The pioneer settlement of modern humans in Asia.

Metspalu M, Kivisild T, Bandelt H-J, Richards M, Villems R

In: Bandelt H-J, Macaulay V, Richards M (eds) Human mitochondrial DNA and the evolution of Homo sapiens. Springer-Verlag, Heidelberg

Supplementary material

MtDNA diversity in Asia

Complete and partial mtDNA coding region sequences have been used to map the backbone and determine the fine-structure of the mtDNA lineages present in East Asia (Kivisild et al. 2002; Yao et al. 2002; Kong et al. 2003). The more recent analysis of complete mtDNA sequences from 672 Japanese individuals has provided a significant refinement of the East Asian mtDNA phylogeny (Tanaka et al. 2004). Combining these and other published data, Figure 1 summarizes the Asian mtDNA tree topology. With many manuscripts that refine the phylogeny being published almost simultaneously it is not surprising then that some confusion regarding to the naming of the haplogroups and their branching order arises. In the following section we shall try to overcome some of these difficulties.

Comments for the Asian mtDNA phylogeny outlined on Figure 1

First of all, we acknowledge that the monophyletic decent of R11 and B; D5 and D6; D4k cannot be taken for granted because their defining positions are known to be highly variable.

Tanaka et al. (2004) have suggested that transition at np 15924 is basal and shared between M11 and M12. In contrast this mutation was not detected in the three complete M11 sequences by Kong et al. (2003) including one M11b sequence (defined by 14790). Because this mutation is also recurrent in other branches of the mtDNA tree (Figure 1) we assume for the moment - pending further information becoming available- the independent origin of the 15924 mutation in M11b and M12 subclades and do not show its position on the tree in Figure 1.

Despite observing the common substitution at np 4491 in haplogroups M9a and E, Tanaka et al. (2004) did not draw E as a subclade of M9. Following the rule of parsimony, Figure 1 shows E nested within M9, but to cope with the unknown phylogenetic placement of mutations at highly variable np 16362 these are not reconstructed in M9.

Haplogroup G1 was originally identified by Schurr et al. (1999) as a clade defined by transition at 16017. Recently three mutations defining G1 in the coding region (at nps 8200, 15232, and 15497) have been acknowledged (Bandelt et al. 2003; Kong et al. 2003). Indeed, all haplogroup G haplotypes (SIB08, SIB35-39, SIB60-61) in Schurr et al. (1999) have the substitution at np 8200 and all but one at np 15497 (a probable back-mutation or mistake since this haplotype -SIB08- has both downstream transitions at np 16017 and 16129) defining all of them in G1. Introducing G5 to replace G1 of Schurr et al. (1999) by Tanaka et al. (2004) is therefore unnecessary. Bandelt et al. (2003) identified G1a and G1b. The former was characterized by motif 15860–16325–150 while

the latter by transition at np 16017. Tanaka et al. (2004) found that substitution at np 7867 is basal for G1 and noticed a new subclade G1a2 while G1 defined by Bandelt et al. (2003) transformed to G1a1 (Figure 1). G4a and G4b introduced by Maruyama et al. (2003) correspond to G1a1 and G1a2 and we suggest retaining the latter nomenclature. G4 used in Derenko et al. (2003) also corresponds to G1a1.

The G subclade with a characteristic mutation at np 16274 has been classified as G3 (Kivisild et al. 2002). In Tanaka 2004 one of its subclades is classified as G4. Here we retain the original G3 and the novel G subclade characterized by Tanaka et al. (2004) as G3 is relabeled as G4 here.

New information presented by Tanaka et al. (2004) suggests that G2a definition by Kong et al. (2003) should be now broadened, so that the previous G2a will become G2a1a (see Figure 1 for details).

D2 and D3 (Derbeneva et al. 2002) are classified within D4e1 and D4b1 in Tanaka et al. (2004). Here we retain the historical haplogroup names within the renewed tree topology.

Comas et al. (2004) identified a subclade of D4 defined by transition at np 16245 and baptized it D4c. The same clade appears in Tanaka et al. (2004) as D4d (probably because the latter manuscript was in press when the former was published). We stick with the labeling proposed by Comas et al. (2004) and rename D4c of Tanaka et al. (2004) D4d.

Tanaka et al. (2004) conclude that the substitution at np 8473 is basal for all D4a lineages. However, the complete mtDNA sequence published by Tawata et al. (2000) lacks it but has the other D4a sites.

The data of Tanaka et al. (2004) suggests a deeper branching of the D4b lineage introducing a new subhaplogroup D4b2. In fact, sample KT12 (Ingman et al. 2000), also falls into D4b2. The new branching supports the renaming of previous D4b1 and D4b2 (Kong et al. 2003) as D4b1a and D4b1b, respectively. Note that D3 is a sister-clade of the latter two. The new topology of D4b points to a likely transition at np 15440 overlooked by Ozawa (1995) (sample Oz5 HCM-P2 also in Kivisild et al. (2002).

The sample MELAS P1 (Ozawa et al. 1991) classified as D4 in Kivisild et al. (2002) falls into D4d1a according to the newly refined classification (Tanaka et al. 2004).

Unclassified D4 lineage Ln7550 in Kong et al. (2003) shares a substitution at np 11696 with D4j of (Tanaka et al. 2004).

The absence of the transition at np 9180 in the sample PD-P2 (Ozawa et al. 1991) is likely a mistake. According to several complete sequences (Tanaka et al. 2004) this position is basal for D5a and D5b, and the latter sample belongs clearly to the former haplogroup.

The topology of haplogroup X follows Reidla et al. (2003).

A sister-group of haplogroup W that shares substitutions at nps. 189, 709, 5046 and 11674 with the latter was identified by Derbeneva et al. (2002). Our unpublished work indicates that, of the additional coding region mutations revealed by full sequencing (Derbeneva et al. 2002), positions 199, 739, 7581, 16106, 16153 and 16223 are characteristic to the newborn haplogroup N2a.

To avoid potential confusion we renamed A5 proposed by Tanaka et al. (2004) as A6 because A5 has been defined before through transition at np 16187. Further, because the original A5 is clearly a subset of the clade defined by 8563 and 11536 (A1 in Tanaka

et al. (2004) we call the whole clade A5 and the original A5 A5a. According to Kivisild et al. (2002) 16362 defines A4 of which A2 defined by 153, 8027, 12007 and 16111 is a subset of. In (Tanaka et al. 2004) A2 appears as A2a and two other A4 lineages as A2. A4 was redefined by Tanaka et al. (2004) as a lineage with three coding area mutations and a characteristic control region motif (16051-16129-16189-16223-16290-16319-146-235). Here we rename this lineage as A7 (keeping the original A4 which comprises A2) and reduce the number of its defining polymorphisms incorporating data from Yao et al. (2002) where substitutions at nps 146 and 10172 were not recorded in a sample (WH6956) with A7 HVSI motif. Further, there might be an even earlier offshoot of A7 in the same dataset. Sample YN271 is lacking transitions at np 16129 and 16189 but has one at np 16051. Tanaka et al. (2004) also baptize A3, although giving names to clades represented by only one complete sequence is probably preliminary.

In haplogroup A the transition at the highly variable np 152 is mapped according to Tanaka et al. (2004), although for the lineages presented here a single gain of the transition on the lineage towards A5c would be more parsimonious.

R10 was defined by Yao and Zhang (2002).

The basal positions defining haplogroup U7 were deduced from the published complete mtDNA sequences (Finnilä et al. 2001; Maca-Meyer et al. 2001).

U2e has been defined by the transversion at np. 16129 (Kivisild et al. 1999). Murci et al. (2004) added to the list two coding are sites 13734 and 15907. However, comparing the full sequence data of Finnilä et al. (2001), Maca-Meyer et al. (2001), Herrnstadt et al. (2002), it is apparent that 13734 characterizes a subset of U2e. Though additional complete sequence data could reduce the number of U2e defining positions, here we use the minimal set required as deduced from the full sequencing data.

The topology of haplogroup R9 follows Kong et al. (2003) while F4 defined by Tanaka et al. (2004) is added. Note that F4 defined by Kong et al. (2004) shares the control region motif with F4a identified by Tanaka et al. (2004).

We have changed the topology of B4c1 (Tanaka et al. 2004) by moving the transition at np 16311 to the basal level. This substitution is present in all the three sub-groups of B4c1 with an exception of one sub-lineage of B4c1b. A gain and a subsequent loss of the substitution is a more parsimonious scenario than three separate gains on three sister branches.

References

Bandelt H-J, Herrnstadt C, Yao Y-G, Kong Q-P, Kivisild T, Rengo C, Scozzari R, Richards M, Villems R, Macaulay V, Howell N, Torroni A, Zhang Y-P (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67:512-524

Comas D, Plaza S, Wells RS, Yuldaseva N, Lao O, Calafell F, Bertranpetit J (2004) Admixture, migrations, and dispersals in Central Asia: evidence from maternal DNA lineages. Eur J Hum Genet 12:495-504

Derbeneva OA, Sukernik RI, Volodko NV, Hosseini SH, Lott MT, Wallace DC (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

Derenko MV, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, Zakharov IA (2003) Diversity of mitochondrial DNA lineages in South Siberia. Ann Hum Genet 67:391-411

Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. Am J Hum Genet 68:1475-1484.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (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.

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

Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, Laos S, Parik J, Watkins WS, Dixon ME, Papiha SS, Mastana SS, Mir MR, Ferak V, Villems R (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. Curr Biol 9:1331-1334

Kivisild T, Tolk H-V, 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 (erratum 1720:1162)

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

Kong QP, Yao YG, Sun C, Zhu CL, Zhong L, Wang CY, Cai WW, Xu XM, Xu AL, Zhang YP (2004) Phylogeographic analysis of mitochondrial DNA haplogroup F2 in China reveals T12338C in the initiation codon of the ND5 gene not to be pathogenic. J Hum Genet 49:414-423

Maca-Meyer N, González AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2:13

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

Ozawa T (1995) Mechanism of somatic mitochondrial DNA mutations associated with age and diseases. Biochim Biophys Acta 1271:177-189

Ozawa T, Tanaka M, Ino H, Ohno K, Sano T, Wada Y, Yoneda M, Tanno Y, Miyatake T, Tanaka T, Itoyama S, Ikebe S, Hattori N, Mizuno Y (1991) Distinct clustering of point mutations in mitochondrial DNA among patients with mitochondrial encephalomyopathies and with Parkinson's disease. Biochem Biophys Res Commun 176:938-946

Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk H, Parik J, et al. (2003) Origin and Diffusion of mtDNA Haplogroup X. Am J Hum Genet 73:1178-1190

Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk Sea-Bering Sea region during the Neolithic. Am J Phys Anthropol 108:1-39.

Tanaka M, Cabrera VM, Gonzalez AM, Larruga JM, Takeyasu T, Fuku N, Guo L-J, et al. (2004) Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res 14:1832–1850

Tawata M, Hayashi JI, Isobe K, Ohkubo E, Ohtaka M, Chen J, Aida K, Onaya T (2000) A new mitochondrial DNA mutation at 14577 T/C is probably a major pathogenic mutation for maternally inherited type 2 diabetes. Diabetes 49:1269-1272

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, 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