Annals of Human Biology, March–April 2007; 34(2): 139–172

informa
healthcare

REVIEW

Y chromosome variation in Europe: Continental and local processes in the formation of the extant gene pool

ANDREA NOVELLETTO

Department of Biology, University 'Tor Vergata', Rome, Italy

(*Received 18 October 2006; revised 20 December 2006; accepted 4 January 2007*)

Abstract
The polymorphism of the male-specific portion of the Y chromosome has been increasingly used to describe the composition of the European gene pool and to reconstruct its formation. Here the theoretical grounds and the limitations of this approach are presented, together with the different views on debated issues. The emerging picture for the composition of the male gene pool of the continent is illustrated, but local peculiarities that represent departures from the main trends are also highlighted, in order to illustrate the main unifying feature, i.e. the overlay of recent patterns onto more ancient ones. A synopsis of the main findings and conclusions obtained in regional studies has also been compiled.

Keywords: *Peopling of Europe, Y chromosome, phylogeography*

Introduction

The reconstruction of the history of human populations by means of genetic markers is a gigantic scientific enterprise which brings together the genetics, anthropology and, more recently, the history, archaeology and forensic medicine communities. In this context, the Y chromosome is playing a pivotal role, for reasons that have been repeatedly authoritatively reviewed (Jobling and Tyler-Smith 1995; Bertranpetit 2000; Renfrew et al. 2000; Cavalli-Sforza and Feldman 2003; Garrigan and Hammer 2006). This review attempts to provide a picture of the state of the art as far as Europe is concerned. For conciseness, only territories strictly within European borders are considered, and also comparisons with the results obtained with other markers are avoided. In order to make the entire context accessible to the non-specialized reader, a summary of the properties of this genomic region and of its place in population genetics theory is also given. An in-depth view of some regions of the continent will follow, as well as the main conclusions obtained for some

Correspondence: Andrea Novelletto, Department of Biology, University 'Tor Vergata', Via della Ricerca Scientifica, Rome 00133, Italy. E-mail: novelletto@bio.uniroma2.it

ISSN 0301–4460 print/ISSN 1464–5033 online
DOI: 10.1080/03014460701206843

'outlier' populations. Whenever possible the reader is addressed to previous reviews and by no means can the literature cited here be considered exhaustive.

Mendelism and the Y chromosome

To the unfamiliar reader, the Y chromosome is known as a chromosome that determines, as a whole, the male sex and has no homologue in the individual's chromosome set. However, from the Mendelian point of view, it consists of three distinct portions, known as PAR1 (pseudo-autosomal region 1), MSY and PAR2 (for a review see Jobling and Tyler-Smith 2003). Here we are concerned only with the MSY (male-specific region; also known as NRY, or non-recombining region) as it is the only portion that is entirely transmitted from male to male, being free from homologue–homologue recombination (i.e. with the X chromosome). With few exceptions DNA variation referred to in the present review resides within the MSY. Since the MSY is present as a single copy in each individual genome, all alleles at variable positions are found *in cis* (i.e on the same DNA molecule), thus constituting a single haplotype. Typically, the subject's haplotype is transmitted unaltered to his male offspring except when a mutation occurs. This haploid state renders markers of the MSY particularly easy to type as only one allele has to be detected and no phase reconstruction is required.

Genetic markers in the MSY

The MSY consists of ~60 Mb of DNA, 30 Mb of which represent the euchromatic portion (Skaletsky et al. 2003). This amount of genetic material contains markers that belong to the same classes observed in the autosomes and thus represent a repertoire larger than in mitochondrial DNA (mtDNA). In the MSY, variation in the modules of alphoid DNA, deletions and inversions of large stretches of DNA are observed, as well as variations of smaller magnitude such as Alu insertions, single nucleotide polymorphisms (SNPs) and variation in the 2–5 bp repeats of microsatellites (short tandem repeats, STRs). All of the above, alone or in combination, have been used for population studies but SNPs and STRs are by far the most popular.

There are >57 000 SNPs of the Y chromosome reported in the SNP database at the time of writing, of which one-third validated. In the Y chromosome literature, SNPs are often referred to as stable binary or biallelic markers as they arise by mutational events that occur with a very low frequency (of the order of 10^{-8} per base pair per generation). The consequence is that the chance of two consecutive events hitting exactly the same nucleotide pair is very low. Moreover, there is only a one in three chance that a putative second event reintroduces the pre-existing nucleotide in a mutated position, making this true reversion a very unlikely event. Inferences drawn on the phylogeny reconstructed with numerous markers confirmed this prediction with few, though notable, exceptions (Hammer et al. 1998; Underhill et al. 2000; YCC 2002). Similar considerations hold true also for Alu insertions, as specific mechanisms to precisely excise an inserted element (reversion) are not known.

A direct extension of these concepts is that all MSY copies (each carried by a different subject) bearing the same allelic variant at a given position can all be considered, as a first approximation, descendants of the first one in which that particular mutational event occurred (i.e. have a monophyletic origin). When considering more than one position on the

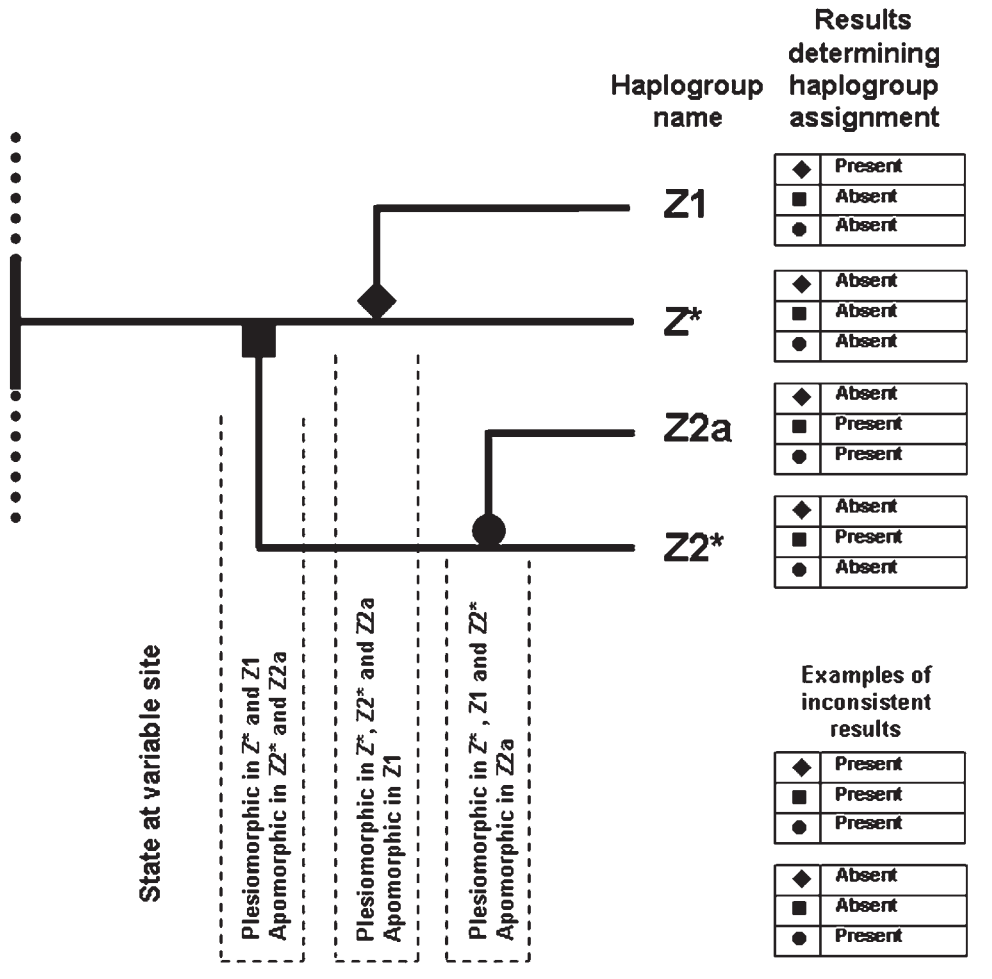

Figure 1. Definition and properties of clades within a hypothetical MSY haplogroup Z. Mutations at different MSY sites are represented by symbols. Each clade is defined by the presence of the ancestral (plesiomorphic) or derived (apomorphic) state at each of the three sites. Haplogroup names are reported at the tip of each clade, and follow the YCC (2002) nomenclature system. A * is used for paragroups defined by the lack of the derived state at some sites; a lowercase letter is used for nested clades (e.g. Z2a into Z2). Experimental results determining the assignment of subjects to a particular clade are reported on the right. Here 'present' and 'absent' are referred to the derived (mutated) state at each site (some of the results not compatible with the tree shown are also reported). Note that, with these population data alone, the relative order of mutations defining Z1 and Z2 cannot be reconstructed, and the particular succession shown here is arbitrary. Dating procedures (see text) can be used to distinguish among different possibilities.

same DNA molecule, the particular combination of allelic variants (the haplotype) thus represents a record of all mutational events that occurred on the lineage leading to that haplotype. Alleles shared by two haplotypes testify that they have common ancestry, whereas alleles that differentiate two haplotypes testify that they belong to lineages that diverged some time in the past and, since then, accumulated a different series of mutations (Figure 1). All haplotypes based on SNPs can be then viewed as the final branches (leaves) of a

phylogenetic tree. The root of the tree is represented by a haplotype (not necessarily found today and thus to be inferred) carrying the ancestral state at all positions found to be variable today. This is also called most recent common ancestor (MRCA), to signify that all variation existing when the MRCA existed, or earlier, has gone extinct. The principles and methods of phylogenetic reconstruction from experimental data can be found in basic books (e.g. Page and Holmes 1998). Hammer and Zegura (2002) also give an instructive parallel with classic cladistic theory.

With the increasing number of known markers, the MSY phylogenetic tree has grown in the number of terminal branches and complexity. After a pioneering era, the search for markers exploited high-throughput methods that were first applied to samples representative of the entire world population and, later, oriented to resolve in finer detail some specific lineages (Underhill et al. 1997, 2000, 2001; Hammer et al. 1998, 2001; Shen et al. 2000; Thomson et al. 2000; Cruciani et al. 2002; Hammer and Zegura 2002). However, due to the composition of the screening panels in each study and the quest for markers that could distinguish populations from specific areas of the world, the current set of MSY markers is still considered to contain ascertainment bias. This can affect quantitative measurements of population affinities, and efforts to overcome this problem with unbiased re-sequencing projects have been carried out (Wilder et al. 2004a).

A general consensus has been reached on the nomenclature of lineages, with alternating letters and numbers from the deepest to the terminal branches (YCC 2002). Each lineage defined by biallelic markers is referred to as a haplogroup, whereas the term haplotype has been restricted to a combination of alleles at STRs (see below). The discovery of new SNPs or of carriers of novel combinations of the previously known SNPs requires a revision of the tree. An advanced version was compiled by Jobling and Tyler-Smith (2003) and tree maintenance is curated at various sites (see Appendix). It is important to realize that, even in the most updated state, the tree includes 'paragroups', i.e. lineages (branches) defined by the absence rather than the presence of a derived allele at (a) certain position(s) (Figure 1). Inferences based on paragroups have to be treated cautiously as these cannot be assumed to have a monophyletic origin (Weale et al. 2003).

Another important class of markers is represented by short tandem repeats (STRs). These include loci with different length of the basic repeat, and extensive searches for developing them as markers have been performed (Kayser et al. 2004 and references therein). Mutation at these loci occurs by addition/subtraction of a number of repeats that in the majority of cases is one. This latter feature fits the theoretical 'Stepwise Mutational Model', which allows the calculation of expectations for the rate of accumulation of diversity and the distribution of allele sizes. In any case, the monophyletic origin of STR alleles cannot be assumed, as any allele of a given size can be generated by a number of events from an entire set of parental alleles. Mutation rates at STR loci are orders of magnitude higher than for SNPs, with a relevant heterogeneity among loci. Estimates of mutation rates can be obtained by a variety of direct methods, i.e. comparison of father's vs. son's haplotypes (see the compilation by Gusmao et al. 2005) as well as changes in allele sizes in deep-rooting pedigrees (Heyer et al. 1997; Bianchi et al. 1998; Foster et al. 1998). Evolutionary methods can be also used. Luca et al. (2005) used coalescent reconstructions and obtained locus-specific values comparable to those of the previous methods. Zhivotovsky et al. (2004) obtained an average mutation rate from population rather than family data, considering known foundation events as starting points for the production of the level of diversity observed today. Perhaps unexpectedly, their estimate was three to fourfold lower than the average of direct measurements, a finding that was recently interpreted in light of fluctuating demographies of subsets of the same population (Zhivotovsky et al. 2006).

Because of the above features, the evolutionary path which led to the observed haplotype diversity at STRs cannot be reconstructed in the form of a tree consisting of only divergent branches. Instead, reticulations must be allowed for, in graphs known as networks (Bandelt et al. 1999).

Nomenclature for haplotypes defined by STR loci varies from paper to paper. It has been proposed that a unified set of markers be utilized for forensic purposes (Pascali et al. 1999). The corresponding population data are deposited at http://www.ystr.org/index.html, which has become a relevant resource to locate populations of the world harbouring haplotypes identical or similar to a particular type, also for evolutionary studies.

Is the MSY a neutral marker?

A thorough discussion of Y-chromosomal lesions and their effects is beyond the scopes of this review. However, evolutionary neutrality is one of the main prerequisites if one wants to use a genetic marker to draw inferences on population processes dependent on time. It is also to be noted that the complete linkage along the MSY (see above) determines that even a single marker subject to selection would impose its own evolutionary trajectory to all markers *in cis*, a phenomenon called hitchhiking. Under this circumstance the entire evolution of the MSY would appear distorted, no matter which marker is observed.

A reproductive advantage of some structural features of the chromosome has been postulated when examining its complete sequence (Rozen et al. 2003; Skaletsky et al. 2003), but this is mainly significant for selection during the emergence of the human lineage, i.e. a time period which long predates that of concern here. Differences in inter-individual fitness due to the standing Y-chromosomal variation are of more direct relevance here. These differences can be classically ascribed to enhancement or reduction either in reproductive success or in viability/mortality. Examples of the first type derive mainly from the common occurrence of variations in copy number and arrangement in the cluster of DAZ genes (Reijo et al. 1995) and their association with sperm counts (Krausz and Degl'Innocenti 2006). These were then shown to be non-randomly associated with specific haplogroups (Paracchini et al. 2000, 2002; Krausz et al. 2001; Fernandes et al. 2004; Repping et al. 2004, 2006). Paradoxically, one such haplogroup not only reaches notable frequencies in Europe, but has attained near-fixation in various Northern populations. In a recent paper Fernandes et al. (2006) confirmed weak or null selection on chromosomes bearing different DAZ haplotypes. In addition, it should be remembered that the practice of monogamy, eventually coupled with control over family size, is able to damp the selective consequences of differential sperm counts (Andersson 1994).

A growing body of literature is also accumulating on the association of Y-chromosomal markers with variables of medical relevance (Charchar et al. 2003; Krausz et al. 2004). The lineage-specific magnitude of differential mortality and its impact in shaping the frequencies of specific lineages are far from being ascertained. Some of the variables that have been scrutinized affect mortality mainly in post-reproductive life and hence have a little selective impact. Moreover, variation of many of them is largely accounted for by environmental factors (e.g. diet or lifestyle), leading to the expectation of possible population-specific selective effects, if any. In any case, it can be predicted that the mounting information on the physiological role of Y-borne genes will promote a burst of studies in this field.

Finally, pure population genetic data can be examined for evidence for selection. The issue is summarized excellently by Jobling and Tyler-Smith (2003), who tentatively concluded in favour of evolutionary neutrality.

The Y chromosome in population genetics theory

The MSY occupies a special place in population genetics theory, as it is a uniparental marker in much the same way as the mtDNA, but with some peculiar features.

In a population of N males and N females, each autosome is represented in $4N$ copies, whereas the effective number of copies of MSY and mtDNA is N each. Thus, at face value, these latter are expected to have an equal effective size and thus to be exposed to similar drift effects assuming a panmictic population, four times stronger than autosomes. Experimental data revealed considerable departures from this expectation. First, the MSY long revealed a low level of diversity per unit of DNA length (Malaspina et al. 1990; Shen et al. 2000), which was compensated only by specific searches in regions particularly prone to mutation (Wilder et al. 2004a). These findings were interpreted as a signature of possible selection on the chromosome with a consequent young genealogy (Thomson et al. 2000), even younger than mtDNA. However, in recent years, it is becoming increasingly clear that the role of other factors that affect the evolutionary rate have not been taken in due account (Wilder et al. 2004b). In fact, multiple features of human reproductive and migratory behaviour cause populations to depart from panmixia. Three deserve attention here.

First, populations with large variances of family size (number of offspring) have reduced effective sizes (Ne). In these populations the null contribution of many subjects who do not reproduce is compensated by a few subjects who leave a very large number of offspring. Inheritance of family size (Austerlitz and Heyer 1998; Helgason et al. 2003) further enhances this effect. Indeed, biological and social factors allow only men to reach very large numbers of children, giving only to the male portion of the population the chance to reach high variances, i.e. small Ne. This is the case for polygyny, and MSY data seem to indicate that this was practised long before the shift to monogamy (Dupanloup et al. 2003). Furthermore, social rank may act in maintaining male reproductive differentials through generations. The extreme cases in which copies of the same MSY have been spread in millions over large geographical areas beginning from founders of ruling clans have been documented (Zerjal et al. 2003; Xue et al. 2005; Moore et al. 2006).

Second, inter-generational time may not be the same for males and females. Also in this case heritability of this trait enhances its genetic consequences (Tremblay and Vezina 2000; Helgason et al. 2003).

Third, some societies practice patrilocality while others are matrilocal, with an expected more even spatial distribution for female-borne and male-borne genes in the two cases, respectively. Oota et al. (2001) first reported MSY and mtDNA findings in sympatric populations with opposite practices. By exploiting the same data, Hamilton et al. (2005) showed that patrilocal and matrilocal societies are less and more open, respectively, to genetic inputs from the corresponding sex.

In summary, while the MSY is intrinsically prone to genetic drift, the above factors sum up in further enhancing it, with the overall result of rendering MSY polymorphism very variable in time (temporal changes in haplotype frequencies) and in space (changes in haplotype frequencies among populations, often between neighbouring ones). This expands the usefulness of MSY diversity in describing population divergence over short periods of

Figure 2. Map of southern Europe showing local vagaries in haplogroup frequencies. Each dot represents a sampling location in Italy and Greece. Upward arrows indicate 19 instances in which one haplogroup (of the 10 searched) had a frequency 0.1 or more higher than its frequency in the corresponding whole national sample. With the sampling size used, only 9.6 such instances were expected by random fluctuations. These vagaries, attributable to founder and/or drift effects, may obscure locally the continent-wide clines. Data from Di Giacomo et al. (2003).

time and limited geographic regions, given that appropriate fast-evolving markers are used (Kayser et al. 2001). However, since the above-mentioned factors act together and are heavily dependent on ethnicity (e.g. Wood et al. 2005), their overall effect on variation of MSY vs. mtDNA vs. autosomes cannot be generalized across populations or societies and cannot be assumed equal at the continental vs. the local scales (Hammer et al. 2001, 2003). Also, parameters of DNA diversity commonly used to infer demographic processes are sensitive to this high level of structuring and an improper pooling of experimental data may produce misleading results (Hammer et al. 2003).

The MSY markers showed the highest quotas of population divergence among continents ever recorded for different portions of the genome (Hammer and Zegura 2002; Romualdi et al. 2002). However, figures for the MSY and mtDNA converged when markers derived from similar re-sequencing surveys were used (Wilder et al. 2004a). When the role of mutation rate was kept putatively constant by using NRY markers and markers with a similar sequence from the X chromosome, world populations showed a higher degree of structuring

for the MSY at the inter- and intra-continental scale (Scozzari et al. 1997; Karafet et al. 1998). Seielstad et al. (1998) compiled data showing a stronger dependence of between-population differentiation on distance for the MSY as compared to both the autosomes and mtDNA, and proposed the idea of a higher proportion of females than males migrating at each generation, increasing the female-mediated gene flow and reducing divergence for all portions of the genome except the MSY (see also Kayser et al. 2001). Others (Malaspina et al. 2000; Karafet et al. 2001) showed that the same regression changes dramatically depending on the geographic area from which populations originated, replicating findings from large datasets of classical genetic markers (figure 2 in Cavalli-Sforza and Feldman 2003). Finally, Wilder et al. (2004a) dismissed the hypothesis that patterns of genetic structure on the continental and global scales are shaped by the higher rate of migration among females than among males.

At a more local scale it is not uncommon to observe high degrees of differentiation (as compared to the average from autosomes) among populations living nearby, or in the same nation. This is partially expected when ethnic, linguistic or tribal affiliations contribute to the partial segregation of subsets of the entire population (e.g. Marjanovic et al. 2005; Sengupta et al. 2006). However, it was shown that also in the absence of these barriers, populations are structured for the MSY at the microgeographic scale, most likely due to local founder and/or drift effects (e.g. Capelli et al. 2003; Di Giacomo et al. 2003) (Figure 2).

The dating revolution

For geneticists, as for archaeologists, it was an ambition to know from when, in the past, the objects (here the MSY haplotypes) under observation come. Genetic dating, i.e. the estimation of the antiquity of an entire phylogenetic tree or of some of its branches based solely on genetic data, is now possible with increasing levels of confidence. This is obtained from the observed level of variation, by means of a variety of methods that consider SNP and other biallelic markers, STRs, or both. Methodological details cannot be given here and the reader is referred to Balding et al. (2001, chapters 7 and 8) and Jobling et al. (2004) (chapter 6.6) for further reading. One important feature is that advanced methods allow for models of populations which varied in size through time and split at some time in the past. As compared to a basic model with a panmictic population of constant size, this simulates much better what is considered to be the rule rather than the exception in many cases worldwide, and in particular in Europe.

Combining stable binary and STR markers seems particularly powerful. The idea (de Knijff 2000) is that the origin of a novel haplogroup is a singular event (see above) and, at that moment, there is a null variance for STR allele sizes on the particular chromosome that undergoes the SNP mutation. From that point on, all descendant chromosomes (that belong to the same new haplogroup) will start to accumulate STR variation. Hence, STR variation can be used to date the nodes of the SNP-based tree, i.e. the haplogroups. A variety of measures and methods have been developed to translate the observed STR variation into an absolute number of generations (Goldstein et al. 1995; Slatkin 1995; Wilson and Balding 1998; Bandelt et al. 1999; Cooper et al. 1999a, 1999b; Stumpf and Goldstein 2001; Zhivotovsky 2001). Although to a different extent, the results obtained with all the above methods depend on the values assumed for the basic determinant in the STR mutational process, i.e. the mutation rate. The use of different figures within the range proposed eventually affects the outcome of dating procedures. Luca et al. (2005) showed that

coalescent methods are able to predict locus-specific figures for mutation rates similar to those obtained in father–son transmissions and that using lower values in input produces convergent results. This adds confidence to the posterior dating estimates returned by these methods.

The possibility of obtaining dating estimates for MSY haplogroups (as well as for other regions of the genome) offers an unprecedented way of testing hypotheses on events that shaped the composition of entire populations' gene pools, by setting at least upper or lower bounds for population splits, admixtures, expansions, bottlenecks, etc. It thus represents a conceptual breakthrough as an escape to the circular argument in which external evidence (such as that provided by archaeology) is used to derive dating estimates for the genetic data and these latter, in turn, are taken as supporting evidence for historical reconstructions. However, great caution must be exerted in using dating results in the reconstruction of population history. A genetic variant can be introduced into a population's gene pool only after it first arose by mutation (the age of the molecule must predate the age of the gene pool harbouring it). However there is a tendency to consider that, if a gene pool contains a given variant, the gene pool dates back to the time of origin of that variant (age of the gene pool equal to the age of the molecule). Such a conclusion is unwarranted, because the gene pool may have formed long after the initial mutation generating the variant. Indeed, a proper phylogenetic resolution, obtained with numerous markers that identify different time depths, is definitely required to associate the composition of the gene pool with a certain antiquity. The issue raised a vivid debate (for an update see Barbujani and Chikhi 2006; Torroni et al. 2006). The above misconception has affected the way MSY data have been interpreted, particularly where Europe is concerned, in view of the mainly-Palaeolithic vs. mainly-Neolithic models for the peopling of the continent (see below).

Detecting MSY diversity in Europe

We are concerned here with how the polymorphism of the MSY can be exploited to reconstruct the build up of the present-day European gene pool. It is obvious that, since the first arrival of modern humans into the continent, the whole process consisted of numerous episodes scattered in time, some of which may have left a signature stronger than others. Basically, this can be viewed as an increase of complexity, due to the repeated addition of new variation to the pre-existing background by two main mechanisms: Immigration of differentiated MSY copies from outer regions, and accumulation of novel MSY variants generated by new mutations *in loco*. There are no *a priori* reasons to adopt a gradualistic view in which this increase in complexity was progressive (linear) in time. For example, population genetic theory leads to the prediction that new mutations accumulate particularly during demographic expansions, as these involve a greater retention of genetic variation (Harpending et al. 1998). Moreover, local gene pools which occupy only a portion of the continent may have been subjected to different events that added to their complexity.

The question for the geneticist is whether a DNA polymorphism which is able to mark a specific episode indeed exists and is known. A high resolution power is obtained with a large number of markers and with a sampling scheme that is able to represent a possible micro-differentiation (e.g. with multiple and closely spaced sampling locations, or accounting for ethnic subdivision, language affiliation or surname etymology). It is then not surprising that, with the increasing availability of markers, recent works have become able to detect the genetic signatures of events that affected local populations (see Table I).

Table I. A synopsis of MSY diversity in regional studies of Europe.

Region	Markers*	Main findings†‡	Main conclusions	Reference
Iberia	SNP	[E*(xE3a)] = 10% in Galicia [E*(xE3a)] = 24% in South Portugal	Clinal distribution attributed to either the Neolithic or to the Islamic rule. Little admixture with North Africa	Pereira et al. (2000)
	SNP + STR + one marker in PAR1	[P*(xR1a)]>58% except in Pasiegos [E] = 42% in Pasiegos [R1a] = 21% in Pasiegos	Heterogeneous resettlement from North Africa for Pasiegos	Scozzari et al. (2001)
	SNP + STR	[R1*(xR1a)]>76% [R1b3f]>11%	Contribution: 78% Upper Palaeolithic, 13% Neolithic, 7% North African (most likely during Islamic occupation)	Bosch et al. (2001)
	SNP + STR	[P*(xR1a)]>57% [E*(xE3a)] increased in Galicia	Strong divergence with North Africa. Genetic homogeneity in the Cantabrian region. Low input during Islamic colonization.	Brion et al. (2003)
	SNP + STR	[P*(xR1a)] reduced and [R1a] + [E*(xE3a)] increased in some Cantabrian locations	Contrast in homogeneity between Galicia and Cantabria	Brion et al. (2004)
	SNP	[R1*(xR1a)]>50% [R1b3f]>30% in Catalans [J2*] > 10% in the South	Low structuring, including the Basques. Prehistoric African contribution not excluded.	Flores et al. (2004)
	SNP + STR	[R1*(xR1a)] = 52–62% in Portugal [E3b1] = 6–8% in northern Portugal [E3b2] = 5–6% in northern Portugal [J1] = 7% but without the Arab STR haplotype	Continuous and regular assimilation of Berbers in north of Portugal prior to Moorish-mediated contributions. J1 as a marker of Sephardim and/or other Near East input.	Goncalves et al. (2005)
	SNP + STR	[E3b1] = 4% in Portugal [E3b2] = 6% [I1b2] = 1.5% [J] = 10.4% [R1*(xR1a)] = 58%	Genetic homogenization of Portuguese sub-populations. Diluted African contribution. Higher diversity in Alentejo.	Beleza et al. (2006)

	STR	Haplotype listing	Gene-frequency gradient, attributed to the post-Moslem reconquest	Pena et al. (2006)
Basques	SNP + STR + MSY minisatellite	R1b3f shared at high frequencies by Basques and Catalans. R1b3f originated few thousands years ago in Iberia	Male-mediated gene flow between Basques and Catalans after the establishment of the languages now spoken (linguistic barrier).	Hurles et al. (1999)
	P49a,f	[XV] = 72%	Frequencies of haplotype XV decreasing on both sides of the Pyrenees	Dieterlen and Lucotte (2005)
	SNP + STR	[R1b3f] = 2.4% 1% North-West African component Low STR diversity	Long history of small effective size. No evidence for the long-range diffusion into north Europe after the retreat of LGM.	Alonso et al. (2005)
	STR	Haplotype listing	Genetic heterogeneity between Resident and Autochthonous Basques	Pena et al. (2006)
Italy	SNP + STR + one marker in PAR1	Decreasing [P*(xR1a)] from north to south. Increasing [J] from north to south	J conveyed by the Neolithic spread of farmers, or later movements	Scozzari et al. (2001)
	SNP + STR	Two spots of high [P*(xR1a)] in the north spots of high [E], [I] and [J2] in the south	Heterogeneity not organized along the lines of continent-wide clines. Local frequencies dominated by drift and founder effects (Figure 2).	Di Giacomo et al. (2003)
	SNP	Decreasing [R1*(xR1a1)] from north to Sicily. Increasing [J2] from north to Sicily	Significant impact of Middle Eastern haplotypes in Sicily	Francalacci et al. (2003)
	SNP + STR	[J2] = 10% in the north but 24% in the south Higher [J2f] and [J2f1] in the south	Excess of derived haplogroups within clade J explained by immigration in the post-Neolithic	Di Giacomo et al. (2004)
	STR	Haplotype listing	Considerable heterogeneity between three southern locations	Capelli et al. (2006a)

(*continued*)

Table I. Continued.

Region	Markers*	Main findings†‡	Main conclusions	Reference
Corsica	SNP	[R1*(xR1a1)] = 50% [I1b2] = 0%	Similarity to northern Italy but strong differentiation with Sardinia	Francalacci et al. (2003)
	P49a,f	Decreasing [XV] from north to south. Increasing [V] from north to south.	Similarity with Italy but Arab input in the south	Lucotte et al. (2002)
Sardinia	YCAII (STR)	Peculiar pattern generated by a multirepeat deletion	Peculiarity of the gene pool composition, generated by founder effect and isolation	Ciminelli et al. (1995); Malaspina et al. (1998); Quintana-Murci et al. (1999)
	STR	Significant difference between the Sardinian and continental gene pool	As above	Caglià et al. (1997)
	SNP + STR + one marker in PAR1	[I1b2] = 17–55%	As above, with a focus in the archaeologically archaic area	Scozzari et al. (2001)
	SNP	[I1b2] > 30%	As above	Francalacci et al. (2003)
	SNP	[I1b2] = 12–27%	As above	Zei et al. (2003)
Central-western Europe and the British Isles	SNP + STR + p49f	[P*(xR1a)] = 73–98% in Gaelic surnames	P*(xR1a) is the indigenous Irish variant. Age within history of human habitation.	Hill et al. (2000)
	SNP + STR	[R1*(xR1a1)] = 89% in Ireland and Wales [AMH + 1] > 44% and 70% in Ireland and Wales [I] ~ 14% in Orkney	Basque–Celtic similarity. Pre-Anglo-Saxon continuity in Ireland and Wales. Norwegian input in Orkney.	Wilson et al. (2001)
	SNP + STR	[P*(xR1a)] from 89% in North Wales to 57% in east Anglia	Strong genetic barrier between central England and North Wales and virtual absence of a barrier between Central England and Friesland. Substantial Anglo-Saxon migration into central England (contributing 50–100% at that time) but not into North Wales.	Weale et al. (2002)

	SNP + STR	[R1*(xR1a1)] = 57–90% [AMH + 1] > 37% [R1a1] = 1–13% [J2] < 7%	Autochthonous background identified and persistent in southern England. Norwegian input in northern isles but not Scotland. German/Danish input in Scotland.	Capelli et al. (2003)
	STR	Most frequent haplotype in Ireland also most common throughout Europe. Six instances of a haplotype observed more than once in Ireland while absent from a worldwide sample.	Lower variability within the Irish sample at both the haplotypic and locus level	Ballard et al. (2005)
	SNP + STR	In Poland: [R1*(xR1a1)] = 7–17% [R1a1] = 53–64% [I] = 11–21% In Germany: [R1*(xR1a1)] = 23–55% [R1a1] = 8–31% [I] = 14–32%	Sharp genetic change between Germany and Poland, established after the continent-wide clines and attributable to resettlements after World War II	Kayser et al. (2005)
	SNP + STR	[P*(xR1a)] = 28% in Czech Rep. [R1a] = 34%	Low gene flow from the Mediterranean. Low internal structure. No frequency shift as between Germany and Poland. Signal of a local population growth.	Luca et al. (2007)
Balkans, Carpathians and eastern Europe	SNP + STR + 50f2 + 49a,f	[E] = 11–18% in Hungary [J] = 2%	Marginal impact of Neolithic demic diffusion in Hungary	Lahermo et al. (1999); Semino et al. (2000b)
	49a,f + STR	[R1a1] = 33–59% in central-eastern Europe, mostly associated with 49a,f Ht 11. Higher diversity in eastern Europe.	R1a1 re-expansion after LGM from an eastern European refuge	Passarino et al. (2001)

(*continued*)

Table I. Continued.

Region	Markers*	Main findings†‡	Main conclusions	Reference
	SNP	[P*(xR1a)] = 16% in Romania [R1a] = 23% Significant heterogeneity between the eastern and western side of the Carpathians	The Carpathians coincide with a genetic boundary. Central Romania poorly influenced by Neolithic demic diffusion.	Stefan et al. (2001)
	SNP + STR	[R1*(xR1a1)] = 0–9% in Russia/Ukraine [N3] = 13–43% [R1a1] = 11–47% [I] = 6–27%	R1a1 shared between Russia/Ukraine and Central Asia resulting from the spread of the Kurgan culture	Wells et al. (2001)
	P49a,f	[XI] = 18–44%	Latitudinal gradient for haplotype XI, tightly associated with R1a	Lucotte et al. (2003)
	STR	[R1*(xR1a)] = 1–19% in Volga-Ural [N3] = 16–56% [N2] = 5–28% [R1a] = 10–34% [I] = 1–19%	Genetic continuity with Siberia marked by haplogroup N2 but low frequencies of Q and C	Tambets et al. (2004)
	STR	[R1a1] = 12–15% in former Yugoslavia [E3b1] = 9–20% [I] = 36–73% [J] = 1–11%	Common background between Serbs, Croats and Bosniacs due to post-LGM re-expansion and migrations from Central Asia. Strong drift in Croats. Overall little Neolithic input.	Marjanovic et al. (2005)
	SNP + STR	[R1*(xR1a)] < 21% in the Balkans [R1a] > 15% [I1b*(xI1b2)] > 18% [E3b1] = 0–24% [J2e] = 0–6.3%	Common paternal history of Slavs. Post-LGM R1a expansion from eastern to western Europe. Younger Dryas-Holocene I1b*(xI1b2) diffusion out of the Balkans. Vardar–Morava–Danube river system as one of major routes for E3b1.	Pericic et al. (2005)
	SNP + STR	[R1*(xR1a)] = 17–21% in Albania [R1a] = 4–10% [I1b] = 2–17% [E3b1] = 27–45% [J2e] = 14–17%	E3b1 as a focal rather than a clinal phenomenon, due most likely to drift. Maritime spread of J2e later than Neolithic demic expansions.	Barac et al. (2003); Pericic et al. (2005)

	SNP + STR	[R1*(xR1a1)] = 8–14% [R1a1] = 6–22% [E3b1] = 10–23% [J2] = 2–20%	Balkan populations genetically homogeneous. Recent and common origin of Aromun communities, with drift.	Bosch et al. (2006)
Greece and Aegean Islands	SNP + STR	[P*(xR1a)] = 13% [R1a] = 10–12% [E] = 20–24% [J2] = 18–25% [I] = 15–18%	Lack of continent-wide gradients detectable within the country. Local heterogeneity among locations. Genetic input from north-eastern Europe.	Di Giacomo et al. (2003)
Crete	SNP + STR	[P*(xR1a)] = 10% [R1a] = 8% [E] = 15% [J2] = 35% [I] = 14%	Overall enrichment in haplogroup J. Local heterogeneity among locations, with multiple drift/ founder effects.	Di Giacomo et al. (2003)
Baltic and Scandinavia	STR + alfoid	Three main lineages detected in Finland. Significant heterogeneity between early and late settlements. Overall low diversity.	Two separate routes for peopling added diversity to a pre-existing background. Male-specific bottleneck.	Kittles et al. (1998, 1999)
	SNP + STR	[P*(xR1a)] = 29% in Norway but < 23% in the rest of populations [N3] = 64% in Finns but <32% in the rest [R1a] = 25–41% in the Baltic Republics and Norway but <21 in the rest	Boundaries across the Baltic sea and between Estonia and Latvia. Western origin for Gotlanders. Stronger geographic than linguistic barriers to gene flow, but detectable differentiation of Estonians from Latvians and Lithuanians.	Zerjal et al. (2001)
	STR	Haplotype listing	Estonians differentiated from Latvians and Lithuanians	Lessig and Edelmann (2001)
	SNP	Low differentiation among Estonians, Latvians and Lithuanians	Common Finno-Ugric ancestry across the present-day linguistic barrier	Laitinen et al. (2002)

(*continued*)

Table I. Continued.

Region	Markers*	Main findings†‡	Main conclusions	Reference
	SNP + STR	[R1*(xR1a1)] = 28% in Norway [I] = 40% [R1a1] = 24%	Gene pool mostly Palaeolithic. No Neolithic contribution. Low input from Uralic speakers.	Passarino et al. (2002)
	SNP + STR	[P*(xR1a)] = 5% in Lithuania [R1a] = 45% [N3] = 37% with low diversity	Lithuanians closely related to Slavs and Finno-Ugrians. No significant differences among ethnolinguistic groups of Lithuanians. Signal of a local population growth.	Kasperaviciute et al. (2004)
	SNP + STR	[P*(xR1a)] = 44–27% from south to north Norway [R1a] = 13–31% from south to north [N3] = 0–10 from south to north	Geographical substructuring between regions coherent with continent-wide clines.	Dupuy et al. (2006)
	SNP + STR	[P*(xR1a)] = 27% in Sweden [N3] = 10% [R1a1] = 12% [I1a] = 37%	R1b3 earliest major lineage. Low structuring except in Vasterbotten, where a German–Dutch input is detectable. Population continuity from the pre-Neolithic.	Karlsson et al. (2006)
	SNP + STR	[I] = 15–52% in Finland [N3] = 25–79% [R1a1] = 2–19%	Sharp genetic border between eastern and western Finns. Early Finno-Ugric migrations in whole country. Subsequent migrations from Scandinavia in western Finland. Early assumptions of the homogeneity of the isolated Finnish population dismissed.	Lappalainen et al. (2006)
Iceland	SNP + STR	[P*(xR1a)] = 41% [R1a] = 34%	Icelanders amongst the more genetically homogeneous populations in Europe. 20–25% of Icelandic founding males had Gaelic ancestry, with the remainder having Norse ancestry.	Helgason et al. (2000)

Faeroe	SNP + STR	[P*(xR1a)] ~ 25% [R1a] ~ 50%	Colonized at approximately the same time as Iceland from the same regions of Scandinavia and, to a lesser extent, from the British Isles.	Jorgensen et al. (2004)
Saami	SNP	[N3] = 42–62% in Inari and Skolt Saami	Dual contribution to the gene pool	Lahermo et al. (1999)
	SNP + STR	[P*(xR1a)] = 6% in Swedish Saami [N3] = 42% [R1a] = 21%	Similar to Estonians, but distinct from Finns	Zerjal et al. (2001)
	SNP + STR	[P*(xR1a)] = 0–4% [N3] = 44–61% [R1a] = 0–10%	Two founding lineages in Saami (and in Finns). Sub-populations with significant differences	Raitio et al. (2001)
	SNP	[P*(xR1a)] = 4% [N3] = 47% [R1a] = 11% [I] = 26%	Large genetic separation of the Saami from other Europeans best explained by assuming that the Saami are descendants of a narrow, distinctive subset of Europeans. Gene flow from Siberians excluded.	Tambets et al. (2004)
Caucasus	STR	[R1*(xR1a1)] = 0–19% [R1a1] = 0–14% [G] = 0–74% [E] = 0–18% [J2] = 0–72%	Caucasus populations more closely related to West Asians than to Europeans. Strong divergence among populations with reduced diversity in the highlands. Language replacements in Armenians and Azeri. Input from Iran in Ossetians.	Nasidze et al. (2003, 2004a, 2004b)
Ashkenazi Jews from central and western Europe	SNP + STR	In Cohanim: [J] = 54–86% [R1a1] = 1–6% In Ashkenazi Levites: [J] = 10% [R1a1] = 52% with low STR diversity	Founding event for Ashkenazi Levites, perhaps at the time of the initial settlement of the community	Behar et al. (2003)

(*continued*)

Table I. Contined.

Region	Markers*	Main findings†‡	Main conclusions	Reference
Ashkenazi Jews	SNP + STR	Higher haplogroup diversity than sympatric non-Jewish populations. Lower within-haplogroup STR diversity	5%–8% of the Ashkenazi gene pool introgressed from non-Jewish European populations. Signature of a bottleneck in population history.	Behar et al. (2004)
Roma	SNP + STR	[H1] = 45% [I] = 23%	Roma sub-populations strongly structured, mainly according to language differences and history of migrations within Europe. Multiple independent admixture events with Europeans.	Gresham et al. (2001)
Vlax Roma	STR	Haplotype listing	Age of the main lineage = 400–500 years. Growth rate lower than the rest of Europe. Population splits at 350 and 100 years ago with low numbers of founders for the present-day groups.	Chaix et al. (2004)

*SNP includes all biallelic markers.
†Brackets indicate 'frequency of'. ‡When only STRs are analysed in a study, the occurrence of particular haplotypes is not reported in detail and is substituted by 'Haplotype listing'.

To achieve this goal, however, it is often necessary to type a large number of markers in a large number of subjects, a requirement that has prompted the development of high throughput methods for both SNPs (reviewed in Sobrino et al. 2005) and STRs (e.g. Bosch et al. 2002; Butler et al. 2002).

Also, the same need prompted two distinct experimental approaches. Most investigations are 'population oriented', i.e. they analyse the entire repertoire of haplogroups in their population samples and discuss them in light of similar data obtained in other populations. This type of studies provides a picture of the composition on entire gene pools, that can be used to partition the total variation in its intra- and inter-population components, and to represent population affinities/differences by a wealth of multivariate methods. Conversely, in the 'lineage oriented' approach a deep analysis of one or few clades of the phylogenetic tree with their internal markers is performed in large collections of samples. This type of phylogeographic studies is able to disclose more immediately the entire home range for a haplogroup and to identify areas with frequency peaks. Also, it is able to highlight spots for the emergence of nested clades (internal lineages defined by additional markers). In summary, they can give a picture for the spread of one or a few components of the male gene pool. The relevance of the underlying population processes may vary from place to place, and must be weighted against information derived from other haplogroups. Underhill et al. (2001) used this approach to re-analyse the data of 2000.

Works performed until 2000 used markers that mostly identified what are now known to be the deepest clades of the tree, and continent-wide samplings that could barely represent heterogeneity at the sub-national level. Nevertheless, they were invaluable as they produced a general view of the genetic landscape of Europe from the MSY perspective, which should be taken in due account by all later works (also those which used other genomic regions). In discussing the data, they considered what are thought to be the most important events in the formation of the gene pool of the entire continent, i.e. the Palaeolithic peopling and the introduction of farming in the Neolithic.

Anatomically and behaviourally modern humans entered Europe from the Levant and rapidly dispersed into the continent between 46 000 and 41 000 ybp (Mellars 2006). This is considered to be the seeding event in the formation of the continental MSY gene pool. In fact, although genomic data are compatible with the presence of ancient molecular types possibly carried by Neanderthals in modern Europeans (Green et al. 2006; Plagnol and Wall 2006), typing of ancient mtDNA has indicated that early modern humans were carriers of molecular types within the range of present-day diversity (Caramelli et al. 2003), and pre-existing types have not survived (Currat and Excoffier 2004). On the other hand, it is entirely possible that molecules older than 46 000 ybp survive in the extant European MSY gene pool. In fact, the number of elapsed generations (<2000) is of the same magnitude as a reasonable effective male population size for the entire continent. Since the expected time to the MRCA (in generations) is twice the number of gene copies for stationary populations, it is possible that the diversity present among all ancient European MSYs has not gone entirely extinct. While the time to the MRCA is shortened in growing populations (as was certainly the case for the European population over this period), even modest levels of ancient gene flow with Asia may have maintained old molecules in Europe.

Farming (including domestication and animal breeding) began in the Fertile Crescent, was imported in Europe in association with Neolithic industries starting from 10 000 ybp, and is now long practised in the entire continent. Different views of this process have been proposed on archaeological grounds (Cavalli-Sforza et al. 1994, chapter 5.2; Jobling et al. 2004, chapter 10, and references therein). The basic question is whether the spread of agricultural knowledge was conveyed mostly by the peoples who practised it

(Demic Diffusion Model, DDM, with the associated movement of genes) or whether it was mainly an acculturation process, with pre-existing hunter–gatherers learning and using techniques from the newcomers (Cultural Diffusion Model, CDM) but with a low degree of interbreeding (SPIWA model in Renfrew 2001). Genetics thus appears to be the optimal tool to solve the issue. While the relevance of the MSY data in favour of one or the other will be discussed below, here it is important to note that agriculture can sustain a much higher population density than the previously practised hunting–gathering. This increase has been estimated at 5–50-fold within the first millennium of change (Cavalli-Sforza and Bodmer 1971, chapter 8.10; Cavalli-Sforza et al. 1994, chapter 2.7), which means a progressive burst of the population as agriculture moved from East to West within Europe.

Europe: The coarse picture

In discussing the distribution of MSY variation, the use of haplogroup names and its technicalities cannot be avoided (YCC 2002) (Figure 1). As some lineages have changed their labels due to advancements in the reconstruction of the phylogeny, for ease of consultation the picture given in Jobling and Tyler-Smith (2003) will be taken as reference unless otherwise specified, by giving the original and unified haplogroup names.

In 2000 two seminal papers produced a largely convergent picture of the MSY landscape of Europe by using large and entirely independent population samples (3677, 48 populations and 1007, 25 populations) and two different sets of markers. Rosser et al. (2000) typed 12 haplogroups using markers described in a number of previous papers. Semino et al. (2000a) reported on the occurrence of 19 haplogroups, the majority of which had been described by Underhill et al. (2000). The results tightly overlapped in showing:

- a frequency of the order of 50% or more for the P clade [hg1 = P*(xR1a) and Eu18 = R1*(xR1a1)] from central Europe to the West;
- a frequency of the order of 30% or more for the R1a clade [hg3 = R1a and Eu19 = R1a1] in central-eastern Europe and the Balkans, to the exclusion of Albania and Greece;
- a frequency of the order of 15–20% or more for the J clade [hg9 = J and Eu9 = J2, plus part of Eu10, which includes J1] in southern Italy, Greece, Albania and south-eastern Balkan peninsula, in continuity with even higher frequencies in Turkey and the Middle East;
- a frequency of the order of 30% or more for the N3 clade [hg16 = N3 and Eu13 + Eu14 = N3] on the eastern coast of the Baltic Sea and at the northern European–Asian border;
- frequencies of roughly 5–15% for the E clade [hg21 = E and Eu2 + Eu3 + Eu4 = E] in Iberia, southern Italy and southern Balkans, rising to approximately 25% in Greece. High frequencies of the E clade were known for Africa but not for the Middle East, interrupting the geographic continuity.

Rosser et al. (2000) analysed in detail the geographical arrangement of haplogroup frequencies and identified a clinal variation for five of them. For hg1 and hg9 the cline spanned the entire continent, and the overall frequency of the latter led to an estimate for the Neolithic contribution similar to that obtained from mtDNA and classical markers. Conversely, in the remaining three cases the cline was more localized and, for hg3 and hg16, may result from recent genetic inputs, mainly in northern and north-eastern Europe. These authors also identified genetic barriers, in the form of lines of abrupt frequency

change, and attributed a greater importance to geographic than linguistic factors in subdividing populations. They hypothesized that many kinds of barriers are probably recent on an evolutionary timescale, as compared to the establishment of clines.

Semino et al. (2000a) stressed on the antiquity of the different haplogroups to obtain a temporal picture. A multivariate analysis showed two clusters of populations geographically consistent with glacial refugia, meaning that re-expansion after the Last Glacial Maximum (LGM) was a major determinant of the present genetic landscape. In doing that, they obscured the genetic input marked by haplogroup N3, considered to be recent (Zerjal et al. 1997). The pooled frequency of three haplogroups (Eu9 + Eu10 + Eu11 = G + J) was regressed on the distance from the Middle East, concluding that they mark the movements of people in the Neolithic. As these, together with Eu4, account for only 22%, they concluded that most of the MSY diversity in Europe is accounted for by haplogroups that date back to a more distant past.

Subsequent works questioned this latter view, re-estimating the Palaeolithic vs. Neolithic contribution in Europeans under the assumption that the extant Basque and Middle Eastern populations are representative of the parental sources (Barbujani and Bertorelle 2001; Chikhi et al. 2002; Dupanloup et al. 2004). This resulted in an average figure of 65% for the Neolithic contribution, decreasing from east to west into the continent and strongly supporting the DDM. While these studies have the intrinsic value of comparing populations rather than molecules, they are subject to the caveat that ancestral traits (haplogroups) shared by an ancient coancestry (symplesiomorphic) enhance the similarity between the parental and the test populations, leading to distorted estimates of admixture.

Based on SNPs, populations facing the Mediterranean basin were clustered into three groupings (Capelli et al. 2006b): Near Eastern Arab, Mediterranean and North African. The observed pattern was ascribed to the Neolithic demographic expansion and subsequent westward migration by Phoenicians and Greeks that contributed to the distribution of Y chromosome types of most likely Near East origin.

Other works addressed specifically individual haplogroups over the continent and beyond. Zerjal et al. (1997) described a variant that identifies haplogroup N3 (=hg16) and determined the frequency of the derived allele in populations worldwide and the associated STR variation. The origin of this allele was most likely located in Asia 2–4 kya. The derived allele turned out to mark a relevant male-mediated gene flow between northern Asia and northern Europe, especially with populations speaking Finno-Permic languages.

The analysis of haplogroup I with newly developed markers (Rootsi et al. 2004) revealed a quota of diversity genuinely aboriginal of Europe. Three internal clades were estimated to have diverged and spread from glacial refugia in the Upper Palaeolithic/Mesolithic. In addition, the divergence of I1b2 within I1b in Iberia/Southern France provided a clue for the first peopling of Sardinia, where I1b2 reaches 40%.

Studies on haplogroup E showed that in Europe this is mostly represented by its internal lineage E3b1, with a focal distribution in the southern Balkans (Semino et al. 2004). By combining SNP and STR typing, it was possible to show that the European focus is phylogenetically distinct from two similar foci centred in Morocco and Ethiopia, respectively (Cruciani et al. 2004). The recent identification of new SNPs marking these phyletic distinctions (Cruciani et al. 2006) also confirms that Europe is characterized by a new molecular type, different from the aboriginal E3b chromosomes from Africa. The clinal frequency distribution of E-M78 within Europe testifies to important dispersal(s), most likely Neolithic or post-Neolithic. These took place from the Balkans, where the highest frequencies are observed, in all directions, as far as Iberia to the west and, most likely, Turkey to the southeast.

Semino et al. (2004) also considered haplogroup J, whose internal structure and distribution have been extensively worked out (Nebel et al. 2000, 2001, 2002; Sengupta et al. 2006). They inferred that two internal subclades had different histories. J2e shows a focal distribution centred in the south-western Balkans from where it probably spread. Conversely, the rest of J2 seems to be the result of population movements from the Middle East, in agreement with the observed similarity between its distribution and that of Neolithic painted pottery (King and Underhill 2002). These conclusions were refined by observing that the dispersal of haplogroup J is characterized by a significant degree of molecular radiation, with a higher incidence of more derived sub-haplogroups from Turkey westward. Here J diversity is not simply a subset of that present in the area where this haplogroup first originated, pointing to a punctuation in the peopling of southern Europe in the post-Neolithic (Malaspina et al. 2001; Di Giacomo et al. 2003). Interestingly, this haplogroup has also been associated with the Jewish diaspora (Hammer et al. 2000).

A general picture of frequency variation in space can also be detected with STR haplotypes. Gusmao et al. (2003) concluded in favour of a long-range cline for a cluster of related haplotypes found to belong to haplogroup P and for a more short-range cline for a cluster belonging to haplogroup E3b. Quintana-Murci et al. (2003) suggested a marked genetic differentiation between the East and the West Mediterranean for DYS392.

A comprehensive picture of Europe was obtained by Roewer et al. (2005) from >12 000 males typed for seven STRs [data deposited in Y Chromosome Haplotype Reference Database (www.ystr.org), see below]. They observed a geographical pattern of diversity resembling that obtained with SNPs and that was, however, attributed to relatively recent events. This conclusion was based on the observation of a major genetic division between a Slavic-speaking cluster to the East from a western-Romance-speaking cluster to the West, separated by a central European block of Germanic- and Italo-Dalmatian-speaking populations (language classification as in http://www.ethnologue.com/). Also, some of the lines of discontinuity recall political systems that were established in historical times. Finally, these authors notice the substantial contribution to the present gene pool which is evident at the eastern and north-eastern edge of the continent, attributing it to the various waves of immigration by nomadic Asian populations, as postulated also by Wells et al. (2001).

Iberia

From the point of view of human peopling, Iberia has long represented a geographical *cul-de-sac*; in fact, land bridges were never present at the Gibraltar Strait (Stringer and Andrews 2005), which implies that peopling was possible only through the Pyrenean gate until the appropriate seafaring technologies enabling a contribution from North Africa became available.

Accordingly, all reports agree in showing strong MSY differentiation between Iberia and North Africa (Capelli et al. 2006b and references therein). The MSY picture in Iberia is dominated by haplogroups belonging to the R1*(xR1a) clade [hg1 + hg22 (Rosser et al. 2000) or Eu18 (Semino et al. 2000a)]. High frequencies of these haplogroups characterize the entire western Europe, but are uncommon in north-western Africa, ruling out a considerable gene flow from this latter area. Within R1, haplogroup R1b3f was found specifically in Iberian populations. This haplogroup is found on both sides of the Pyrenees but particularly in the Basques (11%). By STR typing, the origin of this haplogroup was most likely located in Iberia and its age was estimated at 2300–3500 ybp (Hurles et al. 1999).

It thus represents an ideal tool to explore the degree of gene flow towards the rest of Europe and across the linguistic barrier separating the Basque (a non-Indo-European language) from other languages spoken in Iberia (all Indo-European).

A useful marker to measure the low north African contribution is within haplogroup E. Only with a proper phylogenetic resolution of this clade was it possible to recognize that in the rest of Europe haplogroup E is represented mostly by E3b1, whereas the most represented E branch in Morocco is E3b2 (Bosch et al. 2001; Cruciani et al. 2002, 2004; Arredi et al. 2004). The latter is found at frequencies of 4–12% in Portugal but at only <5.4% in Spain (mainly in the south). The same haplogroup indicates the peculiar ancestry of the Pasiegos, where it represents 28%.

Gene flow from the Mediterranean is testified by the frequencies of haplogroup J2, which are higher on the eastern Spanish coast.

Italy

Italy has long shown a line of genetic discontinuity separating the north from the south of the peninsula (Piazza et al. 1988) and overlapping with a linguistic boundary (Barbujani and Sokal 1990). This is faithfully replicated by MSY data. Northern Italy ranks among central European countries as far as haplogroup R1*(xR1a) is concerned (=Eu18). Conversely, haplogroup J (=Eu9 + part of Eu10) characterizes to a large extent southern locations. There is a consensus in considering the presence of this haplogroup a more or less immediate consequence of the spread of peoples triggered by the Neolithic agricultural revolution. However, uncertainties persist on when it attained the high frequencies observed today in southern Italy. While some papers favour the first wave of Neolithic farmers [documented at the sixth millennium BC (Malone 2003)], others conclude for a much later arrival in bulk numbers, during the Greek colonization (establishment of Magna Graecia) in the first millennium BC. The available data on Sicily favour this latter hypothesis.

Calabria and Sicily are characterized by the presence of haplogroup E3b3 at frequencies of 3–13%, which may testify of direct introgression from North Africa. Also, they indicate a considerable degree of micro-differentiation, in agreement with the long-standing isolation of many villages, due to the mainly mountainous landscape.

MSY analysis has also shown that the MSY gene pool in Corsica is similar to north-central Italy, in sharp discontinuity with the geographically close Sardinia.

Sardinia

Sardinia is known as a genetic outlier (Modiano et al. 1986; Cavalli-Sforza et al. 1994, chapter 5.6), with an additional relevant internal heterogeneity that correlates with linguistic differentiation between dialects (Cappello et al. 1996). The MSY pool of Sardinians is dominated by haplogroup I1b2, which is very rare elsewhere. This haplogroup also carries a distinctive pattern at the YCAII dinucleotide STR. Both these features were noticed early, but important clues to interpret the origins of the Sardinian gene pool derive from recent phylogeographic studies (Rootsi et al. 2004). Outside Sardinia I1b2 is found around the Pyrenees (6–7%) and in the rest of France and of Iberia, the British Isles (Capelli et al. 2003) and the Czech Republic (Luca et al. 2007). This suggests that I1b2 arose somewhere in central-western Europe and was already present among the first humans who colonized

the island about 9000 kya. Later immigrations in historical times (which are known to have affected the coastal region more than the inner region) would have diluted the frequency of I1b2 outside the so-called 'archaic' mountainous area.

Sardinia also shows increased frequencies of haplogroup G, a finding that has no immediate explanations. The dispersal of this haplogroup was initially related to the Neolithic spread of peoples (Underhill et al. 2001) but this view was reconsidered after finding that its distribution in Europe does not overlap that of the alleged 'Neolithic' haplogroup J. A focus of haplogroup G is found in some populations of the Caucasus (Nasidze et al. 2003).

Central-western Europe and the British Isles

The Basques and populations of the British Isles display the highest frequencies of haplogroup R1*(xR1a) [=Eu18 and hg1], with the possible exception of east Anglia. Based on SNP haplogroups, Wilson et al. (2001) identified Wales, Ireland and the Basques as a homogeneous group in which the Neolithic transition did not entail a major demographic shift. Interestingly, the similarity between these populations extends also to STR markers, with the sharing of a predominant haplotype (the so-called Atlantic Modal Haplotype, AMH) at high frequency. The consequence is a remarkable reduction in diversity for both the populations and the entire R1*(xR1a) clade. The relatively young age for the AMH in the entire western European edge remains a puzzling finding if one considers that the R1*(xR1a) clade is a paradigm of the Palaeolithic peopling of the continent.

With the same approach, two additional haplotypes (within haplogroups I and R1a, respectively) enabled the recognition of genetic inputs from mainland Europe in later times.

Taking advantage of the ancient (10th century AD) establishment of surnames as markers of local kinship systems, studies of the MSY in the British Isles have exploited for the first time the association with surnames to understand more precisely the origins of certain haplogroups or STR haplotypes (Hill et al. 2000; Jobling 2001; King et al. 2006; McEvoy and Bradley 2006; McEvoy et al. 2006; Moore et al. 2006). The linguistic origin of surnames was also recently exploited to distinguish the Slavic component into a German population sample (Immel et al. 2006).

In continental central Europe a greater diversity is observed. Germany and Denmark show a focus of haplogroup I1c (10–12%) and share with Scandinavia the highest frequencies of I1a (Rootsi et al. 2004).

Central Europe is crossed by a line which separates high frequencies of haplogroup P*(xR1a) to the west from high frequencies of R1a to the east. This was initially broadly located from the eastern Alps to the Baltic Sea (Malaspina et al. 2000). Recently it was found that this discrepancy in frequencies is more pronounced at the German–Polish border (Roewer et al. 2005), but much smoother more southerly (Luca et al. 2007). Its presence immediately north of the Alps remains to be tested.

The Carpathians and the Balkans

The Carpathian ridge is a semi-circular landmark that separates a continuously hilly/mountainous area to the west from huge steppes to the east, thus representing a strong

ecological discontinuity. The present-day Romania overlaps with both sides and an analysis of multiple Romanian samples showed that a haplogroup frequency discontinuity indeed exists, also in this case accounted for mainly by haplogroup R1a (higher frequencies east of the Carpathians).

Hungary is the sole country in central-southern Europe in which a Uralic language is spoken, after the Magyar conquest in the ninth century AD. Instead, the MSY data and particularly the low frequencies of haplogroup N3 (a very specific marker in this case) rule out a massive genetic contribution of Uralic-speaking populations to the present-day Hungarian gene pool, favouring a process of linguistic replacement (Table I).

The study of populations in the territory of the former Yugoslavia epitomizes the importance of having the proper genetic markers to reconstruct population affinities and splits. This region is strongly characterized by haplogroups I and E3b1 (see above), which differentiate it from central-western Europe. Additional markers within these lineages are able to improve the distinction between the two regions and among Balkan populations. It is to be expected that additional markers that resolve the haplogroups E3b1, I1a and I1b (Underhill et al. 2005; Cruciani et al. 2006) will aid in understanding the formation of the MSY genetic landscape of the region.

It is important to observe that, in spite of a bulk geographic continuity with Greece by land, and through the Ionian Sea, populations of this area display relatively low frequencies of lineages within haplogroup J2 (with the exception of J2e), i.e. little input of what are considered typical markers of the Neolithic diffusion or of post-Neolithic movements ensuing it. Conversely, these haplogroups seem to have undergone a more pronounced entry along the eastern edge of the Balkan peninsula and along the Black Sea coasts.

Greece and Greek Islands

Greece, Crete and the Aegean Islands is a key area to understand the migrations of early farmers to the rest of Europe. A detailed view of the role of Neolithic processes in shaping the Turkish gene pool has been proposed (Cinnioglu et al. 2004). Together with Turkey, Greece and the Aegean appear clearly as a source for haplogroup J2, but the timing for further movements to the west has not yet been fully established. The well documented expansion of the Ancient Greek world, consisting of repeated colonizations, is an immediate candidate process to have spread in the first millennium BC haplogroups that can be dated to an earlier phase of the Neolithic. This punctuation in the dispersal of 'Neolithic' genes has been hypothesized (Malaspina et al. 2001; Di Giacomo et al. 2003) based on haplogroup J, but further phyletic resolution of other haplogroups is needed. The present-day landscape of Greece is also characterized by a small-scale heterogeneity distinct from the continent-wide clines (Figure 2). This potentially provides the possibility of finding haplogroups or STR haplotypes linking the territories of colonies to those of the respective mother cities, as these relationships are historically known.

Greece and Crete also bring the signature of gene flow from north-eastern Europe, mainly represented by frequencies of R1a like nowhere else in southern Europe. This haplogroup is particularly abundant in Thessaly and underwent a further increase in eastern Crete (Di Giacomo et al. 2003).

Scandinavia and the Baltic

The region displays the signature of relatively recent immigration of Finno-Ugric speakers but without extensive contacts with the entirety of Uralic-speaking peoples. In fact, in this latter case higher frequencies of haplogroups N2, Q and C would be observed. The overall pattern is not consistent with isolation by distance but, instead, genetic boundaries testify of colonization of discrete areas with later smoothing of the initial differences. Both language replacement and genetic input from Indo-European speakers explain the detectable divergence between Estonia and Latvia/Lithuania (Table I).

The well studied Saami represent a case in which different samplings have produced drastically divergent conclusions, as could be predicted from a highly fragmented population with different subgroups practising different lifestyles (Cavalli-Sforza et al. 1994, p. 273).

Concluding remarks

At every meiosis, genetic recombination determines that each individual transmits his or her autosomal haploid genome as a set of DNA segments, each of which is extracted with a probability of 0.5 from the two copies received from his or her parents. The MSY is one additional segment and, furthermore, can be transmitted only by males to their male offspring. Taken together, these two mechanisms must sound as a warning against considering the genetic picture of populations provided by the MSY as the sole possible representative of the history of their entire gene pools. Instead, the MSY marks only one of the possible realizations of the evolutionary process, with its own genealogy, age and place of origin for the MRCA, which do not coincide with the genealogies and MRCAs of other genomic regions. The maternally inherited mtDNA represents a mirror situation. In this respect Europe is no exception and it is not surprising that many instances of populations in which different dynamics can be inferred from the two uniparental markers have been documented.

The main unifying feature emerging from the MSY genetic landscape of Europe is the overlay of recent patterns onto more ancient ones. To date, the ancient (pre-Neolithic) genetic landscape has the form of three major components. Some of the subsequent (Neolithic and post-Neolithic) additions have also been identified, but the work is far from complete.

First, the genetic signature of the Neolithic is appearing as a patchwork, in which the seas have represented more of a bridge than a barrier, while the Balkan peninsula (i.e. the most obvious route of entry by land) does not give consistently a signal of strong immigration and flow.

Second, more recent events (from the Bronze age to present) are beginning to demonstrate their effects on the present gene pool. More and more of these processes can be identified, as precise archaeologically-inferred and/or historically-documented events represent useful null hypotheses to be tested. These include back-migrations to Asia, migrations and resettlements at the microgeographic level, and local population expansions attributable to technological developments, favourable local environmental conditions and successful ruling systems. The strategy to arrive at the appropriate markers to highlight them (if not already available) is relatively straightforward; the amelioration of the phyletic resolution in certain haplogroups can be pursued, and sets of STRs can be chosen so that their amount of variation is not saturated and allows the exploration of the time-depth of interest.

Working out the establishment of underlying more ancient clines seems to be a much harder task. The model of three post-glacial re-expansions from a Pyrenean, a Balkan and a yet unidentified refugia to explain the distributions of the three main haplogroups R1b (or P*(xR1a) = Eu18 = Hg 1), I (=Eu7 + Eu8 = part of hg2) and R1a (=Eu19 = hg3), respectively, is not entirely satisfactory. For example, it fits with the observation of a low frequency of R1b in north-eastern Europe despite its origin has to be placed somewhere in Central Asia. Under this hypothesis, the high frequencies in the extreme west of Europe would be the result of further purification by drift or by a bottleneck during the LGM followed by re-expansion. This, however, contrasts with the view of the Basques as a relict population surrounded by other populations that suffered to a much larger extent from Neolithic immigration (see Hurles et al. 1999).

The interpretation of the R1a distribution is also disputable, with at least three competing models. Passarino et al. (2001) propose post-LGM eastward movements. Wells et al. (2001) also place the origin of this haplogroup in southern Russia/Ukraine, but associate its eastward spread with the much later movement of Indo-Iranian speakers 3000–1000 BC. Finally, Quintana-Murci et al. (2001) place the origin of R1a at a relatively recent date, and favour the idea that it marks the spread of Indo-European speakers from Central Asia into modern Iran via an Eastern-Caspian route as well as into India, with an entry into Europe from the north-east.

Here, too, a better phyletic resolution can be useful but in this case the most informative markers are represented by relatively old lineages that have most likely survived at low frequencies and are being replaced by recently derived variants. Searching for markers of this kind requires the identification of populations or isolates that have been relatively spared from recent immigration and drift.

Acknowledgements

I thank Drs F. Cruciani, F. Luca, G. Modiano, G. Passarino and L. Terrenato for critically reviewing a first draft of the manuscript. Work supported by Grant PRIN-MIUR 2005.

References

Alonso S, Flores C, Cabrera V, Alonso A, Martin P, Albarran C, et al. 2005. The place of the Basques in the European Y-chromosome diversity landscape. Eur J Hum Genet 13:1293–1302.

Andersson M. 1994. Sexual selection. Princeton, NJ: Princeton University Press.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, et al. 2004. A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75:338–345.

Austerlitz F, Heyer E. 1998. Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. Proc Natl Acad Sci USA 95:15140–15144.

Balding DJ, Bishop M, Cannings C, editors. 2001. Handbook of statistical genetics. Chichester, UK: Wiley.

Ballard DJ, Phillips C, Thacker CR, Robson C, Revoir AP, Syndercombe Court D. 2005. Y chromosome STR haplotypes in three UK populations. Forensic Sci Int 152:289–305.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Barac L, Pericic M, Klaric IM, Rootsi S, Janicijevic B, Kivisild T, et al. 2003. Y chromosomal heritage of Croatian population and its island isolates. Eur J Hum Genet 11:535–542.

Barbujani G, Sokal RR. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87:1816–1819.

Barbujani G, Bertorelle G. 2001. Genetics and the population history of Europe. Proc Natl Acad Sci USA 98:22–25.

Barbujani G, Chikhi L. 2006. Population genetics: DNAs from the European Neolithic. Heredity 97:84–85.

Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, et al. 2003. Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am J Hum Genet 73:768–779.

Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, et al. 2004. MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. Eur J Hum Genet 12:355–364.

Beleza S, Gusmao L, Lopes A, Alves C, Gomes I, Giouzeli M, et al. 2006. Micro-phylogeographic and demographic history of Portuguese male lineages. Ann Hum Genet 70:181–194.

Bertranpetit J. 2000. Genome, diversity, and origins: The Y chromosome as a storyteller. Proc Natl Acad Sci USA 97:6927–6929.

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, et al. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63:1862–1871.

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–1029.

Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, et al. 2002. High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. Forensic Sci Int 125:42–51.

Bosch E, Calafell F, Gonzalez-Neira A, Flaiz C, Mateu E, Scheil HG, et al. 2006. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. Ann Hum Genet 70:459–487.

Brion M, Salas A, Gonzalez-Neira A, Lareu MV, Carracedo A. 2003. Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. Am J Phys Anthropol 122:147–161.

Brion M, Quintans B, Zarrabeitia M, Gonzalez-Neira A, Salas A, Lareu V, et al. 2004. Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. Gene 329:17–25.

Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. 2002. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 129:10–24.

Caglià A, Novelletto A, Dobosz M, Malaspina P, Ciminelli BM, Pascali VL. 1997. Y-chromosome STR loci in Sardinia and continental Italy reveal islander-specific haplotypes. Eur J Hum Genet 5:288–292.

Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, et al. 2003. A Y chromosome census of the British Isles. Curr Biol 13:979–984.

Capelli C, Arredi B, Baldassari L, Boschi I, Brisighelli F, Caglià A, et al. 2006a. A 9-loci Y chromosome haplotype in three Italian populations. Forensic Sci Int 159:64–70.

Capelli C, Redhead N, Romano V, Cali F, Lefranc G, Delague V, et al. 2006b. Population structure in the Mediterranean basin: A Y chromosome perspective. Ann Hum Genet 70:207–225.

Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, et al. 1996. Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125–141.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, et al. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. Proc Natl Acad Sci USA 100:6593–6597.

Cavalli-Sforza LL, Bodmer WF. (1971). The genetics of human populations. San Francisco: W.H. Freeman & Co.

Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. The history and geography of human genes. Princeton, NJ: Princeton University Press.

Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. Nat Genet 33:(Suppl.)266–275.

Chaix R, Austerlitz F, Morar B, Kalaydjieva L, Heyer E. 2004, Vlax Roma history: What do coalescent-based methods tell us? Eur J Hum Genet 12:285–292.

Charchar FJ, Tomaszewski M, Strahorn P, Champagne B, Dominiczak AF. 2003, Y is there a risk to being male? Trends Endocrinol Metab 14:163–168.

Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci USA 99:11008–11013.

Ciminelli BM, Pompei F, Malaspina P, Hammer M, Persichetti F, Pignatti PF, et al. 1995. Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. J Mol Evol 41:966–973.

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148.

Cooper G, Amos W, Bellamy R, Siddiqui MR, Frodsham A, Hill AV, et al. 1999a. An empirical exploration of the (delta mu)2 genetic distance for 213 human microsatellite markers. Am J Hum Genet 65:1125–1133.

Cooper G, Burroughs NJ, Rand DA, Rubinsztein DC, Amos W. 1999b. Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. Proc Natl Acad Sci USA 96:11916–11921.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70:1197–1214.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. Am J Hum Genet 74:1014–1022.

Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): *a posteriori* evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum Mutat 27:831–832.

Currat M, Excoffier L. 2004. Modern humans did not admix with Neanderthals during their range expansion into Europe. PLoS Biol 2:e421.

de Knijff P. 2000. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 67:1055–1061.

Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, Cresta M, et al. 2003. Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. Mol Phylogenet Evol 28:387–395.

Di Giacomo F, Luca F, Popa LO, Akar N, Anagnou N, Banyko J, et al. 2004. Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. Hum Genet 115:357–371.

Dieterlen F, Lucotte G. 2005. Haplotype XV of the Y-chromosome is the main haplotype in West-Europe. Biomed Pharmacother 59:269–272.

Dupanloup I, Pereira L, Bertorelle G, Calafell F, Prata MJ, Amorim A, et al. 2003. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. J Mol Evol 57:85–97.

Dupanloup I, Bertorelle G, Chikhi L, Barbujani G. 2004. Estimating the impact of prehistoric admixture on the genome of Europeans. Mol Biol Evol 21:1361–1372.

Dupuy BM, Stenersen M, Lu TT, Olaisen B. 2006. Geographical heterogeneity of Y-chromosomal lineages in Norway. Forensic Sci Int 164:10–19.

Fernandes AT, Fernandes S, Goncalves R, Sa R, Costa P, Rosa A, et al. 2006. DAZ gene copies: Evidence of Y chromosome evolution. Mol Hum Reprod 12:519–523.

Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH. 2004. A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180–187.

Flores C, Maca-Meyer N, Gonzalez AM, Oefner PJ, Shen P, Perez JA, et al. 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: Implications for population demography. Eur J Hum Genet 12:855–863.

Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, et al. 1998. Jefferson fathered slave's last child. Nature 396:27–28.

Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, et al. 2003. Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am J Phys Anthropol 121:270–279.

Garrigan D, Hammer MF. 2006. Reconstructing human origins in the genomic era. Nat Rev Genet 7:669–680.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727.

Goncalves R, Freitas A, Branco M, Rosa A, Fernandes AT, Zhivotovsky LA, et al. 2005. Y-chromosome lineages from Portugal, Madeira and Acores record elements of Sephardim and Berber ancestry. Ann Hum Genet 69:443–454.

Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. 2006. Analysis of one million base pairs of Neanderthal DNA. Nature 444:330–336.

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, et al. 2001. Origins and divergence of the Roma (gypsies). Am J Hum Genet 69:1314–1331.

Gusmao L, Sanchez-Diz P, Alves C, Beleza S, Lopes A, Carracedo A, et al. 2003. Grouping of Y-STR haplotypes discloses European geographic clines. Forensic Sci Int 134:172–179.

Gusmao L, Sanchez-Diz P, Calafell F, Martin P, Alonso CA, Alvarez-Fernandez F, et al. 2005. Mutation rates at Y chromosome specific microsatellites. Hum Mutat 26:520–528.

Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. Proc Natl Acad Sci USA 102:7476–7480.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, et al. 1998. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427–441.

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, et al. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc Natl Acad Sci USA 97:6769–6774.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, et al. 2001. Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol 18:1189–1203.

Hammer MF, Zegura SL. 2002. The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. Annu Rev Anthopol 31:303–321.

Hammer MF, Blackmer F, Garrigan D, Nachman MW, Wilder JA. 2003. Human population structure and its effects on sampling Y chromosome sequence variation. Genetics 164:1495–1509.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. 1998. Genetic traces of ancient demography. Proc Natl Acad Sci USA 95:1961–1967.

Helgason A, Sigureth ardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, et al. 2000. Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717.

Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefansson K. 2003. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. Am J Hum Genet 72:1370–1388.

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P. 1997. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803.

Hill EW, Jobling MA, Bradley DG. 2000. Y-chromosome variation and Irish origins. Nature 404:351–352.

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Perez-Lezaun A, et al. 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65:1437–1448.

Immel UD, Krawczak M, Udolph J, Richter A, Rodig H, Kleiber M, et al. 2006. Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population. Eur J Hum Genet 14:577–582.

Jobling MA, Tyler-Smith C. 1995. Fathers and sons: The Y chromosome and human evolution. Trends Genet 11:449–456.

Jobling MA. 2001. In the name of the father: Surnames and genetics. Trends Genet 17:353–357.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: An evolutionary marker comes of age. Nat Rev Genet 4:598–612.

Jobling MA, Hurles ME, Tyler-Smith C. 2004. Human evolutionary genetics. New York: Garland Science.

Jorgensen TH, Buttenschon HN, Wang AG, Als TD, Borglum AD, Ewald H. 2004. The origin of the isolated population of the Faroe Islands investigated using Y chromosomal markers. Hum Genet 115:19–28.

Karafet T, de Knijff P, Wood E, Ragland J, Clark A, Hammer MF. 1998. Different patterns of variation at the X- and Y-chromosome-linked microsatellite loci DXYS156X and DXYS156Y in human populations. Hum Biol 70:979–992.

Karafet T, Xu L, Du R, Wang W, Feng S, Wells RS, et al. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. Am J Hum Genet 69:615–628.

Karlsson AO, Wallerstrom T, Gotherstrom A, Holmlund G. 2006. Y-chromosome diversity in Sweden—a long-time perspective. Eur J Hum Genet 14:963–970.

Kasperaviciute D, Kucinskas V, Stoneking M. 2004. Y chromosome and mitochondrial DNA variation in Lithuanians. Ann Hum Genet 68:438–452.

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, et al. 2001. An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am J Hum Genet 68:990–1018.

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, et al. 2004. A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183–1197.

Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, et al. 2005. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. Hum Genet 117:428–443.

King R, Underhill P. 2002. Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. Antiquity 76:707–734.

King TE, Ballereau SJ, Schurer KE, Jobling MA. 2006. Genetic signatures of coancestry within surnames. Curr Biol 16:384–388.

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, et al. 1998. Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171–1179.

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, et al. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. Am J Phys Anthropol 108:381–399.

Krausz C, Quintana-Murci L, Rajpert-De Meyts E, Jorgensen N, Jobling MA, Rosser ZH, et al. 2001. Identification of a Y chromosome haplogroup associated with reduced sperm counts. Hum Mol Genet 10:1873–1877.

Krausz C, Quintana-Murci L, Forti G. 2004. Y chromosome polymorphisms in medicine. Ann Med 36:573–583.

Krausz C, Degl'Innocenti S. 2006. Y chromosome and male infertility: Update, (2006). Front Biosci 11:3049–3061.

Lahermo P, Savontaus ML, Sistonen P, Beres J, de Knijff P, Aula P, et al. 1999. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet 7:447–458.

Laitinen V, Lahermo P, Sistonen P, Savontaus ML. 2002. Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers. Hum Hered 53:68–78.

Lappalainen T, Koivumaki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, et al. 2006. Regional differences among the Finns: A Y-chromosomal perspective. Gene 376:207–215.

Lessig R, Edelmann J. 2001. Population data of Y-chromosomal STRs in Lithuanian, Latvian and Estonian males. Forensic Sci Int 120:223–225.

Luca F, Basile M, Giacomo FD, Novelletto A. 2005. Independent methods for evolutionary genetic dating provide insights into Y-chromosomal STR mutation rates confirming data from direct father–son transmissions. Hum Genet 118:153–165.

Luca F, Di Giacomo F, Benincasa T, Popa LO, Banyko J, Kracmarova A, et al. (2006). Y-chromosomal variation in the Czech Republic. Am J Phys Anthropol, in press.

Lucotte G, Pinna A, Mercier G. 2002. Y-chromosome haplotypes in Corsica [in French]. C R Biol 325:191–196.

Lucotte G, Mercier G, Dieterlen F. 2003. Y-chromosome DNA haplotype XI in Eastern Europe. Hum Biol 75:405–410.

Malaspina P, Persichetti F, Novelletto A, Iodice C, Terrenato L, Wolfe J, et al. 1990. The human Y chromosome shows a low level of DNA polymorphism. Ann Hum Genet 54:297–305.

Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, et al. 1998. Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. Am J Hum Genet 63:847–860.

Malaspina P, Cruciani F, Santolamazza P, Torroni A, Pangrazio A, Akar N, et al. 2000. Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. Ann Hum Genet 64:395–412.

Malaspina P, Tsopanomichalou M, Duman T, Stefan M, Silvestri A, Rinaldi B, et al. 2001. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. Ann Hum Genet 65:339–349.

Malone C. 2003. The Italian Neolithic: A synthesis of research. J World Prehist 17:235–312.

Marjanovic D, Fornarino S, Montagna S, Primorac D, Hadziselimovic R, Vidovic S, et al. 2005. The peopling of modern Bosnia–Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. Ann Hum Genet 69:757–763.

McEvoy B, Bradley DG. 2006. Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. Hum Genet 119:212–219.

McEvoy B, Brady C, Moore LT, Bradley DG. 2006. The scale and nature of Viking settlement in Ireland from Y-chromosome admixture analysis. Eur J Hum Genet 14:1288–1294.

Mellars P. 2006. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. Nature 439:931–935.

Modiano G, Terrenato L, Scozzari R, Santachiara-Benerecetti AS, Ulizzi L, Santolamazza C, et al. 1986. Population genetics in Sardinia. Proceedings Accademia Nazionale Lincei 18:257–329.

Moore LT, McEvoy B, Cape E, Simms K, Bradley DG. 2006. A Y-chromosome signature of hegemony in Gaelic Ireland. Am J Hum Genet 78:334–338.

Nasidze I, Sarkisian T, Kerimov A, Stoneking M. 2003. Testing hypotheses of language replacement in the Caucasus: Evidence from the Y-chromosome. Hum Genet 112:255–261.

Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, et al. 2004a. Mitochondrial DNA and Y-chromosome variation in the caucasus. Ann Hum Genet 68:205–221.

Nasidze I, Quinque D, Dupanloup I, Rychkov S, Naumova O, Zhukova O, et al. 2004b. Genetic evidence concerning the origins of South and North Ossetians. Ann Hum Genet 68:588–599.

Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, et al. 2000. High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. Hum Genet 107:630–641.

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A. 2001. The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. Am J Hum Genet 69:1095–1112.

Nebel A, Landau-Tasseron E, Filon D, Oppenheim A, Faerman M. 2002. Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. Am J Hum Genet 70:1594–1596.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20–21.

Page RDM, Holmes EC. 1998. Molecular evolution. A phylogenetic approach. Oxford: Blackwell.

Paracchini S, Stuppia L, Gatta V, Palka G, Moro E, Foresta C, et al. 2000. Y-chromosomal DNA haplotypes in infertile European males carrying Y-microdeletions. J Endocrinol Invest 23:671–676.

Paracchini S, Stuppia L, Gatta V, De Santo M, Palka G, Tyler-Smith C. 2002. Relationship between Y-chromosomal DNA haplotype and sperm count in Italy. J Endocrinol Invest 25:993–995.

Pascali VL, Dobosz M, Brinkmann B. 1999. Coordinating Y-chromosomal STR research for the Courts. Int J Legal Med 112:1.

Passarino G, Semino O, Magri C, Al-Zahery N, Benuzzi G, Quintana-Murci L, et al. 2001. The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. Hum Immunol 62:922–932.

Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, Underhill PA. 2002. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur J Hum Genet 10:521–529.

Pena JA, Garcia-Obregon S, Perez-Miranda AM, De Pancorbo MM, Alfonso-Sanchez MA. 2006. Gene flow in the Iberian Peninsula determined from Y-chromosome STR loci. Am J Hum Biol 18:532–539.

Pereira L, Prata MJ, Brion M, Jobling MA, Carracedo A, Amorim A. 2000. Clinal variation of YAP+ Y-chromosome frequencies in Western Iberia. Hum Biol 72:937–944.

Pericic M, Lauc LB, Klaric IM, Rootsi S, Janicijevic B, Rudan I, et al. 2005. High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964–1975.

Piazza A, Cappello N, Olivetti E, Rendine S. 1988. A genetic history of Italy. Ann Hum Genet 52:203–213.

Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations. PLoS Genet 2:e105.

Quintana-Murci L, Semino O, Poloni ES, Liu A, Van Gijn M, Passarino G, et al. 1999. Y-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: A comparative study. Ann Hum Genet 63:153–166.

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, et al. 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. Am J Hum Genet 68:537–542.

Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES. 2003. Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. Am J Phys Anthropol 121:157–171.

Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, et al. 2001. Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 11:471–482.

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, et al. 1995. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383–393.

Renfrew C, Forster P, Hurles M. 2000. The past within us. Nat Genet 26:253–254.

Renfrew C. 2001. From molecular genetics to archaeogenetics. Proc Natl Acad Sci USA 98:4830–4832.

Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, et al. 2004. A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046–1052.

Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, et al. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463–467.

Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, et al. 2005. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291.

Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, et al. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12:602–612.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, et al. 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am J Hum Genet 75:128–137.

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526–1543.

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, et al. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873–876.

Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, et al. 1997. Differential structuring of human populations for homologous X and Y microsatellite loci. Am J Hum Genet 61:719–733.

Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, Moral P, et al. 2001. Human Y-chromosome variation in the western Mediterranean area: Implications for the peopling of the region. Hum Immunol 62:871–884.

Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, et al. 2000a. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. Science 290:1155–1159.

Semino O, Passarino G, Quintana-Murci L, Liu A, Beres J, Czeizel A, et al. 2000b. MtDNA and Y chromosome polymorphisms in Hungary: Inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. Eur J Hum Genet 8:339–346.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, et al. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023–1034.

Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, et al. 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. Am J Hum Genet 78:202–221.

Shen P, Wang F, Underhill PA, Franco C, Yang WH, Roxas A, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. Proc Natl Acad Sci USA 97:7354–7359.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837.

Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.

Sobrino B, Brion M, Carracedo A. 2005. SNPs in forensic genetics: A review on SNP typing methodologies. Forensic Sci Int 154:181–194.

Stefan M, Stefanescu G, Gavrila L, Terrenato L, Jobling MA, Malaspina P, et al. 2001. Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region. Eur J Hum Genet 9:27–33.

Stringer C, Andrews P. 2005. The complete world of human evolution. London: Thames & Hudson.

Stumpf MP, Goldstein DB. 2001. Genealogical and evolutionary inference with the human Y chromosome. Science 291:1738–1742.

Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, et al. 2004. The western and eastern roots of the Saami—the story of genetic 'outliers' told by mitochondrial DNA and Y chromosomes. Am J Hum Genet 74:661–682.

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360–7365.

Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ. 2006. Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345.

Tremblay M, Vezina H. 2000. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am J Hum Genet 66:651–658.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, et al. 2000. Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361.

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, et al. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62.

Underhill PA, King RJ, Chow C-ET, Kivisild T, Rootsi S. (2005). New phylogenetic relationships for Y chromosome haplogroup I: Reappraising its pattern of regional subdivision. In: Mellars P, Stringer C, Bar-Yosef O, Boyle K, editors. Rethinking the human revolution: New behavioural and biological perspectives on the origin and dispersal of modern humans. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 73–75.

Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG. 2002. Y chromosome evidence for Anglo-Saxon mass migration. Mol Biol Evol 19:1008–1021.

Weale ME, Shah T, Jones AL, Greenhalgh J, Wilson JF, Nymadawa P, et al. 2003. Rare deep-rooting Y chromosome lineages in humans: Lessons for phylogeography. Genetics 165:229–234.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, et al. 2001. The Eurasian heartland: A continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249.

Wilder JA, Kingan SB, Mobasher Z, Pilkington MM, Hammer MF. 2004a. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. Nat Genet 36:1122–1125.

Wilder JA, Mobasher Z, Hammer MF. 2004b. Genetic evidence for unequal effective population sizes of human females and males. Mol Biol Evol 21:2047–2057.

Wilson IJ, Balding DJ. 1998. Genealogical inference from microsatellite data. Genetics 150:499–510.

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc Natl Acad Sci USA 98:5078–5083.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, et al. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: Evidence for sex-biased demographic processes. Eur J Hum Genet 13:867–876.

Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, et al. 2005. Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am J Hum Genet 77:1112–1116.

YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 12:339–348.

Zei G, Lisa A, Fiorani O, Magri C, Quintana-Murci L, Semino O, et al. 2003. From surnames to the history of Y chromosomes: The Sardinian population as a paradigm. Eur J Hum Genet 11:802–807.

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, et al. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183.

Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinskas V, et al. 2001. Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. Mol Biol Evol 18:1077–1087.

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. 2003. The genetic legacy of the Mongols. Am J Hum Genet 72:717–721.

Zhivotovsky LA. 2001. Estimating divergence time with the use of microsatellite genetic distances: Impacts of population growth and gene flow. Mol Biol Evol 18:700–709.

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, et al. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61.

Zhivotovsky LA, Underhill PA, Feldman MW. (2006). Difference between evolutionarily effective and germ-line mutation rate due to stochastically varying haplogroup size. Mol Biol Evol, in press.

Appendix—Web resources

http://www.ystr.org/index.html
http://ycc.biosci.arizona.edu/
http://www.familytreedna.com/surname.asp
http://www.isogg.org/
http://freepages.genealogy.rootsweb.com/