Population-scale analysis of inheritance patterns across 858,635 individuals reveals recent historical migration patterns across the North Sea from the Middle Ages to the Industrial Revolution

Xiaolei Zhang[1], Poul Holm[2], Aylwyn Scally[3], Karina Banasik[4,5,9], Laust Hvas Mortensen[6,7,8], Tomas Fitzgerald[1], David Westergaard[4,5,6,9], Piotr Jaroslaw Chmura[4], Sisse Rye Ostrowski[10,11], Erik Sørensen[11], Ole B.V. Pedersen[10,12], Michael Schwinn[11], Christian Erikstrup[13,14], Mie T Bruun[15], Bitten A Jensen[16], Henrik Ullum[17], Christina Mikkelsen[11,18], Thomas Folkmann Hansen[4], DBDS Genetic Consortium, Kari Stefansson[19,20], Søren Brunak[4, 7,*], Ewan Birney[1,*]

1. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK
2. Trinity Centre for Environmental Humanities, School of Histories and Humanities, Trinity College, Dublin, Ireland
3. Department of Genetics, University of Cambridge, Cambridge, UK
4. Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
5. Department of Obstetrics and Gynaecology, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark
6. Statistics Denmark, Copenhagen, Denmark
7. Department of Public Health, University of Copenhagen, Copenhagen, Denmark
8. ROCKWOOL Foundation, Copenhagen, Denmark
9. Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

10. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
11. Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
12. Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark
13. Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark
14. Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
15. Clinical Immunology Research Unit, Department of Clinical Immunology, Odense University Hospital, Odense, Denmark
16. Department of Clinical Immunology, Aalborg University Hospital, Aalborg, Denmark
17. Statens Serum Institut, Copenhagen, Denmark
18. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
19. deCODE genetics/Amgen, Inc, University of Iceland, Reykjavik, Iceland
20. Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland

*Corresponding authors: Søren Brunak (soren.brunak@cpr.ku.dk) and Ewan Birney(birney@ebi.ac.uk).

Abstract

The North Sea's historical migrations have impacted the genetic structure of its neighbouring populations. We analysed haplotype sharing among 858,635 modern individuals from Denmark and Britain to infer migration history from the Middle Ages to the modern day. We estimated the genetic relationship among 370,259 Danes using Danish healthcare registries and validated those with retrievable pedigree relationships from a national family registry. We also compared their IBD sharing with 488,376 British individuals from the UK Biobank.

Our analysis revealed the fine-grained population genetic history of Denmark and identified distinct coastal and island communities with a history of genetic isolation and bottlenecks. We observed a significant population decline in Jutland compared to Zealand during the late medieval period to the start of early modern period, accompanied by migration from Jutland to Zealand, corresponding to historical evidence.

We identified two major IBD sharing patterns between Denmark and Great Britain: early coast-to-coast connections between South Jutland and eastern England, likely driven by Viking settlements, continuous trade and people movements across the North Sea from the Middle Ages through early modern times, and later city-to-city connections such as those between London and København likely influenced by urbanisation and Industrial Revolution. Further comparisons using other North Sea countries showed both shared and unique histories of genetic exchange with Denmark and Britain.

Our study provides novel genetic evidence of migration across the North Sea from the Middle Ages to the Industrial Revolution. It highlights the power of nationwide biobanks in reconstructing fine-scale historical population movements among closely related populations.

Introduction

The North Sea area has a fascinating history due to its location and historical interactions. Its rich maritime heritage, including fishing, trade, and wars, has shaped the populations living around the North Sea. In this study, we focus on the migration between Britain and Denmark, two countries with extensive genetic resources available for responsible research, including historical studies.

Human genetics studies have been previously used to discover and explore historical migrations on the genetic makeup of modern-day populations. The sources of genetic material can either be ancient DNA combined with present-day samples, or just present day samples alone. Previous studies made use of present-day samples and showed that there is a genetic affinity between British and Danish populations (Athanasiadis et al. 2016; Leslie et al. 2015). The studies are only based on a few thousand samples, and so have limited spatial and temporal resolutions. From ancient human DNA samples, there is strong genetic evidence supporting the historical migration between Northern Europe, Scandinavia and Britain during the Late Roman and early Medieval periods (Margaryan et al. 2020; Speidel et al. 2025; Gretzinger et al. 2022; Rodríguez-Varela et al. 2023).

However, there are discrepant conclusions regarding both the strength and the geographic distribution of Viking genetic legacy in Britain among studies based on autosomal and Y-chromosome markers (Leslie et al. 2015; Lall et al. 2021), as well as archaeological and linguistic evidence (J. Kershaw and Røyrvik 2016). There are also limited genetic studies of the genetic exchange post-Viking age across the North Sea. Despite many unclear or unanswered questions in more recent history, it has not been routine to use the extensive genetic resources from modern day samples to inform the movement and links of groups of

people at scale across these timescales. Additionally, Danish populations were previously considered as remarkably homogeneous (Athanasiadis et al. 2016), and there has not been a large-scale genetic study examining the fine-scale population structure and history of Denmark.

In this study, we assembled 858,635 individuals' genetic data from two biobanks located in the UK and Denmark, which is at least 100-fold higher than the sample sizes in previous studies. Due to the scale of our datasets, we can provide an unprecedented geographical and time based view of shared ancestry. At the finest scale that we considered in the study, all NUTS3 (Nomenclature of Territorial Units for Statistics) levels of regions in Britain, at sub-county level, have more than 40 individuals (median=1,256) born at these locations. Similarly in Denmark we can use the municipality level for Danish sub-structure, with at least 73 individuals (median=1,827) (Figure S1). From these biobanks, we identified all the shared genomic segments between any two individuals, inherited from their most recent common ancestors unbroken by recombinations, i.e. haplotype sharing and Identity-by-descent (IBD) sharing. Due to recombinations, the length of shared segments indicates the probable age of their common ancestor, with shorter segments mostly likely inherited from older ancestors and longer segments from recent ones, thus allowing us to identify signals of shared ancestry across different time periods.

In total, we analysed a little over 100 billion IBD segments estimated in the 368 billion pairs of individuals to (1) characterise the population structure and infer demographic history within Denmark; (2) identify the hotspots of IBD sharing between Britain and Denmark across time, stratified by lengths of IBD segments, and (3) analyse the IBD sharing of Britain and Denmark with other neighbouring countries using proxy populations to identify other

potential migration patterns. The proxy populations for non-British or non-Danish European countries were from people born in and inferred to be from other European countries within our assembled UK and Denmark cohorts. We interpreted our genetic findings in comparison with existing historical evidence, where we both support and challenge aspects of the historical record.

Results

Generation of 368.6 billion individual pairs of IBD sharing

We used the large-scale UK BioBank (UKBB)(Bycroft et al. 2018) resource with 488,376 individuals as a sampling of the British population (approximately 0.7% of the total population), and the Danish Blood Donor Study and the Copenhagen Hospital Biobank with a total of 370,259 individuals for a sampling of the Danish population (approximately 6.2% of the total population) (Figure 1 and Methods). Notably, we also include individuals from the Faroe Islands in the Danish cohorts to compare their distinct genetic history with that of Denmark ("continental Denmark"), as shown in previous genetic studies (Gislason 2023; Als et al. 2006). For the population of Britain, we include 410,266 samples of self-reported ethnicity as "White British". As there is no ethnicity information available in the Danish administrative datasets, we refer to the inferred European samples in the Danish cohorts as "Danes of European ancestries" (n=316,512), representing the population with mainly Danish ancestry (see Methods).

We estimated the IBD sharing independently across the samples on three datasets, each defined by a specific set of SNP markers: (1) within the Britain, using the full SNP markers in the UKBB; (2) within Denmark, using the full SNP markers in the Danish samples (DNK);

(3) between the Britain and Denmark, using the intersecting SNP markers shared after combining the UKBB and Danish genotype datasets (UKBB+DNK).

We used the FastSMC (Nait Saada et al. 2020) to estimate the IBD segments for all possible pairs of individuals. The total number of pairwise comparisons is substantial, with 368.6 billion pairs of samples evaluated. Across the three datasets, we discovered and processed 100.8 billion IBD segments in total. To efficiently manage the computational results, we developed a high-performance solution utilizing bespoke data structures, parallel computing and indexing to analyse the IBD data. All computations were performed on the secure computing infrastructure of the Danish National Genome Centre and EMBL's European Bioinformatics Institute.

We consistently use the Danish spelling of city place names in Denmark instead of the anglicised names, with the most obvious difference being København (Copenhagen), or the spelling preferred by the local authorities (e.g Aabenraa and Aarhus). For geographical terms such as the regional names (e.g. Jutland and Zealand), we use the anglicised spelling for ease of comprehension.

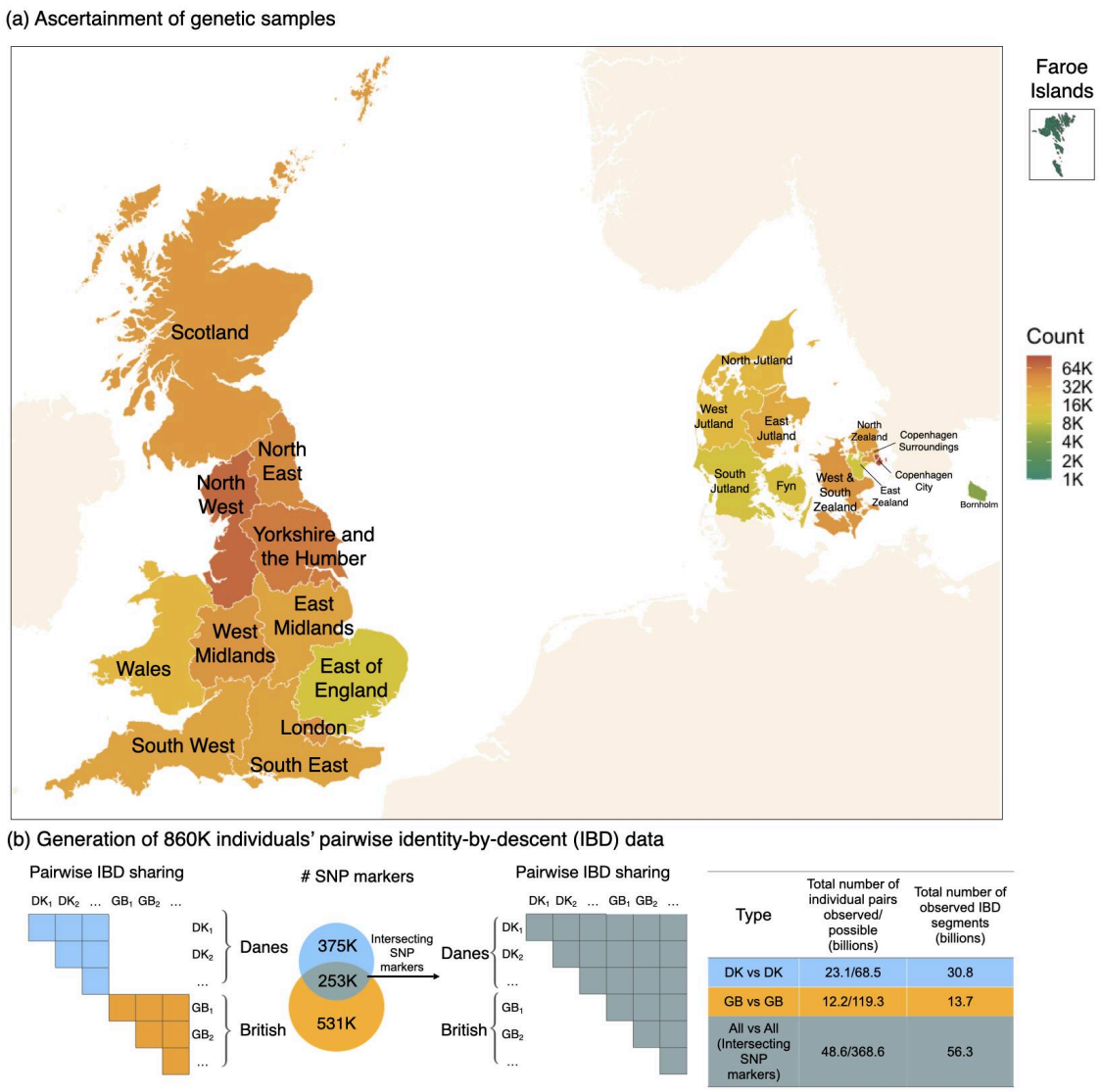

Figure 1. Overview of data collection and study design. (a) Geographic distribution of sample birthplaces and corresponding sample sizes across Britain (including regions of England, Wales and Scotland), Denmark and the Faroe Islands. (b) Diagram for estimating the IBD sharing within the Britain (Britain vs Britain), within Denmark (DK vs DK), and between the Britain and Denmark (included in the All vs All analysis using the intersecting genotype data), along with a summary of the total pairwise comparisons evaluated and the IBD data generated.

Overall properties of IBD sharing via 369 billion pairwise comparisons

We focus on IBD segments greater than 2cM where we are confident that the IBD estimations are sensitive and accurate (see Methods). Figure 2 shows the distribution of IBD sharing among the 860K samples using the intersection marker set. We first considered pairs segregated by their birthplace from the island of Great Britain in the UK (England, Scotland and Wales; Northern Ireland is not considered here due to the low sample size in the UKBB), Denmark, and the Faroe Islands. We see intra-regional IBD sharing is consistently higher than inter-regional, with geographic distinct comparisons such as IBD sharing between Britain and Denmark being the lowest (Figure 2 and Table S1).

Most of the pairs in the populations from Britain and Denmark share at most one short IBD segment, indicating sharing at most one distant common genetic ancestor. The distributions of total amount of IBD sharing and the number of IBD segments are close to exponentially decayed (Figure 2), indicating that the number of detected distant relatives increases far more rapidly than that of closer relatives. This is consistent with established theory and other examples of smaller-scale IBD studies in largely outbred human populations(Huff et al. 2011; Palamara 2014; Henn et al. 2012; Nait Saada et al. 2020; Shemirani et al. 2021). However, most pairs of individuals from the Faroe Islands share multiple IBD segments, with an average of 10.4 segments and 10.7 cM total IBD per pair. This is 4 times and 41 times higher, respectively, than among Denmark (Table S1). This high level of intra-population sharing reflects that this population has been genetically isolated and went through genetic bottlenecks, consistent with previous studies on the Faroe Islands (Gislason 2023).

We leveraged the uniquely comprehensive and digitalised Multi-Generation Register Lite (MGR-lite) (Due et al. 2024; Westergaard et al. 2024) of all Danes (n=7,771,981) to reconstruct pedigree relationships among the 370K Danish genetic samples in our cohorts. This family registry documents the relationships between parents (i.e legal guardians) and their children who were alive in 1968 and born from the year 1953, and is partially complete for people born before 1953. Based on the familial registry, we inferred 56,329 pairs of direct ancestor-offspring relatives and 54,089 pairs of relatives with shared common ancestors (excluding direct ancestor-offspring relatives), including both close and distant cousin relationships up to fifth-degree relatives (Table S2 and Table S3). For these 110,418 pairs of individuals, we can infer the total number of meiosis separating the two individuals from their pedigree relationship. In Figure 2 we show the distance in separating meiosis compared to the total IBD sharing, which follows closely the expectation from theoretical models (Huff et al. 2011). For parent-child and grandparent-grandchild relatives, we observe the long tails on the lower end, suggesting possible unbiological related pairs due to civil registration errors or unrecorded adoptions in the pedigree record, also leading to skewed distributions among shared ancestor pairs. We also observe a minor mode in the shared ancestor pairs, indicating cases with single shared ancestors, such as half-siblings or half-cousins. Overall we have a high confidence that the IBD sharing is exponentially decayed with the increasing number of separating meiosis among genetic relatives, thus helping to infer the age of common ancestors. For ease of presentation, we present data with a time-based estimate using an average human historical generation time (average reproductive age) estimate of 28 years (Moorjani et al. 2016) in subsequent analysis.

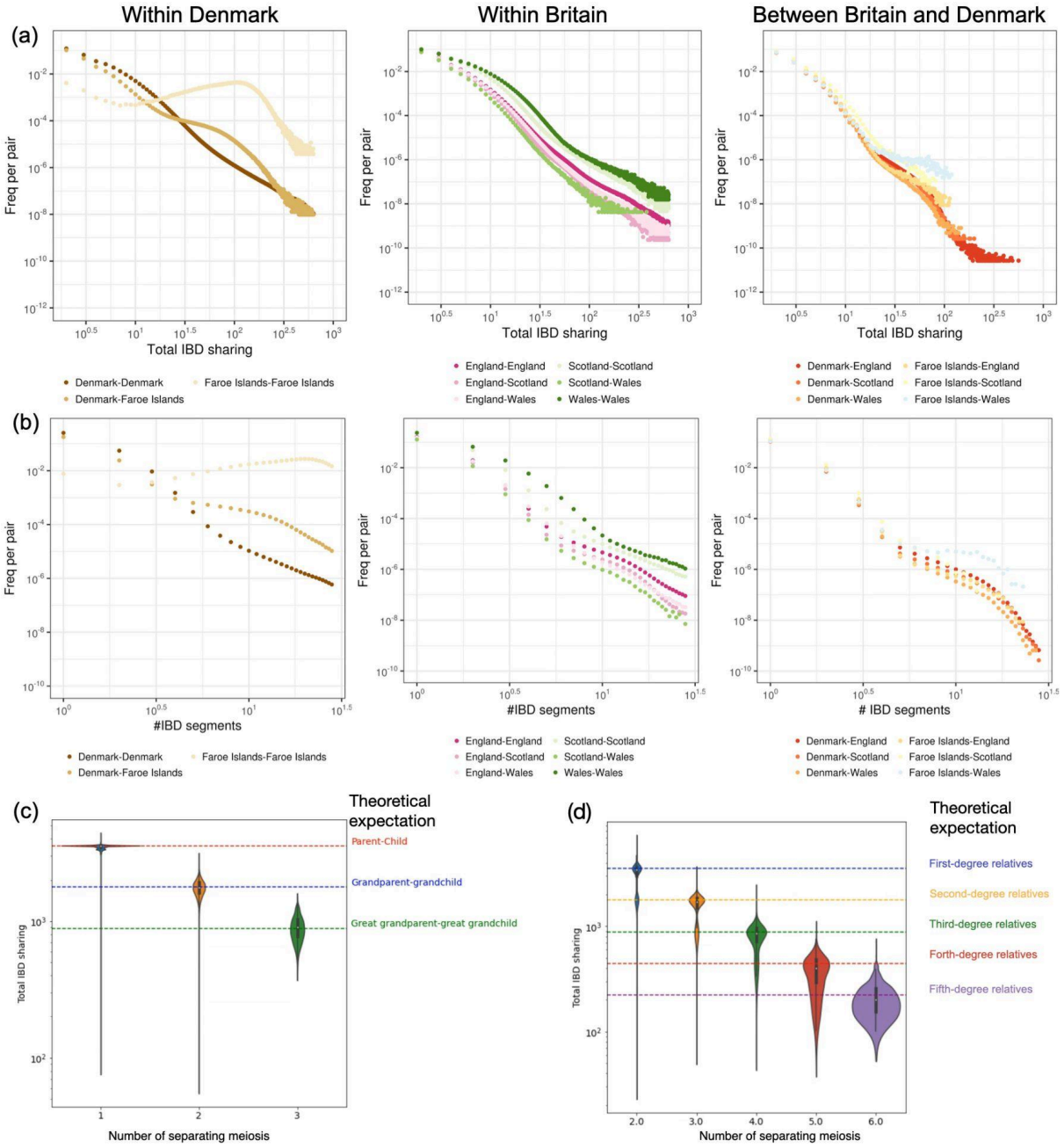

Figure 2. Properties of IBD sharing at population-scale. (a) Frequency distribution of observed total IBD sharing (>2cM) and (b) observed total number of IBD segments (>2cM) within Denmark, within Britain, and between Denmark and Britain. (c-d) Relationship between observed total IBD sharing and expected total IBD sharing: (c) among relatives with direct ancestor-offspring relationship (d) among relatives sharing a common pedigree ancestor, excluding those with direct ancestor-offspring relationship.

Time-resolved fine-scale population structure in Denmark

There have been a number of systematic reconstructions of the population structure and ancestry profile of UK populations (Leslie et al. 2015; Nait Saada et al. 2020; Hu et al. 2025), and our datasets of IBD sharing within the UK are consistent with the previous work (Figure S6-S7). In contrast, our study is the first to use large-scale dataset to look at Danish population history from a genetic perspective. As well as using the entire population sample of individuals with European ancestries, we additionally validated the observed population structure using a refined subset of 28,135 individuals, whom we can identify as having been born in the same provinces as both of their parents, which resulted in no qualitative difference (Figure S19).

We first considered population sharing across the 11 Danish regions at province level. There is substantial intra-regional IBD sharing, and this is most prominent in the rural areas and islands e.g. Bornholm and Læsø (Figure 3a and Figure S8). There is a gradient of decreasing intra-regional sharing from Jutland to Zealand (Figure 3a). Within Jutland, the regions of North and West Jutland demonstrate higher levels of intra-regional IBD sharing compared to South and East Jutland (Figure 3a and Figure S8).

We identified island populations with a high average level of runs of homozygosity (ROHs). Rather than comparing between individuals, ROHs are defined as segments where both parental haplotypes of an individual are identical by descent (>=2cM), thus measuring parental relatedness. These populations, including Sejerø, Fur, Læsø, and Fanø, have average amount of total ROHs and number of ROH segments similar to or higher than the Finnish samples (estimated based on Finnish-born individuals from the DNK cohorts, excluding inferred European ancestry Danes) (Figure S10-S11). We also assessed the ROHs of

individuals with other ancestries (predicted European continental component <90%) in the DNK cohorts. We identified municipalities in the Capital Region especially from Københavns omegn (Copenhagen surroundings) with high levels of mean total ROHs such as Ishøj, Albertslund, Brøndby, Høje-Taastrup,Vallensbæk and Hvidovre (Table S4). These geographically close areas are home to large immigrant populations from countries with high consanguinity such as Pakistan, Nepal, India, Iraq and Turkey ("Statistikbank" 2024; Sahoo et al. 2021). Using data from the 1000 Genomes and Human Genome Diversity Project as reference, we found that individuals with high levels of total ROH (>10cM) in these areas have on average 30%-45% ancestry from Central and South Asia (Figure S13). This is reminiscent of the recent immigration from South Asian communities into the UK with similar high levels of ROHs (Arciero et al. 2021).

When we consider inter-regional IBD sharing, populations from the Jutland Peninsula cluster together and are differentiated from the Islands of Zealand, which is also indicated by the multiple dimensional scaling analysis using regional pairwise IBD sharing (Figure 3a-b). The IBD sharing between Jutland and other areas of Denmark decreases exponentially with their geographic distance (Figure 3c). However, the urban area of København (Byen København and Københavns Omegn) shows higher IBD sharing with geographically distant Jutland than with its own population and the neighboring region of Zealand (Figure 3c), violating the expectation of isolation-by-distance model under limited migration (Ralph and Coop 2013). This supports significant migration from Jutland to København, as discussed below.

The above analysis encompasses all IBD lengths and thus aggregates data across various time periods. We subsequently conducted a time-resolved analysis, estimating population size over time (Figure 3d) using IBDNe (Browning and Browning 2015). Our findings reveal that all

provinces in Denmark, except for København City and its surroundings, exhibit a decreasing growth rate in effective population size (i.e. number of ancestors of current populations) over the past 13 to 20 generations (approximately the end of 14th century to the end of 16th century). Notably, the population of Bornholm experienced a stronger bottleneck with a longer period between 10 to 25 generations ago, which aligns with its high intra-IBD sharing.

In Jutland, the decline of effective population size began significantly earlier, at least 25 generations ago (the lower bound we consider) compared to Zealand. Specifically, North Jutland and West Jutland have experienced the most pronounced decline, reaching their lowest effective population size 13 generations ago (the end of 16th century), at only one third of the size estimated for 25 generations ago. The population decline can also be observed in samples from Aalborg and Aarhus, two of the major cities in Jutland since the Middle Ages (Figure S14). The Black Death in Denmark (1348-1350 CE) occurs just before or at the start of the estimated population decline ("Urban and Rural Survivorship in Pre- and Post-Black Death Denmark" 2021) (21.6 generations ago). Both the Black Death and the associated emigration likely contributed to the reduction in population in Jutland and Zealand, which we discussed below with historical evidence.

Intrigued by this difference, we stratified IBD sharing by segment length to separate out inter-population sharing across different time periods visualised by multidimensional scaling (full time series for each IBD length is shown in Figure S15; the ranges shown in Figure 3e is to illustrate estimated patterns across three different historical periods). We observed elevated sharing between København and Jutland over time, compared to København's sharing with the nearby Region Zealand (Figure 3e). Using København as a focal point, we found a consistent positive correlation between its IBD sharing with other regions and their geographic distance, strongly driven by its high sharing with Jutland, particularly West

Jutland (Figure S16). Other areas of Zealand also exhibit higher IBD sharing with Jutland relative to their geographic distance especially for 2-9cM IBD (Figure S16-S17), corresponding to the late Medieval period through the end of 18th century. This pattern suggests substantial and continuous migration from Jutland to Zealand Island, mostly evident in København, aligning with the significant population decline in Jutland compared to Zealand we have observed in our effective population size analysis. Meanwhile, North Zealand stands out as an outlier in the Capital Region, showing relatively lower genetic exchange with Jutland (visualised in IBD-length stratified multidimensional scaling plots in Figure 3e) and slow population growth following late medieval depopulation (Figure 3c).

Ideally, if we had substantial samples from other European countries, it would help resolve the different possible histories of genetic exchange across Europe with Denmark. In the absence of such data, we constructed proxies of these populations from other European countries by using individuals in the UK and Danish datasets who were born in, and inferred to have substantial ancestry from these countries. This inference was made by using a Leiden community detection approach (Traag, Waltman, and van Eck 2019) based on individual pairwise IBD sharing in our combined UKBB and DNK datasets (UKBB+DNK) (see Methods).

Using this data, we estimated the IBD sharing between the Danish populations and the other neighboring countries (excluding Britain here which we explore in more detail below). We find the pattern of IBD sharing is strongly driven by both the geographic proximity and known migration history (Figure 4). Norway and Sweden have the overall highest IBD sharing with Denmark (Figure 4). Norway is most closely related to the North Jutlandic population while Sweden and Finland are most related to the eastern side of Zealand, especially North Zealand, and Bornholm. This is in contrast with Germany and the

Netherlands, which are most related to the southern border of Denmark at the Jutland Peninsula. Poland shows the highest IBD sharing with the Lolland-Falster islands in the south of Zealand. For Bornholm, compared to Jutland, it has substantially higher sharing with other Baltic countries, including Sweden, Finland and Poland (Figure S21). The results are qualitatively the same if we only use samples born in other countries from either of the cohorts (Figure S22).

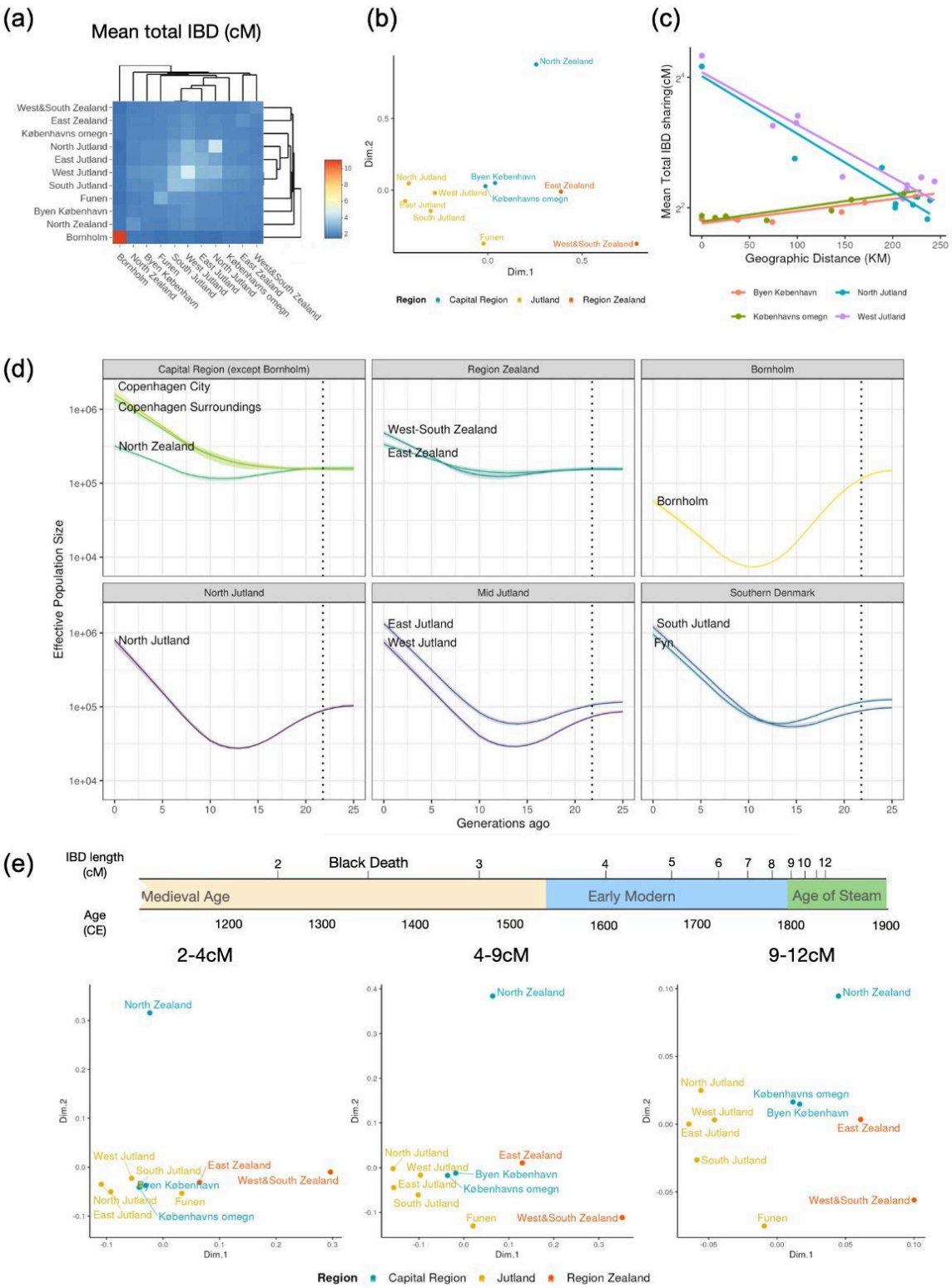

Figure 3. Haplotype-sharing within the Danish population. (a) Pairwise total IBD sharing between provinces of Denmark. (b)Multidimensional scaling on the dissimilarity between provinces of Denmark (except Bornholm) transformed from pairwise total IBD sharing. (c)

The relationship between geographic distance (X-axis) from a focal province to all the other provinces (each dot represents one) and their mean total IBD sharing (Y-axis) (except Bornholm). The linear line indicates the fitted linear regression model. We selected Byen København, Københavns omegn, West Jutland and North Jutland as focal provinces here. (d) Effective population size estimate in regions of Denmark across time (up to 25 generations ago). The dotted line indicates the estimated generation age of the Black Death (21.6 generations ago) based on a generation time of 28 years. (e)Multidimensional scaling on the dissimilarity between provinces of Denmark (except Bornholm) stratified by different ranges of IBD lengths.

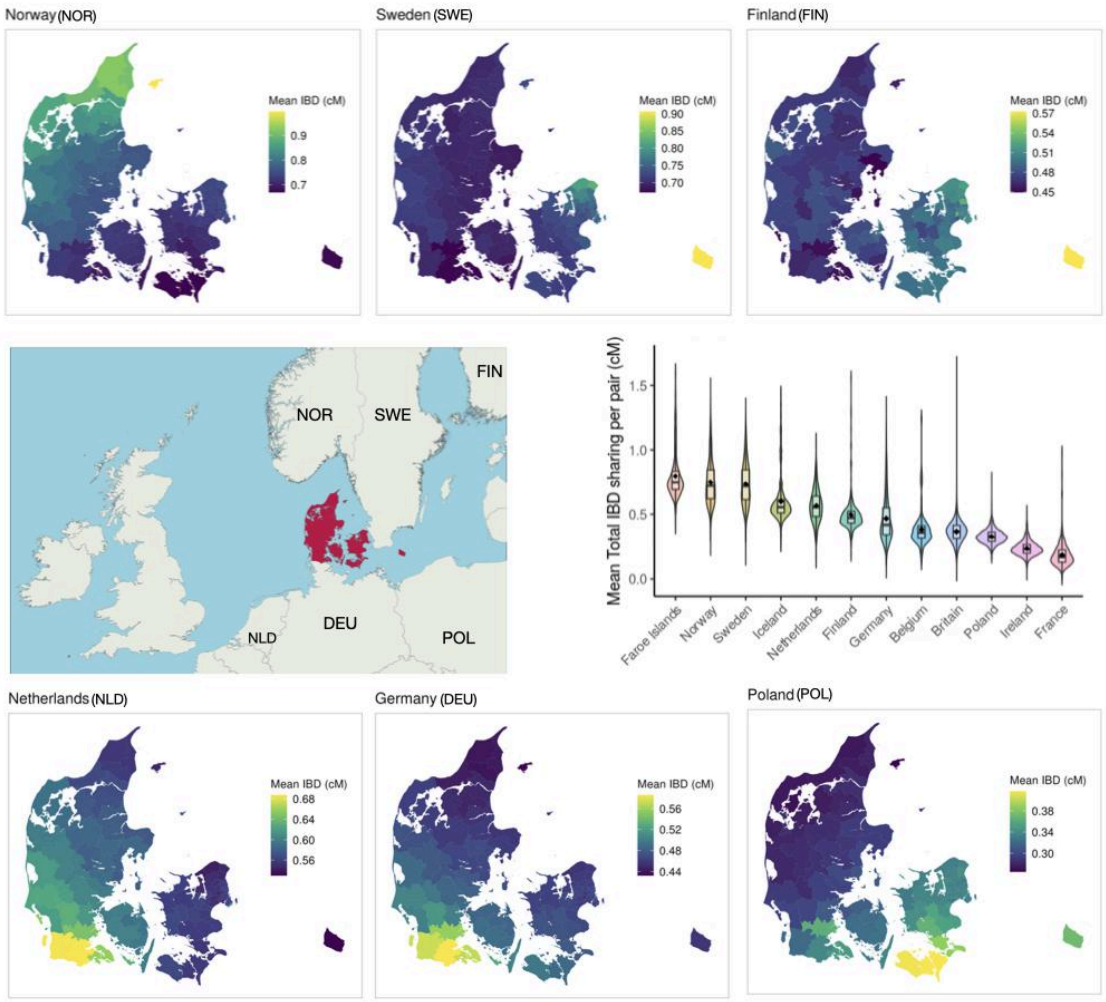

Figure 4. Mean total IBD sharing between Denmark and neighboring European countries. To capture heterogeneity among samples, we calculated the mean total IBD sharing between each sample from neighboring European populations and all samples from Denmark (i.e. one vs all comparisons). The violin plot displays the distribution of average total IBD sharing across individuals (one vs all) in each compared population. The embedded box plot indicates the median, lower, and upper quartiles, while the dot represents the mean total IBD sharing for each European population with Denmark. Norway, Sweden, Finland, the Netherlands, Germany, Poland are highlighted to show their sharing with different regions in Denmark. The color scale of each highlight panel is normalised independently.

Haplotype-sharing across the North Sea

A unique aspect of our study is the ability to look at two substantial regions of Europe which have documented trading and historical links and have accessible genetic resources for research. We analysed 180.8 billion pairs of individuals between Britain and Denmark using the intersection of the two genotyping arrays (as outlined above and in methods). Between Britain and Denmark, 9.5% pairs (17.2 billion) share at least one IBD segment larger than 2cM, indicating the sharing of at least one common ancestor most likely within the past 25 generations. We find that the high IBD sharing with Britain is significantly clustered at the west coast of Jutland populations (Figure 5, Figure S24, local Morans' test *P*-value <0.05; for hotspots/statistically significant clusters, where high values are surrounded by other high values, defined with *P*-value <0.05), in particular the southernmost part of South Jutland - Tønder and Aabenraa municipalities (Figure 4b, local Moran's test *P*-value=0.004). The geographic distribution of total IBD sharing in Denmark is similar across the three British subpopulations English, Welsh and Scottish (Figure S23). Within Britain, the population from the East Midlands (including Derbyshire, Nottinghamshire, Leicestershire, Lincolnshire) and South Yorkshire are the hotspot of IBD sharing with Denmark (Figure 5, Figure S25, local Moran's test *P*-value<0.05), in particular South and West Derbyshire with the strongest positive spatial autocorrelation (local Moran's test *P*-value=0.003).

Similarly to the intra Danish analysis, we then split the analysis into different IBD lengths, corresponding to different time periods (full time series is available in Figure S26-S27; Note that the chosen 3 time windows in Figure 5 are to illustrate the major patterns we see in this analysis and are different from the intra-Danish migration windows.). We see a striking difference in IBD sharing between Britain and Denmark across these three periods. Short IBD segments (defined as 2-4cM, most likely present in distant ancestors estimated as 12.5 to

25 generations ago, dating from 13th century to late 16th century) show the strongest IBD sharing between the East Midlands and Yorkshire in the England, and South Jutland and West Jutland in Denmark (Figure 5c, Figure S24-S25, local Moran's test *P*-value<0.05). From the time period between 7.1-12.5 generations ago (start of 17th century to mid 18th century) corresponding to IBD between 4-7cM, we also observe an elevated sharing between South Jutland and the east coastal area of England including East of England and South East, in particular Norfolk (Figure S25, local Moran's test *P*-value<0.05; Figure S28). In contrast, the long IBD segments (estimated 4.2 to 7.1 generations ago, mid 18 century to mid 19 century) show the highest sharing between London and København (Figure 5c; Figure S25, local Moran's test *P*-value<0.05). Using a fine-grained geographic resolution, we find the capital cities also show accelerated long IBD sharing relative to their short IBD sharing patterns with other metropolitan cities across the North Sea, i.e. London with Aabenraa, Odense and Aarhus, København with Greater Manchester and Glasgow (Figure 5d and Figure 5e). For the Jutland cities Aabenraa, Odense and Aarhus, apart from the link with London, we also observe elevated mid to long IBD sharing between the east coastline of England including Norfolk (East of Anglia) and South East (Figure S30).

We also estimated the IBD sharing between Britain and other neighboring European countries, similar to the Danish analysis (Figure 6). At country-level, we find Ireland has the overall highest sharing with Britain, which is expected given the entangled history of Ireland and Britain and short travel distances. Western Scotland, particularly Glasgow and the Highlands and Islands, and Liverpool from the North West of England have the most IBD sharing with Ireland. Glasgow and Liverpool have a marked increase in early modern and recent generations sharing with Ireland (4.2-12.5 generations ago, Figure S42-S43). Norway, Iceland and the Faroe Islands show a similar geographic distribution of IBD sharing with

Britain, in which the Western Isles, Orkney and Shetland in Scotland show the highest overall IBD sharing across all IBD lengths (Figure S35-S37). This reflects the regions that were under significant Norwegian influence during the medieval time(Margaryan et al. 2020; Als et al. 2006; Jorgensen et al. 2004; Helgason et al. 2000; Ebenesersdóttir et al. 2018).

On the eastern and southern side of the North Sea, Britain has the highest sharing with the Netherlands (Figure 6). Derbyshire shows high IBD sharing with the Netherlands, Belgium and Germany for 2-4cM IBD (Figure S38-S40), as well Norway with England, following a geographic dispersal pattern at this IBD length stratification similar to that of Denmark (Figure S28). However, in contrast with Denmark, the Netherlands also has significant high IBD sharing with Kent in the South East of England consistently throughout time. Long IBD sharing with the Netherlands and Germany clearly shifts towards London and the surrounding areas along the South East coast of England (Figure S43). This aligns with what we have observed with Denmark. There are surprisingly low levels of average IBD sharing with France relative to other countries given its geographic location and entangled history. From the distribution of the mean IBD sharing between each French sample and all British samples (one vs all), we observed a long tail on the higher end, indicating that there could be subpopulations with higher sharing (Figure 6). Overall, samples from France predominantly show the high IBD sharing with London and other parts of South East England (Figure S41), with an increasing degree of sharing with London in recent generations (Figure S43). This signal is robust to whether the source of proxy individuals for France come from either UKBB or the Danish cohorts, and whether we use Leiden clustering of ancestry or not (Figure S33).

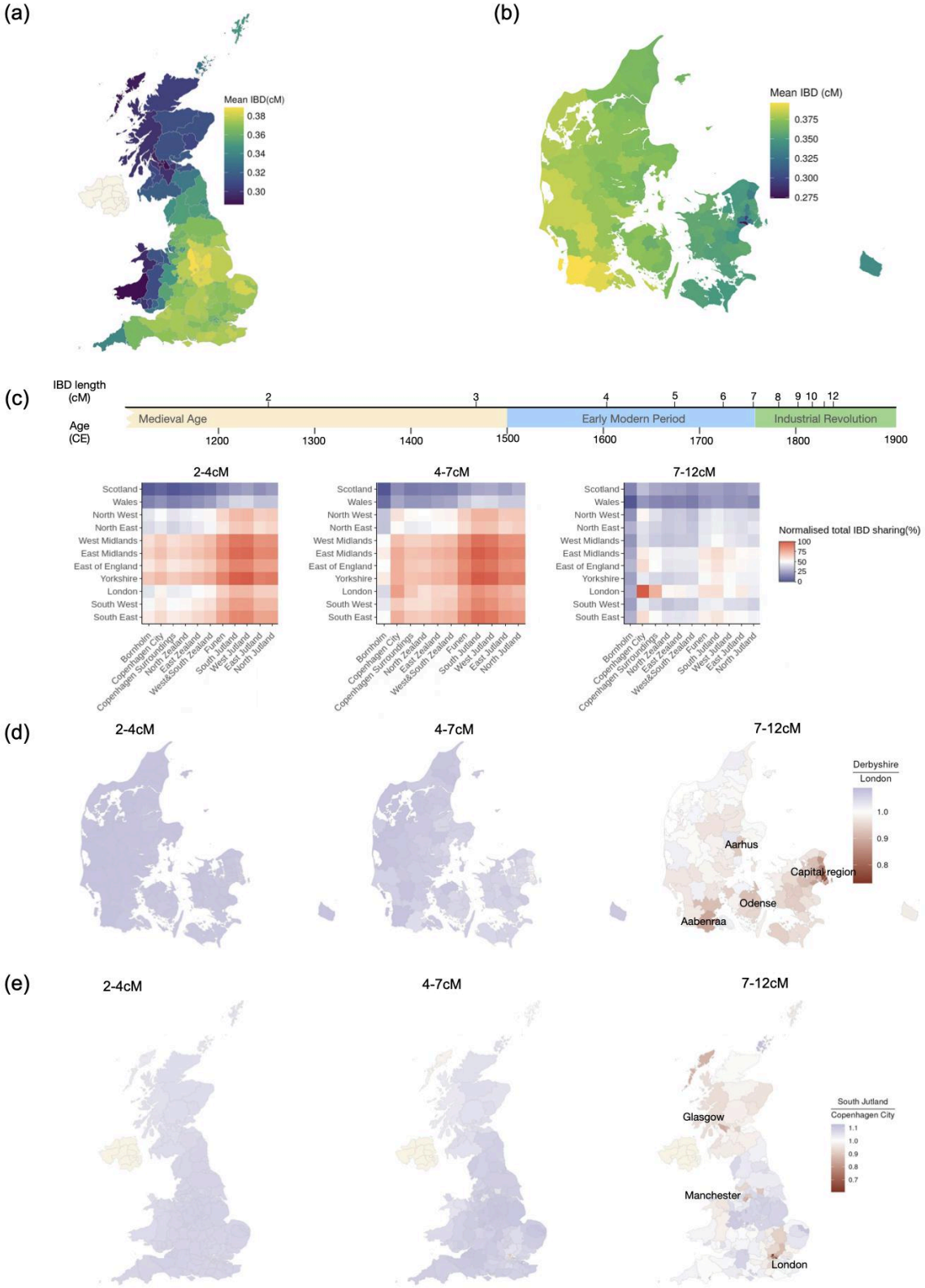

Figure 5. IBD sharing across British and Danish populations. (a).The geographic distribution of mean total IBD sharing between Denmark and regions of Britain (at NUTS3

level). (b) The geographic distribution of mean total IBD sharing between Britain and regions of Denmark (at municipality level). (c) .The IBD sharing across British and Danish regions across time periods through stratification using the length of IBD segments. Each heatmap is normalised independently (scaled the range to be between 0 and 1). (d).The contrast, expressed as a ratio of sharing, of the geographic distribution of IBD sharing of Danish population with Derbyshire and London, stratified by IBD length. (e).The contrast of the geographical distribution of IBD sharing of British population between South Jutland and København, stratified by IBD length.

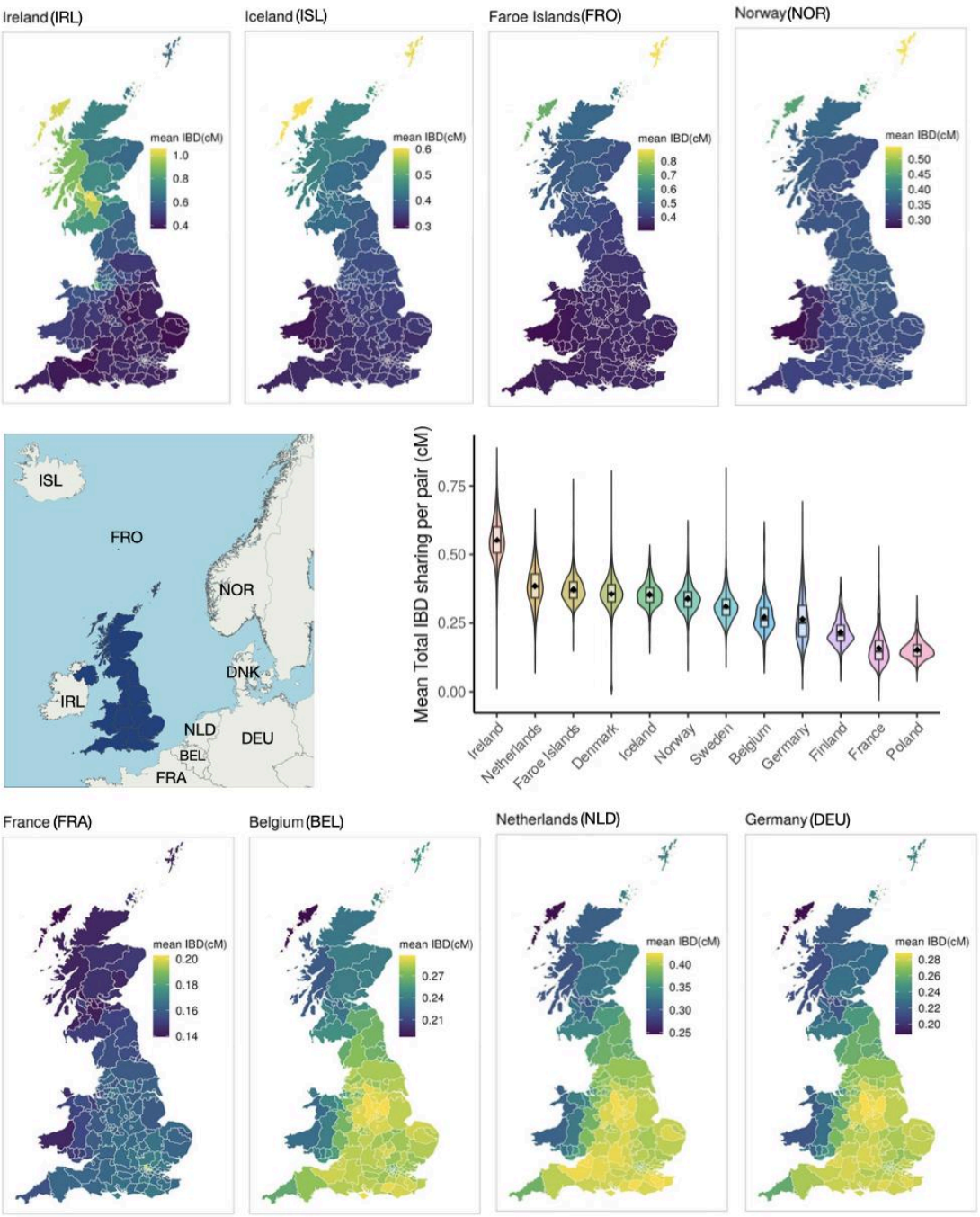

Figure 6. Mean total IBD sharing between British and neighbouring European populations. Here we calculated the mean total IBD sharing between each sample from neighboring European populations and all samples from Britain (i.e. one vs all comparisons). The violin plot displays the distribution of average total IBD sharing across individuals (one vs all) in each compared population. The embedded box plot indicates the median, lower, and upper quartiles, while the dot represents the mean total IBD sharing for each European

population with Britain. Ireland, Iceland, the Faroe Islands, Norway, France, Belgium, the Netherlands, and Germany are highlighted to show their sharing with different regions in Britain. Each highlighted panel is independently normalised.

Comparison to historical evidence and synthesis

The fundamental signal underlying our IBD analysis is shared ancestry between two individuals in our cohorts. When two individuals born nearby share IBD segments, the simplest hypothesis is that their common ancestors were also likely distributed in the same locale, assuming that migration primarily occurred locally. However, alternative hypotheses, such as mass migration events, are also plausible. When two individuals born in geographically distant locations share IBD, multiple hypotheses can explain this sharing. These include individual migration from one location to another (either voluntary or forced), the migration of a related group from a third location to both distant locations, or temporary cohabitation in shared locations where offspring were produced without long-term settlement. All these scenarios and their combinations over time likely contributed to the patterns we observed; the IBD data reflects the average sum of all these genetic exchanges. Comparisons to known or suspected historical events are complex, as these often rely on a limited set of high-profile individuals and proxy data of disputed relevance. Below, we compare IBD data with historical data (archival and archaeological, abbreviated HIS below).

The IBD data suggest a significant demographic event occurred in medieval Denmark, altering both population size and migration patterns within the country.
IBD-based effective population size estimates indicate a higher population decline in Jutland compared to the Zealand Island 13-20 generations ago. A credible event (HIS) corresponding

to this data is the Black Death in 1349-50, and later instances of pestilence continuing through the next century (Hybel and Poulsen 2007). While birth and death records from this period are absent, historians have identified a late medieval decline in the number and density of churches in Jutland. A total of 80 churches were abandoned in the two centuries before the Protestant Reformation in 1536, 75 of these in Jutland and 5 on the islands of Zealand, Funen, Lolland and Møn (Christensen 1938). In 1200, 63.3% of all Danish churches were located in Jutland and by 1500 the Jutland share was down to 50.0% (H. Nielsen 2010). Data for clerical deaths in the two major cities, Ribe in Jutland and Roskilde in Zealand, indicate a higher mortality in the west than in the east (H. Nielsen 2010)). This analysis of IBD data supports a hypothesis that Jutland was hit much harder by the plague and its demographic consequences (which will be discussed below) than the rest of Denmark.

The genetic findings indicate potential migration flows within Denmark pre-1660. Modern Zealanders have distant genetic ancestors across Denmark, whereas the Jutland population is relatively more homogeneous with most ancestors being locally dispersed. HIS evidence shows that degrees of servitude, enforced as a result of the loss of labour after the Black Death, varied considerably. In 1446 at the latest, servitude (vorned) was enforced on adult males in rural areas of Zealand which tied males to the parish (women could move freely). In Jutland there was no similar restriction of male mobility and males are known to have migrated to Zealand in search of better opportunities and better farms. Farmhands from Jutland ('jydekarle') were everywhere (Olsen 1950-2, 203). This may explain why København has higher IBD sharing with Jutland than the nearby regions in Zealand. In the 1699-1705 register of citizens, 19.9% people in København stated their birth place as Jutland and 13.5% originated from Zealand (Olsen 1932, 95).

The IBD and HIS data underline what appears to be a puzzle: it would seem logical to expect that Jutland, which had suffered a severe population decline due to the Black Death, would have experienced considerable immigration to populate deserted farms. Both historical data and IBD data support a large emigration of people post the black death, suggesting that there was a strong motivation for the survivors of the black death to move. This observation raises questions that cannot be explored in this context. One hypothesis is that the population decline may have been followed by changes in agricultural systems that required less labour. For instance, the rise of the north and west Jutland cattle trade, which became a major export trade in the fifteenth century (Enemark, 2003), led to the abandonment of large areas of intensive agriculture in favour of a more extensive pastoral system (Gissel 1981). Pastures produced at most a third of the food of grain fields (Porsmose 1988), and as much of the cattle were exported, Jutland may have experienced food shortages that drove continued emigration. Alternatively, the coastal areas of western and northern Jutland, where we have observed the strongest population decline, have long been characterized by infertile, sandy soils and have periodically suffered from sand drifts between the 6th and 19th centuries, making traditional farming challenging and less productive (Gravesen 2001; Strunge et al., n.d.; Binderup, n.d.; N. K. Nielsen 2023). Another contributing factor may have been repeated military marching and ravaging along the Army Road (Hærvej) in central Jutland, which became severely depopulated (Teitsson 1981). Finally, we emphasize that IBD-based effective population size estimates do not account for migration, potentially underestimating Jutland's historical population size.

The islands around Denmark have varied histories, reflecting different degrees of dependence on agriculture, fishing and international shipping. The islands of Sejerø and Fur relied on a mix of agriculture and local fishing, while the islands of Læsø and Fanø have a strong

maritime past. It might therefore be assumed that the strong international orientation of shipping islands would show up in the genetic makeup. Interestingly our IBD sharing data consistently has the majority of the island communities having high amounts of intra-IBD sharing and ROHs, consistently with a small effective population size. Common to all four islands is the fact that the mean total ROHs is five or more times higher than the national mean total ROHs. This finding suggests that the cultural and trading differences between these islands did not impact the familial structures on the islands. Pontoppidan (1768, 768) and Kyrre (1912, 30) report the aversion of people on Fur to immigrants from the mainland. On the island of Læsø, there is historical evidence that the surplus of females caused by high male mortality at sea was compensated by migrants from Jutland who were held in low esteem (Stoklund 2018, 241-245) but this does not seem to have substantially changed the ROH levels compared to Sejerø or Fur (agricultural) islands.

North Zealand is a consistent outlier with relatively low genetic exchange with other regions of Denmark in our multi-dimensional analysis of IBD sharing, which otherwise recapitulates much of Danish geography (with the exception of København, the capital city and its surrounding area, where immigration seems to be from all other regions). North Zealand was unique in several ways: it was densely forested and depopulated to keep large forests and pastures for the king's hunt; soil productivity was at the low midwest Jutland standard (<10 barrels of hartkorn per km^2) (Dam & Jakobsen 2008, 106), and Helsingør (Elsinore) had a large immigrant population as this was the entry point to the Baltic and the place all ships called to pay the Sound Toll (the main trade route between the North Sea and the Baltic sea, bordering Denmark and Sweden). Between 1556 and 1575, 50-60 (10-13%) of male citizens of Elsinore were of Scottish or English descent (Tønnesen 1985, 77). In later years Dutch and German immigrants became more prominent but Elsinore retained a strong English-Scottish

minority which was reinforced by a new immigration of English merchants around 1700. In 1801 the British colony in the city amounted to 100 people, out of a total population of 5,063 inhabitants. The British component stood out with their wealth and the city was known as 'Little England' (Appel 2012). We attempted to investigate these specific hypotheses related to immigration from Britain but did not identify particular signals driven by exchange with North Zealand. Helsingør also did not show increased IBD sharing compared to other municipalities in North Zealand for long and mid range IBD sharing.

The immigrant flow from neighbouring countries (Figure 4) follows predictable patterns. North Zealand has the higher IBD sharing with Swedish and Finnish relative to other parts of Zealand, which may also contribute to it being an outlier in Zealand in IBD sharing patterns. North Jutland has a distinct Norwegian immigration which is well-documented by eighteenth-century records of individual movements and the 1850 census. This flow of people was strongly related to a grain-for-timber trade which was a major regional activity between c.1600 and 1870 (Holm 1991, 155-157). Conversely, cultural flows and potential migration from Jutland to southern Norway during the 17th century are also documented (Engelstad 1946). The Dutch and German immigration related to a continuous flow from medieval to modern times, originally related to shipping and whaling and in the 20th century to agricultural investment (Guldberg 2020). Until 1920, South Jutland was part of the wider Schleswig-Holstein region of mixed Danish-German populations. Lübeck, located in this region, and its nearby city Hamburg have been important trading centers in Northern Europe connecting the North Sea and the Baltic Sea since the 13th century. It is highly expected that people from Northern Germany largely share their ancestry with Danes from South Jutland although we lack exact birthplace location for people born in other European countries to fully corroborate. Poland has the highest IBD sharing with the Lolland and Falster island

populations in Denmark, likely explained by the history of Polish seasonal workers who settled there to work in sugar beet fields ((“Migration Museum of Denmark” 2024). Bornholm, an island in the Baltic Sea, shows higher ancestry sharing with other Baltic countries, including Sweden, Finland, and Poland in comparison with Jutland. Until 1658, Bornholm was part of the region of Scania that was conquered by Sweden and it is likely that the population will share IBD with southern Sweden. As well as these documented migrations and trade, there is likely a set of earlier migrations from closely related countries giving rise to short IBD segments.

The genetic data reveal clear changes in migratory flows across the North Sea. We observe a striking shift in Danes-British IBD sharing over time. During the medieval period the pair of regions with greatest IBD sharing was South Jutland and Derbyshire, Nottinghamshire, Leicestershire, Lincolnshire and South Yorkshire, while around 1800 the greatest sharing was between London and København as well as Aabenraa, Aarhus, Odense; in addition, København also had a notable IBD sharing with Manchester and Glasgow compared to Derbyshire. In the early modern period we observe an increase in sharing between South Jutland and North Norfolk.

The existing HIS evidence only allows an assessment of migration into and from Denmark post-1660 when Denmark was a protectionist economy under an absolute monarch with considerable control of migration. In a review of the evidence of international migration after 1645, Johansen (Johansen 2002) concluded: “The total number of such immigrants was probably only a few thousand over a long period, and by 1775 the foreign element in the Danish population was, no doubt, very modest.” Migration in the 18th century involved mostly Norway, Germany, and the Netherlands. For the medieval period, when there were

less tight controls on migration and the economy was essentially open, lists of burghers (urban residents in towns) document a substantial component of German and Dutch citizens in Danish cities, while Scottish and English immigrants are documented mainly in the town of Helsingør/Elsinore as mentioned above. In the Viking and early medieval period, flows between Denmark and England in both directions are well-known in the historical record but there is no consensus on quantitative estimates (McLeod 2014).

There is a higher level of IBD sharing between Denmark and Britain in the Derbyshire, Leicestershire, Nottinghamshire, Lincolnshire and South Yorkshire area, and this is predominantly due to older IBD segments (2-3cM) in our dataset. In addition, it is interesting that the less powered Dutch, Belgian and German analysis also show higher old IBD sharing in the same area, pointing to potentially an old shared population structure across these countries. Although the point estimate of the number of generations for 2cM is 25, corresponding to around first half of 13th century, probabilistically, genetic ancestors older than the most likely time can also contribute to an IBD segment of this length with high chance (we estimated it to be at least 0.74 shown in Methods). There is around 50% probability that the posterior estimate of time is older than 40 generations (mid 9th century), encompassing the time of Danish Viking occupation. This indicates that the well established Viking migration in the 8th-11th century that led to the Viking occupation and control of parts of England ("Danelaw") could explain this signal, coupled with Norwegian Viking migrations into the Hebrides, Orkney and Shetland in Scotland. Equally important, we identified South Jutland in Denmark as having the highest IBD sharing with English, which is geographically close to Hedeby (Schleswig-Holstein, now in Germany) and Ribe, Viking Age Danish towns. As two of the most important trading centres in the Viking age, Hedeby

and Ribe have been found to have substantial archaeological links with the Danelaw (J. Kershaw and Røyrvik 2016; J. F. Kershaw 2020).

However, the historically documented Danelaw region extends all of eastern England between River Tees and Thames with York as the Capital (Leslie et al. 2015; Hadley 2006), and these have less IBD sharing with Denmark than Derbyshire and this central region. There could be multiple explanations for the geographic concentration we observe. Firstly, the East Midlands and nearby Yorkshire might have been settled more extensively during the Viking migrations. Secondly, these regions might have had continuing or new demographic events such as trade and population movement lasting beyond the Viking periods. Anglo-Danish trade continued despite the reorientation of England after the Norman conquest (Sawyer 1986). It is known that Danish merchants were allowed to stay for a year and visit inland fairs, and grants of safe-conduct are known from Newcastle, Lynn, Yarmouth, and London through the 12th and 13th centuries (Childs 1998). The toll records of shipping through the Sound document that the Baltic trade had strong participation of skippers originating in Norfolk (East Anglia, England), Firth of Forth regions (Edinburgh, Scotland), Hull (Yorkshire, England) and London (Veluwenkamp, Scheltjens, and Van der Woude 2021). This may explain Norfolk's increased IBD sharing with Denmark as high as Derbyshire for mid-range IBD (4-7cM, late 16th century to mid 18th century). Finally other population movements in Britain after the Viking periods may have diluted the signal in other regions more extensively, especially urban areas compared to the largely rural western and southern Derbyshire("Rural Urban District" 2019). Thus the high IBD sharing between the hotspots could be a residual signal after subsequent migration elsewhere. These explanations are not mutually exclusive - some mixture of all of these might have occurred.

The clear switch to IBD links between the industrial cities of Britain and maritime centres of Denmark around 1800 was probably related to well-documented increased shipping between the two countries (Møller 1988). Potentially, the increased ease of transport followed by movements of workers or trade based movements can lead to higher chance of genetic exchange between cities, though it is notable that Bristol and Newcastle, both well ascertained due to UK BioBank recruitment centres there, do not have the same exchange as Manchester, Glasgow, or London with København (Figure S31). The link to the eastern island towns of København, Aabenraa and Odense is consistent with strong migratory links between the urban region of Southern Denmark and the capital around 1800 (Brock and van Lottum 2019). It is important to stress that this pattern is also consistent with migration from a third location to both the British industrial cities and the Danish industrial cities.

We also observe that Dutch, German and French have clearly increased IBD sharing with London in early modern times and the Industrial Revolution (Figure S42-S43). Though we are limited to consider heterogeneity within these countries and to pinpoint the particular regions driving these links, based on our earlier findings in Denmark, trading with the coastal and urban populations - particularly from regions with trading links to London - could be one of the factors contributing to the patterns. For example, Hamburg, Lübeck, Bruges and port cities in the Netherlands (e.g. Dordrecht) were major trading hubs of the Hanseatic League with connections to London. Compared with the continental Europe nearby, the comparative peace and economic development of England in late medieval times has attracted immigrants from Europe for trade and work. The historical tax records of 64,000 foreign residents in England between 14th-16th century ("England's Immigrants 1330 – 1550," n.d.) shows that these immigrants were mainly servants, merchants, and craft people. While such immigration may have occurred continuously in a small scale until the industrial revolution, there were

also historical mass migrations that could have contributed to the observed IBD sharing signals: Norman-French immigrants replacing most of old English landowners following the Norman Conquest in 11th century and settled mainly in south and east side of England and the majority of 40,000–50,000 French protestants ("Huguenots") migrating to London around 1670 CE.

Discussion

In this study, we performed an extensive inherited by descent (IBD) mapping exercise between two modern populations, Denmark and Britain, that have experienced complex patterns of genetic exchange, both internally and between them. To achieve this, we conducted 369 billion individual pairwise comparisons of genomic sharing, and developed space- and time-efficient methods to store and analyse the results. Using the large and accurate Danish pedigree information, we showed that the amount of IBD sharing we observe closely aligns with theoretical expectations. This demonstrates both that human populations follow classic diploid Mendelian inheritance patterns, as expected and that the calibration of the genetic map (International HapMap 3 Consortium et al. 2010) is consistent with the IBD patterns observed in the population of Denmark. Using the large-scale intra- and inter-genetic exchanges observed between Britain and Denmark, we discovered aspects of Denmark's population genetics history and distinct gene flows between Britain and Denmark from the Middle Ages through to the Industrial Revolution. Further analysis of IBD sharing between Denmark and Britain and other neighbouring countries revealed both shared and unique migration patterns. These findings, coupled with historical evidence, provides novel support of geographical hotspots of Danish Viking genetic legacy in Britain. Our study also

discovered novel genetic evidence of the post-Viking Age migration patterns across the North Sea, likely driven by continuous trade and population movements.

The historical comparison has truly been a cross-disciplinary process of iterating between a genetic and historical perspective. An example of the consistency in information is that the genetic evidence supports a major migration from Jutland to Zealand in mediaeval history, whereas an example of a more complex, partly discrepant aspect is the shift in observed high IBD sharing between Britain and Denmark from East Midlands to South Jutland to specific cities between UK and Denmark. The high IBD sharing hotspots we identified at East Midlands and South Yorkshire with South Jutland support the impact of Danelaw, which provides novel evidence based on autosomal markers of modern genomes. This finding aligns with complementary studies that use Y-chromosome markers, as well as archaeological and linguistic evidence(Lall et al. 2021; J. Kershaw and Røyrvik 2016). For the later shift to city to city migration, it is unclear whether this shift is due to the early increase in trade in the 18th century or the later early 19th century industrial revolution. The variation in IBD sharing across different cities in the UK (London, Leeds, Manchester and Glasgow, but not Bristol, Birmingham or Southampton) suggests that there was heterogeneity in the shift of genetic exchange (Figure S31).

The different small islands across Denmark have far more intra IBD sharing compared with the continental part and large islands, resulting in elevated levels of runs of homozygosity, some of which are similar or even higher than those observed in the proxy population of Finland. Although we observe a larger proportion of individuals from these islands with high amounts of ROHs compared to the national average, the small sample size unavoidably leads to wide confidence intervals in estimating the community average. We also observed an

increased amount of ROH in individuals of non-European ancestries in areas near the capital city, consistent with recent migration of individuals from cultures with higher rates of cousin marriage. The expectation, therefore, is that these areas are likely to have higher rates of certain rare diseases, such as the higher infant mortality and morbidity previously observed (Gundlund et al. 2015), highlighting the potential for precise genetic screening approaches. From a research perspective, these populations present an opportunity for studies similar to those conducted in Finland (Kurki et al. 2023), UK (East London Genes and Health) (Finer et al. 2020; Arciero et al. 2021) and the Middle East (Qatar)(Thareja et al. 2021).

There are other limitations to this analysis. The analysis relies on the sampling of people in the current era. The ascertainment of the cohort will often have geographic and volunteer bias, both of which have been shown in UK BioBank(van Alten et al. 2024; Fry et al. 2017). However, the recruitment centres used for UK BioBank are spread across the country, and many places that might have had a strong exchange, eg, Newcastle, are well sampled. The Danish cohort is a combination of a hospital centred outpatient cohort in København, with a natural bias towards old-age populations in the region and the Danish Blood Donor Cohort, which is across Denmark, but biased towards healthy volunteers(Hansen et al. 2019). Although we would have preferred a more even ascertainment in Britain and Denmark, we do not believe that the analysis of IBD sharing patterns present here are significantly skewed by these factors.

One challenge in bridging genetic methods with historical records is that genetic information naturally tracks with generations, whereas historical records with time. Our genetic estimate of the age of most recent common ancestors has several uncertainties (details described in Methods). These include the underlying statistical properties of length of IBD segments given

the age of common ancestor, particularly for short IBD sharing; the accuracy to estimate IBD segment length, which is limited by the density of SNP markers; and our estimate doesn't account for demographic history. Additionally, there is individual variation in intergenerational time. We are also unclear whether the average intergenerational time has varied over the centuries, given that historical records show variations in the average age at marriage ("What Age Did People Marry in the British Past?," n.d.). These uncertainties limit our current ability to achieve precision in the historical timing or generational resolution for the age of common ancestors, with wider confidence intervals for earlier time estimates. Statistically, the short IBD at 2cM could also be inherited from ancestors older than 50 generations ago though with less probability, which raises the question whether Anglo-Saxon migration also contributes to the short IBD pattern we observe. Given that Anglo-Saxons and Danish Vikings could have been from the very closeby area in South Jutland/Northern Germany and both occupied largely overlapping regions of eastern and southern England, it's challenging to distinguish the contributions of these two migration events based solely on the geographic distribution of IBD sharing using modern genomes. Additionally, considering that short IBD segments are widely dispersed due to continuous migration and mixing onwards in these areas, this further complicates the differentiation between the genetic influences of these two events. Therefore, we cannot exclude the possibility that the observed signal may be partially attributed to Anglo-Saxon migration. However, we can be confident both that there is a strong link between the East Midlands and South Yorkshire of Britain and South Denmark/North Germany and that there have been both subtle and substantial changes in IBD sharing from the Middle Ages to the Modern Period between Britain and Denmark. In our studies we have not used ancient DNA sources, partly due to their low and sparse representation, and partly because it is unclear how extensive IBD mapping at this scale could be biased by the phasing of ancient DNA with modern Northern and Western European

panels. An interesting modelling question in the future is how to integrate substantial IBD data from modern populations with the sparser ancient DNA data that have specific known dates.

Inferring the migration direction or origin of any given IBD sharing pattern based on genetic data is another challenge we have not addressed. The direct pairwise IBD sharing data we generated in this study cannot resolve these questions. Future research using genealogical tree reconstruction methods (Gunnarsson et al. 2024; Zhang et al. 2023; Kelleher et al. 2019; Speidel et al. 2025) might offer opportunities to estimate the direction of gene flow from the tree structure. In addition, while we focus on combining both genetic and historical evidence to interpret the underlying likely demographic events, other characteristics of the demographic events (e.g. migration rate) or complex demographic scenarios (e.g. continuous migrations or pulses) might also be investigated through theoretical approaches(Palamara and Pe'er 2013).

In this paper, we brought together currently two of the largest genetic cohorts, the UK BioBank, and the Copenhagen Hospital Biobank & Danish Blood Donor Study in Denmark. It is conceptually feasible though both technically and legally complex to imagine extending this approach across even more cohorts, such as FinnGen (Kurki et al. 2023), All of Us(All of Us Research Program Genomics Investigators 2024), the Estonian BioBank(Leitsalu et al. 2015), the China Kadoorie Biobank(Chen et al. 2011) and the BioBank Japan(Nagai et al. 2017). To achieve this, we would need further innovation on the computational methods to discover and store IBD regions as well as substantial computing resources. Alongside providing a baseline understanding of these cohorts for basic and medical research, this would provide an unprecedented view of all substantial components of genetic exchange

between the studied populations, allowing a comprehensive map of genetic exchange from medieval times to the modern day. This would serve as a highly complementary resource to the documented histories, and most likely, as we show here, both support and challenge aspects of these histories. Indeed it is feasible to consider generating a comprehensive map of all European migrations from the medieval era onward by performing similar analyses across a pan-European modern cohort, and similarly for other continents and cross-continental migrations. From our study, we estimate that an even sampling of around 0.1% of the population (i.e. median sample size/median population size) would provide a resolution at the NUTS3 level of regionality in Europe. In addition, methodological improvements in integrating ancient genomes with modern ones would not only provide more support for specific migration patterns but also probably deepen the time resolution. We call for more interdisciplinary research between geneticists and historians and the resources to be able to unlock this natural historical record inherent in our DNA.

Method

Data Collection and Processing - UKBB

We used 488,376 participants in UK Biobank (UKBB) with SNP array data available. genotyped using the UK Biobank Axiom Array as a sampling of the UK population. There are 783,966 genotype markers in total after lifting over to the GRCh38 genome build. We define individuals with British ancestry as the ones self-reported as “White British” (UKBB Data Field: 21000) and born in the UK (UKBB Data field: 1647). In total, we have 410,266 individual samples with 354,847, 35,912 and 19,507 from England, Scotland and Wales respectively. While we might have included individuals with recent family immigration

history into the UK based on this definition, we assume the proportion is relatively small given we have a very large sample size. To map individuals' birthplaces into geographic regions, we transformed the birthplace coordinates (UKBB Data field: 129 and 130) into geographical areas including UK nations, UK NUTS level 2 (2016 version), UK NUTS level 3 (2016 version) and councils.

Data Collection and Processing - Genetics data in Danish biobanks

For the Danish population, we included 370,259 individuals from Denmark with genotype data available from previous studies conducted under the Copenhagen Hospital Biobank (CHB) (Sørensen et al. 2021) and the Danish Blood Donor Study (DBDS) (Hansen et al. 2019; Erikstrup et al. 2023) including *Genetics of healthy ageing and specific diseases among blood donors – a GWA study under the Danish Blood Donor Study* (NVK-1700407, P-2019-99), *Genetics of cardiovascular disease - a genome-wide association study on repository samples from Copenhagen Hospital Biobank* (NVK-1708829,P-2019-93), and *Genetics of pain and degenerative musculoskeletal diseases - a genome-wide association study on repository samples from Copenhagen Hospital Biobank* (NVK-1803812, P-2019-51). All the samples were genotyped using the Illumina Global Screening Array with a total of 627,546 genotype markers. Reuse of these genetic datasets for demographic and genealogy studies does not require specific scientific ethical approval.

We retrieved individuals' birthplaces through the Danish civil registration record, i.e. Det Centrale Personregister (CPR register). There are 316,512 and 985 individuals born in Denmark ("continental Denmark") and the Faroe Islands respectively. As we were aware of the evolving discussion about the use of genetic data from Greenland during this analysis we chose not to include samples from Greenland in this study.

Since the Danish administrative data does not contain individual ethnicity information, we adopted genomic approaches and inferred individuals' global ancestries proportions (Method - Global ancestry inference using ADMIXTURE below). We projected the best fit unsupervised clustering via ADMIXTURE trained on a reference panel representing diverse world populations across continents onto Danish samples to infer their ancestral proportions at continental levels. There are 93.1% (294,782 out of 316,512) and 94.6% (932 out of 985) individuals born in mainland Denmark and the Faroe Islands respectively inferred to have more than 90% estimated European ancestral component. Unless otherwise stated, throughout the manuscript, we refer to the inferred European samples as European ancestry Danes ("Danes"). These samples might include individuals with recent immigration history in the last few generations from other European countries given that we have large sample size and the resulting IBD sharing has mirrored the geography of Denmark, we think most of the samples have major family ancestry from Denmark. We adopted the nomenclature of geographic divisions including regions, provinces and municipalities provided by Statistics Denmark (URL: https://www.dst.dk/en/Statistik/dokumentation/nomenklaturer/nuts).

To validate the observed population structure in Denmark, we also curated a subset of individuals, who have both of their parents in the DNK cohorts and were born in the same province as both their parents. As the dataset is much smaller, we also took a rigorous approach to remove one individual in close relative pairs (up to third-degree) to avoid family-specific patterns. This results in a refined subset of 28,135 individuals.

Data Collection and Processing - Danish Multi-Generation Family Registry

For the Danish Multi-Generation Family registry (MGR-lite registry), there are a total of 7,771,981 individuals recorded. MGR-Lite is a registry that documents the complete parental relationship for children (i.e. legal guardian) born in 1953 or later and still alive in 1968 to their parents based on both the CPR register and family relationships recorded in church books (https://www.registerforskning.dk/projekter/mgr-lite/) and partially included parental relationship for children born before 1953. Additionally, we excluded relationships where parental ages were unreasonable or where the relationship was recorded as an adoption built on our previous work (Westergaard et al. 2024). In data preprocessing, we removed 866,638 individuals with no first degree relatives' records (who have neither parents nor children). We also removed five individuals recorded with multiple pairs of parents. In total, 6,845,337 individuals are analysed.

Given the raw record in the MGR-lite registry only includes the parent-child relationship, we infer other familial relationships according to the following steps: (1) extract ancestor-offspring relationships. We use an adjacency matrix to record whether individual i is an ancestor to individual j and the value of the element indicates the number of separating meioses; (2) extract cousin relationships; To reduce the amount of computation, we only extracted the cousin relationships for individuals with both familial record and genetic data available. We iterate over each individual as potential ancestors extracted from step (1) and construct the cousinship among all his/her descendants. For each pedigree relative, we infer the number of separating meiosis. In total, we extracted 56,329 pairs of ancestor-offspring

relatives and 54,089 pairs of cousins (Table S2 and Table S3), which we are going to use to validate the genetic data.

We are limited to extract pedigree relationships for relatives of distant degrees larger than five due to the incomplete record of parental relationships for individuals born before 1953 thus the relatives sharing ancestors born before 1953 are only partially identified in our familial registry data.

Note that we only assume individuals are related through a single genealogical path given that we are assessing close relatives. While it's likely that individuals can be related through multiple genealogical paths, especially for individuals relatively distant related, we are unable to identify them comprehensively given our limited information about distant pedigree relatives in the data thus we ignore this scenario in our analysis.

In our validation, we compared the total IBD sharing observed from IBD estimation with expected given the number of separating meiosis extracted from the pedigree relationship shown in Figure 2. The expectation of IBD sharing is computed based on Huff et.al 2011(Huff et al. 2011). We find the distribution of observed total IBD sharing given the number of separating meiosis largely follows the theoretical expectation. For parent-child and grandparent-grandchildren relatives, we also observe long tails on the lower end, suggesting that there could be unbiological related pairs or possibly unrecorded adoptions by relatives. Consequently, we also observe that the respective distributions among pairs of shared ancestors are skewed towards the lower end. For pairs of shared ancestors, a second mode with less density is also observed, indicating that there are pairs sharing only one ancestor, e.g. half-siblings or half cousins.

We also compared the number of IBD segments observed with expected. The results are shown in Figure S2. For ancestor-offspring pedigree relatives and cousins up to first-degree cousins, we find that the observed number of IBD segments is higher than expected. Given that the total genomic sharing observed for these relatives aligns with expectations, we believe this is likely explained by statistical phasing error. The switch error of the statistical phasing method we applied to the Danish cohorts is likely similar to that in UKBB, occurring at a rate of approximately one error per 20 cM (Loh, Palamara, and Price 2016; Naseri et al. 2019). Thus, we expect phasing errors to more significantly affect close relatives (those less than third-degree) since they are expected to share long IBD segments (≥20 cM). Given this threshold for phasing error, we consider it less of a concern for IBD sharing estimation in distant relatives, which we used in the migration analysis.

Genotype data processing

To process the UKBB SNP array data, we conducted sample-level quality control (QC) and SNP-level quality control. For sample-level quality control, we removed individual outliers due to poor genotyping quality based on heterozygosity and missing rates based on standard QC output provided by UKBB (resource 531 URL: https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=531).

For the SNP array data in Danish biobanks, we conducted the SNP-level QC through the following steps: (1)Filter SNPs with >0.1 missing (2)remove duplicate SNPs.

Haplotype Phasing

We used Eagle v2.4.1 (https://alkesgroup.broadinstitute.org/Eagle/) (Loh et al. 2016) to estimate the haplotype phases. The genetic map was downloaded from HapMap

(ftp://ftp.ncbi.nlm.nih.gov/hapmap//recombination/2011-01_phaseII_B37/genetic_map_HapMapII_GRCh37.tar.gz).

Estimating IBD sharing

We called IBD using FastSMC(Nait Saada et al. 2020) to detect IBD shared within the past 50 generations with a minimum IBD segment length of 1cM using parameters "--time 50 –min_m 1.0". We conducted independent IBD calling based on three different SNP marker sets: (1) within the UK, using all SNP markers from UKBB; (2) within Denmark, using all SNP markers from Danish samples (DNK); (3) between the UK and Denmark, using shared SNP markers from the combined UKBB and Danish datasets (UKBB+DNK).

While imputation could potentially improve the IBD mapping between the UK and Danish samples, we considered and then rejected adopting this approach. On one hand, since British and Danish populations are two closely related populations (Hudson Fst = 0.00075), we were concerned at the biases introduced, for example, the need to use predominantly Northern and Western European, with substantial British and Scandinavian haplotypes included in the imputation panels. On the other hand, our sensitivity analyses of intra-Danish comparisons confirm that using intersecting markers, despite underestimating IBD segment counts, show distributions highly aligned with the full marker set (Spearman *rho* between 0.99 and 1) and have no qualitative differences (Figures S3-S4), supporting their use in downstream analyses.

We computed the joint IBD calling between the UKBB and DNK (UKBB+DNK) by using the following procedures:

(1)Phase each biobank independently

(2)Liftover the UKB phased genotype data from Human reference genome build GRCh37 to GRCh38. We discarded the positions that could not be lifted over.

(3)Extract the subsets of SNP markers in both biobanks

(4)Use the above subsets to call IBD

We conducted additional filtering after IBD calling to reduce the false positives caused by low SNP density on the IBD calling in the joint SNP data. We removed IBDs across areas with SNP density lower than 50 SNPs/cM, which would likely lower the true positive rates but increase the positive predictive value.

In the downstream analysis, we focus on IBD segments larger than 2cM given that the IBD 1-2cM shows lower sharing rate compared to 2-3cM using the DNK IBD call set (Figure S5), reflecting the reduced sensitivity of IBD estimation at 1-2cM. Given we have lower SNP density in the joint IBD calling, IBD 1-2cM might be underestimated to a stronger degree.

To summarise the level of IBD sharing among pairs of individuals at population level, we calculate the total IBD sharing (>2cM) and number of IBD segments (>2cM) by summing over all the identity-by-descent segments across autosomes between the two individuals.To handle the computing efficiently, we developed bespoke analysis code in C based on a previous published software IBDkin (Zhou, Browning, and Browning 2020). Unless specified, we removed pairs of close relatives up to third-degree relatedness (equivalent to first cousins, great-grandparent and great grandchildren) using the lower bound of expected total IBD sharing threshold of 634cM for analysing the total IBD sharing(Manichaikul et al. 2010). This shall reduce the impact of ascertainment bias due to close genetic relatedness.

Similarly, total length of ROHs and number of ROH segments were calculated per individual by summing the total IBD sharing and number of IBD segments between two chromosome copies across autosomes. We first evaluated the ROHs based on all the available samples. Given that we want to reduce the effects of family-specific patterns especially in communities of small size (autozygosity is correlated between siblings though the correlation is highly variable(Kirin et al. 2010; Thompson 2013)), we also took a stringent manner to only include 160,957 unrelated samples in the DNK cohorts after randomly removing one individual in each pair of close relatives (up to third-degree relatedness). We find the results are highly comparable in the two evaluations (Figure S10), suggesting the high ROHs are likely driven by small effective population size instead of specific family patterns.

To study IBD sharing for a specific length, IBD segments were grouped into bins of different lengths defined as half-open intervals, where each bin includes the lower bound but excludes the upper bound, denoted as [a,b). For example, an IBD bin between 2-4cM indicates all IBD segments x such that $2 \leq x < 4$ cM.

Estimate the effective population size

We applied the IBDNe(Browning and Browning 2015) to estimate the trajectories of effective population size over time. The following options were used “filtersamples=TRUE gmax=25 mincm=2” to remove close relatives and only use the minimal IBD segment length at 2cM up to 25 generations ago. 95% confidence intervals of Ne were estimated by using the default options.

Dimension reduction of pairwise IBD sharing via multidimensional scaling

We use the pairwise matrix of mean total IBD sharing (cM) as similarity matrix S. We then convert S into distance matrix D by subtracting each element of S from its maximum element value, i.e.$D_{ij} = max(S) - S_{ij}$. Multidimensional scaling was then computed with the input distance matrix D using the function "cmdscale" in the "stats" R library.

Global ancestry inference using ADMIXTURE

We applied unsupervised clustering implemented in ADMIXTURE(Alexander, Novembre, and Lange 2009) to infer the continental ancestry proportions of individuals in the Danish cohorts. To train the clustering model, we used the a pre-curated harmonised dataset (1kGP+HGDP) with 3,262 individuals(Koenig et al. 2024) which combines both the 1000 Genomes Project (1kGP) and Human Genome Diversity Project (HGDP) as training data. It includes African (AFR), Admixed American (AMR), European (EUR), Middle Eastern (MID), CSA (Central/South Asia), and East Asian (EAS). In the training, the genotype data for each sample individual is modeled as a mixture of *K* ancestral components. We first removed the SNP markers with strong LD in the DNK cohorts by applying PLINK with option "--indep-pairwise 50 10 0.5" (pruning variants within 50kb with a squared correlation greater than 0.5). We used the subset of autosomal genotype markers (n=385,272) from the 1kGP+HGDP dataset that overlap with the LD-pruned DNK cohorts. To estimate the best number of ancestral components *K* in the data, we applied a cross-validation to compare the predictive accuracy on hold-out test data using a range of values from *K*=2 to *K*=8. *K*=5 is chosen given that the slope of the cross-validation error curve is not changing drastically with the increasing *K* (Figure S12). Each ancestral component is annotated based on their

proportions in each continental sample. The ancestral proportions of the Danish samples were predicted by projecting them onto the learned allele frequencies from the training samples.

Identify populations of other major European ancestry in the UKBB+DNK dataset

To investigate the influence of other European countries on Danish and British ancestry, we assembled a reference panel for European populations using the combined samples (UKBB+DNK) from the UK Biobank (UKBB) and Danish cohorts (DNK). Here, we outline the steps involved in our methods:

Sample Selection

To include samples with non-British and non-Danish European ancestry in the UKBB, we selected samples born in Europe (excluding the UK and Denmark) (UKBB Data Field: 20115) who self-reported their ethnicity as “White” or “Any other white background” (UKBB Data Field: 21000). For individuals with Irish ancestry, we specifically selected those born in Ireland and self-reported with “Irish” ethnicity.

In the Danish cohorts, we included individuals with predominantly predicted European (>90%) ancestry who were born outside Denmark and the UK. We exclude any country with less than 100 samples to avoid the bias of small sample sizes, which might have excluded samples from some East European and Southern European countries due to their low ascertainment in our datasets.

To make sure the clustering results would not be driven by the fine-structure of populations of British and Danes given their large sample size in total, we downsampled these groups.

For the British, we randomly selected 9,000 British individuals from the UKBB, with 3,000 from each of England, Scotland, and Wales. For the Danish cohorts, we also randomly selected 3,000 individuals from a subset where both the child and parents were born in the same municipality.

In total, we included 34,555 individuals with 22,555 ones born in 26 European countries (excluding the UK and Denmark) and 12,000 ones born in the UK and Denmark to construct a reference panel of European populations.

Community detection using individuals' pairwise IBD sharing network

Given that we only have information on individuals' countries of birth, it's important to note that these may not fully indicate their ancestral backgrounds. They could be British or Danes born overseas. To identify other European populations, we assume individuals in the same population would highly share their ancestry with each other compared with those in other populations. To identify groups of individuals with shared ancestry, we used an unsupervised community detection approach, assuming that individuals with the similar ancestry background would be closely clustered in the IBD-sharing network. Thus the resulting detected clusters could allow us to identify different populations.

We constructed an IBD-sharing network based on pairwise total IBD sharing among the selected 35K samples. To reduce the impact of recent migration, we applied a more stringent criteria by removing pairs with total IBD sharing greater than 56 cM (the expected total sharing of third-degree cousins), thus removing relatives up to third-degree cousins.

To identify population clusters, we applied the Leiden community detection algorithm(Traag, Waltman, and van Eck 2019), which optimises the modularity to partition the network into distinct clusters. We adopted a multi-level hierarchical approach. We detect sub-clusters from the clusters of one level up by varying the resolution parameter from 1 to 2, increasing by 0.1 at each step until sub-clusters are identified.

Using four levels of clustering, we identified 45 clusters in total with 30 mainly composed of individuals with non-Danish or non-British ancestry. The clustering results captured broadly the ancestry diversity within our samples at country level (Figure S20). We annotated each cluster based on the proportions of members' country of birth. Thus individuals in the same cluster were considered to share a major ancestry background from their country of birth. We also made sure there are both comparable proportions of samples from both the UKBB and DNK cohorts to form the same population cluster. For example, German clusters were identified only if German-born samples from both the UKBB and DNK cohorts clustered together. We only considered the clusters with a sample size larger than 200 to ensure well-supported results. We did not identify further sub-clusters within each country as it's outside the scope of our study and we don't have detailed birthplace data to validate.

Reassuringly, the majority of Danish samples (97.1%) and British samples (91.2%) are grouped within their own clusters. The community detection approach also helps us to remove individuals born in one country but placed in a different cluster, suggesting they are likely to have mixed ancestry or recent immigrant history.

As our clustering results depend on both the sample size and underlying ancestry diversity in our samples, the clusters should be interpreted as groups of individuals with shared ancestry rather than distinct populations.

The resulting clusters are used to identify proxy populations with major ancestry from the neighbouring countries of Britain/Denmark used in Figure 4 and Figure 6. The sample size for each proxy population is shown in Table S5.

Estimate the Age of Most Recent Common Ancestors (TMRCA) using IBD segment length

To convert the IBD sharing to the age of most recent common ancestors (TMRCA), we used the length of IBD segments.

Modeling IBD Segment Length as a Poisson Process

The occurrence of recombination breakpoints along a genome follows a Poisson process (ignoring crossover interference and chromosome ends). For two individuals separated by $2g$ meioses (where g is the number of generations since their most recent common ancestor), the length of shared IBD segments corresponds to intervals between recombination breakpoints over these $2g$ meioses. This length follows an exponential distribution with rate $2g$ (number of breakpoints per Morgan), yielding a mean segment length of *1/(2g)* Morgan (Thompson 2013). The likelihood function is:

$$f(l|g) = 2g\, exp\,(-\,2gl).$$

We have considered the following options to estimate g for a given length l as described similarly in Gunnarsson et.al(2024)(Gunnarsson et al. 2024). Some literature has derived that l follows the Erlang-2 distribution, but this is conditional on a given genomic site(Thompson 2013; Palamara et al. 2012; Huff et al. 2011; Browning and Browning 2015), which we don't think applies here.

Maximum Likelihood Estimation (MLE)

Given we observe l in the data, we can derive the maximum likelihood estimate of g, defined as the g maximizing the likelihood function $f(l|g)$. Taking the derivative of the likelihood (1) with respect to g is $2(1 - 2gl)exp\,(-\,2gl)$. Thus $\widehat{g}_{MLE} = \frac{1}{2l}$ (Gunnarsson et al. 2024; Nait Saada et al. 2020).

Maximum a posteriori (MAP)

The posterior probability of g given l could be derived by taking into account the prior information of the demographic history which would affect the colascent probability between individuals. Under the Wright-Fisher model, for a simple demographic model with a constant population size N (for diploid, it's effectively as number of haploids), the probability density that two individuals would share ancestors g generations ago (here g is used as a continuous scale) follows an exponential distribution with rate $1/N$:

$$f(g) \;= \frac{1}{N} exp(-\frac{1}{N} g).$$

We can then derive the posterior density of g given l:

$$f(g|l) = \frac{f(g)\,f(l|g)}{\int_0^\infty f(l,g)dg} = \frac{\frac{2g}{N}exp[-g(\frac{1}{N}+2l)]}{\int_0^\infty f(g)f(l|g)dg}$$

For the denominator part, it can be simplified as $\frac{2}{N(\frac{1}{N}+2l)^2}$.

Thus, the posterior density $f(g|l)$ can be simplified as a Erlang-2 distribution(Gunnarsson et al. 2024):

$$f(g|l) = \frac{\frac{2g}{N}exp[-g(\frac{1}{N}+2l)]}{\frac{2}{N(\frac{1}{N}+2l)^2}} = g(\frac{1}{N} + 2l)^2\, exp[-\, g(\frac{1}{N} + 2l)].$$

Taking its derivative with respect to g, the maximum a posteriori (MAP) estimator is simplified as:

$$\widehat{g}_{MAP} = \frac{1}{\frac{1}{N}+2l}.$$

For human populations in our studies with large *N* (the scale of $10^5 - 10^8$) and the range of IBD length we considered here *(*0.01-1 Morgan*),* the prior $f(g)$ is weak as it's relatively flat compared to the likelihood function. Therefore, the posterior has the similar shape as the likelihood (Figure S45) and $\widehat{g}_{MAP} \approx \widehat{g}_{MLE}$ for large population size *N*.

Impact of Demographic History

The prior is sensitive to demographic assumptions we use. If the ancestral population size for older generations is smaller than that of recent generations, the coalescent probability at older generations could be underestimated compared to that using a model with constant population

size (e.g. using the mean population size over time). This could lead the posterior mode to shift towards a larger g particularly for short segments. In other words, for short segments, they are more likely inherited from older ancestors than if a constant population size were assumed.

However, since effective population size is inherently uncertain, for simplicity, we prioritize using MLE/MAP with a simple prior. We caution that the generation time estimate is merely used as a guidance with wide uncertainties. As we have demonstrated in Figure S45, the density function is more spread out for short segments thus the generation estimate has broader uncertainty for short segments compared to longer segments.

For the minimal segment size we used in the manuscript (2cM), the MLE/MAP estimate is 25 generations. Given the posterior density, we can also derive the cumulative probability that the coalescent time is larger than the 25 generations given segment length=2cM using a reasonable assumed population size at the scale of 10^5:

$$Pr(g > 25 | l = 2cM,\ N = 10^5) \approx 0.74.$$

We showed in Figure S45 that this estimate is quite similar using reasonable population size between $10^3 - 10^5$.

Therefore, there is a high chance that IBD segments of 2cM are inherited from ancestors older than 25 generations ago.

Converting Generations to Calendar Time

We assume a generation time of 28 years (Moorjani et al. 2016). For UKBB samples, the median year of birth is at 1950 CE. For Danish samples, the median year of birth is 1960 CE. Across the two cohorts, we use the median birth year as 1955. We use this point in time as the reference for the "present" relative to our samples.

In summary, there are various sources of uncertainty in our estimate of TMRCA: (1) the population demographic history; (2) its posterior distribution given length of IBD segments is quite wide; (3)estimating the endpoints of IBD segments depends on the density of SNP markers thus affecting the accurate estimate of IBD length; (4)uncertainty in the generation time, with Moorjani et.al (Moorjani et al. 2016) estimating the mean generation interval to be between 26-30 years based on recombination clock, and Wang et.al estimating 24-30 years based on mutation clock (Wang et al. 2023; Moorjani et al. 2016).

Statistical test of spatial autocorrelation

We used Moran test to assess the spatial association of IBD sharing. Global moran' test is firstly conducted to assess whether the IBD sharing is randomly distributed. The local Moran's test is used to assess whether the degree of IBD sharing of a given area is associated with its neighboring area. We define neighbors as the five areas with the shortest geographic distance to the centroid of a given area. For IBD sharing between Britain and Denmark, the local Moran's test was conducted for each area, and the results allow categorization into the following groups (shown in Figure S24-S25):

Significant spatial associations between the tested area and its neighbors are identified using *P*-value threshold 0.05: High-High(hotspots, high IBD sharing with high IBD-sharing neighbors), Low-Low (coldspots, low IBD sharing with low IBD-sharing neighbors),

High-Low (outliers, high IBD sharing with low IBD-sharing neighbors) and Low-High (outliers, low IBD sharing with high IBD-sharing neighbors); No significant spatial association (annotated as “>0.05”) or missing data.

Ethical Approval

The UK Biobank has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). This research was conducted using the UK Biobank Resource under projects 49978.

All investigations were conducted in accordance with the tenets of the Declaration of Helsinki. The genotype datasets from the Copenhagen Hospital Biobank (CHB) and Danish Blood Donor Study (DBDS) were collected from previous studies (NVK-1700407, P-2019-99; NVK-1708829,P-2019-93; NVK-1803812, P-2019-51). Ethical approval was not required for reusing these datasets in demographic and genealogical research. The use of data from the MultiGeneration Registry-Lite was approved by the Danish National Archives (Case number 22/33918; Reference number KAMB-MOEL). Informed consent or evaluation by a research ethics committee is not required for registry-based research in Denmark.

Conflict of interest

SB has ownerships in Intomics A/S, Hoba Therapeutics Aps, Novo Nordisk A/S, Lundbeck A/S, ALK abello A/S, Eli Lilly and Co. The other authors declare no competing interests.

Acknowledgements

We thank all the volunteers for their participation in our study cohorts, including the UK Biobank, the Copenhagen Hospital Biobank and the Danish Blood Donor Study. This study has also received valuable suggestions and supports from the members of Birney and Brunak groups, particularly Kumar Gaurav, Maria Herrero-Zazo, Troels Siggaard and Lisa Cantwell.

This study was supported by the Novo Nordisk Foundation (grants NNF170C0027594, NNF14CC0001). XZ has also been funded by the Borysiewicz Interdisciplinary Fellowship from the University of Cambridge during the study.

The funding bodies had no role in the design and conduct of the study.

References

"England's Immigrants 1330 – 1550." n.d. Accessed March 19, 2025. http://www.englandsimmigrants.com/.

"Migration Museum of Denmark." 2024. 2024. https://immigrantmuseet.dk/migrationshistorien/roepolakkerne-paa-lolland/tusindvis-af-polakker.

"Rural Urban District." 2019. Derbyshire Observatory -. Derbyshire Observatory. June 24, 2019. https://observatory.derbyshire.gov.uk/rural-urban-district/.

"Statistikbank." 2024. 2024. https://www.statbank.dk/20024.

"Urban and Rural Survivorship in Pre- and Post-Black Death Denmark." 2021. Journal of Archaeological Science: Reports 38 (August):103089.

"What Age Did People Marry in the British Past?" n.d. The Cambridge Group for the History

of Population and Social Structure (blog). Accessed December 13, 2024. https://www.campop.geog.cam.ac.uk/blog/2024/07/11/what-age-did-people-marry/#more-68.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. “Fast Model-Based Estimation of Ancestry in Unrelated Individuals.” Genome Research 19 (9): 1655–64.

All of Us Research Program Genomics Investigators. 2024. “Genomic Data in the All of Us Research Program.” Nature 627 (8003): 340–46.

Als, Thomas D., Tove H. Jorgensen, Anders D. Børglum, Peter A. Petersen, Ole Mors, and August G. Wang. 2006. “Highly Discrepant Proportions of Female and Male Scandinavian and British Isles Ancestry within the Isolated Population of the Faroe Islands.” European Journal of Human Genetics : EJHG 14 (4): 497–504.

Alten, Sjoerd van, Benjamin W. Domingue, Jessica Faul, Titus Galama, and Andries T. Marees. 2024. “Reweighting UK Biobank Corrects for Pervasive Selection Bias due to Volunteering.” International Journal of Epidemiology 53 (3). https://doi.org/10.1093/ije/dyae054.

Appel, Liv. "The Scottish and English Citizens in Elsinore from the 16th to the 18th Century - an Archaeological Approach." Chap. 14 In Across the North Sea: Later Historical Archaeology in Britain and Denmark, C. 1500-2000 Ad, edited by Paul Belford, David Cranstone, Henrik Harnow, Madsen-Host Lene and Myrtue Anders, 187-94. Odense: University Press of Southern Denmark, 2012.

Arciero, Elena, Sufyan A. Dogra, Daniel S. Malawsky, Massimo Mezzavilla, Theofanis Tsismentzoglou, Qin Qin Huang, Karen A. Hunt, et al. 2021. “Fine-Scale Population Structure and Demographic History of British Pakistanis.” Nature Communications 12 (1): 7189.

Athanasiadis, Georgios, Jade Y. Cheng, Bjarni J. Vilhjálmsson, Frank G. Jørgensen, Thomas

D. Als, Stephanie Le Hellard, Thomas Espeseth, et al. 2016. “Nationwide Genomic Study in Denmark Reveals Remarkable Population Homogeneity.” Genetics 204 (2): 711–22.

Binderup, Foto: M. n.d. “Sandflugt.” Naturen i Danmark | Lex. Accessed February 17, 2025. https://naturenidanmark.lex.dk/Sandflugt.

Brock, Aske Laursen, and Jelle van Lottum. 2019. “Rural Maritime Labour Migration to Copenhagen and Stockholm (1700–1800).” Continuity and Change 34 (2): 231–52.

Browning, Sharon R., and Brian L. Browning. 2015. “Accurate Non-Parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent.” American Journal of Human Genetics 97 (3): 404–18.

Bycroft, Clare, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, et al. 2018. “The UK Biobank Resource with Deep Phenotyping and Genomic Data.” Nature 562 (7726): 203–9.

Chen, Zhengming, Junshi Chen, Rory Collins, Yu Guo, Richard Peto, Fan Wu, Liming Li, and China Kadoorie Biobank (CKB) collaborative group. 2011. “China Kadoorie Biobank of 0.5 Million People: Survey Methods, Baseline Characteristics and Long-Term Follow-Up.” International Journal of Epidemiology 40 (6): 1652–66.

Childs, W. R. 1998. “Medieval English Trade in the North Sea: Denmark and Jutland, c 1300-1500.” In A North Sea Region, s. 47–74. Fiskeri- Og Søfartsmuseets Studieserie, Nr. 11 (1998).

Christensen, Aksel E. 1938. Danmarks Befolkning og Bebyggelse i Middelalderen. In: Nordisk Kultur II. Befolkning under medeltiden, ed. Adolf Schück. Stockholm;: Bonniers. Pp. 1-57.

Dam, Peder, and Johnny Grandjean Gøgsig Jakobsen. Atlas over Danmark Serie Ii Bind 7. Copenhagen: Geografforlaget, 2008.

Due, Jeppe Klok, Marianne Giørtz Pedersen, Sussie Antonsen, Joen Rommedahl, Esben Agerbo, Preben Bo Mortensen, Henrik Toft Sørensen, et al. 2024. “Towards More Comprehensive Nationwide Familial Aggregation Studies in Denmark: The Danish Civil Registration System versus the Lite Danish Multi-Generation Register.” Scandinavian Journal of Public Health 52 (4): 528–38.

Ebenesersdóttir, S. Sunna, Marcela Sandoval-Velasco, Ellen D. Gunnarsdóttir, Anuradha Jagadeesan, Valdís B. Guðmundsdóttir, Elísabet L. Thordardóttir, Margrét S. Einarsdóttir, et al. 2018. “Ancient Genomes from Iceland Reveal the Making of a Human Population.” Science (New York, N.Y.) 360 (6392): 1028–32.

Enemark, Poul. Dansk oksehandel 1450-1550. Gylling: Aarhus Universitetsforlag, 2003

Engelstad, Eivind S. 1946. "Hollendere, Friser og Jyder." In: Vestlandske Kunstindustrimuseum Årbok 1943-45, 29-55; Bergen: A.S. John Griegs

Erikstrup, Christian, Erik Sørensen, Kaspar R. Nielsen, Mie T. Bruun, Mikkel S. Petersen, Klaus Rostgaard, Lise W. Thørner, et al. 2023. “Cohort Profile: The Danish Blood Donor Study.” International Journal of Epidemiology 52 (3): e162–71.

Finer, Sarah, Hilary C. Martin, Ahsan Khan, Karen A. Hunt, Beverley MacLaughlin, Zaheer Ahmed, Richard Ashcroft, et al. 2020. “Cohort Profile: East London Genes & Health (ELGH), a Community-Based Population Genomics and Health Study in British Bangladeshi and British Pakistani People.” International Journal of Epidemiology 49 (1): 20–21i.

Fry, Anna, Thomas J. Littlejohns, Cathie Sudlow, Nicola Doherty, Ligia Adamska, Tim Sprosen, Rory Collins, and Naomi E. Allen. 2017. “Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population.” American Journal of Epidemiology 186 (9): 1026–34.

Gislason, Hannes. 2023. “SNP Heterozygosity, Relatedness and Inbreeding of Whole

Genomes from the Isolated Population of the Faroe Islands." BMC Genomics 24 (1): 707.

Gissel, Svend. 1981. Trade and Supply. In: Svend Gissel et al. (eds), Desertion and Land Colonization in the Nordic Countris c. 1300-1600. Stockholm (Almqvist & Wiksell), 188-204.

Gravesen, Palle. 2001. Palle Gravesen: Den Geologiske Udforskning Af Fakse Kalkbrud Fra Midten Af 1700-Tallet Til Nu. Geologisk Tidsskrift.

Gretzinger, Joscha, Duncan Sayer, Pierre Justeau, Eveline Altena, Maria Pala, Katharina Dulias, Ceiridwen J. Edwards, et al. 2022. "The Anglo-Saxon Migration and the Formation of the Early English Gene Pool." Nature 610 (7930): 112–19.

Guldberg, Mette, Asger Nørlund Christensen, Max Pedersen, Christina Folke Ax, Martin Rheinheimer, and Elsemarie Dam-Jensen. Handel Og Vandel: Dansk-Hollandske Kontakter I 1600-1700-Tallet. In: Temp - Tidsskrift For Historie 10 (2020): 36-60.

Gundlund, Anna, Anne Vinkel Hansen, Grete Skøtt Pedersen, Sarah Fredsted Villadsen, Laust Hvas Mortensen, Karen Brøndum-Nielsen, and Anne-Marie Nybo Andersen. 2015. "A Register-Based Study of Diseases With an Autosomal Recessive Origin in Small Children in Denmark According to Maternal Country of Origin." Paediatric and Perinatal Epidemiology 29 (4): 351–59.

Gunnarsson, Árni Freyr, Jiazheng Zhu, Brian C. Zhang, Zoi Tsangalidou, Alex Allmont, and Pier Francesco Palamara. 2024. "A Scalable Approach for Genome-Wide Inference of Ancestral Recombination Graphs." bioRxiv. https://doi.org/10.1101/2024.08.31.610248.

Hadley, D. M. 2006. The Vikings in England. Manchester Medieval Studies. Manchester, England: Manchester University Press.

Hansen, Thomas Folkmann, Karina Banasik, Christian Erikstrup, Ole Birger Pedersen, David

Westergaard, Piotr Jaroslaw Chmura, Kaspar Nielsen, et al. 2019. “DBDS Genomic Cohort, a Prospective and Comprehensive Resource for Integrative and Temporal Analysis of Genetic, Environmental and Lifestyle Factors Affecting Health of Blood Donors.” BMJ Open 9 (6): e028401.

Helgason, A., S. Sigureth ardóttir, J. Nicholson, B. Sykes, E. W. Hill, D. G. Bradley, V. Bosnes, J. R. Gulcher, R. Ward, and K. Stefánsson. 2000. “Estimating Scandinavian and Gaelic Ancestry in the Male Settlers of Iceland.” The American Journal of Human Genetics 67 (3): 697–717.

Henn, Brenna M., Lawrence Hon, J. Michael Macpherson, Nick Eriksson, Serge Saxonov, Itsik Pe’er, and Joanna L. Mountain. 2012. “Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples.” PloS One 7 (4): e34267.

Holm, Poul. Kystfolk. Kontakter og sammenhænge over Kattegat og Skagerrak 1550-1914. Esbjerg: Fiskeri- og Søfartsmuseet, 1991.

Hu, Sile, Lino A. F. Ferreira, Sinan Shi, Garrett Hellenthal, Jonathan Marchini, Daniel J. Lawson, and Simon R. Myers. 2025. “Fine-Scale Population Structure and Widespread Conservation of Genetic Effect Sizes between Human Groups across Traits.” Nature Genetics, February. https://doi.org/10.1038/s41588-024-02035-8.

Huff, Chad D., David J. Witherspoon, Tatum S. Simonson, Jinchuan Xing, W. Scott Watkins, Yuhua Zhang, Therese M. Tuohy, et al. 2011. “Maximum-Likelihood Estimation of Recent Shared Ancestry (ERSA).” Genome Research 21 (5): 768–74.

Hybel, Nils, and Bjørn Poulsen. 2007. The Danish Resources C. 1000-1550: Growth and Recession. BRILL.

International HapMap 3 Consortium, David M. Altshuler, Richard A. Gibbs, Leena Peltonen, David M. Altshuler, Richard A. Gibbs, Leena Peltonen, et al. 2010. “Integrating Common and Rare Genetic Variation in Diverse Human Populations.” Nature 467

(7311): 52–58.

Johansen, Hans Chr. 2002. Danish Population History, 1600-1939.

Jorgensen, Tove H., Henriette N. Buttenschön, August G. Wang, Thomas D. Als, Anders D. Børglum, and Henrik Ewald. 2004. “The Origin of the Isolated Population of the Faroe Islands Investigated Using Y Chromosomal Markers.” Human Genetics 115 (1): 19–28.

Kelleher, Jerome, Yan Wong, Anthony W. Wohns, Chaimaa Fadil, Patrick K. Albers, and Gil McVean. 2019. “Inferring Whole-Genome Histories in Large Population Datasets.” Nature Genetics 51 (9): 1330–38.

Kershaw, Jane F. 2020. Viking Identities. Medieval History and Archaeology. London, England: Oxford University Press.

Kershaw, Jane, and Ellen C. Røyrvik. 2016. “The ‘People of the British Isles’ Project and Viking Settlement in England.” Antiquity 90 (354): 1670–80.

Kirin, Mirna, Ruth McQuillan, Christopher S. Franklin, Harry Campbell, Paul M. McKeigue, and James F. Wilson. 2010. “Genomic Runs of Homozygosity Record Population History and Consanguinity.” PloS One 5 (11): e13996.

Koenig, Zan, Mary T. Yohannes, Lethukuthula L. Nkambule, Xuefang Zhao, Julia K. Goodrich, Heesu Ally Kim, Michael W. Wilson, et al. 2024. “A Harmonized Public Resource of Deeply Sequenced Diverse Human Genomes.” Genome Research 34 (5): 796–809.

Kurki, Mitja I., Juha Karjalainen, Priit Palta, Timo P. Sipilä, Kati Kristiansson, Kati M. Donner, Mary P. Reeve, et al. 2023. “FinnGen Provides Genetic Insights from a Well-Phenotyped Isolated Population.” Nature 613 (7944): 508–18.

Kyrre, Hans. 1912. Furboerne : Dansk fiskerliv I. Copenhagen: Schønbergske Forlag.

Lall, Gurdeep Matharu, Maarten H. D. Larmuseau, Jon H. Wetton, Chiara Batini, Pille Hallast, Tunde I. Huszar, Daniel Zadik, et al. 2021. “Subdividing Y-Chromosome

Haplogroup R1a1 Reveals Norse Viking Dispersal Lineages in Britain." European Journal of Human Genetics: EJHG 29 (3): 512–23.

Leitsalu, Liis, Toomas Haller, Tõnu Esko, Mari-Liis Tammesoo, Helene Alavere, Harold Snieder, Markus Perola, et al. 2015. "Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu." International Journal of Epidemiology 44 (4): 1137–47.

Leslie, Stephen, Bruce Winney, Garrett Hellenthal, Dan Davison, Abdelhamid Boumertit, Tammy Day, Katarzyna Hutnik, et al. 2015. "The Fine-Scale Genetic Structure of the British Population." Nature 519 (7543): 309–14.

Loh, Po-Ru, Petr Danecek, Pier Francesco Palamara, Christian Fuchsberger, Yakir A. Reshef, Hilary K. Finucane, Sebastian Schoenherr, et al. 2016. "Reference-Based Phasing Using the Haplotype Reference Consortium Panel." bioRxiv. bioRxiv. https://doi.org/10.1101/052308.

Loh, Po-Ru, Pier Francesco Palamara, and Alkes L. Price. 2016. "Fast and Accurate Long-Range Phasing in a UK Biobank Cohort." Nature Genetics 48 (7): 811–16.

Manichaikul, Ani, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michèle Sale, and Wei-Min Chen. 2010. "Robust Relationship Inference in Genome-Wide Association Studies." Bioinformatics (Oxford, England) 26 (22): 2867–73.

Margaryan, Ashot, Daniel J. Lawson, Martin Sikora, Fernando Racimo, Simon Rasmussen, Ida Moltke, Lara M. Cassidy, et al. 2020. "Population Genomics of the Viking World." Nature 585 (7825): 390–96.

McLeod, Shane. 2014. The Beginning of Scandinavian Settlement in England: The Viking "Great Army" and Early Settlers, C. 865-900. Turnhout, Belgium: Brepols.

Møller, Anders Monrad. 1988 Jagt og skonnert. Studier i den danske provinssøfart i tiden fra 1814 til 1864. Copenhagen: Falcon.

Moorjani, Priya, Sriram Sankararaman, Qiaomei Fu, Molly Przeworski, Nick Patterson, and David Reich. 2016. “A Genetic Method for Dating Ancient Genomes Provides a Direct Estimate of Human Generation Interval in the Last 45,000 Years.” Proceedings of the National Academy of Sciences of the United States of America 113 (20): 5652–57.

Nagai, Akiko, Makoto Hirata, Yoichiro Kamatani, Kaori Muto, Koichi Matsuda, Yutaka Kiyohara, Toshiharu Ninomiya, et al. 2017. “Overview of the BioBank Japan Project: Study Design and Profile.” Journal of Epidemiology 27 (3S): S2–8.

Nait Saada, Juba, Georgios Kalantzis, Derek Shyr, Fergus Cooper, Martin Robinson, Alexander Gusev, and Pier Francesco Palamara. 2020. “Identity-by-Descent Detection across 487,409 British Samples Reveals Fine Scale Population Structure and Ultra-Rare Variant Associations.” Nature Communications 11 (1): 6130.

Naseri, Ardalan, Xiaoming Liu, Kecong Tang, Shaojie Zhang, and Degui Zhi. 2019. “RaPID: Ultra-Fast, Powerful, and Accurate Detection of Segments Identical by Descent (IBD) in Biobank-Scale Cohorts.” Genome Biology 20 (1): 143.

Nielsen, Helge. 2010. Middelalderen Bag Tallene. Nukøbing Sj.: Bogværket.

Nielsen, Niels Kayser. 2023. “Hedeopdyrkning, ca. 1750-1950,” November. https://danmarkshistorien.dk/vis/materiale/hedeopdyrkning.

Olsen, Albert. Bybefolkningen I Danmark paa Merkantilismens Tid. Copenhagen: C. A. Reitzel, 1932.

Olsen, Gunnar. Stavnsbåndet og tjenestekarlene. Historie / Jyske Samlinger, Ny række, 1, 1950-2, 197-218

P.H. Sawyer, ‘Anglo-Scandinavian trade in the Viking age and after’ in Anglo-Saxon Monetary History, ed. M.S. Blackburn (Leicester, 1986)

Palamara, Pier Francesco, and Itsik Pe’er. 2013. “Inference of Historical Migration Rates via Haplotype Sharing.” Bioinformatics (Oxford, England) 29 (13): i180–88.

Palamara, Pier Francesco, Todd Lencz, Ariel Darvasi, and Itsik Pe'er. 2012. "Length Distributions of Identity by Descent Reveal Fine-Scale Demographic History." The American Journal of Human Genetics 91 (5): 809–22.

Palamara, Pier Francesco. 2014. "Population Genetics of Identity by Descent." arXiv [q-bio.PE]. arXiv. http://arxiv.org/abs/1403.4987.

Pontoppidan, Erich. 1768. Den Danske Atlas. Vol. IV, Copenhagen.

Porsmose, E. 1988. "Det Danske Landbrugs Historie I. 4000 F Kr.-1536. Odense: Landbohistorisk Selskab." Edited by C. Bjørn, 205–416.

Ralph, Peter, and Graham Coop. 2013. "The Geography of Recent Genetic Ancestry across Europe." PLoS Biology 11 (5): e1001555.

Rodríguez-Varela, Ricardo, Kristjan H. S. Moore, S. Sunna Ebenesersdóttir, Gulsah Merve Kilinc, Anna Kjellström, Ludvig Papmehl-Dufay, Clara Alfsdotter, et al. 2023. "The Genetic History of Scandinavia from the Roman Iron Age to the Present." Cell 186 (1): 32–46.e19.

Sahoo, Samali Anova, Rslan A. Zaidi, Santosh Anagol, and Iain Mathieson. 2021. "Long Runs of Homozygosity Are Correlated with Marriage Preferences across Global Population Samples." Human Biology 93 (3): 201–16.

Shemirani, Ruhollah, Gillian M. Belbin, Christy L. Avery, Eimear E. Kenny, Christopher R. Gignoux, and José Luis Ambite. 2021. "Rapid Detection of Identity-by-Descent Tracts for Mega-Scale Datasets." Nature Communications 12 (1): 3546.

Sørensen, Erik, Lene Christiansen, Bartlomiej Wilkowski, Margit H. Larsen, Kristoffer S. Burgdorf, Lise W. Thørner, Janna Nissen, et al. 2021. "Data Resource Profile: The Copenhagen Hospital Biobank (CHB)." International Journal of Epidemiology 50 (3): 719–20e.

Speidel, Leo, Marina Silva, Thomas Booth, Ben Raffield, Kyriaki Anastasiadou, Christopher

Barrington, Anders Götherström, Peter Heather, and Pontus Skoglund. 2025. "High-Resolution Genomic History of Early Medieval Europe." Nature 637 (8044): 118–26.

Stoklund, Bjarne. Læsø Land. Økologi og kultur i et øsamfund 1550-1900. Copenhagen: Museum Tusculanum, 2018

Strunge, Jørgen, Jørgen Strunge Efter Emsholm, Foto: Torben Riis-Nielsen, and Foto: Inger Kappel (b). n.d. "Lynghederne – naturtype – Naturen i Danmark." Naturen i Danmark | Lex. Accessed February 17, 2025. https://naturenidanmark.lex.dk/Lynghederne_-_naturtype.

Teitsson, Björn. 1981. Geographical Variables. In: Svend Gissel et al. (eds), Desertion and Land Colonization in the Nordic Countries c. 1300-1600. Stockholm (Almqvist & Wiksell), 172-187.

Thareja, Gaurav, Yasser Al-Sarraj, Aziz Belkadi, Maryam Almotawa, Qatar Genome Program Research (QGPR) Consortium, Karsten Suhre, and Omar M. E. Albagha. 2021. "Whole Genome Sequencing in the Middle Eastern Qatari Population Identifies Genetic Associations with 45 Clinically Relevant Traits." Nature Communications 12 (1): 1250.

Thompson, Elizabeth A. 2013. "Identity by Descent: Variation in Meiosis, across Genomes, and in Populations." Genetics 194 (2): 301–26.

Tønnesen, Allan. Helsingørs udenlandske borgere og indbyggere ca. 1550-1600. Ringe: Misteltenen, 1985

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." Scientific Reports 9 (1): 5233.

Ulsig, Erik. Pest og befolkningsnedgang i Danmark i det 14. Århundrede. Historisk Tidsskrift 91, 1991, 21-43

Veluwenkamp, Jan Willem, Werner Scheltjens, and Siem Van der Woude. 2021. “Sound Toll Registers Online.” Tijdschrift Voor Sociale En Economische Geschiedenis/ The Low Countries Journal of Social and Economic History 18 (1): 147–60.

Wang, Richard J., Samer I. Al-Saffar, Jeffrey Rogers, and Matthew W. Hahn. 2023. “Human Generation Times across the Past 250,000 Years.” Science Advances 9 (1): eabm7047.

Westergaard, David, Frederik Hytting Jørgensen, Jens Waaben, Alexander Wolfgang Jung, Mette Lademann, Thomas Folkmann Hansen, Jolien Cremers, et al. 2024. “Uncovering the Heritable Components of Multimorbidities and Disease Trajectories Using a Nationwide Cohort.” Nature Communications 15 (1): 7457.

Zhang, Brian C., Arjun Biddanda, Árni Freyr Gunnarsson, Fergus Cooper, and Pier Francesco Palamara. 2023. “Biobank-Scale Inference of Ancestral Recombination Graphs Enables Genealogical Analysis of Complex Traits.” Nature Genetics 55 (5): 768–76.

Zhou, Ying, Sharon R. Browning, and Brian L. Browning. 2020. “IBDkin: Fast Estimation of Kinship Coefficients from Identity by Descent Segments.” Bioinformatics (Oxford, England) 36 (16): 4519–20.