The Dazzling Array of Basal Branches in the mtDNA Macrohaplogroup M from India as Inferred from Complete Genomes

Chang Sun,†‡[1] Qing-Peng Kong,*†‡[1] Malliya gounder Palanichamy,†[1] Suraksha Agrawal,§[1] Hans-Jürgen Bandelt,‖ Yong-Gang Yao,* Faisal Khan,§ Chun-Ling Zhu,* Tapas Kumar Chaudhuri,¶ and Ya-Ping Zhang*†*

*Laboratory of Cellular and Molecular Evolution, and Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; †Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China; ‡Graduate School of the Chinese Academy of Sciences, Beijing, China; §Department of Medical Genetics, Sanjay Gandhi Institute of Medical Sciences, Lucknow, India; ‖Department of Mathematics, University of Hamburg, Hamburg, Germany; and ¶Department of Zoology, North Bengal University, Siliguri West Bengal, India

Many efforts based on complete mitochondrial DNA (mtDNA) genomes have been made to depict the global mtDNA landscape, but the phylogeny of Indian macrohaplogroup M has not yet been resolved in detail. To fill this lacuna, we took the same strategy as in our recent analysis of Indian mtDNA macrohaplogroup N and selected 56 mtDNAs from over 1,200 samples across India for complete sequencing, with the intention to cover all Indian autochthonous M lineages. As a result, the phylogenetic status of previously identified haplogroups based on control-region and/or partial coding-region information, such as M2, M3, M4, M5, M6, M30, and M33, was solidified or redefined here. Moreover, seven novel basal M haplogroups (viz., M34–M40) were identified, and yet another five singular branches of the M phylogeny were discovered in the present study. The comparison of matrilineal components among India, East Asia, Southeast Asia, and Oceania at the deepest level yielded a star-like and nonoverlapping pattern, reflecting a rapid mode of modern human dispersal along the Asian coast after the initial "out-of-Africa" event.

Introduction

The emergence of (nearly) complete mitochondrial DNA (mtDNA) sequences has made it possible to reconstruct the phylogenies of European, African, Oceanian, East Asian, Southeast Asian, and Indian N lineages and to gain detailed insight into the past of modern humans (Ingman et al. 2000; Finnilä, Lehtonen, and Majamaa 2001; Maca-Meyer et al. 2001, 2003; Torroni et al. 2001; Derbeneva et al. 2002; Herrnstadt et al. 2002; Herrnstadt, Preston, and Howell 2003; Ingman and Gyllensten 2003; Kong et al. 2003; Mishmar et al. 2003; Reidla et al. 2003; Achilli et al. 2004, 2005; Howell et al. 2004; Palanichamy et al. 2004; Tanaka et al. 2004; Friedlaender et al. 2005; Macaulay et al. 2005; Merriwether et al. 2005; Thangaraj et al. 2005). However, such systematic analyses have not yet been available for Indian macrohaplogroup M, which is ubiquitous in South Asia and covers more than half of the Indian mtDNA samples (Kivisild et al. 1999, 2003; Metspalu et al. 2004; Quintana-Murci et al. 2004). A recent study (Rajkumar et al. 2005) claimed to have resolved this problem but described only a few Indian M haplogroups. Furthermore, their reported sequences bear the imprints of typical problems of large mtDNA sequencing attempts (Yao et al. 2003; Bandelt et al. 2005). They are, therefore, only of limited help for understanding the Indian M phylogeny.

Based on control-region variation and/or partial coding-region information, previous studies have identified some autochthonous Indian M haplogroups, such as M2, M3, M4, M5, M6 (Kivisild et al. 1999, 2003; Bamshad et al. 2001; Basu et al. 2003) and, more recently M18,

M25, M30, and M33 (Metspalu et al. 2004; Rajkumar et al. 2005; Thangaraj et al. 2005). Most of these haplogroups have not been fully characterized due to limited information available at the time. Moreover, some of them (e.g., M3, M4, and M5) were exclusively defined by "speedy" mutations, such as transitions at sites 16126, 16129, and 16311 (Bandelt et al. 2002), and consequently their monophyletic status has been questioned (Kivisild et al. 2003; Metspalu et al. 2004). In addition, many mtDNAs in our large Indian data set could not be assigned to the recognized haplogroups, pointing to the existence of further basal lineages. To resolve these problems, a systematic study on Indian M haplogroups based on complete sequencing would be indispensable.

Two recent reports provided complete mtDNA genome data from "relict" South(east) Asian populations and deduced the initial migration pattern of modern humans after their exodus from Africa (Macaulay et al. 2005; Thangaraj et al. 2005). In consideration of the pivotal geographic location of India along the "southern route" adopted by modern humans (Stringer 2000) as well as the facts that the offsprings of the earliest settlers composed the most components of Indian mtDNA gene pool and that the recent gene flow into and out of India was quite limited (cf., Kivisild et al. 2003; Metspalu et al. 2004), more Indian complete mtDNA sequences, especially from autochthonous macrohaplogroup M, would be helpful for further elucidating the scenario of a rapid dispersal of modern human along the Asian coast after the initial "out-of-Africa" event (Macaulay et al. 2005; Forster and Matsumura 2005).

In the present study, 56 mtDNAs were selected from our large Indian data collection for complete sequencing, which covered all the recognized M lineages (with the exception of M25 whose distribution frequency is rather low in South Asia; Metspalu et al. 2004) as well as many uncharacterized lineages observed in India. Our results,

[1] These authors contributed equally to this work.

Key words: mitochondrial DNA, India, haplogroup M, complete sequence, phylogeny.

E-mail: zhangyp1@263.net.cn.

together with our recent report about Indian marcohaplogroup N (Palanichamy et al. 2004), provide a comprehensive picture of the Indian matrilineal gene pool. The distinctiveness of the South Asian mtDNA pool in comparison to East Asia and Oceania is further substantiated.

Materials and Methods
Sampling

Over 1,200 Indian samples were collected from the following populations: R—Reddy, T—Thogataveera from Andhrapradesh, South India; A—Bhargava, B—Chaturvedi, C—other Brahmin from Uttarpradesh, North India; SW—Rajbhansi from West Bengal, and S—Khasi population from Meghalaya, both located in Northeast India. Based on the information from the control region and some coding-region segments, each mtDNA was assigned into the known haplogroups (Kivisild et al. 1999, 2003; Bamshad et al. 2001; Basu et al. 2003; Palanichamy et al. 2004) or novel M lineages (data not shown). To cover most of Indian autochthonous M lineages, a total of 56 mtDNAs were selected for complete sequencing, with the purpose that at least one representative was chosen from the known haplogroup and the newly identified novel lineage.

DNA Amplification, Sequencing, and Quality Control

The mtDNA genomes were amplified and sequenced by means of the procedures described in our recent studies (Kong et al. 2003; Palanichamy et al. 2004), including the same data quality assessment, such as at least two independent amplifications. Sequences were edited and aligned by DNASTAR software (DNAStar Inc., Madison, Wisc.), and mutations were scored relative to the revised Cambridge reference sequence (rCRS; Andrews et al. 1999).

Phylogeny Reconstruction and Age Estimation

Besides our newly collected 56 mtDNAs, 10 additional Indian M complete genomes from the literatures (Maca-Meyer et al. 2001; Ingman and Gyllensten 2003; Kivisild et al. 2006) were employed for the tree reconstruction. To distinguish the latter samples from ours, the abbreviations NM, IG, and TK, respectively, representing the aforementioned papers were listed before the original sample name. The phylogeny was reconstructed manually and compared with several (combined) runs of the reduced-median and median-joining algorithms with NETWORK 4.1.0.9 (employing some differential weighting and different parameter settings). In thus exploring the solution space of slightly less parsimonious trees, we found that in particular the seeming 16519 clade appears to be unstable, and the M4a lineage (sample R44) could well belong to a potential clade supported by the 593 transition. We followed the nomenclature system put forward by Richards et al. (1998) and Macaulay et al. (1999), as well as the suggestions in Palanichamy et al. (2004) for the naming of new emerging branches. The available mtDNA complete sequences from India, Southeast Asia, East Asia, and Oceania (Ingman et al. 2000; Maca-Meyer et al. 2001; Ingman and Gyllensten 2003; Palanichamy et al. 2004; Friedlaender et al. 2005;

Kivisild et al. 2005; Macaulay et al. 2006; Merriwether et al. 2005; present study) were used to estimate the age of macrohaplogroup M for each geographic region separately, according to the previously described approach (Kong et al. 2003; Mishmar et al. 2003). The Indian M data reported by Rajkumar et al. (2005) were excluded from these analyses because they bear a number of potential errors (see *Discussion*).

Results
The Identification of Novel M Lineages

As shown in figure 1, seven novel M haplogroups and another five yet singular branches were identified in the present study. In detail, the two mtDNAs, C39 and C56, share six specific mutations (at sites 569, 3010, 6794, 11101, 15865, and 16249) and form a new haplogroup M34. Haplogroup M35 is recognized by two mutations (at sites 199 and 12561), and its major subhaplogroup, M35a, is characterized by another five mutations. Haplogroup M36 is defined by three specific mutations (at sites 239, 7271, and 15110). Samples C26 and R45 share mutations 12007 and 10556 and therefore belong to a novel haplogroup, M37. Samples A24 and T72 constitute another new branch of M, designated as M38, that is supported by nine mutations, among which the 12007 and 246 transitions are in common with haplogroup M18. Note that haplogroups M4, M18, M30, M37, and M38 share the 12007 transition and are thus included in a super-branch nested in M, which is named as M4'30. Haplogroup M39 is characterized by a pronounced hypervariable segment II motif (55+T, 65+T, and 66T) and three coding-region mutations (at sites 1811, 8679, and 15938). Finally, haplogroup M40 has a specific motif composed of four mutations (at 8925, 15721, 15954, and 16463).

Refinement of Previous Haplogroup Definitions

The mtDNA complete sequence information offers us an opportunity to test the phylogenetic status of the haplogroups that were previously recognized only on the basis of control-region and partial coding-region information. The definitions of M3 and M33, for instance, are well supported by the complete sequence information and remain unchanged (Kivisild et al. 1999; Bamshad et al. 2001; Metspalu et al. 2004; Thangaraj et al. 2005). But the definition of other haplogroups needs to be revised. For example, as mutation 16274 was present in all M2 samples of our data set (our unpublished data) and in most of the M2 samples reported elsewhere (Kivisild et al. 2003; Metspalu et al. 2004), we treat this mutation as a basal polymorphism of M2, so that the lack of the 16274 variant in some M2a samples (Kivisild et al. 2003) would be best explained as a back-mutation event. Haplogroup M5, previously identified by the speedy 16129 mutation (Kivisild et al. 1999; Bamshad et al. 2001), is substantiated by an accompanying coding-region mutation (at site 1888). M5a, the major branch of haplogroup M5, is also identified by another three mutations (viz., 709, 3921, and 14323), in addition to 12477 (Rajkumar et al. 2005). Haplogroup M6 is characterized by additional eight mutations (at sites 461, 5082, 5301, 5558, 9329,
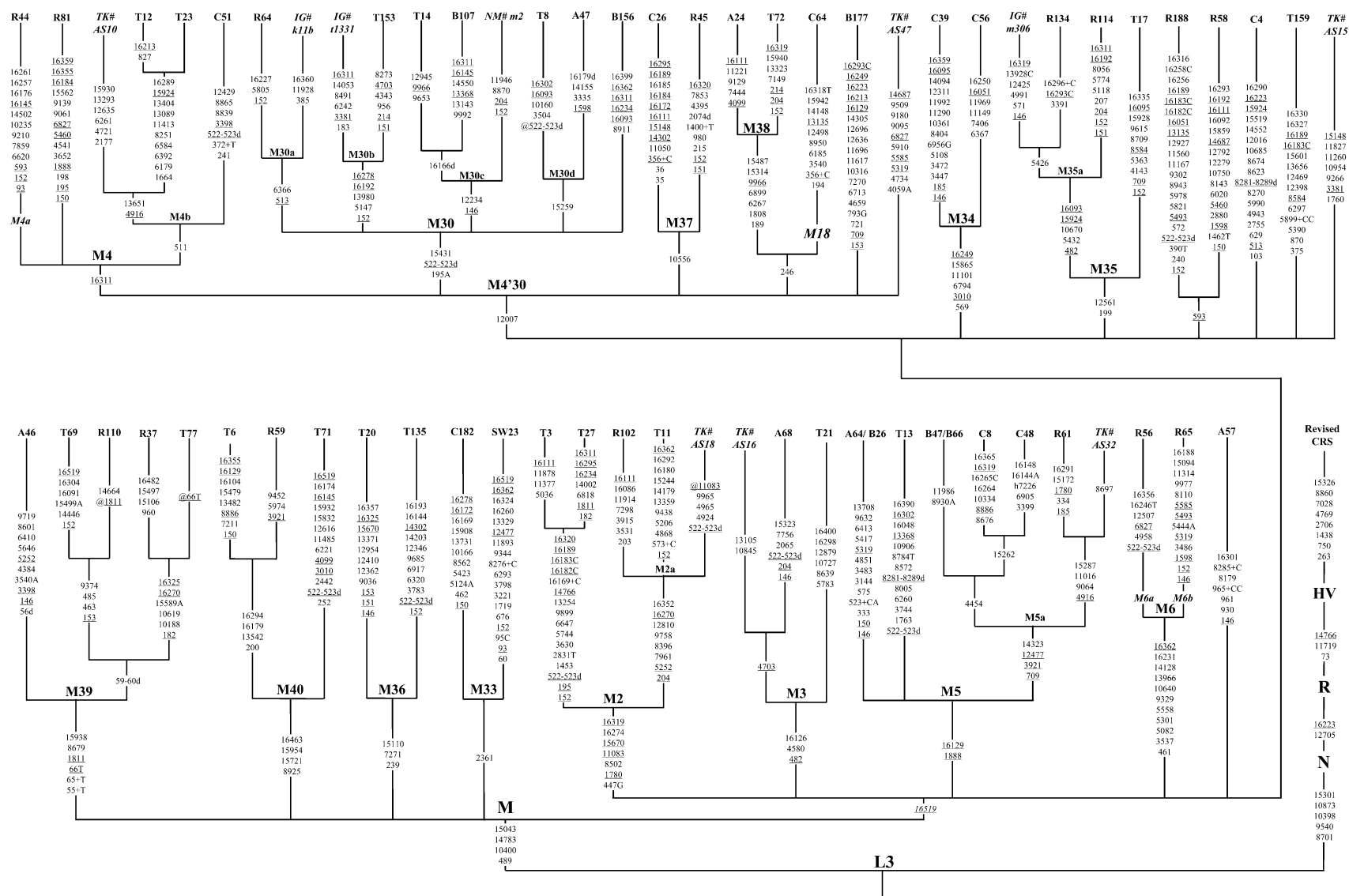
Fig. 1.—The phylogenetic tree of 66 Indian autochthonous mtDNAs from macrohaplogroup M. Mutations are scored relative to the rCRS (Andrews et al. 1999). Samples were taken from the following populations: Bhargava (A), Chaturvedi (B), Brahmin (C) from Uttarpradesh, North India; Reddy (R), Thogataveera (T) from Andhrapradesh, South India; Rajbhansi (SW) from West Bengal, Northeast India. Ten additional samples were collected from published sources (Maca-Meyer et al. 2001; Ingman and Gyllensten 2003; Kivisild et al. 2005) and were referred to by symbols NM, IG, and TK, respectively, followed by "#" and the original sample code. Suffixes A, C, G, and T indicate transversions, "d" signifies a deletion and a plus sign (+) an insertion; recurrent mutations are underlined. The prefix "h" indicates heteroplasmy and "@" highlights back mutation. The motifs of haplogroups in italics (e.g., M6b) are not yet determined. The reconstruction of (in general) highly recurrent mutations (e.g., 146, 150, 152, 195, 16182C, 16183C, 16093, 16129, 16189, 16311, 16362, 16519, and the insertion/deletion of "CA" repeats in region 514–523) is tentative at best.

Table 1
Age Estimation of Haplogroup M

Subcontinent	Sample Size (n)	Distance ($\rho \pm \sigma$)	Age ($\times 10^3$ years)
South Asia	66	8.7 ± 0.6	44.6 ± 3.3
East Asia[a]	27	13.5 ± 1.1	69.3 ± 5.4
Oceania	34	14.2 ± 1.5	73.0 ± 7.9
Southeast Asia	6	10.8 ± 1.4	55.7 ± 7.4

[a] From Kong et al. (2003).

10640, 13966, and 14128) besides the three previously reported ones (viz., 3537, 16231, and 16362; Kivisild et al. 1999; Bamshad et al. 2001; Metspalu et al. 2004).

Because the previously defined haplogroups M4 and M18 would reside in "M30" as proposed by Rajkumar et al. (2005), one would be forced to change the established older nomenclature. Rather, to avoid further confusion, we suggest here to narrow the definition of M30 to represent the branch (viz., the former M30a in Rajkumar et al. [2005]) that is identified by mutations 195A and 15431 besides 12007.

Haplogroup Age Estimation

The ages of macrohaplogroup M in India, Southeast Asia, East Asia, and Oceania are displayed in table 1. With the exception of M in India, the estimated ages are not significantly different and match the postulated age (60×10^3 to 70×10^3 years) of the three Eurasian founder types (Kong et al. 2003; Macaulay et al. 2005). The abnormal younger age ($\sim 44 \times 10^3$ years) of the Indian M lineages may be attributed to the unequal age contributions from different M haplogroups to the total M age estimate. The frequency composition of the particular M haplogroups in a population sample from India then matters a lot because the ages contributed by the different branches range from 21×10^3 years (M30a) to 93×10^3 years (M6b). Our sample drawn for complete sequencing does not reflect the natural frequencies and therefore could bias the age estimation.

To investigate this, we simulated a natural frequency distribution by assigning the mtDNAs sampled from the Chenchu and Koya populations (Kivisild et al. 2003) to their respective haplogroups by (near) matching control-region motifs (Yao et al. 2002). Fewer than 5% of the mtDNAs were virtually unassignable, so that the remaining 142 mtDNAs could be used to evaluate the natural frequency of each M branch in Indian population (table 2). Based on this frequency information, the M age was recalculated by using the complete sequence data. The resulting age, 54.1×10^3 years, is indeed considerably larger than the value obtained for an unweighted estimation and falls into the range of the age estimates from the other geographical regions (see table 1).

The coalescence times of the Indian M haplogroups themselves vary substantially: for those haplogroups represented by more than three mtDNAs in our sample, the age estimates range from 24×10^3 years (M39) to 54×10^3 years (M2). The superhaplogroup, M4'30, has an estimated age of 33×10^3 years. It seems that at the time when the interior of the Indian subcontinent was populated by modern humans ($\sim 30 \times 10^3$ to $\sim 45 \times 10^3$ years), the root

Table 2
Age of Haplogroup M Estimated from the Frequency-Based Composition of M Haplogroups Derived from the Chenchu and Koya Populations

Haplogroup	Haplotype[a]	Frequency	Distance ρ Contribution[b]	Age Contribution ($\times 10^3$ years)
M2a	#20–29	29	13.7	70.2
M2b	#30–32	3	15.0	77.1
M3	#33–37	6	4.7	24.0
M5*	#11	17	9.0	46.3
M5a	#9–10	8	9.5	48.8
M6b	#38	18	18.0	92.5
M18	#13	4	8.0	41.1
M33	#17–19	12	8.5	43.7
M35a	#6	12	7.0	36.0
M38	#7	1	10.5	54.0
M39	#3–5	27	7.6	39.1
M40 (T6/R59)	#12	3	7.5	38.6
M* (T159)	#14	2	8.0	41.1
Total		142	10.5	54.1

[a] From Kivisild et al. (2003).
[b] Calculation based on the complete sequences from this study.

haplotype of M and its one-step derivative, M4'30, were still present and gave rise to numerous haplogroups that seem to deflate the age of macrohaplogroup M, especially when the sampling scheme (as in our case) favored these hitherto poorly characterized subhaplogroups. This staggered scenario would best agree with a two-stage process for the peopling of the Indian subcontinent as was suggested by Kivisild et al. (2003).

Discussion
Complexity of Macrohaplogroup M

The basal variation in macrohaplogroup M of India that has emerged from our study is indeed impressive: it clearly outnumbers that of macrohaplogroup M in East Asia (Kong et al. 2003). The extreme basal multifurcation also poses a particular problem to phylogeny estimation because the long stems of the deepest branches could easily appear to coalesce by way of parallel mutation at a single site. It certainly needs to have many more lineages completely sequenced before we can be reasonably confident about, say, the monophyly of M18'38 or the actual placement of the branch referred to as M4a within M4. Therefore, the current tree of figure 1 cannot yet be regarded as "bulletproof" but rather constitutes a current best working hypothesis.

With the reconstructed phylogeny of macrohaplogroup M in Asia, it is now evident that some common mutations, especially in the control region, of M lineages from East Asia, South Asia, and the Near East are acquired by parallel events. For instance, the 11946 transition shared by an M30c lineage from India and an M7c lineage from the Philippines (Maca-Meyer et al. 2001) turns out to be a parallelism (compare our fig. 1 with fig. 1 from Kong et al. 2003). The global mtDNA phylogeny based on complete mtDNA sequences also allows rejecting premature inferences drawn from few hypervariable control-region sites.

A particular case in question is the origin of haplogroup M1, which is mainly found in Northeast Africa and the Near East (Quintana-Murci et al. 1999). Due to

Fig. 2.—The conflicts between our data and that of Rajkumar et al. (2005). Solid lines represent evolutionary pathways revealed in the present study while broken lines refer to those inferred by Rajkumar et al. (2005). Samples with prefix "RR#" were taken from Rajkumar et al. (2005). Mutations are scored relative to the rCRS (Andrews et al. 1999). Private mutations are not shown (//). Suffixes A, G, and T indicate transversions; "d" denotes deletion and a plus sign (+) denotes an insertion (specified by the nucleotide inserted). Haplogroups are defined and indicated as in figure 1.

Table 3
Additional Indian Samples Screened for Three Particular Coding-Region Sites

Haplogroup	Sample	HVS-I (16000+)	HVS-II	1888	5319	10986
H2b	rCRS			G	A	C
M5	B11	129 223	73 146 150 263 315+C 333	A		C
M5	B64	129 223	73 263 315+C	A		C
M5	C127	129 223 264 265C 311	73 263 309+C 315+C	A		C
M5	C193	129 223 362	73 263 309+C 315+C	A		C
M5	C1	48 129 223 290	73 263 315+C	A		C
M5	C38	129 223 249	73 146 263 309+C 315 +C	A		C
M6	C121	223 231 291 319 362	73 152 194 228 263 309+C 315+C		A	
M6	T167	223 231 362	73 146 263 309+C 315+C		A	
M6b	T151	188 223 231 362	73 146 195 263 309+C 315+C		G	

Note.—HVS, hypervariable segment.

the fact that M1 bears variant nucleotides, for example, at site 16311 in common with haplogroup M4, at 16129 with M5, and at 16249 with haplogroup M34, it has been proposed that M1 might have some affinity with Indian M haplogroups (Roychoudhury et al. 2001). This inference, however, could not receive support from our complete sequencing information. Indeed, the reconstructed ancestral motifs of all Indian M haplogroups turned out to be devoid of those variations that characterized M1, that is, 6446, 6680, 12403, and 14110 (Maca-Meyer et al. 2001; Herrnstadt et al. 2002). Therefore, those common mutations in the control region rather reflect random parallel mutations. There is no evidence whatsoever that M1 originated in India.

Comparison with the Rajkumar et al. (2005) Data

Rajkumar et al. (2005) provided 23 Indian M sequences that were deemed to be complete. On the basis of their phylogenetic tree, the authors negated the haplogroup status of M3 and M4. However, a site-by-site audit of their sequences revealed that the obtained data are problematic, with numerous basal mutations evidently missed as well as some phantom mutations inflicted. To demonstrate these problems and resolve the conflicting information, we compared the evolutionary pathways for haplogroups M2, M3, M5, M6, M35a, and M39 derived from our data and those predicted by Rajkumar et al. (2005). As shown in figure 2, mtDNAs Kur126, Chen, and Katk from Rajkumar et al. (2005) harbored a string of mutations specific for haplogroup M2a, namely, at sites 204, 1780, 5252, 8396, 8502, 9758, 16270, 16319, and 16352 (Kivisild et al. 2003; this study) but lacked the three other diagnostic mutations (at sites 11083, 15670, and 16274), whereas the salient 447G transversion (Kivisild et al. 2003; this study) was misdocumented as a 477 transition in their figure. The two sequences IB306 and Lyn180 from Rajkumar et al. (2005) could be assigned to haplogroup M6 but missed as many as seven M6 characteristic mutations (viz., 461, 5082, 5558, 9329, 10640, 13966, and 14128). Similarly, the M35a sample Lam8 was without the four mutations at sites 482, 12561, 15924, and 16093, and the M39 mtDNA Ho69 lacked a number of characteristic mutations (viz., 55+T, 59–60d, 65+T, 66T, 1811, 8679, and 15938). It is noteworthy that a rare transversion 10986A was found in three haplogroup M5 mtDNAs

(Bho134, Raj90, and Mus112) reported in Rajkumar et al. (2005), whereas this mutation was totally absent in our M5 samples. On the other hand, the basal M5 mutation 1888 identified in our study was not present in the samples of Rajkumar et al. (2005). To resolve these conflicts, we screened both sites in additional six mtDNAs from potentially different branches of haplogroup M5 (table 3). Our results confirmed the presence of 1888 and absence of 10986A in all M5 mtDNAs, thus casting serious doubts on the data provided by Rajkumar et al. (2005). Similarly, the screening for three additional M6 mtDNAs (table 3) confirmed that the 5319 mutation is specific to M6b, which was however regarded as a basal mutation of M6 by Rajkumar et al. (2005). Note that in the first version of the article by Rajkumar et al. (deposited at http://genomebiology.com/content/pdf/gb-2004-6-2-p3.pdf) the 5319 mutation was allocated to only one of the two M6 lineages.

To roughly assess the extent to which Rajkumar et al. (2005) overlooked the basal and private mutations, we calculated the mean distance (rho value: Forster et al. 1996; Saillard et al. 2000) of their reported M lineages to the root of M. The low value (6.7 ± 0.7) observed for their data, compared with ours (8.7 ± 0.6, see table 1), suggests that they might have averagely missed about two mutations in the coding region (577–16023) per sample.

A Rapid Dispersal Along the Asian Coast

It was pointed out that macrohaplogroups M, N, and R are universally distributed in Eurasia but differentiated into distinct haplogroups in East Asia, Oceania, Southeast Asia, and the Andaman Islands in particular (Macaulay et al. 2005; Thangaraj et al. 2005). This finding is further strengthened by our newly obtained Indian M data because the mutations that characterize the basal M lineages in India are virtually unique and not shared by those of East Asian, Oceanian, and Southeast Asian M lineages (Ingman et al. 2000; Ingman and Gyllensten 2003; Kong et al. 2003; Tanaka et al. 2004; Friedlaender et al. 2005; Macaulay et al. 2005). This star-like and nonoverlapping pattern of the mtDNA phylogeny is in good agreement with the proposed scenario that the initial dispersal of modern human into Eurasia some 60×10^3 years ago was rather rapid along the Asian coastline (Macaulay et al. 2005; Thangaraj et al. 2005; Forster and Matsumura 2005).

Supplementary Material

Acknowledgments

Literature Cited

Achilli, A., C. Rengo, V. Battaglia et al. (13 co-authors). 2005. Saami and Berbers—an unexpected mitochondrial DNA link. Am. J. Hum. Genet. **76**:883–886.

Achilli, A., C. Rengo, C. Magri et al. (21 co-authors). 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am. J. Hum. Genet. **75**:910–918.

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. **23**:147.

Bamshad, M., T. Kivisild, W. S. Watkins et al. (18 co-authors). 2001. Genetic evidence on the origins of Indian caste populations. Genome Res. **11**:994–1004.

Bandelt, H. J., A. Achilli, Q. P. Kong, A. Salas, S. Lutz-Bonengel, C. Sun, Y. P. Zhang, A. Torroni, and Y. G. Yao. 2005. Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies. Biochem. Biophys. Res. Commun. **333**:122–130.

Bandelt, H. J., L. Quintana-Murci, A. Salas, and V. Macaulay. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am. J. Hum. Genet. **71**:1150–1160.

Basu, A., N. Mukherjee, S. Roy et al. (12 co-authors). 2003. Ethnic India: a genomic view, with special reference to peopling and structure. Genome Res. **13**:2277–2290.

Derbeneva, O. A., R. I. Sukernik, N. V. Volodko, S. H. Hosseini, M. T. Lott, and D. C. Wallace. 2002. Analysis of mitochondrial DNA diversity in the Aleuts of the commander islands and its implications for the genetic history of Beringia. Am. J. Hum. Genet. **71**:415–421.

Finnilä, S., M. S. Lehtonen, and K. Majamaa. 2001. Phylogenetic network for European mtDNA. Am. J. Hum. Genet. **68**:1475–1484.

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

Forster, P., and S. Matsumura. 2005. Did early humans go north or south? Science **308**:965–966.

Friedlaender, J., T. Schurr, F. Gentz et al. (13 co-authors). 2005. Expanding Southwest Pacific mitochondrial haplogroups P and Q. Mol. Biol. Evol. **22**:1506–1517.

Herrnstadt, C., J. L. Elson, E. Fahy et al. (11 co-authors). 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am. J. Hum. Genet. **70**:1152–1171.

Herrnstadt, C., G. Preston, and N. Howell. 2003. Errors, phantoms and otherwise, in human mtDNA sequences. Am. J. Hum. Genet. **72**:1585–1586.

Howell, N., J. L. Elson, D. M. Turnbull, and C. Herrnstadt. 2004. African haplogroup L mtDNA sequences show violations of clock-like evolution. Mol. Biol. Evol. **21**:1843–1854.

Ingman, M., and U. Gyllensten. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res. **13**:1600–1606.

Ingman, M., H. Kaessmann, S. Pääbo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. Nature **408**:708–713.

Kivisild, T., K. Kaldma, M. Metspalu, J. Parik, S. S. Papiha, and R. Villems. 1999. The place of the Indian mitochondrial DNA variants in the global network of maternal lineages and the peopling of the Old World. Pp 135–152 in R. Deka and S. S. Papiha, eds. Genomic diversity. Kluwer/Academic/Plenum Publishers, New York.

Kivisild, T., S. Rootsi, M. Metspalu et al. (18 co-authors). 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am. J. Hum. Genet. **72**:313–332.

Kivisild, T., P. Shen, D. Wall et al. (17 co-authors). 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics (in press).

Kong, Q. P., Y. G. Yao, C. Sun, H. J. Bandelt, C. L. Zhu, and Y. P. Zhang. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am. J. Hum. Genet. **73**:671–676 (erratum **75**:157).

Maca-Meyer, N., A. M. González, J. M. Larruga, C. Flores, and V. M. Cabrera. 2001. Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. **2**:13.

Maca-Meyer, N., A. M. González, J. Pestano, C. Flores, J. M. Larruga, and V. M. Cabrera. 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genet. **4**:15.

Macaulay, V., C. Hill, A. Achilli et al. (21 co-authors). 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science **308**:1034–1036.

Macaulay, V., M. Richards, E. Hickey, E. Vega, F. Cruciani, V. Guida, R. Scozzari, B. Bonne-Tamir, B. Sykes, and A. Torroni. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. **64**:232–249.

Merriwether, D. A., J. A. Hodgson, F. R. Friedlaender, R. Allaby, S. Cerchio, G. Koki, and J. S. Friedlaender. 2005. Ancient mitochondrial M haplogroups identified in the Southwest Pacific. Proc. Natl. Acad. Sci. USA **102**:13034–13039.

Metspalu, M., T. Kivisild, E. Metspalu et al. (16 co-authors). 2004. Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. BMC Genet. **5**:26 (erratum **6**:41).

Mishmar, D., E. Ruiz-Pesini, P. Golik et al. (13 co-authors). 2003. Natural selection shaped regional mtDNA variation in humans. Proc. Natl. Acad. Sci. USA **100**:171–176.

Palanichamy, M. G., C. Sun, S. Agrawal, H. J. Bandelt, Q. P. Kong, F. Khan, C. Y. Wang, T. K. Chaudhuri, V. Palla, and Y. P. Zhang. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am. J. Hum. Genet. **75**:966–978.

Quintana-Murci, L., R. Chaix, R. S. Wells et al. (17 co-authors). 2004. Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. Am. J. Hum. Genet. **74**:827–845.

Quintana-Murci, L., O. Semino, H. J. Bandelt, G. Passarino, K. McElreavey, and A. S. Santachiara-Benerecetti. 1999.

Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. Nat. Genet. **23**:437–441.

Rajkumar, R., J. Banerjee, H. B. Gunturi, R. Trivedi, and V. K. Kashyap. 2005. Phylogeny and antiquity of M macrohaplogroup inferred from complete mt DNA sequence of Indian specific lineages. BMC Evol. Biol. **5**:26.

Reidla, M., T. Kivisild, E. Metspalu et al. (43 co-authors). 2003. Origin and diffusion of mtDNA Haplogroup X. Am. J. Hum. Genet. **73**:1178–1190.

Richards, M. B., V. A. Macaulay, H. J. Bandelt, and B. C. Sykes. 1998. Phylogeography of mitochondrial DNA in western Europe. Ann. Hum. Genet. **62**:241–260.

Roychoudhury, S., S. Roy, A. Basu, R. Banerjee, H. Vishwanathan, M. V. Usha Rani, S. K. Sil, M. Mitra, and P. P. Majumder. 2001. Genomic structures and population histories of linguistically distinct tribal groups of India. Hum. Genet. **109**:339–350.

Saillard, J., P. Forster, N. Lynnerup, H. J. Bandelt, and S. Nørby. 2000. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. Am. J. Hum. Genet. **67**:718–726.

Stringer, C. 2000. Coasting out of Africa. Nature **405**:24–25, 27.

Tanaka, M., V. M. Cabrera, A. M. González et al. (28 co-authors). 2004. Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res. **14**:1832–1850.

Thangaraj, K., G. Chaubey, T. Kivisild, A. G. Reddy, V. K. Singh, A. A. Rasalkar, and L. Singh. 2005. Reconstructing the origin of Andaman Islanders. Science **308**:996.

Torroni, A., C. Rengo, V. Guida et al. (12 co-authors). 2001. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am. J. Hum. Genet. **69**:1348–1356.

Yao, Y. G., Q. P. Kong, H. J. Bandelt, T. Kivisild, and Y. P. Zhang. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am. J. Hum. Genet. **70**:635–651.

Yao, Y. G., V. Macauley, T. Kivisild, Y. P. Zhang, and H. J. Bandelt. 2003. To trust or not to trust an idiosyncratic mitochondrial data set. Am. J. Hum. Genet. **72**:1341–1346; author reply 1346–1349.