Am. J. Hum. Genet. 73:671–676, 2003

Repor_

Ph logen of Ea_ A_ian Mi_ochondrial DNA Lineage_ Inferred from Comple_e Seq ence_

Qing-Peng Kong,[1,2,3] Yong-Gang Yao,[1,3] Chang Sun,[1,3] Hans-Jürgen Bandelt,[4] Chun-Ling Zhu,[1] and Ya-Ping Zhang[1,2]

[1]Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, and [2]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming, Yunnan, China; [3]Graduate School of the Chinese Academy of Sciences, Beijing; and [4]Fachbereich Mathematik, Universität Hamburg, Hamburg

The now-emerging mitochondrial DNA (mtDNA) population genomics provides information for reconstructing a well-resolved mtDNA phylogeny and for discerning the phylogenetic status of the subcontinentally specific haplogroups. Although several major East Asian mtDNA haplogroups have been identified in studies elsewhere, some of the most basal haplogroups, as well as numerous minor subhaplogroups, were not yet determined or fully characterized. To fill the lacunae, we selected 48 mtDNAs from >2,000 samples across China for complete sequencing that cover virtually all (sub)haplogroups discernible to date in East Asia. This East Asian mtDNA phylogeny can henceforth serve as a solid basis for phylogeographic analyses of mtDNAs, as well as for studies of mitochondrial diseases in East and Southeast Asia.

Recent progress in the analysis of complete or nearly complete mtDNA sequences has provided new insights into the origin and spread of modern humans and the phylogeny of the major African, European, Asian, and Native American mtDNA lineages (Ingman et al. 2000; Finnilä et al. 2001; Maca-Meyer et al. 2001; Torroni et al. 2001; Derbeneva et al. 2002; Herrnstadt et al. 2002; Mishmar et al. 2003). These studies differ in regard to sequencing technique (viz., Maca-Meyer et al. [2001] employed manual sequencing), inclusion of the control region (which was not disclosed by Herrnstadt et al. [2002]), and sampling scheme: mtDNAs either were chosen according to the language spoken by their bearers (Ingman et al. 2000), were randomly selected from a certain geographic range (Finnilä et al. 2001; Herrnstadt

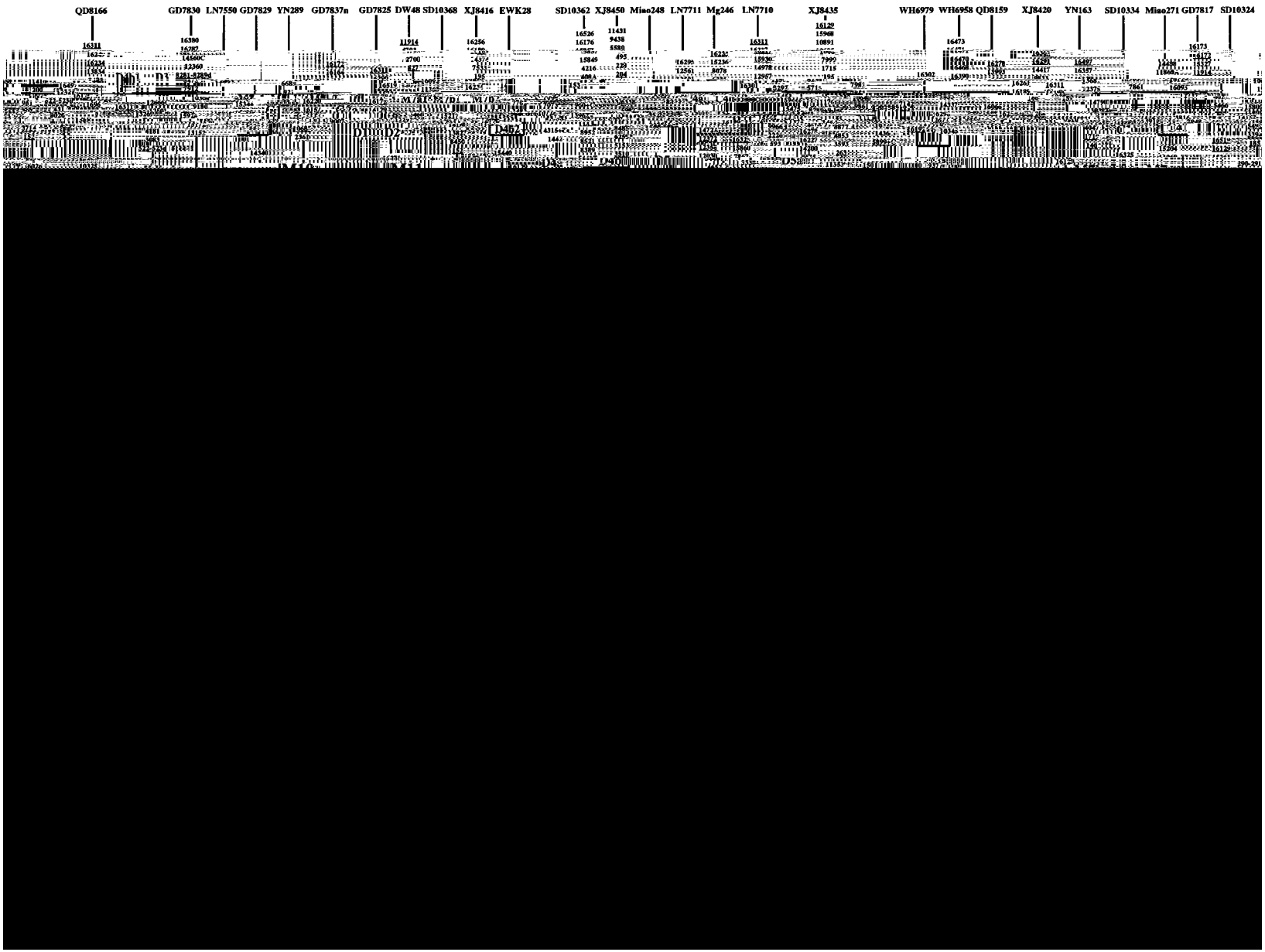
QD8166
GD7830
LN7550
GD7829
YN289
GD7837n
GD7825
DW48
SD10368
XJ8416
EWK28
SD10362
XJ8450
Miao248
LN7711
Mg246
LN7710
XJ8435
WH6979
WH6958
QD8159
XJ8420
YN163
SD10334
Miao271
GD7817
SD10324

Fig re 1 Phylogenetic tree of 48 East Asian mtDNA lineages, which were sampled from various regional Han populations (Yao et al. 2002*a*, 2003*a*), except for those with prefixes DW (Daur, from Inner Mongolia), EWK (Ewenki, from Inner Mongolia), Mg (Mongolian, from Inner Mongolia), and Miao (Miaozu, from Hunan). This tree incorporates the information drawn from previous reports (Derbeneva et al. 2002; Herrnstadt et al. 2002; Kivisild et al. 2002, and references therein; Mishmar et al. 2003) by indicating the roots of several haplogroups from which those Asian sequences branch off. Also indicated are the Nativel. haplogroupsB1(aeltan)-368.e(o)asto pss03) anm

here, we opt for the one that places a forward mutation at site 16304 on the way to haplogroup R9. This prompts yet another broadening of haplogroup F, which is now recognizable by an "A" deletion scored at site 249 and transitions at sites 6392 and 10310. Haplogroup F thus encompasses haplogroups F1, F2, and F3 (originally called "R9a" in Yao et al. [2002*a*]). The definition of

potential oversights and artificial recombination (e.g., as shown in Yao et al. [2003*b*] and Yao and Zhang [2003]). The B5b sequence of a patient suffering from LHON and cardiomyopathy recently reported by Mimaki et al. (2003) evidently missed a batch of mutations relative to the rCRS (73, 204, 263, 1438, 8281–8289del, 8584, 10398, 15223, 16140, and 16189).

The mutational pattern can also be studied in detail with a large complete mtDNA phylogeny at hand. For instance, transversions A→G or T→G are apparently rather rare in the coding region (cf. Herrnstadt et al. [2003]). The only shared transversions to G in the Eurasian mtDNA tree reported to date by more than one lab seem to be 961G in haplogroup H, 12083G in haplogroup I, and 12738G in haplogroup K1 (Ingman et al. 2000; Finnilä et al. 2001; Maca-Meyer et al. 2001; Herrnstadt et al. 2002). Further transversions to G found in lineages from haplogroups J, T, W, and X by Mishmar et al. (2003) may thus be problematic, at least 14974G (Herrnstadt et al. 2003). On the other hand, indels in the coding region seem to occur at an absolute frequency comparable with that of transversions but might be missed occasionally, owing to conservative reading of ambiguous sequencer outputs. For example, only a single private coding-region indel (15944d in an African haplogroup L1c lineage) can be scored in the 53 complete mtDNA sequences of Ingman et al. (2000) (contrast this to nine private indels and five shared ones detected in our 48 complete mtDNA sequences); moreover, their single haplogroup F sequence (closely related to the lineage XJ8440 of Yao et al. [2002*a*]) misses the 249 deletion. We agree with Herrnstadt et al. (2003) that the solution to the problem of mtDNA databases containing errors "is further effort, both at the front end (the sequencing process itself) and at the back end (increased quality control), of mtDNA database construction."

In short, the phylogenetic tree of East Asian mtDNAs obtained in the present study covers all of the major haplogroups in the region and testifies to the phylogenetic status of the newly identified haplogroups (Kivisild et al. 2002; Yao et al. 2002*a*, 2003*a*; authors' unpublished data) that were formerly defined on the basis of control-region and/or only partial coding-region information. This tree, then, can serve as a basis for haplogroup inferences in future studies of East Asian populations and for distinguishing pathogenic mutations from rare polymorphisms in mtDNA medical genetics.

Acknowledgments

We thank Shi-Fang Wu for technical assistance. We also thank Dr. Vincent Macaulay for helpful comments on the manuscript. This study was supported by grants from Chinese Academy of Sciences (KSCX2-SW-2010), Natural Sciences Foundation of China, and Natural Sciences Foundation of Yunnan Province.

Electronic-Database Information

Accession numbers and URL for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for the mtDNA complete sequence data [accession numbers AY255133–AY255180])

References