Population Replacement in Early Neolithic Britain—

SUPPLEMENTARY MATERIAL

Selina Brace[[1]](#footnote-1)\*, Yoan Diekmann[[2]](#footnote-2)\*, Thomas J. Booth1\*, Zuzana Faltyskova2, Nadin Rohland[[3]](#footnote-3), Swapan Mallick3,[[4]](#footnote-4),[[5]](#footnote-5), Kristin Stewardson3,5, Matthew Ferry3,5, Jonas Oppenheimer3,5, Nasreen Broomandkhoshbacht3,5, Megan Michel3,5, Susan Walsh[[6]](#footnote-6), Manfred Kayser[[7]](#footnote-7), Rick Schulting[[8]](#footnote-8), Oliver E. Craig[[9]](#footnote-9), Alison Sheridan[[10]](#footnote-10), Mike Parker Pearson[[11]](#footnote-11), Chris Stringer1, David Reich3,4,5#, Mark G. Thomas2#, Ian Barnes1#

\* these authors contributed equally

# these authors co-supervised the work

[**Pigmentation**](#_j754ou1o1v8q) **2**

[**Supplementary Figure legends**](#_umrmt23ozf6f) **9**

[**Supplementary Table legends**](#_90ev11ak0izv) **10**

## Pigmentation

Susan Walsh, Manfred Kayser

See Supplementary Table S2 for sequencing data and allelic states that entered the analyses below. Predictions are based on methods and tools published in Walsh et al., Forensic Sci Int Genet, 2013, and Walsh et al., Hum Genet, 2017.

**La Braña (Spain, Mesolithic)**  
  
*Eye colour* —All loci are present and have good coverage.  
  
Blue eye 0.459  
Int. eye 0.155  
Brown eye 0.387  
  
Final prediction: Intermediate (hazel/green) eye colour

Explanation: All probabilities are less than 0.5 so it is a combination of all categories, as brown is relatively high, it is more than likely a light hazel eye colour individual, but perceived green (blue/yellow) cannot be ruled out.  
  
*Hair colour*—There is 1 locus (*TYRP1* rs683) with low coverage (1x), hence a heterozygote is possible. Prediction is a range that includes what the 1x coverage found (derived G allele) and the possibility of an A ancestral allele being present.  
   
 *TYRP1* rs683 *TYRP1* rs683  
 (homozygote GG) (heterozygote GA)   
Blond 0.018 0.014  
Brown 0.612 0.595  
Red 0 0  
Black 0.37 0.391  
Light 0.033 0.025  
Dark 0.967 0.975  
  
Prediction range:  
Brown 0.612 – 0.595  
Black 0.37 – 0.391  
  
Final Prediction: Black/Dark Brown hair colour

Explanation: The probability value of black is >0.25 so it has a significant impact on prediction, and will darken the high brown probability. This individual would be perceived to have black hair. Dark Brown however cannot be ruled out.  
  
*Skin pigmentation*—Only 1 locus (*BNC2* rs10756819) is missing, however the profile does contain 3 loci with low coverage (n=1x), hence a heterozygote is possible. When factoring in possible genotype combinations, a prediction range has been generated. The range consists of assuming the 3 loci with low coverage are correct as homozyogote for their sequenced allele (*ASIP* rs1667394 A allele (derived), *OCA2* rs1545397 A allele (ancestral), *TYRP1* rs683 A allele (ancestral)) and omitting *BNC* rs10756819 in the prediction model as it has no coverage, to including this marker with a homozygote ancestral G allele and also derived A allele. The following range for skin colour prediction is possible for this individual with these parameters:  
  
 Omitting rs10756819 G ancestral allele A derived allele  
Very Pale 0 0 0  
Pale 0 0 0  
Intermediate 0.174 0.042 0.205  
Dark 0.463 0.209 0.435  
Dark-Black 0.363 0.749 0.360  
  
Prediction range:  
Very Pale 0   
Pale 0   
Intermediate 0.042 - 0.205  
Dark 0.209 - 0.435  
Dark-Black 0.749 - 0.36   
  
Final prediction: Dark/Dark-to-Black skin

Explanation: The combined effect of probabilities in the dark and dark-to-black colour categories provide an indication that the individual has darkly pigmented skin, it is unlikely that this individual has the darkest possible skin pigmentation, however, it cannot be ruled out as the missing marker does influence that detail, but certainly skin colour is dark in complexion.

**Cheddar Man (UK, Mesolithic)**  
  
*Eye colour* —There is 1 locus (*LOC105374875* (formally known as *SLC24A4*) rs12896399) with low coverage (1x) hence a heterozygote is possible. Prediction includes a range that includes what the 1x coverage found (ancestral G allele) and the possibility of an A derived allele being present.  
  
 *LOC105374875* rs12896399 *LOC105374875* rs12896399  
 (homozygote GG) (heterozygote GA)  
Blue eye 0.564 0.711  
Int. eye 0.189 0.143  
Brown eye 0.247 0.145  
  
Prediction range:  
Blue eye 0.564 - 0.711  
Int. eye 0.189 - 0.143  
Brown eye 0.247 - 0.145  
  
Final prediction: Intermediate (blue/green) eye colour

Explanation: This individual has light or blue/green eye colour, it is not light blue, there are elements of brown/yellow in the eye to give a proposed perceived green colour. Better coverage at the low sequenced marker would clarify this but blue/hazel cannot be ruled out. It is certainly not a brown eyed or clear blue-eyed individual.  
  
*Hair colour*—There is 1 locus *PIGU* rs2378249 with low coverage (1x) hence a heterozygote is possible. Prediction is a range that includes what the 1x coverage found, ancestral A allele, but also includes the possibility of a heterozygote being present.  
   
 *PIGU* rs2378249 *PIGU* rs2378249  
 (homozygote AA ) (heterozygote CA)   
Blond 0.014 0.013   
Brown 0.719 0.759  
Red 0.011 0.023  
Black 0.257 0.205  
Light 1 0.999   
Dark 0 0.001  
  
Prediction range:  
Brown 0.719 – 0.759  
Black 0.257 – 0.205  
  
Final Prediction: Dark Brown/Black hair colour

Explanation: The probability value of black is >0.2 so it has an impact on prediction, and will darken the high brown probability. However there are light pigment alleles indicating a lighter shade phenotype. Better coverage at the low sequenced marker would help clarify this. This individual would be perceived as having dark brown hair. Black however cannot be ruled out.   
  
*Skin pigmentation*—There are 3 loci (*BNC2* rs10756819, *TYR* rs1126809, *MC1R* rs3212355) missing, and the profile does contain 2 loci (*LOC105374875* rs12896399 and *PIGU* rs2378249) with low coverage (n=1x) hence a heterozygote is possible at those sites. When factoring in possible genotype combinations, a prediction range may be generated. The range consists of assuming the two loci with low coverage are correct as homozygote for their sequenced allele (*LOC105374875* rs12896399 G allele and *PIGU* rs2378249 A allele) and omitting the 3 missing loci from the prediction model as they have no coverage, to including these markers with their ancestral (*BNC2* rs10756819-GG, *TYR* rs1126809-GG, *MC1R* rs3212355-CC) and also their derived allele counterparts. The following range for skin colour prediction is possible for this individual with these parameters:

ancestral alleles used derived alleles used  
Very Pale 0 0  
Pale 0 0  
Intermediate 0.394 0.125  
Dark 0 0  
Dark-Black 0.606 0.875  
  
Prediction range:   
Very Pale 0   
Pale 0   
Intermediate 0.394 - 0.125  
Dark 0 - 0  
Dark-Black 0.606 - 0.875

If we omit the three missing alleles, our tool produces 0.891 and 0.109 probabilities for the intermediate and dark-black category respectively, changing the prediction ranges to 0.891-0.125 and 0.109-0.875. However, note that this completely removes the locus from the prediction model, hence the prediction will not perform optimally (how the prediction model was made), therefore it is best to have some allele present to infer the most probable range for Cheddar Man and we therefore derive the ranges above from the extreme allele constellations only.

Final prediction: Dark/Dark-to-black skin

Explanation: The missing loci certainly impact on this prediction; however utilizing the input of all ancestral alleles is the preferred option over the use of the derived alleles at these loci – hence 0.394 for intermediate and 0.606 for Dark-black would be the most probable profile. That being said a broad range is present in both the intermediate and dark-black categories due to the missing loci. Also this effect, of skipping a skin colour prediction category with regards probability values, tends to be observed more often in admixed individuals. What is important to note is the input of the dark-black prediction is significant on the intermediate category and therefore it is acceptable to propose a dark complexion individual over an intermediate/light prediction even though the intermediate range is large. It is unlikely that this individual has the darkest possible pigmentation, however it cannot be ruled out. Better sequencing coverage would clarify to what degree this individual has a dark complexion.  
  
  
  
**Sven (UK, Neolithic)**  
  
*Eye colour* —All loci are present and have good coverage. Artefact bases are proposed for locus *SLC45A2* rs16891982 (T allele).   
  
Blue eye 0.022  
Int. eye 0.090  
Brown eye 0.887  
  
Final prediction: Brown eye colour

Explanation: The highest probability is well above the threshold 0.7p for brown, so a strong brown prediction is proposed.  
  
*Hair colour*—All loci are present and most have good coverage. An artefact base is proposed for locus *SLC45A2* rs16891982 (T allele). For loci *SLC45A2* rs28777, *OCA2* rs12441727, *OCA2* rs1470608, they are assumed to be heterozygotes, although coverage is low (1x) for one of the alleles. Artefact bases are proposed for locus *MC1R* rs1110400 (A allele) and *MC1R* rs885479 (A allele) as there is >40x coverage for the more represented allele, therefore it is assumed that it is not a heterozygote at these loci. There is 1 locus (*TYRP1* rs683) with low coverage (1x) hence a heterozygote is possible. Prediction is given as a range that includes what the 1x coverage found (derived G allele), and the possibility of an A ancestral allele being present at this locus for hair colour prediction.  
  
 *TYRP1* rs683 *TYRP1* rs683  
 (homozygote GG) (heterozygote GA)   
Blond 0.029 0.024  
Brown 0.583 0.566  
Red 0 0  
Black 0.387 0.409  
Light 0.066 0.051  
Dark 0.934 0.949  
   
Prediction range:  
Brown 0.583 – 0.566  
Black 0.387 – 0.409  
  
Final Prediction: Black/Dark Brown hair colour

Explanation: The probability value of black is >0.25 so it has a significant impact on prediction, and will darken the high brown probability. This individual would be perceived to have black hair. Dark Brown however cannot be ruled out.  
  
*Skin pigmentation*—All loci are present and most have good coverage. An artefact base is proposed for locus *SLC45A2* rs16891982 (T allele). For loci *SLC45A2* rs28777, *OCA2* rs12441727, *OCA2* rs1470608, they are assumed to be heterozygotes although coverage is low (1x) for one of the alleles. Additional sequencing of this marker would clarify this. Artefact bases are proposed for locus *MC1R* rs1110400 (A allele) and *MC1R* rs885479 (A allele) as there is >40x coverage for the more represented allele, therefore it is assumed that it is not a heterozygote at these loci. There is 1 locus (*TYRP1* rs683) with low coverage (1x) hence a heterozygote is possible. Prediction is given as a range that includes what the 1x coverage found (derived G allele), and the possibility of an A ancestral allele being present at this locus for skin colour prediction.  
  
 *TYRP1* rs683 *TYRP1* rs683  
 (homozygote GG) (heterozygote GA)   
Very Pale 0 0.008   
Pale 0.142 0.117   
Intermediate 0.673 0.565   
Dark 0.055 0.116   
Dark-Black 0.122 0.195   
  
Prediction range:   
Very Pale 0 – 0.008   
Pale 0.142 – 0.117   
Intermediate 0.673 - 0.565  
Dark 0.055 – 0.116  
Dark-Black 0.122 - 0.195  
   
Final prediction: Intermediate/Dark skin

Explanation: The effect of probability in the dark-to-black colour category has an impact on the high intermediate prediction. It is highly unlikely this individual has the darkest possible skin pigmentation and taken collectively, these probabilities indicate that the individual would fall more into an intermediate skin colour category. However dark cannot be definitively ruled out.  
  
  
  
  
  
**Loschbour (Luxembourg, Mesolithic)**  
  
*Eye colour* —All loci are present and have good coverage.  
  
Blue eye 0.564  
Int. eye 0.189  
Brown eye 0.247  
  
Final prediction: Intermediate (blue/green) eye colour

Explanation: This individual has light or blue/green eye colour, it is not light blue, there are elements of brown/yellow in the eye to give a proposed perceived green colour, however blue/hazel cannot be ruled out. It is certainly not a brown eyed or clear blue-eyed individual.  
  
*Hair colour*—All loci are present and have good coverage. Artefact bases are proposed for locus *TYR* rs1126809 (A allele) and *HERC2* rs1667394 (G allele) as there is >20x coverage for the more represented allele, therefore it is assumed that it is not a heterozygote at these loci.  
   
Blond 0.005   
Brown 0.532   
Red 0   
Black 0.463   
Light 0.022   
Dark 0.978   
  
Final Prediction: Black/Dark Brown hair colour

Explanation: The probability value of black is >0.25 so it has a significant impact on prediction, and will darken the high brown probability. This individual would be perceived to have black hair. Dark Brown however cannot be ruled out.  
  
*Skin pigmentation*—All loci are present and have good coverage.  
There is a similar artefact assessment as above for rs1126809 and rs1667394.   
  
Very Pale 0   
Pale 0   
Intermediate 0.893   
Dark 0.069   
Dark-Black 0.038   
  
Final prediction: Intermediate skin

Explanation: The highest probability of approximately 0.9 for intermediate indicates a light skinned (white) individual. He would not have the darkest possible skin pigmentation but does have tanning ability, so could be perceived as darker than white (pale) in the summer months.

## Supplementary Figure legends

**Supplementary Figure S1**: *Heatmap of Mesolithic individuals.* Heatmap of pairwise outgroup *f3* statistics between Mesolithic individuals presented here and a set of ancient Eurasians from different hunter-gatherer groups such as Western-, Eastern- and Scandinavian hunter gatherers (see Supplementary Table S1 for references).

**Supplementary Figure S2**: *f4 statistics between hunter gatherer groups.* We compare the affinities of the Mesolithic individuals presented here towards El Mirón and Villabruna.

**Supplementary Figure S3**: *f4 statistics between hunter gatherer groups.* We compare the affinities of the Mesolithic individuals presented here towards Eastern- (EHG) and Western hunter gatherers (WHG). See Supplementary Table S1 for the information which individuals were grouped to form EHG and WHG.

**Supplementary Figure S4**: *f4 statistics between hunter gatherer groups.* We compare the affinities of the Mesolithic individuals presented here towards Scandinavian- (SHG) and Western hunter gatherers (WHG). See Supplementary Table S1 for the information which individuals were grouped to form SHG and WHG.

**Supplementary Figure S5**: *f4 statistics between different WHG individuals.* The upper panel compares affinities of each ancient British individual analysed here towards the Mesolithic individuals La Brana and Loschbour. The lower panel repeats the analysis grouping individuals temporally and where possible geographically, see Supplementary Table S1 for the information which individuals were grouped.

**Supplementary Figure S6**: *f4 statistics between different WHG individuals.* The upper panel compares affinities of each ancient British individual analysed here towards the Mesolithic individuals KO1 and Loschbour. The lower panel repeats the analysis grouping individuals temporally and where possible geographically, see Supplementary Table S1 for the information which individuals were grouped.

**Supplementary Figure S7**: *f4 statistics between different WHG individuals.* The upper panel compares affinities of each ancient British individual analysed here towards the Mesolithic British Cheddar Man whose genome is presented here and Loschbour. The lower panel repeats the analysis grouping individuals temporally and where possible geographically, see Supplementary Table S1 for the information which individuals were grouped.

**Supplementary Figure S8**: *Individual f4 admixture proportions*. We estimate the WHG and Anatolian farmer ancestry proportions for each ancient British individual analysed here.

**Supplementary Figure S9**: *f4 statistics between different Central European and Iberian Early Neolithic.* We compare the affinities of all individuals presented here towards Central European (CentralEur EN) and Iberian Early Neolithic (Iberia EN) populations. See Supplementary Table S1 for the information which individuals were grouped to form the CentralEur EN and Iberia EN).

**Supplementary Figure S10**: *Heterozygosity estimates for ancient and modern British individuals*. We compare heterozygosity estimates between Mesolithic Cheddar Man, Neolithic Carsington Pasture 1 (‘Sven’), and two modern British and two modern Yoruba individuals from the 1000 genome project (1000 Genomes Project Consortium *et al.*, Nature, 2015).

## Supplementary Table legends

**Supplementary Table S1**: *Summary of sequencing data.* We list all individuals analysed as part of this paper. Number of reads are before removing duplicates and MAPQ filtering. Coverage is computed as the number of bases divided by the total number of nuclear positions in the reference genome without counting positions that belong to intervals of five or more consecutive N’s. Read depth refers to the average number of times a set of positions is covered given a position is covered at least once. Duplicates are marked with Picard tools and counted with Samtools (Li, Bioinformatics, 2011).

*Abbreviations*: nuclear (NUC), Mitochondrial (MT), number of (nb.)

**Supplementary Table S2**: *Functional variation.* List of SNPs and alleles found in the individuals newly sequenced here based on which phenotypic characteristics have been predicted.

1. Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK [↑](#footnote-ref-1)
2. Research Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, UK [↑](#footnote-ref-2)
3. Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA [↑](#footnote-ref-3)
4. Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA [↑](#footnote-ref-4)
5. Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA [↑](#footnote-ref-5)
6. Department of Biology, Indiana University Purdue University, Indianapolis (IUPUI), Indianapolis, IN, USA [↑](#footnote-ref-6)
7. Department of Genetic Identification, Erasmus MC University Medical Centre Rotterdam, Rotterdam, The Netherlands [↑](#footnote-ref-7)
8. Institute of Archaeology, University of Oxford, 36 Beaumont St, Oxford, OX1 2PG [↑](#footnote-ref-8)
9. BioArCh, Department of Archaeology, University of York, UK [↑](#footnote-ref-9)
10. National Museums Scotland, Edinburgh EH1 1JF, UK [↑](#footnote-ref-10)
11. Institute of Archaeology, University College London, London WC1H 0PY, UK [↑](#footnote-ref-11)