Molecular Biology and Evolution

Research article

Large-Scale mtDNA Screening Reveals a Surprising Matrilineal Complexity in East Asia and Its Implications to the Peopling of the Region

Qing-Peng Kong,*[1,2] Chang Sun,[1] Hua-Wei Wang,[3] Mian Zhao,[1,2] Wen-Zhi Wang,[1,2] Li Zhong,[3] Xiao-Dan Hao,[1,2] Hui Pan,[1,2] Sha-Yan Wang,[4] Yao-Ting Cheng,[1,2] Chun-Ling Zhu,[1] Shi-Fang Wu,[1] Li-Na Liu,[1] Jie-Qiong Jin,[1] Yong-Gang Yao,[5] and Ya-Ping Zhang*[1,2,3]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]KIZ/CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, Kunming, China

[3]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, China

[4]Shenzhen People's Hospital, Clinical Medical College of Jinan University, Shenzhen, China

[5]Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

*__Corresponding author__: E-mail: kongqp@yahoo.com.cn; zhangyp@mail.kiz.ac.cn.

__Associate editor__: Connie Mulligan

Data deposition: All mtDNA genome sequences obtained in this study have been deposited into GenBank under accession numbers: GQ301887 and HM030499–HM030548.

Abstract

In order to achieve a thorough coverage of the basal lineages in the Chinese matrilineal pool, we have sequenced the mitochondrial DNA (mtDNA) control region and partial coding region segments of 6,093 mtDNAs sampled from 84 populations across China. By comparing with the available complete mtDNA sequences, 194 of those mtDNAs could not be firmly assigned into the available haplogroups. Completely sequencing 51 representatives selected from these unclassified mtDNAs identified a number of novel lineages, including five novel basal haplogroups that directly emanate from the Eurasian founder nodes (M and N). No matrilineal contribution from the archaic hominid was observed. Subsequent analyses suggested that these newly identified basal lineages likely represent the genetic relics of modern humans initially peopling East Asia instead of being the results of gene flow from the neighboring regions. The observation that most of the newly recognized mtDNA lineages have already differentiated and show the highest genetic diversity in southern China provided additional evidence in support of the Southern Route peopling hypothesis of East Asians. Specifically, the enrichment of most of the basal lineages in southern China and their rather ancient ages in Late Pleistocene further suggested that this region was likely the genetic reservoir of modern humans after they entered East Asia.

__Key words:__ mtDNA, lineage, matrilineal complexity, peopling, East Asia.

Introduction

In recent years, our knowledge of the fine-detailed mitochondrial DNA (mtDNA) tree of East Asian has been much improved, mainly due to the elaborate analyses of mtDNA genomes from China (Kivisild et al. 2002; Kong, Yao, Sun, et al. 2003; Trejaut et al. 2005; Kong et al. 2006; Zhao et al. 2009) and Japan (Tanaka et al. 2004; Bilal et al. 2008; Ueno et al. 2009). Although the vast majority of mtDNA lineages in East Asia (which is usually understood to encompass China, Mongolia, Korea Peninsula, and Japan) can be allocated into the available mtDNA haplogroup system, there are some infrequent mtDNA lineages whose phylogenetic status remains unknown. Full determination of these uncharacterized mtDNA types has at least two advantages: 1) such information helps to better understand the matrilineal genetic composition and refine the mtDNA phylogeny in East Asia and 2) full recognition of these mtDNA types may have important implications for getting deeper insights into the initial peopling scenario of the region.

Hitherto, debate on whether the earlier hominid populations had been "completely" replaced by modern humans continues, especially for East Asia where abundant and successive hominid fossil records have been discovered and continuity of the morphological characters between *Homo erectus* and modern East Asians has been proposed (Wu 2005, 2006). This proposal receives support from a recent report of the earliest modern human remains (Tianyuan 1) uncovered in Zhoukoudian, China, which shows the mosaic characteristics of both modern and archaic humans and therefore may suggest "substantial" gene flow from the archaic hominid populations (Shang et al. 2007). Despite that Y chromosome study has depicted a clear picture for the recent African origin of modern East Asians and

"completeness of the replacement of modern humans" in the region (Ke et al. 2001); this inference was questioned basically because evidence from paternally inherited Y chromosome could not reject the possibility of maternal contribution from the archaic hominid (Hawks 2001). In this context, the proposed substantial ancient genetic relics contributed by the archaic hominid female (Shang et al. 2007), if survived in modern East Asian matrilineal gene pool, would be naturally expected among these mtDNAs with uncharacterized phylogenetic status.

To achieve this objective, one has to overcome three challenges at first: 1) a large number of samples with extensively regional coverage would be indispensable to sample these infrequent mtDNAs; 2) an efficient strategy should be introduced for quickly pinpointing the unknown mtDNA types from thousands of individuals under study; and 3) an optimized sampling strategy for complete mtDNA sequencing (Kong et al. 2006) should be carried out to minimize the laboratory investment. In the present study, a total of 6,093 individuals, representing 84 populations sampled across China, were analyzed following a well-described strategy that focused on combined mtDNA control region and partial coding region information (Yao, Kong, et al. 2002; Yao et al. 2004). We pinpointed 194 mtDNAs that could not be firmly allocated into the available mtDNA haplogroup classification system. Sequencing the entire mtDNA genomes of 51 representatives selected from these uncharacterized mtDNAs identified a number of subhaplogroups and five novel basal haplogroups that branch directly from the Eurasian founder types (M and N). The current study represents an extensive sampling for the ancient relics of the Paleolithic composition across East Asia and revealed a surprising complexity of East Asian mtDNA gene pool.

Materials and Methods

Sampling

To get a comprehensively geographic coverage of Chinese samples and to collect more uncharacterized mtDNA types, a total of 6,093 individuals from 84 populations across China (fig. 1 and supplementary table S1, Supplementary Material online) were collected. Among them, 2,109 mtDNAs have been reported in our previous studies (Yao, Kong, et al. 2002; Yao, Nie, et al. 2002; Yao and Zhang 2002; Kong, Yao, Liu, et al. 2003; Yao et al. 2003, 2004; Chen et al. 2008; Zhao et al. 2009; Wang et al. 2010) and, if necessary, were further analyzed (for some segment(s), e.g., the second hypervariable segment [HVS-II] of mtDNA control region). To facilitate comparison with the published mtDNA data sets by the other laboratories, additional East Asian 5,065 mtDNAs retrieved from the literature were reanalyzed here (supplementary table S2, Supplementary Material online), including 2,819 samples (among which 68 individuals were in fact from Vietnam; Li et al. 2007) from mainland of China (Kivisild et al. 2002; Wen, Li, et al. 2004; Wen, Xie, et al. 2004; Wen et al. 2005; Zhang, Xu, Cui, et al. 2005; Zhang, Xu, Zheng, et al. 2005; Li et al. 2007), 975 samples from Taiwan (Tsai et al. 2001; Tajima et al. 2003; Trejaut et al. 2005), 311 samples from Japan (Maruyama et al. 2003; Nohira et al. 2010), and 960 samples from Korea (Allard et al. 2004; Jin et al. 2006; Lee et al. 2006), with special attempt to test whether our newly obtained phylogenetic information is helpful for classifying the previously unclassified mtDNAs. To control the sampling bias and appropriately evaluate the distribution of the newly identified basal lineages, the reported 765 Japanese mtDNA genome sequences (Tanaka et al. 2004; Ueno et al. 2009) were taken into account as well, leading the total sample size of Japanese mtDNAs to 1,076.

Haplogroup Classification and Recognition of Uncharacterized Lineages

To quickly determine the phylogenetic status of samples under consideration, we followed the haplogroup classification strategy that was used and optimized in our previous studies (Yao, Kong, et al. 2002; Yao et al. 2004; Ji et al. 2008). Specifically, the control region (commonly covering np 16024-407) was sequenced and analyzed, and at least one specific coding region site was chosen for typing so as to substantiate the haplogroup status that was inferred on the basis of the control region variation motif. The recently established mtDNA trees of East Asian (Kong, Yao, Sun, et al. 2003; Kong et al. 2006; Derenko et al. 2007; Ueno et al. 2009; Zhao et al. 2009), Southeast Asian (Macaulay et al. 2005; Dancause et al. 2009; Peng et al. 2010; Tabbada et al. 2010), South Asian (Palanichamy et al. 2004; Sun et al. 2006; Chandrasekar et al. 2009; Fornarino et al. 2009), and West Eurasian (Finnila et al. 2001; Herrnstadt et al. 2002; Palanichamy et al. 2004; Kivisild et al. 2006; Malyarchuk et al. 2008) were fully referenced across the analysis. To the end, any mtDNAs whose status cannot be recognized relative to the available haplogroup classification system will be regarded as uncharacterized lineages.

mtDNA Genome Typing and Age Estimation

The distilled mtDNA types, which could not be firmly assigned into the available haplogroup scheme (supplementary table S3, Supplementary Material online), were tentatively allocated into different groups according to the specific control region variation(s) they shared, and at least one representative from each group was selected for complete mtDNA sequencing as described elsewhere (Palanichamy et al. 2004; Kong et al. 2006; Sun et al. 2006; Wang et al. 2007; Wang et al. 2008). Quality of the generated genome data was carefully evaluated according to the caveats for the quality control (Yao et al. 2006, 2008, 2009; Kong et al. 2008). The remaining mtDNAs were then determined by genotyping some specific coding region variation according to the newly obtained genome information (fig. 2 and supplementary table S3, Supplementary Material online). Using this succinct approach, the majority of the uncharacterized mtDNAs could then be classified efficiently. For those mtDNAs that remained unclassified after the above analysis, their entire genomes

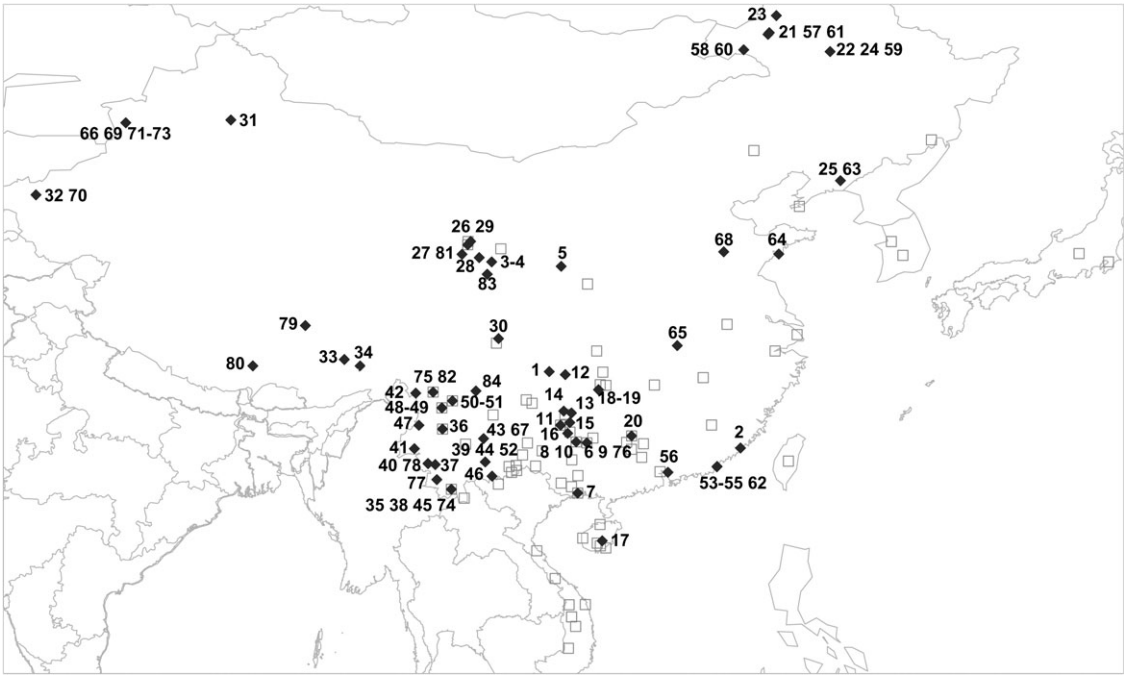

Fig. 1. Map showing the sampling locations of East Asian populations analyzed in this study. Codes 1–84 represent the studied populations from our laboratory (cf. supplementary table S1, Supplementary Material online for details).

were sequenced so as to fully determine their phylogenetic status. To avoid any potential conflict in mtDNA haplogroup nomenclature, we followed the nomenclature system suggested by van Oven and Kayser (2009; http://www.phylotree.org/) and the recently proposed haplogroup naming scheme (Kong et al. 2010) to designate the novel basal lineages in this study.

Age of each novel basal haplogroup was estimated by use of the statistics rho (Forster et al. 1996) and sigma (Saillard et al. 2000), whereas the recently proposed calibration rates for mtDNA were adopted (Perego et al. 2009; Soares et al. 2009).

Shortcut Weighted Near-Matching

Based on the newly acquired mtDNA information and data sets, reported East Asian mtDNAs (containing mainly control region sequences) from other laboratories were scrutinized as well (supplementary tables S2 and S4, Supplementary Material online). For the previously unclassified mtDNAs (supplementary table S4, Supplementary Material online), their phylogenetic status was determined roughly by (near-)matching with our samples that have been unambiguously determined (supplementary table S3, Supplementary Material online), with a special attempt to test whether the previously unassigned mtDNAs could be recognizable by our new data and the updated haplogroup classification system (See Reassignment of the Previously Unclassified mtDNAs for details). To better understand the distribution patterns of the newly identified basal lineages, any newly recognized mtDNAs from the reported data sets (cf. supplementary table S4, Supplementary Material online) will be included in the subsequent analyses such as network reconstruction and age estimation.

Results

Identification of Novel mtDNA Lineages

Detailed information of all mtDNAs with unknown status that were pinpointed from 6,093 Chinese individuals was listed in supplementary table S3, Supplementary Material online. Figure 2 displays the mtDNA variation of the completely sequenced representatives, from which several features can be summarized: 1) 21 unassigned mtDNA types prove to belong to certain undefined intermediate nodes within the East Asian mtDNA tree. For instance, sample GD-Han7832 bearing variants 16189 and 8281–8289d, as well as, a series of specific variations, apparently belongs to a previously unreported branch of haplogroup B; samples YN-Dai125 and GZ-She8 share all M7c diagnostic variants except for transition 5442, indicating the possible existence of an intermediate type between previously defined M7b'c and M7c (Kong et al. 2006); similarly, GX-Mulam24 and SC-DJY463 consist of a novel M7b'c lineage, which share a single variant (viz., 12405) with M7b and is characteristic of additional variants (723-3140-5324-8027-14284-16184 and a back mutation at site 199), indicative of the existence of a novel intermediate node between M7b'c and M7b; XJ-Kaz17, SC-DJY802, and GX-Zhuang12 prove to bear all M7c1 diagnostic variations in the coding region, notwithstanding their control region variation did not show any evident M7c1 motif. To compromise these incompatibilities with the previous M7b'c phylogeny (Kong et al. 2006), we followed the suggestion proposed by Derenko et al. (2007). Therefore, besides the defined haplogroups M7d and M7e (Derenko et al. 2007), the rest two novel M7 lineages are named M7f and M7g, respectively. Samples HN-Yao159 and

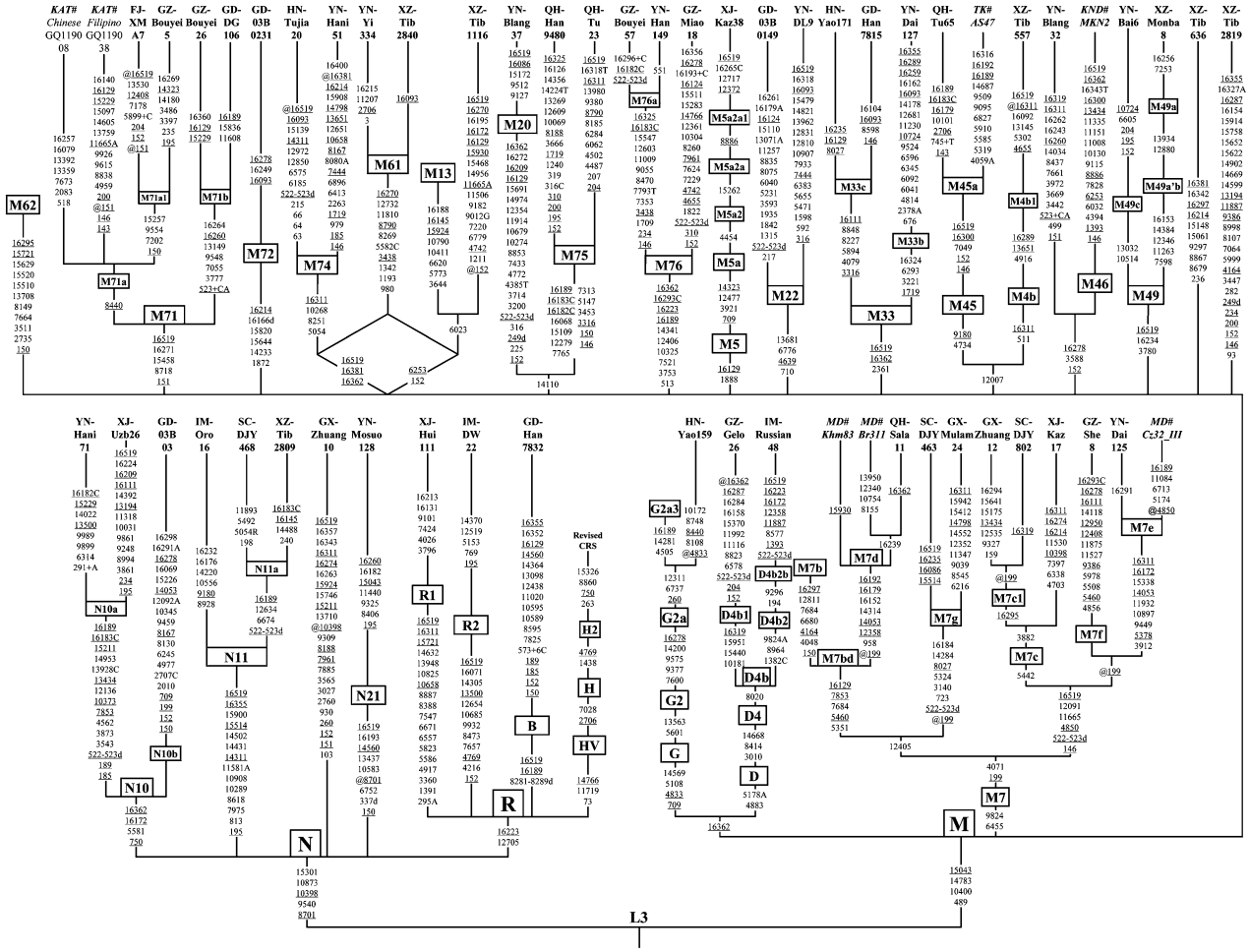

FIG. 2. Phylogenetic tree of the newly recognized mtDNA lineages. Suffixes A, C, G, and T refer to transversions, "d" means deletion, and "+" indicates insertion; recurrent mutations are underlined; "@" means a reverse mutation. The C stretch length polymorphism in regions 303–315 was disregarded for the tree reconstruction. Suffixes *"KAT#," "TK#," "KND#," and "MD#"* attached to sample names indicate the sources of the sequences: Tabbada et al. (2010), Kivisild et al. (2006), Dancause et al. (2009), and Derenko et al. (2007), respectively.

IM-Russian48 belong to haplogroups G2 and D4b2b, respectively, and their apparent bizarre motifs seem to be triggered by the rare back mutation at the relatively conserved sites 4833 and 16223. 2) 42 mtDNAs can be classified into the defined haplogroups prevalent in or specific to the other subcontinents. For instance, HN-Yao171, GD-Han7815, and YN-Dai127 all share variant 2361, suggesting their potential status as haplogroup M33 that was first identified in South Asia under the assumption of a high stability of this position (Sun et al. 2006); likewise, XJ-Kaz38 is assigned into South Asian haplogroup M5a2a1 according to the diagnostic variants it harbored; whereas YN-Mosuo128 can be allocated into haplogroup N21, a unique lineage first recognized in Malay (Macaulay et al. 2005). 3) All of the remaining uncharacterized mtDNAs are fully determined (fig. 2 and supplementary table S3, Supplementary Material online). Besides three lineages (M49, M61, and M71) that were recently defined in Filipino (Tabbada et al. 2010) and Indian (Chandrasekar et al. 2009), herein we have identified five additional novel basal haplogroups, which are tentatively named M74, M75, M76, N10, and N11 here, as well as three novel lineages (XZ-Tib636, XZ-Tib2819, and GX-Zhuang10) that have branched off deeply from the Eurasian founder type M or N (fig. 2).

Reassignment of the Previously Unclassified mtDNAs

Because most of the novel lineages identified in the present study could be effectively classified with the aid of their specific control region motifs (supplementary table S3, Supplementary Material online and fig. 2), we attempt to quickly recognize the previously unclassified East Asian mtDNAs (always labeled as M*, N*, or R*) in the literature (supplementary table S4, Supplementary Material online). Supplementary table S4, Supplementary Material online, displayed the reclassification results. Taken the mtDNA data reported in Wen et al. (2005) as an example, nearly all M*/N*/R* samples can be recognized now by (near-) matching with our newly obtained mtDNAs: nine samples with variations 16223-16311-16362-16381 evidently belong to haplogroup M74; five individuals, two (YLO22 and

YBP28) with motif 16104-16111-16223 and the rest three (MHN43, YLO03, and YLO23) with variations 16111-16223-16235, match well with our GD-Han7815 and HN-Yao171, respectively (fig. 2), thus suggesting their M33 status; YGS03 and MHN22 should be allocated into haplogroup M71 according to the specific variants (16223 and 16271).

Distribution of the Novel Lineages in East Asia

As shown in figure 2, most of newly identified lineages within haplogroup M7 are from southern China, including two Hans (SC-DJY463 and SC-DJY802) from Sichuan Province and four samples (GX-Mulam24, YN-Dai125, GX-Zhuang12, and GZ-She8) belonging to four different ethnic groups (Mulam, Dai, Zhuang, and She). All these ethnic groups trace their origins back to the ancient Bai-Yue tribe and are currently distributed in Guangxi and Yunnan Provinces in southern China. This profile is substantiated when more samples with similar control region motif from the literature were considered (supplementary tables S3 and S4, Supplementary Material online): our sample Hainan-Li66 (supplementary table S3, Supplementary Material online) and nine samples reported by Li et al. (2007) (supplementary table S4, Supplementary Material online) all share the same motif (16172-16223-16291-16311) with YN-Dai125 (fig. 2), suggesting their haplogroup M7e status; these 11 samples all belong to ethnic groups with the Bai-Yue ancestry. In combination with the previous results of haplogroup M7b (Kivisild et al. 2002; Hill et al. 2006), our study confirms the notion that haplogroups M7b and M7c have their origins in southern China.

Most haplotypes within the newly characterized haplogroups M74 and M76 are confined to southern Chinese populations, suggesting their southern China origins (supplementary tables S3, S4 and figs. S1–S6, Supplementary Material online). The distribution patterns for haplogroups M71, N10, and N11 are more complex. Specifically, haplogroup M71 shows restricted distribution to the populations residing in southern China, including those with recorded origins from the ancient Di-Qiang tribe that once resided in northwest China about 3,000–5,000 years ago (Wang 1994) (supplementary fig. S1, Supplementary Material online). For example, two M71 haplotypes (recognized by control region motifs 16223-16271-16311-151 and 16223-16269-16271-16311-151, respectively) are observed in Pumi, Naxi, Lahu, and Mosuo people. Because these four ethnic populations reside specifically in Yunnan Province (southwest China) at the present time and no M71 lineage is observed elsewhere, especially in the populations from northwest China—the homeland of the Di-Qiang tribe, it is then apparent that the presence of M71 in these populations was most likely attributed to gene flow from local residents during the southward migration of Di-Qiang people and/or after their settlement in southwest China (Yao and Zhang 2002). Likewise, half of N10a mtDNAs come from ethnic populations with the Di-Qiang ancestry (such as Hani and Yi) but belong only to two haplotypes (judging from their control region variation; supplementary fig. S5, Supplementary Material online), likely reflecting a founder effect. It seems that most of the N10a haplotypes, including its sister cluster N10b that is identified in Han populations from Guangdong, Shanghai, and Jiangsu, are present in southern Chinese and likely have a southern origin. The distribution pattern of N11 seems to be a little more complex (supplementary fig. S6, Supplementary Material online): most of N11a haplotypes are confined to Yunnan Province and its adjacent regions (e.g., Sichuan Province); whereas IM-Oro16, the only N11b sample, was sampled from Oroqen population from northeast China (Kong, Yao, Liu, et al. 2003). Considering the facts that N11a comprises ten mtDNAs and has highest genetic diversity in southwest China but N11b contains only a single representative (viz., IM-Oro16), it is more likely that southern China might be the origin place of this haplogroup.

To further determine the distribution of these basal lineages outside China, a number of 2,036 samples from Japan (Maruyama et al. 2003; Tanaka et al. 2004; Ueno et al. 2009) and Korea (Allard et al. 2004; Jin et al. 2006; Lee et al. 2006) retrieved from literature were analyzed; none of these lineages were observed. The sample size (2,036) of Korean and Japanese is comparable with the Chinese samples under analysis ($N = 9,887$; including 3,794 samples retrieved from the literature and 6,093 mtDNAs from this study) when taking into account the population sizes of these countries. It is then unlikely that sampling bias could fully explain this observation. Indeed, a screening for the published mtDNA data from northern Asia (comprising 3,824 individuals; Bermisheva et al. 2002; Derenko et al. 2003; Fedorova et al. 2003; Pakendorf et al. 2003; Starikovskaya et al. 2005; Derenko et al. 2007) failed to pinpoint any of these lineages as well, again substantiating our observation.

Discussion

In this study, we have paid special attention to the 194 uncharacterized mtDNA types (supplementary table S3, Supplementary Material online) that were pinpointed from a massive collection (>6,000) of Chinese samples with a broad geographic coverage (fig. 1 and supplementary table S1, Supplementary Material online). Our full determination of the heretofore uncharacterized mtDNAs based on complete genome sequencing and genotyping for specific coding region variants revealed that all of them could be allocated unambiguously into the available mtDNA scheme of modern humans. No archaic mtDNA types with origins from the earliest hominid (e.g., *H. erectus*) were observed.

Applying the newly updated mtDNA tree to the previously reported East Asian mtDNA data sets (comprising of 5,065 individuals; supplementary table S2, Supplementary Material online) has resolved the puzzling status of most of the unclassified mtDNA types (supplementary table S4, Supplementary Material online). However, there are some mtDNAs remaining unrecognizable, partially because of the lack of further information, and they may consist of additional novel mtDNA lineage(s), especially in consideration of the huge population size of East Asians (>1.5 billion).

Moreover, recognition of haplogroup status preliminarily based on specific control region motif may fail simply due to the problem of insufficient information available. Focusing merely on the first hypervariable segment (HVS-I) of mtDNA control region (usually covering 16024–16383) is the most fashionable way in mtDNA studies (cf. Wen et al. 2005; Li et al. 2007). Unfortunately, the control region motifs of some East Asian haplogroups, including the newly identified ones in this study, for example, M71, M75, M76, N10a, and N11, are not located merely in HVS-I; this situation will undoubtedly disable any efforts of haplogroup classification (cf. supplementary table S4, Supplementary Material online).

Evidently, our results showed that, with the exception of a few samples from the reported studies, no archaic types (from the archaic hominid) were observed in the unrecognized mtDNA types identified from an extensive study of 11,158 individuals (6,093 from the present study and 5,065 from the literature) sampled across East Asia, a unique region believed to be the *de novo* place where modern Chinese/East Asians were independently originated or, at least, substantial genetic contribution from the archaic hominid populations had occurred (Shang et al. 2007). Combined with the evidence from the current investigation and the reported Y chromosome study (Ke et al. 2001), it is apparent that no genetic contribution from the earlier hominid populations was detected in the modern East Asian matrilineal pool, or such an ancient contribution is too infinitesimal ($P = 0.0008$, assuming a frequency of 0.1% of the archaic maternal component in East Asian) to be displayed.

Because the newly identified haplogroups such as M74–M76, N10, N11, and M71 have the highest genetic diversity in southern part of China, especially in southwest China, and are virtually absent in Korea, Japan, and northern Asia, we speculate that their presence may be in fact the result of recent gene flow from Southeast Asia. A quick screen of recently published Southeast Asian mtDNA data (containing a total of 3,075 individuals; Fucharoen et al. 2001; Tajima et al. 2004; Hill et al. 2006, 2007; Li et al. 2007; Irwin et al. 2008; Lertrit et al. 2008; Zimmermann et al. 2009; Maruyama et al. 2010; Peng et al. 2010; Tabbada et al. 2010) did identify some sequence matches with the newly identified basal lineages. For instance, seven Philippines (five with motif 16223-16269-16271; Hill et al. 2007 and two with motif 16129-16140-16223-16271; Tabbada et al. 2010), 14 Thais (ten with motif 16223-16260-16264-16271; Fucharoen et al. 2001; Zimmermann et al. 2009; three with motif 16223-16271; Zimmermann et al. 2009; and one with motif 16129-16140-16223-16271; Zimmermann et al. 2009), two Malays (with motifs 16129-16223-16269-16271-151 and 16183-16189-16223-16271-16311-151, respectively; Maruyama et al. 2010), and four Vietnamese (bearing motifs 16223-16269-16271 and 16223-16271; Li et al. 2007; Irwin et al. 2008; Peng et al. 2010) could be assigned into haplogroup M71 (supplementary fig. S1, Supplementary Material online). Five individuals from Borneo, five individuals from Thailand (Hill et al. 2007; Irwin et al. 2008; Zimmermann et al. 2009), and two Vietnamese share either motif 16172-16189-16223-16362 or 16069-16172-16223-16291A-16298-16362 that clearly points to haplogroup N10 (supplementary fig. S5, Supplementary Material online). Five samples from Thailand share a common control region motif (16223-16311-16362) and variant 5054 (Zimmermann et al. 2009), strongly suggesting their M74 status. One individual from Philippine (Tabbada et al. 2010) is characteristic of variation 16126-16189-16223-16319-16325, which is similar to that of our sample QH-Han9480 and thus suggests it plausibly belong to haplogroup M75 (supplementary fig. S3, Supplementary Material online). As for N11 (supplementary fig. S6, Supplementary Material online), no near-match was observed outside China so far. Intriguingly, though M71 was recently regarded as "indigenous lineage" in Philippine (Tabbada et al. 2010), comparison with our current data (supplementary fig. S1, Supplementary Material online) suggests that this notion may be highly implausible, simply because most of the M71 mtDNAs in Southeast Asians (including Filipino) are located virtually at the terminal branches of the haplogroup. This pattern suggests for a recent divergence among these M71 mtDNAs (supplementary fig. S1, Supplementary Material online). As a result, among the five basal haplogroups newly identified in the present study, three (M74, M76, and N10) of them were observed in Southeast Asians (cf. supplementary figs. S2, S4, and S5, Supplementary Material online).

Among the six novel basal haplogroups (including M74–M76, N10, N11, and M71), only four were observed in the previously unassigned Southeast Asian mtDNAs (designated as M* or N*). Although this could be interpreted simply as insufficient sampling for Southeast Asians (3,075 individuals vs. 11,158 East Asians), three lines of evidence obtained so far argued against this interpretation. First, full comparison with >7,000 published (nearly)whole mtDNA genomes (mtDNA tree Build 9; http://www.phylotree.org/), especially those with South Asian (Palanichamy et al. 2004; Thangaraj et al. 2005; Kivisild et al. 2006; Sun et al. 2006; Chandrasekar et al. 2009; Fornarino et al. 2009; Kumar et al. 2009) and Southeast Asian ancestry (Macaulay et al. 2005; Hill et al. 2006; Dancause et al. 2009; Peng et al. 2010; Tabbada et al. 2010), detected no sister clade, suggesting that these six haplogroups did branch directly from the M or N founder node (fig. 2). Second, an extensive searching of the mtDNA lineages in the published Asian populations, including those from Japan, Korea, northern Asia, and Southeast Asia, showed that more root types of these haplogroups were observed in China, whereas the Southeast Asian haplotypes mostly represented the "twig" types (judging from the reconstructed median networks; supplementary figs. S1–S6, Supplementary Material online). Finally, most of the Southeast Asian populations containing haplogroups M71, M74, M76, and N10 (with very low distribution frequencies as well) coincidently had close affinity with some Chinese ethnic populations in the historic time, such as Thai and Vietnamese, the existence of these basal haplogroups in East Asia is unlikely to be introduced from Southeast Asia. Intriguingly, different time estimation methods revealed that the divergence time

Table 1. Estimated Ages of the Identified Basal Lineages in East Asia.

		Control Region Transitions (np 16090–16365)			Coding Region Substitutions			Coding Region Synonymous Transitions			Coding Region Synonymous Substitutions			Complete Genome Substitutions		
		Soares Rate (Soares et al. 2009)			Modified Mishmar Rate (Perego et al. 2009)			Modified Kivisild Rate (Perego et al. 2009)			Soares Synonymous Rate (Soares et al. 2009)			Soares Synonymous Rate (Soares et al. 2009)		
Haplogroup	*n*	$\rho \pm \sigma$	*T* (ky)	*n*	$\rho \pm \sigma$	*T* (ky)	*n*	$\rho \pm \sigma$	*T* (ky)	*n*	$\rho \pm \sigma$	*T* (ky)	*n*	$\rho \pm \sigma$	*T* (ky)	*n*
M71	80	1.74 ± 0.49	32.74 ± 9.30	6	6.83 ± 1.36	31.50 (25.21, 37.79)	6	5.17 ± 1.24	39.53 (30.07, 48.98)	6	5.33 ± 1.25	42.05 (32.21, 51.88)	6	10.83 ± 1.64	30.25 (25.37, 35.22)	6
M74	52	1.21 ± 0.68	22.83 ± 12.77	2	9.50 ± 2.18	43.80 (33.75, 53.84)	2	5.00 ± 1.58	38.25 (26.15, 50.35)	2	5.50 ± 1.66	43.36 (30.29, 56.44)	2	14.50 ± 2.69	41.47 (33.19, 49.97)	2
M75	11	1.64 ± 0.70	30.84 ± 13.27	2	8.50 ± 2.06	39.19 (29.68, 48.69)	2	6.50 ± 1.80	49.73 (35.93, 63.52)	2	7.00 ± 1.87	55.19 (40.44, 69.94)	2	13.00 ± 2.55	36.83 (29.10, 44.76)	2
M76	9	1.78 ± 0.87	33.50 ± 16.49	3	10.00 ± 2.31	46.10 (35.45, 56.75)	3	6.33 ± 1.80	48.45 (34.72, 62.18)	3	6.33 ± 1.80	49.93 (35.78, 64.08)	3	14.00 ± 2.71	39.91 (31.63, 48.43)	3
N10	37	2.35 ± 0.93	44.31 ± 17.46	3	15.00 ± 2.69	69.15 (56.76, 81.54)	3	10.67 ± 2.31	81.60 (63.93, 99.27)	3	10.67 ± 2.31	84.10 (65.89, 102.30)	3	21.33 ± 3.16	63.44 (53.11, 74.00)	3
N11	11	2.00 ± 1.18	37.69 ± 22.21	3	3.67 ± 1.29	16.90 (10.95, 22.85)	3	3.00 ± 1.11	22.95 (14.49, 31.41)	3	3.00 ± 1.11	23.65 (14.94, 32.37)	3	6.00 ± 1.63	16.17 (11.62, 20.83)	3

of all lineages falls in the Late Pleistocene (table 1), indicative of their ancient differentiation and deep roots in the region. However, it has to be pointed out that although there are some debates on the accuracy of molecular dating with the rho statistic (Cox 2008), the estimated ages of the basal lineages only provide a rough time range when they began to differentiate and may serve as some kind of circumstantial evidence in support of their long-time existence in the region. Therefore, our thorough analysis of uncharacterized mtDNA types suggests that most of the identified novel lineages (including M74–M76, N10, N11, and M71) most likely represent the genetic relics of the ancestors of modern humans when they entered and colonized East Asia rather than resulting from recent gene flow from the neighboring regions, for example, Southeast Asia.

In retrospect, the Southern Route model speculated that modern humans first occupied the southern part of East Asia, with subsequent migration northward that eventually led to the colonization of the northern part of the subcontinent (Jin and Su 2000; Abdulla et al. 2009). This scenario received support from studies of Y chromosome (Su et al. 1999; Shi et al. 2005) and mtDNA (Yao, Kong, et al. 2002). Significantly, the observation of some mtDNA types with unknown phylogenetic status (previously designated as M*, N*, and R*) prevalent in southern Han populations was suggested to be in accordance with the hypothesis of southern origin of Chinese (Yao, Kong, et al. 2002). Of note is that this inference reckons much on the exact phylogenetic status of these uncharacterized lineages within the East Asian mtDNA phylogenetic scheme. The phylogenetically basal status of most of these previously uncharacterized mtDNA types, as demonstrated in the present study, substantiated the previous inference and provided additional support for the southern origin of modern East Asian. However, because the Neolithic demographic scenarios arising in northern and central China, including the origin (and dispersal) of agriculture and the expansion of Han people, might have played important roles in shaping the genetic landscape in northern China, which would partially explain the scarceness of these lineages in that region. Considering the facts that 1) the distribution frequencies of these lineages are extremely low, 2) none of them were observed in Korea, Japan, and northern Asia, and 3) these lineages are prevalent in the local populations residing in southern China and Southeast Asia, it is then most unlikely that more recent demographic events in China could lead to the current pattern. Instead, this unique pattern suggests that these basal lineages may find their roots in southern China. In this context, our observations, including an enrichment of most of the basal lineages in southern China (especially its southwestern part) (supplementary figs. S1–S6, Supplementary Material online) and the ancient ages of these newly found basal haplogroups (table 1), raise a possibility that southern (especially southwest) China was probably the genetic reservoir of modern humans when they first populated East Asia. However, one has to admit that the lack of extensive data from Southeast Asia,

especially for some key regions, for example, Myanmar, make any precise localization of this initial scenario at best tentative.

In short, by fully characterizing the genetic relics of the initial modern humans in East Asia based on extensively population sampling and entire mtDNA genome sequencing, our results suggest that the persistence of the infrequent previously uncharacterized mtDNA types in East Asia is largely attributable to the existence of either novel indigenous haplogroups or "foreign" lineages introduced by the ancient and/or recent migration event(s). Significantly, we have observed the enrichment of the novel basal lineages and a higher genetic diversity in populations from south and southwest China. This pattern is in good accordance with the hypothesized southern origin of East Asians and, furthermore, suggests that southern China might serve as the genetic reservoir of modern humans when they first populated East Asia. More information from the other genetic markers would be helpful to validate our conclusion.

Supplementary Material

Supplementary tables S1–S4 and figures S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Prof. Hans-Jürgen Bandelt for his valuable suggestions on this study and the two anonymous reviewers for their helpful comments. This work was supported by grants from National Basic Research Program of China (No. 2007CB507405), Natural Science Foundation of China (30621092, 30900797, and 30925021), Yunnan Province, and the Chinese Academy of Sciences.

References

Abdulla MA, Ahmed I, Assawamakin A, et al. (92 co-authors). 2009. Mapping human genetic diversity in Asia. *Science* 326: 1541–1545.

Allard MW, Wilson MR, Monson KL, Budowle B. 2004. Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. *Leg Med (Tokyo)*. 6:11–24.

Bermisheva MA, Tambets K, Villems R, Khusnutdinova EK. 2002. Diversity of mitochondrial DNA haplogroups in ethnic populations of the volga-ural region. *Mol Biol*. 36:802–812.

Bilal E, Rabadan R, Alexe G, et al. (13 co-authors). 2008. Mitochondrial DNA haplogroup D4a is a marker for extreme longevity in Japan. *PLoS One*. 3:e2421.

Chandrasekar A, Kumar S, Sreenath J, et al. (21 co-authors). 2009. Updating phylogeny of mitochondrial DNA macrohaplogroup m in India: dispersal of modern human in South Asian corridor. *PLoS One*. 4:e7447.

Chen F, Wang S-Y, Zhang R-Z, Hu Y-H, Gao G-F, Liu Y-H, Kong Q-P. 2008. Analysis of mitochondrial DNA polymorphisms in Guangdong Han Chinese. *Forensic Sci Int Genet*. 2:150–153.

Cox MP. 2008. Accuracy of molecular dating with the rho statistic: deviations from coalescent expectations under a range of demographic models. *Hum Biol*. 80:335–357.

Dancause KN, Chan CW, Arunotai NH, Lum JK. 2009. Origins of the Moken Sea Gypsies inferred from mitochondrial hypervariable region and whole genome sequences. *J Hum Genet*. 54:86–93.

Derenko M, Malyarchuk B, Grzybowski T, et al. (12 co-authors). 2007. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet*. 81:1025–1041.

Derenko MV, Grzybowski T, Malyarchuk BA, et al. (11 co-authors). 2003. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet*. 67:391–411.

Fedorova SA, Bermisheva MA, Villems R, Maksimova NR, Khusnutdinova EK. 2003. Analysis of mitochondrial DNA haplotypes in yakut population. *Mol Biol (Mosk)*. 37:643–653.

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

Fornarino S, Pala M, Battaglia V, Maranta R, Achilli A, Modiano G, Torroni A, Semino O, Santachiara-Benerecetti SA. 2009. Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol Biol*. 9:154.

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

Fucharoen G, Fucharoen S, Horai S. 2001. Mitochondrial DNA polymorphisms in Thailand. *J Hum Genet*. 46:115–125.

Hawks J. 2001. The Y chromosome and the replacement hypothesis. *Science* 293:567.

Herrnstadt C, Elson JL, Fahy E, 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.

Hill C, Soares P, Mormina M, et al. (12 co-authors). 2006. Phylogeography and ethnogenesis of aboriginal Southeast Asians. *Mol Biol Evol*. 23:2480–2491.

Hill C, Soares P, Mormina M, et al. (11 co-authors). 2007. A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet*. 80:29–43.

Irwin JA, Saunier JL, Strouss KM, Diegoli TM, Sturk KA, O'Callaghan JE, Paintner CD, Hohoff C, Brinkmann B, Parsons TJ. 2008. Mitochondrial control region sequences from a Vietnamese population sample. *Int J Legal Med*. 122:257–259.

Ji Y, Zhang A-M, Jia X, Zhang Y-P, Xiao X, Li S, Guo X, Bandelt H-J, Zhang Q, Yao Y-G. 2008. Mitochondrial DNA haplogroups M7b1′2 and M8a affect clinical expression of leber hereditary optic neuropathy in Chinese families with the m.11778G→A mutation. *Am J Hum Genet*. 83:760–768.

Jin HJ, Kwak KD, Hong SB, Shin DJ, Han MS, Tyler-Smith C, Kim W. 2006. Forensic genetic analysis of mitochondrial DNA hypervariable region I/II sequences: an expanded Korean population database. *Forensic Sci Int Genet*. 158:125–130.

Jin L, Su B. 2000. Natives or immigrants: modern human origin in east Asia. *Nat Rev Genet*. 1:126–133.

Ke Y, Su B, Song X, et al. (23 co-authors). 2001. African origin of modern humans in East Asia: a tale of 12,000 Y chromosomes. *Science* 292:1151–1153.

Kivisild T, Shen P, Wall DP, et al. (17 co-authors). 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172:373–387.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol*. 19:1737–1751.

Kong Q-P, Bandelt H-J, Sun C, et al. (12 co-authors). 2006. Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet*. 15:2076–2086.

Kong Q-P, Bandelt H-J, Zhao M, Zhang Y-P, Yao Y-G. 2010. Reply to van Oven: suggestions and caveats for naming mtDNA haplogroup. *Proc Natl Acad Sci U S A*. 107:E40–E41.

Kong Q-P, Salas A, Sun C, Fuku N, Tanaka M, Zhong L, Wang C-Y, Yao Y-G, Bandelt H-J. 2008. Distilling artificial recombinants from large sets of complete mtDNA genomes. *PLoS One*. 3:e3016.

Kong Q-P, Yao Y-G, Liu M, Shen S-P, Chen C, Zhu C-L, Palanichamy MG, Zhang Y-P. 2003. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet*. 113:391–405.

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

Kumar S, Ravuri RR, Koneru P, Urade BP, Sarkar BN, Chandrasekar A, Rao VR. 2009. Reconstructing Indian-Australian phylogenetic link. *BMC Evol Biol*. 9:173.

Lee HY, Yoo JE, Park MJ, Chung U, Shin KJ. 2006. Mitochondrial DNA control region sequences in Koreans: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. *Int J Legal Med*. 120:5–14.

Lertrit P, Poolsuwan S, Thosarat R, Sanpachudayan T, Boonyarit H, Chinpaisal C, Suktitipat B. 2008. Genetic history of Southeast Asian populations as revealed by ancient and modern human mitochondrial DNA analysis. *Am J Phys Anthropol*. 137:425–440.

Li H, Cai X, Winograd-Cort ER, et al. (12 co-authors). 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am J Phys Anthropol*. 134:481–488.

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

Malyarchuk B, Grzybowski T, Derenko M, Perkova M, Vanecek T, Lazur J, Gomolcak P, Tsybovsky I. 2008. Mitochondrial DNA phylogeny in Eastern and Western Slavs. *Mol Biol Evol*. 25:1651–1658.

Maruyama S, Minaguchi K, Saitou N. 2003. Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. *Int J Legal Med*. 117:218–225.

Maruyama S, Nohira-Koike C, Minaguchi K, Nambiar P. 2010. MtDNA control region sequence polymorphisms and phylogenetic analysis of Malay population living in or around Kuala Lumpur in Malaysia. *Int J Legal Med*. 124:165–170.

Nohira C, Maruyama S, Minaguchi K. 2010. Phylogenetic classification of Japanese mtDNA assisted by complete mitochondrial DNA sequences. *Int J Legal Med*. 124:7–12.

Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. *Am J Phys Anthropol*. 120:211–224.

Palanichamy MG, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhuri TK, Palla V, Zhang Y-P. 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.

Peng M-S, Quang HH, Dang KP, Trieu AV, Wang H-W, Yao Y-G, Kong Q-P, Zhang Y-P. 2010. Tracing the Austronesian Footprint in Mainland Southeast Asia: a perspective from mitochondrial DNA. *Mol Biol Evol*. (In press).

Perego UA, Achilli A, Angerhofer N, et al. (16 co-authors). 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol*. 19:1–8.

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

Shang H, Tong H, Zhang S, Chen F, Trinkaus E. 2007. An early modern human from Tianyuan Cave, Zhoukoudian, China. *Proc Natl Acad Sci U S A*. 104:6573–6578.

Shi H, Dong Y-L, Wen B, Xiao C-J, Underhill PA, Shen P-D, Chakraborty R, Jin L, Su B. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am J Hum Genet*. 77:408–419.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet*. 84:740–759.

Starikovskaya EB, Sukernik RI, Derbeneva OA, et al. (11 co-authors). 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann Hum Genet*. 69:67–89.

Su B, Xiao J, Underhill P, et al. (21 co-authors). 1999. Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet*. 65:1718–1724.

Sun C, Kong Q-P, Palanichamy MG, Agrawal S, Bandelt H-J, Yao Y-G, Khan F, Zhu C-L, Chaudhuri TK, Zhang Y-P. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol*. 23:683–690.

Tabbada KA, Trejaut J, Loo JH, Chen YM, Lin M, Mirazon-Lahr M, Kivisild T, De Ungria MC. 2010. Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol Biol Evol*. 27:21–31.

Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S. 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet*. 49:187–193.

Tajima A, Sun CS, Pan IH, Ishida T, Saitou N, Horai S. 2003. Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: implications for the population history of aboriginal Taiwanese. *Hum Genet*. 113:24–33.

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

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

Trejaut JA, Kivisild T, Loo JH, Lee CL, He CL, Hsu CJ, Lee ZY, Lin M. 2005. Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biol*. 3:e247.

Tsai LC, Lin CY, Lee JC, Chang JG, Linacre A, Goodwin W. 2001. Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Sci Int Genet*. 119: 239–247.

Ueno H, Nishigaki Y, Kong Q-P, Fuku N, Kojima S, Iwata N, Ozaki N, Tanaka M. 2009. Analysis of mitochondrial DNA variants in Japanese patients with schizophrenia. *Mitochondrion* 9:385–393.

van Oven M, Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 30:E386–394.

Wang C-Y, Wang H-W, Yao Y-G, Kong Q-P, Zhang Y-P. 2007. Somatic mutations of mitochondrial genome in early stage breast cancer. *Int J Cancer*. 121:1253–1256.

Wang F-R. 1994. The origin and blossom of Tubo. In: Wang ZH, editor. History of Chinese ethnic groups. Beijing (China): China Social Sciences Press. p. 363–382.

Wang H-W, Jia X, Ji Y, Kong Q-P, Zhang Q, Yao Y-G, Zhang Y-P. 2008. Strikingly different penetrance of LHON in two Chinese families with primary mutation G11778A is independent of mtDNA haplogroup background and secondary mutation G13708A. *Mutat Res*. 643:48–53.

Wang W-Z, Wang C-Y, Cheng Y-T, Xu A-L, Zhu C-L, Wu S-F, Kong Q-P, Zhang Y-P. 2010. Tracing the origins of Hakka and

Chaoshanese by mitochondrial DNA analysis. *Am J Phys Anthropol*. 141:124–130.
Wen B, Li H, Gao S, et al. (18 co-authors). 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Mol Biol Evol*. 22:725–734.
Wen B, Li H, Lu D, et al. (18 co-authors). 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302–305.
Wen B, Xie X, Gao S, et al. (13 co-authors). 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet*. 74:856–865.
Wu X-Z. 2005. Discussion on the results of some molecular studies concerning the origin of modern Chinese. *Acta Anthropol Sin*. 24:259–269.
Wu X-Z. 2006. New arguments on continuity of human evolution in China. *Acta Anthropol Sin*. 25:17–25.
Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P. 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet*. 70:635–651.
Yao Y-G, Kong Q-P, Man X-Y, Bandelt H-J, Zhang Y-P. 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol*. 20:214–219.
Yao Y-G, Kong Q-P, Salas A, Bandelt H-J. 2008. Pseudomitochondrial genome haunts disease studies. *J Med Genet*. 45:769–772.
Yao Y-G, Kong Q-P, Wang C-Y, Zhu C-L, Zhang Y-P. 2004. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Mol Biol Evol*. 21:2265–2280.
Yao Y-G, Nie L, Harpending H, Fu Y-X, Yuan Z-G, Zhang Y-P. 2002. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *Am J Phys Anthropol*. 118: 63–76.
Yao Y-G, Salas A, Bravi CM, Bandelt H-J. 2006. A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum Genet*. 119:505–515.
Yao Y-G, Salas A, Logan I, Bandelt H-J. 2009. mtDNA data mining in GenBank needs surveying. *Am J Hum Genet*. 85:929–933.
Yao Y-G, Zhang Y-P. 2002. Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet*. 47:311–318.
Zhang Y, Xu Q, Cui H, Cui Y, Lin H, Kim K, Lee J. 2005. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in a Korean ethnic group from northeast China. *Forensic Sci Int Genet*. 151:299–301.
Zhang YJ, Xu QS, Zheng ZJ, Lin HY, Lee JB. 2005. Haplotype diversity in mitochondrial DNA hypervariable region I, II and III in northeast China Han. *Forensic Sci Int Genet*. 149:267–269.
Zhao M, Kong Q-P, Wang H-W, et al. (15 co-authors). 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci U S A*. 106:21230–21235.
Zimmermann B, Bodner M, Amory S, Fendt L, Rock A, Horst D, Horst B, Sanguansermsri T, Parson W, Brandstatter A. 2009. Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). *Int J Legal Med*. 123:495–501.