

Europe and western Asia: genetics and population history

Mark G. Thomas, Toomas Kivisild,
Lounes Chikhi, and Joachim Burger

This chapter examines the genetic history of Europe during the past 20,000 years, looking especially at postglacial recolonization and at the much-debated issue of the origins of the Neolithic populations of the continent. The authors also discuss the different ways geneticists can approach population history, and comment on the importance of natural selection for genes that determine skin pigmentation and enhance the ability to digest the milk sugar lactose.

On a continent-wide scale, genetic data have been used extensively to address three major questions on the demographic history of Europe: (1) What was the relationship between archaic hominins and modern Europeans (discussed in chapter 4); (2) To what extent did the formation and subsequent re-expansion of Ice Age refuge populations shape patterns of genetic variation; and (3) To what extent did Near Eastern farmers replace or interbreed with indigenous hunter-gatherers during the Neolithic transition?

A plentiful supply of samples, well-funded medical studies, and a rich history of archaeological and linguistic research have all contributed to an abundance of both modern and ancient genetic data from various inheritance systems (mitochondrial DNA, Y-chromosome and autosomes) in Europe. What remains more contentious is how such data should be used to infer migration history (Goldstein & Chikhi 2002; Nielsen & Beaumont 2009). While it is beyond the scope of this chapter to discuss genetic inference about population history in detail, we think it is important to outline the three main approaches used, and discuss their strengths and weaknesses. These are (1) phylogeographic analysis, (2) interpretation of summary patterns of genetic data, and (3) population genetic modeling.

Phylogeographic studies (Avise et al. 1987) aim at inferring major demographic events from the geographic distribution of phylogenetically identified clades. This approach is most commonly used with mitochondrial DNA (mtDNA) or Y-chromosome

data, where the clades are usually defined by some shared but otherwise arbitrarily chosen mutations, and are referred to as haplogroups. In humans, such studies have focused on the estimated ages of haplogroups and their spatial distributions. Inference usually proceeds by relating these ages to the presence of some ancient population in the region where they are most common, then treating their current distribution as the signature of an ancient demographic episode, such as an expansion or migration event (Underhill & Kivisild 2007). While there are issues with estimating the shape of a phylogenetic tree within a species and, more importantly, with defining clades and estimating their ages, the main problem is that mutation and the formation of genealogical trees are very “noisy” processes, and branches of a gene tree may actually be very poor proxies of ancestral populations. As a result, there is a tendency in phylogeographic analysis to attribute patterns in the data to interesting aspects of a population’s history, when in fact they may be the result of randomness in the mutation and inheritance processes (Nielsen & Beaumont 2009).

The interpretation of summary patterns is the oldest method of inferring a population’s history from genetic data, and includes approaches such as principal components analysis (Cavalli-Sforza et al. 1994) and the distribution of differences between samples of homologous DNA sequences (Rogers & Harpending 1992). Under ideal conditions, certain of these descriptive statistics can act as convenient proxies for processes such as migration, admixture, and changes in population size. However, because of the “noisy” or stochastic nature of inheritance (*drift*) and mutation, the same patterns of genetic data can arise under a range of very different population histories (*equifinality*), and if we were able to “run” a population’s history identically, multiple times, then each run would give rise to different patterns in the genetic data (evolutionary variance). Additionally, some seemingly simple demographic processes can give rise to surprising or misleading patterns in genetic data (emergence), such as allele surfing, whereby alleles unaffected by natural selection can “ride” on the advancing wave of an expanding population and reach highest frequencies in locations far removed from their origin (Edmonds et al. 2004). This process can also be mistaken for natural selection (Currat et al. 2006).

The final approach, explicit modeling, is statistically the most secure approach to population history inference (Beaumont et al. 2002; Nielsen & Beaumont 2009). Briefly, different population history scenarios are compared statistically, using either analytical or computer simulation results. One branch of population genetics theory that has been particularly useful here is coalescent theory, which is a retrospective model of gene genealogies for a sample under a defined population history, and allows genetic data to be simulated in an efficient manner (Wakeley 2009). Using the coalescent approach, the probability of getting data identical or similar to the real data can be calculated or estimated. The idea here is that the most likely historical scenarios, including values for migration rates and times of population expansion or contraction, are those under which the data observed are most probable. Using this approach, the problems associated with the “noisy” nature of inheritance and mutation can be accounted for explicitly. However, modeling is often difficult to implement for geneticists and non-geneticists alike. Also, the process usually begins with the simplest models since there are an infinite number of different plausible historical scenarios

that could be tested, and use of a complex model when a simple one will do is something of a statistical crime unless there are other prior data that support. However, simple models may appear naive or simplistic to many non-geneticists, and will not reveal the complexities of our demographic past unless the data are very information-rich.

All three approaches have contributed to our understanding of the migration history of Europe. But the first two are somewhat more “interpretative” and are better thought of as means of generating hypotheses from genetic data *post hoc*, rather than independently testing hypotheses that arise from other disciplines such as archaeology and linguistics. The last approach is undoubtedly the best way forward, but challenges remain in developing sufficiently flexible and usable software, and in convincing the wider community of its merits.

Pre-Neolithic contributions to the modern gene pool of Europe and West Asia

Genetic data overwhelmingly indicate that modern European, West Asian, and North African populations are derived from a dispersal of anatomically modern humans (*Homo sapiens*) from sub-Saharan Africa that took place between approximately 50 and 80 kya. However, recent analyses of ancient DNA suggest that Europeans and all other non-African populations can attribute 1–4 percent of their ancestry to early admixture with Neanderthals (Green et al. 2010), as discussed in chapter 4. It is most likely that this occurred before the separation of the East and West Eurasian gene pools because a similar signature is seen in all non-African populations, although Currat & Excoffier (2011) suggest that independent admixture events took place.

A number of phylogeographic studies have attempted to identify the founding mtDNA and Y chromosome lineages of the first *Homo sapiens* populations in Europe and to assess the contribution of these to the present day gene pool (Soares et al. 2010). However, due to genetic drift it is difficult to determine whether more than a handful of such lineages present around 25 kya are likely to have any descendants in Europeans today (Barbujani & Bertorelle 2001). Their current frequencies and diversity distributions have been shaped by random factors such as genetic drift, as well as by allele surfing, admixture, population size fluctuations and migration.

With the benefits of recent and reliable ancient DNA data, some predictions based on phylogeographic criteria (Richards et al. 2000) have found support. For instance, some mtDNA haplogroups, such as U, have been proposed to represent ancestry prior to the Last Glacial Maximum (LGM) in Europe, and haplogroup U2 was recently identified in the 30-kya Kostenki remains in Russia (Krause et al. 2010). Moreover, haplogroups U4 and U5 predominate in a post-LGM hunter-gatherer sample from central Europe (Bramanti et al. 2009) and in a Neolithic hunter-gatherer sample from Sweden (Malmstrom et al. 2009). However, other haplogroups suggested to be pre-Neolithic by Richards and colleagues (2000) are, on the contrary, rare or absent in early Neolithic samples, and some of those found to be common in the latter (e.g. N1a) are equally rare in modern Europeans (Haak et al. 2005; Barbujani & Chikhi 2006).

Attempts to identify Y-chromosome lineages or autosomal variants that might be considered signatures of the first settlers of Europe have met with similar problems, although phylogeographic inferences have been proposed (e.g. Semino et al. 2000). However, due to a lack of corroborating ancient DNA data, claims of Y-chromosome haplotype or autosomal allele signatures for the initial settlement of Europe should be treated with caution. Also, as for mtDNA, genetic drift is likely to make it difficult to make strong inferences without explicit modeling (Chikhi et al. 2002).

Archaeological and palaeoclimatic evidence indicate that the LGM at 24–20 kya led to an almost complete depopulation of northern and central Europe, and humans may have survived only in the glacial refuges of southern Europe and the northern Black Sea (Ramakrishnan & Hadly 2009). The recolonization of central Europe started in a period of post-LGM warming (Banks et al. 2008) that corresponded with the Upper Paleolithic Solutrean culture in France, the Magdalenian in middle Europe and the Epigravettian in south and southeast Europe. But due to the scarcity of ancient DNA evidence from 22–10 kya, it is not known exactly which refuge populations contributed to the overall post-LGM population structure (Street & Terberger 1999).

The distributions of mtDNA and Y variation in extant populations indicate a number of specific haplogroups that may have persisted since this period in a few refuge regions (Richards et al. 2000; Semino et al. 2000). For example, the Y-lineage R-M269, the most common lineage in Western Europe and found in over 110 million men, was initially proposed to have spread from Iberian Paleolithic refuge populations (Semino et al. 2000; Wilson et al. 2001). However, geographic patterns of diversity within this lineage have also been interpreted as indicating a West Asian origin in the Neolithic (Balaesque et al. 2010) – although recent analyses have in turn contested this (Busby et al. 2012). Such views have been debated widely, but ancient DNA studies on human remains of the Holocene period, as discussed below, tend not to favor genetic continuity from Paleolithic/Mesolithic into Neolithic in most parts of Europe.

Demographic events in the Holocene

The major multidisciplinary question concerning the Mesolithic to Neolithic transition in Europe (see chapters 17 and 20) is the extent to which immigrating food producers from the Near East contributed to the ancestry of Europeans after the transition, and until today (Renfrew 1987; Bellwood 2005; Pinhasi & Stock 2011). Unfortunately, and despite many claims to the contrary, genetic analyses have not fully resolved this question. But they have made important contributions and – particularly with recent developments in ancient DNA technology – promise much more in the near future.

One of the first major contributions to the Mesolithic-to-Neolithic debate was made using principal components analysis (PCA) of classical marker frequencies (blood groups and serological variants). In these pioneering and influential studies, spatial clines in major components of genetic variation were quite reasonably assumed to be signatures of past population expansions (Menozi et al. 1978; Cavalli-Sforza et al. 1993, 1994). The first component of variation in Europe showed a southeast to northwest cline – almost identical to the trajectory of the expansion of Neolithic farming

through Europe, and this was interpreted as supporting a major farmer contribution to European ancestry. However, recent simulation studies have revealed some surprising properties of these summary patterns of genetic variation; in particular that under a process of range expansion without admixture with local populations, the first principal component of variation can be perpendicular to the direction of expansion (François et al. 2010). Nonetheless, as we will see later, PCA still has a valuable role to play in understanding genetic history.

MtDNA evidence

Mitochondrial DNA variation in living populations in Europe, western Asia, and northern Africa is characterized by a number of distinct haplogroups (H, I, J, T, U, V, W, X) which all derive from one (N) of the two ancestral clades attributed to the out-of-Africa dispersal of *Homo sapiens* (see chapter 4). Although these haplogroups have pre-Holocene coalescent dates, their presence in North Africa and Europe does not necessarily indicate immigration going back that far. Haplogroups U, H, T, and J, which comprise the majority of European mtDNA lineages, are also found with a substantial degree of variation in western Asia; a likely source of multiple pulses of gene flow into Europe during both the Holocene and the Pleistocene. Although Richards et al. (2000) claimed, on the basis of founder analysis, that less than one-quarter of the mtDNA lineages in Europe had western Asian Neolithic origins, this rather low estimate has not found support from other population geneticists (e.g. Barbujani & Chikhi 2006).

Ancient DNA analysis of human remains from one of the most widespread early Neolithic cultures of Europe, the Linearbandkeramik (LBK, see chapter 20), found that 6 out of 24 samples carried mtDNA haplogroup N1a, which is extremely rare or absent in most modern populations in the region (Haak et al. 2005). The authors concluded that LBK populations were not the direct ancestors of modern Europeans. A later study, which also included mtDNA sequences from late Paleolithic to Neolithic European hunter-gatherer skeletons, as well as modern data, identified large genetic differences between all three groups. A simplified demographic model indicated that such large differences could not be explained by population continuity from Paleolithic times alone (Bramanti et al. 2009). The authors concluded that Neolithic farmers migrated into central Europe around 7,500 years ago.

Mitochondrial DNA sequences identical to those found in the central European LBK sample, as well as in modern Europeans and western Asians, were found in a Neolithic skeletal sample from Hungary. Interestingly, these Neolithic farmers also carried lineages that are present in modern Central and East Asia (Guba et al. 2011). Another ancient DNA study of skeletons of the Pitted Ware culture of southern Scandinavia, late hunter-gatherer people who lived alongside farmers, found a lack of support for population continuity into modern Scandinavians (Malmstrom et al. 2009). In contrast, phylogeographic interpretation of Iberian Late Neolithic ancient DNA data favored continuity between prehistoric and modern inhabitants (Sampietro et al. 2007) although a more recent study including early Neolithic samples favored a major contribution from Near Eastern farmers (Gamba et al. 2012).

As suggested using phylogeographic criteria (Richards et al. 2000), subtypes of haplogroup U have been predominant in all ancient European hunter-gatherer samples analyzed so far (Bramanti et al. 2009; Malmstrom et al. 2009; Krause et al. 2010), despite being relatively rare in Europeans today, except for the Uralic-speaking Sami. Since U4/5-types are present in modern Europeans but absent in Early Neolithic skeletal samples, it now seems likely that modern-day European ancestry has been influenced by regional variation in the extent of admixture between pre-Neolithic hunter-gatherers and Neolithic farmers (Burger 2010). However, spatially explicit modeling of Neolithic admixture that accommodates ancient DNA data is still pending.

Y-chromosome evidence

Similar to mtDNA variation, the Y-chromosome haplogroups in present-day Europe are distinct from those in sub-Saharan Africa and East Asia, but overlap considerably with Middle Eastern and North African populations. Six haplogroups, R1, I, N3, J, G, E3, account for more than 95 percent of the Y-chromosomes in the region (Semino et al. 2000). Notably, while mtDNA haplogroup composition is relatively homogenous across Europe (with the exception of the Uralic-speaking Sami), Y-chromosome variation is considerably more structured geographically, and haplogroup diversity is low in many regions. While attempts have been made to assign haplogroups to European Paleolithic or Near Eastern Neolithic origins (Semino et al. 2000, Soares et al. 2010) – including the widely debated R-M269 lineage (see above) – a lack of well-attested ancient DNA data means that such attempts should be considered speculative.

Despite an inability to confidently assign specific lineages to European Paleolithic or Near Eastern Neolithic origins, it is still possible to use Y-chromosome variation to make estimates of the relative contributions of these groups to modern Europeans. For instance, Semino and her colleagues (2000) considered the Y-chromosome haplogroups with Europe-wide clinal distributions as introduced to Europe by Neolithic farmers, whereas a majority of haplogroups (R1b, R1a, and I, using present-day nomenclature) were interpreted as being derived from pre-Holocene migrations. But instead of focusing on specific haplogroups, Chikhi and his colleagues (2002) used a Bayesian full-likelihood approach that explicitly modeled admixture and genetic drift to infer admixture variation across Europe. They suggested that the contribution of pre-Holocene Y-chromosomes varies enormously across Europe, with a geographic distribution that is best explained by a model of demic diffusion of Neolithic populations through the region (Cavalli-Sforza et al. 1994).

Genome-wide variation and natural selection in European populations

In recent years, the amount of genomic data on Europeans has increased by several orders of magnitude, thanks to improvements in genotyping technology and large-scale medical studies. One way of summarizing variation over many loci is to perform PCA, and when applied to a dataset from living Europeans consisting of around half

a million single nucleotide polymorphisms (SNPs), a plot of the first two components showed that geography has been a major factor in shaping patterns of genetic variation, although it should be noted that the first two components account for less than 1 percent of the total variation (Novembre et al. 2008).

Genetic variants that have unusually high frequencies in Europe when compared to other regions may have been influenced by natural selection. Two genes with clear functional significance stand out in this regard: *SLC24A5*, which is linked to skin colour, and *LCT*, linked to lactose digestion in adulthood (Sabeti et al. 2006). Variation in the *SLC24A5* gene has been shown to explain between 25 and 38 percent of the skin-color differences between individuals of European and African descent (Lamason et al. 2005). The derived allele associated with light skin color is found at near 100 percent in most European populations, but is rare elsewhere, regardless of skin color. It has been suggested that this adaptation favored increased vitamin D synthesis under the lower UVB radiation found at high latitudes (Chaplin & Jablonski 2009). Interestingly, light pigmentation-associated alleles in different genes are common in Asian populations who also have light skin color, indicating that light skin pigmentation has evolved multiple times independently. Both the high haplotype homozygosity and its restricted geographic range suggest that the selective sweep on *SLC24A5* occurred recently, perhaps during the Holocene with the shift from a vitamin D-rich (fish) hunter-gatherer diet to a vitamin D-poor farming diet (but see Craig et al. 2011).

Another gene with characteristic signatures of natural selection in Europeans is the lactase gene, *LCT*. A mutation near *LCT* (called $-13,910^*T$) is strongly associated with lactase persistence – the ability of some adults to digest the milk sugar lactose. The $-13,910^*T$ variant is common in Europe (Enattah et al. 2002), central Asia (Heyer et al. 2011), India (Gallego Romero et al. 2012), and some African populations (Mulcare et al. 2004). Lactase persistence has itself evolved multiple times and, in some sub-Saharan African and Middle Eastern populations, other genetic variants are responsible for the same phenotype. Interestingly, the $-13,910^*T$ variant is found at frequencies of more than 50 percent in most central and northern European populations today, yet was absent from a sample of early Neolithic individuals from the same region (Burger et al. 2007). This suggests that it rose to high frequency during or after the Neolithic. Itan and colleagues (2009), using spatially explicit computer simulations, estimate that the $-13,910^*T$ allele began to be selectively favored around 7,500 years ago in the Balkans and central Europe, possibly among the progenitors of the LBK culture.

Demographic processes after the Neolithic

In the millennia after the introduction of farming, the population of Europe is thought to have grown dramatically (Bellwood 2005). This would have reduced drift and the likelihood that subsequent demographic events could have had major impacts on the European gene pool. Nevertheless, there were further migrations from both within and outside the continent that are likely to have influenced genetic variation to some degree. One enduring question concerns the influence that highly mobile Late Neolithic groups from the Ukrainian steppes might have had on Europe (Rassamakin 2002; Anthony 2007). Some scholars see here not only the possible origins of Indo-European

languages (Gimbutas 1997) but also a major component of European ancestry. This is in contrast to Renfrew's farming-language dispersal hypothesis (Renfrew 1987), which proposes that Indo-European languages spread into Europe with immigrating Neolithic farmers from Anatolia (see chapter 19).

Very little genetic support is available for migrations from the Steppes into Europe, and the evidence is mostly restricted to population dynamics within the Steppes themselves (Lalueza-Fox et al. 2004; Pilipenko et al. 2010). The Hungarians or Magyars are an example of a linguistic isolate in the center of Europe, but genetic analyses consistently place them with their present-day geographic neighbors rather than with their Uralic linguistic relatives (see chapter 21). In such cases, genes and languages do not correlate with each other independently of geography, and therefore genetic data might not be well suited for testing hypotheses of language dispersal.

There is considerable debate on the extent of further migration associated with historically attested cultural transitions. These include Arab expansion, Viking colonization, and the various movements of Germanic peoples during and after the late Roman Empire. Of these, perhaps the most widely debated, in the UK at least, is the extent of migration during the Anglo-Saxon transition in Britain. Two genetic studies using Y-chromosome data and population genetic modeling concluded that the Anglo-Saxon contribution to the English gene pool was greater than 50 percent (Weale et al. 2002; Capelli et al. 2003), an estimate that has proved unpalatable for many archaeologists. Subsequent computer simulations, in combination with historical evidence, indicate that social structure, wealth differentials, and a degree of reproductive isolation between Anglo-Saxons and native Britons could explain this rather high male-line contribution, commencing with a less than 10 percent initial contribution during the migration phase itself (Thomas et al. 2006).

Conclusions

Most genetic variation seen in Europe today derives ultimately, as in other continents, from a Late Pleistocene dispersal of modern humans from Africa. Geography, rather than language, ethnic or cultural identity, appears to be the primary determinant of genetic affinity in Europe. However, ancient DNA studies do indicate genetic discontinuities during and following the Neolithic transition. Most of these changes are likely to have been caused by demographic processes on various scales, including migration, population growth and decline, and admixture, but variation at some loci has been shaped by natural selection. In the near future the application of new genome-wide sequencing technologies, particularly to ancient DNA, as well as the use of more flexible model-based statistical approaches, will hopefully enable us to test questions relating to the peopling of Europe more rigorously.

SEE ALSO: Vol. I: 4 Early Old World migrations of *Homo sapiens*: human biology; 17 Anatolia and the Balkans: archaeology; 19 Europe and western Asia: Indo-European linguistic history; 20 Europe: Neolithic colonization; 21 Northern Europe and Russia: Uralic linguistic migrations; Vol. II: Anglo-Saxon migrations

References

- Anthony, D. (2007) *The Horse, the Wheel, and Language: How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton: Princeton University Press.
- Avise, J., Arnold, J., Ball, R., et al. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18, 489–522.
- Balaresque, P., Bowden, G., Adams, S., et al. (2010) A predominantly Neolithic origin for European paternal lineages. *PLoS Biol* 8, e1000285.
- Banks, W., D’errico, F., Peterson, A., et al. (2008) Human ecological niches and ranges during the LGM in Europe derived from an application of eco-cultural niche modeling. *Journal of Archaeological Science* 35, 481–491.
- Barbujani, G. & Bertorelle, G. (2001) Genetics and the population history of Europe. *Proceedings of the National Academy of Sciences* 98, 22–25.
- Barbujani, G. & Chikhi, L. (2006) Population genetics: DNAs from the European Neolithic. *Heredity* 97, 84–85.
- Beaumont, M., Zhang, W., & Balding, D. (2002) Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Bellwood, P. S. (2005) *First Farmers: The Origins of Agricultural Societies*. Cambridge, MA: Blackwell.
- Bramanti, B., Thomas, M., Haak, W., et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe’s first farmers. *Science* 326, 137–140.
- Burger, J. (2010) Population genetics of the European Neolithic and the role of lactase persistence. In G. Grupe, G. McGlynn, & J. Peters (eds.), *Archaeobiodiversity. A European Perspective*. Rahden, Germany: Verlag Marie Leidorf.
- Burger, J., Kirchner, M., Bramanti, B., et al. (2007) Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proceedings of the National Academy of Sciences* 104, 3736–3741.
- Busby, G., Brisighelli, F., Sanchez-Diz, P., et al. (2012) The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences* 279, 884–892 Online ref. doi: 10.1098/rspb.2011.1044.
- Capelli, C., Redhead, N., Abernethy, J., et al. (2003) A Y chromosome census of the British Isles. *Current Biology* 13, 979–984.
- Cavalli-Sforza, L., Menozzi, P., & Piazza, A. (1993) Demic expansions and human evolution. *Science* 259, 639–646.
- Cavalli-Sforza, L., Menozzi, P., & Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton: Princeton University Press.
- Chaplin, G. & Jablonski, N. G. (2009) Vitamin D and the evolution of human depigmentation. *American J. Physical Anthropology* 139, 451–461.
- Chikhi, L., Nichols, R., Barbujani, G. & Beaumont, M. (2002) Y genetic data support the Neolithic demic diffusion model. *Proceedings of the National Academy of Sciences* 99, 11008–11013.
- Craig, O., Steele, V., Fischer, A., et al. (2011) Ancient lipids reveal continuity in culinary practices across the transition to agriculture in northern Europe. *Proceedings of the National Academy of Sciences* 108, 17910–17915.
- Curat, M. & Excoffier, L. (2011) Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proceedings of the National Academy of Sciences* 108, 15129–15134.

- Currat, M., Excoffier, L., Maddison, W., et al. (2006) Comment on “Ongoing adaptive evolution of *Aspm*, a brain size determinant in *Homo sapiens*” and “Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans.” *Science* 313, 172.
- Edmonds, C., Lillie, A., & Cavalli-Sforza, L. (2004) Mutations arising in the wave front of an expanding population. *Proceedings of the National Academy of Sciences* 101, 975–979.
- Enattah, N., Sahi, T., Savilahti, E., et al. (2002) Identification of a variant associated with adult-type hypolactasia. *Nature Genetics* 30, 233–237.
- François, O., Currat, M., Ray, N., et al. (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution* 27, 1257–1268.
- Gallego Romero, I., Basu Mallick, C., Liebert, A., et al. (2012) Herders of Indian and European cattle share their predominant allele for lactase persistence. *Molecular Biology and Evolution* 29, 249–260.
- Gamba, C., Fernández, E., Tirado, M., et al. (2012) Ancient DNA from an early Neolithic Iberian population supports a pioneer colonization by first farmers. *Molecular Ecology* 21, 45–56.
- Gimbutas, M. (1997) *The Kurgan Culture and the Indo-Europeanization of Europe*. Washington, DC: Institute for the Study of Man.
- Goldstein, D. & Chikhi, L. (2002) Human migrations and population structure: what we know and why it matters. *Annual Review of Genomics and Human Genetics* 3, 129–152.
- Green, R., Krause, J., Briggs, A., et al. (2010) A draft sequence of the Neandertal genome. *Science* 328, 710–722.
- Guba, Z., Hadadi, E., Major, A., et al. (2011) Hvs-I polymorphism screening of ancient human mitochondrial DNA provides evidence for N9a discontinuity and East Asian haplogroups in Neolithic Hungary. *Journal of Human Genetics*, 56, 784–796.
- Haak, W., Forster, P., Bramanti, B., et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites, *Science* 310, 1016–1018.
- Heyer, E., Brazier, L., Segurel, L., et al. (2011) Lactase persistence in central Asia: phenotype, genotype, and evolution. *Human Biology* 83, 379–392.
- Itan, Y., Powell, A., Beaumont, M., et al. (2009) The origins of lactase persistence in Europe. *PLoS Comput Biol*, 5, e1000491.
- Krause, J., Briggs, A., Kircher, M., et al. (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology* 20, 231–236.
- Lalueza-Fox, C., Sampietro, M., Gilbert, M. T., et al. (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. *Proceedings of the Royal Society B: Biological Sciences* 271, 941–947.
- Lamason, R., Mohideen, M., Mest, J., et al. (2005) *Slc24a5*, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* 310, 1782–1786.
- Malmstrom, H., Gilbert, M., Thomas, M., et al. (2009) Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology* 19, 1758–1762.
- Menozi, P., Piazza, A., & Cavalli-Sforza, L. (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201, 786–792.
- Mulcare, C. A., Weale, M. E., Jones, A. L., et al. (2004) The T allele of a single-nucleotide polymorphism 13.9kb upstream of the lactase gene (*Lct*) (C-13.9kbt) does not predict or cause the lactase-persistence phenotype in Africans. *American Journal of Human Genetics* 74, 1102–1110.
- Nielsen, R. & Beaumont, M. (2009) Statistical inferences in phylogeography. *Molecular Ecology* 18, 1034–1047.
- Novembre, J., Johnson, T., Bryc, K., et al. (2008) Genes mirror geography within Europe. *Nature* 456, 98–101.

- Pilipenko, A., Romaschenko, A., Molodin, V., et al. (2010) Mitochondrial DNA studies of the Pazyryk people (4th to 3rd centuries BC) from northwestern Mongolia. *Archaeological and Anthropological Sciences* 2, 231–236.
- Pinhasi, R. & Stock, J. (eds) (2011) *Human Bioarchaeology of the Transition to Agriculture*. Chichester, UK: Wiley.
- Ramakrishnan, U. & Hadly, E. (2009) Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology* 18, 1310–1330.
- Rassamakin, Y. (2002) Aspects of Pontic Steppe development (4550–3000 BC) in the light of the new cultural chronological model. In K. V. Boyle, C. Renfrew, & M. Levine (eds.), *Ancient Interactions: East and West in Eurasia*. Cambridge: McDonald Institute for Archaeological Research.
- Renfrew, C. (1987) *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Cape.
- Richards, M., Macaulay, V., Hickey, E., et al. (2000) Tracing European founder lineages in the near eastern mtDNA pool. *American Journal of Human Genetics* 67, 1251–1276.
- Rogers, A. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9, 552–569.
- Sabeti, P., Schaffner, S., Fry, B., et al. (2006) Positive natural selection in the human lineage. *Science* 312, 1614–1620.
- Sampietro, M., Lao, O., Caramelli, D., et al. (2007) Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proceedings of the Royal Society B: Biological Sciences* 274, 2161–2167.
- Semino, O., Passarino, G., Oefner, P., et al. (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science* 290, 1155–1159.
- Soares, P., Achilli, A., Semino, O., et al. (2010) The archaeogenetics of Europe. *Current Biology* 20, R174–83.
- Street, M. & Terberger, T. (1999) The last Pleniglacial and the human settlement of central Europe: new information from the Rhineland site of Wiesbaden-Igstadt. *Antiquity* 73, 259–272.
- Thomas, M., Stumpff, M., & Harke, H. (2006) Evidence for an apartheid-like social structure in early Anglo-Saxon England. *Proceedings of the Royal Society B: Biological Sciences* 273, 2651–2657.
- Underhill, P. & Kivisild, T. (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* 41, 539–564.
- Wakeley, J. (2009) *Coalescent Theory: An Introduction*. Greenwood Village, CO: Roberts.
- Weale, M., Weiss, D., Jager, R., et al. (2002) Y chromosome evidence for Anglo-Saxon mass migration. *Molecular Biology and Evolution* 19, 1008–1021.
- Wilson, J., Weiss, D., Richards, M., et al. (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proceedings of the National Academy of Sciences* 98, 5078–5083.