In: **Genomic Diversity**. 1999 Edited by Deka, R. Papiha,
S.S.Kluwer/Academic/Plenum Publishers, pp. 135-152

The Place of the Indian mtDNA Variants in the Global Network of Maternal Lineages and the Peopling of the Old World

Toomas Kivisild[1], Katrin Kaldma[1], Mait Metspalu[1], Jüri Parik[1], Surinder Papiha[2], Richard Villems[1]

[1]Institute of Molecular and Cell Biology, Tartu University and Estonian Biocentre
[2]Department of Biochemistry and Genetics, University of Newcastle-upon-Tyne, UK

INTRODUCTION

Both archaeology and genetics suggest that modern humans originated 100,000 to 200,000 years ago in Africa (Cann *et al.* 1987; Stringer 1990). Their first skeletal remains outside Africa are about 100,000 years old but have been found so far only in the immediate vicinity - from the caves in Near East (McDermott *et al.* 1993; Stringer 1992; Stringer *et al.* 1989; Aitken and Valladas 1992). There is no substantial evidence supporting further spatial dispersal of modern humans earlier than around 50,000 years ago. By that time they seem to have reached Papua New Guinea and Australia and soon after that they are found also in Europe. What happened during this 50,000 year long gap and where did the initial radiation of the Eurasian population take place remains largely an open question.

Western Asia and India stand geographically on the road early modern humans had almost inevitably pass to reach eastern Asia, New Guinea and Australia. Did some of the migrating waves of humans settle there instead of going *in corpore* further eastwards? Was it the place where the initial radiation of Eurasian mitochondrial DNA lineages took place? There is a lack of skeletal evidence of modern humans from East Asia older than the Upper Cave Zhoukoudian crania (Foley 1998) that are dated to around 30,000 years before present (BP). The earliest skeletal evidence from South Asia comes from Sri Lanka, where the Fa Hien Lena finds put forward 34,000 year old carbon datings (Deraniyagala 1998). These archeological dates imply the approximate time line for the dispersal of modern humans in Asia.

Driven by fast mutation rate and lack of recombination, distinctive clusters of mtDNA lineages have emerged during the last tens of thousands of years. Low overall population density during Palaeolithic and vast geographic distances favoured the isolation of human populations and thus played an important role in secluding the differences arisen in DNA lineages. The present day mtDNA variability is highly continent-specific (Chen *et al.* 1995; Torroni *et al.* 1996; Wallace 1995). Therefore, already at the level of present day knowledge about the worldwide variation of mtDNA genome, one can reliably distinguish between mtDNAs of eastern Asian, European or sub-Saharan African origin (Fig. 1). Nevertheless, the same knowledge base shows that all mtDNA variants outside Africa, studied so far, derive from a single Pan-African mtDNA cluster L3a (Watson *et al.* 1997).

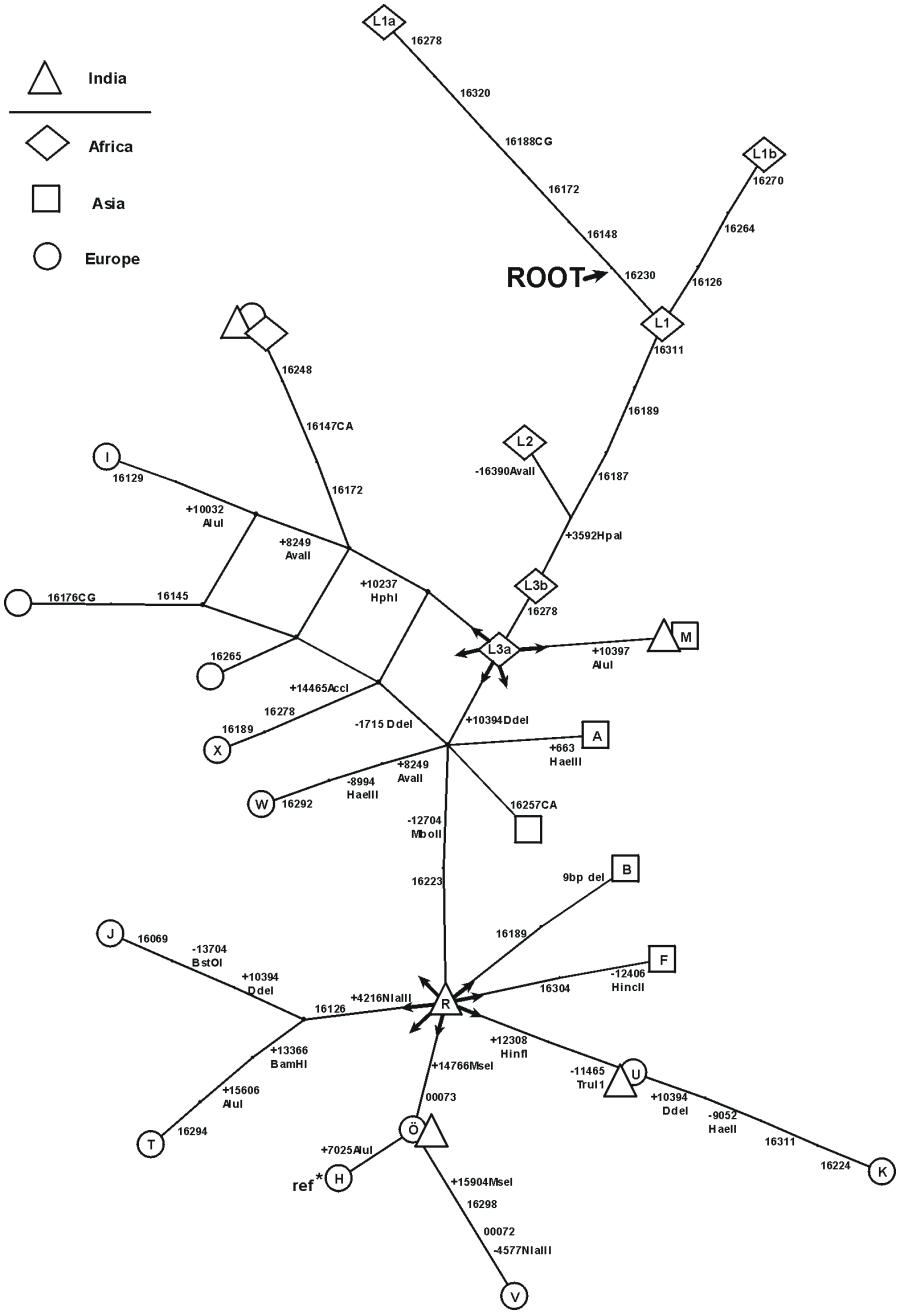

Figure 1. The skeleton network of mtDNA haplogroups and their geographic distribution. Haplogroup defining mutations are shown near lines connecting them, indicating base pair position relative to the Cambridge reference sequence (CRS; Anderson et al. 1981). Presence or absence of restriction enzyme recognition sites is shown also respective to the CRS, not the root, which is proposed using Neandertal sequence (Krings et al. 1997) as an outgroup. The definition and labelling of the haplogroups follows the scheme of earlier studies (Torroni et al. 1996; Richards et al. 1998; Watson et al. 1997; Macaulay et al. 1999). L3a is a super-haplogroup (indicated by arrows) encompassing Eurasian haplogroups. Another super-haplogroup is R, which on the other hand is a sub-cluster to L3a. Several so far undefined clusters are shown unlabelled while a new distinctive haplogroup Ö is defined inside a point previously labelled as HV* (Macaulay et al. 1999).

1. MITOCHONDRIAL DNA HAPLOGROUPS IN INDIA

Both eastern and western Eurasian mtDNA haplogroups have been found among modern Indian populations (Bamshad *et al.* 1997; Passarino *et al.* 1996; Passarino *et al.* 1996; Kivisild *et al.* manuscript in preparation; Bamshad *et al.* manuscript in preparation). Haplogroups M and U constitute the major portion (~75%) of the variation and pooled together, their spread is nearly uniform over India (Table 1). About 10% of the Indian mtDNAs belong to other haplogroups, geographically characteristic to Eurasia, while as many as 15% of the lineages do not belong to any of the continental clusters described before. The lack of African-specific variants in India should be considered with some precaution (Bamshad *et al.* 1997) because African-borne haplogroup L3 lacks any distinctive RFLP or HVS-I sequence motifs. However, the absence of *Hpa*I site at nucleotide position (np) 3592 among Indians seems clearly to negate any recent substatial gene flow from sub-Saharan Africa to the peninsula.

Comparing the frequencies of mtDNA haplogroups present among Indians with those among other Eurasian populations suggests (Table 2) that Indians are simultaneously part of two distinctive meta-populations: eastern and western Eurasians. In that respect Indians appear to be similar to the populations of Central Asia (Comas *et al.* 1998). This is not surprising when bearing in mind the geographical position of both regions. However, haplogroup frequency data alone are insufficient and can be sometimes misleading in choosing between alternative population genetic scenarios: whether the co-presence of these two sets of lineages is due to their initial diversification or because of a recent diffusion. To get deeper insight into the problem a detailed phylogenetic analysis of both eastern and western Eurasian lineages is needed.

Table 1. mtDNA haplogroup frequencies in India

	Pakistan	Kashmir	Uttar Pradesh	Punjab (Lobanas)	Andhra Pradesh (Lambadis)	Other Localities	Total
African							
L	0	0	0	0	0	0	0
e. Eurasian							
A	0	0	1 (1%)	0	0	0	1 (0.3%)
B	0	0	0	0	0	0	0
F	0	4 (21%)	2 (2%)	1 (1%)	1 (1%)	0	8 (2.7%)
M	3 (38%)	5 (26%)	52 (51%)	34 (55%)	55 (64%)	9 (41%)	158 (52.7%)
w. Eurasian							
H	1 (12%)	1 (5%)	3 (3%)	0	2 (2%)	0	7 (2.3%)
I	0	0	2 (2%)	0	2 (2%)	0	4 (1.3%)
J	0	1 (5%)	0	0	1 (1%)	0	2 (0.7%)
K	0	0	0	0	0	0	0
T	0	0	1 (1%)	3 (5%)	1 (1%)	0	5 (1.7%)
U	2 (25%)	2 (11%)	24 (24%)	3 (5%)	11 (13%)	4 (18%)	46 (15.3%)
V	0	0	0	0	0	0	0
W	0	1 (5%)	0	8 (13%)	2 (2%)	0	11 (3.7%)
X	0	1 (5%)	0	0	1 (1%)	0	2 (0.7%)
Ö	1 (12%)	3 (16%)	1 (1%)	0	0	0	5 (1.7%)
Others	1 (12%)	1 (5%)	17 (17%)	13 (21%)	10 (12%)	9 (41%)	50 (16.7%)
n	8	19	103	62	86	22	299

Table 2. mtDNA haplogroup frequencies in some Eurasian populations

	n	**WE**[1]	H	I	J	K	T	U	V	X	W	**EE**[2]	A	B	F	M[C,D,E,G]	M*	**AF**[3]	L1	L2	L3	**Others**
Indians[a]	299	**27.4**	2.3	1.3	0.7	0	1.7	15.3	0	0.7	3.7	**55.7**	0.3	0	2.7	1.7	51.0	**0**	0	0	0	**16.7**
Mongols[b]	103			0	0	1.0	1.0		0	0	0	**77.7**	3.9	9.7	5.8	38.9	19.4					
Tibetians[c]	54											**94.5**	11.1	5.6	14.8	33.4	29.6	**0**	0	0	0	**5.5**
Chinese[d]	66			0	0	0	0		0		0	**94.0**	6.0	20.0	22.0	22.0	24.0	**0**	0	0	0	**6**
Abor. Siberians[e]	259											**97.3**	23.6			49.8	23.9					**2.7**
Native Americans[f]	591											**95.9**	28.7	27.4		39.1		**0**	0	0	0	**4.1**
Central Asians[g]	205	**30.5**	14.0	1.0	2.5	0.5	3.5	8.0	0	0	1.0	**61.0**	6.8	6.8	5.4	29.2	9.3	**0**	0	0	0	**8.5**
Trans-Caucasians[a]	330	**80.9**	24.8	1.8	6.7	8.2	11.8	21.2	0	5.5	0.9	**1.5**	0	0.3	0	1.2	0	**0.3**	0	0	0.3	**17.3**
Italians[h]	99	**93.9**	33.3	4.0	7.1	8.1	9.1	22.2	5.1	3.0	2.0	**0**	0	0	0	0	0	**0**	0	0	0	**6.1**
Finno-Ugries[a]	149	**97.9**	45.6	1.4	12.1	3.3	6.0	22.8	2.0	2.0	2.7	**0.7**	0	0	0	0.7	0	**0**	0	0	0	**1.4**
Slavs[a]	324	**94.7**	41.4	2.8	10.5	3.7	12.3	19.4	3.1	0.6	0.9	**0.9**	0	0	0	0.9	0	**0**	0	0	0	**4.4**
Germans[i]	200	**92.5**	50	2.5	7.5	6.5	8.5	13.5	2.5	0.5	1.0	**0**	0	0	0	0	0	**0**	0	0	0	**7.5**
Ethiopians[j]	74	**13.5**	0	0	0	2.7	1.4	8.1	0	0	1.4	**20.3**						**52.7**				**13.5**
Sub-Saharan Africans[k]	407																		30.2	30.4	39.3	

[1] western Eurasian specific haplogroups

[2] eastern Eurasian specific haplogroups

[3] African specific haplogroups

[a] Kivisild *et al.*, manuscript in preparation

[b] Kolman *et al*. 1996

[c] Torroni *et al*. 1994

[d] deduced from sequence data of Horai et al. 1996

[e] Torroni *et al*. 1993

[f] Baillet *et al*. 1994; Torroni *et al*. 1992; Schurr *et al*. 1990

[g] deduced from sequence data of Comas *et al*. 1998

[h] Torroni *et al*. 1997

[i] deduced from sequence data of Lutz *et al*. 1998

[j] deduced from sequence data of Passarino *et al*. 1998

[k] Watson *et al*. 1997

1.1. Eastern Eurasian mtDNA haplogroups in India

Although the data about Asian populations are still far from being representative, it is already apparent that the structure of Asian mtDNA phylogeny is different from that of European and African populations (Wallace 1995). A common feature for all the Asian populations studied so far is that the major fraction of their mtDNA pool is made up of haplogroups A, B, F and M.

Haplogroup M is a dominant mtDNA cluster among the populations of Mainland Asia as well as among Native Americans (Ballinger *et al.* 1992; Torroni *et al.* 1994; Torroni *et al.* 1994). It is defined by the presence of a *Dde*I site at np 10394 and an *Alu*I site at np 10397. The co-occurrence of these two polymorphisms is highly Asian-specific but Asians lack at the same time any possible progenitor of this haplogroup. Thus, the mtDNA coding region RFLP diversity-based coalescence age of 55,500-73,000 years for haplogroup M (Chen *et al.* 1995) could indicate the time of the initial colonisation of Asia by modern humans.

Haplogroup M has been sub-divided into discrete sub-clusters according to the accumulation of further synapomorphic mutations along Asian maternal lineages. Sub-clusters C, D, E and G, defined by certain RFLP and HVS-I sequence polymorphisms (Torroni *et al.* 1994; Torroni *et al.* 1993; Torroni *et al.* 1993), are spread over vast territories all over Mainland Asia (Fig. 2). These separate sub-clusters display surprisingly similar, approximately 40,000 year old, coalescence times (Table 3). The split of haplogroup M into these subclusters may therefore reflect a secondary expansion event, which led ultimately to the extension of modern Asian populations to northern and central Asia and to the Americas. Besides these haplogroup M sub-clusters, haplogroups A and two basic sub-clusters of haplogroup B display comparable diversity and coalescence ages (Table 3).

Table 3. Coalescence times of eastern Eurasian mtDNA haplogroups in Mongols[1]

Haplogroup[2]	consensus	ρ[3]	time[3]		variance[3]
A	+663*Hae*III; 16223T; 16390T; 16,319G	1.75	35,500	±	13,000
B	9bp del; 16189C	2.5	50,500	±	10,000
B1	9bp del; 16189C; 16217C	2	40,500	±	11,000
B2	9bp del; 16189C; 16243C	1.7	33,500	±	15,000
F	(+12406*Hinc*II); 16304C	2.1	42,000	±	10,000
M	+10394*Dde*I;+10397*Alu*I;16223T	3.7	74,000	±	5,000
C	+10394*Dde*I;+10397*Alu*I;16223T; 16298C; 16327T	2.1	42,000	±	8,000
D	+10394*Dde*I;+10397*Alu*I;16223T; 16362C	2.2	44,500	±	6,000

[1] data from Kolman *et al*. 1996

[2] Haplogroups defined as in Torroni *et al*. 1994

[3] Calculations as in Forster *et al*. 1996

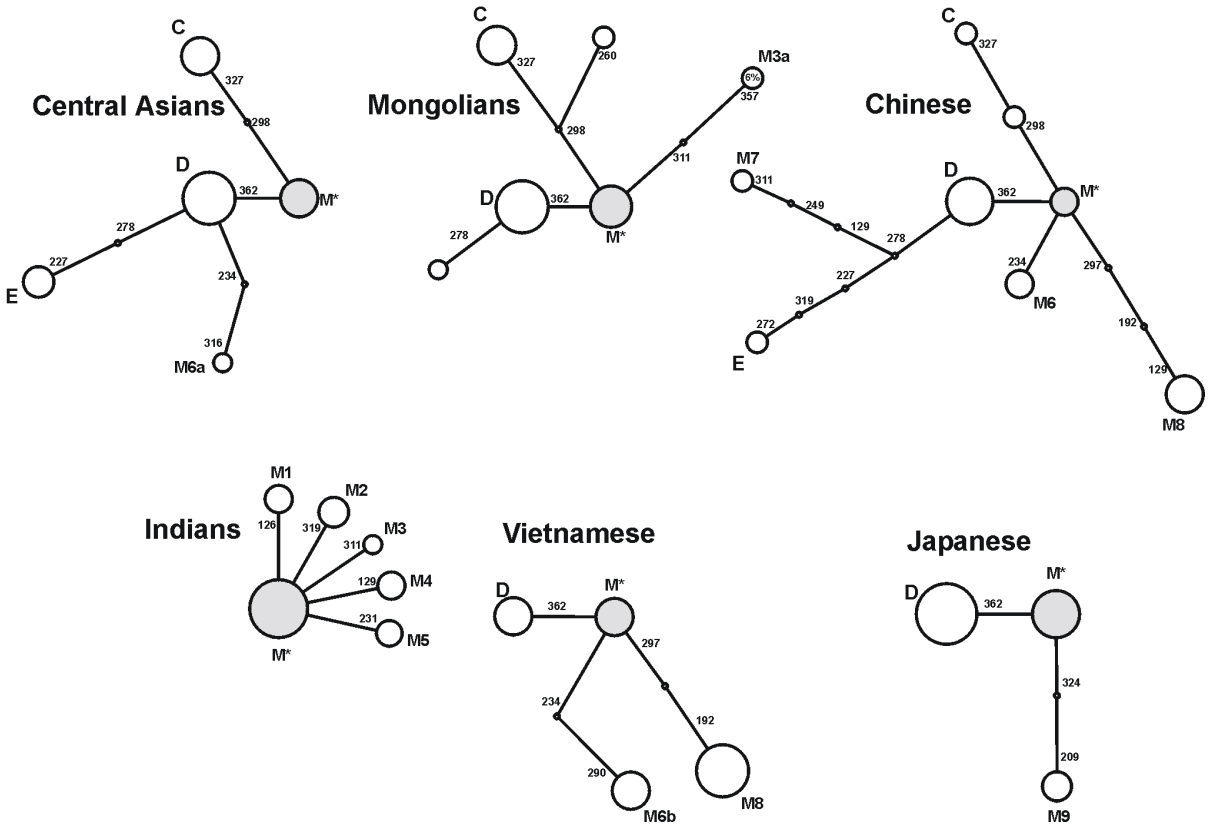

Figure 2. The structure of haplogroup M in some Asian populations. RFLP status of the sequences is known only for Indian (manuscript in preparation) and Mongolian (Kolman *et al.* 1996) sequences. Among Central Asian (Comas *et al.* 1998), Chinese, Japanese (Horai *et al.* 1996), and Vietnamese (Lum *et al.* 1998) populations the sequences were assumed to belong to haplogroup M according to indirect evidence – comparing them with sequences of known status. Only those haplogroup M sub-branches that constitute at least 5% of the total frequency of haplogroup M in the given population are shown. Shaded M* collects all the remaining branches that did not exceed the 5% criterion. The node area reflects the frequency of the sub-branch. Mutations at the given bp in HVS-I (less 16,000) are shown near lines connecting the nodes.

The sub-branching of haplogroup M in India is profoundly different from that described so far for any other Asian locality. Five novel sub-clusters can be defined and they form a half of the Indian haplogroup M lineages (Fig. 2, 3). The other half is scattered between multiple minor branches (Fig. 4). In contrast to that, typical 'Mongoloid' sub-clusters C, D, E and G are found at extremely low frequencies (1.2 % all together, Table 2). Even though, these exceptions are mostly due to the sampling of borderline populations, such as Tharus, whose Oriental origins have been characterised earlier (Brega *et al.* 1986; Passarino *et al.* 1993; Passarino *et al.* 1992). Likewise, the other typically eastern Asian haplogroups A, B and F, are practically absent in India, although B and F are very common in neighbouring southeastern populations of Asia (Wallace 1995). Studies based on mtDNA RFLP polymorphisms in India have shown that haplogroup M divergence predates the separation of proto-Indians from proto-eastern Asians (Passarino *et al.* 1996; Passarino *et al.* 1996). Hence, an important conclusion can be drawn: during the last 50,000 years or so, there has been a very limited admixture of Indian populations with the mongoloid populations living east and north of India. And the other way round: the five major Indian-specific sub-clusters of

haplogroup M are not represented to any significant extent elsewhere in Asia. Thus, it appears that although Indian populations display haplogroup M frequency similar to or even higher than that in other Asian populations, the internal structure of haplogroup M lineages in India reflects their basically autochthonous development.

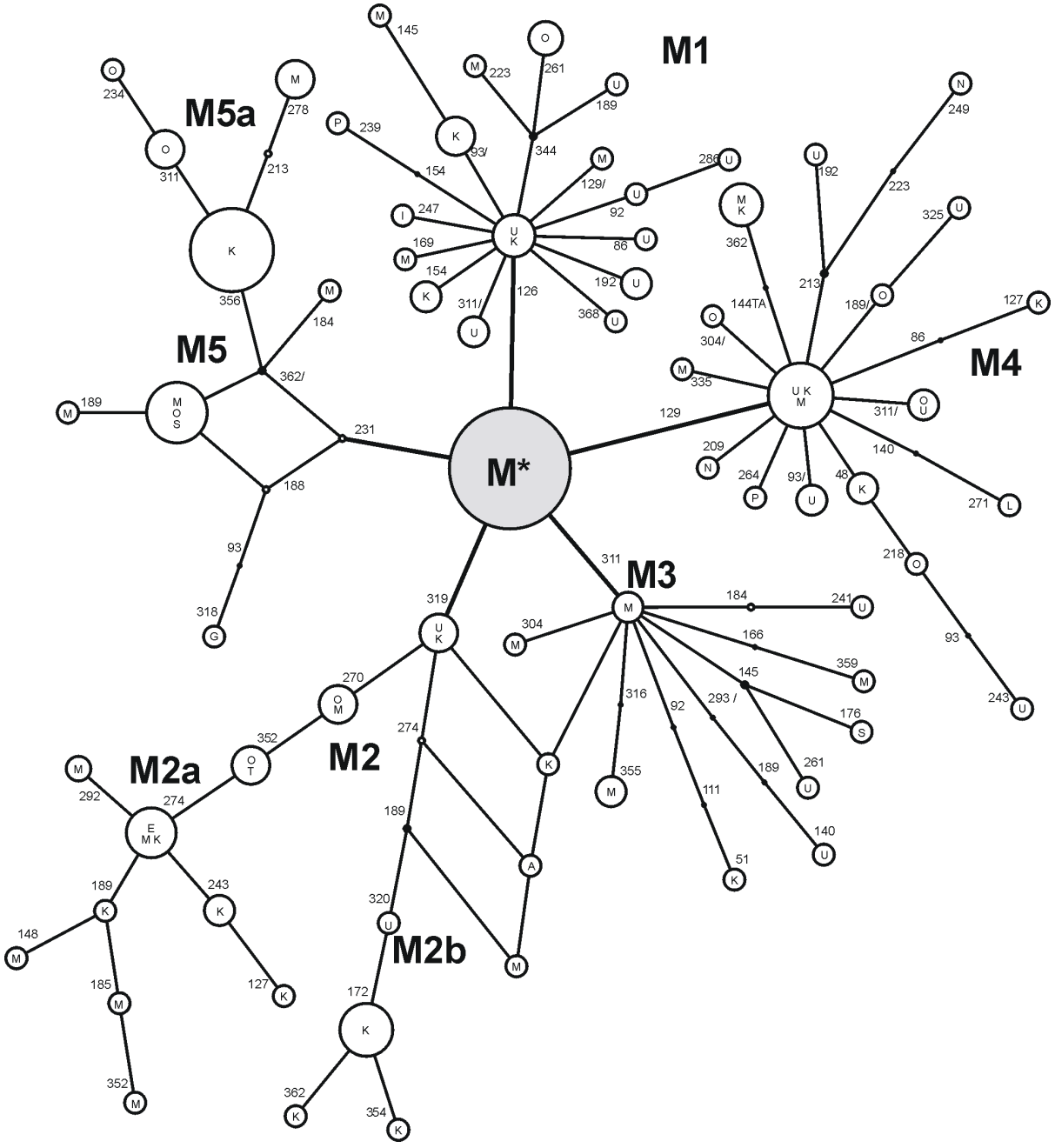

Figure 3. Reconstruction of 5 major haplogroup M subclusters in India. Node areas correspond to haplotype frequencies. Mutations at which two haplotypes differ are shown less 16,000 on lines connecting them. "/" indicates a broken reticulation. Central shaded node corresponds to the ancestral state of haplogroup M as in Fig. 1. U - Uttar Pradesh; O - Lobanas; M – Lambadis; S - Kashmir; K - Kerala; I - Bihar; N - Bangladesh; L - Bengal; A - Andhra Pradesh; P - Pakistan; G – Gujarat.

The spread of haplogroup M variants reveals some characteristic differences among different populations of the Indian Peninsula (Fig. 3). For example, sub-cluster M2 is predominantly found among populations living in the southern parts of the peninsula. Also, an interesting association of sub-cluster M1 with high caste Indians should be mentioned. Our search among the sequences from Kerala and Karnataka populations (Mountain *et al.* 1995) shows that only Brahmin caste Haviks, but not the lower caste Mukris, have a characteristic to M1 motif (transitions at nps 16,126 and 16,223). The distance between higher and lower caste Telugus was also found to be bigger than that between higher and middle or middle and lower castes; also no mtDNA haplotypes were shared exclusively between the upper and the lower castes (Bamshad *et al.* 1998). On the other hand, lineage clusters M3, M4, M5 and haplogroup M minor sub-clusters are spread over India without any evident population or caste-specific sequestration. The general intermixture of lineages of different caste groups at the level of major lineage clusters was shown by Mountain and colleagues (Mountain *et al.* 1995) by a HVR sequence inter-match analysis. These two findings can be explained by a common ancestry of distant caste groups of India, followed by a comparatively recent endogamy within the limits of caste system.

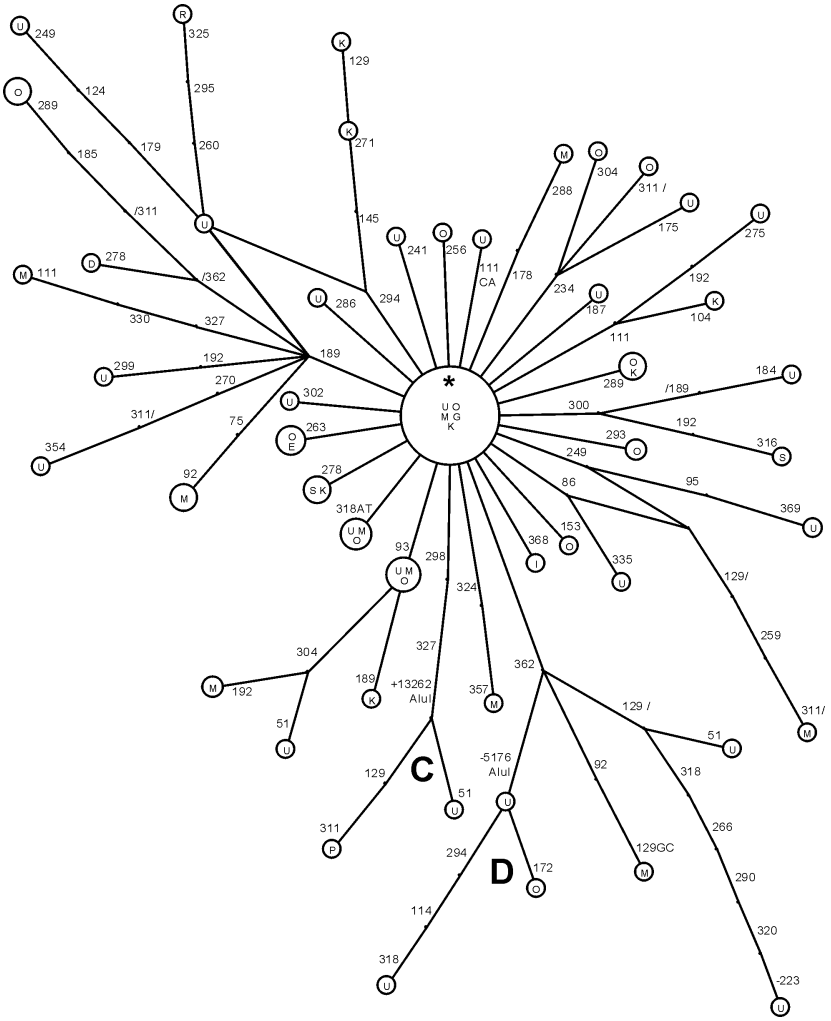

Figure 4. Reconstruction of the M* lineages in India. Populations: U - Uttar Pradesh; O - Lobanas; M - Lambadis; S - Kashmir; K - Kerala; I - Bihar; N – Bangladesh; L - Bengal; A - Andhra Pradesh; P - Pakistan; G - Gujerat; D - Tamil Nadu; R – Orissa. For other details, see Legend to Fig. 3.

We estimated the mean coalescence age of Indian M sequences as 47,000 ± 2,500 years (Tab. 4). This estimate is younger than the date of ~65,000 years proposed earlier (Mountain *et al.* 1995) for an expansion starting from South-Asia. However, the latter date derives from a 'non-African' cluster M2 (Fig. 5 in Mountain *et al.* 1995) that includes also African sequences and could therefore reflect (a possibility considered also by the authors) an expansion occuring in Africa just prior to the exodus. A support for this enigmatic cold and dry period expansion in eastern Africa (60,000-80,000 years ago) was given by a recent analysis of African mtDNAs (Watson *et al.* 1997). When we consider the possibility that more than just a single female walked out of Africa and gave enough progeny, any time calculation based on random sequences from a non-African population should reach the estimate of this initial radiation inside Africa. Therefore, the time for non-African expansions should be calculated from lineage clusters that show autochthonous development outside Africa. Although haplogroup M is predominantly found in Asian populations, a recent study showed also its presence in eastern Africa (Passarino *et al.* 1998). Within the Indian haplogroup M sequences averaging to the 47,000-year-old coalescence point, sub-cluster M2 and many long independent branches (Fig. 4) reveal divergence that exceeds that point. For example the coalescence age of all M2 sequences is 63,000 years. This estimate agrees with the expansion dates discussed above (Mountain *et al.* 1995; Watson *et al.* 1997) and also falls within the limits of the calculations of time depth based on the RFLP data of Asian M (Chen *et al*. 1995).

Table 4. The diversity and age of haplogroup M and its major subclusters in India

	n	ρ	T	±
M-total	158	2.33	47,000	2,500
sub-cluster				
M1	23	1.26	25,500	5,500
M2	34	3.12	63,000	6,000
M2a	13	0.85	17,000	5,000
M2b	11	1.1	22,000	6,500
M3	12	1.58	32,000	7,500
M4	35	0.91	18,500	3,500
M5	31	1.38	28,000	4,500

ρ, t, variance - estimated as in Forster *et al*. 1996

To conclude, two different scenarios can be drawn. First, the origin of haplogroup M lies in eastern Africa and multiple M lineages were present already at the onset of the population movement that finally colonised Asia. Second, the calculation of the initial expansion phase of haplogroup M may be hindered by cumulative demographic turbulences. Figure 5 shows intra- and inter-branch mismatch distribution patterns of five major sub-clusters of haplogroup M in India. Sub-clusters M1, M2a, M2b, M3-M5 all possess star-like topology (Fig. 3) and their expansion dates range from 17,000-32,000 years (Table 4). These star-like clusters thus reflect another influential demographic expansion in the history of Indian populations. In archeology it roughly coincides with the proposed transition from Middle to Upper Palaeolithic in India (Joshi 1996). The inter-match distance of these clusters on the other hand exceeds the mean value of the total M sequence pool. There is also a significant portion of very divergent haplogroup M sub-lineages (Fig. 4) that do not reveal star-like traces of this recent expansion phase. The summary haplogroup M mismatch distribution and

age derived from that could be therefore statistically elusive, affected by the sum of 'noise' from multiple founders and expansion phases. Below we use data from other mtDNA haplogroups to elucidate further the question when the first population expansion started in India.

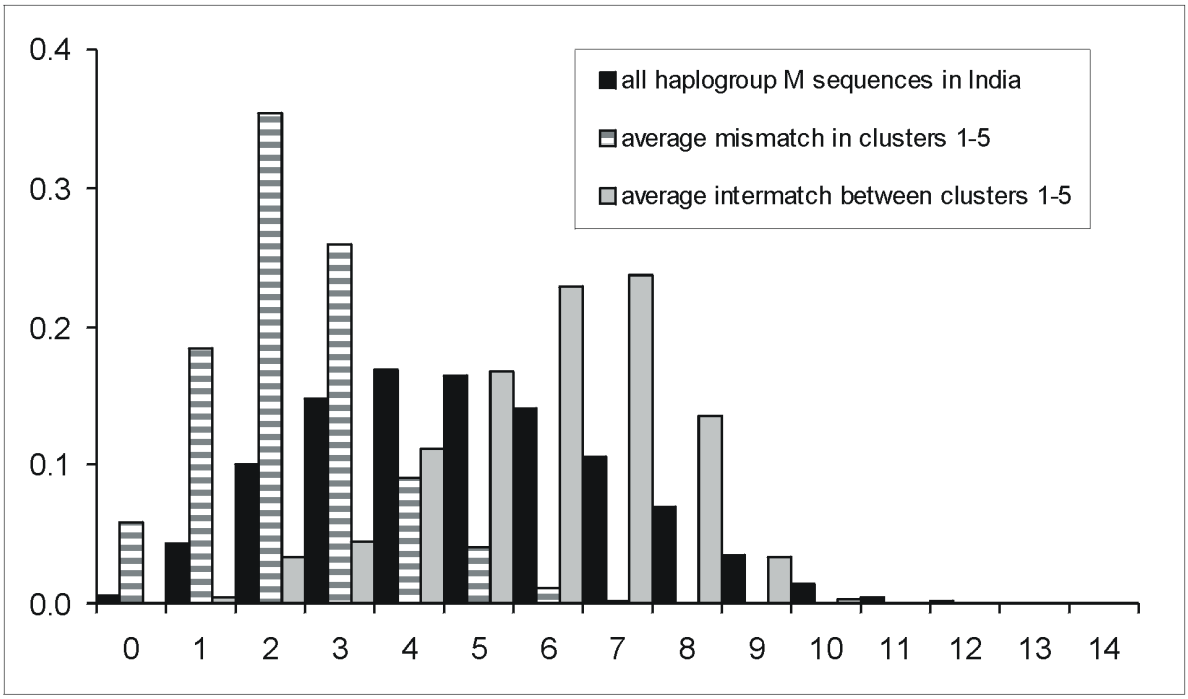

Figure 5. Mismatch distribution of haplogroup M in India. Mismatch and intermatch distribution of Indian-specific haplogroup M major subclusters 1-5 is shown alongside the mismatch distribution of all Indian M sequences.

1.2. mtDNA haplogroups in India which are considered specific for western Eurasia

Haplogroups H, I, J, K, T, U, V, W and X cover about 95% of the mtDNA variation in European populations (Torroni *et al.* 1996). Among them, haplogroup H is the most frequent, followed by U, T and J. Other minor haplogroups hardly ever exceed 5% in their frequency in Europe. Among Indian populations only haplogroup U displays frequency, comparable to that in European populations (Table 2). All other major western-Eurasian specific haplogroups are more than ten times less frequent. Haplogroup U is present also among some African populations (Torroni 1996, Passarino 1998). Hence, the possibility that the Indian haplogroup U varieties could be derived from a non-European gene pool needs to be considered. Figure 6 represents a phylogenetical reconstruction (Kivisild *et al.* manuscript in preparation) of haplogroup U lineages in Indians. This phylogeny is dominated by a single but diverse sub-branch, which is defined by an A to G transition at np 16,051. The transition is rare in European haplogroup U sequences (sub-cluster U2) and is usually associated there with two additional mutations: an A to C transversion at np 16,129 and a T to C transition at np 16,189. The search in other world-wide mtDNA sequence data sets, including African ones, did not reveal sequences similar to those present in India. Thus, the extensive variation within the Indian mtDNA sub-cluster U2 has most likely developed locally in India. The coalescence age of the branches of U2 in India is 53,000 ± 4,000 years. This date is within the range of error limits of the haplogroup M coalescence time discussed above and synchronous with the

oldest layer of haplogroup U found in Europe: sub-cluster U5 was dated 52,000 years old (Richards 1998). It is important to note, however, that these coalescence times pre-date the archeologically accepted time scale of peopling of Europe by modern humans.

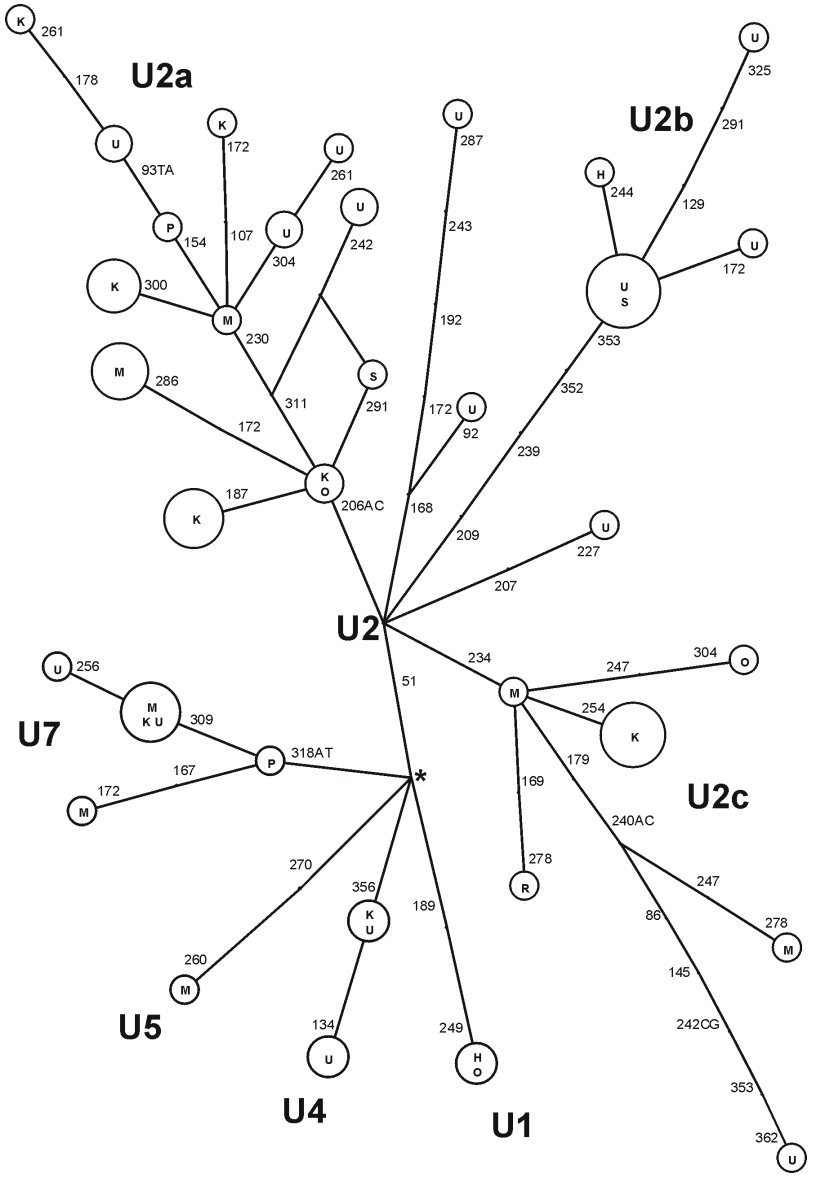

Figure 6. Reconstruction of haplogroup U in India. Populations as in Fig. 4. For other details, see Legend to Fig. 3.

There is still another fraction of haplogroup U, sub-cluster U7, which has likely developed in India from the Palaeolithic (Kivisild *et al.*, manuscript in preparation). It is occasionally found in populations of southern Europe and in the Trans-Caucasus but is present at much higher frequency and diversity in Indian populations. In contrast, other haplogroup U subclusters that are abundant in Europe are rare or missing in India. The same is true for the remaining European-specific mtDNA haplogroups (Table 2). Their combined frequency in India is below 8% and they are spread quite evenly: both in northern and southern regions. None of the extant western Eurasian populations display significantly higher affinity to the Indian "Caucasoid" lineages (Kivisild *et al*, manuscript in preparation). Thus, the putative approximately 4,000 year-old Indo-Aryan invasion can hardly be taken as an explanation for

their spread in India. This 8% fraction of Indian maternal lineages that can be considered as a "loan" from Caucasoid populations is distant from the lineages shared with western Eurasian populations to an extent that suggests 9,000 years for their departure. This mean may and probably does reflect multiple minor immigration waves of nomadic people to India during a long time interval. Any major recent immigration wave should be traceable in a form of frequency gradient. The only geographic variations detected in India are slight spatially ordered frequency changes in haplogroups U and M. However, both demonstrate variation by far older than the 4,000 years under discussion.

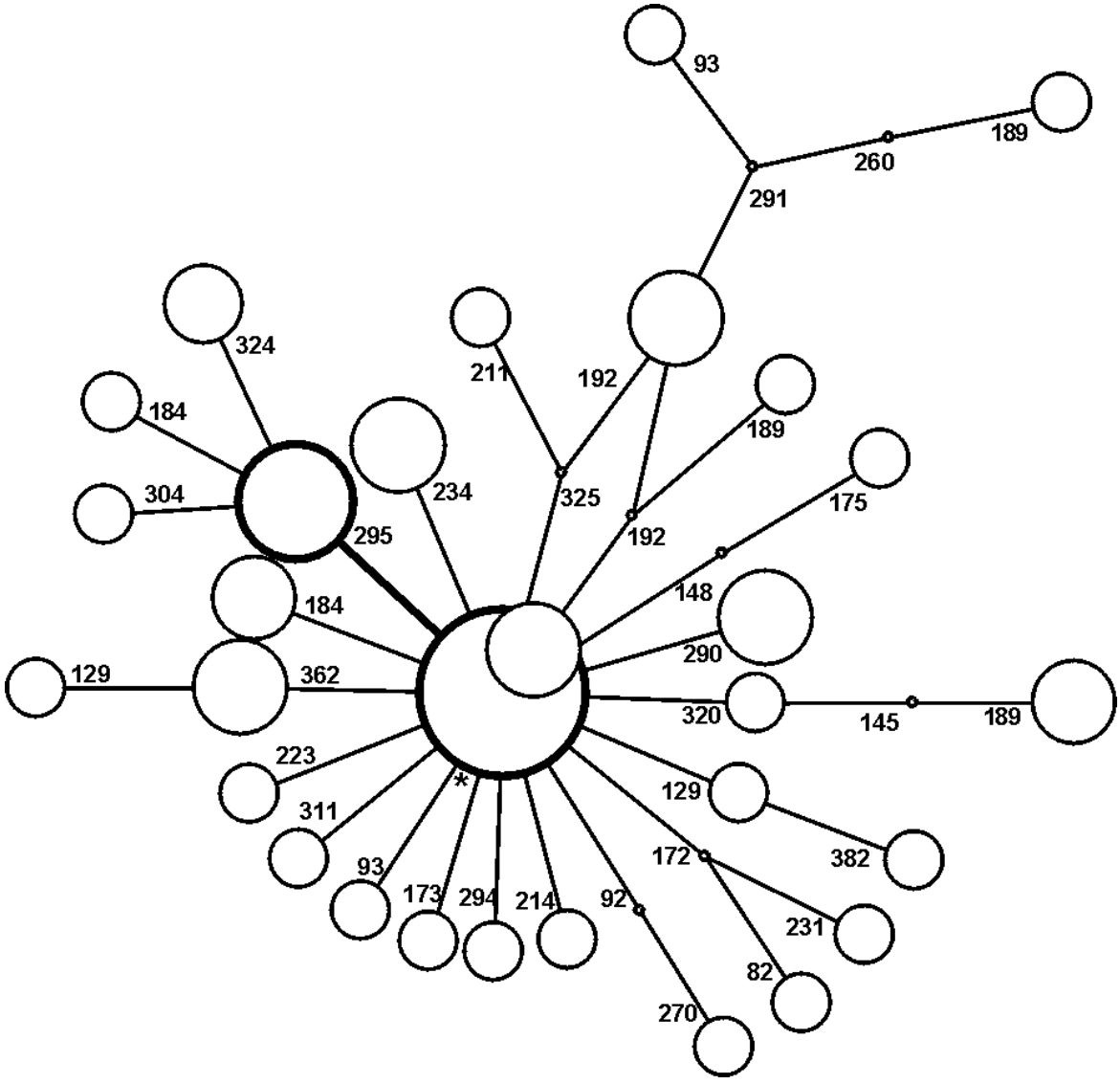

Figure 7. Reconstruction of haplogroup W in western Eurasia (open nodes) and in India (shaded nodes). For other details, see Legend to Fig. 3.

The presence of yet another western Eurasian-specific haplogroup in India deserves special attention. This is haplogroup W, which is a minor mtDNA variety in Europe (Torroni 1996) with a frequency of 3% and less. Haplogroup W was also found in India though with a rather limited and specific distribution (Kivisild *et al.*, manuscript in preparation). It is practically absent among most of the Indian populations but is present both in high frequency and

diversity among the Lamani speaking Lobanas (13%) from Punjab. Another linguistically related Lamani population, the Lambadis from Andhra Pradesh, also possesses haplogroup W, but at several-fold lower frequency (3%). The European and Indian haplogroup W lineages overlap only at their ancestral node (Fig. 7). Coalescence ages calculated separately for Indian and European haplogroup W sequences are very similar: 22,000 and 25,000 years, respectively. The specific distribution of W in India suggests that it could have been imported there by a small number of populations, like the mobile Lobanas, who are known for their past inter-state salt trading occupation (Mastana *et al.* 1991). Nevertheless, as the lineages do not match those found so far in Europeans, their recent European admixture is unlikely. It is reasonable to guess that these lineages can be found also in Iran and/or Afghanistan.

Slightly less than one half of the European mtDNA-s belong to haplogroup H (Torroni *et al.* 1996). Haplogroup H has a minor sister-haplogroup, V, which is present in western Europe but can be found also at an exceptionally high frequency (>40%) among the Saamis (Torroni *et al.* 1998). These two haplogroups share a common node in phylogeny (HV*in Macaulay *et al.* 1999), basic offshoots of which are rare in European populations. The derivatives of this ancestral node are, however, found at a remarkable frequency in the populations of Trans-Caucasus and also in northern India. This cluster, which we propose to define as haplogroup Ö, will be discussed in greater detail in our accompanying paper (E. Metspalu *et al*, this book). The populations of Kashmir and Pakistan are still insufficiently covered in this respect. Finding of four haplogroup Ö mtDNA genomes from a sample of 27 individuals from this region (see Fig. 1 in E. Metspalu *et al.*, this book) hints that the north-western part of the Indian Peninsula may be also rich in derivatives of this ancestral node, which has given rise to a half of European maternal inheritance. Nevertheless, haplogroup Ö has not penetrated deep into India. Taking the presence of nucleotide A at np 00073 as an indirect indicator (Table 5) one can see a clear diminishing south-eastward gradient of this lineage cluster in Asia. Haplogroup Ö reveals traces of a Palaeolithic demic expansion about 35,000 years ago (Kivisild *et al.*, manuscript in preparation). However, the lack of multiple direct matches in haplogroup Ö sequences between Indian samples and the Armenian, Ossetian and Georgian ones suggests that this link does not imply any recent admixture between these groups of populations. Instead, we can hypothesise that India, the Trans-Caucasus and the territories between them were the birthplace of the mitochondrial haplogroups (the place of their expansion) that eventually reached western as well as northern Europe.

Table 5. Frequency of 00073 A in some Asian populations

	n/N	%
East Indonesia[1]	0/142	0
South China (Han)[1]	0/103	0
Java[1]	0/96	0
Malaysia[1]	0/81	0
Taiwan[1]	0/81	0
Bangladesh[1]	0/31	0
Philippines[1]	1/175	0.6
Borneo[1]	1/91	1.1
S.India and Sri Lanka[1]	2/73	2.7
S. India(Lambadis)[2]	2/86	2.3
N. India (Uttar Pradesh)[2]	4/70	5.7
N. India (Sikh)[1]	4/47	8.5
Pakistan[1]	11/73	15

[1] – data from Melton and Stoneking 1996
[2] – Kivisild *et al*., manuscript in preparation

1.3. Maternal lineage clusters, which are highly specific for Indian populations while being rare elsewhere

More than ten per cent of the Indian mtDNA sequences do not belong to any of the continent-specific mtDNA haplogroups characterised so far (Tables 1, 2). Nevertheless, the position of these lineages in the world-wide mtDNA phylogeny (Fig. 1) is not difficult to reveal: they all stem out of a node that occupies a crucial position in the human mtDNA phylogenetic tree (Kivisild *et al.*, manuscript in preparation). It coincides with a hypothetical branching point connecting a large number of distinct, well-characterised mtDNA haplogroups. Theoretically, the existence of such a node was obvious already earlier and in one of the schemes it has been defined as R* (Macaulay *et al.* 1999). However, thus far it has existed as an "empty node". Defining it as the founder of a super-haplogroup of mtDNA lineages allows one to say that it is an ancestral state of all western Eurasian sequences belonging to haplogroups H, V, J, T, U and K and has the same position relative to eastern Eurasian and Amerind sequences belonging to haplogroups F and B (Fig. 1). The western Eurasian haplogroups listed above constitute about 90% of mtDNA variation in Europe, whereas an Asian-specific haplogroup B is close to fixation in some Polynesian populations (Lum *et al.* 1998; Sykes *et al.* 1995) and, together with haplogroup F, makes up a large portion of the mtDNA varieties found in southeastern Asian populations (Ballinger *et al.* 1992).
What is remarkable about the Indian populations is that here and so far only here, one can find at a significant frequency and diversity mtDNA sequences in the vicinity of this central node (Kivisild *et al.*, manuscript in preparation; Fig. 8). Calculation of the beginning of expansion time for R* in India resulted in estimate of 55,000 ± 5,000 years. It is noteworthy that this time estimate practically coincides with those we found for haplogroups M and U2.

2. CONCLUSIONS

Both western and eastern Eurasian-specific mtDNA haplogroups can be found in India together with strictly Indian-specific ones. However, in India the structure of the haplogroups shared either with western or eastern Eurasian populations is profoundly different. This indicates a local independent development over a very long time period. Minor overlaps with lineages described in other Eurasian populations clearly demonstrate that recent immigrations have had very little impact on the innate structure of the maternal gene pool of Indians. Despite the variations found within India, these populations stem from a limited number of founder lineages. These lineages were most likely introduced to the Indian subcontinent during the Middle Palaeolithic, before the peopling of Europe and perhaps the Old World in general. Our demographic analysis reveals at least two major expansion phases that have influenced the wide assortment of the Indian mtDNA lineages. The more recent phase, which according to our estimation started around 20,000-30,000 years ago, seems to correspond to the transition from the Middle to the Upper Palaeolithic. The first expansion phase may reflect a demographic burst immediately after the initial peopling of India around 50 - 60 thousand years ago. This wave of expansion brought forward also those maternal lineages that can rightfully claim the name of Eurasian Eves.

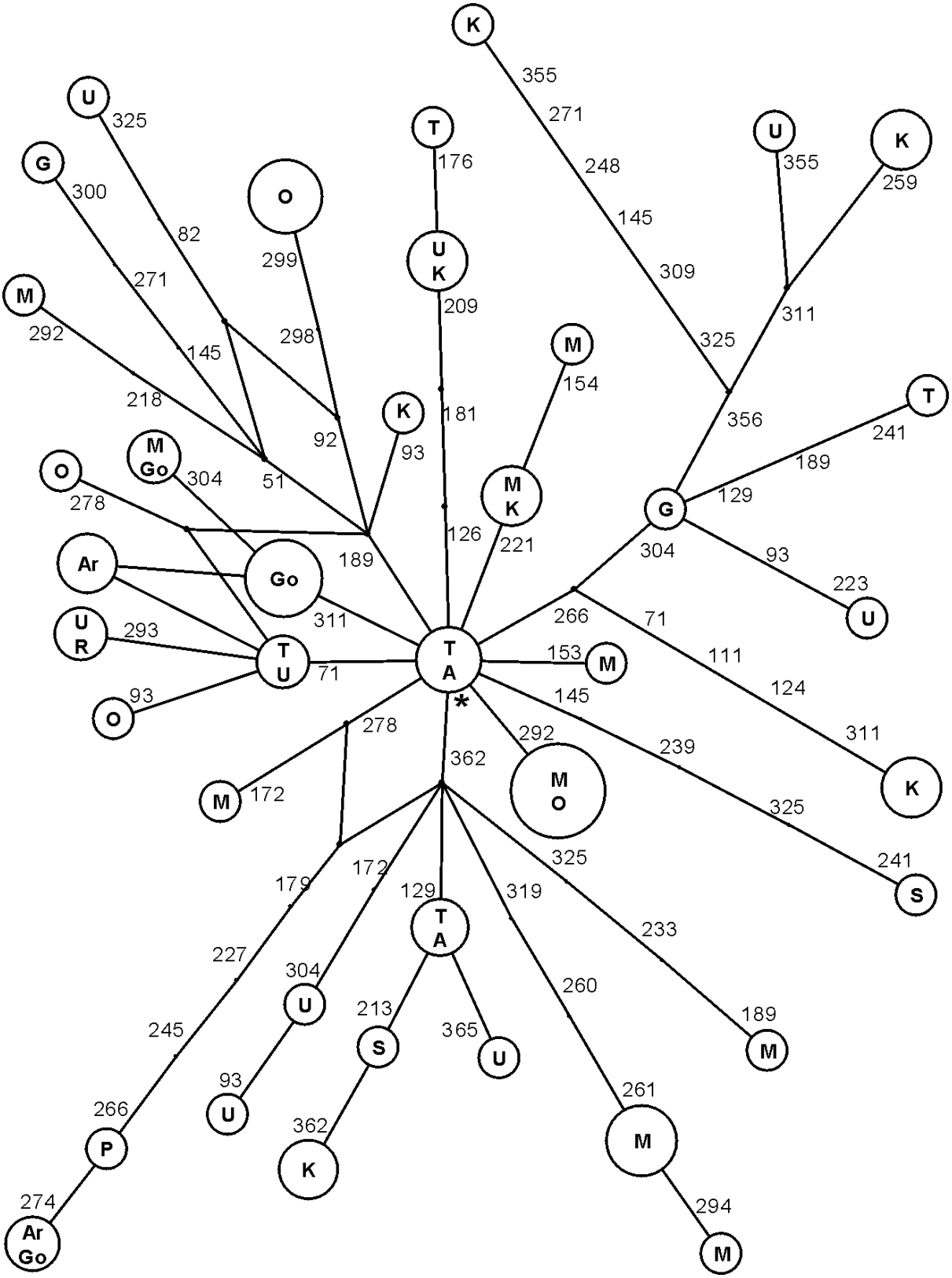

Figure 8. Reconstruction of super-haplogroup R lineages. Populations: U - Uttar Pradesh; O - Lobanas; M - Lambadis; S - Kashmir; K - Kerala; A - Andhra Pradesh; P - Pakistan; G - Gujerat; R – Rajastan; Ar – Armenia; Go – Georgia. For other details, see Legend to Fig. 3.

REFERENCES

Aitken MJ, Valladas H (1992) Luminescence dating relevant to human origins. Philos Trans R Soc Lond B Biol Sci 337:139-44

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457-65

Bailliet G, Rothhammer F, Carnese FR, Bravi CM, Bianchi NO (1994) Founder mitochondrial haplotypes in Amerindian populations. Am J Hum Genet 55:27-33

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, et al (1992) Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. Genetics 130:139-52

Bamshad M, Rao BB, Naidu JM, Prasad BVR, Watkins S, Jorde LB (1997) Response to Spurdle *et al.* Human Biology 69:432-435

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BV, et al (1998) Female gene flow stratifies Hindu castes. Nature 395:651-2

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

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution [see comments]. Nature 325:31-6

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (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

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, 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

Deraniyagala SU (1998) Pre- and protohistoric settlement in Sri Lanka. XIII U.I.S.P. Congress Proceedings. Vol. 5., A.B.A.C.O. s.r.l., Forli, pp. 277-285

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

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

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, et al (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579-90

Joshi RV (1996) SOUTH ASIA in the period of *Homo sapiens neanderthalensis* and contemporaries (Middle Palaeolithic) History of Humanity. Vol. I. UNESCO, pp 162-164

Kolman C, Sambuughin N, 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 RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90:19-30

Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific Island and Asian populations. Am J Hum Genet 63:613-24

Lutz S, Weisser HJ, Heizmann J, 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

Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, et al (1999) The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. Am J Hum Genet 64

Mastana SS, Garlick JP, Papiha SS (1991) Serogenetic studies of the Lobanas in north-west India. Anthropol Anz 49:177-86

McDermott F, Grun R, Stringer CB, Hawkesworth CJ (1993) Mass-spectrometric U-series dates for Israeli Neanderthal/early modern hominid sites. Nature 363:252-5

Melton T, Stoneking M (1996) Extent of heterogeneity in mitochondrial DNA of ethnic Asian populations. J Forensic Sci 41:591-602

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979-92

Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS (1996) 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 LF, Santachiara Benerecetti AS (1996) 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, Santachiara-Benerecetti AS (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, Pepe G, Shrestha SL, Modiano G, Santachiara Benerecetti AS (1992) MtDNA polymorphisms among Tharus of eastern Terai (Nepal). Gene Geogr 6:139-47

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. Am J Hum Genet 62:420-34

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241-60

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN, Knowler WC, et al (1990) Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am J Hum Genet 46:613-23

Stringer CB (1990) The emergence of modern humans. Sci Am 263:98-104

Stringer CB (1992) Reconstructing recent human evolution. Philos Trans R Soc Lond B Biol Sci 337:217-24

Stringer CB, Grun R, Schwarcz HP, Goldberg P (1989) ESR dates for the hominid burial site of Es Skhul in Israel. Nature 338:756-8

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57:1463-75

Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, 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, et al (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835-50

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) 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 JV, Barrantes R, Schurr TG, Wallace DC (1994) 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, 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 TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, et al (1993) Asian affinities and continental radiation of the four founding Native American mtDNAs. Am J Hum Genet 53:563-90

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, 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 RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, et al (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. Am J Hum Genet 53:591-608

Wallace D (1995) 1994 William Allan Award Address. Mitochondrial DNA variation in human evolution, degenerative disease, and aging. Am J Hum Genet 57:201-23

Watson E, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691-704