rana, B.K. *et al.* (1999) 'High Polymorphism at the Human Melanocortin 1 Receptor Locus'. *Genetics*, 151(4), pp. 1547–1557. DOI: 10.1093/genetics/151.4.1547.

Rasmussen, M. *et al.* (2010) 'Ancient Human Genome Sequence of an Extinct Palaeo-Eskimo'. *Nature*, 463(7282), pp. 757–762. DOI: 10.1038/nature08835.

Renaud, G. *et al.* (2015) 'Schmutzi: Estimation of Contamination and Endogenous Mitochondrial Consensus Calling for Ancient DNA'. *Genome Biology*, 16(1), p. 224. DOI: 10.1186/s13059-015-0776-0.

Renaud, G., Stenzel, U. and Kelso, J. (2014) 'LeeHom: Adaptor Trimming and Merging for Illumina Sequencing Reads'. *Nucleic Acids Research*, 42(18), pp. e141–e141. DOI: 10.1093/nar/gku699.

Rivollat, M. *et al.* (2020) 'Ancient Genome-Wide DNA from France Highlights the Complexity of Interactions between Mesolithic Hunter-Gatherers and Neolithic Farmers'. *Science Advances*, 6(22). DOI: 10.1126/sciadv.aaz5344.

Rocha, J. (2020) 'The Evolutionary History of Human Skin Pigmentation'. *Journal of Molecular Evolution*, 88(1), pp. 77–87. DOI: 10.1007/s00239-019-09902-7.

Rogers, A.R., Iltis, D. and Wooding, S. (2004) 'Genetic Variation at the MC1R Locus and the Time since Loss of Human Body Hair'. *Current Anthropology*, 45(1), pp. 105–108. DOI: 10.1086/381006.

Rohland, N. *et al.* (2015) 'Partial Uracil–DNA–Glycosylase Treatment for Screening of Ancient DNA'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1660), p. 20130624. DOI: 10.1098/rstb.2013.0624.

Ross, A.B. *et al.* (2006) 'Lifestyle, Genetics, and Disease in Sami'. *Croatian Medical Journal*, 47(4), pp. 553–565.

RStudio Team. (2020) Available at: http://www.rstudio.com/.

Rubinacci, S. *et al.* (2021) 'Efficient Phasing and Imputation of Low-Coverage Sequencing Data Using Large Reference Panels'. *Nature Genetics*, 53(1), pp. 120–126. DOI: 10.1038/s41588-020-00756-0.

Ruiz, Y. *et al.* (2013) 'Further Development of Forensic Eye Color Predictive Tests'. *Forensic Science International: Genetics*, 7(1), pp. 28–40. DOI: 10.1016/j.fsigen.2012.05.009.

Saag, Lehti. *et al.* (2021) 'Genetic Ancestry Changes in Stone to Bronze Age Transition in the East European Plain'. *Science Advances*, 7(4). DOI: 10.1126/sciadv.abd6535.

Santos, F., Machado, H. and Silva, S. (2013) 'Forensic DNA Databases in European Countries: Is Size Linked to Performance?' *Life Sciences, Society and Policy*, 9(1), p. 12. DOI: 10.1186/2195-7819-9-12.

Schubert, M. *et al.* (2012) 'Improving Ancient DNA Read Mapping against Modern Reference Genomes'. *BMC Genomics*, 13(1), p. 178. DOI: 10.1186/1471-2164-13-178.

Schubert, M., Lindgreen, S. and Orlando, L. (2016) 'AdapterRemoval v2: Rapid Adapter Trimming, Identification, and Read Merging'. *BMC Research Notes*, 9(1), p. 88. DOI: 10.1186/s13104-016-1900-2.

Schulze, T.G. and McMahon, F.J. (2004) 'Defining the Phenotype in Human Genetic Studies: Forward Genetics and Reverse Phenotyping'. *Human Heredity*, 58(3–4), pp. 131–138. DOI: 10.1159/000083539.

Seguin-Orlando, A. *et al.* (2015) 'Amplification of TruSeq Ancient DNA Libraries with AccuPrime Pfx: Consequences on Nucleotide Misincorporation and Methylation Patterns'. *STAR: Science & Technology of Archaeological Research*, 1(1), pp. 1–9. DOI: 10.1179/2054892315Y.0000000005.

Seguin-Orlando, A. *et al.* (2014) 'Genomic Structure in Europeans Dating Back at Least 36,200 Years'. *Science*, 346(6213), pp. 1113–1118. DOI: 10.1126/science.aaa0114.

Seguin-Orlando, A. *et al.* (2021) 'Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France'. *Current Biology*, 31(5), pp. 1072-1083.e10. DOI: 10.1016/j.cub.2020.12.015.

Shriver, M.D. and Parra, E.J. (2000) 'Comparison of Narrow-Band Reflectance Spectroscopy and Tristimulus Colorimetry for Measurements of Skin and Hair Color in Persons of Different Biological Ancestry'. *American Journal of Physical Anthropology*, 112(1), pp. 17–27. DOI: 10.1002/(SICI)1096-8644(200005)112:1<17::AID-AJPA3>3.0.CO;2-D.

Sikora, M. *et al.* (2019) 'The Population History of Northeastern Siberia since the Pleistocene'. *Nature*, 570(7760), pp. 182–188. DOI: 10.1038/s41586-019-1279-z.

Simcoe, M. *et al.* (2021) 'Genome-Wide Association Study in Almost 195,000 Individuals Identifies 50 Previously Unidentified Genetic Loci for Eye Color'. *Science Advances*, 7(11). DOI: 10.1126/sciadv.abd1239.

Sirak, K.A. *et al.* (2017) 'A Minimally-Invasive Method for Sampling Human Petrous Bones from the Cranial Base for Ancient DNA Analysis'. *BioTechniques*, 62(6), pp. 283–289. DOI: 10.2144/000114558.

Skoglund, P., Malmström, H., *et al.* (2014) 'Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers'. *Science*, 344(6185), pp. 747–750. DOI: 10.1126/science.1253448.

Skoglund, P. *et al.* (2012) 'Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe'. *Science*, 336(6080), pp. 466–469. DOI: 10.1126/science.1216304.

Skoglund, P., Northoff, B.H., *et al.* (2014) 'Separating Endogenous Ancient DNA from Modern Day Contamination in a Siberian Neandertal'. *Proceedings of the National Academy of Sciences*, 111(6), pp. 2229–2234. DOI: 10.1073/pnas.1318934111.

Skoglund, P. and Mathieson, I. (2018) 'Ancient Genomics of Modern Humans: The First Decade'. *Annual Review of Genomics and Human Genetics*, 19(1), pp. 381–404. DOI: 10.1146/annurev-genom-083117-021749.

Sousa da Mota, B. *et al.* (2023) 'Imputation of Ancient Human Genomes'. *Nature Communications*, 14(1), p. 3660. DOI: 10.1038/s41467-023-39202-0.

Spichenok, O. *et al.* (2011) 'Prediction of Eye and Skin Color in Diverse Populations Using Seven SNPs'. *Forensic Science International: Genetics*, 5(5), pp. 472–478. DOI: 10.1016/j.fsigen.2010.10.005.

Sturm, R.A. *et al.* (2008) 'A Single SNP in an Evolutionary Conserved Region within Intron 86 of the HERC2 Gene Determines Human Blue-Brown Eye Color'. *The American Journal of Human Genetics*, 82(2), pp. 424–431. DOI: 10.1016/j.ajhg.2007.11.005.

Sturm, R.A. and Duffy, D.L. (2012) 'Human Pigmentation Genes under Environmental Selection'. *Genome Biology*, 13(9), p. 248. DOI: 10.1186/gb-2012-13-9-248.

Sulem, P. *et al.* (2007) 'Genetic Determinants of Hair, Eye and Skin Pigmentation in Europeans'. *Nature Genetics*, 39(12), pp. 1443–1452. DOI: 10.1038/ng.2007.13.

Svensson, E. *et al.* (2021) 'Genome of Peştera Muierii Skull Shows High Diversity and Low Mutational Load in Pre-Glacial Europe'. *Current Biology*, 31(14), pp. 2973-2983.e9. DOI: 10.1016/j.cub.2021.04.045.

The Chimpanzee Sequencing and Analysis Consortium. (2005) 'Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome'. *Nature*, 437(7055), pp. 69–87. DOI: 10.1038/nature04072.

Thomas, R.H. *et al.* (1989) 'DNA Phylogeny of the Extinct Marsupial Wolf'. *Nature*, 340(6233), pp. 465–467. DOI: 10.1038/340465a0.

Tillmar, A. *et al.* (2021) 'The FORCE Panel: An All-in-One SNP Marker Set for Confirming Investigative Genetic Genealogy Leads and for General Forensic Applications'. *Genes*, 12(12), p. 1968. DOI: 10.3390/genes12121968.

Torkamaneh, D. and Belzile, F. (2021) 'Accurate Imputation of Untyped Variants from Deep Sequencing Data'. In pp. 271–281. DOI: 10.1007/978-1-0716-1103-6_13.

Toulza, E. *et al.* (2007) 'Large-Scale Identification of Human Genes Implicated in Epidermal Barrier Function'. *Genome Biology*, 8(6), p. R107. DOI: 10.1186/gb-2007-8-6-r107.

Valenzuela, R.K. *et al.* (2010) 'Predicting Phenotype from Genotype: Normal Pigmentation*'. *Journal of Forensic Sciences*, 55(2), pp. 315–322. DOI: 10.1111/j.1556-4029.2009.01317.x.

Villalba-Mouco, V. *et al.* (2019) 'Survival of Late Pleistocene Hunter-Gatherer Ancestry in the Iberian Peninsula'. *Current Biology*, 29(7), pp. 1169-1177.e7. DOI: 10.1016/j.cub.2019.02.006.

Visconti, A. *et al.* (2018) 'Genome-Wide Association Study in 176,678 Europeans Reveals Genetic Loci for Tanning Response to Sun Exposure'. *Nature Communications*, 9(1), p. 1684. DOI: 10.1038/s41467-018-04086-y.

Walsberg, G.E. (1988) 'Consequences of Skin Color and Fur Properties for Solar Heat Gain and Ultraviolet Irradiance in Two Mammals'. *Journal of Comparative Physiology B*, 158(2), pp. 213–221. DOI: 10.1007/BF01075835.

Walsh, S. *et al.* (2014) 'Developmental Validation of the HIrisPlex System: DNA-Based Eye and Hair Colour Prediction for Forensic and Anthropological Usage'. *Forensic Science International: Genetics*, 9, pp. 150–161. DOI: 10.1016/j.fsigen.2013.12.006.

Walsh, S. *et al.* (2012) 'DNA-Based Eye Colour Prediction across Europe with the IrisPlex System'. *Forensic Science International: Genetics*, 6(3), pp. 330–340. DOI: 10.1016/j.fsigen.2011.07.009.

Walsh, S. *et al.* (2017) 'Global Skin Colour Prediction from DNA'. *Human Genetics*, 136(7), pp. 847–863. DOI: 10.1007/s00439-017-1808-5.

Walsh, S. *et al.* (2011) 'IrisPlex: A Sensitive DNA Tool for Accurate Prediction of Blue and Brown Eye Colour in the Absence of Ancestry Information'. *Forensic Science International: Genetics*, 5(3), pp. 170–180. DOI: 10.1016/j.fsigen.2010.02.004.

Walsh, S. *et al.* (2013) 'The HIrisPlex System for Simultaneous Prediction of Hair and Eye Colour from DNA'. *Forensic Science International: Genetics*, 7(1), pp. 98–115. DOI: 10.1016/j.fsigen.2012.07.005.

Walter, H. (1971) 'Remarks on the Environmental Adaptation of Man'. *Humangenetik*, 13(2), pp. 85–97. DOI: 10.1007/BF00295790.

Wang, K. *et al.* (2023) 'High-Coverage Genome of the Tyrolean Iceman Reveals Unusually High Anatolian Farmer Ancestry'. *Cell Genomics*, 3(9), p. 100377. DOI: 10.1016/j.xgen.2023.100377.

Wang, Z. and Chatterjee, N. (2017) 'Increasing Mapping Precision of Genome-Wide Association Studies: To Genotype and Impute, Sequence, or Both?' *Genome Biology*, 18(1), p. 118. DOI: 10.1186/s13059-017-1255-6.

Webb, A.R., Kline, L. and Holick, M.F. (1988) 'Influence of Season and Latitude on the Cutaneous Synthesis of Vitamin D: Exposure to Winter Sunlight in Boston and Edmonton Will Not Promote Vitamin D Synthesis in Human Skin'. *The Journal of Clinical Endocrinology & Metabolism*, 67(2), pp. 373–378. DOI: 10.1210/jcem-67-2-373.

Westgate, G.E., Botchkareva, N. V. and Tobin, D.J. (2013) 'The Biology of Hair Diversity'. *International Journal of Cosmetic Science*, 35(4), pp. 329–336. DOI: 10.1111/ics.12041.

Wheeler, P.E. (1985) 'The Loss of Functional Body Hair in Man: The Influence of Thermal Environment, Body Form and Bipedality'. *Journal of Human Evolution*, 14(1), pp. 23–28. DOI: 10.1016/S0047-2484(85)80091-9.

Wickham, H. *et al.* (2023) Available at: https://dplyr.tidyverse.org.

Wickham, H., Vaughan, D. and Girlich, M. (2024) Available at: https://tidyr.tidyverse.org.

Willerslev, E. *et al.* (2003) 'Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments'. *Science*, 300(5620), pp. 791–795. DOI: 10.1126/science.1084114.

Willerslev, E. and Cooper, A. (2005) 'Ancient DNA'. *Proceedings of the Royal Society B: Biological Sciences*, 272(1558), pp. 3–16. DOI: 10.1098/rspb.2004.2813.

Woźniak, A. *et al.* (2021) 'Development of the VISAGE Enhanced Tool and Statistical Models for Epigenetic Age Estimation in Blood, Buccal Cells and Bones'. *Aging*, 13(5), pp. 6459–6484. DOI: 10.18632/aging.202783.

Wray, N.R. *et al.* (2013) 'Pitfalls of Predicting Complex Traits from SNPs'. *Nature Reviews Genetics*, 14(7), pp. 507–515. DOI: 10.1038/nrg3457.

Xavier, C. *et al.* (2022) 'Development and Inter-Laboratory Evaluation of the VISAGE Enhanced Tool for Appearance and Ancestry Inference from DNA'. *Forensic Science International: Genetics*, 61, p. 102779. DOI: 10.1016/j.fsigen.2022.102779.

Yang, M.A. and Fu, Q. (2018) 'Insights into Modern Human Prehistory Using Ancient Genomes'. *Trends in Genetics*, 34(3), pp. 184–196. DOI: 10.1016/j.tig.2017.11.008.

Yang, Z. *et al.* (2016) 'A Genetic Mechanism for Convergent Skin Lightening during Recent Human Evolution'. *Molecular Biology and Evolution*, 33(5), pp. 1177–1187. DOI: 10.1093/molbev/msw003.

Yang, Z. *et al.* (2018) 'Darwinian Positive Selection on the Pleiotropic Effects of KITLG Explain Skin Pigmentation and Winter Temperature Adaptation in Eurasians'. *Molecular Biology and Evolution*, 35(9), pp. 2272–2283. DOI: 10.1093/molbev/msy136.

Zihlman, A.L. and Cohn, B.A. (1988) 'The Adaptive Response of Human Skin to the Savanna'. *Human Evolution*, 3(5), pp. 397–409. DOI: 10.1007/BF02447222.