Springer
Berlin
Heidelberg
New York
Hongkong
London
Mailand
Paris
Tokio

Ernst Schering Research Foundation Workshop
Supplement 10

Die Architektur des Lebens – über Gene, Organismen und Personen

Symposium anlässlich des 10-jährigen Bestehens der Schering Forschungsgesellschaft, November 2001

Monika Lessl, Günter Stock
Herausgeber

Springer

Series Editors: G. Stock and M. Lessl

Max-Planck-Institut
für evolutionäre
Anthropologie
Bibliothek

ISSN 1431-7133
ISBN 3-540-43881-5 Springer-Verlag Berlin Heidelberg New York

Die Deutsche Bibliothek - CIP-Einheitsaufnahme
Die Architektur des Lebens : über Gene, Organismen und Personen ; Festvorträge anlässlich des 10jährigen Bestehens der Schering Forschungsgesellschaft November 2001 / Hrsg.: Günter Stock ; Monika Lessl. - Berlin ; Heidelberg ; New York ; Barcelona ; Hongkong ; London ; Mailand ; Paris ; Tokio : Springer, 2003
(Ernst Schering Research Foundation Workshop : Supplement ; 10)
ISBN 3-540-43881-5

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer-Verlag Berlin Heidelberg New York
ein Unternehmen der BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de

© Springer-Verlag Berlin Heidelberg 2002
Printed in Germany

Umschlaggestaltung: design & production, Heidelberg
Satz: Datenkonvertierung im Verlag
Druck: Druckhaus Beltz, Hemsbach
Bindung: J. Schäffer GmbH & Co. KG, Grünstadt
Gedruckt auf säurefreiem Papier SPIN: 10884650 21/3130/ag – 5 4 3 2 1 0

Inhalt

4 Die Wurzeln der Menschheit – die Evolution des humanen Genoms

Svante Pääbo

Svante Pääbo vom Max-Planck-Institut für Evolutionäre Anthropologie erläuterte in seinem Vortrag wie DNA-Sequenzanalysen und Genexpressionsstudien zur Klärung des Verwandtschaftsverhältnisses zwischen Menschen und anderen Organismen, insbesondere Menschenaffen, beitragen können. Der Inhalt seines Vortrages wurde in einem Review-Artikel, der im Journal of Internal Medicine 2002 (Vol. 251:1–18) erschien, umfassend dargestellt. Nachfolgend finden Sie daher einen Abdruck des oben erwähnten Artikels.

The Genetical History of Humans and the Great Apes

Abstract

Kaessmann H, Pääbo S (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany). The genetical history of humans and the great apes. *J Intern Med* 2002; 251: 1–18.

When and where did modern humans evolve? How did our ancestors spread over the world? Traditionally, answers to questions such as these have been sought in historical, archaeological, and fossil records. However, increasingly genetic data provide information about the evolution of our species. In this review, we focus on the comparison of the variation in the human gene pool to that of our closest evolutionary relatives, the great apes, because this provides a relevant perspective on human genetical evolution. For instance, comparisons to the great apes show that humans are unique in having little genetic variation as well as little genetic structure in their gene pool. Furthermore, genetic data indicate that humans, but not the great apes, have experienced a period of dramatic growth in their early history.

4.1 Introduction

Our genome consists of about 3 billion nucleotides which have been passed down to us from our ancestors. In every generation, several of these nucleotides are affected by mutations in the male and female germ-line so that subsequent generations receive slightly different versions of the ancestral genomes. Depending on whether human populations expand, contract, migrate, split or exchange migrants, these mutations accumulate in characteristic ways. Thus, by surveying enough genetic variation in sufficient number of human individuals from around the world, the genetical history of the human species is, in principle, ascertainable. Here, we review what studies of variation in human DNA sequences have revealed, with a particular emphasis on comparisons with the great apes.

4.2 DNA Variation in Humans

4.2.1 Mitochondrial DNA

Most of the early molecular data on human population history were derived from the DNA of mitochondria (mtDNA) [1], which are organelles responsible for much of the energy metabolism in eukaryotic cells[2, 3]. They carry a genome that in humans contains about 16500 base pairs (bp) [4]. It lends itself well to the study of human evolution because of an apparent lack of recombination [5] and a high mutation rate [6]. The absence of recombination (exchange of DNA fragments between maternal and paternal chromosomes in the germline) is an advantage in evolutionary studies because recombination reshuffles ancestral DNA sequences that have been carried by different ancestors. Thus, for sequences that have been affected by recombination it is not possible to reconstruct one unambiguous phylogenetic tree for the entire sequence. The high mutation rate is also an advantage because it has allowed the mtDNAs to accumulate many differences over relatively short time. Hence only short sequences need to be determined from each individual to gather enough information for estimating the relationships of mtDNAs. A further property particular to mtDNA is that it is inherited from mother to offspring while fathers do not pass on any mtDNA [7]. Therefore, the variation of mtDNA sequences in humans reflects the history of females.

mtDNA variation data was initially assessed by restriction fragment length polymorphism (RFLP; a technique based on the sequence specific cleavage of DNA by restriction enzymes) [8–12]. In an early influential paper, Cann *et al.* [9] studied humans from around the world and proposed 'that all contemporary human mtDNAs trace back through maternal lineages to an ancestral mtDNA present in an African population some 200 000 years ago'. This conclusion was strengthened by a subsequent study [13] where two hypervariable (HVRI and HVRII) [14] regions within the noncoding control region of the mtDNA were sequenced. A schematic phylogenetic tree illustrating the general pattern of relationships among mtDNA sequences seen in these studies (Fig. 1) shows that the branches next to the root (i.e. the last common ancestor of all human mtDNA lineages) consist exclusively of African sequences,

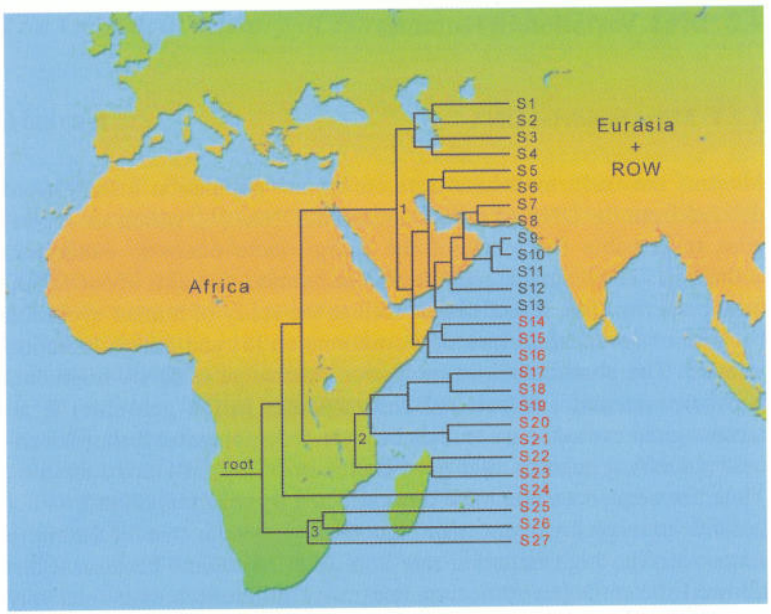


Fig. 1. Schematic tree illustrating the branching pattern seen in trees relating mtDNA and Y chromosome sequences. Several lines of evidence suggest an African origin for the variation among the DNA sequences. All lineages next to the root are African; and of 27 sequences depicted, 14 sequences are carried by Africans (S14-S27) and 13 (S1-S13) by non-Africans (Eurasians and individuals from the rest of the world, ROW). Furthermore, Africans are found in all major clades defined by the nodes 1–3, whereas non-Africans are found in one (node 1). The most parsimonious explanation for this pattern that minimizes the number of migrations and extinctions of mtDNAs is that a subset of African individuals carrying ancestral mtDNAs migrated from Africa into the rest of the world leaving additional variation behind. Thus, non-Africans carry a subset of the African mtDNA variation with additional mutations that occurred after the migration from Africa

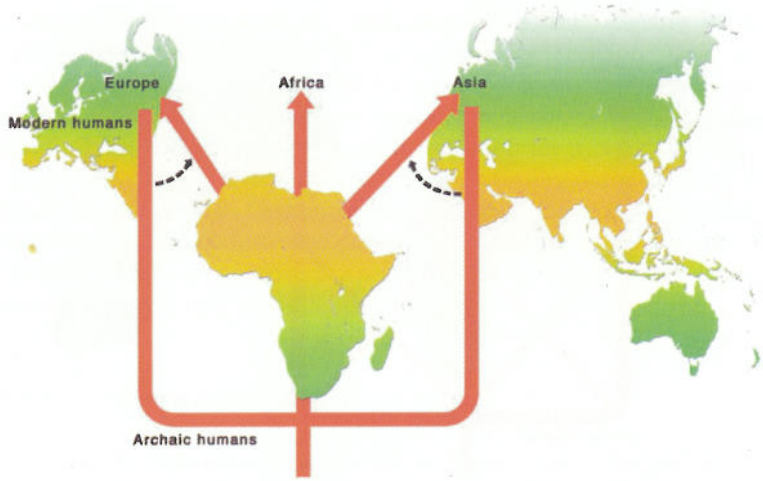


Fig. 2. African replacement hypothesis. According to this hypothesis, modern humans evolved from archaic ancestors (*Homo erectus*) only in Africa. Archaic humans living in Asia (e.g. 'Java Man') and Europe (Neandertals) were replaced by modern humans migrating out of Africa

indicating that the oldest lineages are found among Africans. Furthermore, African lineages are found throughout the tree. The explanation for this pattern that minimizes the number of migrations and extinctions of mtDNAs is that a subset of African individuals carrying ancestral mtDNAs migrated from Africa into the rest of the world leaving additional variation behind. Thus, non-Africans carry a subset of the African mtDNA variation with additional mutations that occurred after the migration from Africa.

As results obtained from the two hypervariable regions of the mtDNA are confounded by recurrent mutations at the same nucleotide site (because of the extremely high mutation rate), it is interesting to note that a recent study where all 16500 bp of 53 mitochondrial genomes were determined concur, in general, with the results obtained from the hypervariable regions alone [15]. Thus, the origin of the mtDNA variation of modern humans is to be found in Africa. When extrapolated to the entire human species, this view is often referred to as the 'Out of Africa' or 'African replacement' [16] hypothesis. It stands,

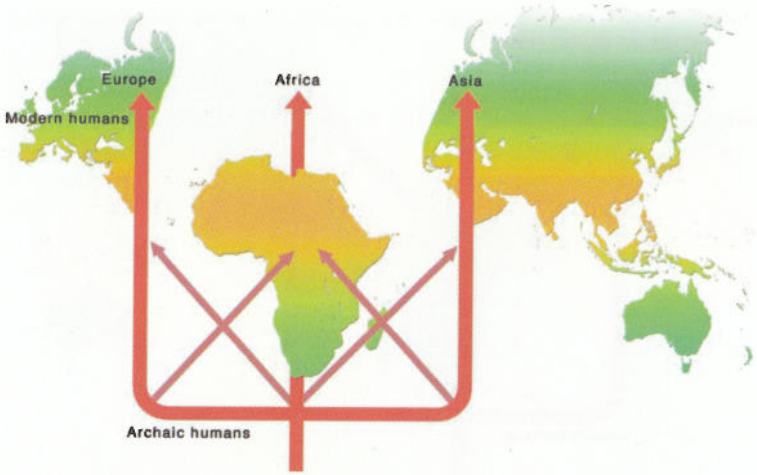


Fig. 3. Multiregional hypothesis. According to this hypothesis, archaic human ancestors (*Homo erectus*) evolved into modern humans in Africa, Asia, and Europe with some migration (gene flow) in between these areas (indicated by thin arrows)

supported by a common interpretation of the fossil evidence [17], that anatomically modern humans originated in Africa as descendants of African *Homo erectus* ancestors (Fig. 2) sometime between one and two hundred thousand years ago [16]. Subsequently, modern humans colonized the rest of the old world and replaced archaic human forms such as the Neandertals in Europe [17,18]. In a variant form of this hypothesis modern humans originated in Africa but may not have completely replaced archaic forms elsewhere [19]. These archaic humans were in turn descendants of *H. erectus* which started to leave Africa almost 2 million years ago [18]. An alternative model for modern human origins (Fig. 3) holds that the major pattern is regional continuity of these archaic human forms in Africa and various regions of Eurasia, with gene flow allowing for some character traits to move between groups [20].

Proponents of these two views often debate their standpoints with great fervour. However, it should not be forgotten that from a genetic perspective, each part of our genome may have its own history, such that (leaving recombination aside) the history of some parts of our genome

could be compatible with the 'African replacement' scenario while others could conform to the 'regional continuity' scenario. Thus, although the history of the mtDNA clearly conforms to the 'African replacement scenario' it represents just one single genetic locus that as a result of either chance occurrences or selection for a certain mtDNA variant could have a history different from the history of the majority of the DNA sequences in the human genome. Since, not every locus must conform to the same historical pattern, it is important to analyse other parts of the genome to arrive at a sense of what a general picture of our genetic history might entail.

4.2.2 Y Chromosome Variation

The locus that offers itself as an obvious counterpart to the mtDNA is the Y chromosome. It has many of the same attractions for evolutionary studies as mtDNA but it lacks recombination and is paternally inherited, i.e. it reflects the history of males just as mtDNA reflects the history of females. However, because of both the small effective population size of the Y chromosome (only males in the population carry a Y chromsome; thus there will be fewer mutations that accumulate over time on all Y chromosomes in the population as compared with, for example, the autosomes for which two copies each are carried by males as well as females) and the overall low mutation rate of nuclear DNA, the extent of DNA sequence variation on the Y chromosome is low compared with the mtDNA. Therefore, elaborate techniques are usually used to identify particular variable nucleotides (SNPs, single nucleotide polymorphisms) that are then studied in populations [21]. In agreement with mtDNA trees, Y chromosome trees show that Africans carry most ancestral lineages as well as considerably more genetic variation than non-African populations [22] (Fig. 1). The majority of Y chromosome studies furthermore provide age estimates of the most recent common ancestor (MRCA), i.e. the age estimate of the oldest divergences currently seen in the gene pool, that agree well with estimates from mtDNA (100 000–200 000years [15, 23, 24]). Thus, the overall picture shows that the Y chromosome and mtDNA experienced similar evolutionary histories. However, a recent study based on a greater number of polymorphisms than previously studied suggests an age of the MRCA for the Y chromo-

some of only 59 000 years [25]. If this point estimate is correct, this may be because of a significantly reduced variability of the Y chromosome relative to other genomic loci caused by selection [26–28]. However, the confidence range around this estimate is large and the upper bound (140 000 years) [25] is well within the range of estimates obtained for the mtDNA ancestor.

Interestingly, some discrepancies occur between the mtDNA and the Y chromosomal pictures that bear on more recent human history. For example, phylogenetic trees based on Y chromosome data show more geographical clustering than mtDNA sequences [29]. It has been suggested, that this reflects a higher female than male migration rate, that is, that females would have moved around more than males among human populations [30]. This 'Women on the move' hypothesis [30] probably reflects the predominance of patrilocal societies, in which wives tend to move into their husband's natal households.

4.2.3 Autosomal DNA Sequence Variation

The vast majority of the nuclear genome is located on the autosomes. In order to assess its diversity approaches other than DNA sequencing have often been used. For instance, studies of microsatellites (short nucleotide repeats that have a high mutation rate and are highly variable in length) show that the majority of length variants and hence the greatest genetic diversity is found among African genomes [31]. Alu-insertion polymorphisms, i.e. the presence or absence of repetitive short interspersed nuclear elements (300 bp) that replicate and 'jump' around in the genome [32], show that the ancestral sequences (absence of the Alu insertion) are typically found in Africans [33, 34].

Until recently, little was known about the extent of variation in single-copy DNA sequences on the autosomes. However, by now population studies of nuclear DNA sequences have been published for genes encoding-globin [35], lipoprotein lipase [36, 37], pyruvate dehydrogenase E1 -subunit (PDHA1) [38] and the melanocortin 1 receptor (MC1R) [39]. In general, these loci show that the ancestral sequence is found among Africans and that the age of the MRCA ranges from approximately 710 000 years (MC1R) to 1 590 000 years (PDHA1) [40]. In addition, major histocompatibility complex (MHC) genes have

been analysed (reviewed in [41]). Interestingly, comparison of human MHC variability with that of other species reveals that there is extensive sharing of MHC polymorphisms [42, 43].

However, all of these autosomal loci may be less than ideal for reconstructing population history because they were originally targeted for study as a result of their involvement in medical conditions, such as haemoglobinopathies (-globin) [44], cardiovascular disease (LPL) [45] and neurological diseases (PDHA1) [46]. Furthermore, MC1R and the MHC genes are involved in skin pigmentation and the immune response, respectively. Thus, these loci are likely to be the direct target of selection which may cause the distribution of different gene variants to reflect selection for certain traits in the carriers of the gene rather than 'population' events such as migrations, bottlenecks, and expansions. For instance, it is known that there is local selection in Africa for certain-globin alleles that confer resistance to malaria [44, 47]. Similarly, it is well established that some form of balancing selection must operate to hold MHC alleles in the population for millions of years [41, 48]. Finally, relatively high recombination rates at many of these genes complicate evolutionary analyses.

4.2.4 Variation at Xq13.3

A nuclear DNA sequence that exhibits several characteristics potentially advantageous for elucidating human population history is a sequence of approximately 10000 bp at Xq13.3 on the long arm of the X chromosome [49]. First, it is noncoding and hence unlikely to be the direct target of selection. Secondly, it displays a recombination rate that is about eight times lower than the average X chromosomal recombination rate [50]. The low recombination rate is unlikely to cause much reshuffling of ancestral DNA sequences that would confound many methods of evolutionary inference [51]. Thirdly, Xq13.3 is advantageous from a practical standpoint because when males, who carry only one X chromosome, are studied the sequencing results are not compounded by the occurrence of two different DNA sequences.

In order to gauge human genetic diversity world-wide at a locus such as Xq13.3, there are at least three possible sampling strategies, all of them with advantages and disadvantages. First, samples could be col-

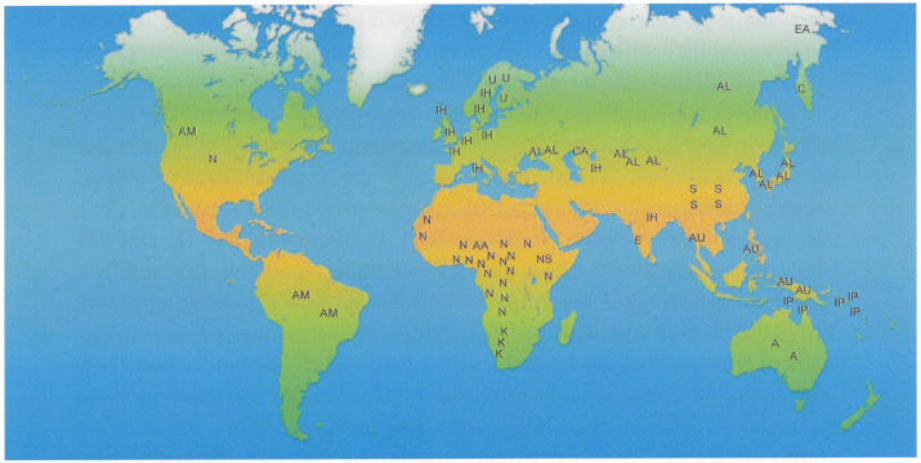
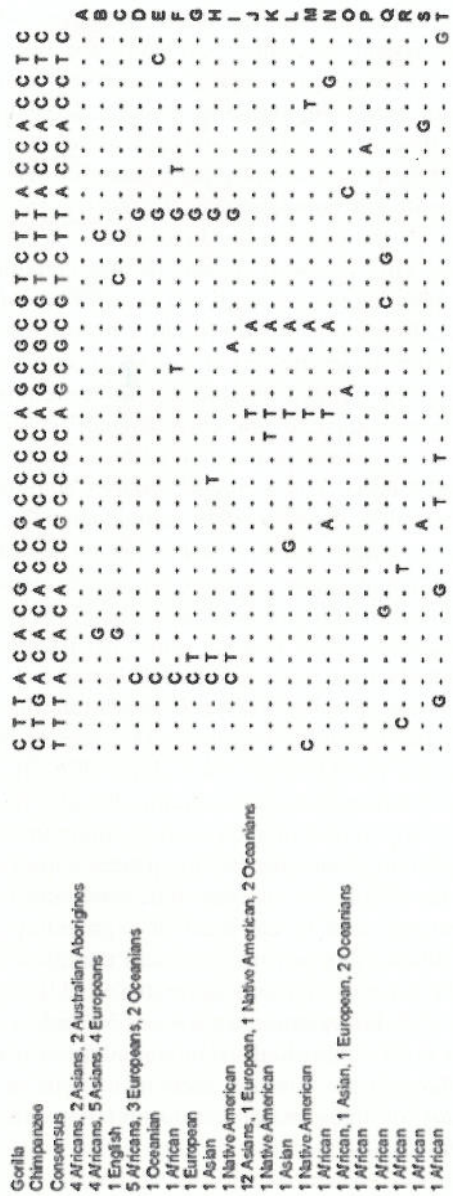


Fig. 4. Map of the world indicating the approximate places of origin of the humans studied with respect to their Xq13.3 DNA sequences. The individuals belong to the following 17 language phyla as defined by Ruhlen [52] (abbreviations in brackets): Khoisan ('K'), Niger-Kordofanian ('N'), Nilo-Saharan ('NS'), Afro-Asiatic ('AA'), Caucasian ('CA'), Indo-Hittite ('IH'), Uralic-Yukaghir ('U'), Altaic ('AL'), Chukchi-Kamchatkan ('C'), Eskimo-Aleut ('EA'), Elamo-Dravidian ('E'), Sino-Tibetian ('S'), Austric ('AU'), Indo-Pacific ('IP'), Australian ('A'), Amerind ('AM'), Na-Dene ('N'). Modified from Kaessmann *et al.Nature Genet*

Fig. 5. Variable positions found in the human Xq13.3 DNA sequences. Individuals with identical sequences are grouped together. Letters on the right (A-T) denominate the different sequences that are defined by the variable positions. '-' indicates the consensus nucleotide. Orthologous nucleotides of a chimpanzee and a gorilla are shown at the top. Modified from Kaessmann *et al.Nature Genet.*, 1999



lected according to current population size. However, when long-term history is studied, recent population expansions, for example associated with the invention of agriculture starting some 10000 years ago, as well as recent colonizations, will influence results. Secondly, individuals could be sampled according to the land area where they live so that a certain number of people are sampled per square mile. However, in such as scheme areas with low population density (such as the Arctic) will be over-represented. A third possibility is to sample according to major linguistic groups. This approach appears to be preferable for our purposes because major language groups are likely to be older than recent expansions and colonizations known from the historic record. Thus, we chose 70 individuals for sequencing at Xq13.3 in such a way that all 17 major language phyla defined by Ruhlen [52] were represented by a minimum of one individual. Within each language phylum individuals were chosen to be as diverse as possible with regard to both linguistic and geographical criteria. This sampling strategy automatically results in a wide geographical distribution of individuals (Fig. 4). For each individual, approximately 10200 consecutive nucleotides at Xq13.3 were determined. An alignment of these sequences reveals 33 variable positions that define 20 different sequences (fig. 5) [49].

The DNA sequences show that the same Xq13.3 sequence variants occur in individuals from very different geographical and linguistic backgrounds. For instance, sequence B is present in four Africans, five Asians and four Europeans, and sequence D is present in five African, three European and two Oceanian individuals (Fig. 5). Extensive genetic homogeneity among human populations has also been observed in other studies [16] which show that on average approximately 90% of the total genetic variation in humans (where genetic variation is measured by estimating the number of polymorphic sites and how often one nucleotide occurs relative to the other at a polymorphic site in the population) is contained in any major human population and that only about 10% of the genetic variation can be attributed to between-population differences [16]. For example, a survey of 30 nuclear restriction-site polymorphisms in 243 individuals from around the world revealed that 89% of the total genetic variation seen in humans at these sites is contained in any of three major continental populations (Africans, Asians, and Europeans) [53].

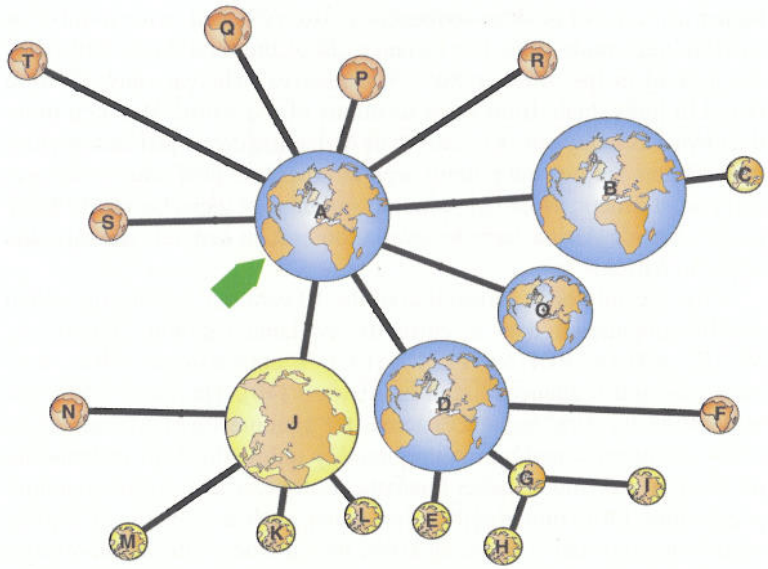


Fig. 6. Phylogenetic network relating human Xq13.3 DNA sequences. Letters refer to sequence designation in Fig. 5. Red circles (showing the African continent) represent sequences observed only in Africa, 'yellow' circle sequences observed only outside Africa, 'blue' circle sequences observed in Africa as well as in other regions of the world. The area of the circles represents the number of individuals carrying each sequence. The location of the ancestral sequence is indicated by the arrow. Figure modified from Kaessmann *et al.Nature Genet.*, 1999

By comparison to the Xq13.3 sequences determined in chimpanzees and gorillas it was found that sequence A carries the ancestral nucleotide state at most of the variable sites while all other sequences differ by at least two additional nucleotides from the ancestral sequence (Fig. 5). Thus, sequence A is the root of the network shown in fig. 6. It is carried by Africans as well as non-Africans and is therefore not informative with regard to the geographical origin of the sequence variation at Xq13.3. However, one or more Africans are represented in all nine branches originating from the ancestral sequence A, whereas non-Africans are present on only four of these lineages. Africans are, therefore, more widely distributed in the tree than non-Africans. A generally

higher diversity of African sequences is also reflected in the number of variable nucleotide sites. For instance, 24 of the total 33 variable sites are present in the African Xq13.3 sequences, whereas only 17 were found in individuals from other locations of the world, although more than twice as many non-Africans than Africans were sequenced. Xq13.3 thus reflects the general pattern seen at other loci [16] where Africans carry most genetic diversity including variants not seen elsewhere, while people outside Africa carry less variation and, in general, variants also found in Africa.

It has recently been shown that of the 10 genomic regions for which suitable population data is currently available (-globin, *dys44*, Gk, MC1R, mtDNA, PDHA1, PLP, Xq13.3, Y chromosome, ZFX), nine have ancestral sequences found in Africans [40]. The observation that 90% of the ancestral sequences across the genome are of African origin is not easily explained by multiregionality unless the African founding population was much larger than the European and Asian founding populations [40]. Importantly, even under such a scenario of a great asymmetry in relative breeding sizes, the genetic contribution of non-African founding populations to the modern human gene pool would be extremely small [40].

4.2.5 Age of the Variation at Xq13.3

In order to date the variation of the Xq13.3 sequences, the evolutionary rate was determined by comparison with the chimpanzee and gorilla sequences. These differ from the human sequences by an average of 94 and 143 substitutions, respectively. Assuming that the human and chimpanzee lineages split about 5million years ago [54–58], about 100 sequence differences accumulated in 5 million years between humans and chimpanzees. This adds up to an average of 1 difference per 50 000 years. Therefore, assuming that half of the mutations occurred in humans and chimpanzees, respectively (evidence for similar mutation rates in humans and chimpanzees is discussed later), approximately 1 mutation per 100 000 years occurred per 10 000 bp at Xq13.3 in humans and chimpanzees.

On the basis of this mutation rate as well as assuming that sequence A is the ancestral sequence, the time of the most recent common ancestor for Xq13.3 can be determined using the coalescent [59, 60], which takes into account not only the number of mutations but also the process of genetic drift, i.e. the stochastic 'survival' or 'death' of alleles caused by the fact that some individuals in every generation have no offspring whereas others have several. The drift process results in the gradual joining of genetic lineages until there is only one lineage left. This is the most recent common ancestor or MRCA (Fig. 7). For Xq13.3, the MRCA was estimated to be approximately 500 000 years. It is important

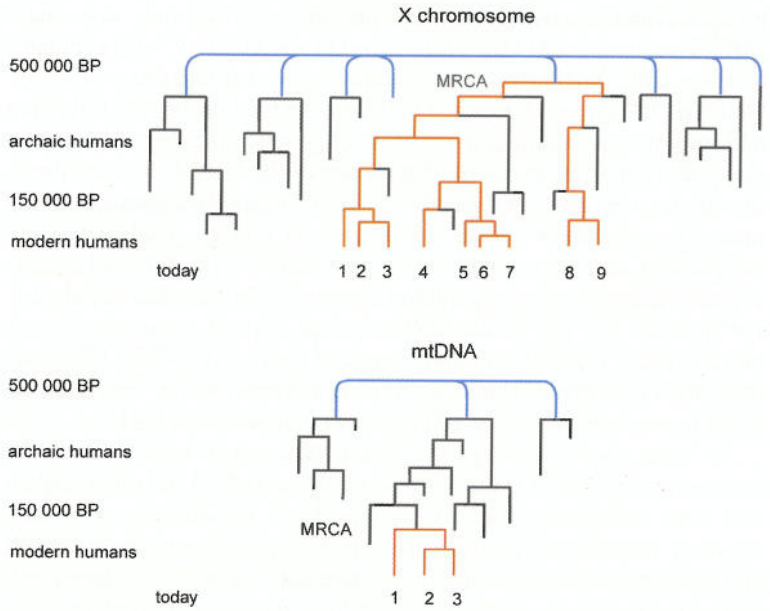


Fig. 7. Illustration of the coalescent process in humans at an X chromosomal locus (such as Xq13.3) as well as for mtDNA. All lineages (in red) present in the human population today trace back to one common ancestor. The threefold difference in the effective number of Xq13.3 versus mtDNA sequences in the population leads to an approximately threefold greater age of the MRCA for the variation in Xq13.3 DNA sequences than in mtDNA. Thus, the age of the variation at Xq13.3 is older than the modern human species that emerged about 100 000 years ago. Note also that other humans lived at the same time as those carrying the ancestral Xq13.3 and mtDNA sequences, but their DNA lineages have not survived until today

to realize that this does not mean that there existed only one single Xq13.3 sequence 500 000 years ago. On the contrary, many humans carrying other Xq13.3 variants doubtlessly lived at that time. However, their Xq13.3 lineages happened to become extinct during the following half million years through the process of genetic drift (Fig. 7). Furthermore, it is important to realize that the age of the MRCA does not imply the emergence of a new 'species' or group of modern humans at that time. The population where one or several changes lead to modern morphology and behaviour (however defined) surely contained genetic variation tracing back to a common ancestor older than that population. For instance, the age of the MRCA for all Xq13.3 sequences seen in humans today is about 500 000 years and hence at least 370 000 years older than the first anatomically modern humans, which appear in the fossil record approximately 100 000–130 000 years ago [61].

As seen above, the MRCAs of mtDNA and Y chromosomal sequences are severalfold younger than those for Xq13.3 and other nuclear loci. One explanation for this discrepancy is the differences in their mode of inheritance. Because the effective population size of the X chromosome (two copies in females, one copy in males) is three times that of mtDNA (one sequence variant from females is passed on to the offspring) and the Y chromosome (one copy in males) and three-quarters of that of autosomes (two copies in males and females), X chromosomal lineages are expected to take three times as long to become extinct and their MRCA is expected to be three times as old (provided that other things are equal) as mtDNA and Y chromosomal DNA sequences (Figs 7 and 8). Nuclear DNA sequences are expected to have MRCAs that are on average four times older than these loci (Fig. 8). The age estimates ranging from approximately 100 000–200 000 years ago for mtDNA and the Y chromosome [15, 24, 62, 63], 500 000 years for Xq13.3 and 750 000 for the -globin gene on chromosome 11 [35] (Fig. 8), are therefore in striking, and in fact surprising, agreement with each other.

If genetic continuity between *H. erectus* and modern human populations in several regions of the world, as suggested by the regional continuity hypothesis (Fig. 3), would apply to Xq13.3 [20], then a time depth of 1 million years or more would be expected because *H. erectus* populations dispersed over a million years ago into the different parts of the world. An age of the MRCA at Xq13.3 of 1 million years or more

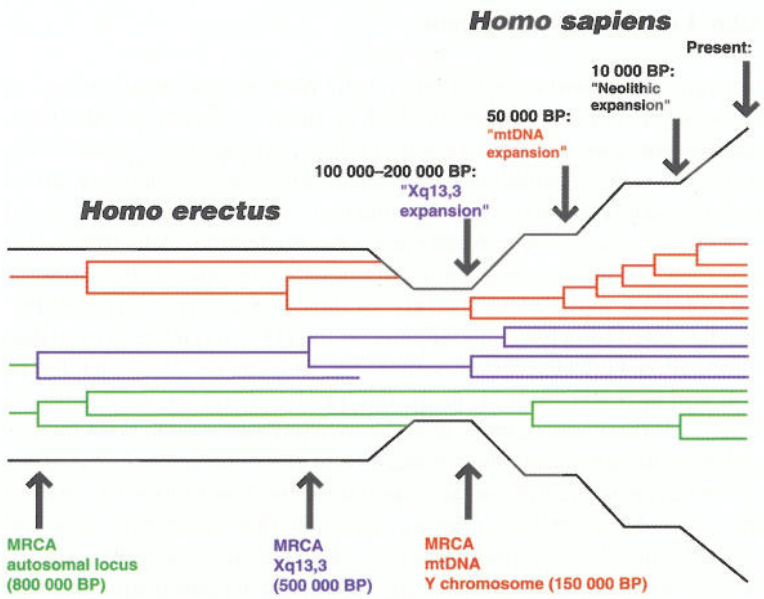


Fig. 8. Illustration of the age of the most recent common human ancestor (MRCA) of different genomic regions, a population bottleneck that may be associated with the origin of modern humans, and subsequent population expansions as reflected by Xq13.3, mtDNA (see text) and allele frequency data ('Neolithic expansion') [109]. The latter reflect a population movement and expansion promoted by the spread of agriculture from the Near East (where it was first developed about 10 000 years ago)

can be excluded at the 1% significance level (P=0.0028). This dating, combined with the relative genetic uniformity of Xq13.3 sequences, the greater African genetic variation and the African origin of most ancestral sequences for different genomic loci (see above), is incompatible with regional continuity, unless extensive migration across the world had occurred among all ancestral populations, and with the major contribution during these migrations coming from Africa.

4.2.6 Linkage Disequilibrium

Not only the extent of variation and the phylogenetic relationships of DNA sequences can be used to estimate the history of a genetic locus, but also the way in which variation along the sequence on a chromosome (so-called 'haplotypes') correlate between sites as a result of mutations and recombination. Assume that, for example, two SNPs exist in a DNA sequence, where one carries the nucleotides A in 10% and G in 90% of chromosomes in a population, while the other SNP carries the nucleotides C in 40% and T in 60% of the chromosomes. If recombination has acted for a long enough time on the DNA sequence to exchange information between chromosomes, the frequencies of the haplotype A-C will be 0.300.70=0.21, the haplotype G-C 0.700.70=0.49 and so on (Fig. 9). If this expectation is met, the two SNPs are said to be in linkage equilibrium, whereas if the frequencies of the observed haplotypes are either larger or smaller, linkage disequilibrium is said to exist between the SNPs. Linkage disequilibrium indicates that not enough time has elapsed since the mutations generated the SNPs (or since the two variants came into contact through migration) for recombination (or new mutations at the SNP sites) to have shuffled them around enough to reach linkage equilibrium. Therefore, the greater the extent of disequilibrium among pairs of variable loci in a population, the younger the genetic variation in that population.

Thus, it was of great interest when it was shown that a polymorphic deletion on chromosome 12 showed hardly any linkage disequilibrium with nearby microsatellite alleles in Africa, whereas extensive disequilibrium existed outside Africa [64]. The fact that the same pattern is seen not only at other loci [65–69] but also on a genome-wide scale [70] is the strongest evidence to date that the gene pool in Africa is older than outside Africa. In principle, this pattern could be generated if the population size over time outside Africa would have been smaller than inside Africa. However, as for the pattern seen for the situation at Xq13.3, this assumes a substantial amount of gene flow among all populations outside of Africa in order to maintain the gene pool homogeneous enough for drift to allow the same lineages to disappear outside Africa. The possibility of extensive gene flow over the entire world, among a probably limited number of human ancestors (effective population size esti-

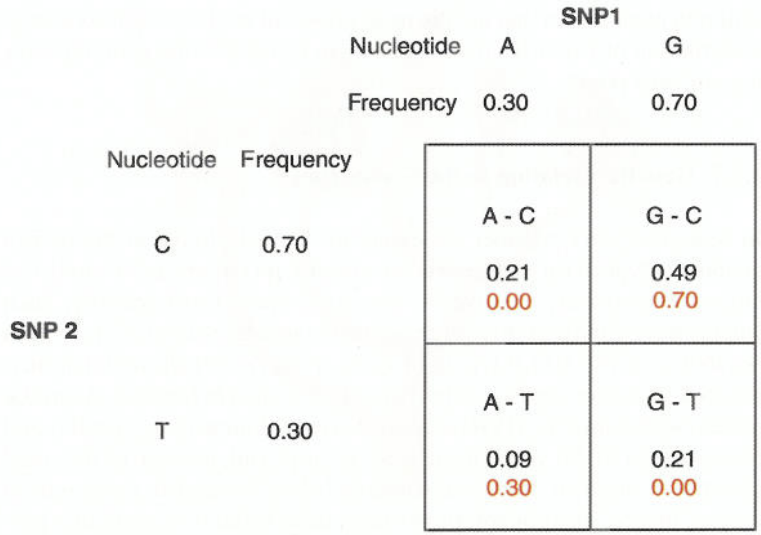


Fig. 9. Linkage disequilibrium illustration. A hypothetical example of two single nucleotide polymorphisms (SNPs) that are either in linkage equilibrium (random association of alleles according to allele frequencies – see text for details) or linkage disequilibrium (nonrandom association of alleles)- shown in red

mates indicate about 10 000 breeding individuals [71]), has been debated [16, 72] but appears implausible [16].

Therefore, on balance, the overall picture of our genetic history conforms much better with an African replacement scenario than with the regional continuity model. However, this does not preclude that some gene variants from archaic human populations penetrated the modern human gene pool and survived until today. Perhaps this may be the case particularly for genes encoding some features that were of selective advantage in a certain region of the world. However, it should also be noted that the variation at most nuclear genes is so old that it predates the separation of many archaic human populations such as Neandertals and modern humans [73]. This means that the nuclear gene variants present in these populations are not expected to have been very different from those seen in modern humans. Therefore, it is the data from loci such as mtDNA, the Y chromosome and linkage disequili-

brium in nuclear loci that are the most powerful evidence against a large contribution of genes from archaic human forms into the contemporary human gene pool.

4.2.7 Genetic Variation in the Great Apes

In order to clarify whether the extent of diversity found in the human genome is typical for a large primate species, it is necessary to study our closest evolutionary relatives – the great apes. Until recently, such studies were almost exclusively confined to protein variants [74, 75] and mtDNA [76, 78]. RFLP typing of great apes [78, 79] showed that their mtDNA variation was two to ten times higher than in humans. A similar pattern was found for HVRI sequences from chimpanzees, gorillas and orang-utans [76, 80, 81]. This may seem surprising because of the small population sizes of the great apes and their restricted geographical ranges. Indeed, electrophoretically detectable variants of particular proteins [74] as well as microsatellite variation [77] have both indicated that the great apes are less variable than humans. However, these studies suffer from an ascertainment bias [82] because the markers studied in the apes had originally been selected to be highly variable in humans. Therefore, it was unclear whether the great apes are more variable than humans on a genome-wide scale.

In order to put the human variation at Xq13.3 into a relevant evolutionary perspective, we sequenced about 10 000 bp at Xq13.3 in chimpanzees (*Pan troglodytes*), bonobos (*P. paniscus*), gorillas (*Gorilla gorilla*), and orang-utans (*Pongo pygmaeus*) [83, 84]. To first ascertain which species is most closely related to humans at Xq13.3, a phylogenetic tree [85] relating sequences from a human, chimpanzee, bonobo, gorilla and orang-utan individual was estimated (Fig. 10a and b) [84]. This tree shows that humans form a clade together with chimpanzees and bonobos to the exclusion of the gorilla. They differ from humans by approximately 0.9% of their nucleotides at Xq13.3, whereas gorillas and orang-utans differ from humans by 1.4 and 2.9%, respectively. Thus, the results obtained for Xq13.3 concur with the majority of studies [86] in placing chimpanzees and bonobos as the closest relatives of humans. To test if evolutionary rates at Xq13.3 are the same among humans and great apes, we performed an evolutionary 'clock-test' [85] which failed

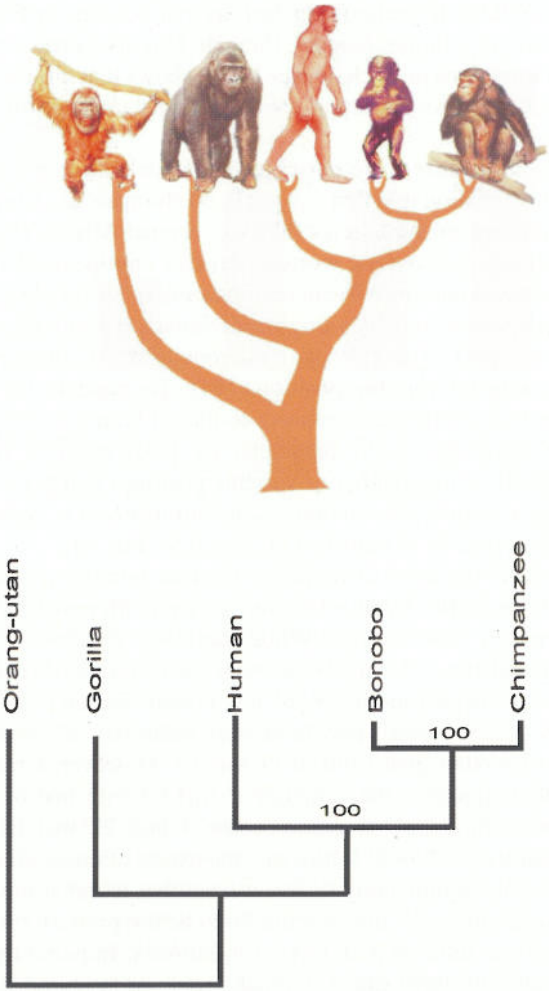


Fig. 10 a,b. Phylogenetic tree relating a human, chimpanzee, bonobo, gorilla, as well as an orang-utan Xq13.3 DNA sequence. Two versions of the tree are shown. One (**a**) is an illustration of the pattern of relatedness seen among the primates studied, the other (**b**) also shows original branch lengths based on the Xq13.3 data. The reliability value of each internal branch indicates how often the corresponding cluster was found among 1000 intermediate trees (in percentage). After Kaessmann *et al.Science*, 1999

to reject the clock hypothesis. In fact, as can be seen in Fig. 10b, all branches are of similar lengths (branch lengths correspond to the number of mutations on each lineage). This shows that differences in the extent DNA sequence variation cannot be attributed to differences in mutation rate.

Because chimpanzees are most closely related to humans at Xq13.3, intraspecific variation was first gauged in 30 chimpanzees from the three currently recognized subspecies [87], i.e. central African chimpanzees (*P. troglodytes troglodytes*), western African chimpanzees (*P. troglodytes verus*), and eastern African chimpanzees (*P. troglodytes schweinfurthii*) (represented only by a single individual), as well as five bonobos (*P. paniscus*) [84]. The chimpanzees contained 24 different Xq13.3 sequences with 84 variable positions [84]. To evaluate the extent of diversity within chimpanzees relative to that of human s, we calculated Watterson's estimate of the parameter $\grave{e}_w$ [88], which is a diversity measure based on the number of variable positions and is corrected for the number of individuals studied. $\grave{e}_w$ in chimpanzees is approximately three times as high as in humans (21.2 vs. 6.8) (Fig. 11).

In principle, the lower diversity in humans relative to chimpanzees may have been shaped by selection and not by a difference in population history between these species. While the DNA segment at Xq13.3 is noncoding and hence is highly improbable to be the direct target of selection, it is located in an area of low recombination [50]. Thus, the Xq13.3 sequences studied may have been indirectly affected by selection acting on other genes linked to Xq13.3. However, a test [89] for selection that compares the variation at Xq13.3 with that of two other nuclear noncoding loci on chromosome 1 and 22 that have higher recombination rates [90, 91] show that the extent of nucleotide diversity at Xq13.3 is not significantly different from that found at the other two loci. Consequently, selection is unlikely to be the primary cause for the low genetic variation seen at Xq13.3 in humans. Importantly, mtDNA has also indicated three times as much variation in chimpanzees as in humans [76, 92]. Furthermore, lower genetic diversity in humans than in chimpanzees has recently been found in a study surveying DNA sequence variation of an autosomal DNA segment (HOXB6) [93]. Therefore, the diversity of the chimpanzee genome seems to be generally greater than that of the human genome. Thus, the difference in diversity

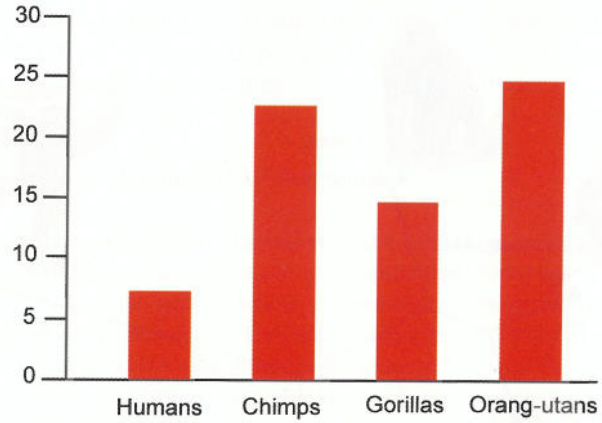


Fig. 11. DNA sequence diversity within humans and great apes. Values are based on the number of variable positions within each species taking the number of sequences determined into account (Watterson's diversity estimator, $\grave{e}_w$)

is the consequence of differences in population history between these species rather than selection acting on individual loci.

The next obvious question is whether humans or chimpanzees are exceptional among primates in having low and high amounts of DNA sequence diversity, respectively. To address this, Xq13.3 DNA was studied in 10 western lowland gorillas (*G. gorilla gorilla*) and one mountain gorilla (*G. gorilla beringei*), thus including two of three currently recognized subspecies [94], which include also eastern lowland gorillas (*G. gorilla graueri*). Xq13.3 was also studied in eight Bornean (*P. pygmaeus pygmaeus*) and six Sumatran (*P. pygmaeus abelii*) orang-utans, i.e. the two orang-utan subspecies [95]. Among the 14 orang-utan and 11 gorilla DNA sequences, 78 and 41 variable positions were identified, respectively [83]. $\grave{e}_w$ was estimated to 24.2 and 14.0, respectively, indicating that orang-utans carry approximately 3.5 times and gorillas twice as much Xq13.3 sequence diversity as humans (Fig. 11). This is illustrated by a phylogenetic tree (Fig. 12) which shows that human Xq13.3 sequences are characterized by short branches, whereas all great apes carry several long branches defining deep splits within the species. The extensive diversity found in the great

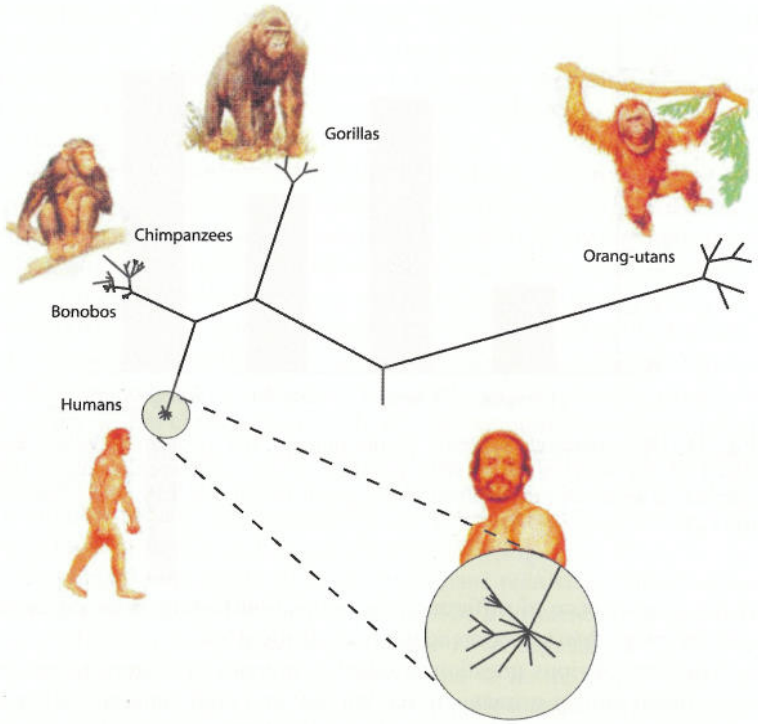


Fig. 12. Phylogenetic tree of 70 human, 30 chimpanzee, 5 bonobo, 11 gorilla and 14 orang-utan Xq13.3 DNA sequences. A gibbon sequence was used as an outgroup. The human cluster is also shown 7.5-fold enlarged (to the right) to illustrate the star-like pattern that is typical of an expansion due to population growth or positive selection affecting the DNA sequence. After Kaessmann *et al.Nature Genet.*, 2001

apes is also reflected in greater ages of their MRCAs which are 1 900 000 years in chimpanzees, 1 160 000 in gorillas, and 2 100 000 years in orang-utans, while the age of the human MRCA is about 500 000 years. Although the absolute ages depend on uncertain fossil calibration points and may thus have to be revised in the future, the relative time depths are not likely to change. Thus, the age of the variation at Xq13.3 in humans is only half to a quarter of that in the great apes in spite of the fact that the human population size currently exceeds that of the great apes by

several orders of magnitude. If applicable to most of the genome, this shows that from a genetical perspective, humans are a very recent species when compared with their closest living relatives.

4.2.8 An Expansion of Xq13.3 Sequences

The short time depth of the variation in the human genome relative to the great apes may be an indication of a population bottleneck (Fig. 8) in humans, i.e. a time period when the number of humans who have descendants today was small. This would cause the diversity of mtDNA, Xq13.3 and other DNA sequences to be reduced because the extinction of genetic lineages through drift would have been extensive during such a period. Because this scenario involves a population expansion subsequent to the period of small population size, it is interesting to test whether the pattern of variation in DNA sequences of contemporary humans indicate that population expansions have happened in the past.

When a population expands, few genetic variants are lost because many individuals get to have descendants. Consequently, many substitutions in such a situation will be found today in only some individuals (Fig. 13). By contrast, in a constant population where more lineages are lost fewer lineages in the past are the ancestors of present-day DNA sequences, past substitutions are found in a larger proportion of contemporary individuals (Fig. 13). One test that detects past expansions [96, 97] therefore compares the number of singletons with the total number of variation among DNA sequences. When it is applied to the Xq13.3 data, a significant excess (P=0.02) of singleton and low frequency substitutions over substitutions of intermediate frequency is seen among humans but not the great apes [83, 98, 99]. This indicates either that population growth or positive selection has affected the human DNA sequences but not those of the great apes [83].

This pattern is also seen in the form of a 'star-like' phylogeny of lineages within a species which reflects the fact that during an expansion, few lineages die out (fig. 13a) [92]. Thus, the star-like phylogeny for human Xq13.3 sequences (Fig. 12) may indicate population growth. An alternative explanation is a selective sweep affecting Xq13.3 sequences in humans but not in the great apes. In this case, an advantageous Xq13.3 variant would have 'swept' to fixation in humans, accu-

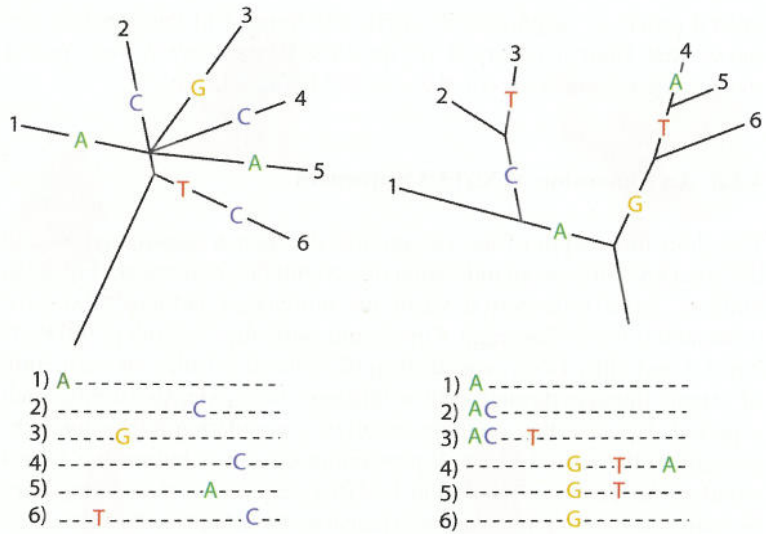


Fig. 13a,b. Illustration of phylogenies (**a**) and DNA sequence patterns (**b**) for a population that has grown in size (to the left) and a population that has been of constant size (to the right). Note that an excess of rare substitutions characterizes the growing population, whereas the stationary population has more shared variants (of intermediate frequency) as well as rarer variants

mulating substitutions in a 'star-like' fashion. In that case, one would expect the expansion signal to be confined to Xq13.3 sequences and not to be seen in other DNA sequences. In this regard, it is noteworthy that expansion signals have so far not been seen in several nuclear DNA sequences studied in humans. However, most of these DNA sequences [35–38] come from transcribed genes (e.g. -globin, *LPL* and PDHA1 genes) carrying alleles implicated, for example, in resistance to malaria (-globin) [44] or diseases [45–46]. Therefore, the distribution of nucleotide substitutions may be influenced not only by demographic phenomena but also by natural selection. By contrast, two loci that like Xq13.3 are noncoding tend to show signals of population expansions in humans [90, 91] as does a survey of variation of gene segments from genomic databases [100]. Thus, the majority of nuclear DNA variation may reflect a human population expansion. However, further work is necessary before this can be stated with confidence.

4.2.9 Dates of Expansions

The onset of population growth (or selection) reflected by Xq13.3 has been dated using two different approaches. One is based on coalescent theory [101] and considers the average pairwise sequence difference of all sequences in the sample, the number of variable positions and the sample size; the other is based on mismatch distributions [102], that is, the distribution of pairwise differences of all sequences in a sample. The latter approach is based on the assumption that if a population has grown in size, many lineages are expected to trace back to common ancestors who lived just before the expansion started. Thus, the most common number of differences among DNA sequences from a growing population corresponds to the number of mutations separating individuals that diverged from common ancestors who lived just before the onset of the expansion. The coalescent analysis gives a maximum likelihood value for the beginning of the putative human population expansion of approximately 190 000years while the mismatch analysis gives a date of 160 000 years. Using another approach, Wooding and Rogers (2000) [98] have dated the Xq13.3 expansion to 120 000 years ago.

Previous mtDNA studies have indicated that modern human populations have expanded approximately 40 000–50 000 years ago [15, 103, 104] and microsatellite data also indicated an onset of population growth around that time [105]. Thus, they seem to reflect an expansion that postdates Xq13.3. However, as all these age estimates have unknown variances that may be large, it is not possible to exclude that the mtDNA and Xq13.3 data indicate the same demographic expansion. It is noteworthy, however, that the mtDNA dates agree well with archaeological evidence indicating the appearance of, for example, art objects and more advanced and diverse tool industries [104, 106, 107]. It is also noteworthy that mtDNA indicates the beginning of expansions in chimpanzee and orang-utan populations for approximately the same time period (data not shown). Thus, it is possible that more general environmental factors such as climatic change have promoted an increase in population size of both humans and great apes about 50000 years ago. For instance, one hypothesis holds that population growth may have commenced after the release from a species-wide population bottleneck associated with a volcanic winter (caused by the eruption of Mount Toba on Sumatra) around 60 000 years ago [108].

For the Xq13.3 sequences it is interesting to note that a date of around 120 000–190 000 years ago coincides with most age estimates of the most recent common mtDNA and Y chromosome ancestors, which are approximately 100 000–200 000 years, respectively [15, 24, 62, 63]. It is also interesting to note that the first anatomically modern humans appeared about 100 000–130 000 years ago in the fossil record [17]. Thus, the ages of the beginning of population growth in humans as seen by Xq13.3 as well as most dates for the MRCAs of mtDNA and the Y chromosome coincide approximately with the emergence of anatomically modern humans as seen in the fossil record. Thus, it is possible that mtDNA and Y chromosomal lineages coalesce to their MRCAs around the time when a small ancestral population began to expand and that expansion is detected by Xq13.3. If that was true, this represents the beginning of the expansion of the African ancestors of modern humans who eventually went on to replace many or most of the other archaic human forms that existed at the time. Future studies of large numbers of nuclear genes will show whether this scenario applies to most of the human genome and thus to humans as a species. If that is true, then the challenge will be to identify the genetic factors that may have been the prerequisites for this expansion.

References

1. StonekingM&SoodyallH. Human evolution and the mitochondrial genome. *Curr Opin Genet Dev* 1996; 6: 731–6.
2. LehningerAL. Some aspects of energy coupling by mitochondria. *Adv Exp Med Biol* 1979; 111: 1–16.
3. LehningerAL, CarafoliE, RossiCS. Energy-linked ion movements in mitochondrial systems. *Adv Enzymol Relat Areas Mol Biol* 1967; 29: 259–320.
4. AndersonS, BankierAT, BarrellBG*etal.*Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290: 457–65.
5. OlivioPD. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 1983; 306: 400–2.
6. BrownWM, GeorgeM Jr, WilsonAC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 1979; 76: 1967–71.
7. GilesRE, BlancH, CannHM, WallaceDC. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 1980; 77: 6715–9.
8. CannRL, BrownWM, WilsonAC. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 1984; 106: 479–99.

9. CannRL, StonekingM, WilsonAC. Mitochondrial DNA and human evolution. *Nature* 1987; 325: 31–6.
10. WhittamTS, ClarkAG, StonekingM, CannRL, WilsonAC. Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. *Proc Natl Acad Sci USA* 1986; 83: 9611–5.
11. JohnsonMJ, WallaceDC, FerrisSD, RattazziMC, Cavalli-SforzaLL. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 1983; 19: 255–71.
12. DenaroM, BlancH, JohnsonMJ*et al.*Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci USA* 1981; 78: 5768–72.
13. VigilantL, StonekingM, HarpendingH, HawkesK, WilsonAC. African populations and the evolution of human mitochondrial DNA. *Science* 1991; 253: 1503–7.
14. KocherTD&WilsonAC. In: Osawa S, Honjo T, eds. *Evolution of Life: Fossils, Molecules and Culture*. Tokyo: Springer-Verlag, 1991; 391–413.
15. IngmanM, KaessmannH, PaaboS, GyllenstenU. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000; 408: 708–13.
16. JordeLB, BamshadM, RogersAR. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 1998; 20: 126–36.
17. StringerCB&AndrewsP. Genetic and fossil evidence for the origin of modern humans. *Science* 1988; 239: 1263–8.
18. StringerC. In: Sykes B, ed. *The Human Inheritance*. Oxford: Oxford University Press, 1999: 33–44.
19. AyalaFJ. The myth of eve: molecular biology and human origins. *Science* 1995; 270: 1930–6.
20. WolpoffM&CaspariR. *Race and Human Evolution.* New York: Simon & Schuster, 1997.
21. UnderhillPA, JinL, LinAA*et al.*Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; 7: 996–1005.
22. UnderhillPA, ShenP, LinAA*et al.*Y chromosome sequence variation and the history of human populations. *Nature Genet* 2000; 26: 358–61.
23. HammerMF, KarafetT, RasanayagamA*et al.*Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 1998; 15: 427–41.
24. StonekingM. In: Donnelly P, Tavare S. eds *Progress in Population Genetics and Human Evolution.* New York: Springer, 1997: 1–13.
25. ThomsonR, PritchardJK, ShenP, OefnerPJ, FeldmanMW. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc Natl Acad Sci USA* 2000; 97: 7360–5.

26. JaruzelskaJ, ZietkiewiczE, LabudaD. Is selection responsible for the low level of variation in the last intron of the ZFY locus? *Mol Biol Evol* 1999; 16: 1633–40.
27. ShenP, WangF, UnderhillPA*et al.*Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci USA* 2000; 97: 7354–9.
28. NachmanMW. Y chromosome variation of mice and men. *Mol Biol Evol* 1998; 15: 1744–50.
29. Ruiz LinaresA, NayarK, GoldsteinDB*etl.*Geographic clustering of human Y-chromosome haplotypes. *Ann Hum Genet* 1996; 60: 401–8.
30. SeielstadMT, MinchE, Cavalli-SforzaLL. Genetic evidence for a higher female migration rate in humans. *Nature Genet* 1998; 20: 278–80.
31. BowcockAM, Ruiz-LinaresA, TomfohrdeJ, MinchE, KiddJR, Cavalli-SforzaLL. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 1994; 368: 455–7.
32. JurkaJ&SmithT. A fundamental division in the Alu family of repeated sequences. *Proc Natl Acad Sci USA* 1988; 85: 4775–8.
33. StonekingM, FontiusJJ, CliffordSL*et al.*Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 1997; 7: 1061–71.
34. WatkinsW, RickerC, BamshadM*et al.*Patterns of ancestral human diversity: an analysis of alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* 2001; 68: 738–52.
35. HardingRM, FullertonSM, GriffithsRC*et al.*Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* 1997; 60: 772–89.
36. ClarkAG, WeissKM, NickersonDA*et al.*Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am J Hum Genet* 1998; 63: 595–612.
37. NickersonDA, TaylorSL, WeissKM*et al.*DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; 19: 233–40.
38. HarrisEE&HeyJ. X chromosome evidence for ancient human histories. *Proc Natl Acad Sci USA* 1999; 96: 3320–4.
39. RanaBK, Hewett-EmmettD, JinL*et al.*High polymorphism at the human melanocortin 1 receptor locus. *Genetics* 1999; 151: 1547–57.
40. TakahataN, LeeS, SattaY. Testing multiregionality of modern human origins. *Mol Biol Evol* 2001; 18: 172–83.
41. HedrickPW. Balancing selection and MHC. *Genetica* 1998; 104: 207–14.
42. KleinJ. Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum Immunol* 1987; 19: 155–62.

43. AyalaFJ, EscalanteA, O'HuiginC, KleinJ. Molecular genetics of speciation and human origins. *Proc Natl Acad Sci USA* 1994; 91: 6787–94.
44. WeatherallDJ&CleggJB 1981*The Thalassemia Syndromes.* Oxford: Blackwell Scientific.
45. BrunzellJ. In: Scriver C Beaudet A Sly W & Valle D, eds. *The Metabolic and Molecular Bases of Inherited Disease.* New York: McGraw-Hill, 1995: 913–33.
46. RobinsonBH, MacKayN, ChunK, LingM. Disorders of pyruvate carboxylase and the pyruvate dehydrogenase complex. *J Inherit Metab Dis* 1996; 19: 452–62.
47. WeatherallDJ. Thalassaemia and malaria, revisited. *Ann Trop Med Parasitol* 1997; 91: 885–90.
48. HughesAL&NeiM. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988; 335: 167–70.
49. KaessmannH, HeissigF, von HaeselerA, PääboS. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genet* 1999; 22: 78–81.
50. NagarajaR, MacMillanS, KereJ*et al.*X chromosome map at 75-kb STS resolution, revealing extremes of recombination and GC content. *Genome Res* 1997; 7: 210–22.
51. FuYX&LiWH. Coalescing into the 21st century. An overview and prospects of coalescent theory. *Theor Popul Biol* 1999; 56: 1–10.
52. RuhlenMA.*Guide to the World's Languages.* Kent: Edward Arnold, 1991.
53. JordeLB, BamshadMJ, WatkinsWS*et al.*Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* 1995; 57: 523–38.
54. AndrewsP. Evolution and environment in the Hominoidea. *Nature* 1992; 360: 641–6.
55. KumarS&HedgesSB. A molecular timescale for vertebrate evolution. *Nature* 1998; 392: 917–20.
56. TakahataN, SattaY, KleinJ. Divergence time and population size in the lineage leading to modern humans. *Theor Popul Biol* 1995; 48: 198–221.
57. SibleyCG&AhlquistJE. DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. *J Mol Evol* 1987; 26: 99–121.
58. AdachiJ&HasegawaM. Improved dating of the human/chimpanzee separation in the mitochondrial DNA tree: heterogeneity among amino acid sites. *J Mol Evol* 1995; 40: 622–8.
59. GriffithsRC&TavareS. Simulating probability distributions in the coalescent. *Theor Popul Biol* 1994; 46: 131–59.
60. GriffithsRC&TavareS. Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Math Biosci* 1995; 127: 77–98.

61. JonesS, MartinR, PilbeamD.*Human Evolution* Cambridge, UK: Cambridge University Press, 1992.
62. WeissG&von HaeselerA. Estimating the age of the common ancestor of men from the ZFY intron. *Science* 1996; 272: 1359–60; discussion 1361–2.
63. HammerMF, SpurdleAB, KarafetT*et al.*The geographic distribution of human Y chromosome variation. *Genetics* 1997; 145: 787–805.
64. TishkoffSA, DietzschE, SpeedW*et al.*Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 1996; 271: 1380–7.
65. TishkoffSA, PakstisAJ, StonekingM*et al.*Short tandem-repeat polymorphism/alu haplotype variation at the PLAT locus: implications for modern human origins. *Am J Hum Genet* 2000; 67: 901–25.
66. TishkoffSA, GoldmanA, CalafellF*et al.*A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. *Am J Hum Genet* 1998; 62: 1389–402.
67. TishkoffSA, RuanoG, KiddJR, KiddKK. Distribution and frequency of a polymorphic Alu insertion at the plasminogen activator locus in humans. *Hum Genet* 1996; 97: 759–64.
68. CalafellF, ShusterA, SpeedWC, KiddJR, KiddKK. Short tandem repeat polymorphism evolution in humans. *Eur J Hum Genet* 1998; 6: 38–49.
69. KiddJR, PakstisAJ, ZhaoH*et al.*Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am J Hum Genet* 2000; 66: 1882–99.
70. ReichDE, CargillM, BolkS*et al.*Linkage disequilibrium in the human genome. *Nature*, 2001; 411: 119–204.
71. HarpendingHC, BatzerMA, GurvenM, JordeLB, RogersAR, SherryST. Genetic traces of ancient demography. *Proc Natl Acad Sci USA* 1998; 95: 1961–7.
72. TempletonAR. Human races: a genetic and evolutionary perspective. *Am Anthropol* 1999; 100: 632–50.
73. PääboS. Human evolution. *Trends Cell Biol* 1999; 9: M13–6.
74. KingMC&WilsonAC. Evolution at two levels in humans and chimpanzees. *Science* 1975; 188: 107–16.
75. SarichVM&WilsonAC. Rates of albumin evolution in primates. *Proc Natl Acad Sci USA* 1967; 58: 142–8.
76. MorinPA, MooreJJ, ChakrabortyR, JinL, GoodallJ, WoodruffDS. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* 1994; 265: 1193–201.

77. WiseCA, SramlM, RubinszteinDC, EastealS. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Mol Biol Evol* 1997; 14: 707–16.
78. FerrisSD, BrownWM, DavidsonWS, WilsonAC. Extensive polymorphism in the mitochondrial DNA of apes. *Proc Natl Acad Sci USA* 1981; 78: 6319–23.
79. ZhiL, KareshWB, JanczewskiDN*et al.*Genomic differentiation among natural populations of orang-utan (*Pongo pygmaeus*). *Curr Biol* 1996; 6: 1326–36.
80. GarnerKJ&RyderOA. Mitochondrial DNA diversity in gorillas. *Mol Phylogenet Evol* 1996; 6: 39–48.
81. WarrenKS, VerschoorEJ, LangenhuijzenS.*et al.*Speciation and intra-subspecific variation of Bornean orang-utans, Pongo pygmaeus pygmaeus. *Mol Biol Evol*, 2001; 18: 472–480.
82. EllegrenH, PrimmerCR, SheldonBC. Microsatellite 'evolution': directionality or bias? *Nat Genet* 1995; 11: 360–2.
83. KaessmannH, WiebeV, WeissG, PaaboS. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nature Genet* 2001; 27: 155–6.
84. KaessmannH, WiebeV, PääboS. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* 1999; 286: 1159–62.
85. StrimmerKS&von HaeselerA. Quartet Puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* 1996; 13: 964–9.
86. RuvoloM. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol Biol Evol* 1997; 14: 248–65.
87. HillWO 1969*The Chimpanzee*. Basel: Karger.
88. WattersonGA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975; 7: 256–76.
89. HudsonRR, KreitmanM, AguadeM. A test of neutral molecular evolution based on nucleotide data. *Genetics* 1987; 116: 153–9.
90. YuN, ZhaoZ, FuYX*et al.*Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol Biol Evol* 2001; 18: 214–22.
91. ZhaoZ, JinL, FuYX*et al.*Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. *Proc Natl Acad Sci USA* 2000; 97: 11354–8.
92. von HaeselerA, SajantilaA, PääboS. The genetical archaeology of the human genome. *Nature Genet* 1996; 14: 135–40.
93. DeinardA&KiddK. Evolution of a HOXB6 intergenic region within the great apes and humans. *J Hum Evol* 1999; 36: 687–703.
94. GrovesCP. Population systematics of the gorilla. *J Zool London* 1970; 161: 287–300.

95. JonesML 1969*The Geographic Races of the Orang-Utan*. Basel: Karger.
96. FuYX&LiWH. Statistical tests of neutrality of mutations. *Genetics* 1993; 133: 693–709.
97. Tajim_aF. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; 123: 585–95.
98. WoodingS&RogersA. A pleistocene population X-plosion? *Hum Biol* 2000; 72: 693–5.
99. WallJD&PrzeworskiM. When did the human population size start increasing? *Genetics* 2000; 155: 1865–74.
100. YangZ, WongGK, EberleMA*et al.*Sampling SNPs. *Nat Genet* 2000; 26: 13–4.
101. WeissG&von HaeselerA. Inference of population history using a likelihood approach. *Genetics* 1998; 149: 1539–46.
102. RogersAR&HarpendingH. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 1992; 9: 552–69.
103. HarpendingHC, SherryST, RogersAR, StonekingM. The genetic structure of Ancient human populations. *Current Anthropol* 1993; 34: 483–96.
104. StonekingM. Mitochondrial DNA and human evolution. *J Bioenerg Biomemb* 1994; 3: 251–9.
105. ZhivotovskyLA, BennettL, BowcockAM, FeldmanMW. Human population expansion and microsatellite variation. *Mol Biol Evol* 2000; 17: 757–67.
106. MellarsPA. Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philos Trans R Soc Lond B Biol Sci* 1992; 337: 225–34.
107. KleinRG 1989*The Human Career: Human Biological and Cultural Origins*. Chicago: University of Chicago Press.
108. AmbroseSH. Late pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans. *J Hum Evol* 1998; 34: 623–51.
109. Cavalli-SforzaLL, MenozziP, PiazzaA. Demic expansions and human evolution. *Science* 1993; 259: 639–46.

5 Genforschung und Embryologie im 20. Jahrhundert

Christiane Nüsslein-Volhard