Medieval social landscape through the genetic history of Cambridgeshire before and after the Black Death

Supplementary Materials

This PDF file includes:

Materials and Methods

Figures S1 – S13

Other Supplementary Materials for this manuscript include the following:

Tables S1 – S13

Materials and Methods

Sample information and ethical statement

All skeletal elements were sampled with permissions from the representative bodies/host institutions. Samples were taken and processed to maximize research value and minimize destructive sampling. Teeth were sampled from skeletons using gloves. Molars were preferred due to having more roots and larger mass, but premolars were also sampled.

Archaeological sites and material
Here we briefly describe the archaeological sites included in this study. More detailed information can be found at (Cessford et al., 2021).

Cherry Hinton

The settlement of Church End Cherry Hinton (Cherry Hinton) is located around six kilometers southeast of Cambridge. In the late 9th to the mid-10th century, a large *thengly* (aristocratic) or proto-manorial center was established (Cessford & Dickens, 2005b; Cessford & Slater, 2014). The associated timber chapel and graveyard were excavated in 1999 by the Hertfordshire Archaeological Trust (subsequently Archaeological Solutions and now Wardell Armstrong) (Lally, 2008).

Only part of the cemetery was investigated, including over 670 graves and the remains of *c.* 980 individuals. The cemetery population was estimated to be *c.* 1000–2000. The cemetery probably served an entire thriving Late Saxon to Norman rural settlement, broadly representative of a mixed rural peasant tenant hierarchy engaged primarily but not exclusively in agricultural labor. Based on radiocarbon dating plus a range of other evidence the cemetery dates to *c.* 940/990–1120/70. As the population from which those buried at Church End were drawn probably increased over time it is likely that the majority, possibly around two thirds, died after the Norman Conquest. As this settlement was located firmly within the rural hinterland of Cambridge this population provides an important comparator to the town, albeit one that is rather earlier than the main populations that have been studied from there.

All Saints

All Saints by the Castle (All Saints) is a medieval parish church and associated cemetery in Cambridge. Although not documented until 1217, the church and cemetery are likely to have been founded *c.* 940–1100 and the parish was amalgamated with St Giles in 1365/6 (HMC, 1877) (299, fol. 1026).

The cemetery of All Saints was first identified in 1972 in evaluation trenches by John Alexander, then partially investigated by Paul Craddock and Vince Gregory in 1973 (Craddock & Gregory, 1975), with further excavations in 1988 and 1994 (Cessford & Dickens, 2005a). In total just over 210 skeletons have been excavated from the cemetery, out of an estimated original overall total of *c.* 2500–3500. The burials at the cemetery of All Saints probably constitute a broadly representative sample of the parishioners' *c.* 940/1100–1365/6, albeit slightly distorted by several factors. Most burials, probably over 80%, are likely to date to the period after the Norman Conquest. Although the population of All Saints was socially and economically mixed, they are probably somewhat poorer than the average for Cambridge and rather more involved in agricultural activities. Radiocarbon dating is compatible with the burials dating to c. 940/1100–1365/6. The burials at All Saints

begin and end earlier than the other main Cambridge urban skeletal assemblages, from the Hospital of St John and the Augustinian Friary.

Hospital of St John

The Hospital of St John the Evangelist in Cambridge (Hospital or Hospital of St John) was established by local townspeople *c*. 1190–1200 to care for, principally in a social and spiritual sense rather than medically, the poor, infirm and sick. It did this until it was dissolved to create St John's College in 1511, although it was allegedly already in decline during the 15th century. Excavations in 2010–2011, conducted by the Cambridge Archaeological Unit in the detached cemetery of the hospital revealed almost 400 complete or partial skeletons, from a likely original population of *c*. 1000–1500 CE burials (Cessford, 2015).

The results of the intensive radiocarbon dating program and other evidence are compatible with burial starting in the early 13th century and indicate that it continued until at least the mid/late 15th century.

The burials at the detached cemetery of the Hospital of St John represent an amalgam of various groups associated with the institution. Most of them were probably charitable inmates of the Hospital.

Augustinian Friary

The Augustinian Friary (Friary) was established in Cambridge between 1279/80 and 1289, when it was first mentioned in a royal pittance. The Friary grew rapidly and thrived, becoming a *studium generale* or national study house with internal connections in 1318 and having 70 friars present in 1328. It continued until the Dissolution 1538, as one of the most important Augustinian friaries in England and one of the largest institutions in Cambridge.

Human remains have been recovered from three locations at the Friary: an early cemetery and the later chapter house and cloister. Burials included in this study were excavated by the Cambridge Archaeological Unit in 2016–2017 (Cessford, 2017, 2020; Cessford & Neil, 2022). Some of the burials were accompanied by single buckles located near the pelvis, indicating that the bodies were buried in a clothed state, with surviving evidence for associated leather girdles and some evidence for textiles. Some of the skeletons definitely lacked buckles and these were probably buried in shrouds. It appears that members of the Augustinian order received clothed burial, while shrouded burials are of lay individuals (Cessford & Neil, 2022). The lay individuals would include patrons and benefactors, as well as lay servants of the Friary and corrodians. The 72 burials recovered from the Friary represent only a portion of the estimated 200–700 individuals likely to have been interred there between 1290 and 1538.

Bene't Street

St Bene't's (a contraction of Benedict's) parochial church in Cambridge was established in *c*. 1000–1050 CE and remains in use, with burial continuing until the 1850s. A strip of land along the western side of the churchyard was transferred to Corpus Christi College between 1352 and 1377, to form an entrance route between Bene't Street and the College. Part of this strip was excavated by the Cambridge Archaeological Unit in 2006, revealing highly truncated individual burials and part of one mass burial (Cessford & Fallon, 2006). Among four skeletons tested from the mass burial, two yielded positive and one tentative identifications of *Y. pestis* (Cessford et al., 2021). It is probable that this mass burial relates to the Black Death as it is likely to predate the construction of the College buildings starting in 1352.

Baptist Chapel

The Providence Calvinistic Baptist Chapel (Baptist Chapel) in Cambridge was in use for just four years 1833–1837 CE. Part of the associated cemetery was excavated in 2012, by Oxford Archaeology East (Rees, 2014). There originally may have been a maximum of 20 graves in the investigated area, with at least 16 graves in five rows 11 of which were recorded. The graves were aligned north-northwest to south-southeast and were mainly earth cut, but two were brick-lined and there is evidence for coffins and shrouds.

Holy Trinity

Holy Trinity Church in Cambridge (Holy Trinity) is first mentioned in 1174, when it is said to have been burnt down, and was probably established *c*. 1050–1150 CE. Excavations by the Cambridge Archaeological Unit in 2016–17 revealed seventeen articulated burials (Newman, 2018). Although these were reburied, they were subject to thorough osteological analysis and sampling. The skeletons included in this study were placed in coffins in the vaults constructed after a vestry building of 1833/4 and prior to the end of burial in the cemetery in 1855.

Midsummer Common

From 1574 until the last outbreak in Cambridge in 1665/6, some individuals infected with the plague were isolated and moved to pest houses, located some distance from the town in its surrounding fields. Although primarily about quarantine and isolation of the living, parish registers record that from at least 1603 onwards some individuals who died at pest houses were buried there rather than returned to their parish cemetery. The location of the pest houses changed over time, with at least four locations known. One of these sites was Midsummer Common, where pest houses are mentioned in 1593 and 1630. Two skulls discovered in the late 19th or early 20th century at Midsummer Common were presented to the Duckworth Collection by the town clerk, John Edleston Ledsam Whitehead (1853–1923, town clerk 1887–1923). A tooth from one of the skulls was dated to 1450–1630 cal ad (Cessford et al., 2021).

Clopton

Clopton was a village in west Cambridgeshire, about 19 kilometers southwest of Cambridge. The village was established by the 10th–11th centuries and appears to have thrived until forcible enclosure for sheep grazing in *c*. 1480–1520 CE. The church is documented as having gone out of use in 1561. There is good evidence that the village fell primarily within the hinterland of Cambridge, although its location meant that it also had links to urban centers in Hertfordshire. The church of St Mary had been established at Clopton by the late twelfth century, but is probably significantly older. A new church was dedicated in 1352 and this continued to stand after the village was deserted.

John Alexander excavated *c*. 70 skeletons from the church and cemetery between 1960 and 1964, out of over 120 burials with records (Alexander, 1968). The excavations focussed primarily on the cemetery to the south of the church, which would have been where the bulk of the parishioners were buried. The limited excavations in and around the church means that the individuals who were buried there, which would have included clerics such as John Thorney as well as wealthy parishioners, are largely absent from the studied human remains. Radiocarbon dating and other evidence is consistent with a date of *c*. 1200–1561 CE for the burials.

Hemingford Grey

Excavations at Meadow Lane, north of Hemingford Grey, 19km northwest of Cambridge by Oxford Archaeology in 2006 revealed sixteen burials in a late 17th–early 18th-century nonconformist cemetery that has been linked to the Society of Friends or Quakers (Sims, 2007). Textual evidence for the cemetery comes from the vicar at Hemingford Grey Parish noting in the Parish register burials that took place away from the Parish cemetery at a site referred to as Wobourn or Oubourne. This lists 17 burials between 1681 and 1721, although it is unclear if the list is complete and earlier and later burials are possible. The individuals buried came from Hemingford Grey and the adjacent Parishes of Fenstanton, St Ives and Hemingford Abbot and there are several examples of multiple members of a single family.

Dating

The radiocarbon determinations were undertaken at the Scottish Universities Environmental Research Centre (SUERC) radiocarbon laboratory and followed their standard procedures (Dunbar et al., 2016). Analysis was undertaken using OxCal v.4.3 (Bronk Ramsey, 2009; Bronk Ramsey & Lee, 2013) using the IntCal13 calibration curve (Reimer et al., 2013). A small number of additional radiocarbon determinations (marked by * in Table S1) were undertaken at a late stage, to date specific individuals that produced evidence for specific pathogen aDNA. These were undertaken at the 14Chrono Centre, Queen's University Belfast. These additional determinations were calibrated using IntCal 2020 (Reimer et al., 2020).

At the Hospital of St John, Augustinian Friary and All Saints by the Castle, detailed Bayesian modeling of these results combined with expert archaeological judgment has helped to inform whether these individuals (along with others in the stratigraphic sequence) died before or after the Black Death pandemic of 1348–9. More details can be found in (Cessford et al., forthcoming).

Sampling, ancient DNA extraction and library preparation

Inside a class IIB hood in the dedicated aDNA facility of the University of Cambridge Department of Archaeology or University of Tartu Institute of Genomics, root portions of teeth were removed with a sterile drill wheel. Petrous was sampled with a 10mm core drill sterilized with bleach followed by distilled water and then ethanol rinse. Root and petrous portions were briefly brushed to remove surface dirt, any varnish or lacquer, and microbial film with full strength household bleach (6% w/v NaOCl) using a disposable toothbrush that was soaked in 6% (w/v) bleach prior to use. They were then soaked in 6% (w/v) bleach for 5 minutes. Samples were rinsed twice with 18.2 MΩcm H2O and soaked in 70% (v/v) Ethanol for 2 minutes, transferred to a clean paper towel on a rack inside a class IIB hood with the UV light on and allowed to dry. They were weighed and transferred to PCR-clean 5 ml or 15 ml conical tubes (Eppendorf) for chemical extraction.

Inside a class IIB hood, per 100 mg of each sample, 2 ml of 0.5M EDTA Buffer pH 8.0 (Fluka) and 50 μl of Proteinase K 10 mg/ml (Roche) was added. Tubes were rocked in an incubator for 72 hours at room temperature. Extracts were concentrated to 250 μl using Amplicon Ultra-15 concentrators with a 30 kDa filter (Millipore). Samples were purified according to manufacturer's instructions using buffers from the Minelute™ PCR Purification Kit (Qiagen) with the following changes: 1) the use of High-Volume spin columns (Roche); 2) 10X PB buffer instead of 5X; and 3) samples incubated with EB buffer (Qiagen) at 37C for 10 minutes prior to elution. The columns were transferred to clean, labeled, 1.5ml Eppendorf tubes. One hundred microlitres EB buffer is added to the membrane and centrifuged at 13,000

rpm for two minutes after the 10-minute incubation and stored at -20 C. Only one extraction was performed per sample for screening and 30µl used for libraries.

Library preparation was conducted using a protocol modified from the manufacturer's instructions included in the NEBNext® Library Preparation Kit for 454 (E6070S, New England Biolabs, Ipswich, MA) as detailed in (Meyer & Kircher, 2010). DNA was not fragmented and reactions were scaled to half volume, adaptors were made as described in (Meyer & Kircher, 2010) and used in a final concentration of 2.5uM each. DNA was purified on MinElute columns (Qiagen). Libraries were amplified using the following PCR set up: 50µl DNA library, 1X PCR buffer, 2.5mM MgCl2, 1 mg/ml BSA, 0.2µM inPE1.0, 0.2mM dNTP each, 0.1U/µl HGS Taq Diamond and 0.2µM indexing primer. Cycling conditions were: 5' at 94C, followed by 18 cycles of 30 seconds each at 94C, 60C, and 68C, with a final extension of 7 minutes at 72C. Amplified products were purified using MinElute columns and eluted in 35 µl EB (Qiagen). Three verification steps were implemented to make sure library preparation was successful and to measure the concentration of dsDNA/sequencing libraries – fluorometric quantitation (Qubit, Thermo Fisher Scientific), parallel capillary electrophoresis (Fragment Analyser, Advanced Analytical) and qPCR.

DNA sequencing

DNA was sequenced using the Illumina NextSeq500/550 High-Output single-end 75 cycle kit at the University of Cambridge Department of Biochemistry DNA Sequencing Facility. As a norm, 20 samples were sequenced together on one flow cell; additional data was generated for samples over 5% endogenous human content to increase coverage.

Mapping

Before mapping, the sequences of the adapters, indexes, and poly-G tails were removed from read ends and reads shorter than 30 bp were removed using cutadapt-1.11 (Martin, 2011).

The sequences were aligned to the reference sequence GRCh37 (hg19) using Burrows-Wheeler Aligner (BWA 0.7.12) (Li & Durbin, 2009) and the command *aln* with re-seeding disabled.

After alignment, the sequences were converted to BAM format and only sequences that mapped to the human genome were kept with samtools 1.3 (Li et al., 2009). Data from different flow cell lanes were merged and duplicates were removed using picard 2.12 (http://broadinstitute.github.io/picard/index.html).

Sequencing coverage was estimated using Qualimap 2.2.1 (Okonechnikov et al., 2015) after applying a custom accessibility mask that includes the 1000 genomes accessibility mask and a composite mappability track (Derrien et al., 2012).

aDNA authentication

As a result of degradation over time, aDNA can be distinguished from modern DNA by certain characteristics: short fragments and a high frequency of C > T substitutions at the 5' ends of sequences due to cytosine deamination. The program mapDamage2.0 (Jónsson et al., 2013) was used to estimate the frequency of 5' C > T transitions. Rates of contamination were estimated from mtDNA using ContamMix (Fu et al., 2013) and from X chromosome using two methods implemented in ANGSD (Korneliussen et al., 2014).

Samtools 1.3 (Li et al., 2009) option *stats* was used to determine the number of final reads, average read length, average coverage etc. The average endogenous DNA content (proportion of reads mapping to the human genome) was 13.22% (0.00% - 81.27%).

Calculating genetic sex estimation

Genetic sex was calculated using the methods and script described in (Skoglund et al., 2013), estimating the fraction of reads mapping to Y chromosome out of all reads mapping to either X or Y chromosome. Genetic sex was calculated for samples with a coverage >0.01x and only reads with a mapping quality >30 were counted for the autosomal, X, and Y chromosome.

Determining mtDNA haplogroups

Mitochondrial DNA haplogroups were determined using Haplogrep2 (Weissensteiner et al., 2016). Subsequently, the identical results between the individuals were checked visually by aligning mapped reads to the reference sequence using samtools-1.3 (Li et al., 2009) command *tview* and confirming the haplogroup assignment in PhyloTree (accessed at: www.phylotree.org). Additionally, private mutations were noted for further kinship analysis.

Y chromosome variant calling and haplotyping

A total of 256,463 binary Y chromosome SNPs that have been detected as polymorphic in previous high coverage whole Y chromosome sequencing studies (Hallast et al., 2015; Karmin et al., 2015; Poznik et al., 2016) and by YFull (https://www.yfull.com/snp-list/) were called in male samples with more than $0.001\times$ autosomal coverage using ANGSD-0.916 (Korneliussen et al., 2014) '-doHaploCall' option. Basal haplogroup affiliations (Table S1) could be determined for 128 samples by assessing the proportion of derived allele calls (pD) in a set of primary (A, B, C...T) haplogroup defining internal branches, as defined in (Karmin et al., 2015), using 4081 informative sites. Haplogroup assignments were confirmed and specified using pathPhynder (Martiniano et al., 2022). Further detailed sub-haplogroup assignments within the phylogeny of the primary haplogroup were determined on the basis of mapping the derived allele calls to the internal branches of the YFull tee (https://www.yfull.com/tree/), requiring support of at least two variants for the terminal branch assignment.

Pseudohaploid genotype calling

Autosomal variants were called with the ANGSD 0.917 (Korneliussen et al., 2014) command `--doHaploCall`, which randomly selects one base at each specified position in the genome. The pseudohaploid genotypes were used in genetic kinship analysis with READ and for PC analyses of individual sites with smartPCA (Figures S1-9).

Imputation

Following (Hui et al., 2020), genotype likelihoods were first updated with Beagle 4.1 (Browning and Browning 2016) from genotype likelihoods produced by ANGSD 0.917 (`-doMajorMinor 3 -GL 1 -doPost 1 -doVcf 1 -sites …`) (Korneliussen et al., 2014) in Beagle -gl mode, followed by imputation in Beagle -gt mode with Beagle 5 (Browning et al., 2018) from sites where the genotype probability (GP) of the most likely genotype reaches 0.99. To balance between imputation time and accuracy, we used 503 Europeans genomes in 1000 Genomes Project Phase 3 (The 1000 Genomes Project

Consortium, 2015) as the reference panel in Beagle -gl step, and 27,165 genomes (except for chromosome 1, where the sample size is reduced to 22,691 due to a processing issue in the release) from the Haplotype Reference Consortium (HRC) (McCarthy et al., 2016) in the Beagle -gt step. After filtering out rare variants with MAC <5, we retained a total of ~37 million variants genome-wide. Because Beagle treats "./." in the VCF input as sporadically missing and imputes them during haplotype phasing, which damages the accuracy when such missing genotypes are common, we imputed each genome individually so that missing genotypes were not included in the VCF input to Beagle 5. The output single-sample VCFs were then merged for downstream analysis. Apart from newly generated genomes in this study, we also imputed published genomes from (Margaryan et al., 2020; Martiniano et al., 2016; Schiffels et al., 2016).

We down-sampled a 10.83x genome that has been previously reported in (Guellil et al., 2022), PSN31, to 0.05x-1x to evaluate our imputation pipeline (Table S13). The accuracy was estimated by comparing the imputed genotypes to genotypes called using GATK HaplotypeCaller (McKenna et al., 2010) (with the flags `--min-pruning 1 --min-dangling-branch-length 1`). For some analysis, we filtered the imputed genotypes further by GP and/or minor allele frequencies at the sites as described in the respective sections.

Principal component analysis

We used FlashPCA2 (Abraham et al., 2017) for principal component analysis (PCA) of imputed genomes (without projection) together with modern reference genomes from selected groups in UK Biobank after excluding variants in linkage disequilibrium with the PLINK --indep-pairwise 1000 50 0.5 option and exclusion of the likely non-neutral regions exclusion_regions_hg19.txt (Figure 1C and Figure S1). For comparison we also performed PCA using modern reference genomes only, before projecting pseudo-haploid genomes (coverage > 0.05X) onto the PC space following (https://github.com/chrchang/eigensoft/blob/master/POPGEN/lsqproject.pdf) (Figures S2-11). The error bars of the projected genomes represent one standard deviation obtained from 20-fold block jackknife.

Detecting IBD and long shared allele interval (LSAI) segments

IBD/LSAI segments (Kivisild et al., 2021) and kinship coefficients were estimated from merged plink files of 61 imputed ancient genomes, 503 Europeans from the 1000 Genome Project and UK Biobank data with IBIS version 1.20.9 (Seidman et al., 2020) using minimum shared segment length (-min_L) thresholds 5 cM and/or 7 cM together with -maxDist 0.1 and -mt 300 parameters. In total, 269,319 binary SNPs with MAF > 0.05 were used.

Although IBIS has the highest IBD inference accuracy for > 7 cM segments, which we use in kinship analyses, we use > 5 cM threshold in our diachronic inferences of population affinities because our focus is on relationships at generational distances > 15 at which longer IBD sharing expectations become relatively low, particularly in combination with the loss of sensitivity to detect long IBD segments from imputed ancient DNA sequences. Because true IBD segments of this length are not expected to be common at these generational distances, we need to consider the detected segments as "long shared allele intervals" (LSAIs) rather than IBD segments sensu stricto. Because they are inferred from unphased data after removal of rare variants (which cannot be imputed with sufficient accuracy), the LSAIs are likely to include undetected recombination points and smaller IBD segments residing on different haplotypes.

The probability of individual connectedness (PiC) score (Kivisild et al., 2021) for individual x in group Z was estimated as the proportion of individuals from group Z with whom individual x shared IBD above the given threshold. In practice, we estimated the count of connected individuals from group Z from sorted IBIS .coef output files by using the linux "join" function to add group codes to individual identifiers and by using the "crosstab" function of datamash[59] to generate the table of counts, each of which we divided by the total number of individuals in group Z to obtain the individual connectedness proportions by groups (the PiC scores).

Unsupervised community extraction analyses on IBIS .coef files calculated from merged plink files including imputed ancient genomes with coverage > 0.2x and modern UK Biobank data were run with the Louvain algorithm (Blondel et al., 2008) implemented in the R library "igraph" (Csardi & Nepusz, 2006) with additional significance tests as described in (Kivisild et al., 2021) using scripts available at https://github.com/SABiagini/Louvain.

Genetic kinship analysis

Pseudohaploid genotypes at 5,494,912 positions with MAF > 0.05 in the UK10K subset of the HRC panel were called with ANGSD version 0.917. For the comparison with published studies, the merged PLINK file from the PCA analysis was used and select populations retained using plink --keep and converted to .tped. The ANGSD output files were converted to .tped format, which was used as an input for kinship analyses with READ (Kuhn et al., 2018). Given highly homogenous ancestry in all study sites, samples from all sites were analyzed together for Figure 2 as well as separately (results not shown) to test for potential biases in normalized P0 calculation. In addition to 1st and 2nd degree relationships we also estimated P0 cutoffs (15/16 = 0.9375 as per (Kuhn et al., 2018)) for the detection of 3rd degree relatives while acknowledging that due the lack of relevant empirical data the false positive and negative error rates for this class of relationship remain unknown.

Runs of homozygosity

We used hapROH (Ringbauer et al., 2021) to detect runs of homozygosity (ROH) in ancient genomes. Using information from a reference panel, hapROH has been shown to work for genomes with more than 400K of the 1240K SNPs panel covered at an error rate lower than 3% in pseudo-haploid genotypes (Ringbauer et al., 2021). We note that the requirement is broadly in line with the imputation accuracy we get from coverages as low as 0.1x, where ~60% of common variants (MAF≥0.05) in the HRC panel are recovered with an overall accuracy around 97% for diploid genotypes (Table S13). Among common variants in the HRC panel, 853,159 overlap with the 1240K SNPs panel.

1000 Genomes Project data were used to construct the reference haplotypes. We kept the standard parameters in hapROH, which had been optimized for 1240K aDNA genotype data:

```
e_model='haploid', post_model='Standard', random_allele=True,
roh_in=1, roh_out=20, roh_jump=300, e_rate=0.01, e_rate_ref=0.0,
cutoff_post=0.999, max_gap=0, roh_min_l=0.01
```

Heterozygosity

Genome-wide heterozygosity and heterozygosity in the HLA region (--chr 6 --from-kb 28477 --to-kb 33448) were estimated with the --het function in PLINK 1.9 (Chang et al., 2015) for imputed genomes with >0.1x coverage and for 5,448,740 sites that had MAF > 0.05 in the

HRC reference panel. Sites where the highest genotype probability is less than 0.99 were set to missing.

Nucleotide diversity

Nucleotide diversity in the HLA region and throughout the autosomes was estimated using vcftools --window-pi (Danecek et al., 2011), which by default outputs result in 10kb windows, in imputed genomes with > 0.05x coverage. Sites where the highest genotype probability is less than 0.99 were set to missing. The results are similar whether we filtered the variants by MAF > 0.05 in the HRC reference panel or not (Figure S13).

Stature estimation

Long bone measurements (clavicle, humerus, radius, ulna, femur, tibia and fibula) were taken as outlined in Buikstra and Ubelaker (Buikstra & Ubelaker, 1994). Measurements were taken using an osteometric board and sliding callipers. Circumferences were taken using a measuring tape. All measurements were taken to 1pd. When left and right bones were measurable, a mean of the two measurements was calculated. Measurements were not taken for the purpose of stature reconstruction for individuals with gross pathologies (e.g. vitamin D deficiency), or from bones with mal-aligned trauma which altered the length of the bone. Only individuals who were developmentally complete (long bone epiphyses fused) were measured for stature estimation. Published measurements were taken for Clopton and the material excavated from the Friary in the 1900s as it was no longer possible to access the post crania of these remains. The appendices in (Inskip, forthcoming) outline all measurements taken.

Stature was calculated from long bone measurements following Trotter (1970) as included in (White & Folkens, 2005). These formulae were selected as many of the comparative collections analyzed in the mid 20th century, for which post crania are no longer available, used these equations. Due to the level of truncation at many of the sites assessed by the ATP project, it was decided to use single bone estimates rather than those that required two bones (e.g. femur and tibia) to maximize sample size. Stature estimates were not derived from the fibulae due to the low number of complete bones. Stature estimated from various long bones can also be found in the appendices of (Inskip, forthcoming).

Polygenic risk score for height

Eight polygenic risk score (PRS) models for human height trained on UK Biobank data were applied on the imputed genomes: PRS, LDpred, MCP, Lasso, PRS+MTAG, LDpred+MTAG, MCP+CTPR and Lasso+CTPR (Chung et al., 2019). Some of the models (MTAG and CTPR) also incorporated information about BMI to aid modeling height. The authors have shown that these models can reach prediction R^2 between 0.22 and 0.43 in modern validation datasets. Coefficients of the models were downloaded from http://lianglab.rc.fas.harvard.edu/CTPR/.

Each model was applied on two versions of the imputed genotypes: dosage data without further filtering, and hardcall genotypes after filtering by GP (max(GP) ≥ 0.99). We calculated the polygenic risk scores from each linear model in PLINK 1.9 (using --score flag) (Chang et al., 2015).

Table S9 shows the correlation coefficients (R^2) between the PRS calculated using different models and stature estimated from long bones in our samples from the later medieval period. We note that when low coverage genomes (< 0.1x) are included, where the imputed

genotypes carry more uncertainty, PRS calculated from the dosage data show higher correlation to stature estimates. The correlation also increases when we raise the coverage threshold, and when we only include stature estimated from the femur. We chose to include 88 genomes above 0.05x with stature estimated from the femurs as a trade-off between the sample size and the predictive power of the PRS to fit the regression models.

Regression models for stature

We fitted a range of Bayesian linear regression models using stature estimation as the outcome variable. The models take the general form

$$\mu = \beta \cdot \mathbf{X} + \alpha$$

$$\mathbf{y} \sim N(\mu, \sigma^2)$$

Where $\mathbf{y}$ denotes femur based stature estimation; $\mathbf{X}$ denotes the matrix of numeric predictor variables, if any; β is the vector of coefficients; α is the vector of intercept. There are multiple α terms when categorical predictor variables (e.g. sex or social group) are included in the model. Both $\mathbf{X}$ and $\mathbf{y}$ were scaled to have mean 0 and standard deviation 1. The prior distribution of the parameters were set to

$$\alpha \sim N(0, 1)$$

$$\beta \sim N(0, 0.5)$$

$$\sigma \sim \exp(1)$$

The "rstan" package (Stan Development Team, 2023) in R (R Core Team, 2022) was used to fit the regression models, and the "loo" package (Vehtari et al., 2022) was used for model comparison. The predictors we included in each model are: 1. Sex only; 2. Sex and PRS; 3. Sex and social group; 4. Sex, social group, and PRS. Since we would like to compare the fitted models, we only included 76 later medieval individuals supported by sufficient contextual information to belong to one of the four social groups: charity cases (a subset of burials from St John's Hospital); friars (a subset from Augustinian Friary); ordinary urban (burials from the parish cemetery of All Saints); ordinary rural (burials from the church at Cherry Hinton). Model 2 was separately fitted to all 88 genomes sequenced to > 0.05x and have stature estimates from the femur to produce Figure 4A.

The posterior distributions of each model we fitted were summarized in Table S10. All models were run with 4 chains of 50,000 iterations (including 25,000 burn-in iterations) each. The R-hat statistic was monitored to check the chains have reached convergence; we also checked that the effective sample size of all parameters is at least around 20,000. When the four models were compared based on the theoretical expected log pointwise predictive density, the models including PRS perform better than those without PRS (Table S10). The model without social group in the predictors (Model 2), appears to outperform the one including social group (Model 4) slightly, but the difference is not conclusive given the standard error of the difference.

To investigate the difference in stature between social groups when controlling for sex and PRS, we extracted the posterior samples from Model 4 and looked at the distribution of the difference between the intercepts associated with social groups (Figure 4B).

For sensitivity analysis, we changed to more diffused prior distributions for all parameters (N(0, 2) for α; N(0, 1) for β; exp(0.5) for σ) and reran the four models. This led to more dispersion in the posterior distributions of α (the intercepts associated with indicator variables for sex and social groups), but our conclusions about model comparison and social group differences remained robust.

Phenotype prediction

Using the imputed genomes generated in this study (143 individuals with a coverage higher than 0.05x) and previously published (71 samples) (Margaryan et al., 2020; Martiniano et al., 2016; Schiffels et al., 2016), we extracted genotype calls for 39 out of the 41 HIrisPlex-S variants and 74 SNPs involved in diet and diseases, coding the allele information as the number of the effective allele (0, 1, 2) using plink (Tables S4-6). The diet and disease set of 74 SNPs was selected starting from lists of variants previously analyzed in ancient DNA studies (Allentoft et al., 2015; Günther et al., 2018; Olalde et al., 2014), prioritizing those with a role in the response to pathogenic infection. In addition to the 70 variants that we have previously used for phenotype prediction (Saag et al., 2021; Saupe et al., 2021), we also analyzed the 4 SNPs in immune loci recently detected to be highly differentiated before or after the Second Pandemic (Klunk et al., 2022) (Table S6). A table with the HIrisPlex-S SNP alleles per sample was uploaded on the HIrisPlex-S webtool (https://hirisplex.erasmusmc.nl/) to obtain probabilities values for each eye, hair and skin color category per sample. This output was then interpreted following the manual to obtain the final pigmentation prediction for each individual (Table S5). We then grouped the individuals by time period. The groups were compared performing an ANOVA statistical test, applying a Bonferroni's correction on each group of variants (carbohydrate metabolism, lipid metabolism, vitamin metabolism, response to pathogens, autoimmune diseases, other diseases, Pigmentation) (Table S4) to set the significance threshold. For only the significant SNPs, we also performed a *post-hoc* Tukey test to identify the significantly different group pairs. For our medieval samples, we also grouped them by relative time to the Second Pandemic (before or after) and performed a statistical t-test as reported in Table S4.

To test the enrichment for higher allele frequency differences at variant positions of immune genes we used Weir and Cockerham (Weir & Cockerham, 1984) Fst as implemented in PLINK (--fst case-control) on lists of variants derived from the Klunk et al. (Klunk et al., 2022) study, as well as an expanded list of 54,931 variants polymorphic in the HRC panel from the 37,574 putatively neutral regions defined by Gronau et al. 2011 that had MAF > 0.1 in 83 imputed genomes from Cambridge dating to before and after the Black Death. As an expanded set of immunity regions, we extracted 19,940 variants with MAF> 0.1 in the 83 Cambridge cohort from 189,173 exonic regions of the 4,723 innate immunity genes curated by InnateDB, https://www.innatedb.com/moleculeSearch.do.

References

Abraham, G., Qiu, Y., & Inouye, M. (2017). FlashPCA2: Principal component analysis of Biobank-scale genotype datasets. *Bioinformatics*, *33*(17), 2776–2778. https://doi.org/10.1093/bioinformatics/btx299

Alexander, J. A. (1968). Clopton: The life-cycle of a Cambridgeshire village. In L. M. Munby (Ed.), *East Anglian Studies* (pp. 48–70). Heffer and Sons.

Allentoft, M. E., Sikora, M., Sjögren, K. G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A. S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. D., Dąbrowski, P., … Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. *Nature*, *522*(7555), 167–172. https://doi.org/10.1038/nature14507

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bronk Ramsey, C. (2009). Bayesian Analysis of Radiocarbon Dates. *Radiocarbon*, *51*(1), 337–360. https://doi.org/10.1017/S0033822200033865

Bronk Ramsey, C., & Lee, S. (2013). Recent and Planned Developments of the Program OxCal. *Radiocarbon*, *55*(2), 720–730. https://doi.org/10.1017/S0033822200057878

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

Buikstra, J. E., & Ubelaker, D. H. (Eds.). (1994). *Standards for data collection from human skeletal remains: Proceedings of a seminar at the Field Museum of Natural History*. Arkansas Archeological Survey.

Cessford, C. (2015). The St. John's Hospital Cemetery and Environs, Cambridge: Contextualizing the Medieval Urban Dead. *Archaeological Journal*, *172*(1), 52–120. https://doi.org/10.1080/00665983.2014.984960

Cessford, C. (2017). *Former Old Examination Hall, North Range Buildings, New Museums Site, Cambridge: An archaeological excavation* (p. 1 pdf file) [PDF]. Cambridge Archaeology Unit. https://doi.org/10.5284/1088728

Cessford, C. (2020). *North Range Buildings, New Museums Site, Cambridge: Further archaeological investigations*. Cambridge Archaeology Unit.

Cessford, C., Alexander, C., Craddock, P., Everson, P., Gregory, G., Newman, R., Samuel, M., & Stocker, D. (forthcoming). Chapter 3. The sites. In J. E. Robb (Ed.), *Health and History in medieval Cambridge*.

Cessford, C., & Dickens, A. (2005a). Castle Hill, Cambridge: Excavation of Saxon, medieval and post-medieval deposits, Saxon execution site and a medieval coinhoard. *Proceedings of the Cambridge Antiquarian Society*, *94*, 73102. https://doi.org/10.5284/1073334

Cessford, C., & Dickens, A. (2005b). The Manor of Hintona: The origins and development of Church End, Cherry Hinton. *Proceedings of the Cambridge Antiquarian Society*, *94*, 5172. https://doi.org/10.5284/1073333

Cessford, C., & Fallon, D. (2006). *Hostel Yard and Environs, Corpus Christi College, Cambridge: An Archaeological Watching Brief* (Cambridge Archaeological Unit Report No. 750).

Cessford, C., & Neil, B. (2022). The people of the Cambridge Austin friars. *Archaeological Journal*, *179*(2), 383–444. https://doi.org/10.1080/00665983.2022.2090675

Cessford, C., Scheib, C. L., Guellil, M., Keller, M., Alexander, C., Inskip, S. A., & Robb, J. E. (2021). Beyond Plague Pits: Using Genetics to Identify Responses to Plague in Medieval Cambridgeshire. *European Journal of Archaeology*, *24*(4), 496–518. https://doi.org/10.1017/eaa.2021.19

Cessford, C., & Slater, A. (2014). Beyond the Manor of Hintona Further thoughts on the development of Church End, Cherry Hinton: The Neath Farm Site. *Proceedings of the Cambridge Antiquarian Society*, *103*, 3960. https://doi.org/10.5284/1073473

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Chung, W., Chen, J., Turman, C., Lindstrom, S., Zhu, Z., Loh, P.-R., Kraft, P., & Liang, L. (2019). Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-08535-0

Craddock, P., & Gregory, V. (1975). Excavation of the Graveyard of All Saints-AD-Castra. *Cambridge. Cambridge Archaeological Unit Unpublished Report*.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1--9.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Derrien, T., Estellé, J., Marco Sola, S., Knowles, D. G., Raineri, E., Guigó, R., & Ribeca, P. (2012). Fast Computation and Applications of Genome Mappability. *PLoS ONE*, *7*(1), e30377. https://doi.org/10.1371/journal.pone.0030377

Dunbar, E., Cook, G. T., Naysmith, P., Tripney, B. G., & Xu, S. (2016). AMS [14] C Dating at the Scottish Universities Environmental Research Centre (SUERC) Radiocarbon Dating Laboratory. *Radiocarbon*, *58*(1), 9–23. https://doi.org/10.1017/RDC.2015.2

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A. M., Martini, F., Cremonesi, R. G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., & Krause, J. (2013). A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology*, *23*(7), 553–559. https://doi.org/10.1016/j.cub.2013.02.044

Guellil, M., Keller, M., Dittmar, J. M., Inskip, S. A., Cessford, C., Solnik, A., Kivisild, T., Metspalu, M., Robb, J. E., & Scheib, C. L. (2022). An invasive Haemophilus influenzae serotype b infection in an Anglo-Saxon plague victim. *Genome Biology*, *23*(1), 22. https://doi.org/10.1186/s13059-021-02580-z

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M.

D., … Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biology*, *16*(1), e2003703. https://doi.org/10.1371/journal.pbio.2003703

Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J. H., Arroyo-Pardo, E., Cavalleri, G. L., de Knijff, P., Destro Bisol, G., Dupuy, B. M., Eriksen, H. A., Jorde, L. B., King, T. E., Larmuseau, M. H., Lopez de Munain, A., Lopez-Parra, A. M., Loutradis, A., Milasin, J., Novelletto, A., … Jobling, M. A. (2015). The Y-Chromosome Tree Bursts into Leaf: 13,000 High-Confidence SNPs Covering the Majority of Known Clades. *Molecular Biology and Evolution*, *32*(3), 661–673. https://doi.org/10.1093/molbev/msu327

HMC. (1877). *Sixth Report of the Royal Commission on Historical Manuscripts. Part I: Report and Appendix*. HMSO.

Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L., & Kivisild, T. (2020). Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Scientific Reports*, *10*(1), 18542. https://doi.org/10.1038/s41598-020-75387-w

Inskip, S. A. (forthcoming). Chapter 5. Demography, social activity, lifestyle and illness. In J. E. Robb (Ed.), *Health and History in medieval Cambridge*.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Järve, M., Talas, U. G., Rootsi, S., Ilumäe, A. M., Mägi, R., Mitt, M., Pagani, L., Puurand, T., Faltyskova, Z., Clemente, F., Cardona, A., Metspalu, E., Sahakyan, H., Yunusbayev, B., Hudjashov, G., … Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, *25*(4), 459–466. https://doi.org/10.1101/gr.186684.114

Kivisild, T., Saag, L., Hui, R., Biagini, S. A., Pankratov, V., D'Atanasio, E., Pagani, L., Saag, L., Rootsi, S., Mägi, R., Metspalu, E., Valk, H., Malve, M., Irdt, K., Reisberg, T., Solnik, A., Scheib, C. L., Seidman, D. N., Williams, A. L., … Metspalu, M. (2021). Patterns of genetic connectedness between modern and medieval Estonian genomes reveal the origins of a major ancestry component of the Finnish population. *The American Journal of Human Genetics*, *108*(9), 1792–1806. https://doi.org/10.1016/j.ajhg.2021.07.012

Klunk, J., Vilgalys, T. P., Demeure, C. E., Cheng, X., Shiratori, M., Madej, J., Beau, R., Elli, D., Patino, M. I., Redfern, R., DeWitte, S. N., Gamble, J. A., Boldsen, J. L., Carmichael, A., Varlik, N., Eaton, K., Grenier, J.-C., Golding, G. B., Devault, A., … Barreiro, L. B. (2022). Evolution of immune genes is associated with the Black Death. *Nature*, *611*(7935), 312–319. https://doi.org/10.1038/s41586-022-05349-x

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356–356. https://doi.org/10.1186/s12859-014-0356-4

Kuhn, J. M., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLOS ONE*, *13*(4), e0195491–e0195491. https://doi.org/10.1371/journal.pone.0195491

Lally, M. (2008). *69 to 115 Church End, Cherry Hinton, Cambridgeshire: Post Excavation Assessment and Updated Project Design. Archaeological Solutions Report 3012*.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

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

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L. M., Jørsboe, E., Ingason, A., Pedersen, M. W., Korneliussen, T., Wilhelmson, H., Buś, M. M., de Barros Damgaard, P., Martiniano, R., Renaud, G., Bhérer, C., Moreno-Mayar, J. V., Fotakis, A. K., … Willerslev, E. (2020). Population genomics of the Viking world. *Nature*, *585*(7825), 390–396. https://doi.org/10.1038/s41586-020-2688-8

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. https://doi.org/10.14806/ej.17.1.200

Martiniano, R., Caffell, A., Holst, M., Hunter-Mann, K., Montgomery, J., Müldner, G., McLaughlin, R. L., Teasdale, M. D., Van Rheenen, W., Veldink, J. H., Van Den Berg, L. H., Hardiman, O., Carroll, M., Roskams, S., Oxley, J., Morgan, C., Thomas, M. G., Barnes, I., McDonnell, C., … Bradley, D. G. (2016). Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature Communications*, *7*(1), 1–8. https://doi.org/10.1038/ncomms10326

Martiniano, R., De Sanctis, B., Hallast, P., & Durbin, R. (2022). Placing Ancient DNA Sequences into Reference Phylogenies. *Molecular Biology and Evolution*, *39*(2), msac017. https://doi.org/10.1093/molbev/msac017

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. https://doi.org/10.1038/ng.3643

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Newman, R. (2018). *Holy Trinity Church, Cambridge: Archaeological Excavation and Monitoring, 2016-2017* [PDF]. Archaeology Data Service. https://doi.org/10.5284/1092761

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, btv566. https://doi.org/10.1093/bioinformatics/btv566

Olalde, I., Allentoft, M. E., Sánchez-Quinto, F., Santpere, G., Chiang, C. W. K., DeGiorgio, M., Prado-Martinez, J., Rodríguez, J. A., Rasmussen, S., Quilez, J., Ramírez, O., Marigorta, U. M., Fernández-Callejo, M., Prada, M. E., Encinas, J. M. V., Nielsen, R., Netea, M. G., Novembre, J., Sturm, R. A., … Lalueza-Fox, C. (2014). Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*, *507*(7491), 225–228. https://doi.org/10.1038/nature12960

Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A., Ayub, Q., McCarthy, S. A., Narechania, A., Kashin, S., Chen, Y., Banerjee, R., Rodriguez-Flores, J. L., Cerezo, M., Shao, H., Gymrek, M., Malhotra, A., Louzada, S., Desalle, R., … Tyler-Smith, C. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*, *48*(6), 593–599. https://doi.org/10.1038/ng.3559

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rees, G. (2014). *A 19th Century Baptist Cemetery at St Matthew's Primary School, Norfolk Street, Cambridge*. Oxford Archaeology East. https://doi.org/10.5284/1030510

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., … Talamo, S. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon*, *62*(4), 725–757. https://doi.org/10.1017/RDC.2020.41

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., … van der Plicht, J. (2013). IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon*, *55*(4), 1869–1887. https://doi.org/10.2458/azu_js_rc.55.16947

Ringbauer, H., Novembre, J., & Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nature Communications*, *12*(1), 5425. https://doi.org/10.1038/s41467-021-25289-w

Saag, L., Vasilyev, S. V., Varul, L., Kosorukova, N. V., Gerasimov, D. V., Oshibkina, S. V., Griffith, S. J., Solnik, A., Saag, L., D'Atanasio, E., Metspalu, E., Reidla, M., Rootsi, S., Kivisild, T., Scheib, C. L., Tambets, K., Kriiska, A., & Metspalu, M. (2021). Genetic ancestry changes in Stone to Bronze Age transition in the East European plain. *Science Advances*, *7*(4), eabd6535. https://doi.org/10.1126/sciadv.abd6535

Saupe, T., Montinaro, F., Scaggion, C., Carrara, N., Kivisild, T., D'Atanasio, E., Hui, R., Solnik, A., Lebrasseur, O., Larson, G., Alessandri, L., Arienzo, I., De Angelis, F., Rolfo, M. F., Skeates, R., Silvestri, L., Beckett, J., Talamo, S., Dolfini, A., … Scheib, C. L. (2021). Ancient genomes reveal structural shifts after the arrival of Steppe-related ancestry in the Italian Peninsula. *Current Biology*, *31*(12), 2576-2591.e12. https://doi.org/10.1016/j.cub.2021.04.022

Schiffels, S., Haak, W., Paajanen, P., Llamas, B., Popescu, E., Loe, L., Clarke, R., Lyons, A., Mortimer, R., Sayer, D., Tyler-Smith, C., Cooper, A., & Durbin, R. (2016). Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature*

Communications, *7*(1), 1–9. https://doi.org/10.1038/ncomms10408

Seidman, D. N., Shenoy, S. A., Kim, M., Babu, R., Woods, I. G., Dyer, T. D., Lehman, D. M., Curran, J. E., Duggirala, R., Blangero, J., & Williams, A. L. (2020). Rapid, Phase-free Detection of Long Identity-by-Descent Segments Enables Effective Relationship Classification. *The American Journal of Human Genetics*, *106*(4), 453–466. https://doi.org/10.1016/j.ajhg.2020.02.012

Sims, S. (2007). *Hemingfords Flood Alleviation Scheme, St Ives, Cambridgeshire: Archaeological watching brief report*. Oxford Archaeology.

Skoglund, P., Storå, J., Götherström, A., & Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*, *40*(12), 4477–4482. https://doi.org/10.1016/j.jas.2013.07.004

Stan Development Team. (2023). *RStan: The R interface to Stan*. https://mc-stan.org/

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, *526*, 68–68.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2022). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*. https://mc-stan.org/loo/

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*, *44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

White, T. D., & Folkens, P. A. (2005). *The human bone manual*. Elsevier Academic.

Supplementary Figures
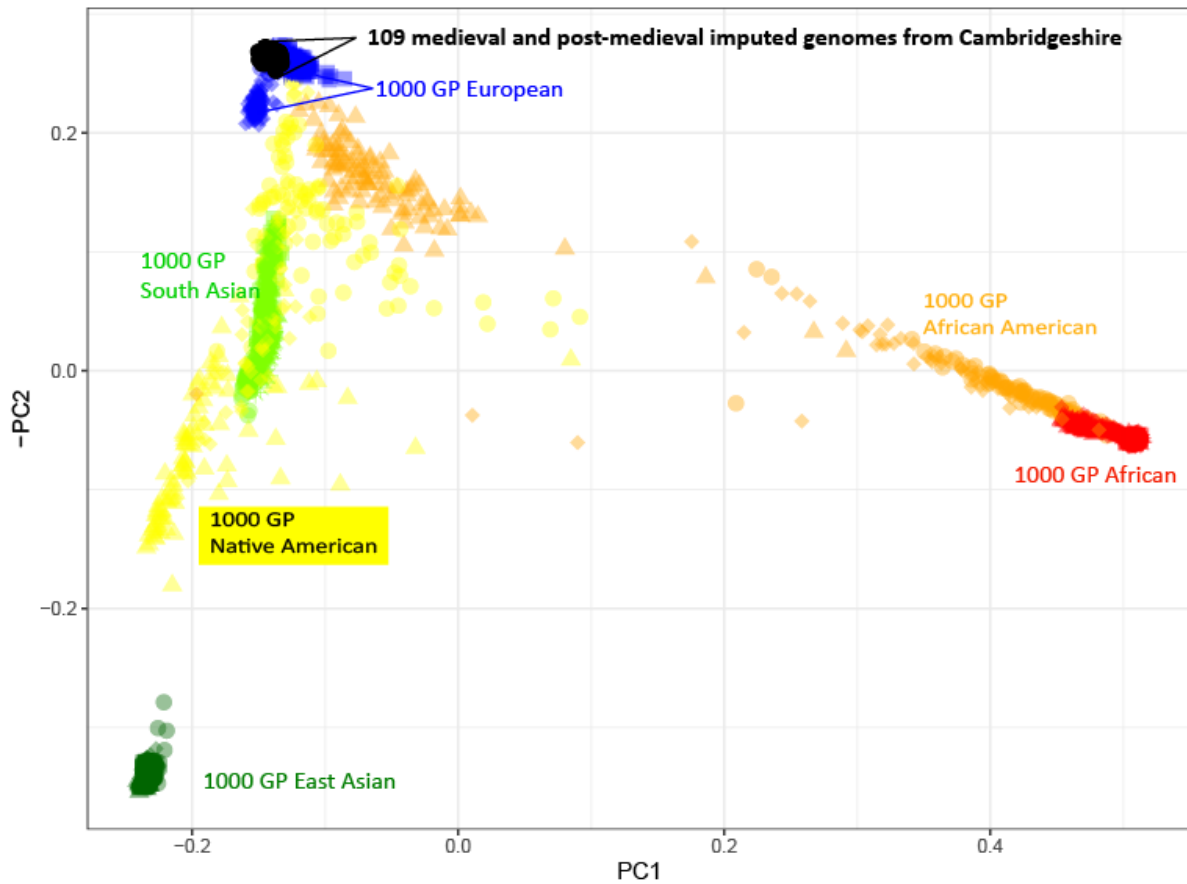
Figure S1. PCA of 109 later medieval and post-medieval genomes (>0.1x) in context of 1000 Genome Project (1000 GP) data.

Figure S2-S11. PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from each site. Only genomes > 0.05x are included; error bars represent one standard deviation estimated from 20-fold block jackknife.
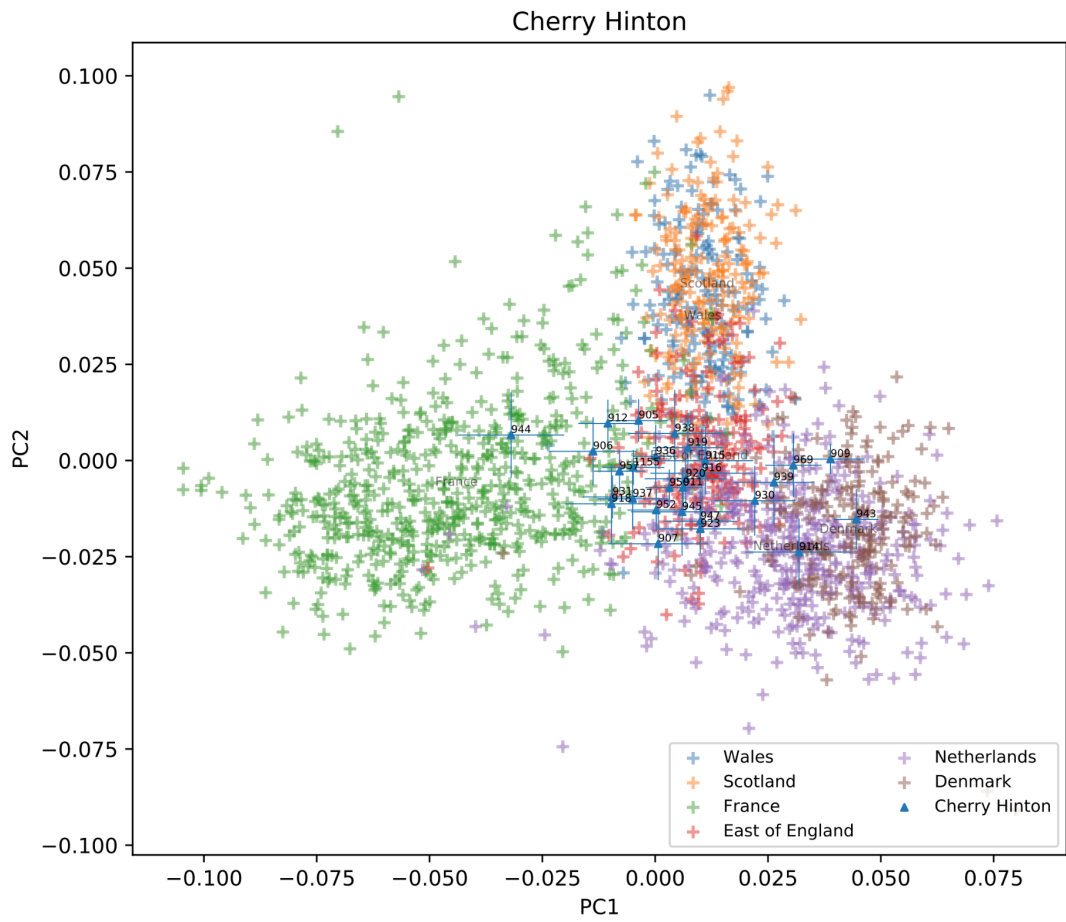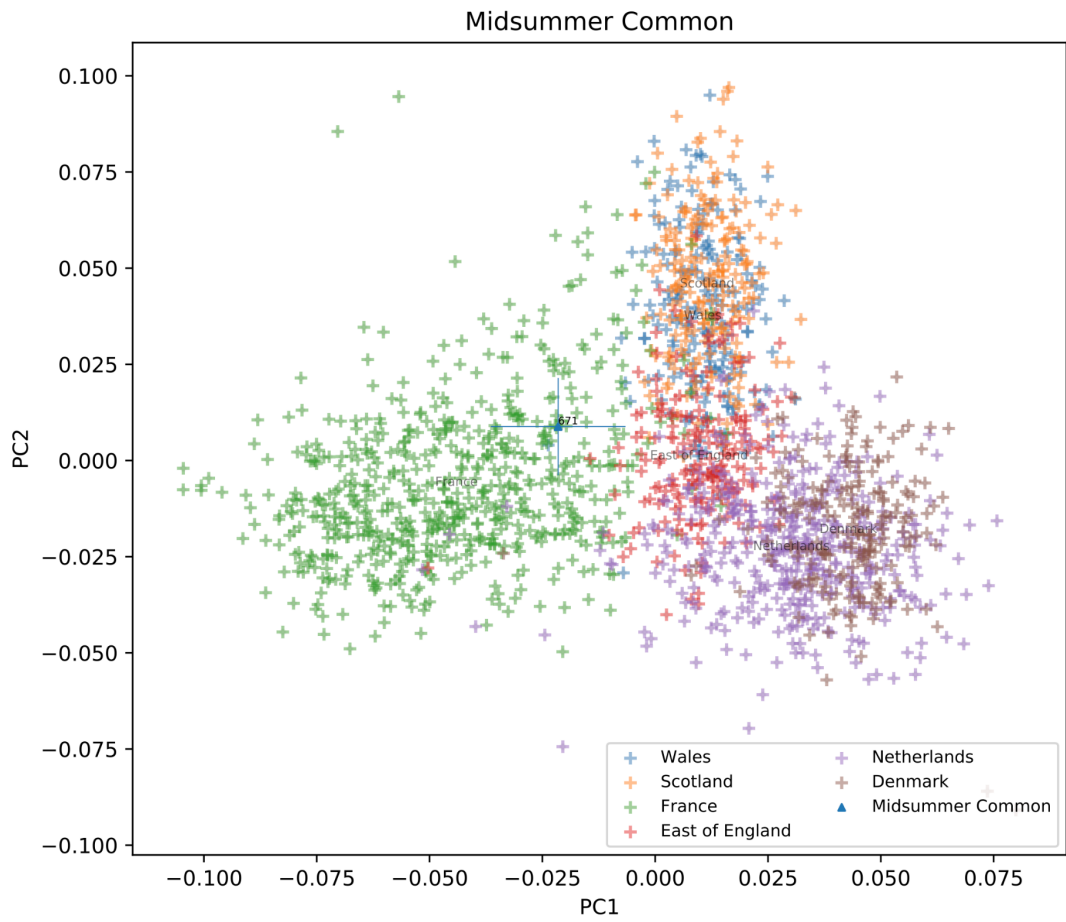
Figure S2: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Cherry Hinton.
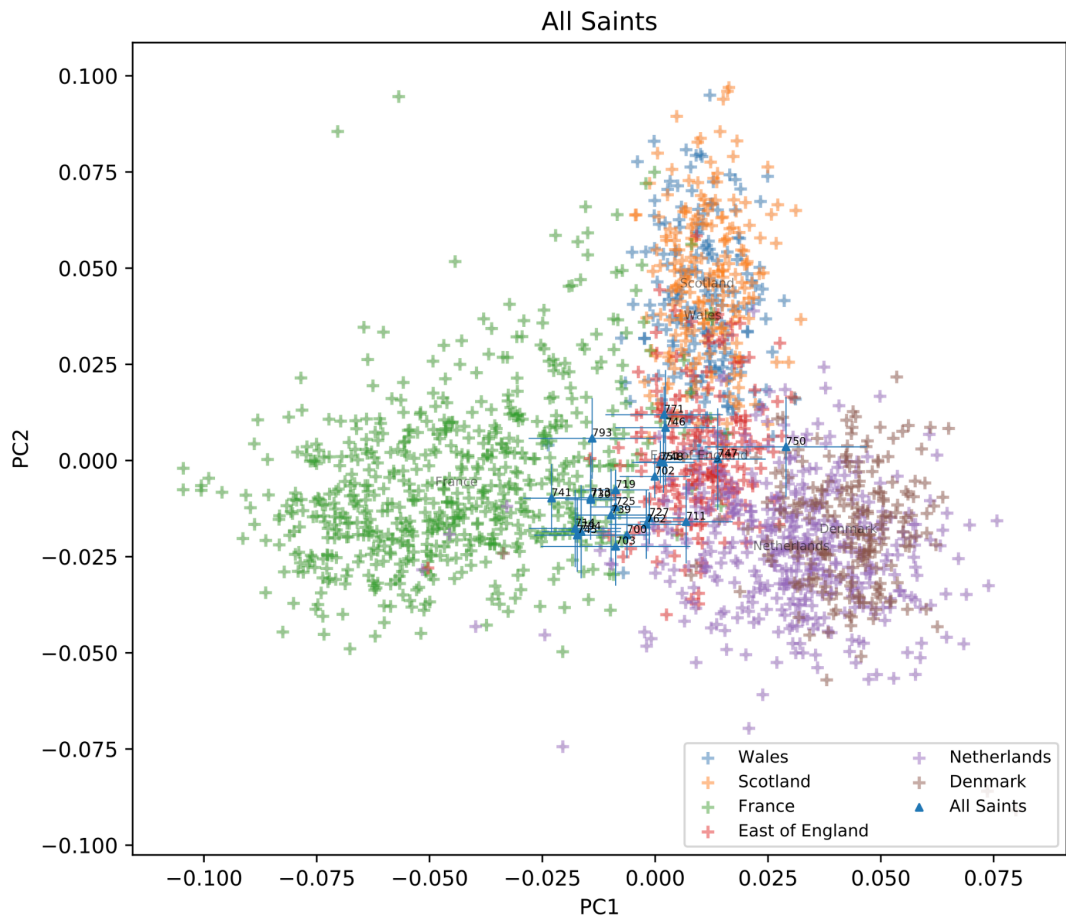
Figure S3: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from All Saints by the Castle.
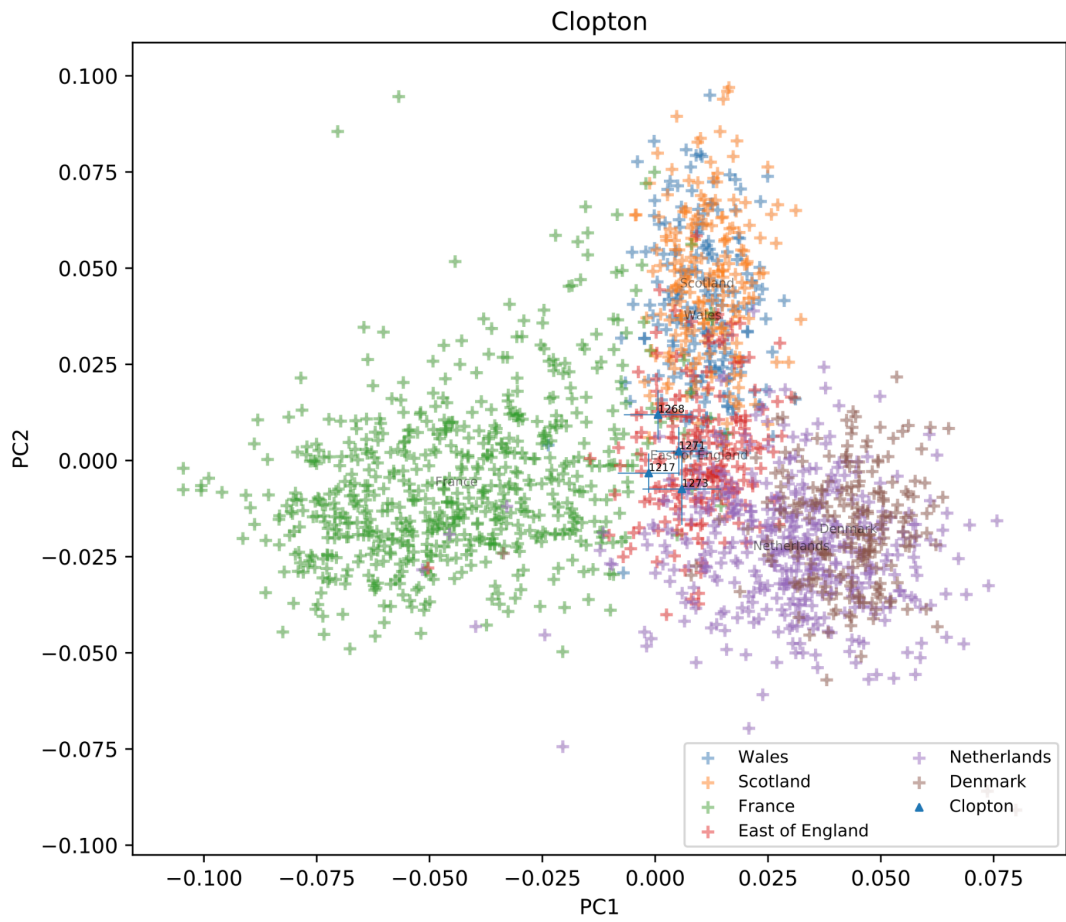
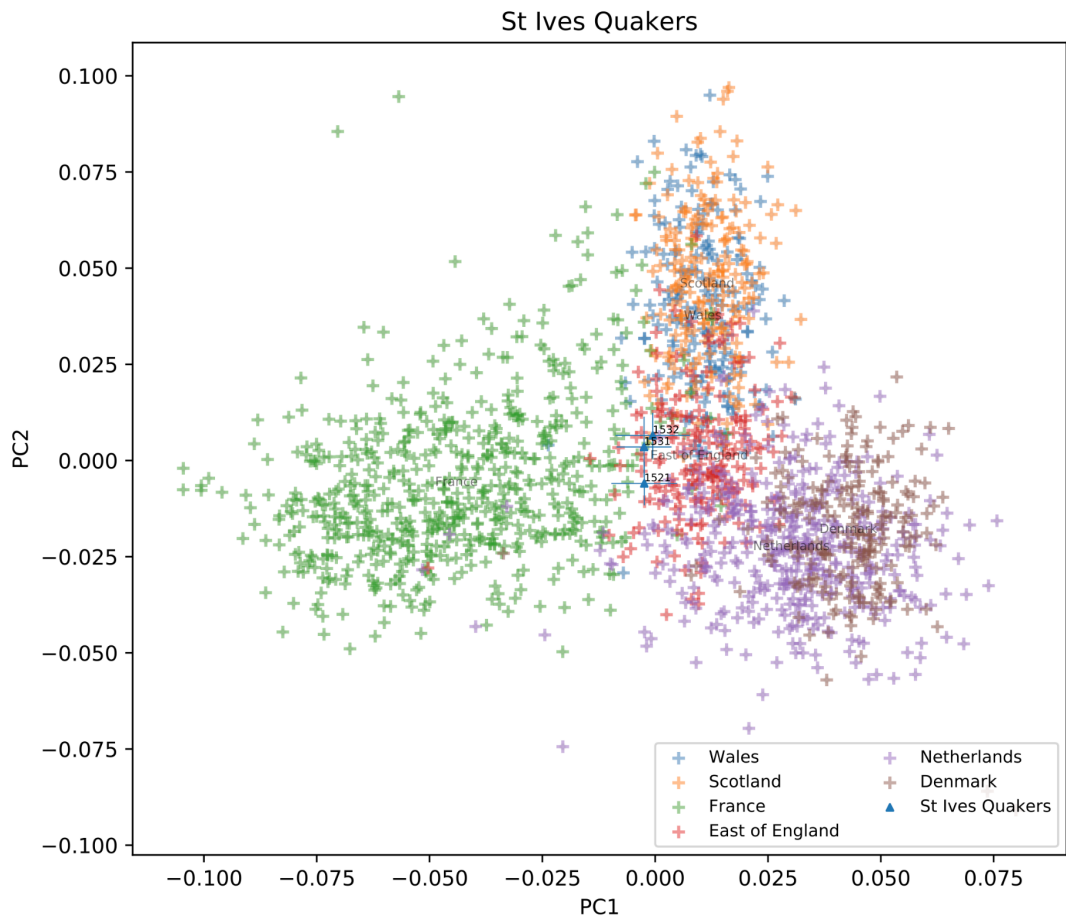**Figure S4: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Clopton.**
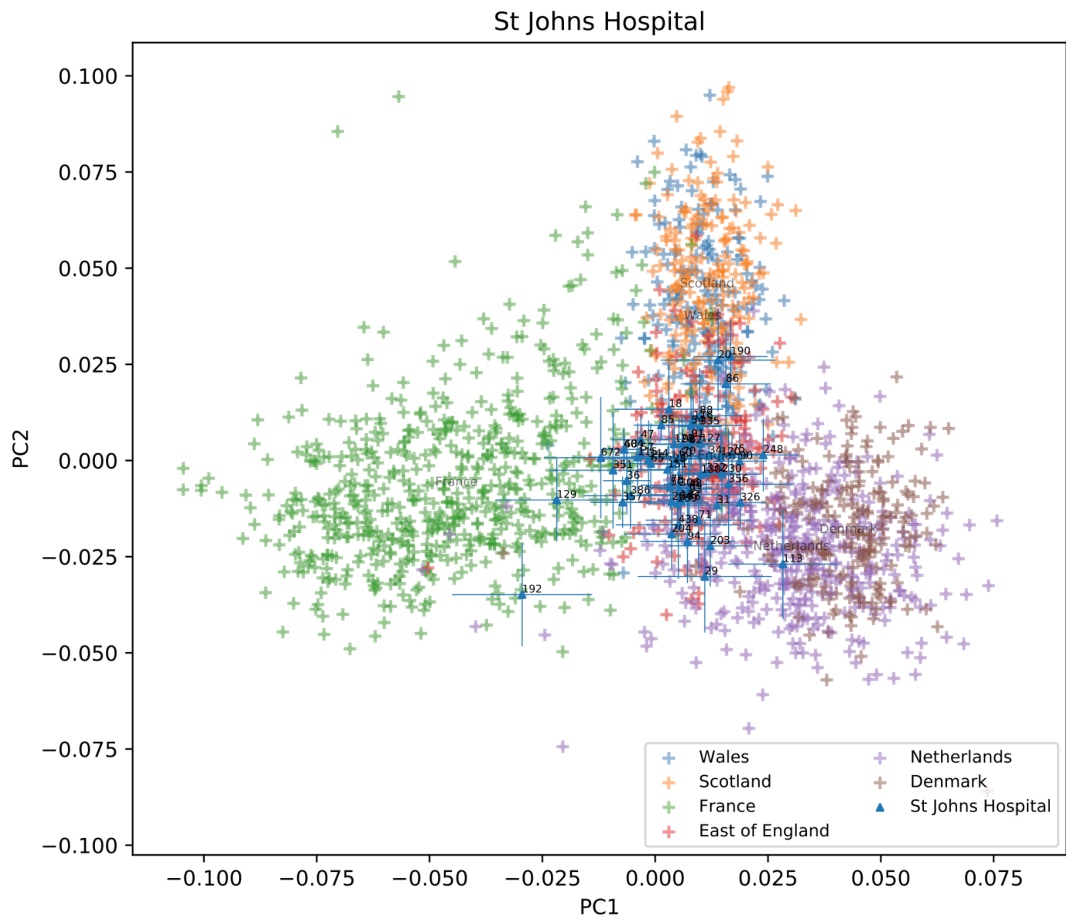
Figure S5: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from the Hospital of St John.
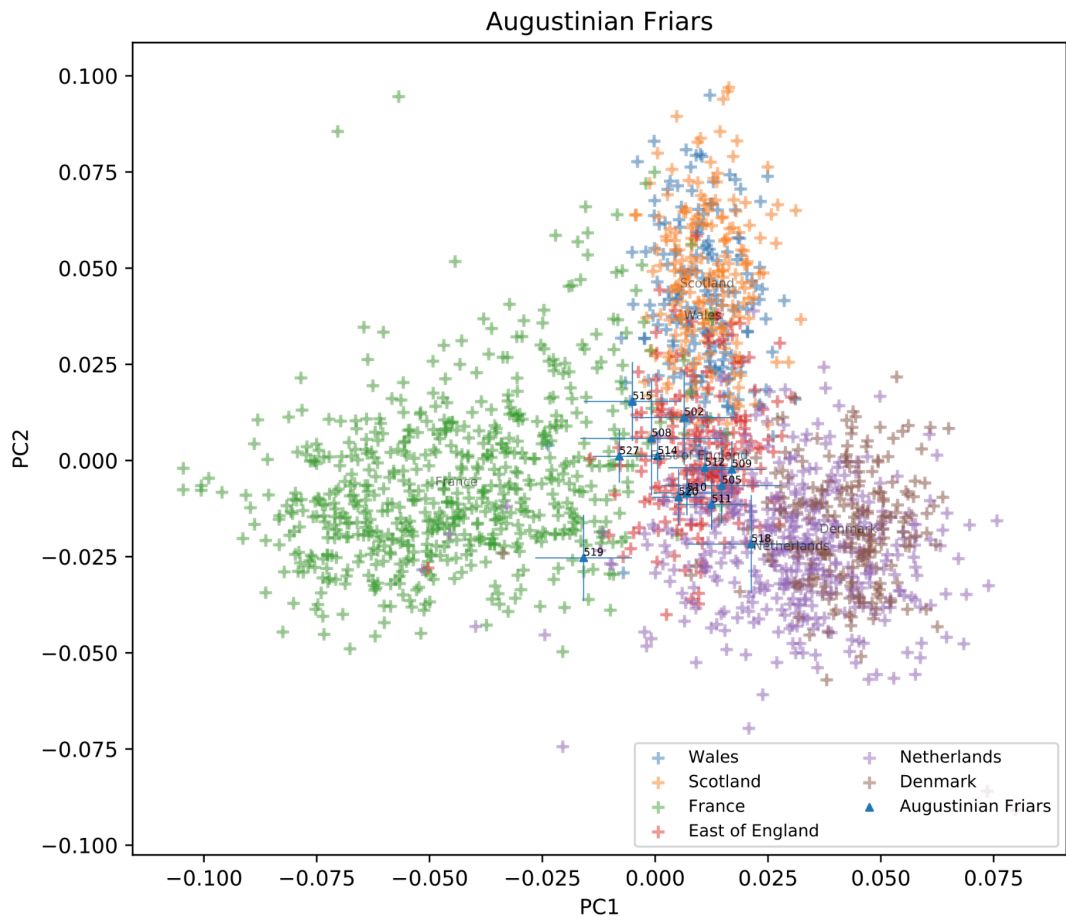
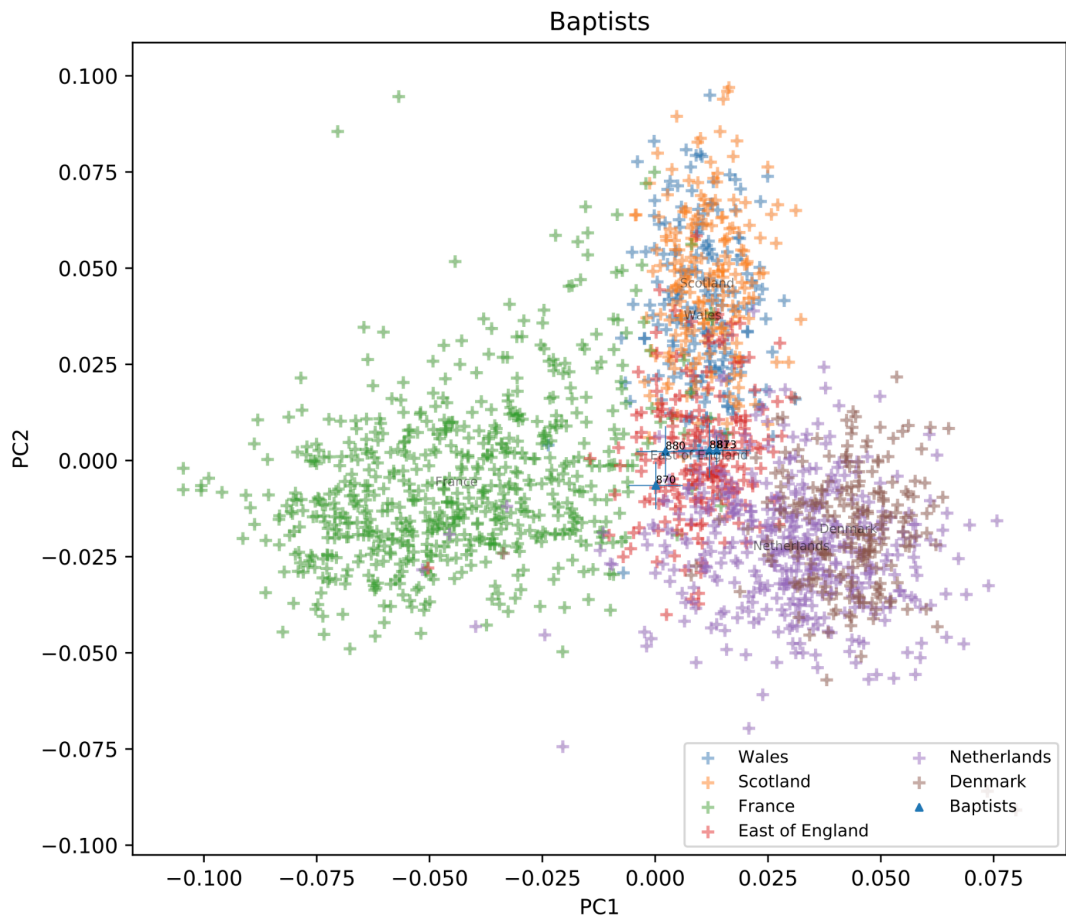**Figure S6: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Augustinian Friary.**
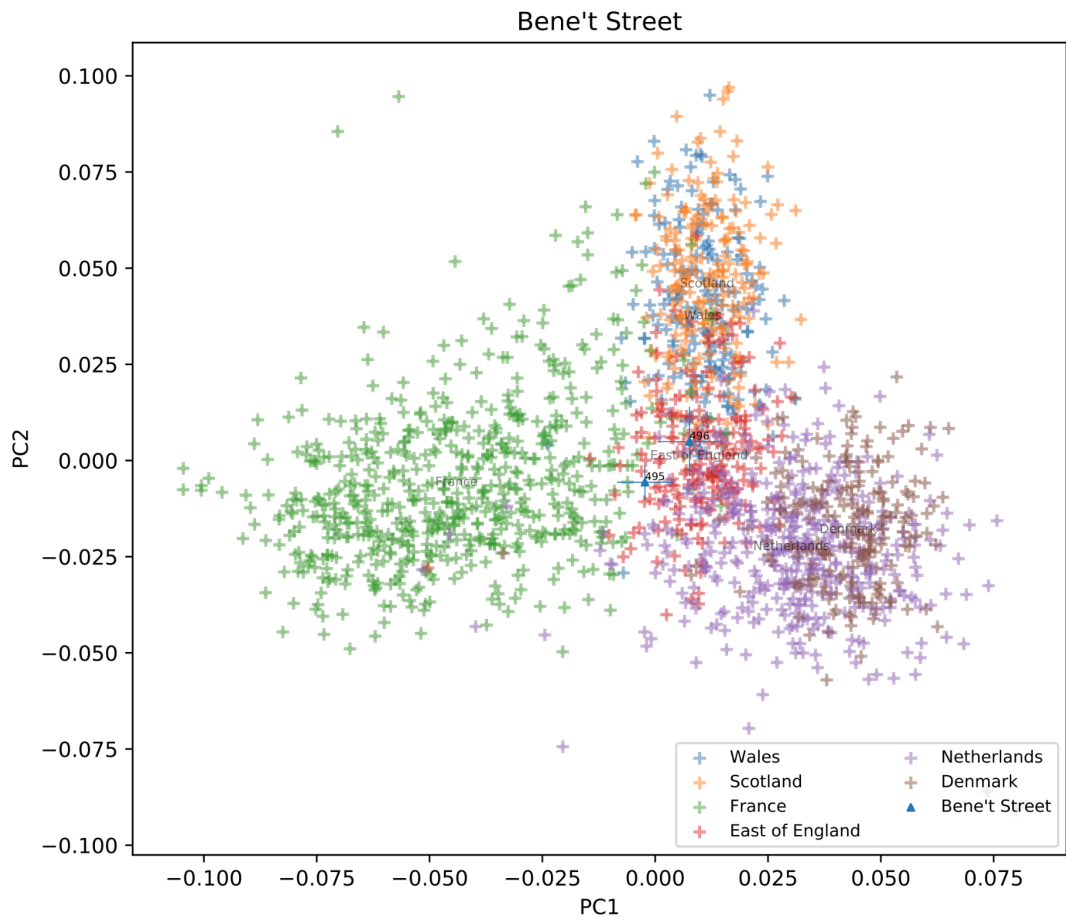
Figure S7: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Bene't Street.

Figure S8: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Midsummer Common.

Figure S9: PCA of West Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Hemingford Gray Quakers.

Figure S10: PCA of west Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Providence Calvinistic Baptist Chapel.
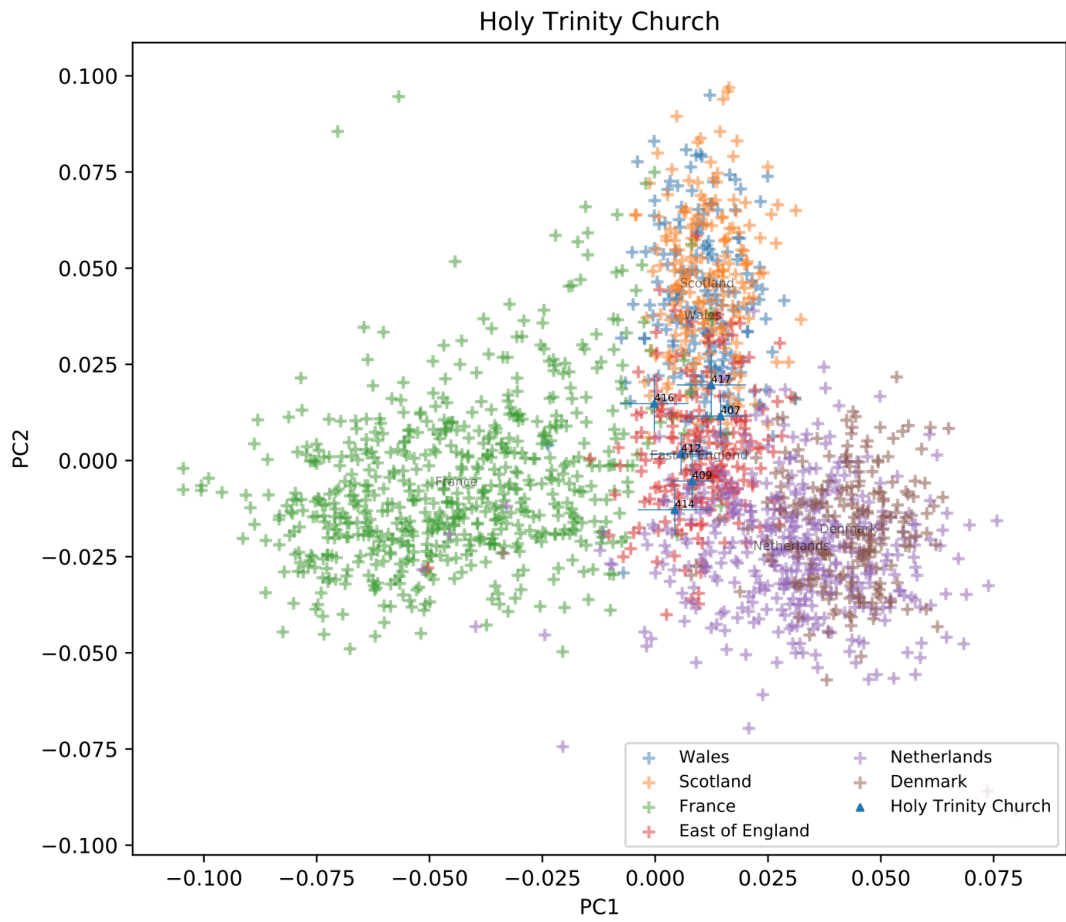
Figure S11: PCA of west Europe individuals from UK Biobank, showing lsq-projection of historical genomes from Holy Trinity Church.
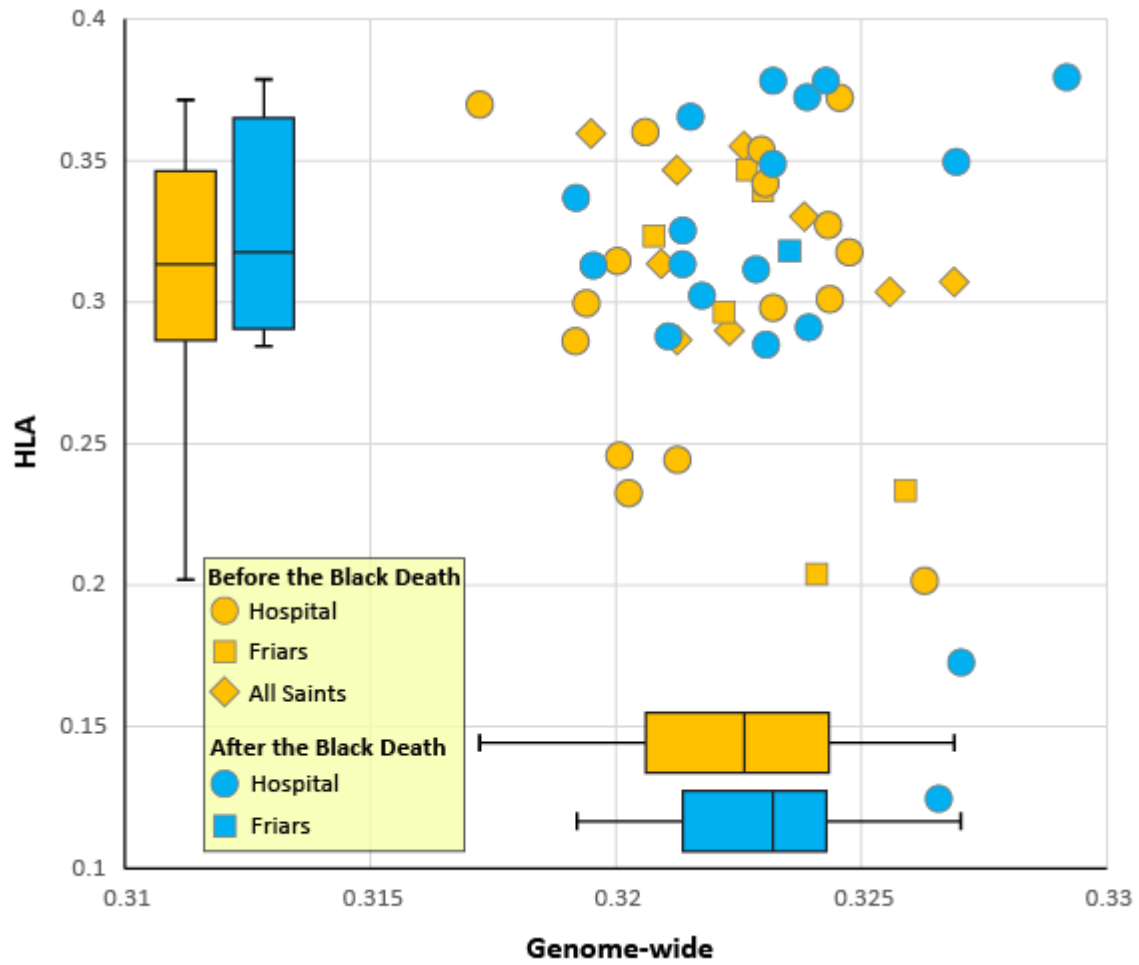
Figure S12. Heterozygosity of the medieval Cambridge genomes from before and after the Black Death.
Average heterozygosity estimates for 5.4 million variants with MAF>0.05 were obtained from imputed
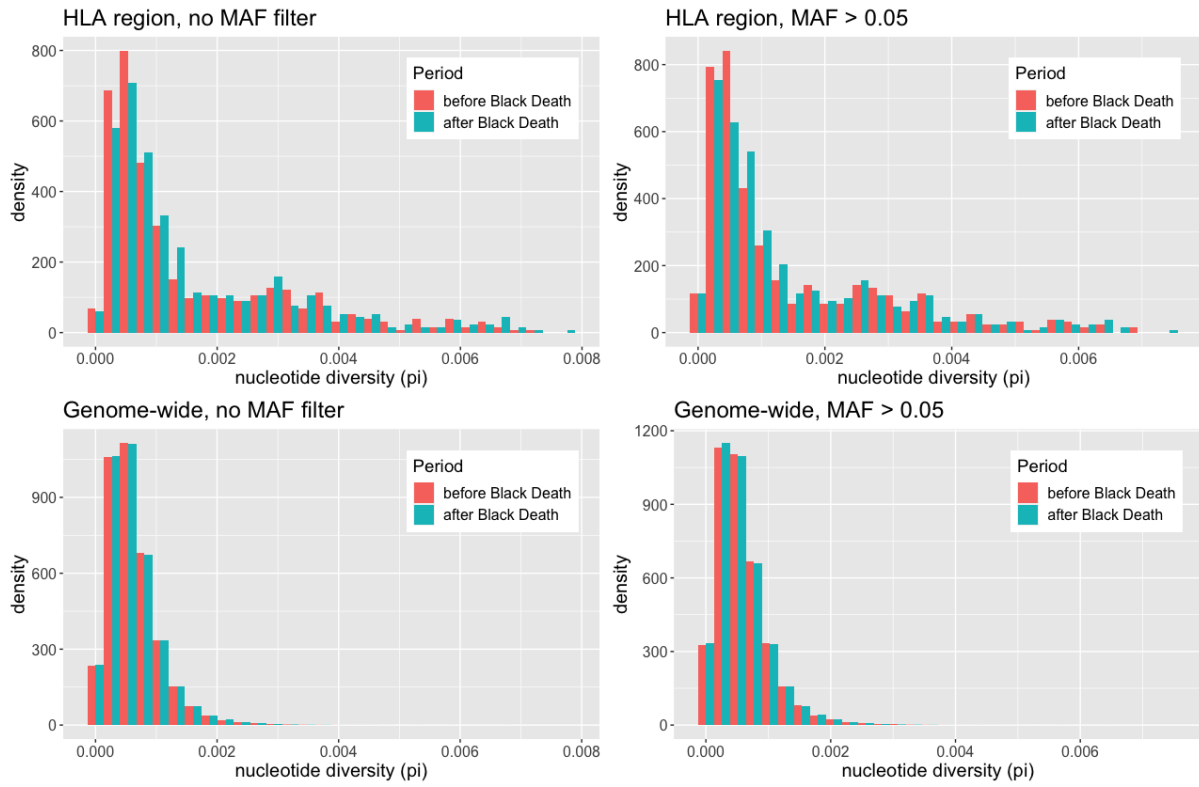genotypes of 50 genomes with coverage >0.1x from 4 sites in Cambridge.

Figure S13. Histograms comparing nucleotide diversity of the medieval Cambridge genomes from before and after the Black Death.