Chapter 8

MtDNA Markers for Celtic and Germanic Language Areas in the British Isles

Peter Forster, Valentino Romano, Francesco Calì,
Arne Röhl & Matthew Hurles

Languages and DNA in Europe

The precise geographical origins of English and Insular Celtic languages on the European continent are obscure, especially so for Celtic languages. In North America, prehistoric language spread can now be traced using state-of-the art genetic markers, for example Na Dene speakers and Eskimo speakers each harbour high frequencies (up to 50 per cent) of distinctive mtDNA types not found elsewhere (Torroni *et al.* 1993; Saillard *et al.* 2000). However, the European situation contrasts with that of America: modern languages and human DNA do not appear to correspond particularly closely. Geographic distance tends to be better at predicting how similar the DNA of any two European populations is (Rosser *et al.* 2000; Zerjal *et al.* 2001).

Nevertheless, it would be overly pessimistic to conclude that in Europe, genetic markers have no hope of shedding light on the prehistory of languages. Genetic markers tracing language migrations may well exist and these markers could then tell us about the routes and, via the molecular clock, the times of such migrations. However, unlike in North America, these markers may represent only a small minority of the gene pool, especially if the spread occurred by élite dominance rather than by pioneer colonization, and the geneticist needs to identify and analyze the markers singly, otherwise their signal may be swamped.

In this paper our interest is to investigate the genetic prehistory of Celtic and Germanic speakers in the British Isles. We aim to show that genetic markers tracing the prehistoric origins of Celtic and English speakers living today indeed exist and will be useful to linguists, archaeologists and historians to provide indirect evidence for the origins of the languages themselves.

In this context, the term 'marker' needs to be explained. Occasionally, in the course of the millennia, mutations may occur in the DNA of an organism, such as a human or a seed of grain, and these mutations are passed down to the descendants. If these descendants remain united by a common characteristic such as a language, a certain geographic area, or a resistance against cereal rust, the DNA mutation can be considered a marker for that characteristic, even though it does not cause the characteristic.

The two best-characterized genetic systems for identifying evolutionary markers in humans are the Y chromosome (passed down exclusively from the father to his male children) and mitochondrial DNA (transmitted exclusively from the mother to her children). Individuals living today differ genetically from each other as a consequence of different mutation events which occurred in their past ancestry, irrespective of whether these mutation events are interesting markers for any characteristic.

To date there have been two approaches to analyzing data for the exploration of genetic patterning: the summary method and the lineage approach. The summary method treats the entire data set in terms of a population concept which considers all the different genetic types as a single heritable unit, which is passed down through time. The lineage method looks at the geographic spread and mutational age of specific lineages which in turn reflect movement of prehistoric individuals (for a discussion of the two approaches see Richards *et al.* 1997; Forster *et al.* 2001). This contrast in approach is also reflected in the choice of methods for detecting geographic patterns in genetic data. We will compare available coarse-grained approaches to detecting genetic patterns in Europe with the finer-scaled methods we propose in the main part of this paper.

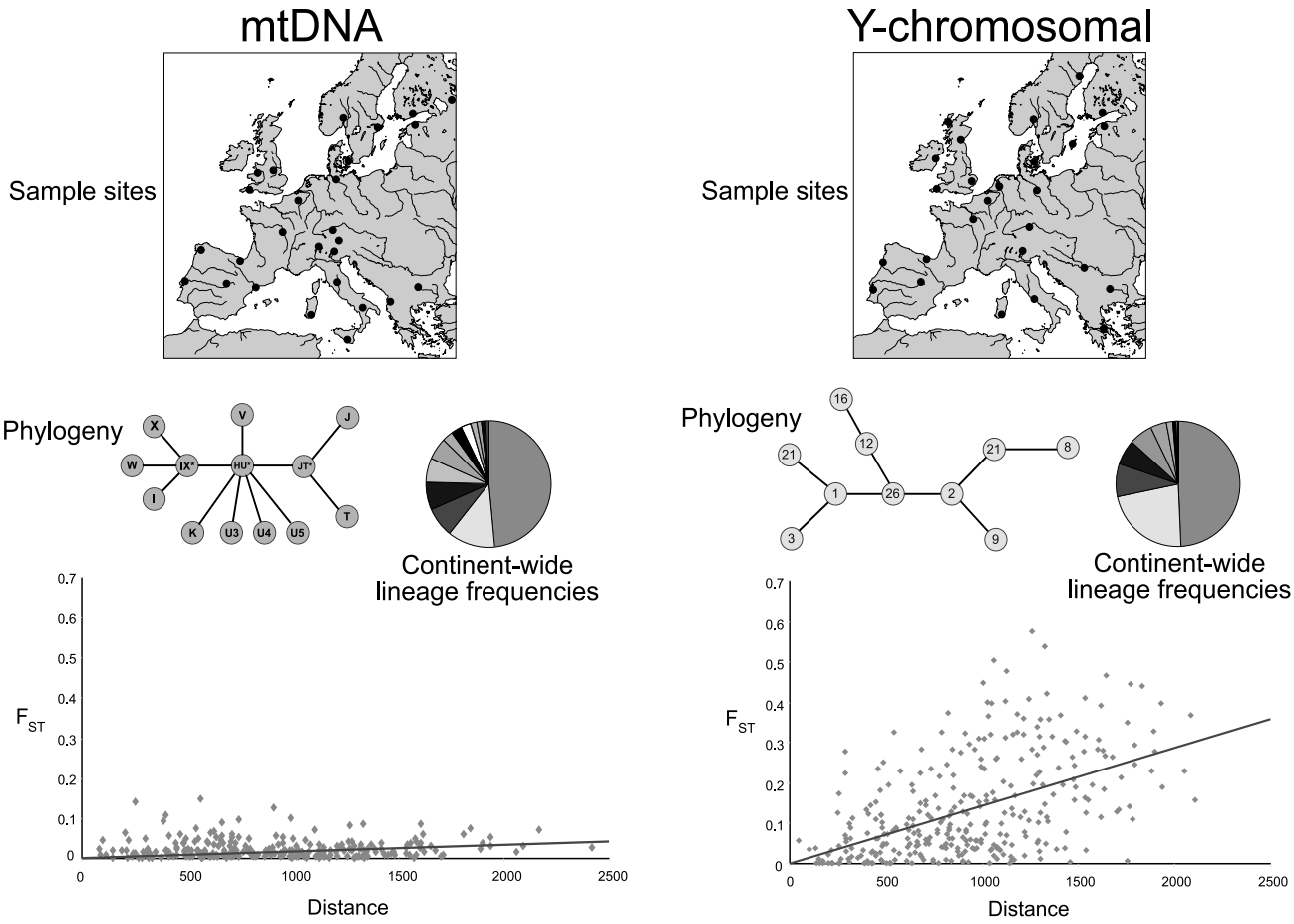

Figure 8.1. *European genetic landscapes of mtDNA and Y chromosomes. Data from ~2000 mtDNAs sampled at 27 locations (Simoni* et al. *2000, with corrections from the subsequent exchange of letters with Torroni* et al. *2000) and ~2300 Y chromosomes sampled at 26 locations (a subset of the data found in Rosser* et al. *2000) are split into 13 and 10 lineages respectively. Whereas 23 per cent of the Y-chromosomal variation is found between populations, only 2 per cent of the mtDNA variation is found between populations. For each locus, genetic distances (FST) and geographical distances between each population are calculated and plotted against one another. A steeper gradient of the regression line between these points indicates that lineage frequencies vary more dramatically between populations separated by a given geographical distance.*

Both approaches benefit from the fact that in Europe particularly, the spatial sampling of populations characterized at the molecular genetic level is becoming ever denser (currently more than 12,000 samples for mtDNA), making it possible to appreciate geographical variation across a 'genetic landscape'.

If we compare the mtDNA and the Y chromosome at a low resolution (i.e. subdivide them into a similar but low number of lineages) it becomes apparent that whereas the Y-chromosomal lineages exhibit dramatic changes in lineage frequencies across Europe, the same is less true of mtDNA lineages (see Fig. 8.1). At this coarse resolution of lineage diversity there is little evidence that there are geographically differentiated mtDNA lineages that could be correlated with different European language groups. Y-chromosomal lineages appear to be patterned more by geographical distance than by linguistic affiliations, although initial attempts are now being made to associate particular paternal and/or maternal lineages with archaeological cultures (Semino *et al.* 2000; Wilson *et al.* 2001). Alternatively, geneticists often avoid associating specific genetic markers with language groups and instead perform more general statistics that summarize the overall diversity. In this vein, Mike Weale and colleagues have identified Y-chromosomal differences within the British Isles apparently corresponding to language barriers (Weale

et al. 2002). British data have been supplemented since by Capelli *et al.* (2003) who focus on potential Scandinavian contributions. In the mtDNA field, summary approaches have been less helpful for our aim of reconstructing the genetic prehistory of the speakers of a language (Sajantila *et al.* 1995; Poloni *et al.* 1997), in part because of the lack of resolution in the genetic picture shown in Figure 8.1, and in part due to technical shortcomings outlined by Bandelt *et al.* (2002).

Before turning to the finer scale of the evolutionary mtDNA tree, the reader must be acquainted with the basis of the mtDNA nomenclature. Mitochondrial DNA variation in Europe is classified into major types, which are labelled as H, U, K, T, J, V, I, W, and X types as originally proposed by Torroni *et al.* (1996). At about 50 per cent, type H is by far the most frequent in Europeans, while the other mtDNA types amount to approximately zero to 15 per cent each, depending on the region. Incidentally, a comparison with current mtDNA diversity in surrounding potential source areas (e.g. Richards *et al.* 1996; 2000) suggests that each of these nine types was carried into Europe, rather than arising by mutation within Europe. It is then trivial to conclude that at least nine prehistoric women settled in Europe, and that each of these nine women had fertile daughters, who in turn had daughters, grand-daughters etc. until the present day. The geographic spread of these nine types includes more or less Europe, the Near East and North Africa. The first European mtDNA candidate for a language marker seems to have been discovered by Richards *et al.* in 1996. What is now known as mitochondrial type J (around 10 per cent of European mtDNA, depending on the region) contains several subtypes (Fig. 8.2) which show some linguistic specificity, especially within the British Isles. In 1996, only 821 European and Middle Eastern mtDNA sequences were available, but for Europe this figure has grown to more than 10,000 as of the year 2004 (according to an update of Röhl *et al.* 2001), partly thanks to forensic interest in mtDNA. So the time is ripe to revisit this potential language marker.

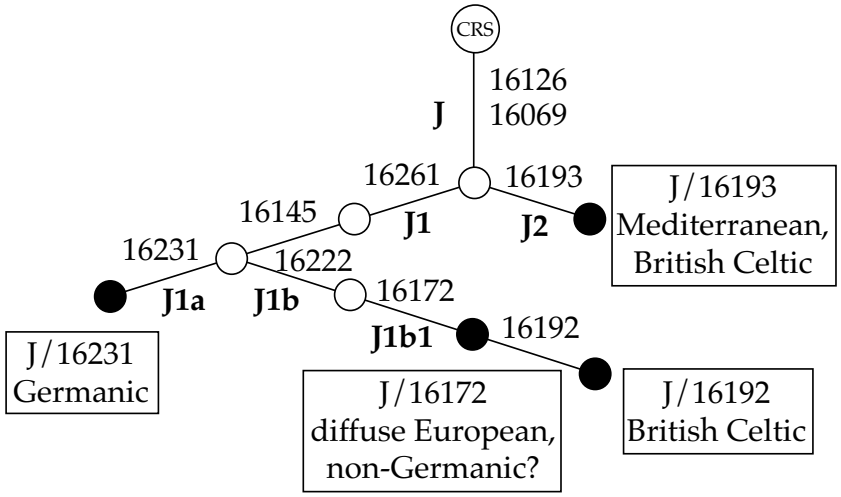

Figure 8.2. *Skeleton phylogeny of mtDNA type J. The numbers 16069, 16126 etc. refer to mutations at mtDNA nucleotide positions numbered as in Anderson* et al. *(1981). Labels J, J1 etc. refer to the branch nomenclature defined by Richards* et al. *(2000) and Richards & Macaulay (2000), whereas labels J/16193, J/16231 etc. refer to the provisional nomenclature used in this paper, as recommended by YCC (2002). Minor branches of J have been omitted here, and the geographic annotations are intentionally simplified. The mtDNA phylogeny and nomenclature is currently in flux due to complete mtDNA sequencing projects, so changes may be expected in the future.*

Linguistic and mtDNA landscape of the British Isles

In historical times, the Brythonic branch of Celtic was spoken in Cornwall, Wales and Brittany, while the Goidelic branch of Celtic was spoken in Ireland, Highland Scotland and the Isle of Man. The relationship of these 'Insular' Celtic languages to the 'Continental' Celtic languages such as extinct Gaulish (formerly spoken in what is now France and north Italy) is controversial due to the fragmentary nature of the Continental Celtic languages, all of which are extinct. Breton in Brittany is closely related to Welsh and Cornish and is generally thought to be due to recent back-migrations from the British Isles (summarized by Dubut *et al.* in press), and is therefore not a Continental Celtic language. Further uncertainty surrounds the timing of the arrival of Celtic languages to the British Isles, and indeed the definition of the term 'Celtic' itself, which was never applied to Britain by Greek and Roman historians (Renfrew 1987). An exploratory phylogenetic analysis of ancient Gaulish inscriptions suggests that Insular Celtic and Gaulish may belong to sister branches which split several thousands of years ago (Forster & Toth 2003).

Most of mainland Britain now is Germanic-speaking, and this is commonly thought to be owing

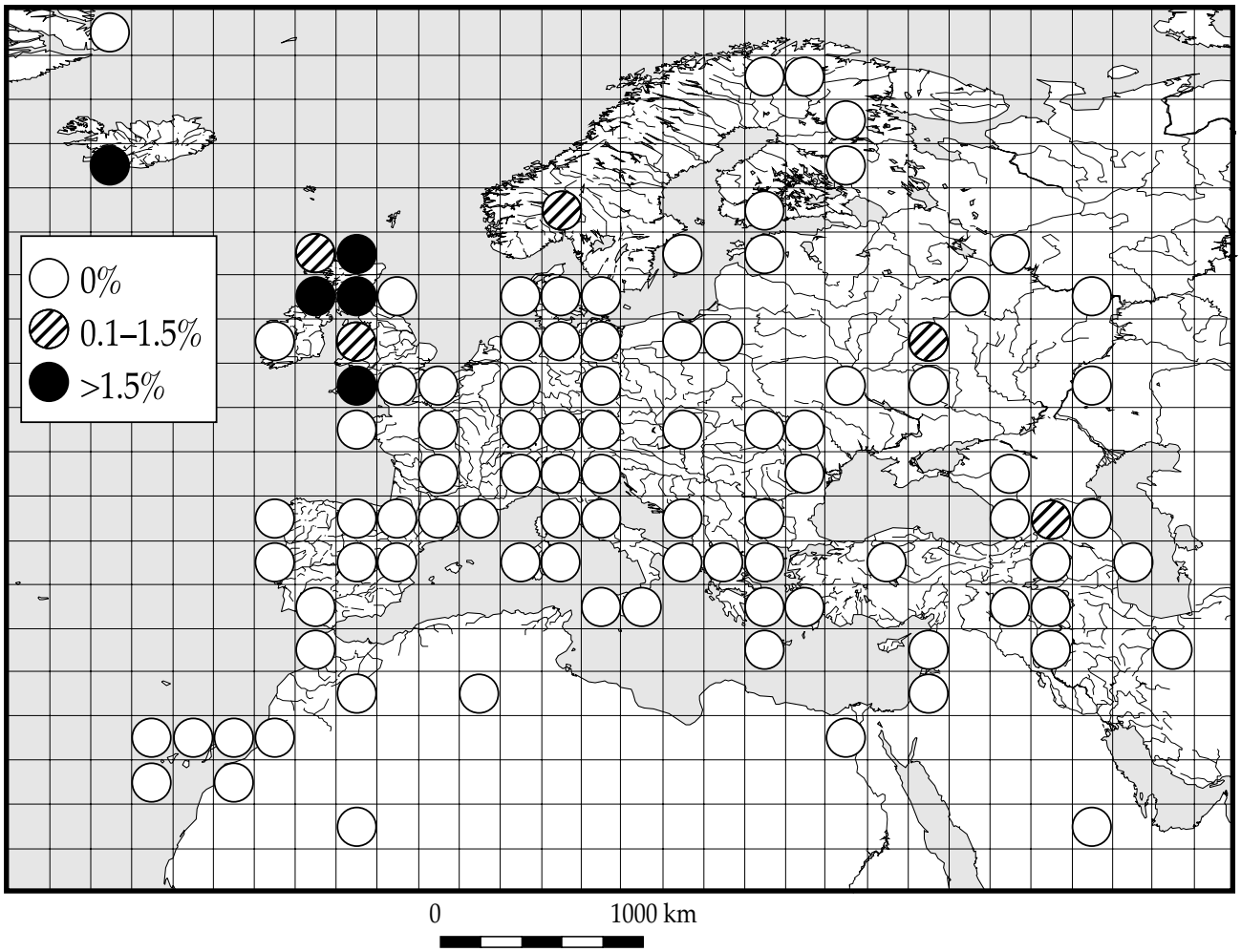

Figure 8.3. *Geographic distribution of mtDNA type J/16192. Symbols in the grid squares depict the local frequency of J/16192. Sample sizes in different grid squares are very uneven, ranging from 11 to nearly 1000 individuals; grid squares with sample sizes <10 individuals are excluded. The uneven and coarse 'grain' of the sample coverage means that percentage values in single grid squares should not be over-interpreted, but rather the broader patterns should be appreciated, as one would in a coarsely grained Mars rover image. In the mtradius data base used for this search, 19,493 individuals from native populations worldwide were activated, of which 12,747 are included in this map. In total, 55 individuals have the J/16192 type and are all shown on this map. There are no J/16192 types elsewhere in the world, according to the current data base.*

to the arrival, in the fifth century AD, of Angles, Saxons and Jutes (according to Beda) or Angles and Frisians (according to Procopius). Their homelands are thought to have been in Schleswig-Holstein and Jutland, and indeed 'Angeln' to this day is the name for a small region in Schleswig, while the homestead of the Frisians, as located by Plinius, is hundreds of kilometres distant (reviewed in Forster 1995). Origin myths tend to crumble under genetic scrutiny, and we shall compare this traditional hypothesis for the arrival of English to Britain with the available north German and English mtDNA data.

Turning first to mtDNA markers for Celtic areas in the British Isles, the most clear-cut Celtic mtDNA type within group J is type J/16192 (Fig. 8.2). In the current mtDNA data base containing around 19,493 representatively sampled individuals, J/16192 types (55 individuals) occurs so far only in British Celtic areas (Cornwall, Wales, Scotland, and Northern Ireland), in Iceland and Norway, and in one Russian and one Georgian individual (Fig. 8.3). It is possible that the Russian and the Georgian matches are due to independent parallel mutations or due to ancient common ancestry rather than representing recent migrants from the British Isles, so they will not be considered further here. The Nor-

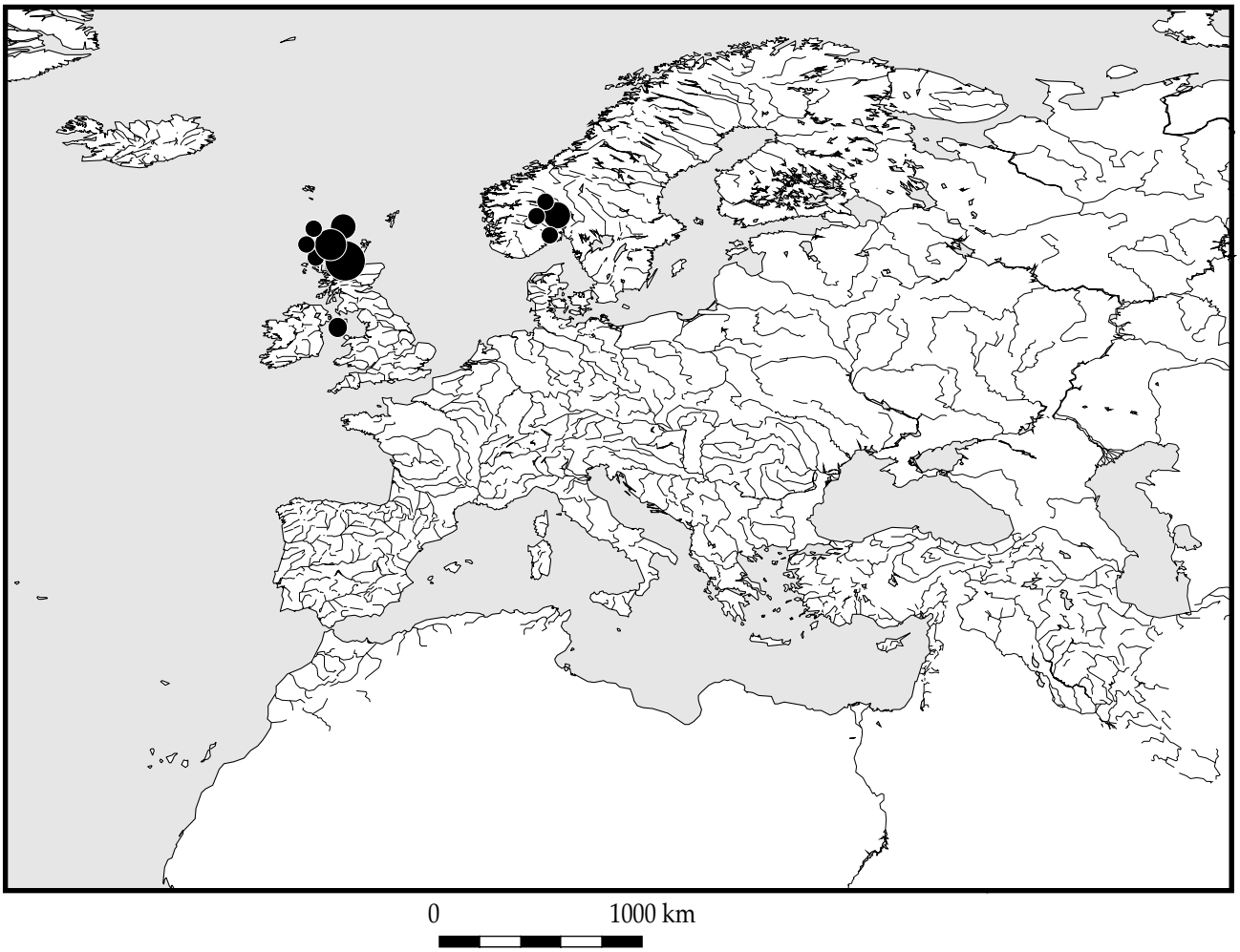

Figure 8.4. *Worldwide centre-of-gravity (COG) analysis of Icelandic mtDNA. Icelandic COGs with high geographic specificity (standard deviation <400 km) were found only in Europe (in the British Isles and in Norway), and hence the rest of the world is not displayed. The size of each circle corresponds to the number of Icelanders with that particular mtDNA type. To generate this plot, Icelandic mtDNA was deleted from the data base, and then closest genetic matches to 447 Icelandic mtDNAs were searched in the rest of the world. COGs were calculated from these matches, and the most specific COGs (standard deviation <400 km) are displayed on the map. COGs are calculated as in P. Forster* et al. *(2002), and are based on at least two closest matches in the mtradius data base. The clustering of Norwegian COGs exclusively in Oslo reflects the fact that other parts of Norway have hardly been sampled to date, apart from Norwegian Saami.*

wegian and Icelandic matches in contrast evidently are true relatives of the British J/16192 types, as they are quite frequent in both these Scandinavian countries (Table 8.1), and Viking raids on the British Isles are known to have occurred since the settlement of Iceland in AD 876. These raids presumably would have involved the abduction of British women and thus British mtDNA (Helgason *et al.* 2001; Arnason *et al.* 2000). What is perhaps surprising is the profound genetic effect of these Viking raids on the mtDNA pool of Iceland: there is more Celtic than Germanic mtDNA type J in Iceland today (Table 8.1). In order to find out whether this Icelandic result might be some artefact peculiar to mtDNA type J, we performed a geographic analysis of all 447 sampled Icelandic mtDNA types, using the centre-of-gravity method (P. Forster *et al.* 2002). The result (Fig. 8.4) confirms that the female ancestry of Icelanders derives mainly from Norwegian and British women, and that the British women contributed a significant proportion if not the majority of Icelandic mtDNA, in agreement with Helgason *et al.* (2001).

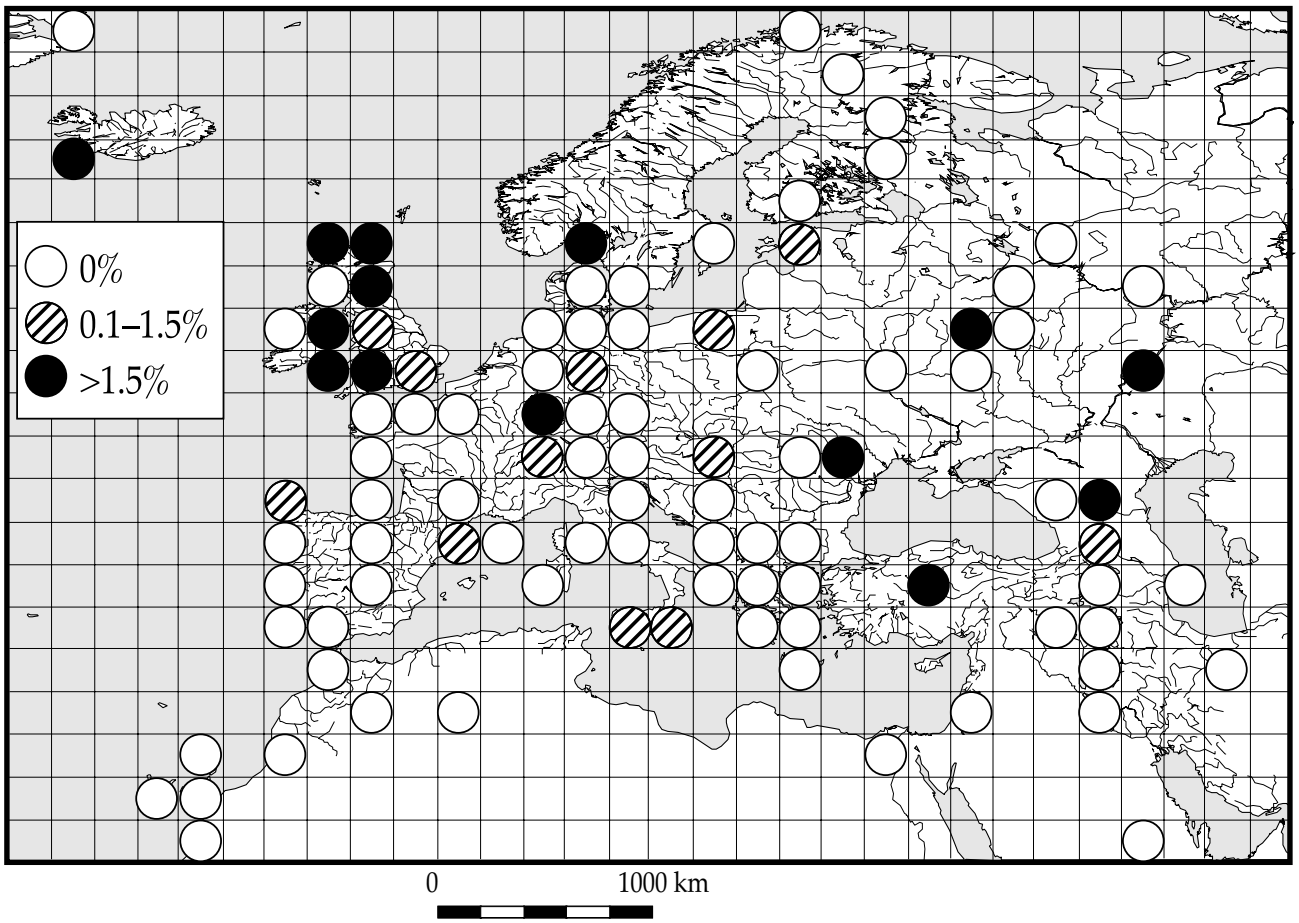

Figure 8.5. *Geographic distribution of mtDNA type J/16172. MtDNA type J/16192 is included because it is a subgroup of J/16172. In the mtradius data base used for this search, 19,493 individuals from native populations worldwide were activated, of which 12,615 are included in this map. In total, 117 individuals have the J/16172 type. Nearly all of these 117 are shown on this map, except for one individual from the Talas valley in northern Kyrgyzstan and one Indian from Andhra Pradesh. For further explanations, consult Figure 8.3.*

Table 8.1. *Frequencies of 'Celtic' and 'Germanic' J types.*

		'Celtic'		'Germanic'
	sample size[1]	**J/16192**	**J/16172**	**J/16231**
England	143	0	0	3
Denmark	50	0	0	2
N Germany[2]	97	0	0	4
Wales	92	2	3	0
Cornwall	105	1	2	0
N&W Scotland[3]	619	10	11	0
mainland Scotland[4]	673	16	22	4
Belfast	34	1	1	0
W Ireland	100	0	0	0
Iceland	447	12	18	1
Norway	543	5	9	5

[1] np16093–16323 in mtradius
[2] Saxon and Frisian placename area as defined in Forster (1995)
[3] Western Isles, Orkneys, NW Scottish coast
[4] mainland Scotland minus NW coast

Note: this table is more stringent than the maps in that forensic DNA data of uncertain provenance are excluded here.

The Icelandic finding contributes to a growing body of genetic evidence (if any genetic evidence were needed) that tribal confrontation between human males has at its core the reproductive control of women, as seen in Brazil (Alves-Silva *et al.* 2000), Central America (Torroni *et al.* 1994), Polynesia (Hurles *et al.* 1998), Greenland (Bosch *et al.* 2003) and the Caribbean (BBC 2003).

Two questions can be asked about the Insular Celtic J/16192 type: when did it arrive in the British Isles, and where did it come from? The geographic origin of J/16192 can in theory be traced by consulting its immediately ancestral mtDNA type J/16172, included in the evolutionary tree of Figure 8.2. Unfortunately, as seen in Figure 8.5, the geographic spread of the J/16172 lineage is rather diffuse, namely all over the Near East and Europe at low frequency.

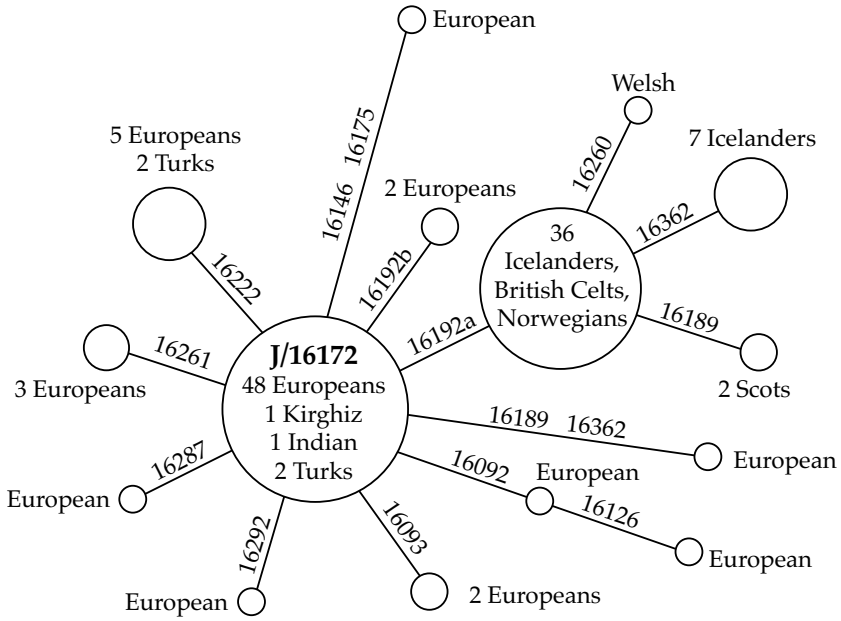

Figure 8.6. *Evolutionary tree of mtDNA type J/16172. This network is the basis for genetic time estimation as explained in the Methods section. Each circle represents an mtDNA type, with the circle area proportional to the number of individuals with that type. The largest circle labelled 'J/16172' is the ancestral root type for the evolutionary tree shown here. Lines represent mtDNA mutation events and are labelled by the mutant nucleotide position. Only mtDNA mutations between np16090 and np16365 are considered in the diagram. The British/Viking-specific subtype J/16192 is the large cluster to the right, designated by the DNA mutation event 16192a. As explained in the text, two Eastern Europeans are assumed to have undergone an independent mutation at np16192, designated here by '16192b'. Non-Europeans (Turks etc.) were excluded from the time estimation in order to focus on European events.*

The probable reasons for this diffuse spread are twofold. Biologically, np16172 is prone to parallel mutations (L. Forster *et al.* 2002) and pseudo-matches may arise by new mutations. Statistically, the frequency of J/16172 is generally very low on the Continent, making local frequency estimations even more erratic than in the other frequency maps presented here. Evidently, better sampling coverage of the Continent is required before firm conclusions can be drawn for the geographic origins of the Insular Celtic J/16192 marker. Perhaps the strongest inference that can be drawn from the available geographic distribution (Fig. 8.5) concerns time scale. The settlement of the Celtic areas of Britain by the J/16192 marker seems to be of considerable antiquity (possibly much more than just two or three millennia), since the J/16192 type has had sufficient time to become a highly successful mtDNA type in all Celtic parts of Britain. Conversely, its ancestral J/16172 type equally must have had ample time to become rare or extinct in potential source areas facing Britain (especially France). Alternatively, the crossing of J/16172 or J/16192 to Britain may have happened at a time in European post-glacial prehistory when low population densities meant that movements even of small groups of women would have had a disproportionate genetic impact. To throw light on these possibilities, we carried out genetic dating based on Figure 8.6 as described in the Methods section, which indicates an absolute age estimate for J/16192 of 4000 years which corresponds to or postdates its arrival time in the British Isles if it arose in the British Isles. The standard error (SE) on this estimate is very large and amounts to 3000 years. On the other hand, if J/16192 was already present in the Continent and then imported into Britain, then the upper age estimate for its arrival in the British Isles would be given by the age of its ancestral lineage J/16172 in Europe, which amounts at the very least to 6000 years (SE: 2000 years). The standard errors are unfortunately very high, so we need to look at further mtDNA markers to decide how ancient the presence of Celtic mtDNA markers are in Britain.

One such further mtDNA marker for British Celtic areas is J/16193 (found in 133 individuals out of 19747 worldwide), also included in the evolutionary tree of mtDNA type J in Figure 8.2. The geographic spread ranges from the Middle East across the Mediterranean (note the lack in the Basque country) to the Goidelic-speaking parts of Britain and Ireland (Fig. 8.7). The map shows that J/16193 is also found, at lower frequencies, in other parts of Europe, but the map in fact understates the rarity of J/16193 in Germanic areas: J/16193 in Germanic areas is present only in forensic samples which reflect the current mixed population rather than individuals genealogically traced for at least two generations, as is usual in anthropological samples. The evolutionary tree of J/16193 (not shown) consists of an ances-

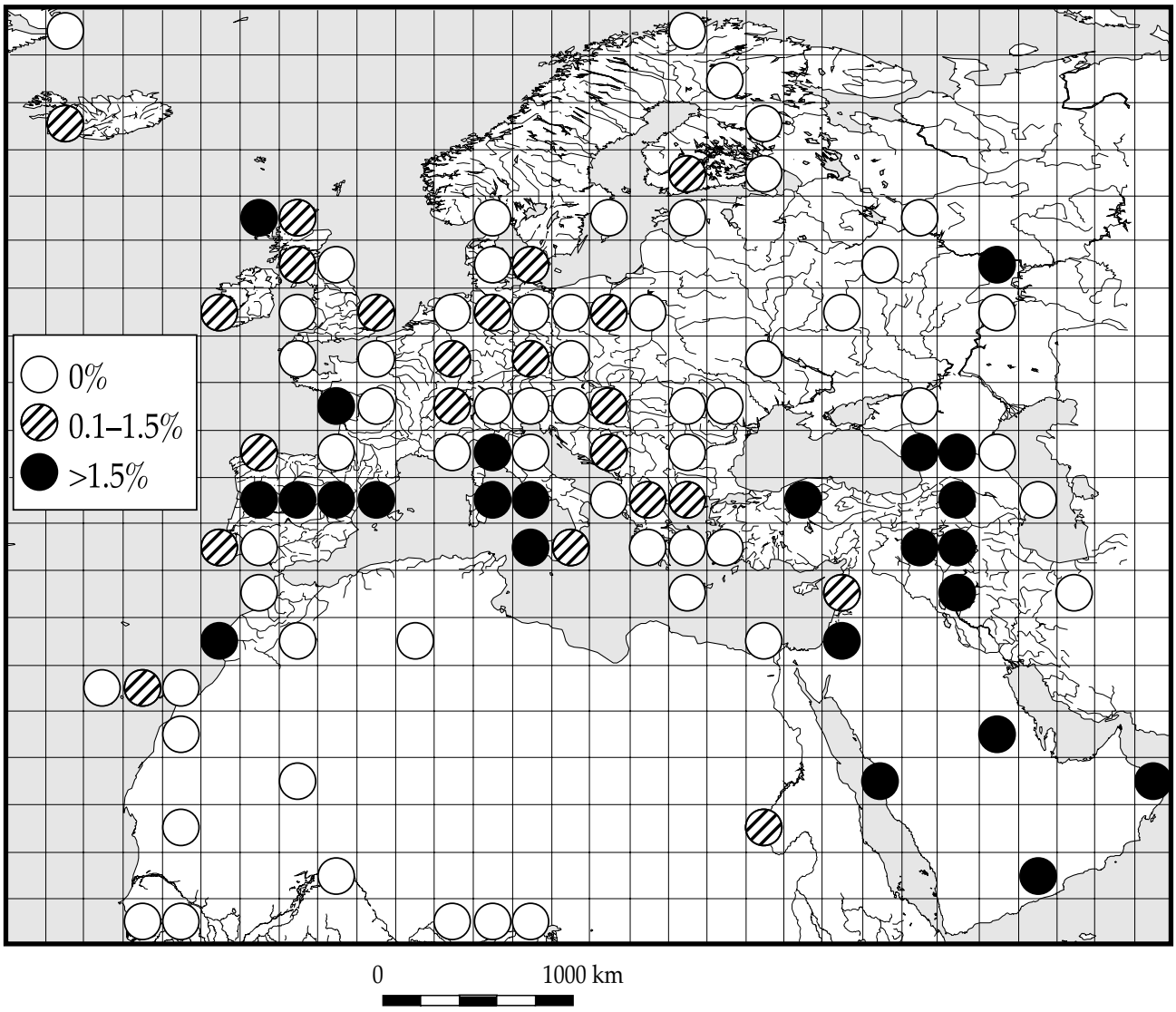

Figure 8.7. *Geographic distribution of mtDNA type J/16193. In the mtradius data base used for this search, 19,493 individuals from native populations worldwide were activated, of which 13,282 are included on this map. In total, 133 individuals have the J/16193 type. Nearly all of these 133 are shown on this map, except for one individual from the Sudanic-speaking Datoga tribe in Tanzania, and another individual from Sary-Tash in the Pamir mountains of southern Kyrgyzstan. For further explanations, consult Figure 8.3.*

tral type which is common in the Middle East and Europe, and a derived type J/16278 which is nearly exclusively European. Via genetic dating we obtain a slightly underestimated age for the arrival of J/16193 in Europe of 7000 years (SE: 2000 years) by discarding non-Europeans and considering only the ancestral type minus the J/16278 type. A maximal (and unrealistic) age estimate for the arrival of J/16193 in Europe would include the J/16278 type and the non-Europeans, and amounts to 17,000 years (SE: 5000 years). We can obtain another minimal age estimate for the arrival of J/16193 in Europe by dating the European-specific J/16278 branch, which amounts to 4000 years (SE: 2000 years). In summary, the age estimates favour a presence of J/16193 in Europe very approximately 7000 years ago, and the specifically coastal distribution of J/16193 along the Mediterranean all the way to the Celtic parts of the British Isles indicates that the spread of J/16193 by prehistoric women was a relatively coherent and rapid process.

Turning now to Germanic mtDNA markers in Britain, or more precisely, mtDNA markers for historically Germanic-speaking areas, the most clear-

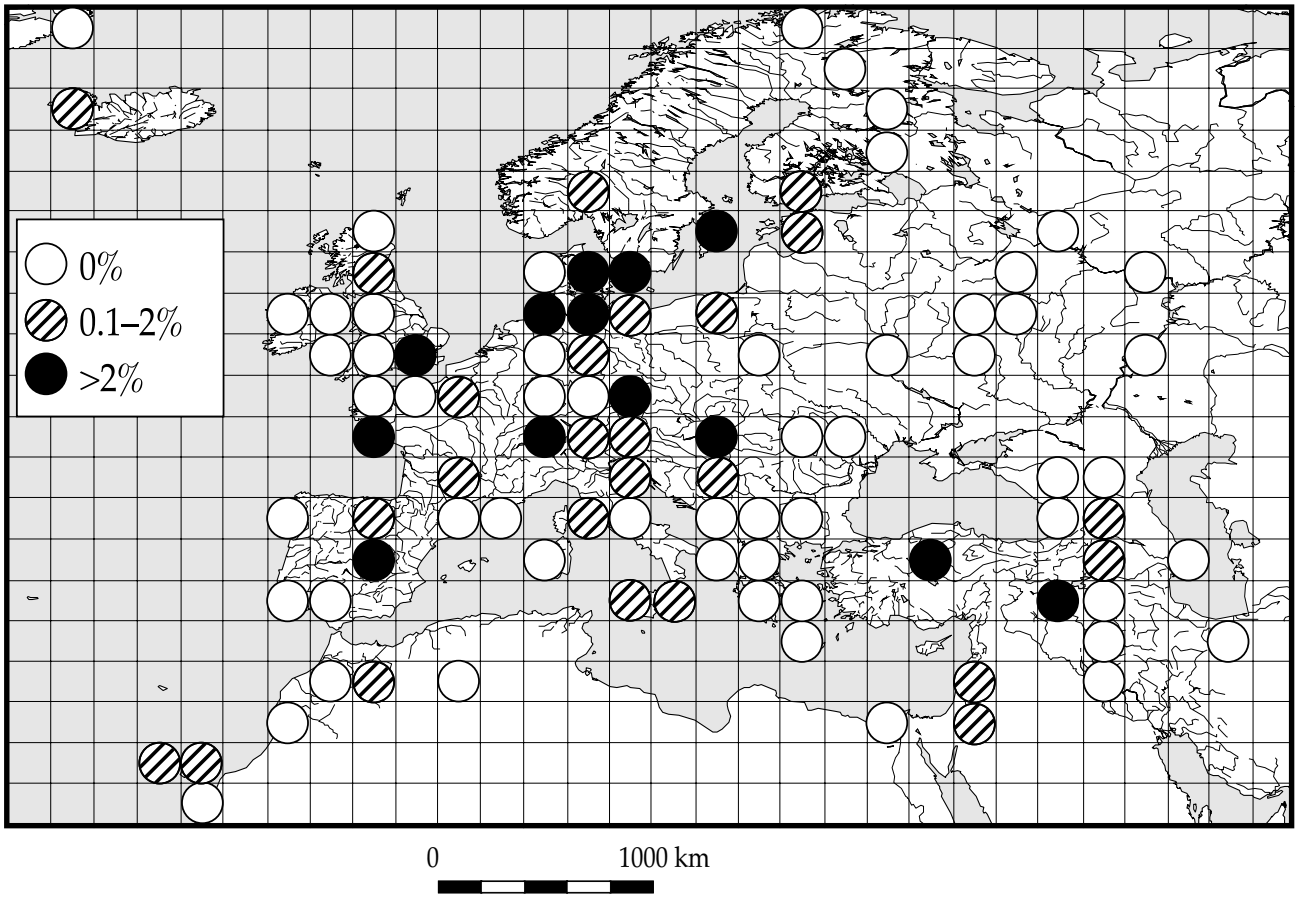

Figure 8.8. *Geographic distribution of mtDNA type J/16231. In the mtradius data base used for this search, 19,493 individuals from native populations worldwide were activated, of which 12,573 are included on this map. In total, 91 individuals have the J/16231 type and are all shown on the map. There are no J/16231 types elsewhere in the world, according to the current data base. For further explanations, consult Figure 8.3.*

cut Germanic mtDNA type within group J is type J/16231. Within the British Isles, J/16231 has been found so far only in traditionally English-speaking areas of England and mainland Scotland, and on the European continent J/16231 is found predominantly in and around Germanic-speaking areas in central and northwestern Europe (Table 8.1 & Fig. 8.8). The mixed Scottish mainland sample consists of both English-speaking and formerly Celtic-speaking areas and thus predictably has both the 'Germanic' and the 'Celtic' J types (Table 8.1). The starlike genealogy (not shown) of J/16231 types today indicates that this mtDNA type increased in number when it was carried into Europe by prehistoric women. The expansion of this type in Europe dates to 5000 years ago, with a high standard error of 3000 years.

Where could the Germanic J/16231 branch in England and Lowland Scotland have come from? An obvious source at first glance might be the traditional Anglo-Saxon settlement of Britain, with Angles and Saxons originating from northwest Germany. However, the Low-German-speaking ('Saxon') areas of the North German Plain harbour a 'Saxon' mtDNA marker H/16189 at about 25 per cent (16/61, updated from Richards *et al.* 1995, although the phylogenetic coherence of this cluster remains to be evaluated), which is rare in England where there is a frequency of only 3.5 per cent (5/143 in the English samples of Helgason *et al.* 2001 and Anderson *et al.* 1981). This low proportion indicates a contribution of zero to maximally 25 per cent of north German women to the native population of England (binomial distribution, 95 per cent confidence). Therefore other potentially Germanic tribes contributing women to the current English mitochondrial DNA pool may have to be considered, such as Jutes, Frisians, and Belgae according to Beda, Procopius and Caesar, respectively. In this context it is note-

worthy that Germans living close to the Dutch border harbour a low percentage of the Saxon marker (6/109 in the data of Pfeiffer *et al.* 1999) which is about as low as the English value of 3.5 per cent. Scant available mtDNA data from Jutland confirm a low frequency of the Saxon marker there, while the neighbouring Benelux countries and northern France, formerly home to the Frisians and Belgae, have not yet been studied for mtDNA.

Discussion

The traditional hypotheses on the arrival of Celtic-speakers and Germanic-speakers to the British Isles do not sit easily with the data from female-born mtDNA presented here. In the traditional, but not uncontested (Renfrew 1987) view, 'Celts' ultimately originating from a peri-Alpine Hallstatt/La Tène culture would have arrived in the British Isles around 600 BC. Neither this date nor an Alpine source area are entirely satisfactory from an mtDNA perspective. We tentatively place the arrival time for Insular Celtic mtDNA markers in the British Isles at thousands rather than hundreds of years BC. Furthermore the mtDNA profile of the 'Celtic' Alps is the opposite of the British Celtic profile as far as J is concerned: in the peri-Alpine region, Insular Celtic J types are absent or rare compared to the Germanic J/16231 marker which is clearly present in the Alps (see Fig. 8.8). As concerns the traditional view that northwest German Angles and Saxons arrived in Britain soon after the departure of the Romans around AD 410, our mtDNA survey reveals a paucity of the northwest German marker H/16189 in England. England does however yield an appreciable percentage of the general Germanic J/16231 marker, for which Germanic tribes other than northwest German Angles and Saxons may well have been responsible, and possibly centuries before the Anglo-Saxon period. The general Germanic marker J/16231 incidentally appears to have a considerable time depth of perhaps 5000 years.

One could argue with some justification that the genetic data are at present too imprecise to deliver reliable dates and geographic origins for fine-grained linguistic studies, and quite reasonably one could go even further and claim that female migrations are largely irrelevant to language spread. Nevertheless, on the basis of the current limited genetic evidence, a Neolithic timescale for the initial spread of Indo-European languages such as Germanic and Celtic within Europe (Renfrew 1987) appears at least as likely as the traditionally assumed shallower time depth. Improved sampling and longer DNA sequences are needed to address the issue of imprecision and to shed further light on the potentially 'Celtic' and 'Germanic' mtDNA markers presented here.

Methods

Nomenclature

When referring to branches (also known as 'lineages', 'clades' or 'haplogroups') in the mtDNA tree, the mtDNA phylogenetic nomenclature initiated by Torroni *et al.* (1993) and updated by Macaulay *et al.* (1999), Richards & Macaulay (2000), and Kivisild *et al.* (2003) is employed. Each branch consists of a number of mtDNA types, either extinct or living. The types discussed in this paper are named after the branches in which they lie. The numbering of a nucleotide position (np) follows the Cambridge reference sequence published by Anderson *et al.* (1981).

Geographic data base

The data base mtradius (Röhl *et al.* 2001), currently containing over 24,000 individuals, was used to quantify and visualize the geographic spread of J types, and to carry out a centre-of-gravity analysis for the 447 Icelandic mtDNAs. For the J analyses, a sequence range of minimally nps16093–16323 was selected, leaving a total of 19,493 individuals active in the data base, of which roughly 12,000 were from Europe and surrounding areas. For the centre-of-gravity analysis, the sequence range was set to nps16093–16362 and Canary Islanders were excluded, which left 17,917 individuals active in the data base. The frequency grid size was set to 4 degrees of longitude by 4 degrees of latitude.

Genetic dating

For age estimation of an ancestral mtDNA type, first the evolutionary tree of all available mtDNA types is reconstructed, typically using a phylogenetic network method. This reconstructed tree contains both present mtDNA types as well as reconstructed ancestral types. Next, the researcher identifies the ancestral mtDNA type in which he is interested (for example, an expansion type, a founder type, or a disease type). The age of the ancestral type is obtained by equating the average length of the descendant branches with elapsed time, measured in number of mutations (Morral *et al.* 1994). A standard error on each date is calculated according to Saillard *et al.* (2000). This error reflects the branching structure; for example, five different branches each leading to one descendant yield a more reliable time estimate than one branch leading to five identical

descendants. Relative 'mutational' time is then converted to absolute time by multiplying it with the mtDNA mutation rate as estimated in Forster *et al.* (1996). Phylogenetic dating software (shareware) is available at www.fluxus-engineering.com. The mutation rate is the Achilles' heel for any absolute DNA chronology. Whereas relative genetic dates and their relative standard errors (both expressed in mutations) are by definition accurate, their conversion to accurate absolute dates (expressed in years) depends entirely on an accurate calibration of the mtDNA mutation rate. Should an improved mtDNA mutation rate become available in the future, all dates presented in this paper can be proportionately adjusted.

Acknowledgements

We thank Lucy Forster, Eleanore Conant and Marta Lahr (Cambridge) for providing access to unpublished mtDNA data, and Alda Ragalmuto and Valeria Chiavetta (Troina) for their technical assistance during data processing. We also thank Martin Richards (Leeds) and Mim Bower (Cambridge) for critically reading the manuscript.

References

Alves-Silva, J., M. da Silva Santos, P.E. Guimaraes, A.C. Ferreira, H.-J. Bandelt, S.D. Pena & V.F. Prado, 2000. The ancestry of Brazilian mtDNA lineages. *American Journal of Human Genetics* 67, 444–61.

Anderson, S., A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden & I.G. Young, 1981. Sequence and organisation of the human mitochondrial genome. *Nature* 290, 457–65.

Arnason, E., H. Sigurgislason & E. Benedikz, 2000. Genetic homogeneity of Icelanders: fact or fiction? *Nature Genetics* 25, 373–4.

Bandelt, H.-J., V. Macaulay & M. Richards, 2002. What molecules can't tell us about the spread of languages and the Neolithic, in *Examining the Farming/Language Dispersal Hypothesis*, eds. P. Bellwood & C. Renfrew. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research, 89–97.

BBC, 2003. *Motherland: a Genetic Journey*. [Documentary on the ancestry of Jamaican descendants of African slaves.]

Bosch, E., F. Calafell, Z.H. Rosser, S. Nørby, N. Lynnerup, M.E. Hurles & M.A. Jobling, 2003. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Human Genetics* 112, 353–63.

Capelli, C., N. Redhead, J.K. Abernethy, F. Gratrix, J.F. Wilson, T. Moen, T. Hervig, M. Richards, M.P.H. Stumpf, P.A. Underhill, P. Bradshaw, A. Shaha, M.G. Thomas, N. Bradman & D.B. Goldstein, 2003. A Y-chromosome census of the British Isles. *Current Biology* 13, 979–84.

Dubut, V., L. Chollet, P. Murail, F. Cartault, E. Beraud-Colomb, M. Serre & N. Mogentale-Profizi, in press. MtDNA polymorphisms in five French groups: importance of regional sampling. *European Journal of Human Genetics*.

Faltings, V.F., A.G.H. Walker & O. Wilts (eds.), 1995. *Northwestern European Language Evolution*, Suppl vol. 12. Odense: University of Odense.

Forster, L., P. Forster, S. Lutz-Bonengel, H. Willkomm & B. Brinkmann, 2002. Natural radioactivity and human mitochondrial DNA mutations. *Proceedings of the National Academy of Sciences of the USA* 99, 13,950–54.

Forster, P., 1995. Einwanderungsgeschichte Norddeutschlands [Immigration history of North Germany], in Faltings *et al.* (eds.), 141–63. [English translation available at www.mcdonald.cam.ac.uk/genetics.]

Forster, P. & A. Toth, 2003. Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *Proceedings of the National Academy of Sciences of the USA* 100, 9079–84.

Forster, P., R. Harding, A. Torroni & H.-J. Bandelt, 1996. Origin and evolution of native American mtDNA variation: a reappraisal. *American Journal of Human Genetics* 59, 935–45.

Forster, P., A. Torroni, C. Renfrew & A. Röhl, 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. *Molecular Biology and Evolution* 18, 1864–81.

Forster, P., F. Calì, A. Röhl, E. Metspalu, R. D'Anna, M. Mirisola, G. De Leo, A. Flugy, A. Salerno, G. Ayala, A. Kouvatsi, R. Villems & V. Romano, 2002. Continental and subcontinental distributions of mtDNA control region types. *International Journal of Legal Medicine* 116, 99–108.

Helgason, A., E. Hickey, S. Goodacre, V. Bosnes, K. Stefansson, R. Ward & B. Sykes, 2001. MtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *American Journal of Human Genetics* 68, 723–37.

Hurles, M.E., C. Irven, J. Nicholson, P.G. Taylor, F.R. Santos, J. Loughlin, M.A. Jobling & B.C. Sykes, 1998. European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *American Journal of Human Genetics* 63, 1793–806.

Kivisild, T., S. Rootsi, M. Metspalu, S. Mastana, K. Kaldma, J. Parik, E. Metspalu, M. Adojaan, H.V. Tolk, V. Stepanov, M. Golge, E. Usanga, S.S. Papiha, C. Cinnioglu, R. King, L.L. Cavalli-Sforza, P.A. Underhill & R. Villems, 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics* 72, 313–32.

Macaulay, V., M. Richards, E. Hickey, E. Vega, F. Cruciani,

V. Guida, R. Scozzari, B. Bonné-Tamir, B. Sykes & A. Torroni, 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *American Journal of Human Genetics* 64, 232–49.

Morral, N., J. Bertranpetit, X. Estivill, V. Nunes, T. Casals, J. Gimenez, A. Reis, R. Varon-Mateeva, M. Macek Jr, L. Kalaydjieva, D. Angelicheva, R. Dancheva, G. Romeo, M.P. Russo, G. Garnerone, M. Ferrari, C. Magnani, M. Claustres, M. Desgeorges, M. Schwartz, M. Schwarz, B. Dallapiccola, G. Novelli, C. Ferec, M. de Arce, M. Nemeti, J. Kere, M. Anvret, N. Dahl & L Kadasi, 1994. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nature Genetics* 7, 169–75.

Pfeiffer, H., B. Brinkmann, J. Huhne, B. Rolf, A.A. Morris, R. Steighner, M.M. Holland & P. Forster, 1999. Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *International Journal of Legal Medicine* 112(5), 291–8.

Poloni, E.S., O. Semino, G. Passarino, A.S. Santachiara-Benerecetti, I. Dupanloup, A. Langaney & L. Excoffier, 1997. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *American Journal of Human Genetics* 61, 1015–35.

Renfrew, C., 1987. *Archaeology and Language*. London: Jonathan Cape.

Richards, M. & V. Macaulay, 2000. Genetic data and the colonization of Europe: genealogies and founders, in *Archaeogenetics: DNA and the Population Prehistory of Europe*, eds C. Renfrew & K. Boyle. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research, 139–51.

Richards, M.B., P. Forster, S. Tetzner, R. Hedges & B.C. Sykes, 1995. Mitochondrial DNA and the Frisians, in Faltings *et al.* (eds.), 141–63.

Richards, M.B., H. Côrte-Real, P. Forster, V. Macaulay, H. Wilkinson-Herbots, A. Demaine, S. Papiha, R. Hedges, H.-J. Bandelt & B.C. Sykes, 1996. Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. *American Journal of Human Genetics* 59, 185–203.

Richards, M.B., V. Macaulay, B.C. Sykes, P. Pettitt , P. Forster & H.-J. Bandelt, 1997. Palaeolithic and Neolithic lineages in the European mitochondrial gene pool: reply to Cavalli-Sforza and Minch. *American Journal of Human Genetics* 61, 247–55.

Richards, M., V. Macaulay, E. Hickey, E. Vega, B. Sykes, V. Guida, C. Rengo, D. Sellitto, F. Cruciani, T. Kivisild, R. Villems, M. Thomas, S. Rychkov, O. Rychkov, Y. Rychkov, M. Golge, D. Dimitrov, E. Hill, D. Bradley, V. Romano, F. Calì, G. Vona, A. Demaine, S. Papiha, C. Triantaphyllidis, G. Stefanescu, J. Hatina, M. Belledi, A. Di Rienzo, A. Novelletto, A. Oppenheim, S. Nørby, N. Al-Zaheri, S. Santachiara-Benerecetti, R. Scozari, A. Torroni & H.-J. Bandelt, 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *American Journal of Human Genetics* 67, 1251–76.

Röhl, A., B. Brinkmann, L. Forster & P. Forster, 2001. An annotated mtDNA database. *International Journal of Legal Medicine* 115, 29–39.

Rosser, Z.H., T. Zerjal, M.E. Hurles, M. Adojaan, D. Alavantic, A. Amorim, W. Amos, M. Armenteros, E. Arroyo, G. Barbujani, G. Beckman, L. Beckman, J. Bertranpetit, E. Bosch, D.G. Bradley, G. Brede, G. Cooper, H.B. Corte-Real, P. de Knijff, R. Decorte, Y.E. Dubrova, O. Evgrafov, A. Gilissen, S. Glisic, M. Golge, E.W. Hill, A. Jeziorowska, L. Kalaydjieva, M. Kayser, T. Kivisild, S.A. Kravchenko, A. Krumina, V. Kucinskas, J. Lavinha, L.A. Livshits, P. Malaspina, S. Maria, K. McElreavey, T.A. Meitinger, A.V. Mikelsaar, R.J. Mitchell, K. Nafa, J. Nicholson, S. Nørby, A. Pandya, J. Parik, P.C. Patsalis, L. Pereira, B. Peterlin, G. Pielberg, M.J. Prata, C. Previdere, L. Roewer, S. Rootsi, D.C. Rubinsztein, J. Saillard, F.R. Santos, G. Stefanescu, B.C. Sykes, A. Tolun, R. Villems, C. Tyler-Smith & M.A. Jobling, 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American Journal of Human Genetics* 67, 1526–43.

Saillard, J., P. Forster, N. Lynnerup, H.-J. Bandelt & S. Nørby, 2000. MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *American Journal of Human Genetics* 67, 718–26.

Sajantila, A., P. Lahermo, T. Anttinen, M. Lukka, P. Sistonen, M.L. Savontaus, P. Aula, L. Beckman, L. Tranebjaerg, T. Geddedahl, L. Isseltarver, A. DiRienzo & S. Pääbo, 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Research* 5, 42–52.

Semino, O., G. Passarino, L. Quintana-Murci, A. Liu, J. Béres, A. Czeizel & A. Silvana, 2000. MtDNA and Y chromosome polymorphisms in Hungary: inferences from the Palaeolithic, Neolithic and Uralic influences on the modern Hungarian gene pool. *European Journal of Human Genetics* 8(5), 339–46.

Simoni, L., F. Calafell, D. Pettener, J. Bertranpetit & G. Barbujani, 2000. Geographic patterns of mtDNA diversity in Europe. *American Journal of Human Genetics* 66, 262–78.

Torroni, A., T.G. Schurr, M.F. Cabell, M.D. Brown, J.V. Neel, M. Larsen, D.G. Smith, C.M. Vullo & D.C. Wallace, 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *American Journal of Human Genetics* 53, 563–90.

Torroni, A., Y.S. Chen, O. Semino, A.S. Santachiara-Beneceretti, C.R. Scott, M.T. Lott, M. Winter & D.C. Wallace, 1994. MtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *American Journal of Human Genetics* 54, 303–18.

Torroni, A., K. Huoponen, P. Francalacci, M. Petrozzi, L. Morelli, R. Scozzari, D. Obinu, M.L. Savontaus &

D.C. Wallace, 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144, 1835–50.

Torroni, A., M. Richards, V. Macaulay, P. Forster, R. Villems, S. Nørby, M.L. Savontaus, K. Huoponen, R. Scozzari & H.-J. Bandelt, 2000. MtDNA haplogroups and frequency patterns in Europe. *American Journal of Human Genetics* 66, 1173–7.

Weale, M.E., D.A. Weiss, R.F. Jager, N. Bradman & M.G. Thomas, 2002. Y chromosome evidence for Anglo-Saxon mass migration. *Molecular Biology and Evolution* 19, 1008–21.

Wilson, J.F., D.A. Weiss, M. Richards, M.G. Thomas, N. Bradman & D.B. Goldstein, 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proceedings of the National Academy of Sciences of the USA* 98, 5078–83.

Y Chromosome Consortium, 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12, 339–48.

Zerjal, T., L. Beckman, G. Beckman, A.V. Mikelsaar, A. Krumina, V. Kucinskas, M.E. Hurles & C. Tyler-Smith, 2001. Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in northern European populations. *Molecular Biology and Evolution* 18, 1077–87.