

# The genetical archaeology of the human genome

Arndt von Haeseler<sup>1</sup>, Antti Sajantila<sup>2</sup> & Svante Pääbo<sup>1</sup>

Palaentology and archaeology are disciplines that traditionally deal with the reconstruction of human origins and history. Recently, however, molecular genetics has come to make increasing contributions to this area. In particular, several data sets indicate that variation of the human gene pool originated in Africa within the last 200,000 years. Furthermore, the study of DNA sequences allows the detection of expansions in population size. Here we briefly summarize and exemplify how DNA sequences can be used to reconstruct the history of populations.

Many of us, especially in our youth, are interested in the lives of our parents and immediate family; then again, as members of a particular group or population, we like to know about the life of our ancestors; finally, as members of the human race, we are fascinated with the question of human origins. Palaentology and archaeology are the disciplines that traditionally deal with our history prior to the availability of written or oral records. They address questions such as when and where our species first emerged, how our ancestors spread over the globe and what major events occurred in their history. However, early humans left traces of their activities not only in the form of their bones and artefacts. They also passed on to us their genomes. Every genome is made up of about three billion base pairs, several of which experience mutations in each generation, and, as the way in which these mutations accumulate in populations are influenced by how populations expand, contract, split and merge, the study of genetic variation has the potential to yield a great deal of information regarding our history.

Thanks to the development of methods that allow quick and cost-effective determination of DNA polymorphisms in many individuals, molecular genetics has recently begun to tap into this vast source of information. For example, molecular data have suggested that our species originated in Africa very recently, just 100,000 to 200,000 years ago, and from there went on to colonize the world, replacing other human forms, such as Neanderthals, in the process. Other studies have dealt with more recent events, such as major expansions in the human population and the colonization of areas where earlier humans had not ventured, such as Oceania and the Americas.

Below, we outline how DNA sequence variation can be used to reconstruct the history of populations and discuss some current views on the origin of modern humans and the major processes that have since shaped our gene pool.

## A bit of theory

All individuals have parents, and some individuals have the same parent(s). The consequence of these trivial facts is that as genealogical lineages in a population trace back over generations, they will occasionally coalesce to common ancestors. There will be fewer and fewer ancestors as one goes back. Eventually, all female lineages will trace back through a series of consecutive mothers to one single mother and all male lineages will similarly trace back to a single father — that is, the most recent common ancestors (MRCAs) on the maternal and paternal side (Fig. 1). If the demographic history of a population is known, the expected number of generations back to the MRCAs can be calculated: in a population that has been of constant size over time, it is approximately twice the number of reproducing females or males in the population. Thus, the time back to the MRCAs will be longer in a large population than in a small population (Fig. 1).

Variations in population size over time will affect the time to the MRCAs. For example, populations may be reduced dramatically in size and subsequently recover (a “bottleneck effect”) (Fig. 1*a*). In such a case, the MRCA may be as recent as in a smaller population that remained constant in size. However, in a population that has expanded, many lineages will coalesce in a rather short time frame after the ‘bottleneck’ and thus the genealogy of the lineages will be star-like. In contrast, in a population of constant size, the two oldest lineages will on average need as much time to coalesce as all the other lineages and the genealogy of the lineages will consequently show a deep ‘split’ (Fig. 1*b*).

If the demographic history of a population is unknown, it can be reconstructed from the patterns of nucleotide substitutions in the genome. DNA sequences from the mitochondrial (mt) genome and those from the majority of the Y chromosome are particularly useful as they are passed on without recombination from mother to daughter and from father to son. Consequently, these sequences can be traced back directly to the genealogical maternal or paternal MRCAs. Autosomal DNA sequences, which are inherited through both males and females and occur in two copies per individual, trace back to ‘biparental’ MRCAs that are on average four times as old as maternal and paternal MRCAs. If a population has grown in size, many lineages are expected to trace back to common ancestors close to the MRCA (Fig.

<sup>1</sup>Institute of Zoology,  
University of Munich,  
PO Box 202136,  
D-80021 Munich,  
Germany

<sup>2</sup>Present address:  
Department of  
Forensic Medicine,  
University of  
Helsinki, Kytösuontie  
11, FIN-00300  
Helsinki, Finland

e-mail: [paabo@zi.biologie.uni-muenchen.de](mailto:paabo@zi.biologie.uni-muenchen.de)

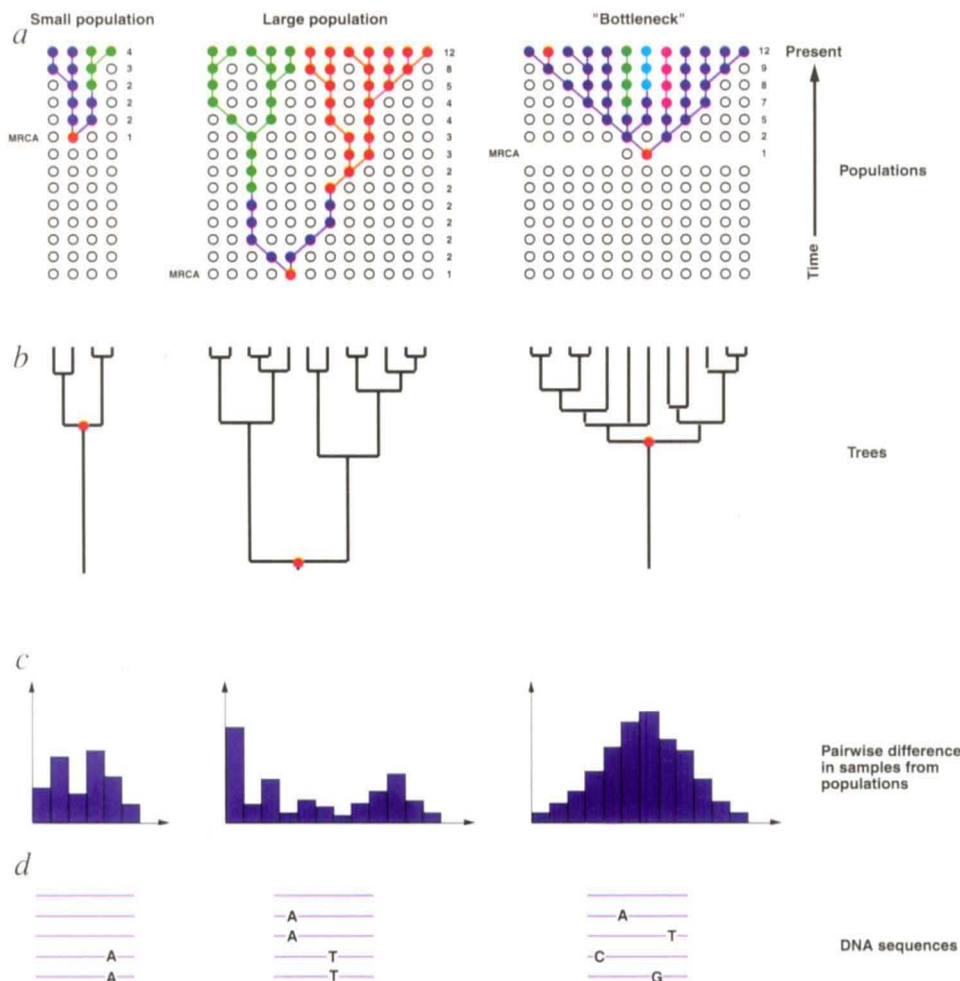


Fig. 1 A synopsis of some features of coalescent theory. *a*, Three populations are shown, one that has been of small and constant size over time, one that has been of large and constant size, and one that has been of small size during a 'bottleneck' and subsequently grown. Coalescence is the process by which the individuals in the present time trace back in the genealogy to common ancestors (the numbers of which are given for each generation), eventually reaching a single most recent common ancestor (MRCA) (red). If the demographic history is unknown, the pattern of genetic variation in the populations can be used to reconstruct the history. Other things being equal, the smaller population has a more recent MRCA than the larger population, and thus the deepest divergence, as well as the average amount of diversity in the population, is smaller. The population that has grown may be similar to the small population in that it has a recent MRCA but will display a number of additional features. First, the phylogeny of the genetic lineages in the population with size fluctuations will be star-shaped whereas it will show deep 'clusters' in the constant populations (*b*). Such clusters can therefore not, as is sometimes done<sup>57</sup>, be taken as indications of ancient 'populations' that have merged in the past. Second, the distribution of pairwise sequence differences (*c*) will tend to be bell-shaped whereas it tends to be ragged in the constant populations. Third, the number of variable positions (*d*) relative to the mean sequence difference will be larger in the population that has expanded than in the constant populations. The reason is that many mutations (illustrated by colours in (*a*) and (*d*)) will have occurred during a time when only few surviving lineages existed in the constant populations whereas many substitutions occur only in a few individuals in the growing population.

1). Therefore the distributions of pairwise sequence differences in such populations will tend to be bell-shaped, whereas those that have been of constant size will tend to have a ragged shape (Fig. 1c; refs 1, 2). A further difference between populations that have been of constant size and those that have expanded, is that in constant-size populations more substitutions have occurred when few of the currently surviving lineages existed. As such lineages are ancestral to many individuals in the population, these substitutions are found in many individuals today (Fig. 1d). In contrast, in a growing population, more substitutions have occurred in lineages that exist in only a few individuals today. They therefore affect many different sequence positions. However, the number of variable positions also depends on the over-all genetic diversity of the population reflected, for example, by the mean pairwise sequence difference. Thus, in populations that have been constant in size, the ratio of variable positions to mean pairwise sequence difference will be smaller than in populations that have increased in size.

In this way we can determine the approximate size of a population and whether it has been growing, and, if so, approximately when it started growing<sup>3</sup>. A major advantage is that, in principle, this approach is based on no assumptions external to the molecular data themselves. However, a limitation here is that as chromosomal segregation and recombination reshuffle the genome in every generation, different parts of the genome may have different histories. It is therefore of pivotal importance to study many genetic regions using the coalescent theory approach.

Another approach looks at the genetic relationships of populations rather than DNA sequences of individuals. To do this, many loci are generally studied and a distance measure is calculated that takes into account frequency of alleles in different populations, and sometimes the degree of difference between alleles in terms of mutational steps. Generally, the relationships between the populations are then depicted in the form of 'trees' where populations that have smaller genetic distances are shown as more closely related 'twigs.' The population approach has an advantage over the coalescent approach in that it uses information from many parts of the genome and takes into consideration the population aspect of human history. However, this approach also has limitations. First, as there is no genetic definition of a 'population'; the units of study are defined based on geographic, linguistic or other cultural ideas about 'groupings' of humans which may or may not be valid. Second, trees based on gene frequencies depict human history as a process of consecutive population splits. This, of course, is not the whole truth as populations exchange individuals and sometimes merge. Third, the time at which two populations diverged cannot, without the introduction of unrealistic assumptions, be gauged from such trees, as changes in population sizes — especially if populations are small — will cause dramatic shifts in the frequencies of alleles. However, together with the coalescent approach based on the analysis of individual DNA sequences, the population approach can provide a new understanding of our history.

### The age of the human gene pool

The mitochondrial genome has led the way in many studies of genetic diversity as its high evolutionary rate allows many substitutions to be observed even when rather short pieces of DNA are sequenced. In particular, attempts have been made to date the greatest divergence between mtDNAs of all humans currently existing in order to estimate the date of the maternal MRCA. Using restriction site data<sup>4</sup>, DNA sequence data<sup>5–7</sup> and different evolutionary rate calibrations<sup>8</sup>, dates on the order of 200,000 years ago have been obtained. Using particular models<sup>9</sup>, or slowly occurring transversal substitutions<sup>10</sup>, the date may be as far back as 300,000–800,000 years, but the majority of analyses still point to dates on the order of 200,000 years ago.

Similarly, DNA sequence variation on the Y chromosome has been used to calculate dates for the MRCA of Y chromosomes. In one study, an intron of the *ZFY* gene was sequenced from 38 men. Despite the lack of observed differences between the sequences, a date of around 200,000 years was calculated<sup>11,12</sup>. In another study, sequences of 2,600 bp on the Y chromosomes of 16 humans were determined. Here, variation did exist and a similar date was derived<sup>13,14</sup>. Thus, the data from the Y chromosomes agree with the 200,000-year-dates established by the mtDNA data.

Does this tell us that our species originated when the common ancestors of our mitochondrial genomes and Y chromosomes lived? Not necessarily. Genetic variation at many loci indicate that the size of the human population was around 10,000 for most of its history<sup>15</sup>. As discussed above, we can therefore expect to find maternal and paternal MRCAs around 200,000 years ago even if no event whatsoever occurred at that time (the 'small population' scenario in Fig. 1). As the patterns and modes of change in our genome are still not sufficiently characterized to allow us to distinguish a bottleneck from a 'constant' scenario for very old events, we can not yet tell if our species originated close to the dates of the MRCAs for the mtDNA and Y chromosomes. However, it is of interest that a comparison with the apes suggests that we may differ from our closest non-human relatives in having maternal and paternal MRCAs as recent as 200,000 years ago. When DNA sequences of a hypervariable region of the mitochondrial genome are compared in 3,263 humans, they are found to differ at an average of 5.4 positions. In contrast, 60 chimpanzees<sup>16</sup> differ by 16.6 in the same region. Similarly, studies of restriction site polymorphisms and DNA sequences in the mitochondrial genomes of humans, chimpanzees, gorillas and orang-utans, have shown humans to be much less diverse than apes<sup>17,18</sup>. This does not seem to be due to a lower mutation rate in humans<sup>19</sup>. Thus, the data suggest that the variation in the mitochondrial gene pool of our species is about three times younger than that of the chimpanzees. A recent study of microsatellite loci on chromosome 4 has similarly shown higher diversity among chimpanzees than among humans<sup>20</sup>. So, although more data are clearly needed, the human gene pool may indeed be younger than that of our closest ape relatives.

The dates for mtDNA and Y chromosomal sequences are in fair agreement with studies of variability in nuclear genes<sup>21,22</sup>. However, data on the variability of the major histocompatibility complex (MHC) stand in apparent contradiction to this picture. The function of most of the proteins encoded in this region is to bind fragments of antigens and present them to the immune system, and it has been shown that the genes in this region are under selection pro-

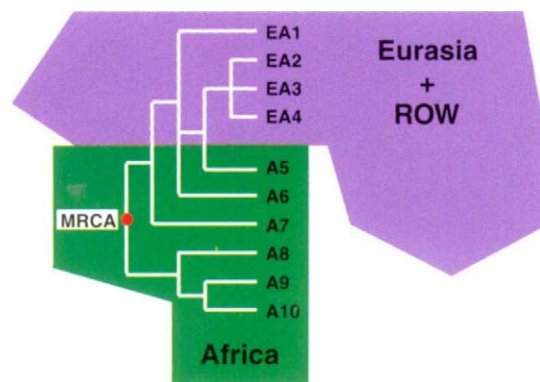


Fig. 2 Schematic illustration of how a tree relating human DNA sequences can point to the geographic origin of the DNA sequence. Of ten DNA sequences studied, six (A5–10) are detected in Africa (green), and four (EA1–4) in Eurasia and the rest of the World (ROW) (purple). If the depicted tree is correct, the explanation of the distribution of sequences around the world that minimizes the numbers of migrations is that the most recent common ancestor of these sequences (red) lived in Africa. In Africa, there exist sequences without close relatives that have emigrated to Eurasia (A8–10) as well as those with close relatives that have emigrated (A5 and A6) may have closer 'relatives' outside Africa (EA1–4) than inside Africa. Similarly, many sequences in Eurasia (EA2–4) are more closely related to sequences in Africa (A5) than to others in Eurasia (EA1). This general tree structure is compatible with data for mitochondrial DNA<sup>4,5,38</sup> as well as for the CD4 locus<sup>40</sup> and a minisatellite locus on chromosome 16 (ref. 45). Whether it is typical of most loci in our genome remains to be determined.

moting variation<sup>23</sup>. The large number of alleles present at the *DRB1* and *DQB1* loci in humans and the fact that many of them have their closest relatives in other primate species<sup>24</sup>, has been taken to mean that the long-term effective population size of humans is on the order of 100,000 (refs 25–27). Simulations show, however, that a brief bottleneck with a population of just 4,500 women and men would allow much of the diversity in the MHC to be maintained<sup>25–27</sup>. Thus, the MHC data are not necessarily in contradiction to the data for other loci. Furthermore, the rate of evolution of MHC genes may be underestimated. For the *DRB1* locus, which is used to infer the largest effective population size<sup>26</sup>, convergence between two alleles found in Africa and South America has been reported<sup>28</sup>. Extensive studies of introns in the *DRB1* locus also show that the evolutionary rate in the exons may be much larger than expected (Gyllenstein, U.B. *et al.*, manuscript in preparation). More work is needed before we can exclude that a high rate of mutation and convergent evolution<sup>29,30</sup> may partly explain the similarities between human and primate alleles.

### A place of origin

Initial analyses of mitochondrial genomes from around the world indicated that the greatest divergence in a phylogenetic tree was between a branch leading to exclusively African mitochondrial types and branches leading to African as well as other types<sup>4,5</sup>. As indicated in Fig. 2, such a branching structure of the tree would indicate that the ancestor lived in Africa. However, reanalyses of the nucleotide sequence data showed little statistical support for this conclusion<sup>31–34</sup>. One problem was that the tree contained more sequences than variable nucleotide positions, so that tree reconstruction algorithms could not point to one well-supported tree. Another problem was that the only available outgroup,



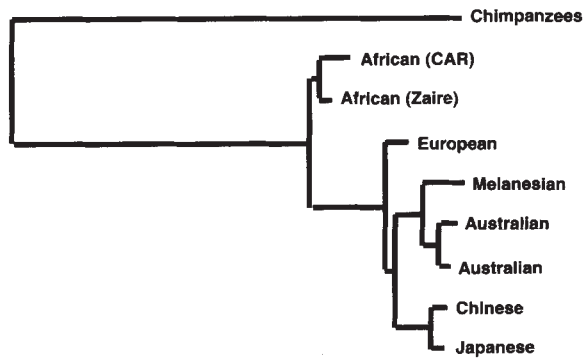


Fig. 3 A population tree relating population based on restriction fragment length polymorphisms (after ref. 42). The root of the tree is determined by the chimpanzee and falls between African and non-African populations, indicating that Africans have the longest independent history among the populations analysed.

chimpanzees, is so distantly related to humans that multiple substitutions have affected many positions, which make it difficult to find out how chimpanzee mitochondrial sequences relate to human sequences.

Various novel methods have now been used to analyse these data for a third time. Penny *et al.*<sup>35</sup> located the root among African sequences by using an approach that allows many trees of equal quality to be evaluated, giving more weight to nucleotide positions that are less affected by multiple substitutions. Furthermore, a new tree reconstruction program that allows a maximum likelihood approach to be used on large data sets<sup>36</sup> shows that the root supports an African origin with 85% reliability (Strimmer, K. *et al.*, pers. comm.) as does a principal components analysis of the data<sup>37</sup>. Finally, Zischler *et al.*<sup>38</sup> provided an outgroup that is less affected by multiple substitutions than the chimpanzees by isolating a partial mitochondrial control region sequence that was inserted in chromosome 11 shortly before the origin of the modern human mitochondrial gene pool. This sequence again indicates that the root lies in Africa. Thus, after the third generation of analyses of the mitochondrial data, current evidence suggests that the best placement of the root is in Africa.

Other data come from studies of haplotypes in the nuclear genome. In a pioneering study, J.B. Clegg's group<sup>39</sup> studied haplotypes of restriction site polymorphisms in the  $\beta$ -globin locus 601 chromosomes from eight populations. They found that of the 14 haplotypes detected, three were frequent in all non-African populations. In Africa, these haplotypes were rare, while two other haplotypes were frequent. This difference between African and non-African populations was interpreted as evidence of a migration out of Africa by a rather small population.

Recently, Tishkoff *et al.*<sup>40</sup> studied a haplotype located within an intron of the *CD4* gene on chromosome 12 consisting of a microsatellite and a deletion present in some humans. Within Africa, they found that three different microsatellite alleles occur in conjunction with the deletion at approximately equal frequencies. In contrast, outside Africa the deletion is associated with one of these microsatellite alleles in 98% of all cases. This argues that there has been substantially more time for mutations to occur in Africa than outside Africa, and that non-African populations represent a subset of Africans that colonized the world. They estimated that populations carrying the deletion left Africa at most 102,000 years ago.

An alternative explanation is that the population size outside Africa was small—this would also result in reduced variation relative to Africa. However, such a small population, spread over Eurasia, would need to be united by extensive gene flow and simultaneously be isolated from Africa in order for the same haplotype to become common everywhere. This seems unrealistic<sup>41</sup> and thus the *CD4* data are most easily explained by a colonization from Africa of the rest of the world.

Recently, Nei and Takezaki<sup>42</sup> reanalysed five different data sets based on frequencies of microsatellites, restriction fragment length polymorphisms, protein polymorphisms and insertions of *Alu* elements in different human populations, using chimpanzees as an outgroup. In all five cases, the roots of the trees relating the human populations to each other fell on the branch connecting Africans and non-Africans (Fig. 3), indicating that the former have the longest independent history. In four of the data sets support for this conclusion was statistically significant. Furthermore, in an attempt to date the divergence between African and non-African populations based on microsatellites, Goldstein *et al.*<sup>43</sup> have dated the divergence to 56,000–213,000 years by assuming that microsatellites evolve by the gain and loss of single repeat units. Although such a mutational model is highly simplified, it lends further support to the recent African origin of our gene pool.

In summary, it is the study of many different genetic regions with a coalescent approach that will eventually tell us if all, most, or just a fraction of our gene pool traces its ancestry back to Africa. The admittedly scanty data available to date indicate that an African origin is likely for many genetic regions<sup>44,45</sup>.

**Agricultural expansions**

Similar approaches to the study of the patterns of genetic diversity can be used to reveal more recent events in our history; for example, the relationship between the Saami and Finns (Fig. 4). These two groups live in the same region of Northern Fenno-Scandia and are linguistically closely related, yet differ in that the Saami rely on reindeer herding and, traditionally, hunting and

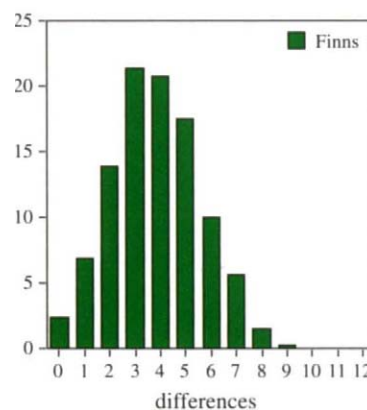
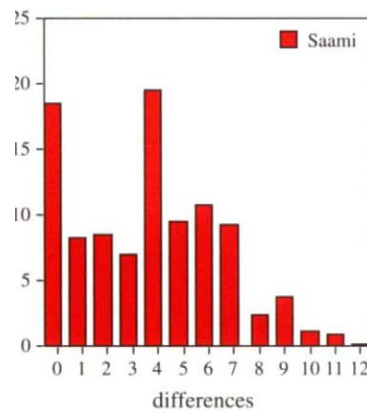


Fig. 4 Distributions of pairwise sequence difference for the Saami and the Finns.

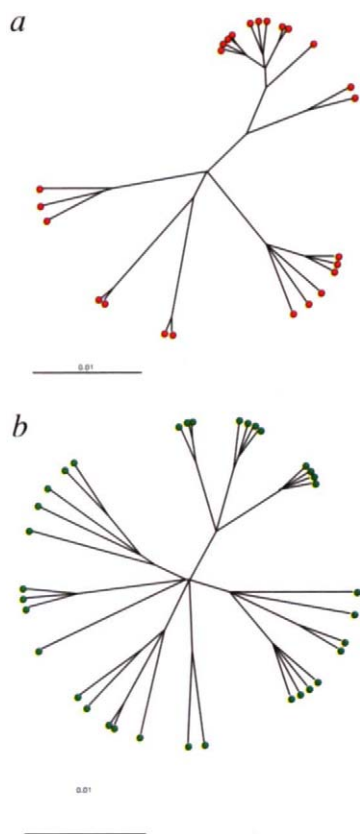


Fig. 5 Phylogenetic trees estimating the relationships of mitochondrial DNA sequences in the Saami (a) and the Finns (b). The trees are reconstructed using DNAML of PHYLIP 3.57c (ref. 58) with a clock assumption and a transition/transversion ratio of 15.

gathering for their subsistence whereas the Finns rely on agriculture and animal husbandry. It can be seen that the Saami display a ragged distribution compatible with a constant population size over time, whereas the Finns show a bell-shaped distribution suggestive of a past population expansion (compare Fig. 4 and Fig. 1c). This is reflected also in the phylogenetic trees estimating the relationships of Saami and Finnish sequences — the Finns show a star-like tree whereas the Saami show a tree with a few deep divergences (Fig. 5). The number of variable positions observed in the Saami sequences is compatible with a constant population size, whereas variation in the Finnish sequences falls outside the 95% confidence interval of what is expected for a population of constant size, indicating that the Finns have expanded (Table 1).

In addition, the Finns appear to share many mitochondrial lineages with other agricultural groups in Europe that speak very different languages, such as the English, Germans and Italians, while the Saami have a long independent history<sup>46</sup>. This implies that the close linguistic affiliation between Finns and Saami is due to the Finns having borrowed their language from Saami ancestors in prehistoric times<sup>47</sup>. More importantly, other groups that, like the Saami, have a non-agricultural mode of subsistence — for example, the Mukri in India<sup>48</sup> and several groups in Africa<sup>49</sup> — show patterns of genetic diversity compatible with constant population size whereas agriculturalists show signs of expansion. This agrees well with studies of gene frequencies in Eurasia that have revealed gradients that can be associated with the spread of agriculture from the Middle East into Europe and western Asia<sup>50</sup>.

Of course, if the peak in pairwise sequence differences (Fig. 4) seen in agricultural populations is caused by the increase in population size that followed the domestication of major crops, then the date derived from the peak should be from less than 10,000 years ago. In contrast to this, the increase has been dated to around 50,000–60,000 years ago<sup>51,52</sup> — much earlier than the origin of agriculture. However, in these calculations, calibrations of the evolutionary rate of the mitochondrial control region based on comparisons of chimpanzee and human mitochondrial sequences were used. Recently, studies of mutations occurring in fami-

lies<sup>53,54</sup> have indicated that the rates may be several-fold faster than those derived from comparisons between species. Much more work is needed before this is understood but it would seem that a likely explanation is that a fraction of the sites in the mitochondrial genome evolve very quickly. Consequently, the rates derived from comparisons with chimpanzees may be too slow for the purpose of dating events that took place comparatively recently<sup>55</sup>. In fact, preliminary, crude rates derived from family studies seem to indicate an age for the Finnish population expansion compatible with the agricultural expansion<sup>56</sup>. Indeed, the invention of agriculture, which led to a dramatic expansion in population size whose end is still not in sight, may well be the event that has had the most profound influence on human gene pool since the origin of our species.

### Perspective

While there is wide-spread excitement and enthusiasm for our genome as a new source of historic information, most of the insights in the field have been widely and hotly debated. The main reason for this is that the technical possibilities to analyse variation in the human genome has grown so quickly that the theoretical tools for their analysis have tended to stay one step behind. In fact, the analysis of the data is currently in a state of flux, where many ideas are emerging, most of them holding only a part, or even just a grain, of the truth.

We believe that besides the acquisition of new data, improved data analysis is crucial for this field. Molecular studies offer a new view of our history that can emancipate us from age-old prejudice that taint many traditional sources of knowledge. However, this potential can only be realized if the data are analysed in a way that is free from preconceived ideas about what constitutes, for example, 'nations,' 'peoples,' or 'races.' This is all the more urgent because we are all so easily drawn into believing 'stories' about our history, particularly if they speak to our preconceived notions of who we are and where we come from. Therefore, careful analysis of the data, and of our prejudices, take on a particular importance in this field.

Furthermore, the fact that molecular data offer a fresh view of our history can lead to an unjustified arrogance, where geneticists may feel that they contribute an 'ultimate truth' whereas in fact they study just one, often very limited, aspect of history. For example, while it is true that many political and cultural concepts in Europe trace their origins to ancient Greece, only very few European genetic lineages may be derived from the Aegean area. This, of course, does not negate or diminish the cultural affiliation that many Europeans feel with ancient Greece. Thus, in many instances, biological history is of much lesser significance for our understanding of ourselves than other aspects of history. With this in mind, however, molecular genetics provides a powerful tool to investigate the origin and history of the human species, and of populations. Combined with solid data analysis, it can yield a truly unique, and indeed emancipatory, contribution to our understanding of ourselves.

### Acknowledgements

We apologise to many colleagues whose contributions could not be mentioned due to space limitations. We thank J. Castresana, B. Cohen, T. Imanishi, M. Laan, P. Morin, J. Rydström, M. Stoneking, N. Takahata, R. Ward, S.-I. Yokobori and others for comments and help with the manuscript. Our work is supported by the Deutsche Forschungsgemeinschaft.

**Table 1** Some features of mtDNA diversity in two Scandinavian populations

Population	Mean sequence difference	Variable positions	
		Expected	Observed
Finns	3.8	7–35	37
Saami	3.8	9–38	29

The mean pairwise sequence difference, the 95% confidence interval of the expected number of variable positions given the assumption of constant population size and the observed number of variable positions are shown for Finns ( $n = 50$ ) and Saami ( $n = 115$ ). More variable positions than expected are found in the Finns, indicating population growth.

1. Slatkin, M. & Hudson, D. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562 (1991).
2. Majoram, P. & Donnelly, P. Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. *Genetics* **136**, 673–683 (1994).
3. Rogers, A. Genetic evidence for a Pleistocene population explosion. *Evolution* **49**, 608–615 (1995).
4. Cann, R.L., Stoneking, M. & Wilson, A.C. Mitochondrial DNA and human populations. *Nature* **325**, 31–36 (1987).
5. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A.C. African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503–1507 (1991).
6. Ruvolo, M. *et al.* Mitochondrial COII sequences and modern human origins. *Mol. Biol. Evol.* **10**, 1115–1135 (1993).
7. Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92**, 532–536 (1995).
8. Stoneking, M., Sherry, S.T., Redd, A.J. & Vigilant, L. New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil. Trans. R. Soc. Lond. B* **337**, 167–175 (1992).
9. Pesole, G., Sbisà, E., Preparata, G. & Saccone, C. The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol. Biol. Evol.* **9**, 587–598 (1992).
10. Wills, C. When did Eve live? An evolutionary detective story. *Evolution* **49**, 593–607 (1995).
11. Dorit, R.L., Akashi, H. & Gilbert, W. Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* **268**, 1183–1185 (1995).
12. Weiss, G. & von Haeseler, A. How many Ys? *Science* **272**, 1358–1360 (1996).
13. Hammer, M.F. A recent common ancestry for human Y chromosomes. *Nature* **378**, 376–378 (1995).
14. Fu, Y.-X. & Li, W.-H. The age of the common ancestor of human Y chromosomes. *Nature* (in the press).
15. Takahata, N. Allelic genealogy and human evolution. *Mol. Biol. Evol.* **10**, 2–22 (1993).
16. Morin, P.A. *et al.* Kin selection, social structure, gene flow and the evolution of chimpanzees. *Science* **265**, 1193–1201 (1994).
17. Ferris, S.D., Brown, W.M., Davidson, W.S. & Wilson, A.C. Extensive polymorphism in the mitochondrial DNA of apes. *Proc. Natl. Acad. Sci. USA* **78**, 6319–6323 (1991).
18. Ruvolo, M. *et al.* Gene trees and hominoid phylogeny. *Proc. Natl. Acad. Sci. USA* **91**, 8900–8904 (1994).
19. Blanchard, J.L. & Schmidt, G.W. Mitochondrial DNA migration events in yeast and humans: Integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.* **13**, 537–548 (1996).
20. Crouau-Roy, B., Service, S., Slatkin, M. & Freimer, N. A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. *Hum. Mol. Genet.* **8**, 1131–1137 (1996).
21. Nei, M. & Graur, D. Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17**, 73–118 (1984).
22. Li, W.H. & Sadler, L.A. Low nucleotide diversity in man. *Genetics* **129**, 513–523 (1991).
23. Hedrick, P. Evolutionary genetics of the major histocompatibility complex. *Am. Naturalist* **143**, 945–964 (1994).
24. Gyllensten, U.B., Sundvall, M. & Erlich, H.A. Allelic diversity is generated by intraexon exchange at the *DRB1* locus of primates. *Proc. Natl. Acad. Sci. USA* **88**, 3686–3690 (1991).
25. Takahata, N., Satta, Y. & Klein, J. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**, 925–938 (1992).
26. Ayala, F.J. The myth of Eve: molecular biology and human origins. *Science* **270**, 1930–1936 (1995).
27. Ayala, F.J. & Escalante, A.A. The evolution of human populations: A molecular perspective. *Mol. Phylog. Evol.* **5**, 188–201 (1996).
28. Titus-Trachtenberg, E.A., Rickards, O., De Stefano, G.F. & Erlich, H.A. Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: A novel *DRB1* allele reveals evidence of convergent evolution and balancing selection at position 86. *Am. J. Hum. Genet.* **55**, 160–167 (1994).
29. Gustafsson, K. *et al.* Structure of miniature swine class II *DRB* genes: conservation of hypervariable amino acid residues between distantly related species. *Proc. Natl. Acad. Sci. USA* **87**, 9798–9802 (1990).
30. Gustafsson, K. & Andersson, L. Structure and polymorphism of horse MHC class II *DRB* genes: convergent evolution in the antigen binding site. *Immunogenet.* **39**, 355–358 (1994).
31. Maddison, D.R. African origin of human mitochondrial DNA reexamined. *Syst. Zool.* **40**, 355–363 (1991).
32. Templeton, A.L. Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737 (1992).
33. Hedges, S. B., Kumar, S., Tamura, K. & Stoneking, M. Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737–739 (1992).
34. Maddison, D.R., Ruvolo, M. & Swofford, D.L. Geographic origins of human mitochondrial DNA phylogenetic inference from control region sequences. *Syst. Biol.* **41**, 111–124 (1992).
35. Penny, D., Steel, M., Waddell, P.J. & Hendy, M.D. Improved analyses of human mtDNA sequences support a recent African origin of *Homo sapiens*. *Mol. Biol. Evol.* **12**, 863–882 (1995).
36. Strimmer, K. & von Haeseler, A. (1996) Quartet puzzling: A maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964–969 (1996).
37. D'Andrade, R. & Morin, P.A. A principal components and individual-by-site analysis of chimpanzee and human mitochondrial DNA. *Am. Anthropol.* **98**(2), 2–20 (1996).
38. Zischler, H., Geisert, H., von Haeseler, A. & Pääbo, S. A nuclear "fossil" of the mitochondrial D-loop and the origin of modern humans. *Nature* **378**, 489–492 (1995).
39. Wainscoat, S. *et al.* Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* **319**, 491–493 (1986).
40. Tishkoff, S.A. *et al.* Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science* **271**, 1380–1387 (1996).
41. Manderscheid, E.J. & Rogers, A.R. Genetic admixture in the late Pleistocene. *Am. J. Phys. Anthropol.* (in the press).
42. Nei, M. & Takezaki, N. The root of the phylogenetic tree of human populations. *Mol. Biol. Evol.* **13**, 170–177 (1996).
43. Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. & Feldman, M.W. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* **92**, 6723–6727 (1995).
44. Tishkoff, S.A. *et al.* A global haplotype analysis of myotonic dystrophy CTG-repeats in humans and other primates. *Am. J. Hum. Genet.* **57**, A42 (1995).
45. Armour, J.A.L. *et al.* Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet.* **13**, 154–160 (1996).
46. Sajantila, A. *et al.* Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* **5**, 42–52 (1995).
47. Sajantila, A. & Pääbo, S. Language replacement in Scandinavia. *Nature Genet.* **11**, 359–360 (1995).
48. Mountain, J.L., Hebert, J.M., Bhattacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M. & Cavalli-Sforza, L.L. Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.* **56**, 979–992 (1995).
49. Watson, L., Bauer, K., Aman, R., Weiss, G., von Haeseler, A. & Pääbo, S. MtDNA sequence diversity in Africa. *Am. J. Hum. Gen.* **59**, 437–444 (1996).
50. Cavalli-Sforza, L.L. & Ammermann, A. *The neolithic transition and the genetics of populations in Europe*. (Princeton University Press, 1984).
51. Rogers, A. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
52. Sherry, S.T. *et al.* Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* **66**, 761–775 (1994).
53. Ivanov, P.L. *et al.* Mitochondrial sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genet.* **12**, 417–420 (1996).
54. Howell, N., Kubacka, I. & Mackey, D.A. How rapidly does the human mitochondrial genome evolve? *Am. J. Hum. Genet.* **59**, 501–509 (1996).
55. Pääbo, S. Mutations in the mitochondrial microcosm. *Am. J. Hum. Genet.* **59**, 493–496 (1996).
56. Sajantila, A. *et al.* Mitochondrial and Y chromosome lineage analyses reveal founder effect in the Finnish population. *Proc. Natl. Acad. Sci. USA*, (in the press).
57. Richards, M. *et al.* Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* **59**, 185–203 (1996).
58. Felsenstein, J. Phylogenetic Inference Programs (PHYLP), University of Washington, Seattle and University Herbarium, University of California, Berkeley. (1993).