PhD thesis

The origins of southern and western Eurasian populations: an mtDNA study

Toomas Kivisild

Department of Evolutionary Biology, Institute of Molecular and Cell Biology, Tartu University, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Genetics) on March 7th, 2000 by the Council of the Institute of Molecular and Cell Biology, Tartu University.

Opponent: Dr. Peter Forster (McDonald Institute for Archaeological Research, Cambridge University, UK)

Commencement: May 9th, 2000

The publication of this dissertation is granted by the University of Tartu

Contents

List of original publications

The current dissertation is based on the following publications referred to in the text by their Roman numbers:

I. **Kivisild, T**., Bamshad, M., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., Dixon, M.E., Papiha, S.S., Mastana, S.S., Mir, M.R., Ferak, V., Villems, R. (1999). Deep common ancestry of Indian and western Eurasian mtDNA lineages. *Current Biology* 9: 1331-1334.

II. Bamshad, M, **Kivisild, T.**, Watkins, W. S., Dixon, M. E., Ricker, C. E., Rao, B. B., Naidu, J. M., Prasad, V. R., Reddy, P. G., Rasanayagam, A., Papiha, S. S., Villems, R. Batzer, M. A., Jorde, L. B. Genetic evidence on the origins of Indian caste populations. (Manuscript).

III. **Kivisild, T.**, Kaldma, K., Metspalu, M., Parik, J., Papiha, S.S., Villems, R. (1999). The Place of the Indian mtDNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World. in *Genomic Diversity*. (Kluwer Academic/Plenum Publishers). 135-152.

IV. Metspalu, E., **Kivisild, T**., Kaldma, K., Parik, J., Reidla, M., Tambets, K., Villems, R. (1999). The Trans-Caucasus and the expansion of the Caucasoid-specific human mtDNA lineages. in *Genomic Diversity*. (Kluwer Academic/Plenum Publishers). 121-134.

Abbreviations

bp	base pair
COII	cytochrome oxydase subunit II
CRS	Cambridge Reference Sequence (Anderson *et al*. 1981)
D-loop	displacement loop (=control region) of mtDNA
HVS-I	the first hypervariable segment of the control region
LGM	the Last Glacial Maximum
LHON	Leber's hereditary optic neuropathy
MP	maximum parsimony
MRCA	the most recent common ancestor
mtDNA	mitochondrial DNA
np	nucleotide position
PNG	Papua New Guinea
RFLP	Restriction Fragment Length Polymorphism
$tRNA^{Lys}$	lysyl transfer RNA

Definition of basic terms used in the thesis

haplotype	a sequence type that comprises all identical sequences
haplogroup	a group of haplotypes that share a common ancestor defined by an array of synapomorphic substitutions
lineage	any array of characters/mutations shared by more than one haplotype
star-like tree	a set of sequences is said to have a pattern of star-like phylogeny if most (ideally all of them) coalesce to one and the same haplotype (that has not necessarily been observed in the sample)
founder haplotype	haplotype to which all sequences under concern coalesce to
time depth = age	time calculated to the MRCA of a group of sequences
expansion time	time calculated to the founder that displays star-like phylogeny
Caucasoids = western	Eurasians, including populations from Near East, West Asia, Europe and northern Africa
Mongoloids = eastern	Eurasians, including populations east of India and Central Asia.

Introduction

A general intellectual curiosity, inherent for our species, has made humans now and then in quest of answers related to our origins. While there is little doubt that the main driving force behind the search for human genetic variation is associated with an attempt to understand the genetics of diseases, it has also a much wider aspect: this variation reflects the past of mankind, the origin and spread of anatomically modern humans, their demographic history. It is increasingly more evident that knowledge about these processes is equally indispensable for a proper understanding of the genetics of complex diseases.

More than 80 years of research of genetic variation have revealed that a substantial degree of diversity lies both within and between extant human populations. Despite the huge amount of data gathered during this relatively long time span of classical ("pre-DNA") genetics, recently reviewed in a fundamental way by Cavalli-Sforza *et al*. (1994), the "classical era" raised many basic problems, leaving them largely unsolved. Increasingly more resourceful "DNA era", rapidly expanding during the last 20 years, took up the same list of problems and is formulating new ones (for recent reviews see Foley 1998; Jorde *et al*. 1998; Paabo 1999; Relethford 1998; Sherry *et al*. 1998)

Largely irrespective of what kind of general questions related to demographic history are being asked, the present-day genetics investigates variation in three different systems: autosomal chromosomes, paternally inherited Y chromosome and maternally inherited mitochondrial DNA. While the size of autosomal genes exceeds manifold that of Y chromosome and vastly the content of mtDNA, the latter two have powerful advantages, making them the favorite tools in the hands of population geneticists. Besides uniparental mode of inheritance, the list of advantages includes also the lack of recombination. Taken together, they allow, at least in theory, to reconstruct genetic lineages back to the most recent common ancestors - our Y-chromosomal Adam and mitochondrial Eve. And in combination with information about the variation in autosomal genes, a promising synthesis is hopefully possible in the future. The first part of the literature overview in the current thesis is therefore dedicated to the properties of mtDNA as a tool in population genetics.

My thesis is about mtDNA variation in places remote from Estonia: India and Trans-Caucasus. In order to understand mtDNA variation and its origin in European populations additional background knowledge of those living outside Europe is needed. Due to its particular properties mtDNA allows one to construct individual genealogies and connect them to a global phylogenetic tree of all humans. Basal branches of such mtDNA tree have been shown to be highly continent specific - making mitochondria a more valuable tool compared to most other markers where their overall variation is hidden mainly within populations. The study of the spread and variability of mtDNA clusters enables one to address questions related to the early population movements of our species. While the recent out-of-Africa scenario of modern human origins is rather well supported, the exact time, number, extent and directions of these movements still need to be dealt with.

Through the collective effort of many labs, mitochondrial DNA clusters (haplogroups) have been defined and studied in different continental regions. The central focus of the literature overview of the current thesis will be therefore the general skeleton tree connecting and characterizing mtDNA haplogroups worldwide. The reason why this particular aspect may and should be considered as central is that without understanding its topology, much of the empirical results would merely be a phenomenological description of the populations under study. Regrettably, numerous recent publications still suffer from a clearly phylogenetic approach,

making otherwise interesting data much less useful. Furthermore, an understanding of the skeleton of the mtDNA phylogeny is very much needed to plan experiments: to choose positions which are possibly informative in tree building and to analyze them cladistically, not as a pool of unrelated point mutations.

Therefore, the main part of the literature overview is a critical analysis and also synthesis of the data published by others on this issue. It will serve as a phylogenetic background and repository of questions on which the extent of variation in populations studied by us is compared to and to which answers are searched for.

Literature Overview

The properties of mtDNA

Several properties of mtDNA need to be considered when using it in phylogenetic studies.

1. The fast mutation rate of mtDNA has been taken as one of the main advantages for reconstructing human phylogenies (Wilson *et al*. 1985). MtDNA diverges at the rate of 2-4% per million years (Brown *et al*. 1979; Cann *et al*. 1987) which is on the average 10 to 100 fold faster than the rate in the nuclear genome. High evolutionary rate enables one to get more refined information of recent events in the phylogeny of the species, e.g. of the evolution of anatomically modern humans during the past 150,000 years (Stoneking 1994). On the other hand, fast rate creates its own problems, like multiple hits on the same sites that disturb the overall structure of the phylogenetic tree and are expressed in that, for example, millions of most parsimonious trees can be drawn from a data set of approximately 100 sequences (Cann *et al*. 1987; Templeton 1992; Vigilant *et al*. 1991).

2. Maternal inheritance of mtDNA in animals (Giles *et al*. 1980) together with the lack of recombination (Merriwether *et al*. 1991; Olivo *et al*. 1983) makes the effective population size (N_e) of mitochondrial genome smaller compared to that of any autosomal nuclear locus. Smaller N_e increases the sensitivity of mtDNA diversity to fluctuations of population size and enables to detect bottlenecks that are not necessarily expressed in nuclear markers with three- (X chromosome) or four-fold higher N_e. While paternal leakage of mtDNA in inter-species crosses of mice has been detected (Gyllensten *et al*. 1991), in intra-species crosses paternal mtDNA is selectively eliminated (Kaneda *et al*. 1995). It is likely that one of the signals for the destruction of paternal mtDNAs is the ubiquination of the mid-piece of sperm (Hopkin 1999). Moreover, the number of mtDNA molecules in an oocyte exceeds hundreds of times that of a sperm (Michaels *et al*. 1982), making an effective penetrance of paternal mtDNA very unlikely. Although, some studies have questioned the clonal inheritance mode on either phylogenetic (Eyre-Walker *et al*. 1999; Hagelberg *et al*. 1999) or statistical (Awadalla *et al*. 1999) grounds, most of the currently available evidence from both population genetics (see Macaulay *et al*. 1999a) and pedigree studies (e.g. Bendall *et al*. 1996; Jazin *et al*. 1998; Parsons *et al*. 1997; Soodyall *et al*. 1997) is consistent with strictly maternal inheritance of mitochondria.

3. Neutrality of the markers being used is a necessary precondition for many methods in phylogenetic analysis. Several tests of neutrality have been performed, measuring the difference between the fixation of non-synonymous and synonymous mutations in mtDNA at intra- and interspecies level (e.g. Excoffier 1990; Graven *et al*. 1995; Nachman 1998; Nielsen and Weinreich 1999; Whittam *et al*. 1986). Torroni *et al*. (1994) showed that mtDNA mutations are unlikely to be related to the Tibetans' adaptation to living in high altitudes. Those non-synonymous substitutions that are known to be under selection are normally excluded from phylogenetic analyses. Several substitutions have been found in association with the inheritance of diseases (for review see Wallace *et al*. 1999, and references therein).

Mildly deleterious substitutions under selection should get eliminated in long run. However, when a mildly deleterious substitution occurs in a small expanding population, it can, on the contrary, be maintained (Excoffier 1990; Merriwether *et al*. 1991). A good example for the latter case could be haplogroup J in Europe. Several mutations specific for haplogroup J, which has probably expanded in Europe together with the Neolithic spread of farmers (Richards *et al*. 1996), have been suggested to play a background role in LHON expression (Torroni *et al*. 1997). Otherwise, the frequency of neutral substitutions when linked with a substitution under selection can be shifted up or down by hitch-hiking (Smith and Haigh 1974), as far as mtDNA is a single

locus. Yet, being as well the subject of the cumulative chance fixation of mildly deleterious mutations (because of the lack of recombination and clonal inheritance), mitochondria are likely to be led to Muller's ratchet (Lynch 1996).

The properties of mtDNA mentioned above, with the given restrictions, allow one to reconstruct genealogies of individuals through maternal descent. This option for 'trees of individuals' makes mitochondria different from markers whose variation is expressed only in allele frequencies and/or evolve too slowly for revealing genealogies through their mutational pattern.

Phylogenetic tree building methods
Different tree-building methods exist for the study of phylogenetic relatedness between DNA samples (e.g. (Li 1997). Most generally the algorithms can be divided into groups: those relying either on (i) distance, e.g. neighbor joining (NJ), or on (ii) character state, e.g. maximum parsimony (MP) and maximum likelihood (ML), analyses.

Neighbor joining method
NJ tree construction (Saitou and Nei 1987) starts with a starlike tree and a distance matrix inferred from pairwise comparison of all sequences. Sequentially, the search for the closest neighbors in the matrix defines additional interior branches in the tree, minimizing its total length.

Parsimony method
MP methods (Fitch 1977; Swofford 1993) search for tree(s) that require the smallest number of substitutions. Tree construction employs only informative sites - substitutions that favor certain branching order; and from the set of all possible or heuristically chosen sample of trees the one(s) with the smallest number of substitutions (character state changes) are chosen.

Maximum likelihood method
ML methods (Felsenstein 1988) assign likelihood values to trees or subsets of a tree to choose the best tree by the value. The likelihood estimates are derived from a probabilistic model that is specified for character state changes. Such models take into account substitution rate from one character state to another. The rates can be taken as uniform for all substitution types (Jukes and Cantor's 1-parameter model), or different values can be given for transitions and transversions (Kimura's 2-parameter model). Rates can be further refined by different substitution types and GC content. Unlike MP method ML analysis makes use of all sites available in the sequences.

Phylogenetic networks
Multiple hits are a common problem in all phylogenetic analyses as they cause incompatible character states. The higher the number of pairwise incompatible (homoplasious) sites the higher is the number of trees with equal length that can be drawn from the data set. One particular tree from such a forest of MP trees alone, thus, can be misleading as far as character conflicts are resolved arbitrarily. At the cost of lower resolution the given homoplasies can be expressed by alternative pathways in a form of networks. Fast algorithms of network construction have been developed, relying either on sequential split decomposition of each informative character in the sequence matrix (Bandelt 1994; Bandelt *et al*. 1995), ending up with a (reduced) median network containing (almost) all MP trees, or on sequential introduction of inner branches between components of shortly connected nodes (Bandelt *et al*. 1999).

Rooting of phylogenetic trees
Any tree construction produces unrooted trees which do not define directions of evolutionary

relationships nor make any assumptions about the common ancestor of the group. The tree can be rooted if additional knowledge is supplied from an outgroup. The latter is usually chosen from another group that is externally known to have separated from the common phylogenetic lineage before the existence of the most recent common ancestor of the group under study. If such information is not available, an alternative, yet not so powerful, way of rooting a tree is to specify a node in the tree from which the distance to all terminal nodes is minimal as the mid-point root. Insertion of parts of mtDNA into the more conservative nuclear genome can serve as a special case of outgroup - a recent insertion of a D-loop segment into chromosome 11 has been proved very useful for rooting human mtDNA tree (Zischler *et al*. 1995).

Mutation rate calibration

One of the crucial points in mtDNA based population genetics is the calibration of the molecular clock. Since timing is decisive for the interpretation of the demographic history of populations, a reliable relation between sequence diversity and the time scale is needed. Several approaches have been taken, all based on assumptions that can be quantitatively checked, like (i) constant rate in different lineages, (ii) neutrality of the mutations being used. Two of them rely on 'outside' information, like fossils or historical records, which have to be taken as they are.

1. One way of clock calibration is to consider variation in populations from specific geographical regions with a well-known time of colonization. The extent of differentiation within clusters specific to New Guinea, Australia and the New World have been used to estimate the mean rate of mtDNA divergence within humans. It has been found that the divergence rate (twice the substitution rate), averaged over the whole mtDNA molecule, lies between 2-4% (Cann *et al*. 1987; Torroni *et al*. 1994c; Wilson *et al*. 1985) and for transitions in a HVS-I segment (16,090-16365) is about 36% per million years (Forster *et al*. 1996).
2. An out-group or interspecies calibration method considers differentiation within one species compared to the distance from the MRCA with another related species. The date for the split of these species is taken from paleontological evidence and related with the sequence diversity. Horai (Horai *et al*. 1995), for example, considered the divergence time of African apes from fossil record as being 13 million years and estimated from that figure that the human/chimpanzee split occurred 4.9 million years ago. This calibration gave the substitution rate at synonymous sites and in the D-loop region as 3.89 and 7.00 x 10^{-8}/site/year, respectively. From these estimates the age of the last common ancestor of the human mtDNAs was inferred as 143,000 ± 18,000 years. Similarly, the distance between humans and chimpanzees was used to calibrate the rate of most widely used 360 bp segment of HVS-I (Ward *et al*. 1991), yielding the divergence rate of 33% per million years. For the whole control region, with a total of 751 nps, 23% per million years of divergence has been estimated (Stoneking *et al*. 1992).
3. Pedigree calibrations try to estimate the mutation rate directly from samples of known genealogy. Fast rates (260% divergence per million years) obtained from the first few pedigree studies (Howell *et al*. 1996; Parsons *et al*. 1997) were taken as being in conflict with those calculated phylogenetically. It was argued (Paabo 1996) that the pedigree rate could be more appropriate to use in studies focused on events in a timescale of hundreds or thousands of years. Now, when more data have became available (e.g. Bendall *et al*. 1996; Jazin *et al*. 1998; Soodyall *et al*. 1997) the pedigree estimates have become close to the conventional ones and an improved understanding of the differences between the mutation and fixation rates, and differences in rates between sites in control region, suggests that there is no need to revise the mutation rate estimates (Macaulay *et al*. 1997).

Homoplasies and multiple hits

In 1983 Aquadro and Greenberg (Aquadro and Greenberg 1983) found that the sequence diversity in the D-loop region of mtDNA is several fold higher than in the coding parts and that

there exists a 32-fold bias of transitions over transversions. Wherefore transitions, in a relatively long run, tend to get 'saturated' and are about to undergo multiple hits, which means that one site may have gone through many substitutions and yet be at the same state. Multiple hit corrections therefore are to be made whenever attempting to draw phylogenetically reasonable conclusions, especially those concerning time estimates. One way of solving this problem for transitions is to compare the observed difference between two species with the one expected from transversional distance under a model of known ratio of transitions and transversions (Vigilant *et al*. 1991). Another way to cope with the high levels of homoplasy in HVS-I haplotypes, caused by either 'visible' parallelisms or 'hidden' multiple hits, is a combined application of both RFLP and HVS-I data (Graven *et al*. 1995; Helgason *et al*. 2000; Kolman *et al*. 1996; Macaulay *et al*. 1999b; Schurr *et al*. 1999; Starikovskaya *et al*. 1998; Torroni *et al*. 1996).

Rate variation

Besides 20-30 fold transitional bias it has been noted that the rate variation between sites from the control region is also significantly high (Excoffier and Yang 1999; Hasegawa *et al*. 1993; Ohno *et al*. 1991; Wakeley 1993). Transitions at sites like 16093, 16129, 16209, 16311 and 16362 from HVS-I and 00146, 00150, 00152 and 00195 from HVS-II occur in many different lineages and these sites can be considered as mutational hotspots. The heterogeneity is higher in HVS-II (Aris-Brosou and Excoffier 1996) that contains a handful of extremely fast evolving sites and a long list of those that rarely undergo any mutation. Rate variation can be taken into account in phylogenetic homoplasy solving by giving different weights to sites according to known rate variation (Helgason *et al*. 2000; Richards *et al*. 1998).

Mismatch distributions and demographic expansions

Pairwise comparison of sequence types gives a frequency distribution of distances ("mismatch distributions", Harpending *et al*. 1993) between all possible pairs of sequences in the sample. When a population goes through a rapid expansion then most of the variation within that population is preserved and new mutations get fixed with higher probability. Given the random nature of mutation cumulation, the frequency distribution of pairwise distances is thus expected to be unimodal and fit the Poisson process. Contrarily, in populations with constant size the distribution is expected to be a multi-modal or "bumpy" one. Thus, in the ideal case there is a downright correlation between the population demographic history and the mismatch distributions. In reality, very often, the population history can be a sum of several fusions, splits and multiple expansions or bottle-necks. Hence, a simplistic analysis of mismatch distributions in a whole sample can be misleading. For example, African populations were first interpreted (Graven *et al*. 1995; Watson *et al*. 1996) as stationary because of their bumpy mismatch distributions. Yet, when phylogenetically properly dissected (Watson *et al*. 1997) several African-specific lineages clusters were detected that exhibited unimodal mismatch frequency pattern reflecting possible ancient expansions. These separate clusters had expanded and were further fused together before the emergence of modern ethnic groups.

There is a contributive correlation between the coalescence age estimate of a tree and the effective population size for a given population since the coalescence. By theory, the smaller the effective population size is the sooner a relevant tree is expected to coalesce in the past. For example, human mtDNA diversity compatible with the African "Eve's" age of 200,000 years has been argued to infer that the minimal effective number of females in the late Middle Pleistocene could have been in the order of only one to ten thousands (Harpending *et al*. 1993). Further, the peaks of mismatch distributions suggest that a major population expansion in human history took place long after that time, around 60,000 years ago (Harpending *et al*. 1993;

Mountain *et al*. 1995; Rogers and Harpending 1992; Rogers and Jorde 1995; Sherry *et al*. 1994).

Classification of mtDNA variation world wide

As early as in 1980, a study of 21 humans from diverse racial and geographic origin revealed that mitochondrial DNA restriction enzyme cleavage pattern can be used in human population genetics to trace the history of our species (Brown 1980). On the basis of observed diversity in a worldwide sample a surprisingly recent 180,000 year old coalescence age estimate (taking 1 % per million years as an estimated rate of base substitution) was obtained for the global mtDNA variation. Shortly after the publication of the first complete sequence of human mtDNA (Anderson *et al*. 1981), Denaro and colleagues (Denaro *et al*. 1981) found that a *Hpa*I site presence (3594 C/T) separates most of Africans from Caucasians and Orientals. Interestingly, this key polymorphism in the human mtDNA phylogeny stood in a derived state among Africans in respect to other primate species. Two years later it became in a similar way apparent (Blanc *et al*. 1983) that an ancestral state of a *Hinc*II site (at 12406) is frequent in some Asian populations, suggesting that one specific Oriental lineage could be ancestral to all human mitochondrial descent. The Oriental origins of human mtDNA variation were further concluded in a study of Tharus from Nepal (Brega *et al*. 1986) and in an analysis of several Chinese national groups (Yu *et al*. 1988). High mtDNA diversity was found to be present among Japanese populations as well (Horai *et al*. 1984; Horai and Matsunaga 1986). Besides Asian populations and Africans, Cann et al. (Cann *et al*. 1982) reported that native Australians display equally high diversity estimates. Johnson and colleagues (Johnson *et al*. 1983) observed the highest diversity in Africans but could not exclude the interpretation that it could be due to differences in evolutionary rates among different mtDNA lineages. Another explanation for higher African mtDNA diversity - the existence of differential selective mechanisms in mitochondria - was given by Excoffier and Langaney (Excoffier and Langaney 1989). They put forward an intriguing hypothesis for human origins suggesting that Caucasoid populations were the closest to the ancestral population from which all other continental groups diverged. Higher genetic diversity in Africa is compatible also with the reasoning that effective population size has been higher in Africa during recent human evolution (Relethford and Jorde 1999).

The debate around modern human mtDNA origins centered around two opposing models when Alan Wilson's group published in 1987 the results of RFLP mapping of 147 mtDNAs taken worldwide (Cann *et al*. 1987) and concluded that the origins of the human mtDNA tree lay in Africa. It was at about the same time that Chris Stringer suggested a recent replacement model on archeological grounds (Stringer and Andrews 1988).

African root of the human mtDNA tree and the "mitochondrial Eve"

By assuming that all human mtDNAs stem (they do so by theory anyway, provided that recombination does not exist) from one woman Cann *et al.* (1987) calculated her to have lived around 200,000 years ago. Their argumentation for the single African origin was based on two findings: (i) the midpoint-rooting split the MP tree relating all human mtDNA sequence types into two basic branches, one of which contained exclusively Africans, the other one Africans and all the non-African mtDNAs; (ii) the higher diversity, represented by deepest branches, characterized African lineages. Cann *et al.* paper was criticized for some of its weaknesses, namely that: (i) it used the RFLP approach instead of direct sequencing; (ii) a small sample of African Americans was taken to represent native Africans; (iii) it used an inferior rooting method, (iv) it lacked statistical justification for African origins of human mtDNA and (v) it provided an inadequate calibration of the rate of mtDNA evolution (Darlu and Tassy 1987; Excoffier and Langaney 1989; Saitou and Omoto 1987; Wolpoff 1989). Furthermore, Maddison (Maddison 1991) found 10,000 trees with geographically mixed basal clades more parsimonious

by five steps than the MP tree given by Cann *et al*. (1987).
Four years later a study from the same group (Vigilant *et al*. 1991) tried to consider previous criticism and provided the results of a study of 189 sequences, including 121 native Africans, of both hypervariable regions of mtDNA. They also employed now an outgroup rooting making use of a published by then (Foran *et al*. 1988) chimpanzee sequence and different statistical tests to evaluate the geographic origin of the mtDNA ancestor. The results confirmed the previous ones in both that the outgroup split the tree into exclusively African and African plus non-African branches and that the time to MRCA, dubbed "Eve", stayed recent - slightly more than 200,000 years old. Yet, again, from the same data set more parsimonious trees were found (Templeton 1992) with mixed basal clades and it was to be concluded (Hedges *et al*. 1992) that the available sequence data were insufficient to solve statistically the geographical origin of the human mtDNA.

Multiregional model of human evolution
Nearly at the same time when Chris Stringer and Alan Wilson molded the theory of recent African origins, Milford Wolpoff, Alan Thorne and Wu Xinzhi put forward another explanation for the pattern of human evolution (Thorne and Wolpoff 1992). The multiregional model of evolution traces all human populations back to a *Homo erectus* ancestor that supposedly left Africa at least 1 million years ago. Further, despite an everlasting network of intercontinental gene flow, each continental group of populations developed to *sapiens* in its own way, retaining in the process distinguishable features (and genes) of the particular local pre-*sapiens* populations.

By the use of different analysis techniques and extended mtDNA data sets the theory of recent African origins of human mtDNA phylogeny has found overwhelming support (e.g. Chen *et al*. 1995; Horai *et al*. 1995; Penny *et al*. 1995; Watson *et al*. 1997). Moreover, most nuclear data sets also back up the theory (Armour *et al*. 1996; Bowcock *et al*. 1991; Jin *et al*. 1999; Jorde *et al*. 2000; Kaessmann *et al*. 1999; Nei and Takezaki 1996; Tishkoff *et al*. 1996) as do the paleoanthropological findings (Foley 1998). The global data base of mitochondrial DNA variation allows now to face in more detail the questions related to the MRCA, her age and the time and route(s) of expansions and dispersals in and out of Africa.

MtDNA variation in Africa
A *Hpa*I site at np 3592 (3594T/C) separates most Africans from all non-African populations, defining thereby the only African-specific haplogroup L. This one letter has undesirably to cope with the highest African mtDNA variation, compared to, for example, 9 letters reserved for relatively low variation observed in Europe. Three major clusters: L1a, L1b and L2 are spread widely all over Africa and are specified as those which share the presence of the *Hpa*I site at 3592 (Fig. 1). Those African lineages that lack the *Hpa*I polymorphism have been designated as L3 - the clade which also incorporates all non-African clusters (Chen *et al*. 1995). These four major African clusters reveal traces of ancient demographic expansions and cover altogether 87% of the African mtDNAs (Watson *et al*. 1997); the leftover consists mainly of individual isolated lineages from super-cluster L1.

The presence of distinguishable differences between African populations was first revealed by RFLP variation when Scozzari and colleagues compared Senegalese populations with Bantus and Bushmen (Scozzari *et al*. 1988). The spread of major African-specific haplogroups and their inner sub-structure varies geographically and also by tribes. All eastern and western Pygmies and the majority of all other sub-Saharan Africans harbor the *Hpa*I site at 3592 (Chen *et al*. 1995; Johnson *et al*. 1983; Scozzari *et al*. 1994; Scozzari *et al*. 1988; Soodyall and Jenkins

1992a; Soodyall and Jenkins 1993). Disjoint location of distinctive clusters specific for the Mbuti (in L2 -13065 *Dde*I and +11776 *Rsa*I) and the Biaka (in L1c +10319 *Alu*I) Pygmies suggests that the "Pygmy" phenotype has originated more than once (Bandelt *et al*. 1995; Chen *et al*. 1995; Wallace *et al*. 1999).

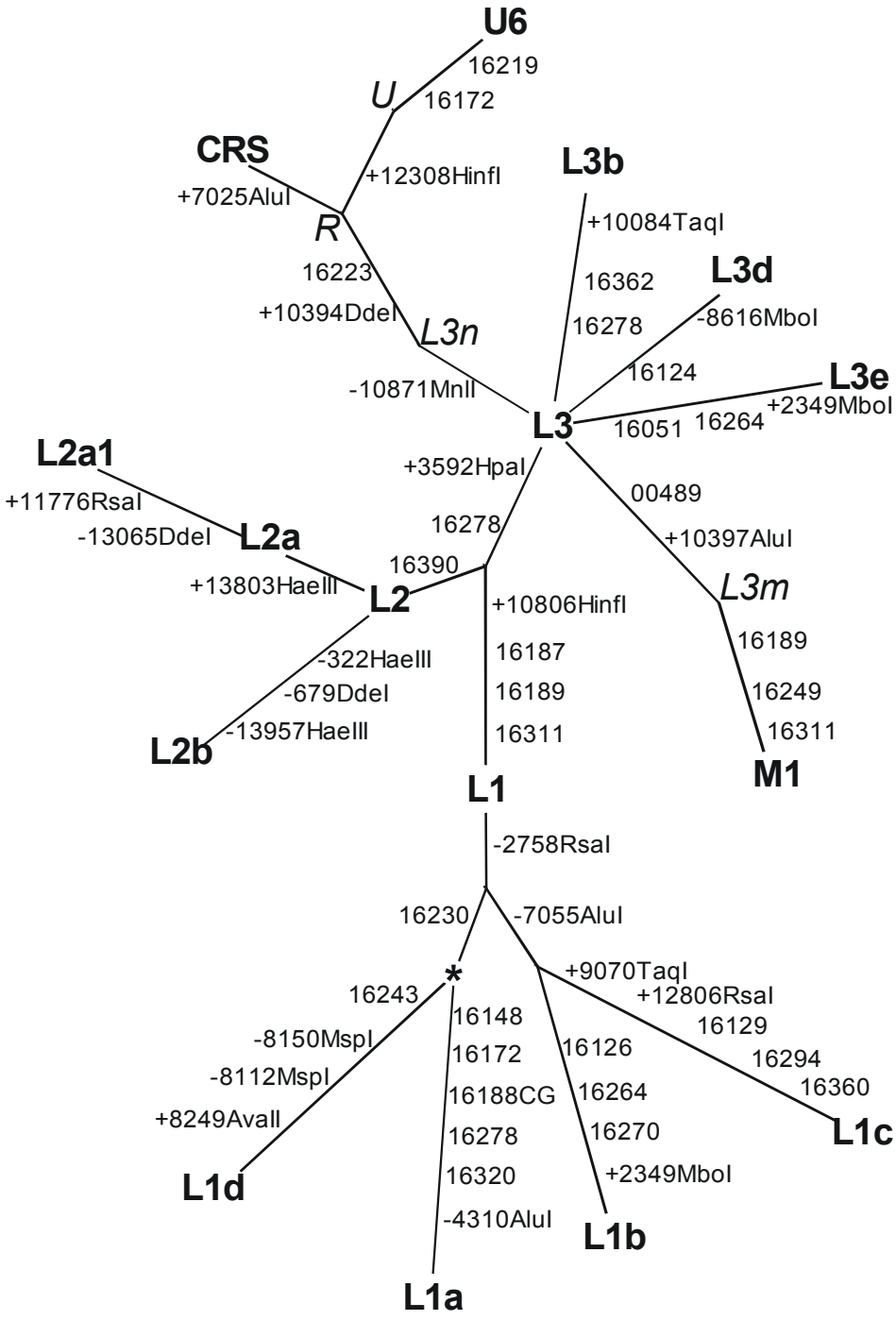

Figure 1. Mitochondrial DNA haplogroups in African populations, a synthesis from Chen *et al*.1995, Watson *et al*. 1997, Bandelt *et al*. 1997 and Rando *et al*. 1998. Each haplogroup is separated by selected sequence or restriction enzyme polymorphisms at np positions shown beside branches. HVS-II variation is not included. Only in case of transversions the change of nucleotide is specified. Polarity of RFLP sites is shown compared to the reference sequence (CRS). Intermediate supergroup labels are shown in italic. Root, noted by an asterisk, is inferred (as in Watson *et al*.1997) by chimpanzee outgroup.

Haplogroup L1 (+3592 *Hpa*I; +10806 *Hinc*II; HVS-I motif 187-189-223-278-311) diverges at the root of human mtDNA tree when chimpanzee mtDNA is used as an outgroup (Cann *et al*. 1987; Chen *et al*. 1995; Watson *et al*. 1997). All !Kung sequences and most of the Pygmies (Vigilant *et al*. 1991) belong to L1. Lack of the *Rsa*I site at 2758 ('A' shared with chimpanzee at np 2758) defines further the largest fraction of L1, encompassing sub-groups L1a, L1b, L1c and L1d (Fig. 1). L1b and most lineages from L1c share a further loss of an *Alu*I site at np 7055 (Chen *et al*. 1995).

L1a (-2758 *Rsa*I; -4310 *Alu*I; HVS-I motif 16148-16172-16188CG-16230-16278-16320; compared to L1) lineages are spread widely through eastern, central and southern Africa (Chen *et al*. 1995; Soodyall and Jenkins 1992a; Soodyall *et al*. 1996; Watson *et al*. 1997) and coalesce to a common node with age about 52,000 years (Watson *et al*. 1997). A sub-set of this haplogroup harboring the COII/tRNA^{Lys} 9-bp deletion is thought to have been exposed in the expansion of Bantu-speakers (Bandelt *et al*. 1995; Chen *et al*. 1995; Soodyall *et al*. 1996), probably together with cluster L3b (Watson *et al*. 1997) and a sub-cluster of L2 (see below).

L1b (-2758 *Rsa*I; -7055 *Alu*I, +2349 *Mbo*I, +185 *Taq*I; HVS-I motif 16126-16264-16270) variation is mainly restricted to western Africa (Graven *et al*. 1995; Mateu *et al*. 1997; Rando *et al*. 1998; Watson *et al*. 1997). HVS-I variation shows that its phylogeny is star-like, with two founders separated by a transition at 16293. The diversity of the cluster gives an age of 17,000 years for its expansion, starting likely from the Sahel Zone. Its spread to Northwest Africa was probably more recent, in the Neolithic or accompanied by the slave trade (Rando *et al*. 1998).

L1c (-2758 *Rsa*I; -7055 *Alu*I, +9070 *Taq*I, +12806 *Rsa*I; HVS-I motif 16129-16294-16360; HVS-II motif 00186A-00189C) is a cluster spread mainly in Central and Western Africa (Rando *et al*. 1998; Watson *et al*. 1997). A further subset of this group, characterized by an *Alu*I site at np 10319, is common in Biaka Pygmies (Wallace *et al*. 1999).

L1d (-2758 *Rsa*I; -8112 *Msp*I, -8150 *Msp*I, +8249 *Ava*II; HVS-I motif 16230-16243) is a cluster restricted to South African Khoisan populations (Bandelt and Forster 1997; Soodyall and Jenkins 1992a; Wallace *et al*. 1999) with a great time depth. It coalesces near to the root of the human mtDNA tree, (e.g., the !Kung cluster in Vigilant *et al*. 1991; cluster L1α in Wallace *et al*. 1999).

Haplogroup L2 (+3592 *Hpa*I; 16223-16278-16390) as a whole is pan-African in its spread (Watson *et al*. 1997). Both HVS-I (53,000-59,000 years; Watson *et al*. 1997) and RFLP (59,000-78,000 years; Chen *et al*. 1995) based time estimates suggest an ancient, late Pleistocene expansion for this cluster. The most frequent and widely spread sub-cluster of L2 (L2a in Fig. 1) is defined by an additional *Hae*III site at np 13803 (Chen *et al*. 1995). Yet, several of L2 sub-clusters reveal more recent expansion phases. One distinct sub-branch of L2 (defined by -322 *Hae*III, -679 DdeI, -13957 *Hae*III) is the dominant mtDNA cluster in western Africans (Chen *et al*. 1995; Rando *et al*. 1998) with coalescence age 17,000 years and suggests that it has co-expanded in western Africa together with L1b. Another specific sub-group within L2a is restricted (Chen *et al*. 1995; Wallace *et al*. 1999) mainly to Mbuti Pygmies (-13065 *Dde*I, +11776 *Rsa*I).

Haplogroup L3 (-3592 *Hpa*I, +10394 *Dde*I; HVS-I motif 16223) reveals an eminent east to west frequency gradient in Africa and the fact that its diversity in east exceeds that of the rest of Africa suggests an eastern-African origin for this key cluster that mothers all non-Africans

(Watson *et al*. 1997). The hypothesis of eastern-African origins for Eurasians is supported also by nuclear markers, e.g. the linkage-disequilibrium patterns in the CD4 locus on chromosome 12 (Tishkoff *et al*. 1996). L3 reveals together with its sister-cluster L2 traces of an ancient population expansion at around 60,000 years ago in eastern Africa. Three sub-groups specific of western Africa, one of Asia and one with global spread have been characterized (Fig. 1) so far.

Sub-group L3b (+10084 *Taq*I; HVS-I motif 16278-16362) is spread mostly in western Africa and its younger coalescence time (around 30,000 years) compared to L3 diversity in eastern Africa indicates that L3 could have expanded in western Africa comparatively recently (Watson *et al*. 1997).

Sub-group L3d (-8616 *Mbo*I; HVS-I motif 16124) is starlike for both HVS-I (Rando *et al*. 1998) and RFLP data (Chen *et al*. 1995), with a coalescence time around 30,000 years. Its spread, like for L3b, is restricted to western sub-Saharan Africa (Rando *et al*. 1998).

Sub-group L3e (+2349 *Mbo*I; HVS-I motif 16051-16264) is another western African variety of L3 found so far mainly in Senegalese populations (Chen *et al*. 1995; Rando *et al*. 1998).

Sub-cluster M1 (+10397 *Alu*I; HVS-I motif 16129-16223-16249-16311) derives from L3m (M) - the dominant mtDNA haplogroup in Asian populations (see below). Its spread in Africa is restricted mostly to the east (Passarino *et al*. 1998; Quintana-Murci *et al*. 1999) although it is occasionally found also in northern Africa (Rando *et al*. 1998) and Middle East (Macaulay *et al*. 1999b). According to RFLP and control region variation this group seems to have expanded in eastern Africa 36,000-48,000 years ago and is thought to have survived there as a relic of an early population movement from the Horn of Africa to Asia (Quintana-Murci *et al*. 1999).

Haplogroup U6 (12308G, +2349 *Mbo*I; HVS-I motif 16172-16219) was first identified by its distinct control region motif in a sample of Berbers from Algeria (Corte-Real *et al*. 1996). This 'Berber motif' is mainly found in northern and northwestern Africa among Moroccan and Algerian Berbers and Mauritanians. By its phylogenetic position this group is disjoint to other African clusters (Fig. 1), featuring several substitutions that are commonly found in Eurasia. The phylogeny of U6 is split into two branches that are both present in the Near East (Di Rienzo and Wilson 1991; Rando *et al*. 1998) suggesting that U6 may have originated from there or from Northeast Africa and subsequently dispersed to Northwest Africa. The age of the two branches is around 30,000 years although a more recent, probably Neolithic, date of its dispersal should be considered as plausible (Rando *et al*. 1998).

The age of the MRCA calculated from the African (haplogroup L) diversity alone ranges roughly between 100-170 thousand years (Chen *et al*. 1995; Wallace *et al*. 1999; Watson *et al*. 1997). This estimate agrees with that for the whole humankind, "the mitochondrial Eve" (Cann *et al*. 1987), suggesting that (i) the origin of human species lies in Africa, (ii) the colonization of other continents was recent (< 200,000 years) and (iii) other pre-existing hominid *speciae* were replaced by modern humans.

MtDNA variation in eastern Asians and Native Americans

Haplogroups A, B, C and D were the first Asian specific haplogroups recognized among Native American populations (Torroni *et al*. 1993a; Torroni *et al*. 1992). Subsequently, it has been shown that C and D share a more recent common ancestor in haplogroup M, while haplogroups B and A coalesce to another common node, termed N. Thus, phylogenetically mtDNA variation of present-day Asian populations can be classified under these two macro-clusters (Fig. 2). Both

M and N coalesce to the African macro-cluster L3, the consensus sequence of which can be considered as an the MRCA of all non-Africans. Yet the geographical spread and the most likely origin thereby of L3 lineages, not belonging to clusters N and M, seems to be restricted solely to Africa (Watson *et al*. 1997). Substantial frequency and diversity of haplogroup M lineages both among Asians and Ethiopians (Passarino *et al*. 1998) and their rarity in Europe support (Quintana-Murci *et al*. 1999) the hypothesis that Asia was colonized first, by a separate early migration starting from eastern Africa (see above) following the southern route (Cavalli-Sforza *et al*. 1994; Lahr and Foley 1994). Considering the haplogroup M diversity in Asians this migration dates more than 50,000 years back (Quintana-Murci *et al*. 1999).

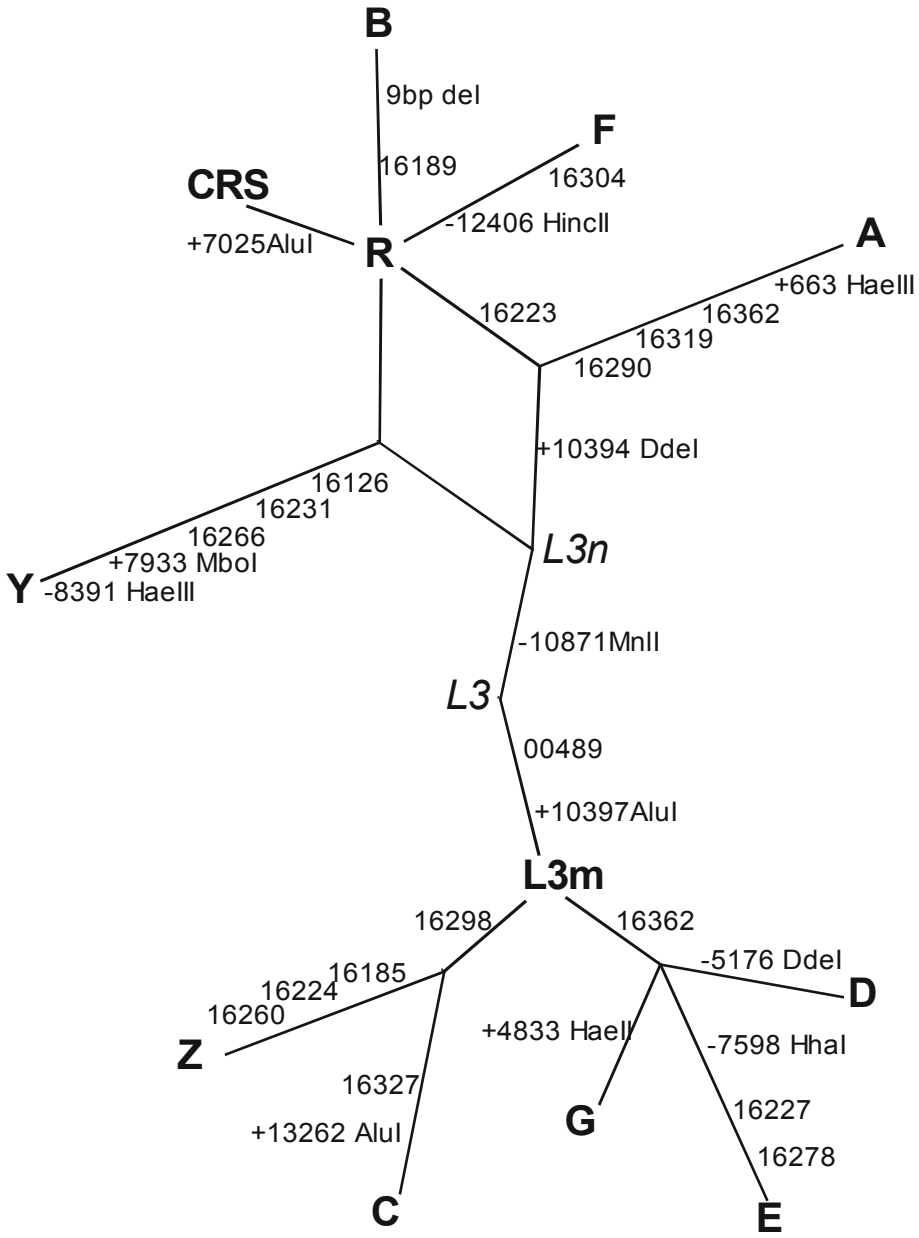

Figure 2. Mitochondrial DNA haplogroups in Asian populations, a synthesis from Ballinger *et al*. 1992; Torroni *et al*. 1994, Schurr *et al*. 1999 and Quintana-Murci *et al*. 1999. Each haplogroup is separated by selected sequence or restriction enzyme polymorphisms at np positions shown beside branches. HVS-II variation is not included. Only in case of transversions the change of nucleotide is specified. Polarity of RFLP sites is shown compared to CRS. Intermediate supergroup labels are shown in italic.

Table 1. mtDNA haplogroup frequencies in Asian and Amerindian populations

	n	A	B	C	D	E	F	G	Z	Y	M*	N*	R*
South-East Asia[1]	153	2.6%	22.2%	0.6%	3.3%	8.5%	16.3%	1.9%	ND	0.0%	30.7%	2.6%	11.7%
Tibetans[2]	54	11.1%	5.6%	3.7%	16.7%	7.4%	14.8%	5.6%	ND	0.0%	31.4%	0.0%	3.7%
Mongols[3]	103	3.9%	9.7%	14.6%	24.3%	2.9%	8.7%	ND	3.9%	1.9%	13.6%	5.8%	10.7%
Siberians[4]	153	1.3%	0%	33.3%	13.7%	0%	0.6%	2.0%	ND	26.8%	13.1%	11.1%	ND
NE Siberians[5]	202	5.4%	0%	31.2%	1.0%	0%	0.0%	48.0%	5.9%	8.4%	0.0%	0.0%	0.0%
Eskimos[6]	50	80%	0%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%
Na-Dene[7]	130	76.9%	16.9%	2.3%	2.3%	0%	0%	0%	0%	0%	0%	1.5%[8]	0%
Amerinds[7]	398	39.2%	22.1%	17.6%	17.3%	0%	0%	0%	0%	0%	0%	3.8%[8]	0%
Northern	115	39.1%	19.1%	20.9%	8.7%	0%	0%	0%	0%	0%	0%	12.2%[8]	0%
Central	137	65.7%	28.5%	2.9%	2.2%	0%	0%	0%	0%	0%	0%	0.7%[8]	0%
Southern	146	14.4%	18.5%	28.8%	38.4%	0%	0%	0%	0%	0%	0%	0%	0%

[1] Ballinger *et al*. 1992

[2] Torroni *et al*. 1994b

[3] Kolman *et al*. 1996

[4] Nivkhs, Evenks and Udegeys; Torroni *et al*. 1993b

[5] Koryaks and Itel'men; Schurr *et al*. 1999

[6] Torroni *et al*. 1993a

[7] Torroni *et al*. 1992

[8] Caucasian mtDNA types; Torroni *et al*. 1992

* unclassified derivatives

Haplogroup M is defined (Chen *et al*. 1995) by the presence of two closely located restriction sites, +*Dde*I 10394 (10398 A to G) and +10397 *Alu*I (10400 C to T) and was first detected among South-East Asian populations (Ballinger *et al*. 1992). About 50% of native Asians belong to this haplogroup (Table 1) which is further divided into a number of recognized sub-haplogroups (C, D, E, G and Z; Fig. 2). Most of these sub-groups are characterized by their specific RFLP and HVS-I signatures. Haplogroups D, E and supposedly also G, share a C at np 16362 in HVS-I, known however to be one of the evolutionary hot spots (Hasegawa *et al*. 1993). Haplogroups C and Z share a C at np 16298.

Haplogroup C (+13262 *Alu*I; HVS-I motif 16223-16298-16327) is widely spread among northern Asian populations. Its geographic ditribution extends from Central Asia to the Kamchatkan peninsula and Japan. Frequency of haplogroup C is highest among Evenks (84.3%) and Yukaghirs (59.3%) (Torroni *et al*. 1993b) and very low or marginal in southern Asian populations (group R in Ballinger *et al*. 1992), where its occasional presence in Taiwanese Han could probably be ascribed to recent migrations. C is also one of the four basic Native American mtDNA haplogroups (Torroni *et al*. 1993b). Characteristic of most Native American haplogroup C mtDNAs is an additional transition at np 16325 (Forster *et al*. 1996) which has occasionally been found in some Asian populations (Comas *et al*. 1998; Horai *et al*. 1996; Kolman *et al*. 1996) as well.

Haplogroup D (-5176 *Alu*I; HVS-I motif 16223-16362) reveals also a strong south to north frequency difference in eastern Asians. Similarly to haplogroup C it is widely spread and frequent in Central Asian populations and native Siberians but sampled only in low frequencies in southeastern Asians (Table 1). More than one third of the maternal lineages of Nganasans and Yukaghirs (Torroni *et al*. 1993b) belong to haplogroup D. Among Native Americans, haplogroup D is most frequent in Southern American native populations (Table 1).

Haplogroup E (-7598 *Hha*I; HVS-I motif 16223-16362) is another widely spread mtDNA cluster in Asian populations. In southeastern Asia its spread is quite uneven (group G in Ballinger *et al*. 1992): frequency of E is above 25% in Sabah aboriginals and Malays while it is absent among other southeastern populations of Asia. Its uneven distribution is further manifested in the fact that E is absent among most Siberian populations and Native Americans but found at recognizable frequencies in Tibetans. Haplogroup E is frequently associated with two additional HVS-I mutations, namely, 16227 and 16278. The presence of this D-loop signature among Central Asian populations (Comas *et al*. 1998), Mongolians (Kolman *et al*. 1996), Chinese (Horai *et al*. 1996) and Koreans (Lee *et al*. 1997) shows the extent of its geographical spread together with other haplogroups specific of Mongoloid populations.

Haplogroup G (+4830 *Hae*II) is frequent in northeastern Siberian populations (Table 1), reaching the frequency of 40-60% in Koryaks and Itel'men (Schurr *et al*. 1999). It has been found also in Tibetans and Koreans but has not been sampled as yet among most other southeastern Asian populations (Ballinger *et al*. 1992; Starikovskaya *et al*. 1998; Torroni *et al*. 1994b).

Haplogroup Z (HVS-I motif 16129-16185-16223-(16224)-16260-16298) was recently defined by Schurr *et al*. (1999) as yet another haplogroup M sub-group among native Siberian populations. Its highest frequency is observed in Tungusic-speaking populations. However, the spread of haplogroup Z is not restricted to northern Siberian populations only (Table 1). The characteristic D-loop motif has been occasionally found among Mongolians (Kolman *et al*.

1996), Central Asians (Comas *et al*. 1998) and also, interestingly, among the Saami (Lahermo *et al*. 1996; Sajantila *et al*. 1995). The presence of haplogroup Z among a Finno-Ugric speaking population can perhaps be best explained by comparatively recent contacts of the Saami with other Circum-Arctic populations rather than be taken for a trace of an ancient link between Uralic and Altaic languages, because other Finno-Ugric populations lack these signs of Mongoloid pattern in their mtDNAs and most other mtDNA haplotypes found among the Saami belong to typical European-specific haplogroups (Villems *et al*. 1998).

Some additional M sub-groups are likely to be defined in the future, when more data of both coding and control region sequence variations will become available. For example, HVS-I motif 16209-16223, common in Japanese and Ainus (Horai *et al*. 1996), seems to be associated with a number of coding region substitutions (Ozawa 1995) and deserves a separate haplogroup status. Another motif, 16223-16297, is common in both Japanese and Chinese (Horai *et al*. 1996), but is also found in Koreans (Lee *et al*. 1997) and Mongolians (Kolman *et al*. 1996). Furthermore, these two motifs seem to derive from a common ancestor because they share a T to C transition at np 9824 (Ozawa 1995).

Super-cluster N in Asia is subdivided into distinct mtDNA haplogroups A, B and F (Fig. 2). Haplogroup A has maintained the ancestral T at np 16223, while B and F derive from N through a common ancestor with most Europeans, haplogroup R. Although the ancestral nodes N and R are both shared by European and eastern Asian mtDNA clusters, it is observed very rarely that haplogroups characteristic of Europeans are found in native Asian population and *vice versa*.

The geographic spread of haplogroup A (+663 *Hae*III; HVS-I motif 16223-16290-16319-16362) is similar (Wallace 1995; Wallace *et al*. 1999) to that of haplogroups C and D: all of them are more frequent in northern Asians and rare or even absent in southeastern Asians. Haplogroup A is the most frequent mtDNA group among Chukchis and Eskimos (Schurr *et al*. 1999; Starikovskaya *et al*. 1998). In Native Americans haplogroup A frequency is also the highest in north among the Na-Dene speakers. It has been suggested that an expansion of a subset A2 in these populations occurred after Younger Dryas glacial relapse (Forster *et al*. 1996).

Haplogroups B, F and Y derive from super-cluster R (Fig. 2) which defines also most of European mtDNA variation with haplogroups specific of western-Eurasian populations.

Haplogroup B (9-bp deletion between COII and $tRNA^{Lys}$, HVS-I motif 16189) is common in central and southeastern Asia, reaching almost fixation in some Oceanian populations (Ballinger *et al*. 1992; Hertzberg *et al*. 1989; Lum *et al*. 1998; Melton *et al*. 1995; Redd *et al*. 1995; Stoneking *et al*. 1990; Sykes *et al*. 1995). B is also one of the founder haplogroups of Native Americans (Torroni *et al*. 1993a; Torroni *et al*. 1992). Its lack in most of the Siberian populations has raised speculations that the 9-bp deletion could have arrived to the Americas via a separate and later migration either along the coastal regions of East Asia (Torroni *et al*. 1994c; Torroni *et al*. 1993a; Torroni *et al*. 1992) or even through the Pacific (Cann 1994). However, the latter view has been put under doubt because Polynesian and Amerind B sequences are different (Bonatto *et al*. 1996) and because haplogroup B was recently found in some Siberian populations, like the Buryats and Tuvinians (Derenko *et al*. 1999).

Haplogroup F (-*Hpa*I/*Hinc*II 12406) is most common in populations of southeastern Asia (Table 1), like the Vietnamese and Malays (Ballinger *et al*. 1992). The characteristic HVS-I motif (16172-16304) indicates that F is also spread in Mongolia and Central Asia (Comas *et al*. 1998; Kolman *et al*. 1996). Interestingly, this cluster (group III in (Sykes *et al*. 1995) is quite rare in

Polynesians (0.6%), although its sister group B is very frequent there. Haplogroup F is not found among Native Americans either.

Haplogroup Y (+7933 *Mbo*I, -8391 *Hae*III, +10394 *Dde*I; HVS-I motif 16126-16231-16266) is a recently defined (Schurr *et al*. 1999) lineage group that has spread mostly in northeastern Asia - it is absent among most Siberian populations, but appears frequently among populations living in the Far East, in Kamchatka (Schurr *et al*. 1999), Sakhalin (64.9% in Nivkhs; Torroni 1993) and Korea. Schurr et al. (1999) suggest that haplogroup Y has originated in the Amur River region. Its D-loop motif suggests that this haplogroup is also very common among Ainus (Horai *et al*. 1996) and it seems to have spread as far west as to Central Asia (Comas *et al*. 1998) and Anatolia (Comas *et al*. 1996), perhaps together with the spread of the Altaic languages.

MtDNA variation in South Asia and genetic affinities of the Indian populations

The population of India is divided by numerous social, cultural, ecological and biological characteristics into at least 4 thousand communities that behave more or less as endogamous units (Singh 1997). The linguistic division recognizes most people speaking either Sanskrit or Dravidic derived languages, while a minor part of the population speaks Austroasiatic and Sino-Tibetan languages. An extensive literature exists on classical genetic studies carried out among Indian populations (reviewed, e.g,. in Cavalli-Sforza *et al*. 1994; Papiha 1996), revealing high heterogeneity among different caste and tribal groups. Inbreeding and social restrictions in marriage customs have apparently lead to the formation of isolated groups with the result that castes of the same status from different regions may look extremely different in certain allele frequencies. The genetic structure underlying India's modern populations is generally thought to consist of at least three basic components, although Risley (Risley 1903) describes 7 and Majumder (Majumder 1990) claims that 'people of India cannot be classified into a fixed set of ethnic categories'. The components are: (1) a minor *Australoid* component, represented now by only a few tribes and dating back probably to the Palaeolithic period; (2) an extensive 'Caucasoid' component, thought to derive mainly from Neolithic and Bronze age migrations; and (3) a 'Mongoloid' component, expressed mainly in northeastern and eastern territories and probably resulting from multiple recent migrations of small tribes from the east. Some well distinguishable populations with possibly African origins are also found in India but, overall, Indians are genetically characterized as Caucasoids (Cavalli-Sforza *et al*. 1988; Mourant *et al*. 1976; Nei and Roychoudhury 1982).

The commonly held theories relate the dominant 'Caucasoid share' in Indians to the Indo-European languages spoken in India as well as in most countries of Europe. The spread of Indo-European speaking populations to India is thought to have occurred recently through the putative Indo-Aryan invasion around 4,000 years ago (e.g. Poliakov 1974; Thapar and Rahman 1996). Recent limited mtDNA analyses (Barnabas *et al*. 1996; Passarino *et al*. 1996a) have favored this view by finding in Indians haplotypes that are uncommon among Mongoloids.

The mtDNA pool of the southeastern Asians contains haplogroup B at high frequencies (see above), whereas in Indians this haplogroup is absent or rare (Harihara *et al*. 1988; Passarino *et al*. 1993; Watkins *et al*. 1999). At the first glance, this finding supports the cited above suggestions about the Caucasoid genetic affinity of Indians. Meanwhile, a majority of Indian populations, irrespective of their linguistic or geographic origin, are characterized by a high frequency of combined presence of a *Dde*I site at 10394 and an *Alu*I site at 10397 (Bamshad *et al*. 1997; Passarino *et al*. 1996a; Passarino *et al*. 1996b). These restriction sites define haplogroup M which is spread widely among other Asian populations. Passarino and colleagues (Passarino *et al*. 1996a; Passarino *et al*. 1996b) noticed that the frequency of haplogroup M is

the lowest in Punjab and increasingly higher in southern India. They suggested that the putative Caucasoid invasion had perhaps penetrated into India from that direction and by introducing non-M haplotypes was thus responsible for the gradient observable in the spread of M lineages in India now.

Mountain *et al*. (1995) noticed a starlike non-African lineage cluster when combining Indian mtDNA control region sequences to those from other continental populations. That cluster, representing supposedly a major expansion starting when or after modern humans initially left Africa, was dated to have a coalescence time approximately 65,000 years (Mountain *et al*. 1995).

The apparent lack of any eminent clustering of Indians by caste rank or tribe has been witnessed in several studies employing different regions of mtDNA (Bamshad *et al*. 1996; Barnabas *et al*. 1996; Mountain *et al*. 1995; Soodyall and Jenkins 1992b) or markers from nuclear genome (Bhattacharyya *et al*. 1999; Majumder *et al*. 1999). Yet differences by caste rank exist in mismatch distribution patterns (Mountain *et al*. 1995) indicating possible differences in the growth rate of distinct caste groups. Also, genetic distances between caste populations from the same geographic region reveal a particular pattern: middle castes stand between lower and higher castes and there is a lack of haplotype sharing between distinct caste groups, indicating a restriction of gene flow between castes (Bamshad *et al*. 1998; Bhattacharyya *et al*. 1999).

Several questions relating to the origins of Indian populations are open yet: whether the common share of haplogroup M in Indians and eastern Asians is due to a common late Pleistocene ancestor or recent contacts? whether the non-M lineages are Indian specific or short derivative branches of lineages spread among other Eurasian populations? and in what extent do the mtDNA lineages of Sanskrit and Dravidic speaking populations differ?

MtDNA diversity in native Australians and Papuas

There is still a need for a systematic approach, combining mtDNA RFLP and control region sequence data, to classify populations from Papua New Guinea and Australia. The data available from the literature up to now (Redd and Stoneking 1999; Stoneking *et al*. 1990; Sykes *et al*. 1995; van Holst Pellekaan *et al*. 1997; van Holst Pellekaan *et al*. 1998) show that the aboriginal populations of Australia and PNG are very diverse and stand quite distant from other populations in the world, a fact that can be explained by a comparatively long isolation. Distance calculations based on control region variation connects Aboriginal Australian populations most closely with populations from the Asian continent (Redd and Stoneking 1999).

MtDNA variation in Western Eurasia

The current nomenclature of European mtDNAs relies on the synthesis (Macaulay *et al*. 1999b; Richards *et al*. 1998; Torroni *et al*. 1996) of both control region (Richards *et al*. 1996) and RFLP-based (Torroni *et al*. 1994a) classifications. The systematics of mtDNA clusters is more refined for Europeans thanks to the existance of more extensive studies of European than of other world populations. Most haplogroups that are generally defined by specific coding region variants, have already been split further into minor sub-groups, sometimes defined by a few or even single control region polymorphisms.

More than 90% of the European mtDNAs belong to 9 distinct haplogroups (Fig. 3) that are highly specific of western Eurasia in general (Macaulay *et al*. 1999b; Torroni *et al*. 1996; Torroni *et al*. 1994a). All of them descend from a common node in the human mtDNA phylogeny - L3n (N). This super-clade incorporates also several clusters specific of Asian populations (A, B, F and Y), meaning that all European mothers together do not share an

exclusive common maternal ancestor of their own. Super-cluster N has one major branch R (Fig. 3) in common with Asian populations, defined by transitions at nps 12705 and 16223 (Hofmann *et al*. 1997; Macaulay *et al*. 1999b). This derivative sub-section hosts most of the variation confined to Europeans. Five traditional haplogroups that form this cluster are H, J, T, V, and U. The latter includes also haplogroup K as its sub-group. Haplogroups H and V share an ancestral node, termed HV, a number of unclassified derivatives of which have been sampled in the Near East (Macaulay *et al*. 1999b), where the origin of this cluster could lie.

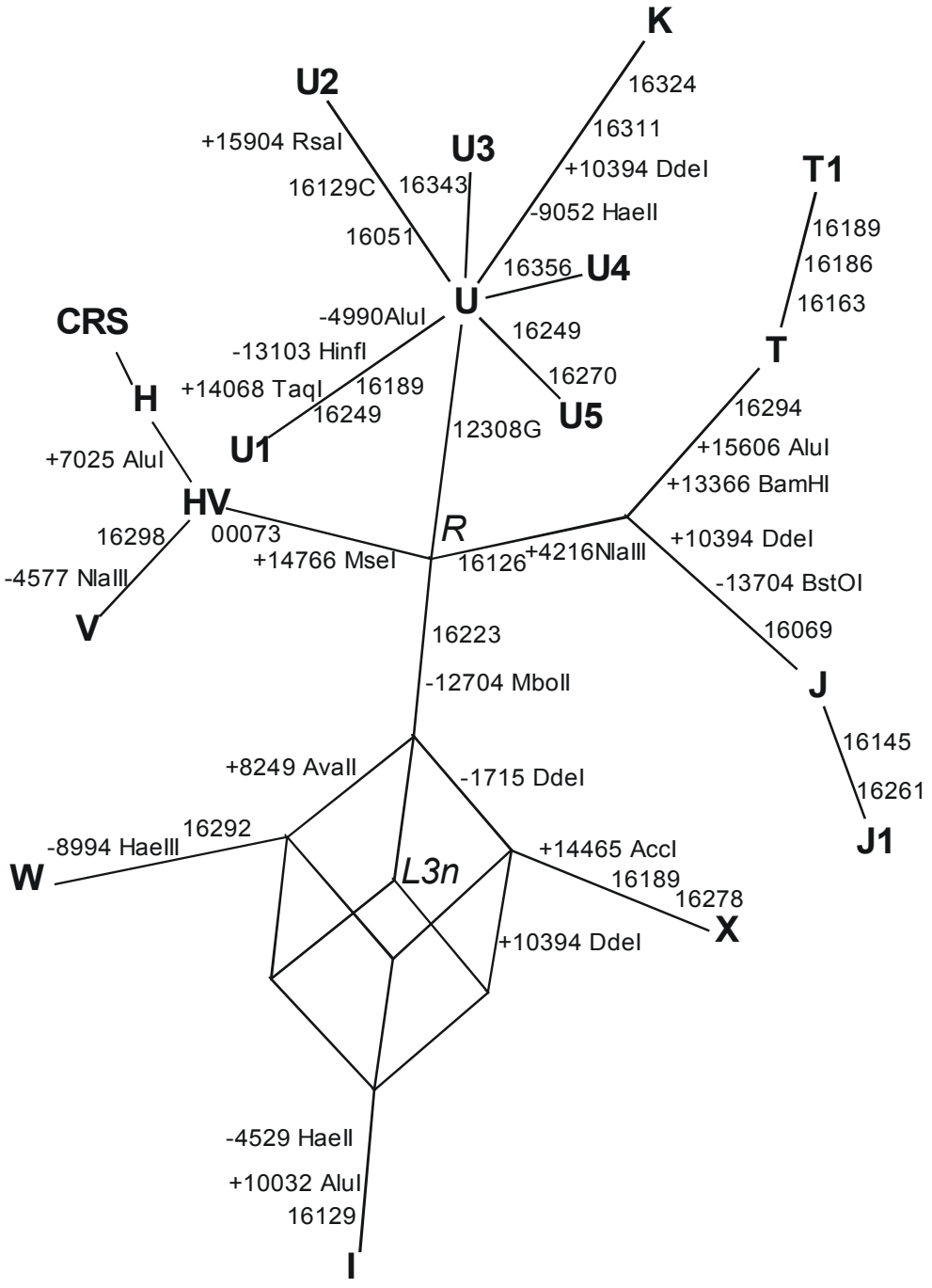

Figure 3. Mitochondrial DNA haplogroups in European populations, adapted from Macaulay *et al*. 1999. Each haplogroup is separated by sequence or restriction enzyme polymorphisms at np positions shown beside branches. Only in case of transversions the change of nucleotide is specified. Polarity of RFLP sites is shown compared to CRS. Intermediate supergroup labels are shown in italic.

Haplogroup H (-7025 *Alu*I) is the largest mtDNA haplogroup spread in Europe in a wide range of populations with frequencies around 40% (Table 2). The lack of *Alu*I site at 7025 (7028C) correlates with A at position 00073 in HVS-II, known to be characteristic of European populations (Wilkinson-Herbots *et al*. 1996) and rare or absent elsewhere. That substitution thus builds a bridge between RFLP and control region-based studies. Relying on HVS-I sequence types, this cluster is most frequent among the Basques (group 1 in Richards *et al*. 1996). Haplogroup H includes the CRS (Anderson *et al*. 1981) at the center of its phylogeny. That is overall the most common HVS-I haplotype in Europeans. Numerous short derivatives identified by few HVS-I mutations (Richards *et al*. 1998) or restriction sites (Torroni *et al*. 1998) give the cluster a star-like appearance. Haplogroup H divergence time in Europe is 20,000 - 30,000 years (Richards *et al*. 1998; Torroni *et al*. 1998; Torroni *et al*. 1996) and, given its lower frequency in the Near East, suggests that it reached Europe quite early, possibly during the Upper Palaeolithic (Richards *et al*. 1998).

Haplogroup V (-4577 *Nla*III; +15904 *Mse*I; HVS motif 72C- 73A-16298C) is a sister group to H. Yet, V is much younger and has been suggested to reveal traces of a late Upper Palaeolithic population expansion starting from Iberia (Torroni *et al*. 1998).

Haplogroup U (12308G) is the most ancient haplogroup found in Europeans, with an estimated time depth over 50,000 years (Richards *et al*. 1998; Torroni *et al*. 1996). It is split into sub-clusters U1-U5 and K in Europe; an additional distinctive U sub-branch, U6, is mainly restricted to western and northern Africans (Corte-Real *et al*. 1996; Rando *et al*. 1998). Sub-cluster U5 (group 5 in Richards *et al*. 1996) catches up most of U sequences in Europe and is the most divergent cluster there, dating back to the times when Europe was first colonized by modern humans (Richards *et al*. 1998). Several U5 sub-clusters are specific for certain populations or geographical regions. For example many Saami and some other northeastern Europeans share a specific for them "Saami motif" (Lahermo *et al*. 1996; Sajantila *et al*. 1995). Other haplogroup U sub-clusters are less evenly spread in Europe and date to more recent periods. The eldest among them, U4, dates to more than 25,000 years, suggesting another pre-LGM expansion (Richards *et al*.1998). Haplogroup K (-9052 *Hae*II, +10394; HVS-I motif 16224-16223), recognized now as a sub-group of U (Hofmann *et al*. 1997; Richards *et al*. 1998) shows divergence time of 13,000-18,000 years (Richards *et al*. 1998; Torroni *et al*. 1994a).

Haplogroup J (+10394 *Dde*I, -13704 *Bst*NI; HVS-I motif 16069-16126) has several distinct founder types in Europe. Although its overall time depth is definitely pre-LGM and its geographic spread restricted to western Eurasia, estimation of variance within distinct sub-founders indicates that it could have been introduced to Europe from the Near East within the last 10,000 years (Richards *et al*. 1996; Richards *et al*. 1998).

Haplogroup T (+4216 *Nla*III, +13366 *Bam*HI, +15606 *Alu*I, -15925 *Msp*I; HVS-I motif 16126-16294) comprises about 8% of the mtDNA diversity in Europe and can be split into two sub-clusters. Again, haplogroup T time depth extends back to 46,500 years (Richards *et al*. 1998), yet, when dissected, one of the sub-groups, T1 (16126-16163-16186,-16189-16294), is only 9,000 years old in Europe. The rest of T dates back then to 32,000 years (Richards *et al*. 1998). Thus, it seems most likely that haplogroup T has migrated to Europe at least twice.

Table 2. mtDNA haplogroup frequencies in European and Near Eastern populations

	n	H	I	J	T	U	U-K	V	W	X	L	M	Other
Italians[1,2]	147	36.1%	4.1%	9.5%	9.5%	18.4%	7.5%	3.4%	2.0%	4.8%	0.7%	0.7%	3.4%
Spanish[3]	182	45.1%	1.1%	4.4%	9.9%	14.3%	3.3%	5.5%	1.1%	1.6%	3.3%	0.5%	9.9%
Swedes[1]	37	40.5%	0.0%	2.7%	21.6%	16.2%	13.5%	5.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Finns[1]	49	40.8%	2%	14.3%	6.1%	16.3%	4.1%	4.1%	4.1%	4.1%	0.0%	2.0%	2.0%
Icelanders[4]													
American Caucasians[5]	175	40.0%	6.9%	9.1%	3.4%	ND	7.4%	ND	1.7%	1.7%	0.0%	0.0%	ND
Germans[6]	67	46.3%	4.5%	7.5%	11.9%	11.9%	4.5%	4.5%	1.5%	2.9%	0.0%	0.0%	4.5%
Adygei[7]	50	30.0%	0%	4.0%	14.0%	32.0%	2.0%	0%	2.0%	0%	0%	6.0%	10.0%
Druze[7]	45	13.3%	2.2%	6.7%	4.4%	11.1%	15.6%	0%	0.0%	26.7%	0%	2.2%	17.8%
Europe, average[8]	942	ND	1.6%	11.1%	7.5%	ND	6.8%	4.2%	1.3%	1.7%	ND	ND	ND

[1] Torroni *et al.* 1996
[2] Torroni *et al.* 1997
[3] Torroni *et al*. 1999
[4] Helgason *et al*. 2000
[5] Torroni *et al*. 1994
[6] Hofmann *et al*. 1997
[7] Macaulay *et al*. 1999
[8] as inferred by Richards *et al*. 1998

Haplogroups coalescing (Fig. 3) to node L3n (12705T and 16223T) are comparatively rare (~5%) in Europe (Richards *et al*. 1998) compared to the derivative haplogroups of its daughter cluster R (~85%). Since N and R are also the main phylogenetic nodes to which all Asian haplogroups, except those belonging to the M cluster, coalesce to, it seems likely that both eastern and western Eurasian populations may have shared a common source population with both N and R founding types.

Haplogroup X (-1715 *Dde*I, +14465 *Acc*I, -10394 *Dde*I; HVS-I motif 16189-16223-16278) is a minor mtDNA haplogroup in Europeans with an age of 24,000 years, found at frequencies under 5% in Europe. Surprisingly it has been found not only in Caucasoids but also - rarely, though - among some Native Americans (Torroni *et al*. 1993a). It has been shown, however, that the presence of haplogroup X in the Americas, restricted mainly to northern Amerind populations, is not due to a recent admixture with descendants of Europeans but that X is probably yet another, the fifth, Native American founder haplogroup (Brown *et al*. 1998; Forster *et al*. 1996; Smith *et al*. 1999; Torroni and Wallace 1995).

Haplogroup I (-1715 *Dde*I, -4529 *Hae*II, +8249 *Ava*II, +10032 *Alu*I, +10394 *Dde*I; HVS-I motif 16129-16223) is another rare mtDNA cluster. Its time depth of around 35,000 years suggests an Upper Palaeolithic origin. It is spread mostly in northern and western Europe (Richards *et al*. 1998).

Haplogroup W (+8249 *Ava*II, -8994 *Hae*III, -10394 *Dde*I; HVS-I motif 16223-16292) has a more recent age of 18,500 years (Richards *et al*. 1998) and appears more diverse in southern than in northern Europe.

The temporal phases of the colonization of Europe by modern humans and their relative impact to the present day populations has been an issue of the issue of an intense debate over the last decade. First, there is the question of the continuity of Neanderthal genes in modern Europeans. Analyses of present-day populations have shown that mtDNA variation in Europeans is comparatively low and no outliers, identifiable as putative Neanderthal lineages, can be found (Richards *et al*. 1996; Torroni *et al*. 1994a). Most mtDNA-based studies of European populations estimate the ages of population expansions (Calafell *et al*. 1996; Sajantila *et al*. 1995) and those of major lineage clusters (Bertranpetit *et al*. 1996; Richards *et al*. 1996; Torroni *et al*. 1996) typically at 30,000 to 50,000 years, in agreement with the total replacement of pre-modern populations. Yet there is now more direct evidence to the matter. In 1997 Krings and colleagues (Krings *et al*. 1997) sequenced a part of the Neanderthal-type specimen mtDNA control region and found that the Neanderthal sequence fell outside the variation of modern humans. The Neanderthal and modern humans coalesce to a common ancestor who is about 4 times older than the common ancestor of all human mtDNAs.

The other topic, still under dispute, is the contribution to the European gene pool of a migration associated with the onset of farming in the Near East/Anatolia. Some authors have suggested nearly total replacement of the indigenous European Palaeolithic populations by neolithic farmers (Barbujani *et al*. 1998; Chikhi *et al*. 1998), while a classical demic diffusion scenario, based on principal component analyses of classical genetical markers (Ammerman and Cavalli-Sforza 1984; Cavalli-Sforza *et al*. 1994; Menozzi *et al*. 1978) proposed that a substantial minority of the European gene diversity originates from the western Asian neolithic farmers. Thirdly, a scenario of pioneer colonization involving only a few newcomers and an extensive continuity of the indigenous populations (Zvelebil 1986) has found support from mtDNA

analyses (Richards *et al*. 1996). There are some important differences (e.g. see Cavalli-Sforza and Minch 1997 and Barbujani *et al*. 1998 vs Richards *et al*. 1997; Richards *et al*. 1998) in the interpretations of different sets of markers affecting the estimates of time and contribution assigned to separate migrations. While two or more migrations at different times but from the same source population can leave a similar or even cumulative trace as a gradient of allele frequencies, proper founder analysis relying on mapping the differential cumulation of sequence diversity in a source and a recipient population can be used to quantify the contribution and time of each separate migration and/or back migration (Richards *et al*. 1998; Sykes 1999; Richards et al., in preparation).

Conclusions and the formulation of some questions

The power of the tools for interpreting genetic variation in human mitochondrial DNA has increased very significantly during the last five years. This progress became possible thanks to the flow of new experimental data, covering large and diverse populations worldwide and the employment of new phylogenetical methods. The sequencing of mtDNA control region in combination with the typing of a "standard" large set of RFLP markers in coding regions allows now nearly unambiguously to classify phylogenetically most of the lineages found in the Eurasian mtDNA pool. In turn, this emerging skeleton of the mtDNA tree feeds back to the phylogeographic interpretation of large data banks about the spread of the mtDNA variation globally and in a historic perspective.

Of course, several important and a large number of minor problems are unsolved yet. To mention just one: as far as western Eurasian populations are concerned, we still lack an understanding detailed enough of the phylogeny of the dominant haplogroup H. And there are still major problems in discerning the dates of expansion times and time depths through current tools of coalescence calculations and in finding the ways to consider "fast" and "slow" variant positions etc. Nevertheless, as already indicated, the progress is encouraging and suggests that it may now be possible to approach questions which only recently remained largely beyond our reach.

In my PhD work, I became interested in understanding the origin and rise of the genetic diversity of Eurasian mtDNA lineages. This problem became approachable thanks to an ever increasing evidence in favor of the "recent out-of-Africa" scenario of the spread of modern humans. Furthermore, as it was shown above, the analysis of the African varieties of mtDNA suggested that the global mtDNA variation outside Africa could perhaps trace back to the diversification of one single sub-Saharan African branch of the phylogenetic tree of mtDNA. Yet large gaps were apparent, the largest among them southern Asia, from where only a limited number of control region sequences were published without information about the coding region variation, or some restricted Indian populations characterized by a set of RFLP polymorphisms. Even less was known about the populations living in the Trans-Caucasus area. One may say that equally intriguing Iraqi, Iranian, Afghan, Syrian and southern Arabian populations have not been studied either, but as in many such investigations, availability of relevant samples is decisive. For the analysis of Indian and Trans-Caucasian populations the following questions were raised:

1. what mtDNA clusters are spread among the Indian and Trans-Caucasian populations?
2. how do these clusters connect to the global tree of mtDNA?
3. what is the diversity of mtDNA clusters present in Indian and the Trans-Caucasian populations as compared to related clusters in European and eastern Asian populations?
4. when and where did the separation of eastern and western Eurasian mtDNA lineages take place?
5. was there a single migration out of Africa or multiple dispersals?

6. what has been the impact of succeeding gene flow between geographically distinct populations on the example of Indian subcontinent?
7. what mtDNA lineages can be found to expose the link between Indo-European speakers from western Eurasia and India?
8. what has been their contribution to the maternal gene pool of Indians?

Results and discussion

The experimental basis of the current thesis relies on the analysis of DNA samples from 300 predominantly Indo-Aryan speaking Indian, 536 Trans-Caucasian and 424 eastern European individuals. Most of the experimental work done on Indian and a smaller part on the Trans-Caucasian and eastern European samples was performed by the author. References I and II are based on our joint research project with our colleagues from the University of Utah and include in addition the data from 250 Dravidic speaking Telugus. Each sample was PCR amplified and sequenced for 360 bp in the HVS-I region of mtDNA, followed by a search for known haplogroup-specific RFLP polymorphisms from the coding region. Sequences were aligned and identical sequences pooled as haplotypes. Haplotypes were connected through the reduced median network approach (Bandelt *et al*. 1995). Diversity and age estimates in each cluster and sub-cluster of the network were calculated through the estimator ρ, taking 20,180 years as an average time for the fixation of one transition between nps 16090-16365 (Forster *et al*. 1996). Haplogroup frequencies and diversities were compared by populations with one another and with the data available from the literature.

The place of Indian mtDNA variation in global scale and the link between Indian and West Eurasian mtDNA lineages (Reference I)

An analysis of mtDNA variation in Indian populations of two different language groups (Dravidic and Hindi) revealed the presence of haplogroups specific to both eastern and western Eurasian populations as well as of Indian-specific clusters not characterized before (Ref I; Fig. 1). All mtDNA lineages in Indians trace back to an African mtDNA ancestor in super-cluster L3, supporting the hypothesis of recent African origins of modern humans and the replacement of any pre-existing hominid species in the Indian sub-continent. Both East and West Eurasian-specific haplogroups in Indians are represented by deep Indian-specific branches. The primary clustering of mtDNA lineages is not language-specific (Hindi, Dravidic) and only a small fraction of Indian mtDNA lineages (<10%) can be ascribed to a relatively recent admixture with western Eurasians (Ref. I; Table 1). An attempt to date their arrival was made that yielded an estimate of about 9,000 years. This date, however, is most likely an average of a number of different West Eurasian donations to the Indian gene pool. Yet, it is more consistent with the time when domesticated cereals could have reached India from the Fertile Crescent (Cavalli-Sforza *et al*. 1994) than with later Bronze age migrations and our finding of these haplotypes also among Dravidic speakers further supports the linguistic connection of Elamite and Dravidic populations (Renfrew 1989).

Indians and western Eurasian populations share a deep late Pleistocene link through mtDNA haplogroup U which is the eldest group in western Eurasia and the second most frequent in both Indians and Europeans. Most of the Indian haplogroup U lineages (Ref. I; Table 2, Fig. 2) coalesce to a founder haplotype (U2) which dates back to around 53,000 years. This estimate falls to the same period when the European-specific U5 lineages started to diverge, around 52,000 years ago (Richards *et al*. 1998). Given the absence of U lineages in East Asians, the hypothesis of multiple dispersals, developed recently on archeological and paleoanthropological evidence (Lahr and Foley 1994), explains best the spread of both eastern and western Eurasian mtDNA lineages, inferring that the southern route migration brought the proto-Asian haplogroups (M and R) to India first and was then followed by a second late Pleistocene migration from the west carrying mostly haplogroup U. The first wave had to disperse quite rapidly further east to explain the lack of western Eurasian haplogroups east of India and the high frequency of East Asian-specific lineage clusters in India. Yet, the admixture of the two colonizing populations in India must have occurred quite early to explain, on the other hand, the lack of language specific mtDNA clusters in India.

Proto-Asian origins of Indians revealed through the structure of haplogroup M and caste affinities (Reference II)

Haplogroup M is the most frequent mtDNA haplogroup in Indian (Ref. II; Table 4) as well as among most Mongoloid populations. However, the phylogenetic structure of M in India (Ref II; Fig. 1) is distinct from that in other Asian populations and Ethiopians. Five major sub-clusters account for about a half of haplogroup M variation in Indians, and these clusters are different from those represented in other Asian populations studied so far (Ref. II; Fig. 2). The average distance of haplogroup M haplotypes from the founder motif puts the age of M in India to 48,000 years. This estimate suggests that Indian and East Asian haplogroup M lineages were split apart already early in the late Pleistocene and there has been only a limited gene flow between the populations thereafter.

In genetic distances calculated from mtDNA HVS-I data, Indian castes were found to be closer to eastern Asian populations, while in contrast Y chromosomal STR polymorphisms and autosomal markers revealed closer affinities of castes to Europeans than to eastern Asian populations (Ref. II; Tables 1-3). Indian specific lineages from sub-clusters M3 and U2 but also haplotypes closely related to western Eurasian counterparts from haplogroups H, J, K, and T, account mostly for the differences between caste ranks. Thus, although mtDNA analysis suggests largely proto-Asian origins of Indian castes, Y chromosomal and autosomal loci show that the process of caste formation could have been more complex, involving higher proportion of western Eurasian admixture in the paternal line. It is not possible yet to conclude whether the sex-specific differences were fixed in Indian populations during the initial settlement of the Indian subcontinent through two separate late Pleistocene migrations or whether the Neolithic and Bronze age migrations originating from the west were male-dominated.

The expansion nodes in Indian mtDNA variation (Reference III)

As Indian and eastern Asian lineages of haplogroup M coalesce in late Pleistocene, the deep time gap between the differentiation of Indian and eastern Asian populations is further manifested by the lack in India of other haplogroups specific to East Asian populations like A, B, and F. Yet haplogroups B and F together with western Eurasian haplogroups H, J, T, U, and V share a common ancestor: as a phylogenetic node R (Ref. III; Fig. 1) which is present in Indians with a number of Indian-specific descendant lineages (Ref. III; Fig. 8). The dating of the coalescence of these Indian specific derivatives of R node yielded an estimate of 55,000 years, which overlaps within error margins with the estimates of time depths of haplogroups M and U2 and suggests therefore that the first settlement of India by modern humans could have occurred at that time.

Besides the late Pleistocene overlap in the time depths of Indian lineages belonging to haplogroups M, U2 and R* and that lineages from haplogroups like H, J, K, and T can be seen in India to derive quite recently from their West Eurasian counterparts, there is yet another intermediate time period that is scheduled out by the ages of certain Indian-specific sub-clusters within haplogroup M and others. All 5 major sub-clusters of haplogroup M (Ref. III; Fig. 3) possess a starlike topology and their expansion phases range between 17,000-32,000 years (Ref. III; Table 4.). Lineage clusters U7, W (Ref. III; Fig. 7) and Ö (=HV in Macaulay *et al*. 1999b) that are infrequently spread in western Eurasia as well, reveal coalescence times in India that fall again into that period. Moreover, archeologically the time between 20,000 and 30,000 years is known as the period when Upper Palaeolithic artifacts are starting to spread all over the Indian peninsula (Joshi 1996). These mtDNA clusters could therefore correspond to another

demographic expansion in South Asia triggered either by climatic change or by the spread of a new Palaeolithic culture.

Inner nodes of the mitochondrial DNA tree in the Trans-Caucasus area (reference IV)
The first systematic attempts to classify European mtDNA diversity (Richards *et al*. 1996; Torroni *et al*. 1996) in the global context revealed that the majority of the lineages present in Europe today coalesce in Upper Palaeolithic and only a minor part can be assigned to a recent introduction by Neolithic Farmers. Whenever and no matter by how many distinct migrations Europe was colonized, the most likely route to be considered is from the south-east. While the European mtDNA pool has been extensively studied (Bertranpetit *et al*. 1995; Corte-Real *et al*. 1996; Francalacci *et al*. 1996; Lutz *et al*. 1998; Parson *et al*. 1998; Piercy *et al*. 1993; Pult *et al*. 1994; Richards *et al*. 1996; Sajantila *et al*. 1995; Torroni *et al*. 1996; Torroni *et al*. 1994a) there are only a few, and limited in size and geography, studies (Calafell *et al*. 1996; Comas *et al*. 1996; Di Rienzo and Wilson 1991; Macaulay *et al*. 1999b) on mtDNA diversity in the Near East, the Trans-Caucasus region and Anatolia. To further investigate the origins of mtDNA haplogroups spread in Europe we analyzed three populations from the Trans-Caucasus region - Armenians, speaking one of the earliest offshoots of the Indo-European language; Georgians, whose language belongs to Kartvelian, a Caucasian language family; and Ossetes who are supposed to be the direct descendants of an Indo-Iranian group called Alans.

The populations living in the Caucasus area are characterized, first, by the same basic haplogroups that are spread in Europe. Compared to Central Asian populations (Comas *et al*. 1998) who share a substantial part of their maternal lineages with East Asian Mongoloid populations, the Caucasians reveal only a few exceptional lineages which can be considered either East Asian or African in origin.

The second characteristic feature of the Caucasus populations is that the same lineage clusters (like T1, U3, J or others) that are comparatively young in Europe in their diversity, are all much more divergent and reveal traces of ancient expansion preceding the LGM.

Thirdly, an ancestral node Ö from which a large fraction of European lineages (H and V) descend from is represented at considerable frequency (~8%) and diversity in both Armenians and Georgians while being rare or absent in the European populations that we have studied. This node links populations from the Trans-Caucasus and Anatolia with Indians, among whom the node with specific derivative lineages also exists. This link gets further substantiated by the presence of several distinct clusters of haplogroup U (like U7) that are rarely sampled in Europe.

These findings are consistent with the scenario that the western Eurasian mtDNA gene pool as a whole, including populations from Europe, Trans-Caucasus and Anatolia has been generally the same in haplogroup structure since the Upper Palaeolithic with very limited impact of eastern and African lineages, and that the birthplace of most European lineage groups lies outside Europe, most probably between the Caucasus/Anatolia and North-West India.

Conclusions

1. Populations inhabiting the Indian peninsula share in maternal descent lineage groups both with western and eastern Eurasian populations. All Indian mtDNA lineages coalesce ultimately to an African-specific lineage group L3. These findings support the scenario of replacement of any pre-existing hominid species in southern Asia by modern humans originating from Africa.
2. Coalescence calculations show that Indian and eastern Asian haplogroup M lineages share a ~50,000 year old common ancestor and the gene flow between Mongoloids and Indians thereafter has been very restricted.
3. The maternal link between Indian and West Eurasian populations goes primarily through a share of haplogroup U which is the second most frequent lineage cluster in both regions. Haplogroup U sub-groups in India and western Eurasia are different and date to a common ancestor around 50,000 years ago.
4. The presence in India of two distinct, eastern and western Eurasian specific lineage groups, dating back to late Pleistocene, suggests that there were at least two separate migration events to India roughly 50,000 years ago. The first wave followed probably the southern route and inhabited the Indian sub-continent to extend further to eastern Asia. The second migration, from which most of western Eurasian populations descend from, reached India also during late Pleistocene and admixed with the first one.
5. Less than ten per cent of the extant mtDNA lineages in India can be explained by gene flow from the west during the last 10,000 years. This low percentage suggests that the total impact of Neolithic, Bronze Age and more recent migrations on Indian maternal gene pool has been limited in scale.
6. Both Indian and Trans-Caucasian populations are characterized by generally higher diversity than European populations and their mtDNA pool contains lineages that stand in or derive from the ancestral nodes (R and HV) that are absent or rare in Europe. Thus it is highly suggestive that India, Trans-Caucasus and the regions between them were the birthplace of the mitochondrial DNA haplogroups which are now widely spread throughout Europe.

References

Ammerman, A. J., and Cavalli-Sforza, L. L. (1984). The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., et al. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290**:**457-65.

Aquadro, C. F., and Greenberg, B. D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* 103**:**287-312.

Aris-Brosou, S., and Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol* 13**:**494-504.

Armour, J. A., Anttinen, T., May, C. A., Vega, E. E., Sajantila, A., Kidd, J. R., Kidd, K. K., Bertranpetit, J., Paabo, S., and Jeffreys, A. J. (1996). Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* 13**:**154-60.

Awadalla, P., Eyre-Walker, A., and Smith, J. M. (1999). Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286**:**2524-5.

Ballinger, S. W., Schurr, T. G., Torroni, A., Gan, Y. Y., Hodge, J. A., Hassan, K., Chen, K. H., and Wallace, D. C. (1992). Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130**:**139-52.

Bamshad, M., Fraley, A. E., Crawford, M. H., Cann, R. L., Busi, B. R., Naidu, J. M., and Jorde, L. B. (1996). mtDNA variation in caste populations of Andhra Pradesh, India. *Hum Biol* 68**:**1-28.

Bamshad, M., Rao, B. B., Naidu, J. M., Prasad, B. V. R., Watkins, S., and Jorde, L. B. (1997). Response to Spurdle et al. *Human Biology* 69**:**432-435.

Bamshad, M. J., Watkins, W. S., Dixon, M. E., Jorde, L. B., Rao, B. B., Naidu, J. M., Prasad, B. V., Rasanayagam, A., and Hammer, M. F. (1998). Female gene flow stratifies Hindu castes. *Nature* 395**:**651-2.

Bandelt, H. J. (1994). Phylogenetic networks. *Verh. Naturwiss. Ver. Hamburg* 34**:**51-71.

Bandelt, H. J., and Forster, P. (1997). The myth of bumpy hunter-gatherer mismatch distributions. *Am J Hum Genet* 61**:**980-3.

Bandelt, H. J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16**:**37-48.

Bandelt, H. J., Forster, P., Sykes, B. C., and Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141**:**743-53.

Barbujani, G., Bertorelle, G., and Chikhi, L. (1998). Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* 62**:**488-92.

Barnabas, S., Apte, R. V., and Suresh, C. G. (1996). Ancestry and interrelationships of the Indians and their relationship with other world populations: a study based on mitochondrial DNA polymorphisms. *Ann Hum Genet* 60**:**409-22.

Bendall, K. E., Macaulay, V. A., Baker, J. R., and Sykes, B. C. (1996). Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* 59**:**1276-87.

Bertranpetit, J., Calafell, F., Comas, D., Perez-Lezaun, A., and Mateu, E. (1996). Mitochondrial DNA sequences in Europe: an insight into population history. In: A. J. Boyce, and C. G. N. Mascie-Taylor (ed.), *Molecular biology and human diversity*. Cambridge University Press, Cambridge.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P., and Comas, D. (1995). Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* 59**:**63-81.

Bhattacharyya, N. P., Basu, P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S., and Majumder, P. P. (1999). Negligible male gene flow across ethnic boundaries in

India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res* 9**:**711-9.

Blanc, H., Chen, K. H., D'Amore, M. A., and Wallace, D. C. (1983). Amino acid change associated with the major polymorphic Hinc II site of Oriental and Caucasian mitochondrial DNAs. *Am J Hum Genet* 35**:**167-76.

Bonatto, S. L., Redd, A. J., Salzano, F. M., and Stoneking, M. (1996). Lack of ancient Polynesian-Amerindian contact. *Am J Hum Genet* 59**:**253-8.

Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K., and Cavalli-Sforza, L. L. (1991). Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci U S A* 88**:**839-43.

Brega, A., Gardella, R., Semino, O., Morpurgo, G., Astaldi Ricotti, G. B., Wallace, D. C., and Santachiara Benerecetti, A. S. (1986). Genetic studies on the Tharu population of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA. *Am J Hum Genet* 39**:**502-12.

Brown, M. D., Hosseini, S. H., Torroni, A., Bandelt, H. J., Allen, J. C., Schurr, T. G., Scozzari, R., Cruciani, F., and Wallace, D. C. (1998). mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet* 63**:**1852-61.

Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci U S A* 77**:**3605-9.

Brown, W. M., George, M., Jr., and Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76**:**1967-71.

Calafell, F., Underhill, P., Tolun, A., Angelicheva, D., and Kalaydjieva, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60**:**35-49.

Cann, R. L. (1994). mtDNA and Native Americans: a Southern perspective. *Am J Hum Genet* 55**:**7-11.

Cann, R. L., Brown, W. M., and Wilson, A. C. (1982). Evolution of human mitochondrial DNA: a preliminary report. *Prog Clin Biol Res* 103**:**157-65.

Cann, R. L., Stoneking, M., and Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature* 325**:**31-6.

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). The History and geography of human genes.Princeton University Press, Princeton.

Cavalli-Sforza, L. L., and Minch, E. (1997). Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61**:**247-54.

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., and Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci U S A* 85**:**6002-6.

Chen, Y. S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S., and Wallace, D. C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57**:**133-49.

Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V., and Barbujani, G. (1998). Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. *Proc Natl Acad Sci U S A* 95**:**9053-8.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., and Bertranpetit, J. (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13**:**1067-77.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., et al. (1998). Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 63**:**1824-38.

Corte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J., and Sykes, B. C. (1996). Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60**:**331-50.

Darlu, P., and Tassy, P. (1987). Disputed African origin of human populations. *Nature* 329**:**111-2.

Denaro, M., Blanc, H., Johnson, M. J., Chen, K. H., Wilmsen, E., Cavalli-Sforza, L. L., and Wallace, D. C. (1981). Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 78**:**5768-72.

Derenko, M. V., Dambueva, I. K., Malyarchuk, B. A., Dorzhu, C. M., and Zakharov, I. A. (1999). Structure and diversity of the mitochondrial gene pool in indigenous populations of Tuva and Buryatia based on restriction polymorphism data. *Genetika* 35**:**1706-1712.

Di Rienzo, A., and Wilson, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci U S A* 88**:**1597-601.

Excoffier, L. (1990). Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J Mol Evol* 30**:**125-39.

Excoffier, L., and Langaney, A. (1989). Origin and differentiation of human mitochondrial DNA. *Am J Hum Genet* 44**:**73-85.

Excoffier, L., and Yang, Z. (1999). Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. *Mol Biol Evol* 16**:**1357-68.

Eyre-Walker, A., Smith, N. H., and Smith, J. M. (1999). How clonal are human mitochondria? *Proc R Soc Lond B Biol Sci* 266**:**477-83.

Felsenstein, J. (1988). Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet* 22**:**521-65.

Fitch, W. (1977). On the problem of discovering the most parsimonious tree. *Am. Nat.* 111**:**1169-1175.

Foley, R. (1998). The context of human genetic evolution. *Genome Res* 8**:**339-47.

Foran, D. R., Hixson, J. E., and Brown, W. M. (1988). Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nucleic Acids Res* 16**:**5841-61.

Forster, P., Harding, R., Torroni, A., and Bandelt, H. J. (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59**:**935-45.

Francalacci, P., Bertranpetit, J., Calafell, F., and Underhill, P. A. (1996). Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* 100**:**443-60.

Giles, R. E., Blanc, H., Cann, H. M., and Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77**:**6715-9.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A., and Excoffier, L. (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Mol Biol Evol* 12**:**334-45.

Gyllensten, U., Wharton, D., Josefsson, A., and Wilson, A. C. (1991). Paternal inheritance of mitochondrial DNA in mice. *Nature* 352**:**255-7.

Hagelberg, E., Goldman, N., Lio, P., Whelan, S., Schiefenhovel, W., Clegg, J. B., and Bowden, D. K. (1999). Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc R Soc Lond B Biol Sci* 266**:**485-92.

Harihara, S., Saitou, N., Hirai, M., Gojobori, T., Park, K. S., Misawa, S., Ellepola, S. B., Ishida, T., and Omoto, K. (1988). Mitochondrial DNA polymorphism among five Asian populations. *Am J Hum Genet* 43**:**134-43.

Harpending, H., Sherry, S., Rogers, A., and Stoneking, M. (1993). The genetic structure of ancient human populations. *Current Anthropology* 34**:**483-496.

Hasegawa, M., Di Rienzo, A., Kocher, T. D., and Wilson, A. C. (1993). Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37**:**347-54.
Hedges, S. B., Kumar, S., Tamura, K., and Stoneking, M. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science* 255**:**737-9.
Helgason, A., Sigurdadottir, S., Gulcher, J., Ward, R., and Stefanson, K. (2000). mtDNA and the origins of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 66: 999-1016.
Hertzberg, M., Mickleson, K. N., Serjeantson, S. W., Prior, J. F., and Trent, R. J. (1989). An Asian-specific 9-bp deletion of mitochondrial DNA is frequently found in Polynesians. *Am J Hum Genet* 44**:**504-10.
Hofmann, S., Jaksch, M., Bezold, R., Mertens, S., Aholt, S., Paprotta, A., and Gerbitz, K. D. (1997). Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* 6**:**1835-46.
Hopkin, K. (1999). Death to sperm mitochondria. *Sci Am* 280**:**21.
Horai, S., Gojobori, T., and Matsunaga, E. (1984). Mitochondrial DNA polymorphism in Japanese. I. Analysis with restriction enzymes of six base pair recognition. *Hum Genet* 68**:**324-32.
Horai, S., Hayasaka, K., Kondo, R., Tsugane, K., and Takahata, N. (1995). Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc Natl Acad Sci U S A* 92**:**532-6.
Horai, S., and Matsunaga, E. (1986). Mitochondrial DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. *Hum Genet* 72**:**105-17.
Horai, S., Murayama, K., Hayasaka, K., Matsubayashi, S., Hattori, Y., Fucharoen, G., Harihara, S., Park, K. S., Omoto, K., and Pan, I. H. (1996). mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet* 59**:**579-90.
Howell, N., Kubacka, I., and Mackey, D. A. (1996). How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* 59**:**501-9.
Jazin, E., Soodyall, H., Jalonen, P., Lindholm, E., Stoneking, M., and Gyllensten, U. (1998). Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat Genet* 18**:**109-10.
Jin, L., Underhill, P. A., Doctor, V., Davis, R. W., Shen, P., Cavalli-Sforza, L. L., and Oefner, P. J. (1999). Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proc Natl Acad Sci U S A* 96**:**3796-800.
Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C., and Cavalli-Sforza, L. L. (1983). Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 19**:**255-71.
Jorde, L. B., Bamshad, M., and Rogers, A. R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20**:**126-36.
Jorde, L. B., Watkins, W. S., Bamshad, M., Dixon, M. E., Ricker, C. E., Seielstad, M. T., and Batzer, M. A. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosomal data. *Am J Hum Genet* 66.
Joshi, R. V. (1996). SOUTH ASIA in the period of *Homo sapiens neanderthalensis* and contemporaries (Middle Palaeolithic). In: *History of Humanity*. UNESCO, pp. 162-164.
Kaessmann, H., Heissig, F., von Haeseler, A., and Paabo, S. (1999). DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat Genet* 22**:**78-81.
Kaneda, H., Hayashi, J., Takahama, S., Taya, C., Lindahl, K. F., and Yonekawa, H. (1995). Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse

embryogenesis. *Proc Natl Acad Sci U S A* 92**:**4542-6.
Kolman, C., Sambuughin, N., and Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142**:**1321-34.
Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., and Paabo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell* 90**:**19-30.
Lahermo, P., Sajantila, A., Sistonen, P., Lukka, M., Aula, P., Peltonen, L., and Savontaus, M. L. (1996). The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 58**:**1309-22.
Lahr, M., and Foley, R. (1994). Multiple dispersals and modern human origins. *Evolutionary Anthropology* 3**:**48-60.
Lee, S., Shin, C., Kim, K., Lee, Y., and Lee, J. (1997). Sequence variation of mitochondrial DNA control region in Koreans. *Forensic Sci Int* 87**:**99-116.
Li, W. (1997). Molecular Evolution. In: (ed.), Sinauer Associates, Inc., Sunderland, pp. .
Lum, J. K., Cann, R. L., Martinson, J. J., and Jorde, L. B. (1998). Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. *Am J Hum Genet* 63**:**613-24.
Lutz, S., Weisser, H. J., Heizmann, J., and Pollak, S. (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111**:**67-77.
Lynch, M. (1996). Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol Biol Evol* 13**:**209-20.
Macaulay, V., Richards, M., and Sykes, B. (1999a). Mitochondrial DNA recombination - no need to panic. *Proc R Soc Lond B Biol Sci* 266**:**2037-2039.
Macaulay, V. A., Richards, M. B., Forster, P., Bendall, K. E., Watson, E., Sykes, B., and Bandelt, H. J. (1997). mtDNA mutation rates--no need to panic. *Am J Hum Genet* 61**:**983-90.
Macaulay, V. A., Richards, M. B., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonné-Tamir, B., Sykes, B., and Torroni, A. (1999b). The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. *Am J Hum Genet* 64**:**232-49.
Maddison, D. (1991). African origin of human mitochondrial DNA reexamined. *Syst Zool* 40**:**355-363.
Majumder, P. (1990). Anthropometric variation in India: A statistical Appraisal. *Current Anthropology* 31**:**94-103.
Majumder, P. P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B., Mukherjee, N., Roy, M., Thakurta, P. G., and Sil, S. K. (1999). Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur J Hum Genet* 7**:**435-46.
Mateu, E., Comas, D., Calafell, F., Perez-Lezaun, A., Abade, A., and Bertranpetit, J. (1997). A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tome, Gulf of Guinea. *Ann Hum Genet* 61**:**507-18.
Melton, T., Peterson, R., Redd, A. J., Saha, N., Sofro, A. S., Martinson, J., and Stoneking, M. (1995). Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* 57**:**403-14.
Menozzi, P., Piazza, A., and Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science* 201**:**786-92.
Merriwether, D. A., Clark, A. G., Ballinger, S. W., Schurr, T. G., Soodyall, H., Jenkins, T., Sherry, S. T., and Wallace, D. C. (1991). The structure of human mitochondrial DNA variation. *J Mol Evol* 33**:**543-55.

Michaels, G. S., Hauswirth, W. W., and Laipis, P. J. (1982). Mitochondrial DNA copy number in bovine oocytes and somatic cells. *Dev Biol* 94**:**246-51.
Mountain, J. L., Hebert, J. M., Bhattacharyya, S., Underhill, P. A., Ottolenghi, C., Gadgil, M., and Cavalli-Sforza, L. L. (1995). Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 56**:**979-92.
Mourant, A., Kopec, A., and Domaniewska-Sobczak, K. (1976). The distribution of the human blood groups and other polymorphisms. In: (ed.), Oxford monograph on medical genetics. Oxford University Press, London.
Nachman, M. W. (1998). Deleterious mutations in animal mitochondrial DNA. *Genetica* 103**:**61-9.
Nei, M., and Roychoudhury, A. (1982). Genetic relationship and evolution of human races. *Evol Biol* 14**:**1-59.
Nei, M., and Takezaki, N. (1996). The root of the phylogenetic tree of human populations. *Mol Biol Evol* 13**:**170-7.
Nielsen, R., and Weinreich, D. M. (1999). The age of nonsynonymous and synonymous mutations in animal mtDNA and implications for the mildly deleterious theory. *Genetics* 153**:**497-506.
Ohno, K., Tanaka, M., Suzuki, H., Ohbayashi, T., Ikebe, S., Ino, H., Kumar, S., Takahashi, A., and Ozawa, T. (1991). Identification of a possible control element, Mt5, in the major noncoding region of mitochondrial DNA by intraspecific nucleotide conservation. *Biochem Int* 24**:**263-72.
Olivo, P. D., Van de Walle, M. J., Laipis, P. J., and Hauswirth, W. W. (1983). Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 306**:**400-2.
Ozawa, T. (1995). Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. *Biochim Biophys Acta* 1271**:**177-89.
Paabo, S. (1996). Mutational hot spots in the mitochondrial microcosm. *Am J Hum Genet* 59**:**493-6.
Paabo, S. (1999). Human evolution. *Trends Cell Biol* 9**:**M13-6.
Papiha, S. S. (1996). Genetic variation in India. *Hum Biol* 68**:**607-28.
Parson, W., Parsons, T. J., Scheithauer, R., and Holland, M. M. (1998). Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 111**:**124-32.
Parsons, T. J., Muniec, D. S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M. R., Berry, D. L., Holland, K. A., Weedn, V. W., Gill, P., et al. (1997). A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15**:**363-8.
Passarino, G., Semino, O., Bernini, L. F., and Santachiara-Benerecetti, A. S. (1996a). Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* 59**:**927-34.
Passarino, G., Semino, O., Modiano, G., Bernini, L. F., and Santachiara Benerecetti, A. S. (1996b). mtDNA provides the first known marker distinguishing proto-Indians from the other Caucasoids; it probably predates the diversification between Indians and Orientals. *Ann Hum Biol* 23**:**121-6.
Passarino, G., Semino, O., Modiano, G., and Santachiara-Benerecetti, A. S. (1993). COII/tRNA(Lys) intergenic 9-bp deletion and other mtDNA markers clearly reveal that the Tharus (southern Nepal) have Oriental affinities. *Am J Hum Genet* 53**:**609-18.
Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M., and Santachiara-Benerecetti, A. S. (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62**:**420-34.
Penny, D., Steel, M., Waddell, P. J., and Hendy, M. D. (1995). Improved analyses of human

mtDNA sequences support a recent African origin for Homo sapiens. *Mol Biol Evol* 12**:**863-82.
Piercy, R., Sullivan, K. M., Benson, N., and Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106**:**85-90.
Poliakov, L. (1974). The Aryan Myth. In: (ed.), Basic Books, **New York**, pp. 190.
Pult, I., Sajantila, A., Simanainen, J., Georgiev, O., Schaffner, W., and Paabo, S. (1994). Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol Chem Hoppe Seyler* 375**:**837-40.
Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K., and Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* 23**:**437-41.
Rando, J. C., Pinto, F., Gonzalez, A. M., Hernandez, M., Larruga, J. M., Cabrera, V. M., and Bandelt, H. J. (1998). Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* 62**:**531-50.
Redd, A. J., and Stoneking, M. (1999). Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. *Am J Hum Genet* 65**:**808-28.
Redd, A. J., Takezaki, N., Sherry, S. T., McGarvey, S. T., Sofro, A. S., and Stoneking, M. (1995). Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12**:**604-15.
Relethford, J. H. (1998). Mitochondrial DNA and ancient population growth. *Am J Phys Anthropol* 105**:**1-7.
Relethford, J. H., and Jorde, L. B. (1999). Genetic evidence for larger African population size during recent human evolution. *Am J Phys Anthropol* 108**:**251-60.
Renfrew, C. (1989). The origins of Indo-European languages. *Sci. Am.* 261**:**82:90.
Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. J., and Sykes, B. (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59**:**185-203.
Richards, M., Macaulay, V., Sykes, B., Pettit, P., Forster, P., Hedges, R., and Bandelt, H. J. (1997). Reply to Cavalli-Sforza and Minch. *Am J Hum Genet* 61**:**251-254.
Richards, M. B., Macaulay, V. A., Bandelt, H. J., and Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62**:**241-60.
Risley, H. (1903). Census of India. In: (ed.), India. Ethnographic appendices, Calcutta, pp. .
Rogers, A. R., and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9**:**552-69.
Rogers, A. R., and Jorde, L. B. (1995). Genetic evidence on modern human origins. *Hum Biol* 67**:**1-36.
Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4**:**406-25.
Saitou, N., and Omoto, K. (1987). Time and place of human origins from mt DNA data. *Nature* 327**:**288.
Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M. L., Aula, P., Beckman, L., Tranebjaerg, L., Gedde-Dahl, T., et al. (1995). Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 5**:**42-52.
Schurr, T. G., Sukernik, R. I., Starikovskaya, Y. B., and Wallace, D. C. (1999). Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. *Am J Phys Anthropol* 108**:**1-39.
Scozzari, R., Torroni, A., Semino, O., Cruciani, F., Spedini, G., and Santachiara Benerecetti, S. A. (1994). Genetic studies in Cameroon: mitochondrial DNA polymorphisms in

Bamileke. *Hum Biol* 66**:**1-12.
Scozzari, R., Torroni, A., Semino, O., Sirugo, G., Brega, A., and Santachiara-Benerecetti, A. S. (1988). Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am J Hum Genet* 43**:**534-44.
Sherry, S. T., Batzer, M. A., and Harpending, H. C. (1998). Modelling the genetic architecture of modern populations. *Annual Rev. Anthrop.* 27**:**153-169.
Sherry, S. T., Rogers, A. R., Harpending, H., Soodyall, H., Jenkins, T., and Stoneking, M. (1994). Mismatch distributions of mtDNA reveal recent human population expansions. *Hum Biol* 66**:**761-75.
Singh, K. S. (1997). The Scheduled Tribes. In: K. S. Singh (ed.), *People of India*. Oxford University Press, Oxford, pp. 1266.
Smith, D. G., Malhi, R. S., Eshleman, J., Lorenz, J. G., and Kaestle, F. A. (1999). Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* 110**:**271-84.
Smith, J. M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet Res* 23**:**23-35.
Soodyall, H., and Jenkins, T. (1992a). Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. *Ann Hum Genet* 56**:**315-24.
Soodyall, H., and Jenkins, T. (1992b). Mitochondrial DNA studies in the South African Indian population. *Gene Geogr* 6**:**127-37.
Soodyall, H., and Jenkins, T. (1993). Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo. *Ann Hum Biol* 20**:**477-85.
Soodyall, H., Jenkins, T., Mukherjee, A., du Toit, E., Roberts, D. F., and Stoneking, M. (1997). The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. *Am J Phys Anthropol* 104**:**157-66.
Soodyall, H., Vigilant, L., Hill, A. V., Stoneking, M., and Jenkins, T. (1996). mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 58**:**595-608.
Starikovskaya, Y. B., Sukernik, R. I., Schurr, T. G., Kogelnik, A. M., and Wallace, D. C. (1998). mtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of Ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63**:**1473-91.
Stoneking, M. (1994). Mitochondrial DNA and human evolution. *J Bioenerg Biomembr* 26**:**251-9.
Stoneking, M., Jorde, L. B., Bhatia, K., and Wilson, A. C. (1990). Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124**:**717-33.
Stoneking, M., Sherry, S. T., Redd, A. J., and Vigilant, L. (1992). New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil. Trans. Royal Soc.* 337**:**167-175.
Stringer, C. B., and Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science* 239**:**1263-8.
Swofford, D. L. (1993). PAUP: Phylogenetic Analysis Using Parsimony. In: (ed.), Illinois Natural History Survey, Champaign, pp. .
Sykes, B. (1999). The molecular genetics of European ancestry. *Philos Trans R Soc Lond B Biol Sci* 354**:**131-8; discussion 138-9.
Sykes, B., Leiboff, A., Low-Beer, J., Tetzner, S., and Richards, M. (1995). The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet* 57**:**1463-75.
Templeton, A. R. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science* 255**:**737.

Thapar, B. K., and Rahman, A. (1996). The post-Indus cultures. In: A. H. Dani, and M. J.-P. (ed.), History of Humanity. Clays Ltd., St. Ives plc., UK.

Thorne, A. G., and Wolpoff, M. H. (1992). The multiregional evolution of humans. *Sci Am* 266**:**76-9, 82-3.

Tishkoff, S. A., Dietzsch, E., Speed, W., Pakstis, A. J., Kidd, J. R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A. S., Moral, P., and Krings, M. (1996). Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271**:**1380-7.

Torroni, A., Cruciani, F., Rengo, C., Sellitto, D., Lopez-Bigas, N., Rabionet, R., Govea, N., Lopez De Munain, A., Sarduy, M., Romero, L., et al. (1999). The A1555G mutation in the 12S rRNA gene of human mtDNA: recurrent origins and founder events in families affected by sensorineural deafness. *Am J Hum Genet* 65**:**1349-58.

Torroni, A., Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonne-Tamir, B., et al. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62**:**1137-52.

Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L., and Wallace, D. C. (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144**:**1835-50.

Torroni, A., Lott, M. T., Cabell, M. F., Chen, Y. S., Lavergne, L., and Wallace, D. C. (1994a). mtDNA and the origin of Caucasians: identification of ancient Caucasian- specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* 55**:**760-76.

Torroni, A., Miller, J. A., Moore, L. G., Zamudio, S., Zhuang, J., Droma, T., and Wallace, D. C. (1994b). Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 93**:**189-99.

Torroni, A., Neel, J. V., Barrantes, R., Schurr, T. G., and Wallace, D. C. (1994c). Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci U S A* 91**:**1158-62.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., et al. (1997). Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. *Am J Hum Genet* 60**:**1107-21.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M., and Wallace, D. C. (1993a). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53**:**563-90.

Torroni, A., Schurr, T. G., Yang, C. C., Szathmary, E. J., Williams, R. C., Schanfield, M. S., Troup, G. A., Knowler, W. C., Lawrence, D. N., Weiss, K. M., et al. (1992). Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130**:**153-62.

Torroni, A., Sukernik, R. I., Schurr, T. G., Starikorskaya, Y. B., Cabell, M. F., Crawford, M. H., Comuzzie, A. G., and Wallace, D. C. (1993b). mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* 53**:**591-608.

Torroni, A., and Wallace, D. C. (1995). MtDNA haplogroups in Native Americans. *Am J Hum Genet* 56**:**1234-8.

van Holst Pellekaan, S., Frommer, M., Sved, J., and Boettcher, B. (1997). Mitochondrial D-loop diversity in Australian riverine and Australian desert Aborigines. *Electrophoresis* 18**:**1538-43.

van Holst Pellekaan, S., Frommer, M., Sved, J., and Boettcher, B. (1998). Mitochondrial control-region sequence variation in aboriginal Australians. *Am J Hum Genet* 62**:**435-49.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., and Wilson, A. C. (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253**:**1503-7.
Villems, R., Adojaan, M., Kivisild, T., Metspalu, E., Parik, J., Pielberg, G., Rootsi, S., Tambets, K., and Tolk, H.-V. (1998). Reconstruction of maternal lineages of Finno-Ugric speaking people and some remarks on their paternal inheritance. In: K. Wiik, and K. Julku (ed.), *The roots of peoples and languages of Northern Eurasia* I. *Societas Historiae Fenno-Ugricae*, Turku.
Wakeley, J. (1993). Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* 37**:**613-23.
Wallace, D. (1995). Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet* 57**:**201-23.
Wallace, D. C., Brown, M. D., and Lott, M. T. (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238**:**211-30.
Ward, R. H., Frazier, B., Dew-Jager, K., and Paabo, S. (1991). Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci U S A* 88**:**8270-8274.
Watkins, W. S., Bamshad, M., Dixon, M. E., Bhaskara Rao, B., Naidu, J. M., Reddy, P. G., Prasad, B. V., Das, P. K., Reddy, P. C., Gai, P. B., et al. (1999). Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Am J Phys Anthropol* 109**:**147-58.
Watson, E., Bauer, K., Aman, R., Weiss, G., von Haeseler, A., and Paabo, S. (1996). mtDNA sequence diversity in Africa. *Am J Hum Genet* 59**:**437-44.
Watson, E., Forster, P., Richards, M., and Bandelt, H. J. (1997). Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61**:**691-704.
Whittam, T. S., Clark, A. G., Stoneking, M., Cann, R. L., and Wilson, A. C. (1986). Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. *Proc Natl Acad Sci U S A* 83**:**9611-5.
Wilkinson-Herbots, H. M., Richards, M. B., Forster, P., and Sykes, B. C. (1996). Site 73 in hypervariable region II of the human mitochondrial genome and the origin of European populations. *Ann Hum Genet* 60**:**499-508.
Wilson, A., Cann, R., Carr, S., George, M., Gyllensten, U., Helm-Bychowski, K., Higuchi, R., Palumbi, S., Prager, E., Sage, R., et al. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* 26**:**375-400.
Wolpoff, M. (1989). Multiregional evolution: the fossil alternative to Eden. In: P. Mellars , and C. Stringer (ed.), *The human revolution: Behavioral and biological perspectives on the origins of modern humans*. Edinburgh University Press, Edinburgh.
Yu, M. S., Qiu, X. F., Xue, J. G., Liu, Z. D., Tan, J. Z., Li, H. J., Liu, D. X., Li, L., Yu, W. Z., Tang, X. Z., et al. (1988). Mitochondrial DNA polymorphism in Chinese. *Sci Sin [B]* 31**:**860-72.
Zischler, H., Geisert, H., von Haeseler, A., and Paabo, S. (1995). A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* 378**:**489-92.
Zvelebil, M. (1986). Mesolithic prelude and Neolithic revolution. In: M. Zvelebil (ed.), *Hunters in transition: Mesolithic societies of temperate Eurasia and their transition to farming*. Cambridge University Press, Cambridge.

Kokkuvõte

Mitokondriaalne DNA pärandub emaliini pidi ning tänu oma suhteliselt suurele evolutsioneerumise kiirusele võimaldab tema populatsiooni-geneetiline kasutamine käsitleda küsimusi, mis on seotud kaasaegse inimese tekke aja ja koha ning edasise levikuga. India ning Taga-Kaukaasia rahvaste võrdlev analüüsimine teiste Euroopa ja Aasia populatsioonidega võimaldab kvalitatiivselt ja kvantitatiivselt hinnata hüpoteese kaasaegse inimese viimatise Aafrikast väljarände aja ning teekondade kohta.

Minu dissertatsiooni eksperimentaalne osa põhineb 300 hindu, 536 taga-kaukaaslase ja 424 ida-eurooplase mtDNA analüüsil. Enamuse India ning väiksema osa Taga-Kaukaasia ja Ida-Euroopa proovidega seotud eksperimentaalsest tööst teostasin ise. Igast DNA proovist amplifitseeriti mtDNA esimene hüpervarieeruv piirkond ning 360 aluspaari sellest sekveneeriti. Restriktsioon-analüüsiga klassifitseeriti iga proov vastavalt mtDNA kodeeriva osa polümorfismide jaotusele. Iga proovi järjestus joondati ning identsed järjestused koondati haplotüüpideks. Haplotüüpide vaheliste fülogeneetiliste seoste analüüsimiseks kasutasin mediaan-võrgustiku meetodit. Saadud võrgustikus iga klastri ning alam-klastri varieeruvuse hindamiseks kasutasin meetodit, milles arvutatakse iga järjestuse põhjal keskmine transitsiooniline kaugus klastri sügavaimast hargnemispunktist, oletatavast fülogeneetilisest eellas-haplotüübist.

Nii hindi kui draviidi keeli kõnelevate hindude mtDNA analüüs näitas, et sõltumata keelest, kastist või asualast on neile omased samad põhiklastrid, mis jaotuvad fülogeneetiliselt kahte superklaadi. Esimene superklaad, L3m (M), on haruldane Lääne-Euraasias, kuid esineb sagedasti kogu Aasias ning selle ida-aasialikud alamklastrid on levinud ka amerindidel. Samas aga koalestseeruvad India ning mongoloidsete populatioonide mtDNA liinid ühisele eellasele alles hilises pleistotseenis, ligikaudu 50 tuhat aastat tagasi. Leitud erinevused ja aeg lubavad oletada, et haplogrupp M levis Ida-Aasiasse läbi India või isegi tekkis seal, vahetult pärast hilist kaasaegse inimese Aafrikast väljarännet ning pärast seda on haplogrupp M liinid divergeerunud nii Indias kui mongoloidsetes populatsioonides omasoodu.

Antropoloogilised ning klassikalised geneetilised võrdlused kaasaegsete inimpopulatsioonide vahel loevad India poolsaare rahvaid sarnasteks Lääne-Euraasia rahvastega, ühendades neid koondnimetuse ‘kaukasoidid’ alla. Toetudes 19-nda ning 20-nda sajandi esimese poole filoloogilistele ning arheoloogilistele uurimustöödele on levinud hüpotees, et kaukasoidsed geenid levisid Indiasse eeskätt läbi aarialaste, kelle oletatav sissetung pärast Harappa kultuuri langust tõi Indiasse ka indo-euroopa keeled.

Mitokondriaalse DNA analüüs näitab vastupidiselt, et enamus hindusi evib liine, mis on proto-Aasia päritolu. Vaid 20.5% hindude emaliinidest kuulub haplogruppidesse, mis on levinud Lääne-Euraasias ning haruldased või puuduvad mongoloididel. Üle poole sellest ühisosast ehk 13.1% kõigist liinidest kuulub alamklastrisse U2 ning on oma mutatsioonimustrilt omased üksnes India populatsioonidele. India- ja Lääne-Euraasia-pärased U liinid koalestseeruvad jällegi hilises pleistotseenis ligikaudu 50 tuhat aastat tagasi ning lubavad oletada teist suuremat migratsiooni lainet Aafrikast, millest pärinevad peaaegu kõik eurooplaste mtDNA tüübid.

Vähem kui 10% India mtDNA liinidest sarnaneb lähedalt Lääne-Euraasias levinutega, kattudes nendega täpselt või erinedes ühe-kahe asenduse poolest. Katse dateerida selliste liinide kaudu hiljutist geenivoolu Indiasse andis neile keskmiseks eaks 9,000 aastat. Saadud dateering on tõenäoliselt paljude sõltumatute migratsioonide keskmine, kuid toetab pigem neoliitilise kui pronksiaegse migratsiooni osatähtsust. Seega pärineb mitokondriaalse DNA varieeruvus Indias

enam kui 90% osas paleoliitikumist ning oletatav indo-aarialaste sissetung, kui see ka aset leidis, ei jätnud olulist jälge India rahvaste emaliinidesse.

Mitokondriaalse DNA varieeruvus nii India kui Taga-Kaukaasia rahvastel on oluliselt suurem kui eurooplastel. Klastrid, mille ekspansiooniaeg Euroopas ulatub vaid viie kuni kümne tuhande aastani, on tagakaukaaslastel esindatud mitmekesiste liinidega, mille koalestsentsi ajad ulatuvad 20,000 aasta kaugusele. Samuti on nii Taga-Kaukaasia rahvastel kui hindudel levinud liine, mis divergeeruvad mitmete Euroopas levinud haplogruppide ühisest eellas-haplotüübist. Saadud tulemused näitavad, et Euroopas levinud mitokondriaalse DNA klastrite sünnipaik võis ulatuda Kaukaasiast ja Anatooliast Indiani.

Acknowledgments
First I would like to thank my supervisor Professor Richard Villems for leading me in and through this extremely interesting work. This work became essentially possible through the collaboration with several people who helped us to obtain the samples: I would thank Professor Surinder S. Papiha and Dr. Sarabjit S. Mastana for the expertise in collecting the highly valuable and large sample of Indian populations. I am also indebted to Dr. Arpita Pandya for the northern Indian collection. The Trans-Caucasian samples were obtained through expeditions of our Institute and I wish to thank cordially Dr. Armen Torosjan for his efforts in gathering Armenian, Georgian and Ossetian samples in most complicated local conditions. I acknowledge and wish to thank all the colleagues and fellow students in our lab for a friendly atmosphere and on-going discussions. Continuous discussions over the Net with Mike Bamshad and many colleagues from Europe were most useful and I learned a lot. For linguistic help I would like to express my gratitude to Triinu Pakk-Allmann. Still, all possible faults and lapses in this work are entirely mine. Finally, but not least, I would thank my family - my wife and children and my parents for caring.

Curriculum vitae

Date of birth:	11. august 1969, Tapa
Citizenship:	Estonian
Family status:	married, 3 children
Work address:	Institute of Molecular and Cell Biology Riia 23, 51010 Tartu
Telephone:	372 7 375 053
Fax:	372 7 420 286
e-mail:	tkivisil@ebc.ee
Home address:	Alasi 31-19, Tartu

Education and professional employment

1976-1987	Tartu 2nd Secondary School
1987-1992	Tartu University, Faculty of Biology and Geography, graduated *cum laude* as zoologist
1992-1994	Tartu University, Institute of Zoology and Hydrobiology, MSc student in zoology, title of thesis "Non-metrical cranial traits in Estonian mediaeval skeletal material"
1993-1998	teaching Human Anatomy for biologists in Institute of Zoology and Hydrobiology
1995	working and learning for 2 months in Göttingen University in Ullrich Grossbach's group of developementtal biology
1995-2000	PhD student in the Institute of Molecular and Cell Biology, Department of Evolutionary Biology. Research subject: human mitochondrial DNA variation.

Scientific work

Since 1995 I have been working in the group of Prof. Richard Villems. Main focus of my work has been human mtDNA variation in different populations from Europe, Asia and Africa.

Curriculum vitae

Sünniaeg:	11. august 1969, Tapa
Kodakondsus:	Eesti
Pererkonnaseis:	abielus, 3 last
Aadress tööl:	Eesti Biokeskus Riia 23, 51010 Tartu
Telefon:	372 7 375 053
Fax:	372 7 420 286
e-mail:	tkivisil@ebc.ee
Kodune aadress:	Alasi 31-19, Tartu

Haridus ja erialane teenistuskäik

1976-1987	Tartu 2. Keskkool
1987-1992	Tartu Ülikool, bioloogia-geograafiateaduskond, lõpetanud *cum laude*
1992-1994	TÜ ZHI magistrant, MSc zooloogia erialal teemal "Kolju mittemeetrilised tunnused ja nende sagedused Eesti keskaegses materjalis"
1993-1998	Inimese anatoomia praktikumide juhendamine ZHI-s
1995	kahekuuline täiendõpe Göttingeni Ülikoolis prof. Grossbachi laboris
1995-2000	TÜ MRI doktorant, evolutsioonilise bioloogia õppetooli juures

Teadustegevus

Alates 1995. aastast olen töötanud prof. Richard Villemsi töögrupis. Olen uurinud inimese mitokondriaalse DNA varieeruvust mitmetel Euroopa, Aasia ja Aafrika populatsioonidel.