Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages

T. Kivisild*, M.J. Bamshad†, K. Kaldma*, M. Metspalu*, E. Metspalu*, M. Reidla*, S. Laos*, J. Parik*, W.S. Watkins†, M.E. Dixon†, S.S. Papiha‡, S.S. Mastana§, M.R. Mir¶, V. Ferak¥ and R. Villems*

About a fifth of the human gene pool belongs largely either to Indo-European or Dravidic speaking people inhabiting the Indian peninsula. The 'Caucasoid share' in their gene pool is thought to be related predominantly to the Indo-European speakers. A commonly held hypothesis, albeit not the only one, suggests a massive Indo-Aryan invasion to India some 4,000 years ago [1]. Recent limited analysis of maternally inherited mitochondrial DNA (mtDNA) of Indian populations has been interpreted as supporting this concept [2,3]. Here, this interpretation is questioned. We found an extensive deep late Pleistocene genetic link between contemporary Europeans and Indians, provided by the mtDNA haplogroup U, which encompasses roughly a fifth of mtDNA lineages of both populations. Our estimate for this split is close to the suggested time for the peopling of Asia and the first expansion of anatomically modern humans in Eurasia [4–8] and likely pre-dates their spread to Europe. Only a small fraction of the 'Caucasoid-specific' mtDNA lineages found in Indian populations can be ascribed to a relatively recent admixture.

Addresses: *Department of Evolutionary Biology, Tartu University, Riia 23, Tartu 51010, Estonia. †Departments of Pediatrics and Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA. ‡Department of Human Genetics, University of Newcastle-upon-Tyne, Newcastle, UK. §Department of Human Sciences, Loughborough University, Loughborough, UK. ¶Veterinary College of Srinagar, Kashmir 190003, India. ¥Faculty of Natural Sciences, Comenius University, 842 15, Bratislava, Slovakia.

Correspondence: R. Villems
E-mail: rvillems@ebc.ee

Received: 10 August 1999
Revised: 28 September 1999
Accepted: 29 September 1999

Published: 8 November 1999

Current Biology 1999, 9:1331–1334

0960-9822/99/$ – see front matter

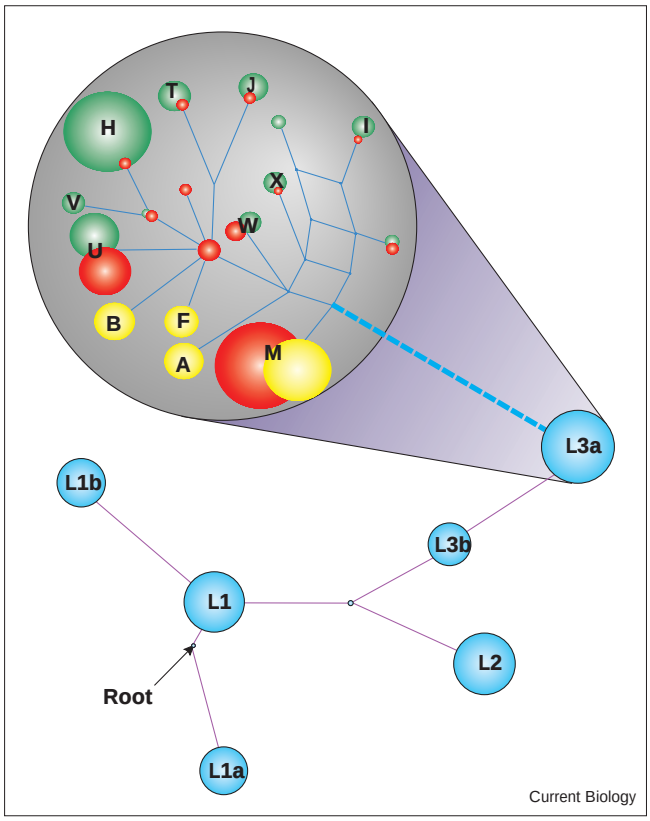
Results and discussion

The recent African origin of modern humans is now supported by palaeoanthropological, as well as sex-specific and autosomal genetic, evidence (for recent reviews, see [8,9]). The concordance between the interpretation of data obtained by mtDNA, Y-chromosomal and most of the autosomal markers is encouraging and suggests that, irrespective of the differences in the mode of inheritance, these three genetic approaches produce consistent overall findings in this central issue.

We sequenced the mitochondrial hypervariable region I (HVR I) and performed extensive restriction fragment length polymorphism (RFLP) analysis of 550 Indian mtDNA samples. We inferred a parsimonious phylogenetic tree from the data using the median network approach [10], which is particularly suitable for intraspecies analysis of mtDNA lineages and other highly variable data sets. Figure 1 is an outline of this Indian mtDNA tree within the background of the previously defined global mtDNA lineage clusters (haplogroups) [11–13]. Consistent with the recent out-of-Africa model of human origins [14], all of the Indian mtDNA lineages we inferred can be seen as deriving from the African mtDNA lineage cluster L3a, described in [15]. We found that more than 80% of the Indian mtDNA lineages belong to either Asian-specific haplogroup M (60.4%) or western-Eurasian-specific haplogroups H, I, J, K, U and W (20.5%), while the remaining 19.1% of lineages do not belong to any of the previously established mtDNA haplogroups (Table 1). We note that haplogroup K should now be considered a sub-cluster of haplogroup U [13].

The first and the most profound layer of overlap between the western-Eurasian and the Indian mtDNA lineages relates to haplogroup U, a complex mtDNA lineage cluster with an estimated age of 51,000–67,000 years [16]. Until now, this haplogroup has not been reported to occur in India nor east of India and was considered a western-Eurasian-specific haplogroup. Surprisingly, we found that haplogroup U is the second most frequent haplogroup in India as it is in Europe (Table 1). Nevertheless, the spread of haplogroup U subclusters in Europe and India differs profoundly (Figure 2). The dominant subcluster in India is U2. Although rare in Europe, the South-Asian form differs from the western-Eurasian one: western-Eurasian U2 includes a further characteristic transversion at nucleotide position (np) 16,129 [12], which is absent in Indian U2 varieties (Figure 2). We calculated the coalescence age essentially as described in [15,17] and estimate the split between the Indian and western-Eurasian U2 lineages as 53,000 ± 4,000 years before present (BP). We note that U5, the most frequent and ancient subcluster of haplogroup U

Figure 1



The skeleton network of Indian lineage clusters on the background of continent-specific mtDNA haplogroups. Red, Indians; green, western Eurasians; yellow, eastern Eurasians; blue, Africans. Haplogroup frequencies are proportional to node sizes. All Indian, eastern-Eurasian and western-Eurasian mtDNA lineages coalesce finally to the African node L3a. The former are shown magnified to account for higher mtDNA diversity in sub-Saharan Africans. The most likely root of the tree [15] is indicated within a pan-African cluster L1. The dashed line leading from the African external node L3a to the Eurasian mtDNA varieties identifies the position of L3a in the magnified part of the tree.

Table 1

MtDNA haplogroup frequencies (%) among some Indian and Eurasian populations.

	Population or group of populations						
	India			Western and eastern Eurasia			
Haplogroup affiliation	*1*	*2*	*3*	*4*	*5*	*6*	*7*
African*	0	0	0	0	0	0	0
Eastern Eurasian	53.3	65.7	60.4	1.5	0.7	61	94.5
Western Eurasian	29.3	14.5	20.5	80.9	95.4	30.5	0
H	3	1.2	1.8	24.8	41.1	14	0
I	2	0	0.7	1.8	2.6	1	0
J	0	0.4	0.5	6.7	10.3	2.5	0
K	0	0.4	0.2	8.2	4.4	0.5	0
T	1	1.7	1.8	11.8	10.1	3.5	0
U	23.3	10.3	13.1	21.2	20.8	8	0
V	0	0	0	0	3.1	0	0
W	0	0.4	2.2	0.9	1.6	1	0
X	0	0	0.2	5.5	1.4	0	0
Others†	17.4	19.8	19.1	17.6	3.9	8.5	5.5

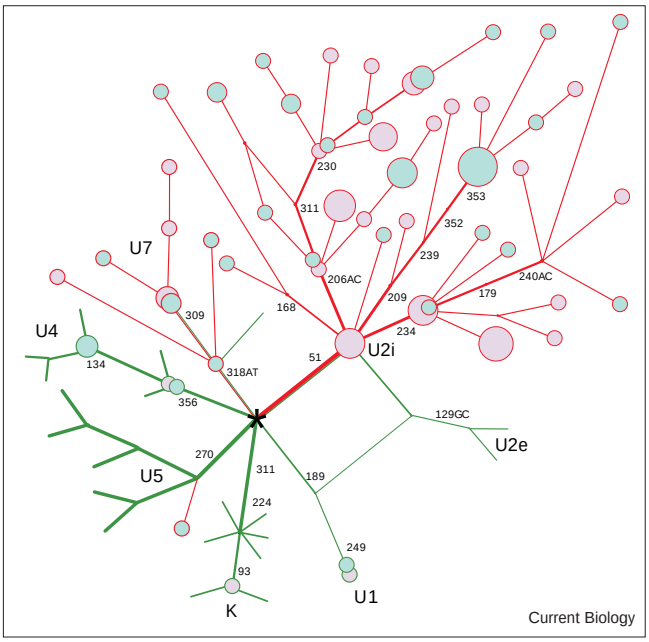
The numbers in italics represent the following populations: *1,* North India (Uttar Pradesh, *n* = 103, this study); *2*, South India (Andhra Pradesh Telugus, *n* = 250, this study); *3*, India total (*n* = 550, this study); *4*, The Caucasus – Armenians (*n* = 192, this study), Georgians (*n* = 138, this study); *5,* Europe – Slovaks (*n* = 129, this study), Russians (*n* = 100, this study), Czechs (*n* = 95, this study), Estonians (*n* = 100, this study), Italians (*n* = 99 [27]), Finns (*n* = 49 [16]); *6,* Central Asia – Kirghiz (*n* = 95, deduced from [18]), Kazakhs (*n* = 55, deduced from [18]), Uighurs (*n* = 55, deduced from [18]); *7*, Tibet (*n* = 54 [26]). *L1 and L2 defined by +3592 *Hpa*I. †Lineages that do not belong to any of the previously established haplogroups.

in Europe, has an almost identical coalescence age estimate [13]. Still, despite their equally deep time depth, the Indian U2 has not penetrated western Eurasia, and the European U5 has almost not reached India (Table 2).

Subcluster U7 (among U* in [12,13]) is another variety of haplogroup U present in India (Figure 2). Unlike the Indian U2, it has been sampled, albeit rarely, in southern Europe, the Near East [12,13] and (according to HVR I sequence identification only) also in Central Asia [18]. We calculated the coalescence age of this subcluster in India as 32,000 ± 5,500 years: still deep in late Pleistocene but considerably younger than that for U2. Table 2 compares the frequency of varieties of haplogroup U in India, in the Trans-Caucasus populations and in Europe.

Typical western-Eurasian mtDNA lineages found in India belong to haplogroups H, I, J, T, X and to subclusters U1, U4, U5 and K of haplogroup U (Figure 1; Tables 1,2). Frequencies of these lineages in Indian populations are more than an order of magnitude lower than in Europe: 5.2% versus 70%, respectively (normalised from Table 1). This finding might be explained by gene flow, as suggested previously [2]. Nevertheless, we note that the frequency of these mtDNA haplogroups reveals neither a strong north–south, nor language-based gradient: they are found both among Hindi speakers from Uttar Pradesh (6%) and Dravidians of Andhra Pradesh (4%). Assuming that they are largely of western-Eurasian origin, we may ask when their spread in India started. To assign a tentative date for their introduction, we calculated the averaged minimal distance of the corresponding mtDNA hypervariable region sequences in Indians from the branches shared with western Eurasians. We obtained a value for the statistic ρ (see Materials and methods) equal to 0.46, consistent with a

Figure 2



Reconstruction of haplogroup U lineages found in India. Green bold lines, the background of previously characterized haplogroup U lineages from western Eurasia; red lines, lineages and haplotypes found only in India; pink nodes, Dravidic speakers; blue nodes, Hindi speakers. The HVR I mutations at given nucleotide positions compared with the Cambridge Reference Sequence [28] are shown less the 16,000 prefix near the lines connecting the nodes. Only transversions are specified (for example, 318AT defines an A to T transversion at np 16,318). The ancestral node of haplogroup U, marked with an asterisk, differs from the reference sequence by transitions at nps 00073 (+*Alw*44I) , 7028 (+*Alu*I), 12308 (+*Hinf*I), 11467 (–*Tru*I).

Table 2

Frequencies (%) of subclusters of haplogroup U in India and in some western-Eurasian populations.

	Population or group of populations		
Subcluster	Indians	Armenians, Georgians	Estonians, Russians, Slovaks
U1	2.3	14.4	1.2
U2i	77.9	1.0	NF
U2e	NF	5.2	10.6
U3	NF	15.5	4.7
U4	4.7	18.6	20.0
U5	1.2	11.3	45.9
U6	NF	NF	NF
U7	12.7	4.1	NF
K	1.2	28.9	12.9
Other U	NF	1.0	4.7

Population sizes and their absolute U frequency as in Table 1. Subclusters of haplogroup U are defined as in [12,13]. U2i and U2e indicate Indian and western-Eurasian varieties of subcluster U2, respectively (see Figure 2). NF, not found.

divergence time of 9,300 ± 3,000 years BP. This is an average over an unknown number of various founders and, therefore, does not tell us whether there were one or many migration waves, or whether there was a continuous long-lasting gradual admixture. Their low frequency but still general spread all over India plus the estimated time scale, does not support a recent massive Indo-Aryan invasion, at least as far as maternally inherited genetic lineages are concerned. We note, however, that within an error margin this time estimate is consistent with the arrival to India of cereals domesticated in the Fertile Crescent [4,19]. Furthermore, the spread of these western-Eurasian-specific mtDNA clusters also among Dravidic-speaking populations of India lends credence to the suggested linguistic connection between Elamite and Dravidic populations [20].

Thus, we have shown that the overwhelming majority of the so-called western-Eurasian-specific mtDNA lineages in Indian populations, estimated here to be carried by more than a hundred million contemporary Indians, belong in fact to an Indian-specific variety of haplogroup U of a late Pleistocene origin. The latter exhibits a direct common phylogenetic origin with its sister groups found in western Eurasia (Figure 1), but it should not be interpreted in terms of a recent admixture of western Caucasoids with Indians caused by a putative Indo-Aryan invasion 3,000–4,000 years BP. From the deep time depth of the split between the predominant Indian and European haplogroup U varieties, it could be speculated that haplogroup U arose in neither of the two regions. This split could have already happened in Africa, for example, in Ethiopia, where haplogroup U was recently described [21].

Although there is no strong evidence yet for the presence of anatomically modern humans in India before 35,000–40,000 years ago [22], the earliest estimates of the presence of modern humans in Australia [23] make it very likely that the subcontinent served as a pathway for eastward migration of modern humans somewhat earlier and that it could have been inhabited by them *en route*, as suggested by the 'Southern Route' hypothesis [24,25]. Our coalescence age estimate for the mtDNA sub-cluster U2 overlaps not only with the corresponding value for the European U5, but with the suggested coalescence age of the Indian-specific subset of the predominantly Asian haplogroup M lineages as well (M.J.B., T.K., W.S.W., M.E.D., B.B. Rao, J.M. Naidu, *et al.*, unpublished observations). Taken together, these data suggest that a common denominator — most likely beneficial climate conditions — led to the expansion of populations all over Eurasia, including the ancestors of those who now encompass most of the mtDNA genome pool of the extant

Indians. Furthermore, this specific distribution of mtDNA varieties in India compared with the distribution observed among Mongoloids and the Caucasoid populations of western Eurasia (Figure 1) is, at present, best explained by two separate late Pleistocene migrations of modern humans to India. One of them, possibly arriving by the southern route, brought to India an ancestral population carrying haplogroup M and was spread further eastward. The second migration brought the ancestors of haplogroup U. Although the admixture of these major waves started perhaps very early — explaining the spread of these major mtDNA varieties all over the subcontinent — it is likely that it happened after the carriers of haplogroup M found their way further east, explaining the absence of haplogroup U lineages among Mongoloid populations studied so far.

Materials and methods

Samples from 86 Lambadi, 62 Lobana (Lamani speakers; Indo-Aryan languages), 12 Tharu and 18 Buksa (Indo-Aryan languages), 122 predominantly Indo-Aryan language speakers from Uttar Pradesh (GenBank accession numbers AJ234902–AJ235201) and a set of 250 Telugu samples (Dravidic speakers) were sequenced for hypervariable region I of mtDNA and typed for the presence of major continent-specific markers, described in [11,16,26]. The HVR I polymorphic sites of all 550 Indian mtDNAs sequenced by us are provided in the Supplementary material. The phylogenetic analysis also included 101 published HVR I sequences from south-western India [6].

Phylogenetic analysis was performed by reduced median networks [10], applied here using parsimony analysis of the data. The median network analysis allows one to reveal simultaneously multiple parallel, equally probable, phylogenetic pathways in the form of reticulations induced by highly variable markers. The distinct mtDNA lineage clusters are referred to here as haplogroups. The time to the most recent common ancestor of a cluster of lineages (haplogroup) or, where appropriate, a sub-cluster inside a particular haplogroup, was calculated as described [17], using an estimator ρ, which is the average transitional distance from the founder haplotype sequence.

Supplementary material

Supplementary material including a table listing the mtDNA HVR I sequence polymorphisms in different Indian populations and a more detailed description of the materials and methods is available at http://current-biology.com/supmat/supmatin.htm.

Acknowledgements

We thank Lynn B. Jorde and Paul M.A. Broda for valuable comments, and Jaan Lind and Ille Hilpus for technical assistance. This work was supported by Citrina Foundation UK and by Estonian Science Fund grants 1669 and 2887 to R.V., NSF grants SBR-9514733 and SBR-9512178, and NIH grant PHS MO1-00064.

References

1. Thapar BK, Rahman A: The post-Indus cultures. In *History of Humanity*, Volume II. Edited by Dani AH, Mohen J-P. UK: Clays Ltd, St Ives plc.; 1996:266-279.
2. Passarino G, Semino O, Bernini LF, Santachiara-Benerecetti AS: Pre-Caucasoid and Caucasoid genetic features of the Indian population, revealed by mtDNA polymorphisms. *Am J Hum Genet* 1996, 59:927-934.
3. Barnabas S, Apte RV, Suresh CG: Ancestry and interrelationships of the Indians and their relationship with other world populations: a study based on mitochondrial DNA polymorphisms. *Ann Hum Genet* 1996, 60:409-422.
4. Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes.* Princeton: Princeton University Press; 1994.
5. Stringer CB, Andrews P: Genetic and fossil evidence for the origin of modern humans. *Science* 1988, 239:1263-1268.
6. Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, *et al.*: Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 1995, 56:979-992.
7. Seielstad M, Bekele E, Ibrahim M, Tour A, Traor M: A view of modern human origins from Y chromosome microsatellite variation. *Genome Res* 1999, 9:558-567.
8. Foley R: The context of human genetic evolution. *Genome Res* 1998, 8:339-347.
9. Disotell TR: Human evolution: sex-specific contributions to genome variation. *Curr Biol* 1999, 9:R29-31.
10. Bandelt HJ, Forster P, Sykes BC, Richards MB: Mitochondrial portraits of human populations using median networks. *Genetics* 1995, 141:743-753.
11. Wallace D: Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet* 1995, 57:201-223.
12. Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, *et al.*: The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999, 64:232-249.
13. Richards MB, Macaulay VA, Bandelt HJ, Sykes BC: Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 1998, 62:241-260.
14. Cann RL, Stoneking M, Wilson AC: Mitochondrial DNA and human evolution. *Nature* 1987, 325:31-36.
15. Watson E, Forster P, Richards M, Bandelt HJ: Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 1997, 61:691-704.
16. Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, *et al.*: Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996, 144:1835-1850.
17. Forster P, Harding R, Torroni A, Bandelt HJ: Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996, 59:935-945.
18. Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, *et al.*: Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 1998, 63:1824-1838.
19. Diamond J: *Guns, Germs and Steel: The Fates of Human Societies.* London: Jonathan Cape; 1997:99-101.
20. Renfrew C: The origins of Indo-European languages. *Sci Am* 1989, 261:82-90.
21. Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS: Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 1998, 62:420-434.
22. Deraniyagala SU: Pre- and protohistoric settlement in Sri Lanka. In *XIII U.I.S.P.P. Congress Proceedings* Volume V. Forli: A.B.A.C.O. s.r.l; 1998:277-285.
23. Thorne A, Grun R, Mortimer G, Spooner NA, Simpson JJ, McCulloch M, *et al.*: Australia's oldest human remains: age of the Lake Mungo 3 skeleton. *J Hum Evol* 1999, 36:591-612.
24. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J: Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc Natl Acad Sci USA* 1988, 85:6002-6006.
25. Lahr M, Foley R: Multiple dispersals and modern human origins. *Evol Anth* 1994, 3:48-60.
26. Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, *et al.*: Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 1994, 93:189-199.
27. Torroni A, Petrozzi M, D'Urbano L, Sellitto D, Zeviani M, Carrara F, *et al.*: 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* 1997, 60:1107-1121.
28. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, *et al.*: Sequence and organization of the human mitochondrial genome. *Nature* 1981, 290:457-465.

Supplementary material

Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages

T. Kivisild, M.J. Bamshad, K. Kaldma, M. Metspalu, E. Metspalu, M. Reidla, S. Laos, J. Parik, W.S. Watkins, M.E. Dixon, S.S. Papiha, S.S. Mastana, M.R. Mir, V. Ferak and R. Villems

Current Biology 8 November 1999, 9:1331–1334

Supplementary materials and methods

Samples from 86 Lambadi, 62 Lobana (Lamani speakers; Indo-Aryan languages), 12 Tharu and 18 Buksa (Indo-Aryan languages) were collected as part of the ongoing genetic studies of the populations of the Indian subcontinent at the Division of Human Genetics of the University of Newcastle-upon-Tyne. In addition, 122 samples with mixed-caste status, predominantly Indo-Aryan language speakers from Uttar Pradesh and Kashmir, were included in the analyses (GenBank accession numbers AJ234902–AJ235201). The Uttar Pradesh sequences are from our independent Gypsy study (K.K., F. Calafell, T.K., M.J.B., J.P., E.M. *et al*., unpublished observations). The set of 250 Telugu samples that was used to represent Dravidic speakers will be published elsewhere. Altogether, 550 samples from the Indian peninsula were sequenced for hypervariable region I of mtDNA and typed for the presence of major continent-specific markers, described in [S1–S3].

Phylogenetic analysis was performed by reduced median networks [S4], applied here using parsimony analysis of the data. In general, parsimony methods for inferring phylogenies operate by selecting trees that minimize the total tree length [S5]. In particular, the median networks approach allows one to reveal simultaneously multiple parallel, equally probable, phylogenetic pathways in the form of reticulations induced by highly variable markers. These reticulations reflect either parallel mutations or, more often, ambiguities in the branching pattern of a phylogenetic tree. Compared with any 'single tree' method, the network approach does not increase the phylogenetic resolution artificially. Nevertheless, considerable reduction of the network towards a tree can be achieved by giving higher weight to conservative markers versus hypervariable ones. Here, the reduced median networks were constructed using RFLP-typed conservative markers from the mtDNA coding region, with haplogroups specified according to the nomenclature proposed in [S1–S3,S6] and were further refined using sequence data from HVR I of the D-loop of the mtDNA genome. Every cluster of mtDNA lineages thus inferred should, in theory, constitute a monophyletic clade in the human mtDNA pool. These distinct clusters are referred to here as mtDNA haplogroups. The time to the most recent common ancestor of a cluster of lineages (haplogroup) or, where appropriate, of a sub-cluster inside of a particular haplogroup, was calculated as described [S7], using an estimator ρ, which is the average transitional distance from a founder haplotype sequence. We considered only transitions between nucleotide positions 16,090–16,365 in the HVR I of mtDNA and one substitution per 20,180 years was taken as an average distance from a specified founder [S8]. The phylogenetic analyses also included 101 published D-loop sequences from southwestern India [S9]. Western Eurasian samples that were used as a comparator included our unpublished sequences and RFLP data on the Caucasus area (n = 330), Slavic populations (n = 324) and approximately 2000 sequences retrieved from data banks [S10,S11] and recent publications [S12,S13].

Supplementary references

S1. Wallace D: Mitochondrial DNA variation in human evolution, degenerative disease, and aging. *Am J Hum Genet* 1995, 57:201-223.

S2. Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, *et al.*: Classification of European mtDNAs from an analysis of three European populations. *Genetics* 1996, 144:1835-1850.

S3. Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, *et al.*: Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* 1994, 93:189-199.

S4. Bandelt HJ, Forster P, Sykes BC, Richards MB: Mitochondrial portraits of human populations using median networks. *Genetics* 1995, 141:743-753.

S5. Hillis DM, Moritz C, Mable BK: *Molecular Systematics, 2nd edition.* Sunderland: Sinauer Associates; 1996:415-430.

S6. Richards MB, Macaulay VA, Bandelt HJ, Sykes BC: Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 1998, 62:241-260.

S7. Forster P, Harding R, Torroni A, Bandelt HJ: Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 1996, 59:935-945.

S8. Watson E, Forster P, Richards M, Bandelt HJ: Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 1997, 61:691-704.

S9. Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, *et al.*: Demographic history of India and mtDNA-sequence diversity. *Am J Hum Genet* 1995, 56:979-992.

S10. Handt O, Meyer S, von Haeseler A: Compilation of human mtDNA control region sequences. *Nucleic Acids Res* 1998, 26:126-129.

S11. Miller KW, Dawson JL, Hagelberg E: A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. *Int J Legal Med* 1996, 109:107-113.

S12. Macaulay VA, Richards MB, Hickey E, Vega E, Cruciani F, Guida V, *et al.*: The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999, 64:232-249.

S13. Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, *et al.*: Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. *Am J Hum Genet* 1998, 63:1824-1838.

S14. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, *et al.*: Sequence and organization of the human mitochondrial genome. *Nature* 1981, 290:457-465.

Table S1

The mtDNA HVR I sequence polymorphisms in different Indian populations.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
Ind625	UP	51 145 179 234 240AC 242CG 353 362	U2i	+12308 *Hinf*I; -11465 Tru1I
Ind873	UP	86 126 223	M1	+10394*Dde*I; +10397*Alu*I
Ind872	UP	234	H	-00073*Alw*44I; -7025*Alu*I
Ind624	UP	126 223 311	M1	+10394*Dde*I; +10397*Alu*I
Ind874	UP	223 319	M2	+10394*Dde*I; +10397*Alu*I
Ind623	UP	223 290 319 362	A	+663 *Hae*III
Ind868	UP	51 129 223 362	M4	+10394*Dde*I; +10397*Alu*I
Ind869	UP	126 362	pTJ	-00073*Alw*44I
Ind621	UP	refer	R*	
Ind871	UP	172 304 362	R*	
Ind620	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind616	UP	51 93TA 154 206AC 230 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind622	UP	184 189 223 300	M*	+10394*Dde*I; +10397*Alu*I
Ind618	UP	189 223 294	L3a	
Ind612	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind614	UP	51 92 168	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind613	UP	256 309 318AT	U7	+12308 *Hinf*I; -11465 *Tru*1I
Ind619	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind617	UP	93 223 266 304	R1	
Ind615	UP	51 207 227	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind604	UP	126 223	M1	+10394*Dde*I; +10397*Alu*I
Ind605	UP	51 206AC 242 291 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind606	UP	71	R*	
Ind607	UP	86 223 335	M*	+10394*Dde*I; +10397*Alu*I
Ind608	UP	93 129 223	M4	+10394*Dde*I; +10397*Alu*I
Ind388	Guj	93 188 223 231 318	M5	+10394*Dde*I; +10397*Alu*I
Ind610	UP	92 126 223	M1	+10394*Dde*I; +10397*Alu*I
Ind538	UP	93 172 304 362	R*	
Ind429	Mah	153	R*	
Ind404	Pun	179 227 245 266 278 362	R*	
Ind600	UP	129 192 213 223	M4	+10394*Dde*I; +10397*Alu*I
Ind601	UP	51 172 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind602	UP	111 192 223 275	M*	+10394*Dde*I; +10397*Alu*I
Ind603	UP	189 223 254 270 311	M*	+10394*Dde*I; +10397*Alu*I
Ind609	UP	189 304	F	-12704*Hinc*II
Ind438	Kash	223 278	M*	+10394*Dde*I; +10397*Alu*I
Ind611	UP	266 304 311 355 356	R1	
Ind508	UP	51 206AC 230 304 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind537	UP	316	H	-00073*Alw*44I; -7025*Alu*I
Ind590	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind498	UP	51 206AC 230 261 304 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind499	UP	129 213 249	M4	+10394*Dde*I; +10397*Alu*I
Ind509	UP	51 82 92 189 325	R*	
Ind502	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind503	UP	134 356	U4	+12308 *Hinf*I; -11465 *Tru*1I
Ind506	UP	223	M*	+10394*Dde*I; +10397*Alu*I
Ind540	UP	189 192 223 299	M*	+10394*Dde*I; +10397*Alu*I
Ind539	UP	126 192 223	M1	+10394*Dde*I; +10397*Alu*I
Ind541	UP	51 206AC 242 291 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind500	UP	48 93 129 218 223 243	M4	+10394*Dde*I; +10397*Alu*I
Ind591	UP	223 286	M*	+10394*Dde*I; +10397*Alu*I
Ind592	UP	223	M*	+10394*Dde*I; +10397*Alu*I
Ind593	UP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Ind596	UP	126 189 223 344	M1	+10394*Dde*I; +10397*Alu*I
Ind594	UP	134 356	U4	+12308 *Hinf*I; -11465 *Tru*1I
Ind595	UP	51 206AC 230 304 311	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Ind597	UP	129 189 260 362	F	-12704*Hinc*II
Ind598	UP	51 93 223 304	M*	+10394*Dde*I; +10397*Alu*I
Ind599	UP	187 223	M*	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
Ind505	UP	214	H	-00073*Alw*44I; -7025*Alu*I
Pak454	Pak	126 154 223 239	M1	+10394*Dde*I; +10397*Alu*I
Ind455	Ben	129 140 223 271	M4	+10394*Dde*I; +10397*Alu*I
Ind456	Ori	51 169 234 278	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ind458	UP	129 169 223	I	+10237*Hph*I; +10394*Dde*I; +8249*Ava*II; –1715*Dde*I
Ind459	Tamil	189 223 278 362	M*	+10394*Dde*I; +10397*Alu*I
Ind462	UP	184 223 241 311	M3	+10394*Dde*I; +10397*Alu*I
Ind464	Bih	126 223 247	M1	+10394*Dde*I; +10397*Alu*I
Ind465	Bang	129 209 223	M4	+10394*Dde*I; +10397*Alu*I
Ind466	Mah	51 209 239 244 352 353	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ind467	AP	129 362	R*	
Ind489	Mah	172 278	R*	
Ind491	UP	129 362 365	R*	
Ind487	Raj	71 293	R*	
Ind488	Guj	223	M*	+10394*Dde*I; +10397*Alu*I
Ind492	UP	126 192 223	M1	+10394*Dde*I; +10397*Alu*I
Ind493	UP	129 169 223	I	+10237*Hph*I; +10394*Dde*I; +8249*Ava*II; –1715*Dde*I
Ind494	UP	129 223	M4	+10394*Dde*I; +10397*Alu*I
Ind495	UP	223 319	M2	+10394*Dde*I; +10397*Alu*I
Ind496	UP	51 168 172 192 243 287	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ind497	UP	223	M*	+10394*Dde*I; +10397*Alu*I
Ind264	Guj	51 145 189 271 300	R*	
Ind431	AP	refer	R*	
Ind426	UP	145 223 261 311	M3	+10394*Dde*I; +10397*Alu*I
Pak260	Pak	refer	H	–00073*Alw*44I; –7025*Alu*I
Ind423	Bih	223 368	M*	+10394*Dde*I; +10397*Alu*I
Ind428	Tamil	217 243	R*	
Pak267	Pak	223 362	M*	+10394*Dde*I; +10397*Alu*I
Ind507	UP	126 181 209	pTJ	
Pak451	Pak	129 223 264	M4	+10394*Dde*I; +10397*Alu*I
Pak403	Pak	92	Ö	–00073*Alw*44I
Ind504	UP	129 266 290 318 320 362	M4	+10394*Dde*I; +10397*Alu*I
Pak453	Pak	51 154 206AC 230 311	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Pak261	Pak	318AT	U7	+12308 *Hinf*I; –11465 *Tru*1I
Ind439	Ori	189 223 260 294 295 325	M	+10394*Dde*I; +10397*Alu*I
Ind510	UP	309 318AT	U7	+12308 *Hinf*I; –11465 *Tru*1I
Ind387	Guj	266 304	R1	
Ind433	Mah	189 249	U1	+12308 *Hinf*I; –11465 *Tru*1I
Pak425	Hary	51 93TA 154 178 206AC 230 261 311	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Pak262	Pak	129 223 298 311 327	M-C	+10394*Dde*I; +10397*Alu*I; +13262*Alu*I
Ind501	UP	189 223 274 319 320	M2b	+10394*Dde*I; +10397*Alu*I
Lam1	AP	189 223 274 311 319	M2b	+10394*Dde*I; +10397*Alu*I
Lam2	AP	129 223	M4	+10394*Dde*I; +10397*Alu*I
Lam3	AP	223 311	M3	+10394*Dde*I; +10397*Alu*I
Lam4	AP	189 233 304 325 362	R*	
Lam5	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Lam6	AP	223 270 311 319 352	M2a	+10394*Dde*I; +10397*Alu*I
Lam7	AP	189 223 278	X	–1715*Dde*I; +14465*Acc*I
Lam8	AP	129 223	M4	+10394*Dde*I; +10397*Alu*I
Lam9	AP	223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
Lam10	AP	292	R*	
Lam11	AP	223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
Lam12	AP	126 275 294 296 325	T	+13366*Bam*HI; +15606*Alu*I
Lam13	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
Lam14	AP	93 192 223 311	M3	+10394*Dde*I; +10397*Alu*I
Lam15	AP	167 172 318AT	U7	+12308 *Hinf*I; –11465 *Tru*1I
Lam16	AP	51 172 206AC 286	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam17	AP	126 129 223	M1	+10394*Dde*I; +10397*Alu*I
Lam18	AP	126 344	M1	+10394*Dde*I; +10397*Alu*I
Lam19	AP	93 192 223 311	M3	+10394*Dde*I; +10397*Alu*I
Lam20	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
Lam21	AP	260 261 319 362	R*	
Lam22	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I
Lam23	AP	86 129 223 249 259 311	M3	+10394*Dde*I; +10397*Alu*I
Lam24	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I
Lam25	AP	166 223 311 359	M3	+10394*Dde*I; +10397*Alu*I
Lam26	AP	188 189 223 231 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lam27	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lam28	AP	223 311 316 355	M3	+10394*Dde*I; +10397*Alu*I
Lam29	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I
Lam30	AP	189 192 223 292	W	+8249*Ava*II; –8994*Hae*III
Lam31	AP	260 261 319 362	R*	
Lam32	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Lam33	AP	69 126 145 222 261	J	–13704*Bst*NI; +10394*Dde*I
Lam34	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I
Lam35	AP	223 270 319	M2a	+10394*Dde*I; +10397*Alu*I
Lam36	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lam37	AP	51 179 234 247 278 /240 /1	U2i	+12308 *Hinf*I; -11465 *Tru*1I
Lam38	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lam39	AP	75 92 189 223	M*	+10394*Dde*I; +10397*Alu*I
Lam40	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Lam41	AP	309 318AT	U7	+12308 *Hinf*I; –11465 *Tru*1I
Lam42	AP	260 261 319 362	R*	
Lam43	AP	260 261 319 362	R*	
Lam44	AP	51 172 206AC 286	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam45	AP	223 324 357	M*	+10394*Dde*I; +10397*Alu*I
Lam46	AP	213 223 231 278 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lam47	AP	213 223 231 278 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lam48	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Lam49	AP	93 126 145 223	M1	+10394*Dde*I; +10397*Alu*I
Lam50	AP	213 223 231 278 318 324 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lam51	AP	51 206AC 230 311	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam52	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lam53	AP	223 270 274 292 319 352	M2b	+10394*Dde*I; +10397*Alu*I
Lam54	AP	223 304 311	M3	+10394*Dde*I; +10397*Alu*I
Lam55	AP	223 311 316 355	M3	+10394*Dde*I; +10397*Alu*I
Lam56	AP	51 189 218 292	R	
Lam57	AP	51 172 206AC 286	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam58	AP	260 261 294 319 362	R*	
Lam59	AP	75 92 189 223	M*	+10394*Dde*I; +10397*Alu*I
Lam60	AP	51 172 206AC 286	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam61	AP	260 270	U5	+12308 *Hinf*I; –11465 *Tru*1I
Lam62	AP	126 169 223	M1	+10394*Dde*I; +10397*Alu*I
Lam63	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Lam64	AP	184 274	H	–00073*Alw*44I; –7025*Alu*I
Lam65	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Lam66	AP	129 223 304 311	I	+10237*Hph*I; +10394*Dde*I; +8249*Ava*II; –1715*Dde*I
Lam67	AP	51 234	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lam68	AP	154 221	R*	
Lam69	AP	185 223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
Lam70	AP	178 223 288	M*	+10394*Dde*I; +10397*Alu*I
Lam71	AP	129 223 335	M4	+10394*Dde*I; +10397*Alu*I
Lam72	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lam73	AP	129 144TA 223 362	M4	+10394*Dde*I; +10397*Alu*I
Lam74	AP	129 223 304 311	I	+10237Hphl; +10394*Dde*I; +8249*Ava*II; –1715*Dde*I
Lam75	AP	304 311	R*	
Lam76	AP	185 189 223 270 274 319	M2b	+10394*Dde*I; +10397*Alu*I
Lam77	AP	223 292	W	+8249*Ava*II; –8994*Hae*III
Lam78	AP	111 189 223 327 330	M*	+10394*Dde*I; +10397*Alu*I
Lam79	AP	221	R*	
Lam80	AP	129 242 356	Ö	–00073*Alw*44I
Lam81	AP	223 311	M3	+10394*Dde*I; +10397*Alu*I
Lam82	AP	147CA 172 223 248 294 320 355	O	+10394*Dde*I; +10237*Hph*I; –1715*Dde*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
Lam83	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
Lam84	AP	148 189 223 270 274 319	M2b	+10394*Dde*I; +10397*Alu*I
Lam85	AP	92 223 362	M*	+10394*Dde*I; +10397*Alu*I
Lam86	AP	129 223 304 311	M*	+10394*Dde*I; +10397*Alu*I
Lob2	Pun	189 192 223 260 291 292 325 355	W	+8249*Ava*II; –8994*Hae*III
Lob3	Pun	185 223 289 311 362	M*	+10394*Dde*I; +10397*Alu*I
Lob4	Pun	93 223	M*	+10394*Dde*I; +10397*Alu*I
Lob5	Pun	223 263	M*	+10394*Dde*I; +10397*Alu*I
Lob6	Pun	51 234 247 304	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lob7	Pun	189 283 304	F	–12704*Hinc*II
Lob8	Pun	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Lob9	Pun	145 189 223 292 320	W	+8249*Ava*II; –8994*Hae*III
Lob10	Pun	223 289	M*	+10394*Dde*I; +10397*Alu*I
Lob11	Pun	223 234 304	M*	+10394*Dde*I; +10397*Alu*I
Lob12	Pun	223 263	M*	+10394*Dde*I; +10397*Alu*I
Lob14	Pun	223 293	M*	+10394*Dde*I; +10397*Alu*I
Lob17	Pun	223 292	W	+8249*Ava*II; -8994*Hae*III
Lob19	Pun	92 189 298 299	F	–12704*Hinc*II
Lob21	Pun	189 249	U1	+12308 *Hinf*I; –11465 *Tru*1I
Lob24	Pun	292	R*	
Lob26	Pun	126 294 296 304	T	+13366*Bam*HI; +15606*Alu*I
Lob27	Pun	172 223 362	M-D	+10394*Dde*I; +10397*Alu*I
Lob28	Pun	185 223 289 311 362	M*	+10394*Dde*I; +10397*Alu*I
Lob30	Pun	126 223 261 344	M1	+10394*Dde*I; +10397*Alu*I
Lob33	Pun	129 223 304	M4	+10394*Dde*I; +10397*Alu*I
Lob35	Pun	223 270 319 352	M2a	+10394*Dde*I; +10397*Alu*I
Lob36	Pun	223	M*	+10394*Dde*I; +10397*Alu*I
Lob38	Pun	126 223 261 344	M1	+10394*Dde*I; +10397*Alu*I
Lob40	Pun	126 223 261 344	M1	+10394*Dde*I; +10397*Alu*I
Lob41	Pun	223 290 292	W	+8249*Ava*II; –8994*Hae*III
Lob42	Pun	223	M*	+10394*Dde*I; +10397*Alu*I
Lob43	Pun	223 293	M*	+10394*Dde*I; +10397*Alu*I
Lob45	Pun	223	M*	+10394*Dde*I; +10397*Alu*I
Lob46	Pun	223	M*	+10394*Dde*I; +10397*Alu*I
Lob47	Pun	292	R*	
Lob52	Pun	223 231 234 311 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lob54	Pun	223 231 311 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lob55	Pun	292	R*	
Lob56	Pun	223 231 311 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lob57	Pun	51 206AC	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Lob58	Pun	223 231 311 356 362	M5	+10394*Dde*I; +10397*Alu*I
Lob59	Pun	223 256	M*	+10394*Dde*I; +10397*Alu*I
Lob60	Pun	129 189 223	M4	+10394*Dde*I; +10397*Alu*I
Lob62	Pun	292	R*	
Lob63	Pun	217	R*	
Lob64	Pun	48 129 218 223	M4	+10394*Dde*I; +10397*Alu*I
Lob65	Pun	223 290 292	W	+8249*Ava*II; -8994HaeIII
Lob67	Pun	126 294 296 304	T	+13366*Bam*HI; +15606*Alu*I
Lob68	Pun	223 289	M*	+10394*Dde*I; +10397*Alu*I
Lob69	Pun	145 189 223 292 320	W	+8249*Ava*II; –8994*Hae*III
Lob70	Pun	92 189 298 299	F	–12704*Hinc*II
Lob71	Pun	223 292	W	+8249*Ava*II; –8994*Hae*III
Lob72	Pun	129 223 311	M4	+10394*Dde*I; +10397*Alu*I
Lob74	Pun	71 189 278	R*	
Lob77	Pun	126 294 296 304	T	+13366*Bam*HI; +15606*Alu*I
Lob78	Pun	217	R*	
Lob79	Pun	92 189 298 299	F	–12704*Hinc*II
Lob81	Pun	92 189 298 299	F	–12704*Hinc*II
Lob86	Pun	292	R*	
Lob89	Pun	223 290 292	W	+8249*Ava*II; –8994*Hae*III
Lob91	Pun	71 93	R*	
Lob92	Pun	223 234 311	M*	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
Lob94	Pun	223 270 319	M2a	+10394*Dde*I; +10397*Alu*I
Lob95	Pun	223 270 319	M2a	+10394*Dde*I; +10397*Alu*I
Lob97	Pun	111 184 189 223 274 295	M*	+10394*Dde*I; +10397*Alu*I
Lob101	Pun	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Bog1	UP	124 179 189 223 249 294	M*	+10394*Dde*I; +10397*Alu*I
Bog2	UP	223 241	M*	+10394*Dde*I; +10397*Alu*I
Bog3	UP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
Bog4	UP	223	M*	+10394*Dde*I; +10397*Alu*I
Bog5	UP	129 223 311	M4	+10394*Dde*I; +10397*Alu*I
Bog6	UP	93 223	M*	+10394*Dde*I; +10397*Alu*I
Bog8	UP	95 223 249 359	M*	+10394*Dde*I; +10397*Alu*I
Bog10	UP	129 362	R*	
Bog12	UP	223 270 319 352	M2a	+10394*Dde*I; +10397*Alu*I
Bog14	UP	111CA 223	M*	+10394*Dde*I; +10397*Alu*I
Bog15	UP	refer	R*	
Bog17	UP	126 223 311	M1	+10394*Dde*I; +10397*Alu*I
Bog20	UP	129 189 223 325	M4	+10394*Dde*I; +10397*Alu*I
Bog21	UP	95 223 249 359	M*	+10394*Dde*I; +10397*Alu*I
Bog22	UP	140 189 223 293 311	M*	+10394*Dde*I; +10397*Alu*I
Bog23	UP	51 93TA 154 206AC 230 311	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Bog29	UP	126 176 181 209	pTJ	
Bog99	UP	92 126 223 286	M1	+10394*Dde*I; +10397*Alu*I
UP106	UP	175 223 234	M*	+10394*Dde*I; +10397*Alu*I
UP111	UP	51 129 209 239 291 325 352 353	U2i	+12308 *Hinf*I; –11465 *Tru*1I
UP183	UP	223	M*	+10394*Dde*I; +10397*Alu*I
Tha2	UP	51 223 298 327	M-C	+10394*Dde*I; +10397*Alu*I; +13262*Alu*I
Tha4	UP	114 223 294 318 362	M-D	+10394*Dde*I; +10397AluI; –5176*Alu*I
Tha5	UP	223 362	M-D	+10394*Dde*I; +10397*Alu*I; –5176*Alu*I
Tha15	UP	refer	R*	
Tha25	UP	93 126 163 186 189 294	T1	+13366*Bam*HI; +15606*Alu*I
Tha36	UP	71	R*	
Tha46	UP	223 302	M*	+10394*Dde*I; +10397*Alu*I
Tha47	UP	356	U4	+12308 *Hinf*I; –11465 *Tru*1I
Tha48	UP	129 189 241 266 304	R1	
Tha49	UP	126 223 368	M1	+10394*Dde*I; +10397*Alu*I
Tha50	UP	93 129 223	M4	+10394*Dde*I; +10397*Alu*I
Tha51	UP	51 209 239 352 353	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ksh1	Kash	189 304	F	–12704*Hinc*II
Ksh2	Kash	309 325	Ö	–00073*Alw*44I
Ksh3	Kash	refer	Ö	–00073*Alw*44I
Ksh4	Kash	189 304	F	–12704*Hinc*II
Ksh5	Kash	129 213 362	R*	
Ksh6	Kash	69 126 145 261	J	–13704*Bst*NI; +10394*Dde*I
Ksh7	Kash	189 304	F	–12704*Hinc*II
Ksh8	Kash	189 304	F	–12704*Hinc*II
Ksh9	Kash	309 325 362	Ö	–00073*Alw*44I
Ksh10	Kash	148 175 223 292	W	+8249*Ava*II; –8994*Hae*III
Ksh11	Kash	refer	H	–00073*Alw*44I; –7025*Alu*I
Ksh12	Kash	51 206AC 291	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ksh13	Kash	145 176 223 311	M3	+10394*Dde*I; +10397*Alu*I
Ksh14	Kash	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Ksh15	Kash	145 239 241 325	R*	
Ksh16	Kash	51 209 239 352 353	U2i	+12308 *Hinf*I; –11465 *Tru*1I
Ksh17	Kash	192 223 300 316	M*	+10394*Dde*I; +10397*Alu*I
Ksh18	Kash	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
APx	AP	223 274 311 319	M2b	+10394*Dde*I; +10397*Alu*I
BV10	AP	146 311	H	–7025*Alu*I
BV12	AP	179 227 245 266 278 362	R*	
BV14	AP	51 234		+10394*Dde*I; +10397*Alu*I
BV15	AP	223 257	M*	+10394*Dde*I; +10397*Alu*I
BV16	AP	126 223 311		+12308 *Hinf*I
BV19	AP	126 223 271	M1	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
BV2	AP	51 172 286 291 206AC	U2i	+12308 *Hinf*I
BN22	AP	189 197CA 223 287 327 330	M*	+10394*Dde*I; +10397*Alu*I
BN24	AP	129 223 311	M4	+10394*Dde*I; +10397*Alu*I
BV25	AP	93 224 311	U-K	P+12308 *Hinf*I;+10394*Dde*I; –9025*Hae*II
BN26	AP	126 266 304 309 325 356	R	
BN28	AP	129 223 286 291	M4	+10394*Dde*I; +10397*Alu*I
BV29	AP	69 274 280 318AT	U7	+12308 *Hinf*I
BN30	AP	223	M*	+10394*Dde*I; +10397*Alu*I
BN31	AP	172 223 243 270 319 352	M2a	+10394*Dde*I; +10397*Alu*I
BV32	AP	223 324 362	M*	+10394*Dde*I; +10397*Alu*I
BV33	AP	223 274	M*	+10394*Dde*I; +10397*Alu*I
BV34	AP	129 223	M4	+10394*Dde*I; +10397*Alu*I
BV35	AP	126 223 311	M1	+10394*Dde*I; +10397*Alu*I
BV36	AP	104 189 223 243 319 362	M*	+10394*Dde*I; +10397*Alu*I
BV38	AP	178 223 256	M*	+10394*Dde*I; +10397*Alu*I
BV4	AP	223	M*	+10394*Dde*I; +10397*Alu*I
BV40	AP	51 172 206AC	U2i	+12308 *Hinf*I
BV42	AP	145 176 223 261 311	M3	+10394*Dde*I; +10397*Alu*I
BV43	AP	304 311	R*	
BV44	AP	207 309 318AT	U7	+12308 *Hinf*I
BV45	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
BV47	AP	223 240AC 274 311 319	M2b	+10394*Dde*I; +10397*Alu*I
BV48	AP	37 51 234 325	U2i	+12308 *Hinf*I
BV49	AP	207 309 318AT 352	U7	+12308 *Hinf*I
BV5	AP	126 223 271	M1	+10394*Dde*I; +10397*Alu*I
BN50	AP	184 214 357	R*	
BN51	AP	223 274 319	M2b	+10394*Dde*I; +10397*Alu*I
BN52	AP	172 278		+10394*Dde*I; +10397*Alu*I
BN53	AP	126 185 223		
BN54	AP	311 320		+10394*Dde*I; +10397*Alu*I
BN55	AP	54AT 223 325		–7025*Alu*I
BN56	AP	223 325	M*	+10394*Dde*I; +10397*Alu*I
BN57	AP	93 223 278	M*	+10394*Dde*I; +10397*Alu*I
BN9	AP	223 304	M*	+10394*Dde*I; +10397*Alu*I
J11	AP	223 256 294	M*	+10394*Dde*I; +10397*Alu*I
J13	AP	51 86 209 239 354 362	U2i	+12308 *Hinf*I
J15	AP	92 145 223 261 311	M3	+10394*Dde*I; +10397*Alu*I
J21	AP	92 145 223 261 311	M3	+10394*Dde*I; +10397*Alu*I
J24	AP	223	M*	+10394*Dde*I; +10397*Alu*I
J25	AP	93 104 234 243 244	M*	+10394*Dde*I; +10397*Alu*I
J26	AP	93 104 108 234 243 244	M*	+10394*Dde*I; +10397*Alu*I
KT1	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
KT10	AP	126 294 296 325	T	
KT11	AP	184 189 223 288TA 300	M*	+10394*Dde*I; +10397*Alu*I
KT12	AP	223 274 301	M*	+10394*Dde*I; +10397*Alu*I
KT13	AP	51 93TA 154 206AC 230 311	U2i	+12308 *Hinf*I
KT14	AP	223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
KT15	AP	189 223 270 278	M*	+10394*Dde*I; +10397*Alu*I
KT16	AP	126 223 309	M1	+10394*Dde*I; +10397*Alu*I
KT17	AP	188 189 223 231 355 362	M5	+10394*Dde*I; +10397*Alu*I
KT2	AP	129 213 319 362	R*	
KTK22	AP	42 51 93 179 234 240AC	U2i	+12308 *Hinf*I
KTK23	AP	69 126 145 172 222 261	J	
KTK24	AP	refer	U*	+12308 *Hinf*I
KTK25	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
KTK27	AP	266 304 311 356	R1	
KTK28	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
KTK30	AP	51 93TA 154 206AC 230 311	U2i	+12308 *Hinf*I
KT32	AP	223 263	M*	+10394*Dde*I; +10397*Alu*I
KTK33	AP	223 270 274 292 319 352	M2b	+10394*Dde*I; +10397*Alu*I
KTK34	AP	refer	H	–7025*Alu*I
KTK35	AP	292	R*	

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
KTK36	AP	129 223	M4	+10394*Dde*I; +10397*Alu*I
KT38	AP	51 189 223 274 319 320 362	M2b	+10394*Dde*I; +10397*Alu*I
KT39	AP	172 278 344CA	R*	
KTK4	AP	223 304	M*	+10394*Dde*I; +10397*Alu*I
KT40	AP	189 223 300	M*	+10394*Dde*I; +10397*Alu*I
KT41	AP	179 223 294 319 356	M*	+10394*Dde*I; +10397*Alu*I
KTK42	AP	51 230 311 206AC	U2i	+12308 *Hinf*I
KT43	AP	93 223 258AC 274	M*	+10394*Dde*I; +10397*Alu*I
KT44	AP	129 223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
KT45	AP	129 223 264 265AC	M4	+10394*Dde*I; +10397*Alu*I
KTK46	AP	223	M*	+10394*Dde*I; +10397*Alu*I
KT47	AP	86 221 278	R*	
KT48	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
KT49	AP	223 234	M*	+10394*Dde*I; +10397*Alu*I
KTK5	AP	93 189 223 232 270 311	M*	+10394*Dde*I; +10397*Alu*I
KT50	AP	37 51 154 206AC 230 311	U2i	+12308 *Hinf*I
KT51	AP	126 223 312	M1	+10394*Dde*I; +10397*Alu*I
KTK52	AP	189 192 223 300	M*	+10394*Dde*I; +10397*Alu*I
KT54	AP	223 234	M*	+10394*Dde*I; +10397*Alu*I
KT55	AP	223	M*	+10394*Dde*I; +10397*Alu*I
KTK57	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
KT59	AP	119 193 223 303GT	M*	+10394*Dde*I; +10397*Alu*I
KTK6	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
KTK60	AP	260 261 319 362	R*	
KTK64	AP	111 223	M*	+10394*Dde*I; +10397*Alu*I
KTK66	AP	111 223	M*	+10394*Dde*I; +10397*Alu*I
KT67	AP	274	R*	
KT68	AP	92TA 145 223 227 245 290CA 291		
KTK69	AP	75 223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
KT8	AP	304 311	R*	
KT9	AP	169 172 223	M*	+10394*Dde*I; +10397*Alu*I
KS1	AP	223	M*	+10394*Dde*I; +10397*Alu*I
KS10	AP	179 223 294	M*	+10394*Dde*I; +10397*Alu*I
KS11	AP	126 223 265	M1	+10394*Dde*I; +10397*Alu*I
KS2	AP	270	R*	
KS3	AP	270	R*	
KS4	AP	270	R*	
KS7	AP	223	M*	+10394*Dde*I; +10397*Alu*I
KS8	AP	179 223 294	M*	+10394*Dde*I; +10397*Alu*I
KS9	AP	129 223 311	M4	+10394*Dde*I; +10397*Alu*I
M11	AP	126 181 209 362	R*	
M12	AP	188 189 223 256 311	M*	+10394*Dde*I; +10397*Alu*I
M13	AP	104 223 234 243 244	M*	+10394*Dde*I; +10397*Alu*I
M14	AP	129 223 291	M4	+10394*Dde*I; +10397*Alu*I
M15	AP	129 223 291	M4	+10394*Dde*I; +10397*Alu*I
M16	AP	172 223 270 274 319 352	M2b	+10394*Dde*I; +10397*Alu*I
M17	AP	172 278	R*	
M18	AP	129 223 291	M4	+10394*Dde*I; +10397*Alu*I
M19	AP	189 223 248 274 291 319 320	M2b	+10394*Dde*I; +10397*Alu*I
M2	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
M21	AP	223 301	M*	+10394*Dde*I; +10397*Alu*I
M22	AP	223 231 362	M5	+10394*Dde*I; +10397*Alu*I
M23	AP	93 145 189 223 290 312 355 381	M*	+10394*Dde*I; +10397*Alu*I
M24	AP	129 223 264 265AC	M4	+10394*Dde*I; +10397*Alu*I
M25	AP	129 223 264 265AC	M4	+10394*Dde*I; +10397*Alu*I
M26	AP	221	R*	
M27	AP	126 223	M1	+10394*Dde*I; +10397*Alu*I
M29	AP	75 92 189 223 270	M*	+10394*Dde*I; +10397*Alu*I
M3	AP	37 51 234 283AC	U2i	+12308 *Hinf*I
M30	AP	266 289 304	R1	
M31	AP	75 92 189 223 270	M*	+10394*Dde*I; +10397*Alu*I
M32	AP	223	M*	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
M4	AP	304 311	R*	
M7	AP	145 189 192 266 304 309 325 356	R1	
M8	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
M9	AP	126 163 186 189 294	T1	
ML1	AP	124 179 209 223 294 311 319 356	M*	+10394*Dde*I; +10397*Alu*I
ML12	AP	223 324	M*	+10394*Dde*I; +10397*Alu*I
ML13	AP	184 223	M*	+10394*Dde*I; +10397*Alu*I
ML14	AP	93 134 223 318AT	M*	+10394*Dde*I; +10397*Alu*I
ML15	AP	193 223 256	M*	+10394*Dde*I; +10397*Alu*I
ML16	AP	129 179 189 223 292 294 355	M*	+10394*Dde*I; +10397*Alu*I
ML17	AP	129 223	M4	+10394*Dde*I; +10397*Alu*I
ML18	AP	309 318AT	U7	+12308 *Hinf*I
ML19	AP	172 223	M*	+10394*Dde*I; +10397*Alu*I
ML2	AP	189 223 327 330	M*	+10394*Dde*I; +10397*Alu*I
ML20	AP	136 248 266 304 325 356	R1	
ML21	AP	223 234 318AT	M*	+10394*Dde*I; +10397*Alu*I
ML23	AP	223 304	M*	+10394*Dde*I; +10397*Alu*I
ML24	AP	51 114 193 278 357	M*	+12308 *Hinf*I
ML25	AP	223 231 356 362	M5	+10394*Dde*I; +10397*Alu*I
ML26	AP	223	M*	+10394*Dde*I; +10397*Alu*I
ML27	AP	116AC 169 234 283 317AT 351AT	R*	
ML3	AP	256 266 304 356	R1	
ML4	AP	145 185 189 239 325	R*	
ML5	AP	136 248 266 304 325 356	R1	
ML6	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
ML7	AP	266 304 311	R1	
ML8	AP	223	M*	+10394*Dde*I; +10397*Alu*I
ML9	AP	189 223 270 287 296 318AT	M*	+10394*Dde*I; +10397*Alu*I
R1	AP	189 218 223 274 319 320	M2b	+10394*Dde*I; +10397*Alu*I
R10	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
R11	AP	189 223 228CG 242 274 319 320 355	M2b	+10394*Dde*I; +10397*Alu*I
R12	AP	189 218 223 274 319 320	M2b	+10394*Dde*I; +10397*Alu*I
R13	AP	147 189 223 243 278 355 362	M*	+10394*Dde*I; +10397*Alu*I
R15	AP	93 292	R*	
R16	AP	178 223 288	M*	+10394*Dde*I; +10397*Alu*I
R17	AP	178 223 288	M*	+10394*Dde*I; +10397*Alu*I
R18	AP	93 172 223 327	M*	+10394*Dde*I; +10397*Alu*I
R19	AP	223 274 319 357	M2b	+10394*Dde*I; +10397*Alu*I
R2	AP	93 301 317AT	R*	
R20	AP	93 292	R*	
R21	AP	178 223 288	M*	
R3	AP	209 223 156GT	M*	+10394*Dde*I; +10397*Alu*I
R4	AP	223 318AT	M*	+10394*Dde*I; +10397*Alu*I
R5	AP	189 218 223 274 319 320 228CG	M2b	+10394*Dde*I; +10397*Alu*I
R6	AP	51 247 254 362	U2i	+12308 *Hinf*I
R7	AP	156GT 184 189 223 316AT	M*	+10394*Dde*I; +10397*Alu*I
R8	AP	156GT 223 318AT	M*	+10394*Dde*I; +10397*Alu*I
R9	AP	48 63 129 223 362	M4	+10394*Dde*I; +10397*Alu*I
VS1	AP	223 292	W	
VS10	AP	117 126 223 278	M1	+10394*Dde*I; +10397*Alu*I
VS2	AP	126 223 278	M1	+10394*Dde*I; +10397*Alu*I
VS3	AP	51	U2i	+12308 *Hinf*I
VS4	AP	292	R*	
VS5	AP	51	U2i	+12308 *Hinf*I
VS6	AP	126 223 278	M1	+10394*Dde*I; +10397*Alu*I
VS7	AP	126 223 278	M1	+10394*Dde*I; +10397*Alu*I
VS8	AP	51	U2i	+12308 *Hinf*I
VS9	AP	126 223 278	M1	+10394*Dde*I; +10397*Alu*I
WB15	AP	172 304	R*	
WB17	AP	104 223 234 243 244	M*	+10394*Dde*I; +10397*Alu*I
WB18	AP	223	M*	
WB19	AP	92 145 223 261 311	M3	+10394*Dde*I; +10397*Alu*I

Table S1 cont.

Sample number	Origin	HVS-I sequence polymorphisms*	Haplogroup	RFLP polymorphisms*
WB20	AP	266 304 311 355 356		+10394*Dde*I; +10397*Alu*I
WB21	AP	155AT 356	R*	
WB22	AP	104 223 234 243 244		
WB23	AP	104 223 234 243 244	M*	+10394*Dde*I; +10397*Alu*I
Y1	AP	184 223 311	M3	+10394*Dde*I; +10397*Alu*I
Y11	AP	192 223 231 356 362	M5	+10394*Dde*I; +10397*Alu*I
Y15	AP	126 275 294 296 325	T	
Y17	AP	51 172 209 224 239 352 353	U2i	+12308 *Hinf*I
Y18	AP	126 223 311	M1	+10394*Dde*I; +10397*Alu*I
Y19	AP	129 242 356	R*	
Y2	AP	129 193 223	M4	+10394*Dde*I; +10397*Alu*I
Y20	AP	189 222 223 256 274 319 320	M2b	+10394*Dde*I; +10397*Alu*I
Y21	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Y22	AP	129 193 223	M4	+10394*Dde*I; +10397*Alu*I
Y23	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Y24	AP	193 223	M*	+10394*Dde*I; +10397*Alu*I
Y25	AP	129 242 356	R*	
Y26	AP	93 223	M*	+10394*Dde*I; +10397*Alu*I
Y27	AP	86 172 189 223 227 278 362	M-E	+10394*Dde*I; +10397*Alu*I; –7598*Hha*I
Y28	AP	223 304	M*	+10394*Dde*I; +10397*Alu*I
Y29	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Y3	AP	223 304	M*	+10394*Dde*I; +10397*Alu*I
Y30	AP	189 231 270 319 362	R*	
Y31	AP	129 242 356	R*	
Y32	AP	104 223 234 243 244 354	M*	+10394*Dde*I; +10397*Alu*I
Y33	AP	172 189 228CG 270 278 296 355	R*	
Y34	AP	126 294 296 304	T	
Y35	AP	309 318 AT	U7	+12308 *Hinf*I
Y36	AP	51 234	U2i	+12308 *Hinf*I
Y37	AP	193 223	M*	+10394*Dde*I; +10397*Alu*I
Y38	AP	188 223 231 362	M5	+10394*Dde*I; +10397*Alu*I
Y39	AP	145 221 260 261 311 319 343 362	R*	
Y4	AP	223	M*	+10394*Dde*I; +10397*Alu*I
Y40	AP	213 223 231 356 362	M5	+10394*Dde*I; +10397*Alu*I
Y41	AP	304 311	R*	
Y42	AP	223 270 362 381	M*	+10394*Dde*I; +10397*Alu*I
Y43	AP	266 304 310 311 356	R1	
Y44	AP	221	R*	
Y45	AP	172 278	R*	
Y46	AP	126 223	M1	+10394*Dde*I; +10397*Alu*I
Y47	AP	184 223 311	M3	+10394*Dde*I; +10397*Alu*I
Y49	AP	129 193 223	M4	+10394*Dde*I; +10397*Alu*I
Y51	AP	51 126 179 227 234 240AC	U2i	+12308 *Hinf*I
Y52	AP	223 270 319 352	M2a	+10394*Dde*I; +10397*Alu*I
Y53	AP	93 223 243	M*	+10394*Dde*I; +10397*Alu*I
Y54	AP	189 212 223 228CG 270 327 330	M*	+10394*Dde*I; +10397*Alu*I
Y55	AP	86 172 189 223 227 278 362	M-E	+10394*Dde*I; +10397*Alu*I; –7598*Hha*I
Y56	AP	126 275 294 296 325 357	T	
Y6	AP	129 193 223	M4	+10394*Dde*I; +10397*Alu*I
Y7	AP	129 213 320 362	R*	
Y9	AP	250 260 261 293 311 319	R*	

All mtDNAs were compared to the R node which equals [S14] in HVS-I sequence but differs from it by the presence of an *Alw*44I site at nucleotide position (np) 00073 and the presence of an *Alu*I site at np 7025. Origin of the samples: AP, Andhra Pradesh; Bang, Bangladesh; Ben, Western Bengal; Bih, Bihar; Guj, Gujarat; Hary, Haryana; Kash, Kashmir; Mah, Maharashtra; Ori, Orissa; Pak, Pakistan; Pun, Punjab; Raj, Rajasthan; Tamil, Tamil-Nadu; UP, Uttar Pradesh. For those data sets, published by others, with which the Indian data were compared, see Supplementary material and methods section.