

ESM1: 31 mtGenomes rejected in the first step that were again added to the EMMA database

Accession number	Original publication	Haplogroup in Phylotree Build 15	Initial reason for exclusion from EMMA
AF346973	Ingman et al. 2000	F4a1	highlighted as problematic in Yao et al. 2009
DQ523642	Fraumene et al. 2006	X2b3	highlighted as problematic in Yao et al. 2009
DQ523643	Fraumene et al. 2006	H3x1	highlighted as problematic in Yao et al. 2009
DQ523647	Fraumene et al. 2006	H3u	highlighted as problematic in Yao et al. 2009
DQ523649	Fraumene et al. 2006	T2b3a	highlighted as problematic in Yao et al. 2009
DQ523662	Fraumene et al. 2006	H3x1	highlighted as problematic in Yao et al. 2009
EF660936	Gasparre et al. 2007	HV1a'b'c	highlighted as problematic in Yao et al. 2009
EF660956	Gasparre et al. 2007	H13b1	highlighted as problematic in Yao et al. 2009
GU733719	Gunnarsdottir et al. 2011	M74b1	2nd generation sequencing data
GU733724	Gunnarsdottir et al. 2011	B4b1a2c	2nd generation sequencing data
GU733734	Gunnarsdottir et al. 2011	M74b1	2nd generation sequencing data
GU733737	Gunnarsdottir et al. 2011	R9b1a1	2nd generation sequencing data
GU733740	Gunnarsdottir et al. 2011	N11b	2nd generation sequencing data
GU733744	Gunnarsdottir et al. 2011	B4b1a2c	2nd generation sequencing data
GU733776	Gunnarsdottir et al. 2011	N11b	2nd generation sequencing data
GU733790	Gunnarsdottir et al. 2011	B4b1a2b	2nd generation sequencing data
GU733814	Gunnarsdottir et al. 2011	B4b1a2d	2nd generation sequencing data
GU733815	Gunnarsdottir et al. 2011	B4b1a2b	2nd generation sequencing data
GU733825	Gunnarsdottir et al. 2011	B4b1a2d	2nd generation sequencing data
HM596670	Gunnarsdottir et al. 2011	M26	2nd generation sequencing data
HM596703	Gunnarsdottir et al. 2011	N9a6	2nd generation sequencing data
HM596710	Gunnarsdottir et al. 2011	M26	2nd generation sequencing data
HM852761	Schönberg et al. 2011	HV1a1	2nd generation sequencing data
HM852777	Schönberg et al. 2011	U7a4	2nd generation sequencing data
HM852790	Schönberg et al. 2011	U1a	2nd generation sequencing data
HM852797	Schönberg et al. 2011	U3c	2nd generation sequencing data
HM852803	Schönberg et al. 2011	U3c	2nd generation sequencing data
HM852823	Schönberg et al. 2011	U7a4	2nd generation sequencing data
HM852844	Schönberg et al. 2011	U1a'c	2nd generation sequencing data

References:

- Fraumene,C., Belle,E.M., Castri,L., Sanna,S., Mancosu,G., Cosso,M., Marras,F., Barbujani,G., Pirastu,M. and Angius,A., High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates, *Mol. Biol. Evol.* 23 (11), 2101-2111 (2006)
- Gasparre,G., Porcelli,A.M., Bonora,E., Pennisi,L.F., Toller,M., Iommarini,L., Ghelli,A., Moretti,M., Betts,C.M., Martinelli,G.N., Ceroni,A.R., Curcio,F., Carelli,V., Rugolo,M., Tallini,G. and Romeo,G., Disruptive mitochondrial DNA mutations in complex I subunits are markers of oncocytic phenotype in thyroid tumors, *Proc. Natl. Acad. Sci. U.S.A.* 104 (21), 9001-9006 (2007)
- Gunnarsdottir,E.D., Li,M., Bauchet,M., Finstermeier,K. and Stoneking,M., High-throughput sequencing of complete human mtDNA genomes from the Philippines, *Genome Res.* 21 (1), 1-11 (2011)
- Ingman,M., Kaessmann,H., Paabo,S. and Gyllensten,U., Mitochondrial genome variation and the origin of modern humans, *Nature* 408 (6813), 708-713 (2000)
- Schönberg A, Theunert C, Li M, Stoneking M, Nasidze I., High-throughput sequencing of complete human mtDNA genomes from the Caucasus and West Asia: high diversity and demographic inferences, *Eur J Hum Genet.* 19 (9), 988-94 (2011)