Tracing the spread of Celtic languages using ancient genomics

McColl, H.; Kroonen, G.J.; Pinotti, T.; Koch, J.; Ling, J.; Demoule, J.-P.; ... ; Willerslev, E.

Tracing the Spread of Celtic Languages using Ancient Genomics

Hugh McColl[1,2*^], Guus Kroonen[3,4*^], Thomaz Pinotti[1,5], John Koch[6], Johan Ling[2], Jean-Paul Demoule[7,8], Kristian Kristiansen[1,2*] Martin Sikora[1], Eske Willerslev[1,9,10*]

[1]Lundbeck Foundation GeoGenetics Center, Globe Institute, University of Copenhagen, Copenhagen, Denmark, [2]Department of Historical Studies, University of Gothenburg, Gothenburg, Sweden, [3]Leiden University Centre for Linguistics, Leiden University, Leiden, The Netherlands, [4]Department of Nordic Studies and Linguistics, University of Copenhagen, Copenhagen, Denmark, [5]Laboratório de Biodiversidade e Evolução Molecular (LBEM), Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, [6]University of Wales Centre for Advanced Welsh and Celtic Studies, Aberystwyth, UK, [7]University of Paris I - Panthéon-Sorbonne, France, [8]UMR 8215 CNRS, [9]Department of Genetics, University of Cambridge, Cambridge, UK, [10]MARUM, University of Bremen, Germany

^These authors contributed equally

*Corresponding authors; email: ewillerslev@sund.ku.dk, hugh.mccoll@sund.ku.dk, guus@hum.ku.dk, kristian.kristiansen@archaeology.gu.se

Summary

Celtic languages, including Irish, Scottish Gaelic, Welsh and Breton, are today restricted to the Northern European Atlantic seaboard. However, between 3 and 2 thousand years before present (BP) Celtic was widely spoken across most of Europe. While often associated with Bell Beaker-related populations, the spread of this prominent Indo-European linguistic cluster remains debated[1–7,] . Previous genomic investigations have focused on its arrival to specific regions: Britain[8], Iberia[9] and Southwestern Germany[10]. Here, we utilize new genomic data from Bronze and Iron Age Europe to investigate the population history of historically Celtic-speaking regions, and test different linguistic theories on the origins and early spread of the Celtic languages. We identify a widespread demographic impact of the Central European Urnfield Culture. We find ancestry associated with its Knovíz subgroup in the Carpathian Basin to have formed between 4 – 3.2 kyr BP, and subsequently expanded across much of Western Europe between 3.2 and 2.8 kyr BP. This ancestry further persisted into the Hallstatt Culture of France, Germany and Austria, impacting Britain by 2.8 kyr BP and Iberia by 2.5 kyr BP. These findings support models of an Eastern Central rather than a Western European center of spread for a major component of all the attested Celtic languages. Our study demonstrates, yet again, the power of ancient population genomics in addressing long-standing debates in historical linguistics.

Main

Ancient DNA studies have transformed our ability to track human migrations and interactions, radically improving our understanding of the archaeological record and the formation of linguistic landscapes. Despite these advances, significant challenges remain, particularly in more recent prehistory and at which time various regions were affected by consecutive and increasingly subtle genetic shifts. One key question concerning the cultural formation of Bronze and Iron Age Europe involves the origins and dispersal of the Celtic languages.

Europe's present-day linguistic landscape is dominated mainly by Romance, Germanic and Slavic languages, which expanded relatively recently during the Iron Age, mainly following the rise of the Roman Empire and the so-called Barbarian Invasions[11–14]. Prior to these events, however, Celtic languages featured prominently among the continent's linguistic identities. These included Lepontic in the Alpine region, Celtiberian in Iberia, and a diverse range of Gaulish or closely related dialects extending across present-day France, Central Europe, Northern Italy, the Balkans, and even parts of Anatolia[15,16]. In the modern era, Celtic languages have retreated to the Northwestern Atlantic fringe, where they persist as the Goidelic varieties Irish and Scottish Gaelic, and the Brittonic varieties Welsh and Breton, the latter having back-migrated to the continent possibly reflecting pressure from the Anglo-Saxon settlements[16–20].

Although alternative interpretations exist[21–23], several recent genomic studies have linked the broader spread of the Indo-European language family to migrations of Steppe-derived populations of the Yamnaya Culture during the 5th millennium BP (calibrated years before present/1950 CE)[24–29]. These migrations introduced Steppe-related ancestry into Europe and through admixture with populations of Neolithic farmer-related ancestry (including the Globular Amphora Culture with Northeast European Hunter-Gatherer ancestry, European Farmers with Western Hunter-Gatherer ancestry, and Bronze Age Anatolians with Caucasus Hunter-Gatherer ancestry[30]), gave rise to key groups relevant to Bronze and Iron Age Europe.

Genetic analyses reveal admixture between Corded Ware and Globular Amphora-related groups during the Bronze Age, with widespread dissemination of this ancestry across Europe by 4,800 BP[30]. In Western Europe, Bell Beaker populations incorporated additional European Farmer-related ancestry, which remained absent in northern Europe until the Migration Period[31].

Steppe-impacted populations of the archeological Bell Beaker phenomenon have been proposed as carriers of several Indo-European dialects, potentially also including the linguistic ancestor of Celtic[1,3,7,32–34]. Recent studies on Britain have identified genetic transitions around 4500 BP and again between 3200–2800 BP, marking potential entry points for Celtic languages in this part of Europe[8,35,36]. While there have been suggestions of links between the later of these migrations to the British Isles and the Urnfield-associated Knovíz Culture of the Czech Republic[8] , studies of Iron Age genomes from the intermediate regions, France[37] and Southwest Germany[10], have found no general evidence of discontinuity from the Bronze Age to the Iron Age. As a result, the exact area in which Celtic formed and the timeframe of its wider European expansion remain contentious[38]. The debate is currently focused around three main models: (1) a Late Neolithic/Early Bronze Age spread along the Atlantic seaboard linked to the Bell Beaker Culture[5,6,39,40]; (2) a Bronze Age spread from France, Iberia or Northern Italy[1,4,41]; and (3) a Late Bronze Age/Early Iron Age spread from Central Europe associated with the Hallstatt and La Tène Cultures[42–45]. Thus, the appearance and dispersal of the Celtic language constitutes a key question concerning the cultural formation of Bronze and Iron Age Europe.

In an accompanying paper, we report the generation of 752 new genomes, including 126 originating from France, Germany, Austria, and the British Isles, which are of immediate relevance to these questions (Fig. 1). The final filtered dataset included imputed genomes from 4587 individuals (Supplementary Table S1.1), which were clustered using a hierarchical clustering method leveraging the total length of shared identity-by-descent (IBD) fragments between all individuals. We expand upon the results of[31] by focussing on the broad region spanning Western, Southern and Central Eastern Europe.

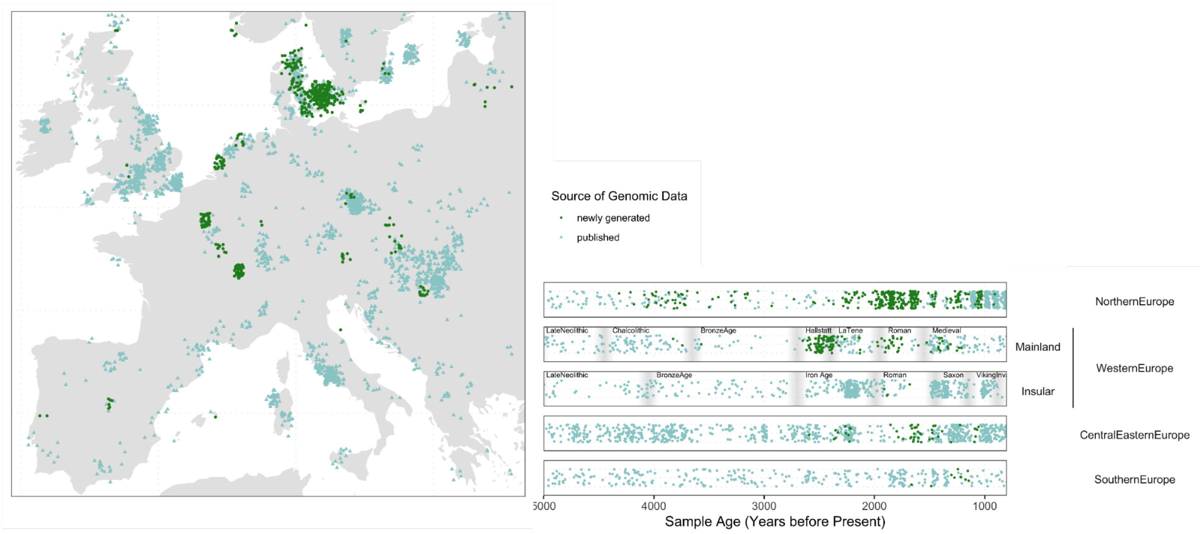


Fig. 1. Geographic and temporal sampling of the subset of ancient individuals included in the final dataset. Newly generated (green) and published (light blue) ancient individuals from the Late Neolithic / Early Bronze Age to the Viking Age. Grey bars on the timeline represent the boundary between historical periods for Western Europe.

The spatio-temporal distributions of the Bell Beaker-related populations

Our understanding of the early spatio-temporal distribution of the Bronze Age antecedents of the attested Celtic languages is limited by the absence of a written record. As recent studies on western Europe have shown a link between Celtic and Bell Beaker ancestry[31,46], insight into the formation of this ancestry profile and its geographical distribution through time may further our understanding of Celtic linguistic prehistory. We first repeated the IBD Mixture

Modelling analyses and ordinary global spatio-temporal kriging from[31] that revealed population structure that distinguishes northeastern from southwestern Europe from the Bronze Age onwards and found a strongly significant association linking individuals from Corded Ware archaeological contexts and later Germanic-speaking regions, and Bell Beaker archaeological contexts and later Celtic-speaking regions.

In brief, through IBD Mixture Modelling a palette is created for every ancient individual in the dataset, representing the proportion of IBDshared with all clusters in the dataset. Individuals from clusters are then chosen as 'sources' and others as 'targets', who are modelled using non-negative least squares. The use of non-negative least square modelling has been routinely applied using chromopainter chunks shared between ancient source and target sets[47], however recent studies have shown that using shared IBD segments between ancient individuals is effective at recovering finescale detail[30]. A strength of this method is that target palettes are modelled as combinations of potential sources palettes, meaning information from all other individuals in the dataset is incorporated into the modelling, rather than relying solely on the sharing of segments between the source and target individuals. When applied to densely sampled regions like Holocene Western Eurasia, sources and targets separated by 10,000 years continue to be modelled reliably[31] . By then performing spatio-temporal kriging[48] using the results from the IBD Mixture Modelling, we interpolate ancestry proportions for points in time and space where genetic data is not available.

We first compared the spatio-temporal kriging results at around ~4300 BP, when ancestry associated with Bell Beaker population first peaks, to ~2500 BP, where the written record documents the geographical distribution of Celtic across Europe. During this period, the range of Corded Ware-related ancestry remains stable (Fig. 2A, D), but Bell Beaker-related ancestry expands (Fig. 2B, E) and European Farmer-related ancestry contracts, correspondingly (Fig. 2C, F).

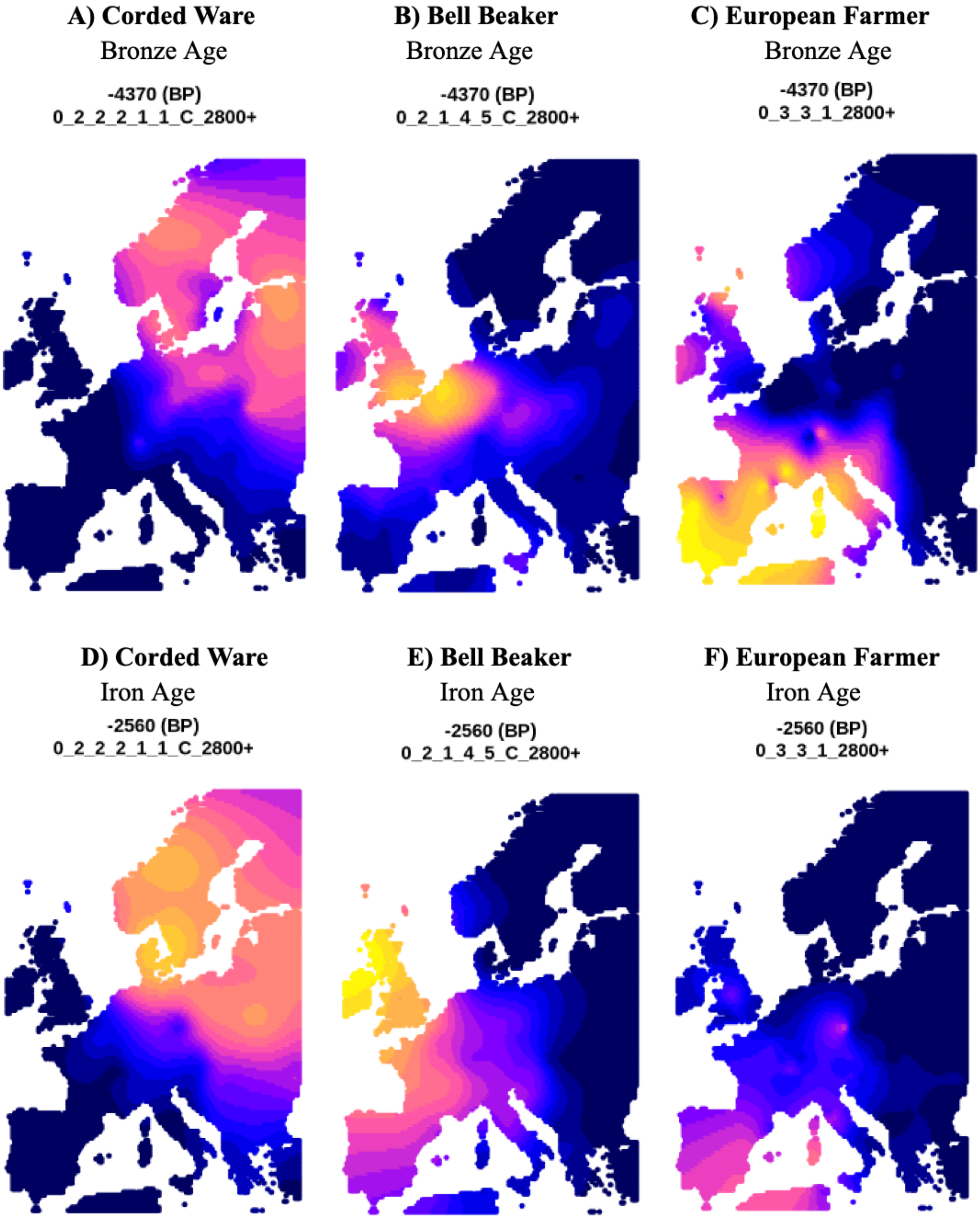


Fig. 2. Spatio-temporal kriging results showing the distribution of Corded Ware- (A,D) , Bell Beaker- (B, E) and European Farmer-related (C,F) ancestry at 4300 BP (upper

panel) and 2500 BP (lower panel). Brighter colours indicate a higher proportion of ancestry modelled. Kriging was performed on the IBD Mixture Modelling set C2 results.

Based on direct sampling, we find early evidence of Bell Beaker-related ancestry being present in France by 4463 BP, the Netherlands by 4384 BP, on the British-Irish Isles by ~4300 BP (England: 4333 BP, Scotland: 4312 BP, Ireland: 3906), and reaching southern Europe by ~4300 BP (Spain 4367 BP, Portugal 4158 BP, Italy 4212 BP). Despite occurring at a similar time, the proportion of Bell Beaker-related ancestry in France, Spain and Italy is less than that of Britain and the Netherlands (Supplementary Fig. S1.5). Further east, the Bell Beaker period was preceded by the Corded Ware Culture. This transition is mirrored genetically in which the Bell Beaker-related individuals follow, or overlap with the end of, the temporal distribution of Corded Ware-related individuals. This is seen in Germany (CWC-related 4703–4358 BP, BB-related by 4439), Hungary (BB-related by 4299 BP, other CWC-related from the same time), Czech Republic (CWC-related 4739–4477 BP, BB-related by 4333), and Poland (CWC-related 4705–4239 BP, BB-related by 4250 BP). Additionally, in all regions, we found a strong temporal correlation between the arrival of Beaker-related ancestry and Y-chromosome haplogroup R-P312, confirming their association (Supplementary Note S3).

Throughout the Bronze and Iron Age, multiple eastward and westward migrations are evident at the peripheries of the region of Bell Beaker-related ancestry. This can be seen, for example, in the Czech Republic, with the expansion of Bell Beaker-related ancestry evident during the Bell Beaker period in place of Corded Ware-related ancestry, and the subsequent reduction during the Únětice period with the reappearance of Corded Ware-related ancestry from the east (Extended Data Fig. 1). However at present, our understanding of migrations occurring within the Bell Beaker-related region remains restricted to specific areas[8–10], rather than the region as a whole.

Population dynamics during the Bronze and Iron Age

While the genetic borders of Bell Beaker-related ancestry place a likely temporal and geographical boundary within which Celtic formed, understanding the population dynamics within this region may provide insight into whether this Iron Age linguistic landscape developed during a period of genetic continuity or disruption. We hence first attempted to identify the early genetic structure of the region. The occurrence of admixture between Steppe-related and European Farmer-related ancestry in Southern Europe has been well documented in ancient DNA studies[24,25,35,49]; therefore, we set out to assess whether this potentially distinct Farmer ancestry from across Western and Southern Europe could be used as a later proxy for migrations in the region.

Across Europe, we found a series of IBD clusters containing individuals of primarily European Farmer ancestry restricted to specific regions of Europe. To assess how these groups contributed to Bronze and Iron Age Bell Beaker-related populations, we performed IBD Mixture Modelling (Supplementary Note S1). We included source individuals from clusters associated with Neolithic British-Irish Isles (0_3_1_2800+), France and Iberia (0_3_2_4_4_1_2800+), Italy (0_3_2_5_2_2_2_2800+) and the Czech Republic (0_3_3_1_2800+), encompassing proposed homelands for Celtic. In addition, we included a series of Neolithic-related outgroups, Bronze Age Anatolia (0_3_4_1_1_2800+), Polish Globular Amphora Culture (0_3_3_3_C_2800+) and early Anatolian Farmers (Early Anatolian 0_3_4_3_3_2_2_2800+).

Despite relatively dense sampling now existing across the region of interest during the Bronze and Iron Age, very few genomes exist from the Late Bronze Age as a result of the widespread practice of cremation associated with the Urnfield Culture. We therefore first focus on the peripheries of the region of Bell Beaker-related ancestry, England and the Czech Republic, where dense sampling exists throughout the entire period, and France, where the previously sparse sampling has been greatly improved.

To visualise these changes through time, we first binned samples into the Early Bronze Age (4800–4000 BP), Middle Bronze Age (4000–3200 BP), Late Bronze Age (3200–2800 BP) and two Iron Age bins (2800–2470 BP and 2470–2000 BP), and plotted the proportion of ancestry modelled for each individual in (Fig. 3.).

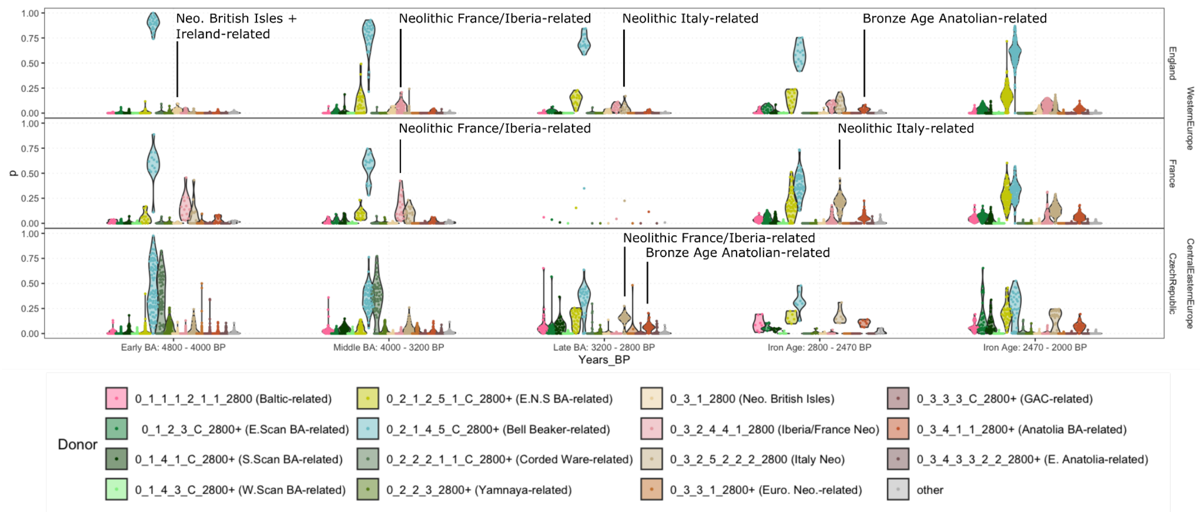


Fig. 3. IBD Mixture Modelling results (Set C3) for Bronze and Iron Age England, France and the Czech Republic. The shifts in relevant Neolithic Farming-related ancestries have been annotated. The full set of IBD Mixture Modelling results can be found in Supplementary Fig. S1.1.

For England, we find a series of transitions in the prominent farming ancestries present for each time slice (Fig. 3). Initially, between 4800–4000 BP, we find a individuals are modelled with a high proportion of Bell Beaker related ancestry, and the tendency to have a slightly higher proportion of local British-Irish Isles Neolithic ancestry, relative to the other Neolithic-related ancestries. By the Middle Bronze Age (4000–3200 BP), the highest Farmer-related ancestry is French/Iberian Neolithic-related rather than the local Neolithic ancestry, consistent with recent studies suggesting migrations from the mainland[8,36]. This migration, specifically the Iberian connection, is further supported by evidence that the UK received copper from Iberia during this phase (3350/3250–750 BP)[41]. However, in the Late Bronze Age, we see a shift, in which the proportion of Italian Neolithic ancestry has increased to similar proportions to that of French/Iberian Neolithic. In the Iron Age, similar patterns are seen, with the additional appearance of Bronze Age Anatolian-related ancestry. The changes in Farming ancestry present are suggestive of migrations from distinct regions of Europe in which local farming ancestry was incorporated.

In France, we see a similar transition (Fig. 3). During the Early and Middle Bronze Age, more local French/Iberian- than Italian Neolithic-related ancestry tends to be present. By the Iron Age, the relative proportions have swapped, so Italian Neolithic-related ancestry is the highest, accompanied by Bronze Age Anatolian ancestry. Due to the lack of samples from the Late Bronze Age, the time of this transition cannot be directly measured. However, the increased proportion of Italian Neolithic-related ancestry during the Late Bronze Age on the British Isles suggests it was present in France by this time.

Further east, in the Czech Republic, we see the increase of Italian Neolithic-related and Bronze Age Anatolian-related ancestry between 3200–2800 BP (Fig. 3). We also note that we detect no evidence of French/Iberian Neolithic ancestry. By splitting further into the cultural phases for the region, we find that this ancestry profile in the Czech Republic occurred by 3300 BP, in individuals associated with the Tumulus Culture and continuing into the Knovíz and Hallstatt Periods (Extended Data Fig. 1).

To understand how these Neolithic Farmer-related ancestries spread during the Bronze and Iron Age more broadly, we performed spatio-temporal kriging on the IBD Mixture Modelling results for these ancestries (Extended Data Fig. 2). For the three European Farmer related ancestries, we see the proportion of ancestry modelled decreasing through time. However, while we see a general range reduction of the British and French/Iberian Neolithic-related ancestries, we find an increase in the geographical range of the Italian Neolithic-related and Bronze Age Anatolian-related ancestries throughout these periods.

To model more directly, we included Bronze Age sources from the regions in which the distinct farming ancestries associated with the migrations were found. Relevant to the Middle Bronze Age migrations to the British-Irish Isles, we included individuals from France and Iberia, who are modelled primarily as Bell Beaker- and the local French/Iberian Bronze Age-related ancestry (0_3_2_2_2800+). Relevant to the appearance of Italian Neolithic and Bronze Age Anatolian-related ancestry in Western Europe by the Iron Age, we included individuals from Hungary/Serbia (0_3_4_2_2_C_2800+). Consistent with the results found from using the Farmer-related ancestries as a proxy, we find the appearance of Bronze Age French/Iberian ancestry appearing in England during the Middle Bronze Age, and the Hungarian/Serbian Bronze Age reaching widespread distributions during the Iron Age (Fig. 4). In the Czech Republic, we find almost all individuals being modelled with a large proportion of Hungarian/Serbian ancestry during the Late Bronze Age.

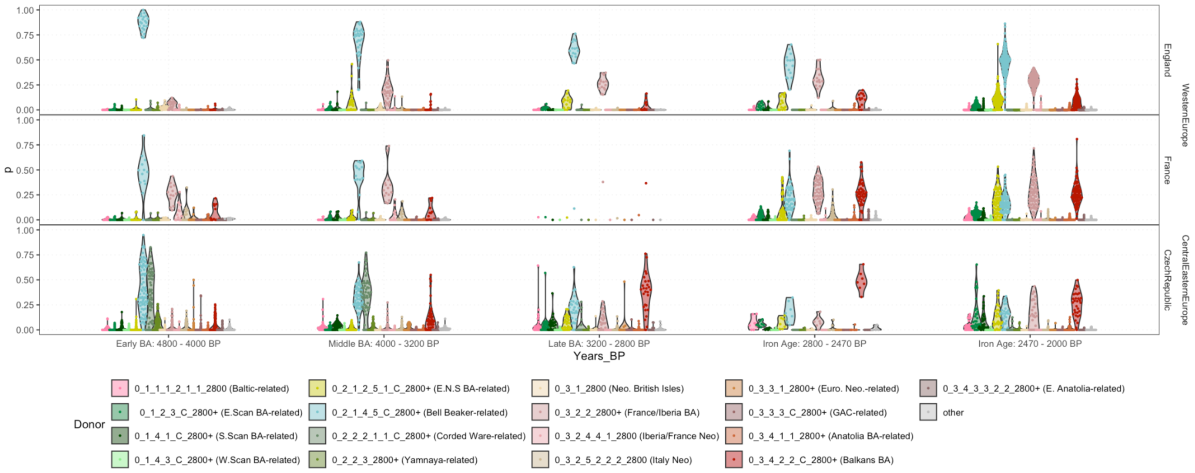


Fig. 4. IBD Mixture Modelling results for England, France and the Czech Republic during the Bronze and Iron Age, highlighting the Southeast (Hungarian/Serbian) and Southwest (France Iberian) European Bronze Age ancestries.

To understand how these Bronze Age-related ancestries spread during the Bronze and Iron Age more broadly, we performed spatio-temporal kriging on the IBD Mixture Modelling results for these ancestries (Fig. 5). The full set of IBD Mixture Modelling results can be found in Supplementary Fig. S1.1.

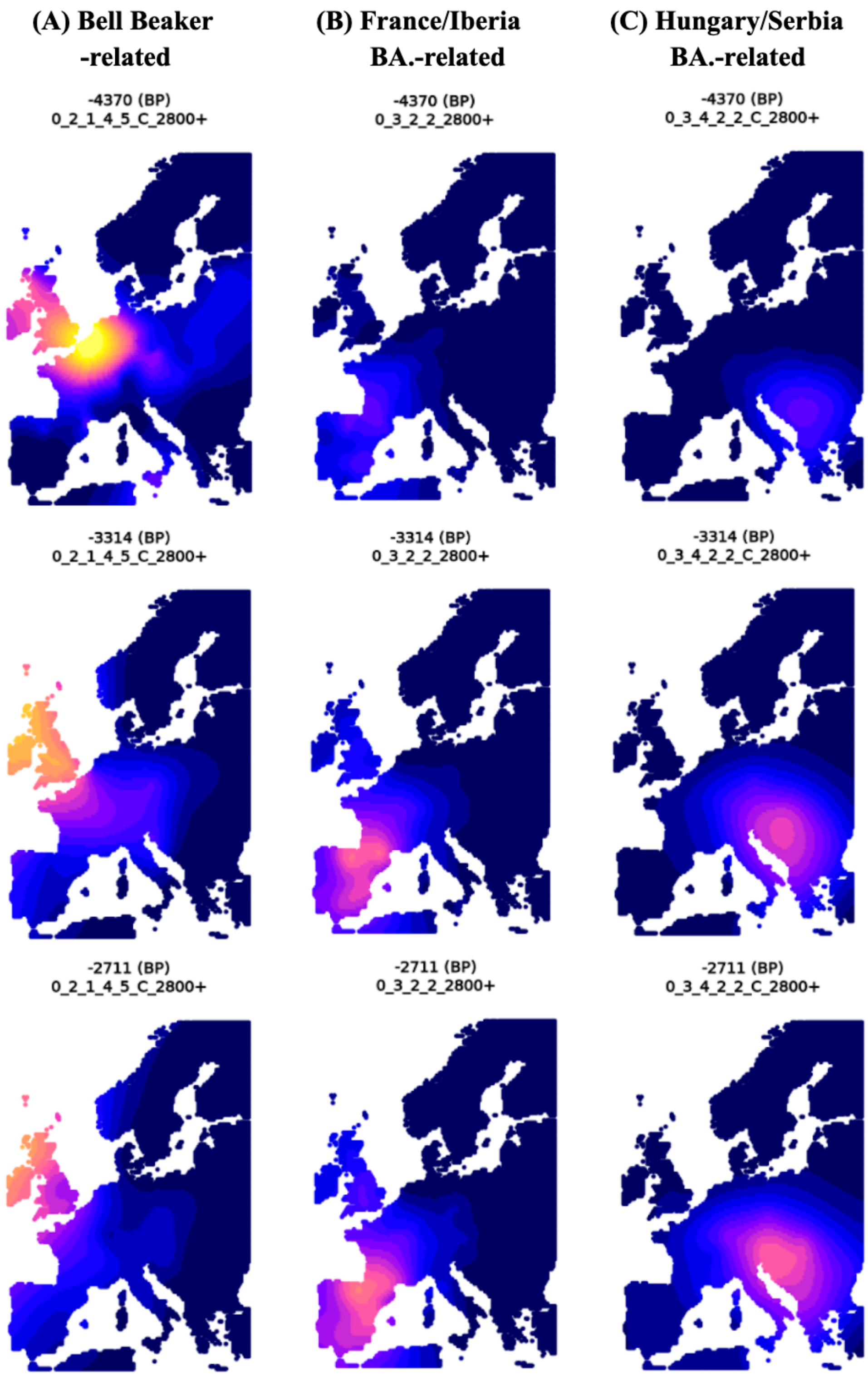


Fig. 5. Spatio-temporal kriging results for (A) Bell Beaker-related, (B) Bronze Age France/Iberia related, and (C) Hungary/Serbia-related sources from IBD Mixture Modelling set C4.

Next, we included individuals from the Late Bronze Age from the Czech Republic, associated with the Urnfield subgroup of the Knovíz Culture, as a source, who were modelled above with high proportions of Hungarian/Serbian ancestry (Extended Data Fig. 1). In the early Iron Age (2800–2470 BP), we find this ancestry modelled across Western Southern and Eastern Central Europe in varying proportions, complemented by more local sources (Extended Data Fig. 3.). In England, the highest proportion tends to be modelled as Bell Beaker-related, followed by Knovíz-related and French/Iberian Bronze Age-related. In contrast, on the mainland, the proportion of ancestry modelled as Bell Beaker-related tends to be low or absent, with the Steppe ancestry in these individuals better modelled by the other Bronze Age sources, i.e. Bronze Age Knovíz-related, French/Iberia and Hungary/Serbia. Compared to England, the impact of Knovíz-related ancestry is particularly high in France, Germany and the Czech Republic. In Austria, we note considerable diversity at the eponymous Hallstatt site, with individuals modelled with particularly high proportions of either Knovíz or Hungarian/Serbia Bronze Age-related ancestry, more similar to Hungary, Slovenia and Slovakia, where the Hungarian/Serbian Bronze Age-related ancestry is modelled in high proportions.

Finally, we included a Bronze Age source from Britain and Ireland (Set C6), corresponding to the fixed ‘left’ population described in the qpAdm analyses below. In England from the Late Bronze Age until the Migration Period, individuals tend to be modelled primarily by the British-Irish Bronze Age ancestry, with some Knovíz or French/Iberian Bronze Age-related ancestry (Supplementary Fig. S1.9). Notably, the East Yorkshire individuals are modelled with very high proportions of the local British-Irish Ancestry (Supplementary Fig. S1.1). In France, Iron Age individuals tend to be modelled primarily by the Knovíz and French/Iberian Bronze Age-related ancestries, but with little to no British-Irish Bronze Age-related ancestry.

Notable outliers with high British-Irish Bronze Age ancestry in France include two individuals from the Urville-Nacqueville necropolis in Normandy, a site archaeologically linked with Southern England[37], one of whom carries a typical British-Irish Beaker paternal lineage (R1b-L21).

The resolution to distinguish between these regions is also informative in relation to Migration Period movements. Some of the Migration Period individuals from Britain carry high proportions of Knovíz-related ancestry and little to no British-Irish Bronze Age-related, suggestive of migrations from or admixture on the continent (Supplementary Figure S1.1). Migration Period migrations into Denmark and Sweden detected elsewhere revealed that people carrying some continental ancestry[12,50], but primarily of Scandinavian ancestry[31], arrived in Denmark and Southern Sweden by the Viking Period; here we provide further insight into the source of the continental ancestry: the influx of small proportions of continental ancestry is modelled as Knovíz and Hungary/Serbia Bronze Age-related ancestry, generally lacking British-Irish Bronze Age and French/Iberian Bronze Age. As such, we can exclude the Netherlands, France, Britain and Ireland as a source of this continental ancestry, and infer a source region further east. This stands in direct contrast to Norway, where high proportions of the British-Irish Bronze Age-related ancestry are detected in most individuals with non-local ancestry, consistent with previous studies[50].

To confirm the occurrence of these migrations using a more conventional method, we ran a series of qpAdm analyses (Supplementary Note S2). In all models, we fixed a British-Irish Isles Bronze Age group as a ‘left’ source population (included in set C6 of the IBD Mixture Modelling). We performed a series of 1, 2 and 3 source population models. For the 2 and 3 source population models we used a rotational approach, in which all possible combinations of a series of Neolithic Farming-related (British-Irish Isles: 0_3_1_2800+, France/Iberia: 0_3_2_4_4_1_2800+, Denmark: 0_3_3_2_2_2_C_2800+, Italy (early): 0_3_4_3_5_4_C_2800+, Italy (late): 0_3_2_5_2_2_2_2800+, Czech Republic: 0_3_3_1_2800+, Poland (Globular Amphora Culture): 0_3_3_3_C_2800+, Bronze Age Anatolia: 0_3_4_1_1_2800+) and Bronze Age European (Bronze Age France/Iberia: 0_3_2_2_2800+, Hungary/Serbia: 0_3_4_2_2_C_2800+) populations were placed either as the 'left' or 'right' groups. As outgroups, we included populations from across Eurasia and Africa. The full list of individuals can be found in Supplementary Table S2.4.

From the IBD Mixture Modelling results, the occurrence of three distinct migrations is apparent on the British-Irish Isles. When using qpAdm, only in the earliest period (4800–4000 BP) do we see individuals being consistent under the 1 pop model, with the local Bell Beaker source (Fig. 6). In contrast, in the Middle Bronze and Late Bronze Age, only 2 population models are not rejected, with multiple possible Neolithic sources, or Bronze Age France/Iberia, but not Bronze Age Serbia/Hungary. In the Iron Age transects, no 1 or 2 population models are consistent with the data. In the first (2800–2470 BP), the only models not rejected include Bronze Age Serbia/Hungary. In the second (2470–2200 BP), the only non-rejected model (with marginal significance) includes Bronze Age Serbia/Hungary.

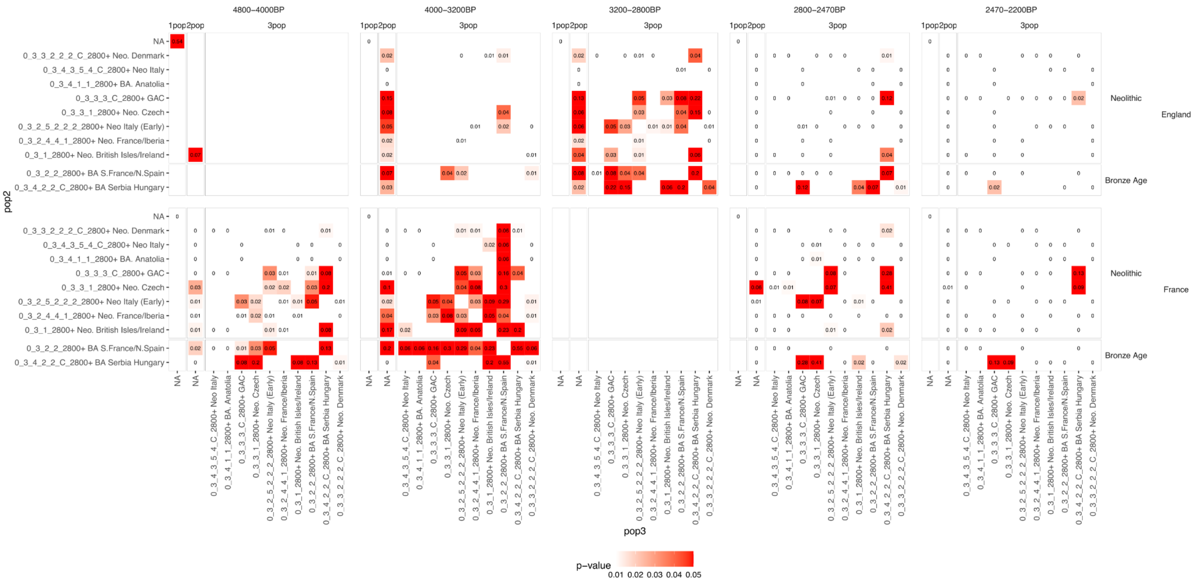


Fig. 6. P-values for rotating qpAdm models for the five time transects from England and France. For all models, Pop1 is fixed as the English Bell Beaker-related cluster. For the 2 population model, the second pop is listed on the Y axis. For the three population mode, the third population is listed on the X axis. P-values for feasible models are shown for each model.

In France, no 1 population model is consistent with the data (Fig. 6.). For the 4800–4000 BP bin, a Neolithic farming group or the local Bronze Age source are weakly significant. For the 4000–3200 BP bin, multiple Neolithic farming groups or the local Bronze Age group are not rejected, but not the Bronze Age Hungary/Serbia group. During the Iron Age a single Neolithic Farmer is feasible in a 2 population model for the earlier bin, but not the later. In both, multiple 3 population models including Bronze Age Hungary/Serbia but not France/Spain fail to be rejected.

We find these results do not contradict, and are broadly consistent with the IBD Mixture Modelling results. We therefore conclude that in England, migrations from at least three distinct populations occurred. Following the first arrival of Steppe ancestry, we find Middle and Late Bronze Age groups can be modelled only with additional ancestry from Mainland Southern and Western Europe. In contrast the Iron Age population requires ancestry from more Eastern Central Europe to be well modelled, likely occurring at the end of the Bronze Age. In France a similar result is found; in the Iron Age, populations can only be modelled when including a Bronze Age source from Eastern Central Europe, indicating here too, migrations occurred from the east.

Additionally, we find some level of support to these migrations from Y-chromosome haplogroups. While R-P312 is closely linked spatiotemporally with Bell Beaker ancestry, its subhaplogroups exhibit great geographical structure, particularly in the Bronze Age, with R-L21 dominant in Britain and Ireland[3,8], R-DF27 in Iberia and France[9,51,52], and R-U152 in Central Europe and Northern Italy[24,35,53,54] (Supplementary Note S3). This distribution provides a framework for understanding later population movements. Consistent with IBD Mixture Modelling results, we see R-DF27 appearing in England during the Middle Bronze Age[8,37], likely from Iberia/France, and R-U152 reaching both England and France at the end of the Bronze Age[8,37,55]. In France, most occurrences are associated with Hallstatt and Gaulish contexts, further reinforcing this connection.

Discussion

The Late Bronze Age was a period of cultural transitions that were formative to the successive Iron Age, including the rise of intensified agriculture, metal craftsmanship and new forms of warfare. Archaeologically, this era saw the emergence and expansion of the Urnfield Culture (3300–2750 BP) across much of Europe[56,57]. Several factors contributed to its expansion, including the aforementioned advancements in metalworking and agricultural techniques, which provided economic stability and surplus[58]. The introduction of new crops, such as millet, may have further aided population growth[59,60]. This fostered expansion and facilitated the long-distance exchange of important commodities, including bronze, salt, and other raw materials[61]. In contrast, the northward expansion from Central Europe into northern Germany and southern Scandinavia appears to have culminated in a significant defeat, as suggested by the Tollense battle around 3200 BP[62] and evident also from the lack of Knovíz-related ancestry in Scandinavia prior to the Migration Period.

Notably, the Urnfield Period saw the introduction of a new collective burial rite based on cremation to different regions of Europe[63]. We therefore lack direct genomic evidence for this culture, with the exception of the Urnfield subgroup of the Knovíz Culture, known for its settlement expansion and metallurgical advancements[58]. We nevertheless find that the sporadic appearance of this ancestry in England during the Urnfield Period, followed by its widespread presence by the Hallstatt Period, suggests that populations associated with the wider Urnfield Culture were genetically similar to those of the Knovíz Culture. Our results are therefore in accordance with evidence of the westward expansion of the Urnfield Culture into France[58] and Northeast Iberia[64], suggested previously only by limited ancient DNA data[9,65].

Our results stand in contrast to recent studies on Iron Age France[37] and Southwest Germany[10], where qpAdm analyses do not reject models in which early Iron Age populations are modelled by local Bronze Age populations. This different outcome likely results from the application of less fine-scale methods and inclusion of fewer ancient samples. The benefit of IBD Mixture Modelling in providing resolution at the individual level is also apparent from the ability to detect diversity within a population, when compared to qpAdm analyses which are typically run at a population level to achieve statistical significance. Notably, the identification of the outlier individuals from Iron Age Normandy, France with elevated British-Irish Bronze Age ancestry, is consistent with archaeological evidence for this site suggesting origins in Southern England[37].

Consistent with previous studies on the British-Irish Isles[10], we observe the appearance of Neolithic-related ancestry during the Middle to Late Bronze Age. Here, however, we are able to distinguish between multiple migration waves: the initial Bell Beaker migrations, a second French/Iberian Bronze Age migration, and a final Knovíz-related migration from Eastern Central Europe. We further confirm links between a Middle Bronze Age outlier from Margetts Pit with high Early European Farming ancestry and the Knovíz individuals of Central Europe; we find this individual modelled with a high proportion of Knovíz-related ancestry (Supplementary Figure S1.1, Set C5). However, we note that this individual is also modelled with additional Italian- and France/Iberia Neolithic-related ancestry, likely representing admixture from mainland Western Europe. When using the C5 source set, later individuals from Iron Age Britain generally lack Italian Neolithic-related ancestry, indicating that this individual was not representative of the migration into the British-Irish Isles by the Iron Age.

Unlike the Steppe migrations across Neolithic Europe, now widely associated with the onset of the Indo-European language dispersal[24,25], the Bronze Age migrations discussed here likely involved groups already speaking diverged Indo-European dialects. While historically Germanic-speaking groups primarily carried Corded Ware-associated ancestry and expanded from Northern Europe during the Iron Age[13,31,66], Bell Beaker-associated groups, although culturally and linguistically diverse[38], more likely mediated Indo-European dialects ancestral to Italic and Celtic ("Italo-Celtic", Supplementary Note S4)[46]. A potential remnant of these dialects is Lusitanian in Iberia. Under a model of multiple Bell-Beaker related migrations into Ibera, the presence of this Italo-Celtic language, whose specific relation to Celtic remains ambiguous (Supplementary Note S4), may be the result of the initial intrusion into Iberia.

In the British-Irish Isles, as well as in Iberia and France, a large proportion of the local Bronze Age ancestries persists throughout the Iron Age, strengthening models in which the historically known varieties of Italo-Celtic evolved locally during these eras. Previous studies have indeed highlighted the possibility that Celtic languages, specifically the linguistic predecessors of Goidelic and Brittonic, were introduced to Britain from France during the Middle to Late Bronze Age[7]. This is supported by our own detection of France/Iberia Bronze Age-related ancestry and consistent with Atlantic models suggesting that Celtic emerged in Central-West Europe[1,41]. However, we find no evidence of this ancestry reaching Eastern Central Europe, where Celtic was spoken during the Iron Age, indicating that this migration does not account for the spread of Celtic as a whole.

Furthermore, the expansion of Knovíz-related ancestry from Eastern Central Europe detected here, continuing into Hallstatt and La Tène populations, provides new evidence supporting the linguistic model in which Celtic languages were mediated to France, Britain, Iberia and Italy during the Late Bronze Age by populations associated with the Urnfield Culture[67,68]. This aligns with the association of the Lepontic language with the Urnfield-derived Golasecca Culture of Northern Italy[69], as well as the Urnfield-type weaponry of ~3200–3000 BP depicted on Late Bronze Age warrior stelae of the Southwestern Iberian Peninsula[70–72]. Late Bronze Age Urnfield influences were notably highlighted in the discussion on the arrival of the Celtic languages in Britain by archaeologist and linguist V. G. Childe[68]. However, further intrusions to Ireland, suggested to have occurred either in the Late Bronze Age or Late

Iron Age[7], cannot be evaluated here due to the absence of genomes from Ireland for the period between 3500 and 1200 BP.

Conclusion

Here, we demonstrate that it is feasible to resolve various Bell Beaker-associated migrations. By exploring the hitherto elusive genomic impact of the Urnfield Culture, we highlight the potential for further understanding the formation of the distinct linguistic landscapes under scrutiny, once a sufficient number of ancient genomes from these regions is available.

Concretely, our findings support a major component of the Celtic languages forming in, and spreading from, Bronze Age Eastern Central Europe in association with the archaeological Urnfield, Hallstatt and La Tène Cultures[42,68,73]. They do not, in contrast, support a dispersal from Southern France or Iberia linked with Bronze Age migrations to the British Isles[1,2,41] or along the Atlantic seaboard having arrived there with the Late Neolithic Bell Beaker Culture[5,6,39]. Such scenarios require a spread of Celtic contrary to the identified demographic movements. Thus, the evaluation of contrastive hypotheses concerning the emergence and early spread of the Celtic languages yet again demonstrates the power of past population genomics in addressing long-standing debates in historical linguistics.

References

1. Sims-Williams, P. An alternative to 'Celtic from the East' and 'Celtic from the West'. *Camb. Archaeol. J.* **30**, 511–529 (2020).
2. Koch, J. T. The Arrival or Emergence of Celtic in the Iberian Peninsula in the Light of Recent Discoveries. in *Filosofia da Cultura Celta* (eds. Natário, C., Pinto, J., Borges, L. & Epifânio, R.) 150–179 (Instituto de Filosofia da Universidade do Porto & DG Edições, Porto, 2024).
3. Cassidy, L. M. *et al.* Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. *Proceedings of the National Academy of Sciences* **113**, 368–373 (2016).
4. Schrijver, P. Sound change, the Italo-Celtic linguistic unity, and the Italian homeland of Celtic. in *Celtic from the West 3* (eds. Koch, J. T. & Cunliffe, B.) 489–502 (Oxbow, Oxford, 2016).
5. Cunliffe, B. W. *Facing the Ocean: The Atlantic and Its Peoples, 8000 BC-AD 1500.* (Oxford University Press, 2001).
6. Brun, P. L'origine des Celtes. Communautés linguistiques et réseaux sociaux. in *Celtes et Gaulois, l'Archéologie face à l'Histoire. 2 La Préhistoire des Celtes* (ed. Vitali, D.) 29–44 (Glux-en-Glenne, Bibracte, 2006).
7. Mallory, J. P. From the steppe to Ireland: The impact of aDNA research. in *The Indo-European Puzzle Revisited* 129–145 (Cambridge University Press, 2023).
8. Patterson, N. *et al.* Large-scale migration into Britain during the Middle to Late Bronze Age. *Nature* **601**, 588–594 (2022).
9. Olalde, I. *et al.* The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**, 1230–1234 (2019).
10. Gretzinger, J. *et al.* Evidence for dynastic succession among early Celtic elites in Central

Europe. *Nature Human Behaviour* **8**, 1467–1480 (2024).

11. Stolarek, I. *et al.* Genetic history of East-Central Europe in the first millennium CE. *Genome Biology* **24**, (2023).
12. Speidel, L. *et al.* High-resolution genomic history of early medieval Europe. *Nature* **637**, 118–126 (2025).
13. Gretzinger, J. *et al.* The Anglo-Saxon migration and the formation of the early English gene pool. *Nature* **610**, 112–119 (2022).
14. Antonio, M. L. *et al.* Stable population structure in Europe since the Iron Age, despite high mobility. *Elife* **13**, (2024).
15. Jørgensen, A. R. Celtic. in *The Indo-European language family: a phylogenetic perspective* (ed. Olander, T.) 135–151 (Cambridge University Press, Cambridge, United Kingdom ; New York, NY, 2023).
16. *The Celtic World: Critical Concepts in Historical Studies*. (Routledge, London, 2007).
17. *The Celtic Languages*. (Cambridge University Press, Cambridge, 1992).
18. Ball, M. J. & Muller, N. *The Celtic Languages*. (Routledge, London / New York, 2012).
19. Cunliffe, B. *Bretons and Britons: The Fight for Identity*. (Oxford University Press, 2021).
20. Fleuriot, L. *Les origines de la Bretagne: l'émigration*. (1980).
21. Renfrew, C. The origins of Indo-European languages. *Scientific American* **261**, 106–115 (1989).
22. Demoule, J.-P. *The Indo-Europeans: Archaeology, Language, Race, and the Search for the Origins of the West*. (Oxford University Press, 2023).
23. Heggarty, P. *et al.* Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *Science* **381**, eabg0818 (2023).
24. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–

172 (2015).

25. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).

26. De Barros Damgaard, P. *et al.* The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* **360**, eaar7711 (2018).

27. Lazaridis, I. *et al.* The genetic history of the Southern Arc: A bridge between West Asia and Europe. *Science* **377**, eabm4247 (2022).

28. Hovhannisyan, A. *et al.* Demographic history and genetic variation of the Armenian population. *Am. J. Hum. Genet.* (2024) doi:10.1016/j.ajhg.2024.10.022.

29. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).

30. Allentoft, M. E. *et al.* Population genomics of post-glacial western Eurasia. *Nature* **625**, 301–311 (2024).

31. McColl, H. *et al.* Steppe Ancestry in Western Eurasia and the Spread of the Germanic Languages. *bioRxiv* 2024.03.13.584607 (2025) doi:10.1101/2024.03.13.584607.

32. Gallay, A. Bell Beakers Today. Pottery, people, culture, symbols in prehistoric Europe: Proceedings of the international colloquium Riva del Garda (Trento, Italy) 11-16 May 1998. in *L'énigme campaniforme* (ed. Nicolis, F.) 41–57 (Provincia Autonoma di Trento, Servizio Beni Culturali, Ufficio Beni Archeologici, Trento, 2001).

33. Vander Linden, M. The band vs. the cord, or can Indo-European reconstructed institutions be tested against archaeological data? in *Proceedings of the Fourteenth Annual UCLA Indo-European Conference* (eds. Jones-Bley, K., Della Volpe, A., Huld, M. & Robbins Dexter, M.) 284–303 (Institute for the Study of Man, Washington, DC, 2003).

34. Falileyev, A. A Folk who will never speak: Bell Beakers and linguistics. in *The Bell*

Beaker transition in Europe: Mobility and local evolution during the 3rd Millennium BC (eds. Pilar Prieto Martínez, M. & Salanova, L.) 1–7 (Oxbow Books, Oxford, England, 2015).

35. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196 (2018).
36. Patterson, N. What genetics can say about Iron Age and Bronze Age Britain. in *Presenting Counterpoints to the Dominant Terrestrial Narrative of European Prehistory* (eds. Koch, J. T., Fauvelle, M., Cunliffe, B. & Ling, J.) (2025).
37. Fischer, C.-E. Origin and mobility of Iron Age Gaulish groups in present-day France revealed through archaeogenomics. *iScience* **25**, 104094 (2022).
38. Gibson, C. & Wodtko, D. S. *The Background of the Celtic Languages: Theories from Archaeology and Linguistics*. (University of Wales Centre for Advanced Welsh and Celtic Studies, Aberystwyth, 2013).
39. Cunliffe, B. W. & Koch, J. T. *Celtic from the West: Alternative Perspectives from Archaeology, Genetics, Language, and Literature*. (Oxbow Books, 2010).
40. Koch, J. T. & Fernández Palacios, F. A Case of Identity Theft? Archaeogenetics, Beaker People, and Celtic Origins. in *Exploring Celtic Origins: New Ways Forward in Archaeology, Linguistics, and Genetics* (eds. Cunliffe, B. & Koch, J. T.) vol. 22 38–79 (Oxbow Books, Oxford, 2019).
41. Ling, J., Díaz-Guardamino, M., Horn, C., Koch, J. & Stos-Gale, Z. A. *Bronze Age Rock Art in Iberia and Scandinavia: Words, Warriors, and Long-Distance Metal Trade*. (Oxbow Books, 2024).
42. Chadwick, N. K. *The Celts*. (Penguin Books, Harmondsworth, 1970).
43. Powell, T. G. E. *The Celts*. (Thames & Hudson, London, 1958).
44. Schmidt, K. H. History and culture of the Celts, draft plan of a comprehensive survey. in

Geschichte und Kultur der Kelten (ed. Schmidt, K. H.) 14–24 (Universitätsverlag Winter, Heidelberg, 1986).

45. Schmidt, K. H. The Celts and the ethnogenesis of the Germanic people. *Historische Sprachforschung / Historical Linguistics* **104**, 129–152 (1991).
46. Yediay, F. E. *et al.* Ancient genomics support deep divergence between Eastern and Western Mediterranean Indo-European languages. *bioRxiv* 2024.12.02.626332 (2024) doi:10.1101/2024.12.02.626332.
47. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet* **8**, e1002453 (2012).
48. Racimo, F. *et al.* The spatiotemporal spread of human migrations during the European Holocene. *Proceedings of the National Academy of Sciences* **117**, 8989–9000 (2020).
49. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
50. Margaryan, A. *et al.* Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).
51. Seguin-Orlando, A. *et al.* Heterogeneous Hunter-Gatherer and Steppe-Related Ancestries in Late Neolithic and Bell Beaker Genomes from Present-Day France. *Curr Biol* **31**, 1072–1083.e10 (2021).
52. Villalba-Mouco, V. *et al.* Genomic transformation and social organization during the Copper Age-Bronze Age transition in southern Iberia. *Sci Adv* **7**, eabi7038 (2021).
53. Damgaard, P. de B. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369–374 (2018).
54. Antonio, M. L. *et al.* Ancient Rome: A genetic crossroads of Europe and the Mediterranean. *Science* **366**, 708–714 (2019).
55. Brunel, S. *et al.* Ancient genomes from present-day France unveil 7,000 years of its

demographic history. *Proc Natl Acad Sci U S A* **117**, 12791–12798 (2020).

56. Müller-Karpe, H. Beiträge zur Chronologie der Urnenfelderzeit nordlich und südlich der Alpen. *Römisch-Germansiche Forschungen* **22**, (1959).
57. Sørensen, M. L. S. & Rebay-Salisbury, K. A brief history of urns, urnfields, and burial in the Urnfield culture. in *Death and the body in Bronze Age Europe: from inhumation to cremation* 15–35 (Cambridge University Press, Cambridge, 2023).
58. Kristiansen, K. *Europe Before History*. (Cambridge University Press, 2000).
59. Filipović, D. *et al.* New AMS C dates track the arrival and spread of broomcorn millet cultivation and agricultural change in prehistoric Europe. *Sci Rep* **10**, 13698 (2020).
60. Orfanou, E. *et al.* Biomolecular evidence for changing millet reliance in Late Bronze Age central Germany. *Sci Rep* **14**, 4382 (2024).
61. Kristiansen, K. & Suchowska-Ducke, P. Connected histories: The dynamics of Bronze Age interaction and trade 1500–1100bc. *Proc. Prehist. Soc.* **81**, 361–392 (2015).
62. Brinker, U. *et al.* The Bronze Age battlefield in the Tollense Valley, Mecklenburg-Western Pomerania, Northeast Germany – Combat marks on human bones as evidence of early warrior societies in northern Middle Europe? in *Late Prehistory and Protohistory: Bronze Age and Iron Age (1. The Emergence of warrior societies and its economic, social and environmental consequences; 2. Aegean – Mediterranean imports and influences in the graves from continental Europe – Bronze and Iron Ages)* 39–56 (Archaeopress Publishing Ltd, 2016).
63. Falkenstein, F. The development of burial rites from the Tumulus to the Urnfield Culture in Southern Central Europe. in *Ancestral landscapes: burial mounds in the Copper and Bronze Ages* (eds. Borgna, E. & Müller Celka, S.) 329–340 (MOM, Lyon, 2012).
64. Ruiz-Zapatero, G. Bronce final-hierrola naturaleza de los campos de urnas. in *La*

transició bronze final-1a edat del ferro en els Pirineus i territoris veïns (ed. Mereadal, O.) 635–658 (Institut d'Estudis Ceretans, Puigcerdà, 2014).

65. Olalde, I. Estudio del origen ancestral de los celtíberos con datos paleogenómicos. *Antropo* **49**, 1–12 (2023).
66. Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nat Commun* **7**, 10408 (2016).
67. Hoz, J. de. Lepontic, Celtiberian, Gaulish and the archaeological evidence. *Etudes Celtiques* **29**, 223–240 (1992).
68. Childe, V. G. *The Bronze Age*. (Cambridge University Press, Cambridge, 1930).
69. Hoz, J. de. The Celts of the Iberian Peninsula. *Zeitschrift für celtische Philologie* **45**, 1–37 (1992).
70. *Bronze Age Rock Art in Iberia and Scandinavia: Words, Warriors, and Long-Distance Metal Trade*. (Oxbow Books, Oxford, 2024).
71. Brandherm, D. Westward ho? Sword-bearers and all the rest of it. in *Celtic from the West 2: Rethinking the Bronze Age and the arrival of Indo-European in Atlantic Europe* (eds. Koch, J. T. & Cunliffe, B.) 147–156 (Oxbow Books, Oxford, 2013).
72. Montero Ruiz, I., Hunt Ortiz, M. A. & Santos Zalduegui, J. F. El depósito de la Ría de Huelva: procedencia del metal a través de los resultados de análisis de isótopos de plomo. in *El hallazgo leones de Valdevimbre y los depósitos del Bronce Final Atlántico en la Península Ibérica* (eds. Celis, J., Delibes de Castro, G., Fernández Manzano, J. & Grau Lobo, L.) 194–209 (Estudios y Catálogos Museos de Castilla y León, Léon, 2007).
73. Green, M. *The Celtic World*. (Routledge, 2012).
74. Damgaard, P. B. *et al.* Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* **5**, 11184 (2015).
75. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed

target capture and sequencing. *Cold Spring Harb Protoc* **2010**, db.prot5448 (2010).

76. Gansauge, M.-T., Aximu-Petri, A., Nagel, S. & Meyer, M. Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nature Protocols* **15**, 2279–2300 (2020).

77. Hanghøj, K., Moltke, I., Andersen, P. A., Manica, A. & Korneliussen, T. S. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. *Gigascience* **8**, (2019).

78. Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet* **53**, 120–126 (2021).

79. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

80. Browning, B. L. & Browning, S. R. Detecting identity by descent and estimating genotype error rates in sequence data. *Am J Hum Genet* **93**, 840–851 (2013).

81. Soetaert, K., Van den Meersche, K. & van Oevelen, D. limSolve: Solving Linear Inverse Models. *CRAN: Contributed Packages* The R Foundation https://doi.org/10.32614/cran.package.limsolve (2008).

82. Pebesma, E. & Graeler, B. gstat: Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation. *CRAN: Contributed Packages* The R Foundation https://doi.org/10.32614/cran.package.gstat (2003).

Methods

Dataset preparation

The 578 new ancient individuals included in the analyses here are part of a larger dataset of 712 ancient shotgun sequenced genomes from [31]. Details on the sampling, data generation, quality checks, site description and detailed metadata can be found in [31]. In brief, ancient samples were visually identified as suitable for processing. Samples were drilled, pre-digested [74] , extracted [24], built into libraries [75,76], and sequenced on Illumina Hiseq 4000 and Novaseq 6000 platforms. Sequenced reads were mapped against the human reference genome (hg19), and authenticated based on the length of reads, damage patterns, endogenous rates and presence of contamination. All samples with a genomic coverage greater than 0.1X were imputed, and close relatives identified using ngsRelate (v2) [77] were excluded for IBD based analyses.

Genomes were imputed using Glimpse[78] against the 1000 Genomes (1000G) v5 phase 3 dataset[79]. Shared IBD segments were called using IBDseq [80] using default parameters. Segments with a LOD score <3 and hotspot regions [30] with excess sharing were excluded. A detailed description can be found in [31].

IBD Mixture Modelling

To identify fine-scale structure, we applied IBD Mixture Modelling[30]. We relied on the dataset and IBD Clustering from [31]. The cluster IDs for all individuals included in the dataset can be found in Supplementary Table S1.1. The 4587 ancient individuals in the dataset were determined to be suitable for IBD Mixture Modelling based on a series of filtering steps for genomic coverage (0.1X for shotgun sequenced data, 1X for 1240K capture data), average genotype probability for the 1240K captures sites (0.9) and the removal of close relatives. The final dataset included 690,211sites, being those that intersected with the sites 1000

Genomes panel and the 1240K sites, that pass the 1000 Genomes mappability mask, and an imputation INFO score of 0.5.

For each of the 4587 ancient individuals in the dataset, a palette was created, representing the proportion of sharing with all other clusters, and with the other individuals in their own cluster. The proportions were normalised by the number of individuals in each cluster. To account for the fact that the sharing within the individuals own cluster has n-1 individuals, sharing with all other clusters excludes one individual at random. As a result, clusters with a single individual do not contribute to other individuals' palettes. All segments greater than 1cM, with a lod score greater than 3 that did not originate from a hotspot of IBD sharing [30] were included in the palettes. IBD Mixture Modelling was then undertaken using these palettes, by modelling the target individual using non-negative least squares[81], as described in [30]. Standard errors of the ancestry proportions were calculated using a weighted block jackknife, leaving out each chromosome.

Spatio-temporal kriging

To understand the trends in ancestry through time, we performed spatiotemporal kriging [48] to interpolate IBD Mixture Modelling results for points in time and space where genetic data is not available. For the ancestry of particular relevance for Set 2 (Corded Ware-, Bell Beaker- and European Farmer-related) and Set C3 (Neolithic British Isles/Ireland-, Neolithic France/Iberia-, Neolithic Italy- and Bronze Age Anatolia-related), we fit spatiotemporal variograms via a metric covariance model [82].

qpAdm analyses

qpAdm analyses were undertaken on the 690,211 sites of the dataset described above, using admixtools2[83]. Populations were placed into three sets: left_fixed, right_fixed or rotational (Supplementary Table S2.4). The 'left_fixed' population was included as a potential source in all qpAdm analyses. The 'right_fixed' population was included as outgroups in all analyses. In 1 population models, all 'rotational' populations were included as outgroups. For 2 population models, one 'rotational' population was placed together with the 'left_fixed' population as potential sources, and all others as outgroups, in all possible combinations. For 3 population models, two 'rotational' populations were placed together with the 'left_fixed' population as potential sources, and all others as outgroups, in all possible combinations.

Data availability

Sequence data for the new 578 ancient genomes can be found in the ENA under accession: xxxxxxxx. The public 4009 ancient genomes included in the dataset are listed in Supplementary Table S1.1, with additional details provided in Supplementary Note S4 of the accompanying paper [31].

Methods References

References 74 - 82.

Acknowledgements

GK, JK, JL and MS are supported by Riksbankens Jubileumsfond *(grant M 21-0018 for the 6-year programme MARITIME ENCOUNTERS: a counterpoint to the dominant terrestrial* *narrative of European prehistory)*. KK is supported by Riksbankens Jubileumsfond *(M16-0455:1, Rise II)*. The Lundbeck Foundation GeoGenetics Centre is supported by grants from the Lundbeck Foundation *(R302-2018-2155, R155-2013-16338)*, the Novo Nordisk Foundation (*NNF18SA0035006*), the Wellcome Trust (*WT214300*), Carlsberg Foundation (*CF18-0024*), the Danish National Research Foundation (*DNRF94, DNRF174*), the University of Copenhagen (*KU2016 programme*) and Ferring Pharmaceuticals A/S.

Author contributions

EW initiated the study. HMC, GK, EW led the study. HMC, GK, MS, EW conceptualised the study. KK, MS, EW supervised the research. GK, JK, JL, KK, MS, EW acquired funding for research. HMC, TP, MS were involved in data analysis. HMC, GK, JK, JL, KK, EW drafted the main text. HMC, GK, TP, JK, JL, KK, MS drafted supplementary notes and materials. HMC, GK, TP, JK, JL, JPD, KK, MS, EW involved in reviewing and editing drafts.

Ethics declarations

The authors declare no competing interests

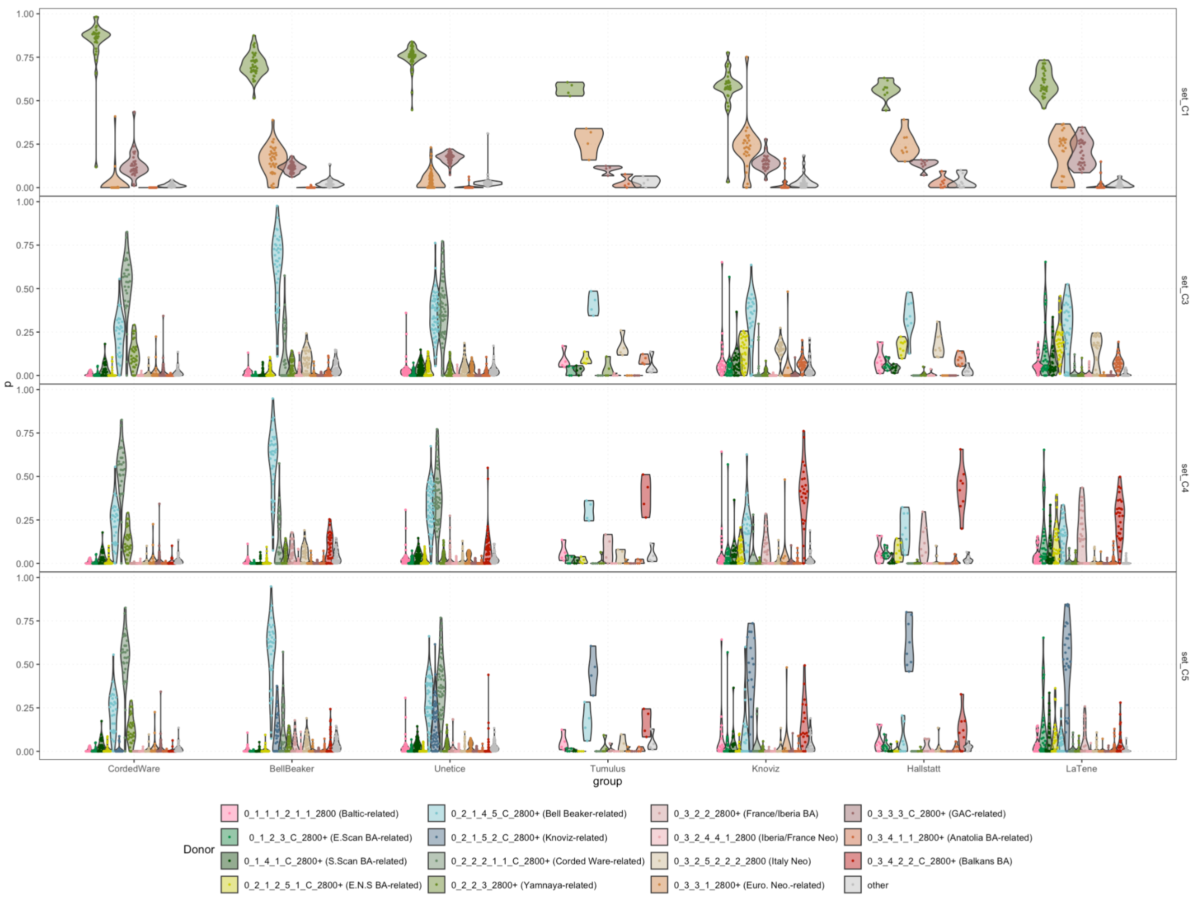
Additional Information

Supplementary Information is available for this paper.

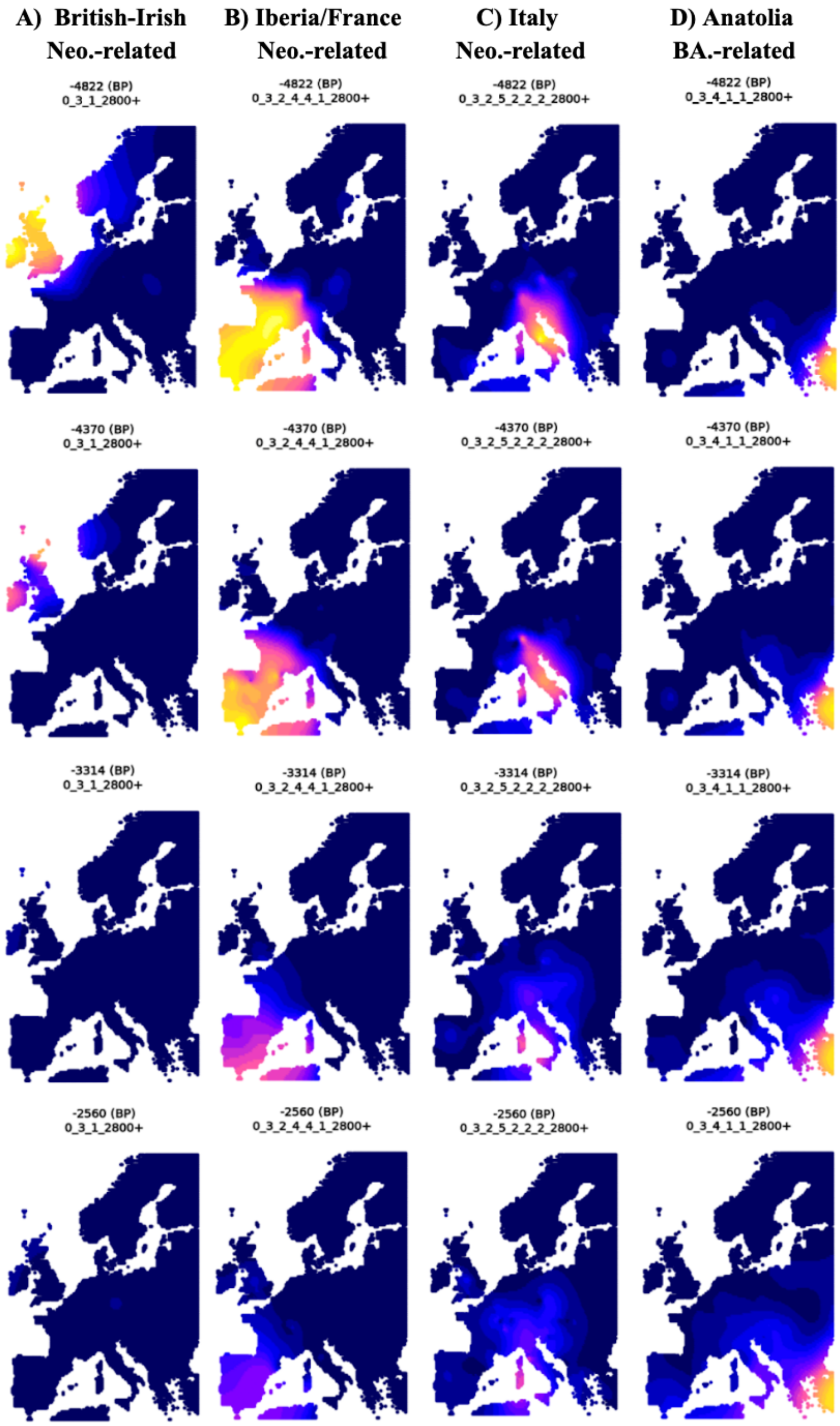
Correspondence and requests for materials should be addressed to: ewillerslev@sund.ku.dk, hugh.mccoll@sund.ku.dk, guus@hum.ku.dk, kristian.kristiansen@archaeology.gu.se

Reprints and permissions information is available at www.nature.com/reprints.

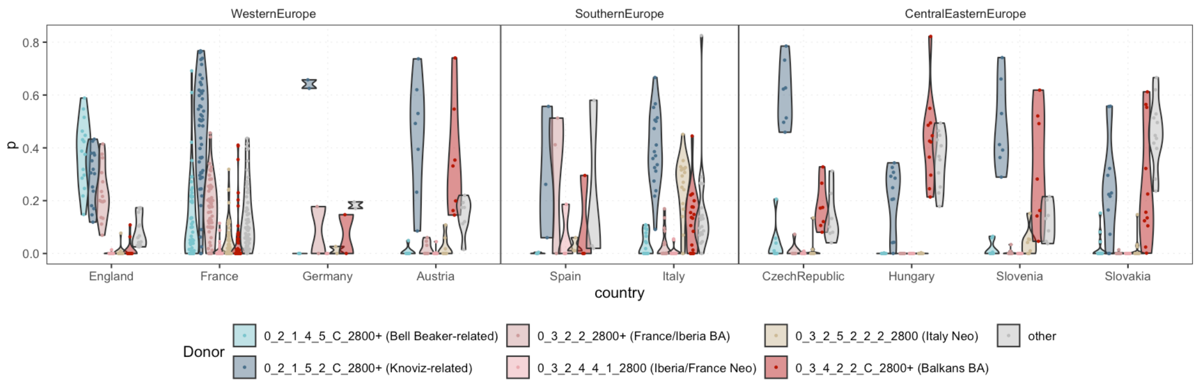
Extended Data Figures



Extended Data Fig. 1. IBD Mixture Modelling Results for samples associated with archaeological cultures/complexes from the Czech Republic for different archeological periods. Cultures are ordered from oldest (left) to youngest (right)**.** The top panel (set C1) highlights decreasing Yamnaya ancestry (olive green) and varying Farming (related ancestry) through time. The second panel (set 3) highlights transitions between Corded Ware (grey-green) and Bell Beaker (light blue) ancestry, and the various farming-related ancestries. The third panel (set C4) highlights the appearance of the Balkans Bronze Age (red) ancestry from the Tumulus Culture onwards. The lower panel (set C5) models the Knovíz-related ancestry (dark blue) directly.



Extended Data Fig. 2. Spatio-temporal kriging of from IBD Mixture Modelling for European Farmer-related source groups from (A) British Isles, (B) Iberia/France, (C) Italy and (D) Bronze Age Analolia. Kriging results from around the time of arrival of Steppe ancestry (4822 BP), the start of the Bell Beaker migrations (4300 BP), the Middle/Late Bronze Age (3314 BP) and the Iron Age (2560 BP) are shown (top to bottom). Kriging was performed on the IBD Mixture Modelling set C3 results.



Extended Data Fig. 3. Ancestry Proportions from IBD Mixture Modelling set C5 for the early Iron Age (2800 -2470 BP), showing the relevant ancestry sources.