Genomic signals of migration and continuity in Britain before the Anglo-Saxons

Authors: Rui Martiniano[1], Anwen Caffell[2,3], Malin Holst[2], Kurt Hunter-Mann[4], Janet Montgomery[3], Gundula Müldner[5], Russell L McLaughlin[1], Matthew D Teasdale[1], Wouter van Rheenen[6], Jan H Veldink[6], Leonard H van den Berg[6], Orla Hardiman[7], Maureen Carroll[8], Steve Roskams[9], John Oxley[10], Colleen Morgan[9], Mark G Thomas[11], Ian Barnes[12], Christine McDonnell[4], Matthew J Collins[9]*, Daniel G Bradley[1]*

Affiliations:

[1]Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland;

[2]York Osteoarchaeology Ltd, 75 Main St, Bishop Wilton, York, North Yorkshire YO42 1SR, UK

[3]Department of Archaeology, Dawson Building, Durham University, South Road, Durham, DH1 3LE, UK

[4]York Archaeological Trust for Excavation and Research Limited, 47 Aldwark, York YO1 7BX, UK

[5]Department of Archaeology, University of Reading, Whiteknights PO Box 227, Reading RG6 6AB, UK

[6]Department of Neurology, Neurogenetics Unit, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

[7]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Pearse Street, Dublin 2, Ireland

[8]Department of Archaeology, University of Sheffield Northgate House, West Street, Sheffield S1 4ET UK

[9]Department of Archaeology, University of York, York, UK

[10]City of York Council, York, North Yorkshire, YO1 6GA, UK;

[11]Research Department of Genetics, Evolution and Environment, University College London, London, UK

[12]Department of Earth Sciences, Natural History Museum, London, UK.

ABSTRACT

The purported migrations that have formed the peoples of Britain have been the focus of generations of scholarly controversy. However, this has not benefited from direct analyses of ancient genomes. Here we report nine ancient genomes (~1X) of individuals from northern Britain: seven from a Roman era York cemetery, bookended by earlier Iron Age and later Anglo-Saxon burials. Six of the Roman genomes show affinity with modern British Celtic populations, particularly Welsh, but significantly diverge from populations from Yorkshire and other eastern English samples. They also show similarity with the earlier Iron Age genome, suggesting population continuity, but differ from the later Anglo-Saxon genome. This pattern concords with profound impact of migrations in the Anglo-Saxon period. Strikingly, one Roman skeleton shows a clear signal of exogenous origin, with affinities pointing towards the Middle East, confirming the cosmopolitan character of the Empire, even at its northernmost fringes.

Introduction

Ancient genomics has the power to anchor the emergence of modern genetic patterns to archaeological events but, to date, no such genome-scale data has emerged for the Romano-British world, or indeed from any era in the British past. Extensive surveying of modern genomic variation in the British Isles has produced divergent interpretations of the migratory history of the islands. An east-west gradient of Y chromosome, autosomal and mtDNA allele frequencies has been interpreted as reflecting the genetic legacy of substantial Anglo-Saxon invasions following the Roman period [1–4]. However, it is difficult to distinguish the effects of this much-debated event from other migratory influences from northwest continental Europe, whether these are, for example, Germanic elements in the Late Roman army predating the Anglo-Saxon migrations or Scandinavian settlers arriving some centuries afterwards [5].

At its maximum, the Roman Empire stretched from Atlantic Europe to the Near East and from Northern Britain to the Sahara, incorporating an advanced transport infrastructure which would have enabled previously unprecedented levels of mobility [6,7]. Evidence for the presence of foreigners in Britain has been based on epigraphic sources [8], material culture [9] and, more recently, bioarchaeological (isotopic) data [10–13]. However, there is no way of knowing how representative the people mentioned in inscriptions might be; artefactual imports indicate contact beyond the province but may not point to movement of people; and isotopic baseline values in British data overlap considerably with numerous other regions, including much of Western Europe and the Mediterranean littoral [13]. Thus measuring who moved about the Roman Empire, and on what scale, remains challenging.

In order to investigate the genetic identity of Britain in late BC and the the early centuries AD we report shotgun genomic sequencing of nine human genomes to coverage depth of ~1X and analysis of these in the context of extensive genome wide data from modern populations. Seven ancient genomes are sampled from a cemetery in Roman York between the 2nd into the 4th century AD, one from an earlier Yorkshire Iron Age burial (210 BC - 40 AD) and one from a later neighbouring Anglo-Saxon burial (650-910 AD).

Results

Archaeological samples

York (*Eboracum*), founded c. AD 71, became the Roman empire's northernmost provincial capital in about AD 200. Its southwest approach road was lined with tombstones and mausolea [14] and excavations conducted there between 2004 and 2005 at Driffield Terrace revealed a cemetery dating from the 2nd into the 4th century AD with a high incidence of decapitated remains [15,16]. After initial screening of eight individuals chosen for DNA extraction, seven presented superior endogenous DNA content. Near Melton, East Yorkshire, a first century AD late Iron Age settlement gave human remains predominantly of adult females and non-adults [17] from which 5 individuals were sampled. From these, the best preserved individual with known date was M1489 (between 210 BC and 40 AD). Norton Bishopsmill, dates to 650-910AD and was a Christian Anglo-Saxon cemetery excavated in the village of Norton, Teesside, north east England [18]. We sampled 3 from burials of 100 skeletons and selected the best preserved, NO3423, for the present study (Supplementary Note 1, Supplementary Fig. 1, Supplementary Methods).

Sequencing results and sample contamination

Gamba *et al*. [19] established that the dense internal centre of the inner ear petrous bone is an excellent source of preserved ancient DNA suitable for high throughput sequencing. Accordingly, we sampled single petrous bones, extracted DNA and made indexed Illumina sequencing libraries. After preliminary screening for endogenous content, nine samples were chosen for genome level sequencing. On alignment to the human reference genome, reads showed mismatch patterns typical of archaeological DNA indicating deamination damage (Supplementary Fig. 6) [20]. Contamination estimated from both mtDNA heterozygosity (1.82±0.47%) and X-chromosome contamination in male samples (0.79±0.21%) was low (Supplementary Methods, Supplementary Tables 5-7).

Sex determination and uniparental marker analysis

Using the ratio between sequencing reads aligned to the X and Y chromosomes [21], it was possible to assign biological sex to each individual, confirming skeletal assessments: the Anglo-Saxon and each Roman period sample were male whereas the Iron Age sample was female (Supplementary Fig. 7). Mitochondrial genomes were retrieved for each sample with between 39X and 98X coverage and were assigned to known haplogroups (Table 1) which are common in present-day European populations [22]. Y-chromosome haplogroups were determined for each male (Table 1): the majority (6/7) of Driffield Terrace samples belong to sub-lineages of R1b-L52/L11, which reaches its highest frequencies (>70%) in Western European countries [23]. Sample 3DRIF-26, on the other hand, despite belonging to

the same burial context, presented a lineage consistent with haplogroup J2-L228, which has a modern distribution centred on the Middle East but which is also present in the Caucasus region, the Balkans and Italy [24]. The Anglo-Saxon (NO3423) sample was assigned to haplogroup I1-S107 which is widespread in Nordic countries [25].

{Table 1 about here}

Affinity with global populations

We called between ~210 and ~400 thousand single nucleotide polymorphisms within our ancient samples that had previously been genotyped in a dataset of 780 European, West Asian, North African and Middle Eastern individuals [26]. Fig. 1a shows a principal components analysis (PCA) where eight of nine ancient genomes cluster close to a collection of northwest European samples. One York Roman, 3DRIF-26, gives a clear Middle Eastern signal, with closest neighbours of Palestinian, Jordanian and Syrian origin. This dichotomy is also apparent in maximum likelihood estimation of individual ancestries using NGSadmix (Fig. 1b). In this, when a model of three ancestral populations is imposed across the entire sample, this analysis highlights three major geographical foci: Europe, North Africa and West Asia/Middle East. The European ancestral component predominates in the majority of ancient samples (which show similar profiles to modern northwestern Europeans), whereas 3DRIF-26 again shows a majority West Asian/Middle Eastern component. Isotopic analyses of the skeletons support this genetic differentiation of 3DRIF-26 from the remainder of the individuals sampled. Strontium isotope ratios ($^{87}Sr/^{86}Sr$) vary mainly according to geological substrate, while oxygen isotope values ($\partial^{18}O$), which track locally available drinking water, reflect climatic and geographic variables such as temperature, rainfall levels or distance from the coast [27,28]. When we compared these ratios in our seven samples to other British Romans 3DRIF-26 showed both an unusually low $^{87}Sr/^{86}Sr$ ratio and an extreme $\partial^{18}O_p$ value (Supplementary Fig. 2).

{Figure 1 about here}

In order to maximise resolution of genetic affinity, for each ancient sample we performed pairwise comparisons with each modern sample and calculated the proportion of SNP positions at which these were identical by state (IBS). Only single SNP alleles were considered at each locus and were randomly sampled from the biallelic genotype. Taking the median IBS score for each modern population sample, we then ranked these for similarity to each ancient genome in turn. Interestingly, the top ranked modern sample for IBS for each of these ancient British samples was one of the formerly Celtic language speaking regions of the British Isles, with the single exception of 3DRIF-26

which showed highest IBS with samples from Saudi Arabia. We gauged the sensitivity of this approach by checking whether individual modern samples were assignable to their region of origin. When tested, local individuals were assigned with high frequency (0.97) to the British Isles and also most often to their correct country. The method showed lower sensitivity for Middle Eastern genotypes, with primary assignment to that region in only 39% of instances (Supplementary Fig. 15). Nevertheless, outside assignments tended to be to Cypriot, Sardinian and Druze; never to Northern Europe. In contrast, specificity of a Middle Eastern assignment was high - only three individuals, from Iran, Tunisia and Morocco, were incorrectly assigned to that region. Thus assignment of 3DRIF-26 to the Middle East region seems secure but resolution to an individual population may not be possible. Specificity in assignment to the British Isles was lower, with about half of assignments (0.53) derived from elsewhere, most often from neighbouring populations such as France (0.28) and Norway (0.15). Small sample sizes (~10 per population) render individual scores only weakly informative but, when we compared the six Roman burials after excluding the outlying 3DRIF-26, their rank orders across the geographical sample were highly correlated (Spearman rank correlation coefficients ($r=0.982$; $p<0.01$, Supplementary Methods, Supplementary Figs. 12-14, Supplementary Tables 11-12). This allowed us to consider these together and generate a combined percentile score by scoring each comparator population as the product of the rank percentiles vs. each Roman genotype (Fig. 2). The Welsh were most consistently ranked as highly identical by state, followed by Irish and Scottish scores, a result strongly supporting an origin within the British Isles for this Roman sample majority. Interestingly, the modern English sample was ranked only ninth for IBS to the Romans from York, at a level similar to German, Norwegian, Orcadian and Basque samples.

{Figure 2 about here}

Ancient sample ancestry within Britain

In order to place our ancient genomes within a detailed British context, we next plotted these in a background PCA using 3,075 published genotypes from British [3], Irish [29] and southern Netherlands samples [30]. The modern samples were analysed using SNP genotypes at ~250,000 loci and projected into a single plot using smartpca (Fig. 3a). As in Burton *et al*. [3] the first component of the variation was informative for the structure within Britain. Given the close ancestral relationships between these populations and their well-known history of migrational exchange, a substantial overlap between regional groups was both expected and observed. However, by considering median values one can see a clear progression from Irish samples at one pole through Scottish, Welsh, English to the Dutch

cohort at the other extreme. In this plot the York Romans cluster centrally close to the modern Welsh median value, along with the Iron Age genome. The local Anglo-Saxon is placed differently, closest to modern East Anglians between the English and Dutch medians.

This first component also offers an opportunity to compare within the English sample. Fig. 3b shows a boxplot of PC1 values for each subsample and structure is evident, with higher median values in Eastern regions such as East Anglia, East Midlands, intermediate values in the southern and western parts and lower values in the North and North West. This pattern is more clearly seen in a geographical plot of interpolated values (Fig. 5a). When the York Romans are compared together with each modern cohort, they are most similar to the Welsh distribution of PC1 values and differ significantly from all other regional groups, apart from those from North and North West England (Mann-Whitney test; Fig. 3b; Supplementary Methods, Supplementary Table 12). An interesting difference is the marked one between the Driffield Terrace ancient and contemporary Yorkshire samples ($p=0.003$), implying regional discontinuity. It is also worth noting that the PC1 coordinate of the Anglo-Saxon individual is closer to the median PC1 value of East Anglians, possibly reflecting a more pronounced contribution of Germanic immigrants to eastern British populations. However, we note the inherent uncertainty in drawing inference from a single sample.

{Figure 3 about here}

Leslie et al. [4] used a haplotype-based clustering method, implemented in fineSTRUCTURE [31] to deconstruct British populations into geographically and historically meaningful clusters. Accordingly, we used this approach to search for substructure within the modern British genotypes used here and to identify patterns of allele sharing between the ancient samples and the clusters identified. Structure was apparent with separation into subgroups of predominantly Welsh, English and Scottish provenance. The six Romans shared most alleles on average with those clusters consisting primarily of Welsh individuals (Fig. 4 and Supplementary Fig. 16). Allele sharing patterns also allowed comparison with the other ancients; when median IBS values across the clusters were compared between those for the Iron Age genome and those for the Roman cohort these correlated strongly ($r=0.74$, $P=0.004$), supporting continuity. However, a comparison between the Anglo Saxon and the Romans showed no correlation ($r=0.06$, $P=0.842$) (Fig. 4).

{Figure 4 about here}

Genomic change in Yorkshire between the early centuries AD and modern sampling is further illustrated by both the oldest and most extensively typed genetic system (the ABO blood group) and

the system known to show maximal differentiation within the British Isles (chromosome Y haplogroup R1b1a2-M269). The plots in Fig. 5 show imputed blood group O and observed chromosome Y haplogroup R1b1a2-M269 frequencies for the Roman genomes (excluding the immigrant outlier) contrasting sharply with interpolated allele frequencies for modern eastern Britain [1,32–34].

{Figure 5 about here}

Imputation and phenotype determination

Using a similar approach to that of Gamba et al. [19], we used phased reference genomes from the 1000 Genomes Project to impute genotypes associated with phenotypic traits. In particular, we inferred genotypes at SNP positions to predict eye and hair pigmentation [35]. The most common predicted phenotype in the Roman burial samples is brown eyes and black/brown hair. However, one sample, 6DRIF-18, was estimated to have had a distinctive appearance with blue eyes and blonde hair, as did the single Anglo-Saxon individual. We also inferred that blood group O is the most common in the Roman samples (Supplementary Table 17). The Iron Age sample is also estimated as blood type O and the Anglo-Saxon is likely to have been type B or possibly type A. Five samples returned imputed lactase persistence genotypes: two Roman burials and the Iron Age individual were likely to have been lactase persistent while two Romans, 6DRIF-22 and the suspected migrant, 3DRIF-26, were homozygous for the ancestral non-persistence variant.

Discussion

Combined genomic and isotopic evidence support the inference that the origins and childhood of individual 3DRIF-26 lay far outside Britain. His modern genomic affinities clearly lie with the Middle East. Isotopically, the most plausible suggestion is an arid environment on igneous or limestone geology, which is consistent with the same regions (Supplementary Fig. 2, Supplementary Methods). Hence, although this individual is indistinguishable from the other inhumations in terms of burial practice and osteology, the analyses show that, even in its northernmost provincial capital, the profoundly cosmopolitan nature of the Roman Empire suggested by documentary and epigraphic sources continued to hold sway.

The peoples of Britain show marked genetic structure (Fig. 5) which has been the focus of generations of investigation [36]. A common theme in this research has been a contrast between a southern and eastern lowland zone and a western and northern upland zone. This patterning resembles the

geography of Anglo-Saxon settlement in the 5th-7th centuries AD, inviting the conclusion that the cultural and linguistic change effected by this migration was also reflected by major genetic change [1,2,37].

Projections from modern data to the past are, however, subject to considerable uncertainties and may be compounded by unknown complexities which do not feature in their underlying models. Prehistorians point out that the Germanic affinity of eastern Britain could also be a result of earlier communications with the north-west European mainland. For example, there may have been 'Belgic' peoples in Britain at the time of the Claudian conquest in AD43 [38], and the Roman army which arrived in Britain was composed of recruits from various provinces [39]. Recently Leslie *et al*. [4] have used haplotype-based statistical methods applied to modern genome-wide SNP genotypes to infer several distinct ancestral influences from migratory events into Britain. This included a major 35% contribution to modern Central and Southern English populations from a German source which, they surmise, occurred in the century after AD800, some 200 years or more after archaeological evidence for initial Anglo-Saxon influence. Evidence from direct observations of ancient genomes is required, however, if we are to draw conclusions about genetic exchange which distinguish between closely-dated events.

Six of the seven individuals sampled here are clearly indigenous Britons in their genomic signal. When considered together, they are similar to the earlier Iron Age sample, whilst the modern group with which they show closest affinity are Welsh. These six are also fixed for the Y chromosome haplotype R1b-L51, which shows a cline in modern Britain, again with maximal frequencies among western populations. Interestingly, these people differ significantly from modern inhabitants of the same region (Yorkshire and Humberside) suggesting major genetic change in Eastern Britain within the last millennium and a half. That this could have been, in some part, due to population influx associated with the Anglo-Saxon migrations is suggested by the different genetic signal of the later Anglo-Saxon genome. Iron Age, Roman, Anglo-Saxon, Viking and other migrations have all been proposed as contributors to the genetic structure in modern UK [40].

The thesis that the mountainous regions of Wales may have held populations that are representative of earlier more widely dispersed indigenous British genetic strata is not new, yet it finds some support in our analyses. The genomes of modern Scottish and Irish populations diverge from this group of early inhabitants of northern Britain, whereas their Welsh counterparts do not. Modern data for genetic structure among non-Saxon samples from the British Isles is said to deny the existence of a single "Celtic" population [4]. Our data indicate that differentiation within such groups may have happened

before the early centuries AD. By the same token, it lends support for genetic exchange between Scotland and Ireland, as attested in some historical sources and mirrored by linguistic affinity: Irish and Scottish Gaelic are sister P-Celtic languages, whereas Welsh is a divergent Q-Celtic language, akin to that thought to have been spoken throughout pre-Roman Britain [41].

In the Roman York burials at Driffield Terrace, the majority were adults under 45 years old, male and most had evidence of decapitation [42]. They were slightly taller than average for Roman Britain, displayed a high occurrence of trauma, potentially related to interpersonal violence, and evidenced childhood stress and infection (Supplementary Note 1). This demographic profile resembles the population structure in a recently excavated burial ground of the second and third century A.D. at Ephesus [43] which has been interpreted to be a burial ground for gladiators [44]. However, the evidence could also fit with a military context; the Roman army had a minimum height for recruitment [45] and fallen soldiers would match the young adult profile of the cemetery. In this later Roman period increasingly large numbers of soldiers were enlisted locally [39].

Whichever the identity of the enigmatic headless Romans from York, our sample of the genomes of seven of them, when combined with isotopic evidence, indicate six to be of British origin and one to have origins in the Middle East. This is the first refined genomic evidence for far-reaching ancient mobility and (although from an unusual context) also the first snapshot of British genomes in the early centuries AD, indicating continuity with an Iron Age sample prior to the migrations of the Anglo-Saxon period.

Methods

Isotope analysis

Partial isotope data for six of the seven Driffield Terrace individuals has been previously published [12,46 47,48]. We sampled a molar tooth from the remaining individual 3DRIF-26 for isotope analysis and re-attempted on long bones of individuals where rib samples had previously failed to yield a viable product. Analysis were done at the NERC Isotope Geosciences Laboratory, Keyworth and Stable Isotope Laboratories, Universities of Bradford and Reading (UK). (Supplementary Methods, Supplementary Figs. 2-5, Supplementary Tables 3-4).

DNA sampling, extractions and sequencing

Ancient DNA sample processing was done at the Ancient DNA lab, Smurfit Institute, Trinity College Dublin (Ireland), in clean-room facilities exclusively dedicated for this purpose. We extracted DNA from ~150 mg of 9 temporal bone samples belonging to Iron Age (1), Roman (7), and Anglo-Saxon (1) burial sites in York (UK; Supplementary Note 1) using a modified [49] silica-column-based method [50]. DNA libraries were constructed from extracted DNA using the method described in [51] with modifications [19,52]. We amplified the aDNA (ancient DNA) libraries with 3-4 distinct indexing oligos for each sample in order to increase index diversity in each lane. PCR products were then purified (Qiagen MinElute PCR Purification Kit, Qiagen, Hilden, Germany), quantified (Agilent Bioanalyzer 2100) and pooled. Each sample was sequenced to approximately 1X in a Illumina HiSeq 2000 (100 cycle kit, single-end reads mode; Macrogen) (Supplementary Methods).

Read processing and analysis

Next-generation sequencing reads were trimmed with Cutadapt v. 1.3[53]. Two bases from each side of reads were removed with seqtk (https://github.com/lh3/seqtk). Reads were aligned to the human reference genome (UCSC hg19) and mtDNA (rCRS, NC_012920.1) with Burrows-Wheeler Aligner (BWA) v.0.7.5a-r405 [54], filtering by base quality 15 and disabling seed length as recommended for aDNA data [55], discarding PCR duplicates and reads with mapping quality inferior to 30 using SAMtools v.0.1.19-44428cd [56] (Table 1). Base quality scores were rescaled with mapDamage v.2.0 [57] to exclude potential deamination residues from subsequent analysis.

Contamination Estimates and Authenticity

To determine the extent of contamination in the ancient samples sequenced, we calculated the number of mismatches in mtDNA haplotype defining mutations [19] and X-chromosome polymorphisms in samples determined to be male (Supplementary Methods, Supplementary Tables 5-8) [58,59]. We also used PMDtools [60] to select reads with evidence of deamination and compared sex determination and PCA (Supplementary Methods, Supplementary Figs. 7-8). Finally, we confirmed the presence of aDNA misincorporations by analysing a subset of 1 million reads for each sample with mapDamage 2.0 [57].

Sex determination and uniparental lineage determination

We used the method published in reference [21] to determine the sex of the ancient individuals (Supplementary Fig. 7). Y-chromosome lineages of ancient male samples were identified using clean_tree software [61] (http://www.erasmusmc.nl/fmb/resources/cleantree/, Supplementary Table 10). Regarding mtDNA analysis, NGS reads were separately aligned to the revised Cambridge Reference

Sequence (rCRS; NC_012920.1) [62], filtering for base ($q \geq 20$) and mapping ($q \geq 30$) quality and duplicate reads as above. We then used SAMtools to obtain a consensus sequence in fasta format which we uploaded to HaploFind [63], which determines mtDNA haplogroups based on Phylotree build 16 [64] (Table 1, Supplementary Methods). Mitochondrial mutations detected in each sample are shown in Supplementary Table 9.

Population genetics analysis and datasets

Alleles identified with Genome Analysis Toolkit v.2.5 (GATK) [65], Pileup mode, were haploidized following reference [66]. For comparisons with modern human populations, we used two datasets: 780 individuals of European, Middle Eastern, West Asian and North African populations from reference [26] and the other of Wellcome Trust Case Control Consortium (WTCCC1) 1958 British Birth Cohort SNP genotype data [3] with Dutch [30] and Irish [29,67] genotypes (Supplementary Methods). Principal Component Analysis (Fig. 1a, Fig. 3a, Supplementary Figs. 8-9) were performed using smartpca from the EIGENSOFT v.5 package [68,69]. Model-based clustering analysis was done with ADMIXTURE v.1.23 [70]. We then extracted Genotype Likelihoods (GLs) from aDNA data with ANGSD v.0.592 [59] (Supplementary Fig. 11), which we combined with genotype data of present-day populations. We analysed this data with NGSadmix v.32 [71] (K=3) and plotted with distruct v.1.1 [72] (Fig 1b, Supplementary Methods).

Identity-by-state analysis

Identity-by-state (IBS) between ancient and present-day samples was estimated with PLINK v.1.9 [73]. Median IBS proportions between aDNA samples and European, Middle Eastern and North African populations were obtained and plotted individually on maps (Supplementary Fig. 12). Then, we selected only the Roman York samples, except for 3DRIF-26 (Middle Eastern affinity), ranked their median IBS score in relation to modern populations and combined these ranks by calculating their product (Fig. 2). Spearman rank correlations were estimated with R [74] (Supplementary Methods, Supplementary Table 11). Regarding the WTCCC1 genotypes, we followed the same approach as described above, but present scaled median IBS values to the 0-1 range on Fig. 4.

fineSTRUCTURE analysis

We randomly selected 100 individuals from each region of the WTCCC1 dataset, excluding SNPs with missing genotypes, which resulted in a total of 431,366 variants and 1,000 samples. We used SHAPEIT2 to phase genotypes [75] and ran the ChromoPainter pipeline [31] with default parameters as implemented by fineSTRUCTURE v.2. For the fineSTRUCTURE analysis, the following settings were used: 3,000,000 burn in iterations, 1,000,000 sample iterations for the MCMC and 10,000,000

tree comparisons (Supplementary Methods). We then called genotypes in our ancient samples for estimation of IBS between these and the inferred fineSTRUCTURE population clusters (Fig 4, Supplementary Fig. 16-17).

Imputation of phenotype associated loci

In order to investigate loci associated with selective sweeps, we took a similar approach as in reference [19], where alleles observed in the 1000 Genomes Project [76] were called with GATK, from which we extracted genotype likelihoods and converted to BEAGLE format. BEAGLE v.3.3.2 [77] was used to phase and subsequently impute genotypes at SNP positions described in the HIrisplex system [35], loci associated with blood groups [78,79], lactase persistence [80,81] and pigmentation phenotypes [82,83]. Only posterior genotype probabilities greater or equal than 0.85 were kept [19] (Supplementary Methods; Supplementary Table 14-17). We generated interpolated frequency maps of blood group [32] and Y-chromosome frequency data [1,33,34] with ArcMap v.10.1 from the ArcGis suite (Environmental Systems Research Institute) using the default settings of the geospatial analysis plugin (Fig. 4).

Acknowledgements

This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. The authors wish to acknowledge the DJEI/DES/SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities and support. This study has been financed by the BEAN project of the Marie Curie ITN (grant number 289966). We thank the authors of Hellenthal et al.[26] for providing their genotype dataset, V. Mattiangeli for assistance with sequencing, as well as E. Jones and L. Cassidy for advice and help provided regarding data processing.

Accession codes

Raw Illumina sequencing reads can be downloaded at http://www.ebi.ac.uk/ena/data/view/PRJEB11004.

Contributions:

DGB and MJC supervised the study. RM, KHM, CMD, MJC, DGB designed research. RM processed aDNA samples and prepared NGS libraries, RM, RLM, MDT and DGB analysed genetic data, AC, MH, KHM, CMD, MJC provided samples and interpretation of the archaeological context. JM, GM produced and analysed isotopic data, RM, AC, GM, JM, MC, SR, MJC and DGB wrote the paper with help from all authors.

Competing financial interests

The authors declare no conflict of interest.

Corresponding authors

Correspondence to: Daniel G Bradley (daniel.bradley@tcd.ie) or Matthew J Collins (matthew.collins@york.ac.uk)

Figure legends

Figure 1 - a) Principal Component Analysis (PCA) and **b)** Model based clustering using NGSadmix (K=3) of Driffield Terrace, Iron Age (IA) and Anglo-Saxon (AS) samples merged with European, West Asian, Middle Eastern and North African populations [26].

Population key: Ad - Adygei; Ar - Armenian; Ba - Basque; Bed - Bedouin; Be - Belorussian; Bu - Bulgarian; Ch - Chuvash; Cy - Cypriot; Dr - Druze; Ea - East Sicilian; Eg - Egyptian; En - English; Fi - Finnish; Fr - French; Ge - Georgian; GA - Germany/Austria; Gr - Greek; Hu - Hungarian; Ira - Iranian; Ir - Ireland; Jo - Jordanian; Le - Lezgin; Li - Lithuanian; Mo - Moroccan; Moz - Mozabite; NIt - North Italian; Nor - Norwegian; Or - Orcadian; Pa - Palestinian; Po - Polish; Ro - Romanian; Ru - Russian; Sa - Sardinian; Sau - Saudi; Sc - Scottish; So - South Italian; Sp - Spanish; Sy - Syrian; Tun - Tunisian; Tk - Turkish; Tu - Tuscan; UAE - United Arab Emirates; We - Welsh; WS - West Sicilian; Ye - Yemeni.

Figure 2 - Combined percentile scores of modern European samples ranked by identity by state to the Roman York genotypes reveals strongest affinity to modern Welsh, followed by Irish and Scottish. One outlier, 3DRIF-26 was excluded from this analysis.

Figure 3 - a) Principal Component Analysis (PCA) of the Roman samples from Driffield Terrace (excluding one outlier), one Iron Age individual and one Anglo-Saxon merged with modern Irish, British and Dutch genotype data. **b)** Boxplot of PC1 broken down by subregion. The symbols on the left represent the significance of a Mann-Whitney test performed to compare the Roman population with all other populations in the dataset. There were no significant differences between the Roman sample and the present-day Welsh, Northern and Northwestern English samples included in this analysis; all other regions had significantly different median values for PC1.

Population-key: Du - Dutch; En - English; Ir - Irish; Sc - Scottish; Wa - Wales

N.S - Non-significant; N.S - $p>0.05$; * $0.05>p>0.01$; ** $0.01>p>0.0001$; *** $p<0.0001$.

Figure 4 - FineSTRUCTURE analysis of modern British genotypes and IBS affinity to the British Roman cohort. Panel a shows the inferred clusters of moderns, their regional origins, the order of emergence of these groups and numbers of individuals in each. Below, mean IBS between each cluster and the ancient Roman samples is plotted; the most prominent feature is their relative similarity to the predominantly Welsh clusters. Panel b shows plots of median cluster IBS values of the Romans Vs. the single Iron Age genome and, secondly Vs. the Anglo Saxon sample. An indication of continuity is a strong correlation with the former and later discontinuity is supported by the lack of correlation with the latter.

Figure 5 - Interpolated maps of allele frequency comparing Roman York samples and modern populations from the British Isles: a) PC1 median values; b) blood group O frequency; c) Y-chromosome haplogroup R1b1a2-M269 frequency.

Tables

Table 1 - Result summary for the samples analysed in the present study.

Sample	Period	Excavation Site	Total Reads	Mapped reads *	Duplication %	Endogenous % *	Mean Coverage (x)	chrY hg	mtDNA hg
3DRIF-16	Roman	Driffield Terrace	63341920	28683678	2.92	43.96	0.67	R1b1a2a1a1-M405	H6a1a
3DRIF-26	Roman	Driffield Terrace	207248970	50652260	8.2	22.44	1.13	J2-L228	H5
6DRIF-18	Roman	Driffield Terrace	98083358	41157853	3.78	40.38	1.07	R1b1a2a1a-L52/L11	H1bs
6DRIF-21	Roman	Driffield Terrace	91887701	48712821	5.19	50.26	1.16	R1b1a2a1a2c2-DF63	J1c3e2
6DRIF-22	Roman	Driffield Terrace	115324680	45995965	2.45	38.91	1.12	R1b1a2a1a2b-S28	H2+195
6DRIF-23	Roman	Driffield Terrace	117230764	25256982	2.85	20.93	0.65	R1b1a2a1a-L52	H6a1b2
6DRIF-3	Roman	Driffield Terrace	112316793	68421310	2.59	59.34	1.67	R1b1a2a1a1-M405	J1b1a1
M1489	Iron Age	Melton	81838435	21802991	2.00	26.64	0.56	-	U2e1e
NO3423	Anglo-Saxon	Norton on Tees	89918177	43369123	2.00	48.23	1.05	I1-S107	H1a

* Reads were filtered by base quality 15, mapping quality 30 and duplicates removed.

References

1. Weale, M. E., Weiss, D. A., Jager, R. F., Bradman, N. & Thomas, M. G. Y chromosome evidence for Anglo-Saxon mass migration. *Mol. Biol. Evol.* **19,** 1008–1021 (2002).
2. Thomas, M. G., Stumpf, M. P. H. & Härke, H. Evidence for an apartheid-like social structure in early Anglo-Saxon England. *Proc. Biol. Sci.* **273,** 2651–2657 (2006).
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447,** 661–678 (2007).
4. Leslie, S. *et al.* The fine-scale genetic structure of the British population. *Nature* **519,** 309–314 (2015).
5. Sims-Williams, P. Bronze- and Iron-Age Celtic-speakers: what don't we know, what can't we know, and what could we know? Language, genetics and archaeology in the twenty-first century. *J. R. Soc. Antiq. Irel.* **92,** 427–449 (2012).
6. Foubert, L. L. & Breeze, D. in *Past Mobilities. ArchaeologicalApproaches to Movement and Mobility* (ed. Leary, J.) 175– 86 (Ashgate, 2014).
7. Adams, C. & Laurence, R. *Travel and geography in the Roman Empire*. (Routledge, 2012).
8. Noy, D. Epigraphic evidence for immigrants at Rome and in Roman Britain. *Roman Diasporas: Archaeological Approaches to Mobility and Diversity in the Roman Empire* 13–26 (2010).
9. Eckardt, H. *Roman Diasporas: Archaeological Approaches to Mobility and Diversity in the Roman Empire*. (Journal of Roman Archaeology, 2010).
10. Leach, S., Lewis, M., Chenery, C., Müldner, G. & Eckardt, H. Migration and diversity in Roman Britain: a multidisciplinary approach to the identification of immigrants in Roman York, England. *Am. J. Phys. Anthropol.* **140,** 546–561 (2009).
11. Chenery, C., Müldner, G., Evans, J., Eckardt, H. & Lewis, M. Strontium and stable isotope evidence for diet and mobility in Roman Gloucester, UK. *J. Archaeol. Sci.* **37,** 150–163 (2010).
12. Müldner, G., Chenery, C. & Eckardt, H. The "Headless Romans": multi-isotope investigations of an unusual burial ground from Roman Britain. *J. Archaeol. Sci.* **38,** 280–290 (2011).
13. Eckardt, H., Müldner, G. & Lewis, M. People on the move in Roman Britain. *World Archaeol.* **46,** 534–550 (2014).

14. Ottaway, P. *Book of Roman York*. (BT Batsford Ltd, 1993).

15. Hunter-Mann, K. Romans Lose Their Heads: An Unusual Cemetery at The Mount. *The Archaeology of Roman York* (2007). at <http://www.iadb.co.uk/driffield6/index.php>

16. Ottaway, P. *1–3 Driffield Terrace, York. Assessment Report on an Archaeological Excavation, 2005*. (York Archaeological Trust for Excavation and Research, 2005).

17. Fenton-Thomas, C. *Where Sky and Yorkshire and Water Meet: The Story of the Melton Landscape from Prehistory to the Present: Archaeological Excavations on the A63 Grade Separated Junction at Melton, East Yorkshire*. (On-Site Archaeology, 2011).

18. Sherlock, S. J. & Welch, M. G. *An Anglo-Saxon Cemetery at Norton, Cleveland*. **82,** (Council for British Archaeology York, 1992).

19. Gamba, C. *et al.* Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* **5,** 5257 (2014).

20. Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A. von & Pääbo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29,** 4793–4799 (2001).

21. Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40,** 4477–4482 (2013).

22. Kivisild, T. Maternal ancestry and population history from whole mitochondrial genomes. *Investig. Genet.* **6,** 3 (2015).

23. Myres, N. M. *et al.* A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* **19,** 95–101 (2011).

24. Semino, O. *et al.* Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* **74,** 1023–1034 (2004).

25. Lappalainen, T. *et al.* Migration waves to the Baltic Sea region. *Ann. Hum. Genet.* **72,** 337–348 (2008).

26. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343,** 747–751 (2014).

27. Dansgaard, W. Stable isotopes in precipitation. *Tell'Us* **16,** 436–468 (1964).

28. Bentley, A. R. Strontium Isotopes from the Earth to the Archaeological Skeleton: A Review. *J Archaeol Method*

Theory **13,** 135–187 (2006).

29. McLaughlin, R. L. *et al.* A second-generation Irish genome-wide association study for amyotrophic lateral sclerosis. *Neurobiol. Aging* **36,** 1221.e7–1221.e13 (2015).
30. Van Es, M. A. *et al.* Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41,** 1083–1087 (2009).
31. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8,** e1002453 (2012).
32. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The history and geography of human genes*. (Princeton university press, 1994).
33. Moore, L. T., Mc Evoy, B., Cape, E., Simms, K. & Bradley, D. G. Report A Y-Chromosome Signature of Hegemony in Gaelic Ireland. *Am. J. Hum. Genet.* **78,** 334–338 (2006).
34. Capelli, C. *et al.* A Y chromosome census of the British Isles. *Curr. Biol.* **13,** 979–984 (2003).
35. Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7,** 98–115 (2013).
36. Mourant, A. E. & Others. The Distribution of the Human Blood Groups. *The Distribution of the Human Blood Groups.* (1954). at <http://www.cabdirect.org/abstracts/19552900220.html>
37. Töpf, A. L., Gilbert, M. T. P., Dumbacher, J. P. & Hoelzel, A. R. Tracing the phylogeography of human populations in Britain based on 4th-11th century mtDNA genotypes. *Mol. Biol. Evol.* **23,** 152–161 (2006).
38. Miles, D. *The tribes of Britain*. (Weidenfeld & Nicolson, 2005).
39. Mattingly, D. *An imperial possession: Britain in the Roman Empire, 54 BC-AD 409*. **1,** (Penguin, 2008).
40. Cunliffe, B. *Britain begins*. (Oxford University Press, 2013).
41. Schmidt, K. H. 3 Insular Celtic: P and Q Celtic. *The Celtic Languages* 64 (1993).
42. Caffell, A. & Holst, M. *Osteological Analysis, 3 and 6 Driffield Terrace, York, North Yorkshire*. (York Osteoarchaeology, 2012).
43. Kanz, F. & Grossschmidt, K. Head injuries of Roman gladiators. *Forensic Sci. Int.* **160,** 207–216 (2006).
44. Redfern, R. & Bonney, H. Headhunting and amphitheatre combat in Roman London, England: new evidence from the Walbrook Valley. *J. Archaeol. Sci.* **43,** 214–226 (2014).

45. Roth, J. P. *The Logistics of the Roman Army at War: 264 B.C. - A.D. 235*. (Brill, 1999).

46. Montgomery, J., Knüsel, C. J. & Tucker, K. in *The bioarchaeology of the human head : decapitation, decoration and deformation.* (ed. Bonogofsky, M.) 141–178 (University Press of Florida, 2011).

47. Montgomery, J., Knüsel, C. & Katie, T. Identifying the origins of decapitated male skeletons from 3 Driffield Terrace, York, through isotope analysis: reflections of the cosmopolitan nature of Roman York in the time of Caracalla. 141–178 (2011).

48. Müldner, G., Chenery, C. & Eckardt, H. The "Headless Romans": multi-isotope investigations of an unusual burial ground from Roman Britain. *J. Archaeol. Sci.* **38,** 280–290 (2011).

49. MacHugh, D. E., Edwards, C. J., Bailey, J. F., Bancroft, D. R. & Bradley, D. G. The Extraction and Analysis of Ancient DNA From Bone and Teeth: a Survey of Current Methodologies. *Anc. Biomol.* **3,** 81 (2000).

50. Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105,** 539–543 (1998).

51. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010,** db.prot5448 (2010).

52. Martiniano, R. *et al.* Genetic evidence of African slavery at the beginning of the trans-Atlantic slave trade. *Sci. Rep.* **4,** 5994 (2014).

53. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).

54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

55. Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13,** 178 (2012).

56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

57. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29,** 1682–1684 (2013).

58. Rasmussen, M. *et al.* An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* **334 ,** 94–98 (2011).

59. Korneliussen, T., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15,** 356 (2014).

60. Skoglund, P. *et al.* Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 2229–2234 (2014).

61. Ralf, A., van Oven, M., Zhong, K. & Kayser, M. Simultaneous analysis of hundreds of Y-chromosomal SNPs for high-resolution paternal lineage classification using targeted semiconductor sequencing. *Hum. Mutat.* **36,** 151–159 (2015).

62. Andrews, R. M. *et al.* Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23,** 147 (1999).

63. Vianello, D. *et al.* HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34,** 1189–1194 (2013).

64. Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30,** E386–94 (2009).

65. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

66. Skoglund, P. *et al.* Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* **336,** 466–469 (2012).

67. Cronin, S. *et al.* A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum. Mol. Genet.* **17,** 768–774 (2008).

68. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2,** 2074–2093 (2006).

69. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).

70. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19,** 1655–1664 (2009).

71. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195,** 693–702 (2013).

72. Rosenberg, N. A. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4,**

137–138 (2004).

73. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *arXiv [q-bio.GN]* (2014).
74. R Development Core Team, R. *R: A Language and Environment for Statistical Computing*. (2011).
75. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nat. Methods* **9,** 179–181 (2012).
76. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature* **135,** 0–9 (2012).
77. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81,** 1084–1097 (2007).
78. Heit, J. A. *et al.* A genome-wide association study of venous thromboembolism identifies risk variants in chromosomes 1q24.2 and 9q. *J. Thromb. Haemost.* **10,** 1521–1531 (2012).
79. Nakao, M. *et al.* ABO blood group alleles and the risk of pancreatic cancer in a Japanese population. *Cancer Sci.* **102,** 1076–1080 (2011).
80. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74,** 1111–1120 (2004).
81. Enattah, N. S. *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* **30,** 233–237 (2002).
82. Beleza, S. *et al.* Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genet.* **9,** e1003372 (2013).
83. Canfield, V. A. *et al.* Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3* **3,** 2059–2067 (2013).

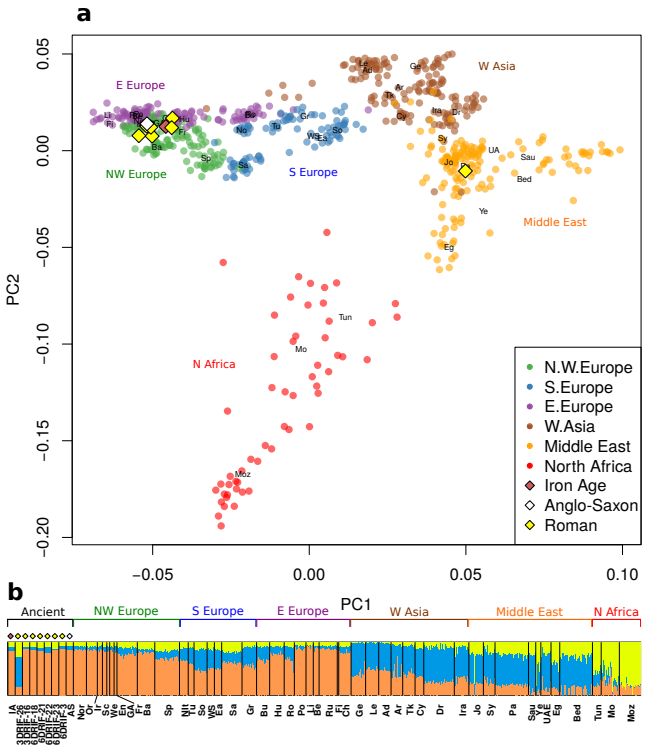

Figure 1 - a) Principal Component Analysis (PCA) and **b)** Model based clustering using NGSadmix (K=3) of Driffield Terrace, Iron Age (IA) and Anglo-Saxon (AS) samples merged with European, West Asian, Middle Eastern and North African populations.

Population key: Ad - Adygei; Ar - Armenian; Ba - Basque; Bed - Bedouin; Be - Belorussian; Bu - Bulgarian; Ch - Chuvash; Cy - Cypriot; Dr - Druze; Ea - East Sicilian; Eg - Egyptian; En - English; Fi - Finnish; Fr - French; Ge - Georgian; GA - Germany/Austria; Gr - Greek; Hu - Hungarian; Ira - Iranian; Ir - Ireland; Jo - Jordanian; Le - Lezgin; Li - Lithuanian; Mo - Moroccan; Moz - Mozabite; NIt - North Italian; Nor - Norwegian; Or - Orcadian; Pa - Palestinian; Po - Polish; Ro - Romanian; Ru - Russian; Sa - Sardinian; Sau - Saudi; Sc - Scottish; So - South Italian; Sp - Spanish; Sy - Syrian; Tun - Tunisian; Tk - Turkish; Tu - Tuscan; UAE - United Arab Emirates; We - Welsh; WS - West Sicilian; Ye - Yemeni.

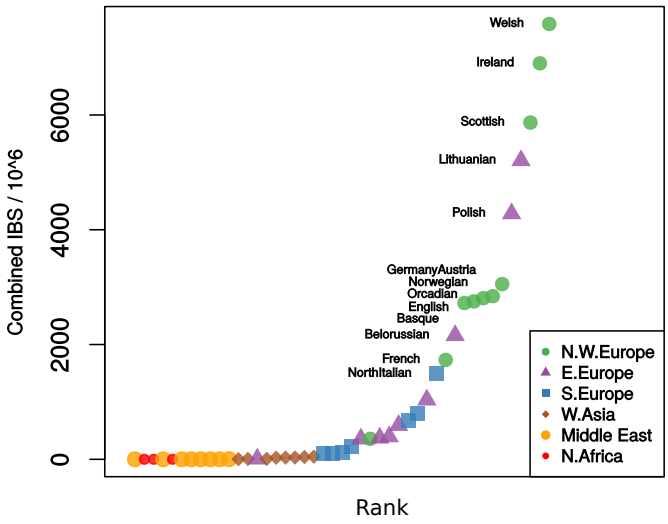

Figure 2 - Combined percentile scores of modern European samples ranked by identity by state to the Roman York genotypes reveals strongest affinity to modern Welsh, followed by Irish and Scottish. One outlier, 3DRIF-26 was excluded from this analysis.

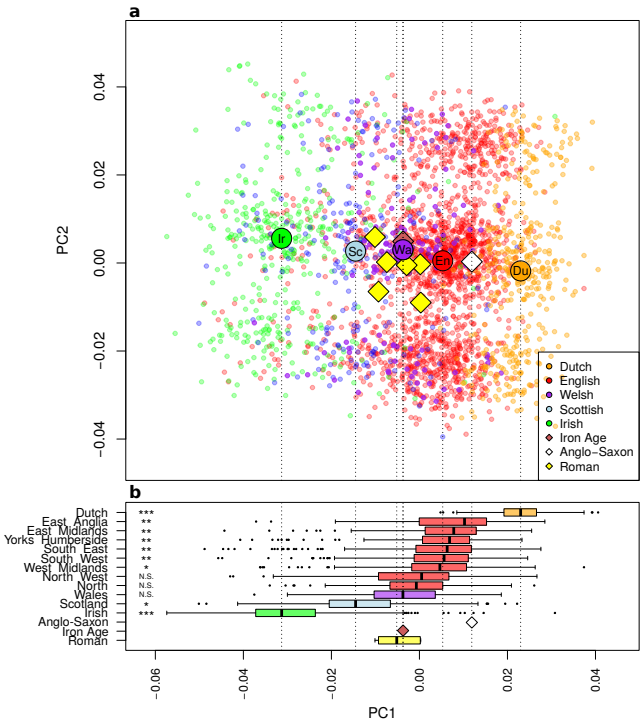

Figure 3 - a) Principal Component Analysis (PCA) of the Roman samples from Driffield Terrace (excluding one outlier), one Iron Age individual (IA) and one Anglo-Saxon (AS) merged with modern Irish, British and Dutch genotype data. **b)** Boxplot of PC1 broken down by subregion. The symbols on the left represent the significance of a Mann-Whitney test performed to compare the Roman population with all other populations in the dataset. There were no significant differences between the Roman sample and the present-day Welsh, Northern and Northwestern English samples included in this analysis; all other regions had significantly different median values for PC1.
Population-key: Du - Dutch; En - English; Ir - Irish; Sc - Scottish; Wa - Wales
N.S - Non-significant; N.S - p>0.05; * 0.05>p>0.01; ** 0.01>p>0.0001; *** p<0.0001.

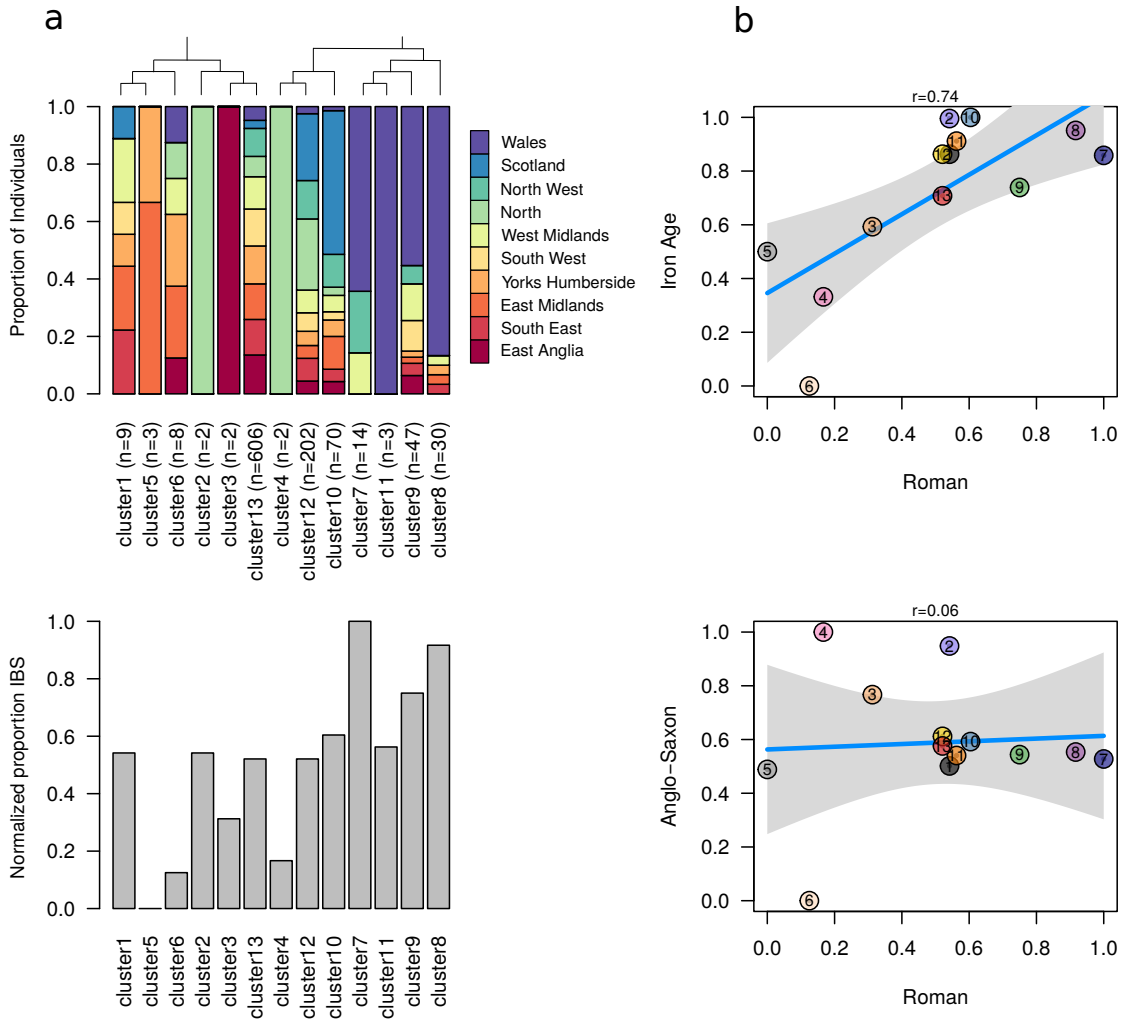

Figure 4 - FineSTRUCTURE analysis of modern British genotypes and IBS affinity to the British Roman cohort. Panel a shows the inferred clusters of moderns, their regional origins, the order of emergence of these groups and numbers of individuals in each. Below, mean IBS between each cluster and the ancient Roman samples is plotted; the most prominent feature is their relative similarity to the predominantly Welsh clusters. Panel b shows plots of median cluster IBS values of the Romans Vs. the single Iron Age genome and, secondly Vs. the Anglo Saxon sample. An indication of continuity is a strong correlation with the former and later discontinuity is supported by the lack of correlation with the latter.

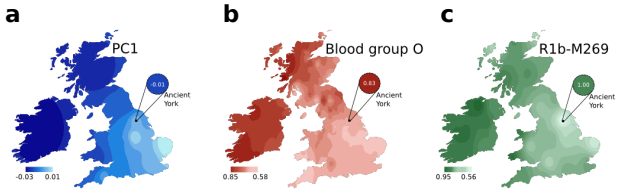

Figure 5 - Interpolated maps of allele frequency comparing Roman York samples and modern populations from the British Isles: a) PC1 median values; b) blood group O frequency; c) Y-chromosome haplogroup R1b1a2-M269 frequency.