

Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA

The Human Genome Project, from one perspective, began in 1981 with the publication¹ of the complete sequence of human mitochondrial DNA (mtDNA). The Cambridge reference sequence (CRS), as it is now designated, continues to be indispensable for studies of human evolution, population genetics and mitochondrial diseases. It has been recognized for some time, however, that the CRS differs at several sites from the mtDNA sequences obtained by other investigators^{2,3}. These discrepancies may reflect either true errors in the original sequencing analysis or rare polymorphisms in the CRS mtDNA. A further complication is that the original mtDNA sequence was principally derived from a single individual of European descent, although it also contained some sequences from both HeLa and bovine mtDNA (ref. 1). To resolve these uncertainties, we have completely resequenced the original placental mtDNA sample.

The results of the resequencing confirm that there are both errors and rare polymorphisms in the CRS (Table 1). There are 11 nucleotide positions at which the original sequence contains the incorrect nucleotide (all sequences here refer to that of the CRS mtDNA L-strand), one of which involves the CC doublet at posi-

tions 3,106 and 3,107 in the original sequence, which is actually a single cytosine residue. The other errors are mistakes in the identification of single base pairs and typically involve the incorrect assignment of a guanine residue. Errors at nt 14,272 and 14,365 result from the use of the bovine mtDNA sequence at ambiguous sites¹. There are an additional seven nucleotide positions at which the original CRS is correct and which represent rare (or even private) polymorphic alleles. Six of these polymorphisms are single base pair substitutions, although one involves the simple repeat of cytosine residues at nt 311–315. The CRS has five residues in this repeat, whereas most other human mtDNAs have six.

To determine which errors in the CRS resulted from the use of HeLa mtDNA sequence, we sequenced the HeLa mitochondrial genome in its entirety (data not shown). It is of African origin and divergent from the published CRS at a number of sites. The only error that we were able to explain using the HeLa sequence was that at nt 14,766 (T versus C, respectively).

The revised CRS mtDNA belongs to European haplogroup H on the basis of the cytosine at nt 14,766 (instead of the thymine in the original CRS) and the cyto-

sine at nt 7,028 (instead of a thymine^{4–6}). The assignment of the revised CRS to haplogroup H is confirmed by the absence of any of the predicted restriction site changes that characterize the other European mtDNA haplogroups^{4,6}.

If we include both the technical errors (8) and the errors (3) introduced by the assumption of the bovine and HeLa mtDNA sequences, the overall error frequency for the original CRS analysis was 0.07%, emphasizing the remarkable *tour de force* by Sanger and colleagues. Nevertheless, because of its widespread use, we recommend that the CRS for human mtDNA should be revised as follows: (i) the ten simple substitution errors should be corrected; (ii) the rare polymorphic alleles should be retained (that is, the revised CRS (RCRS) should be a true reference sequence and not a consensus sequence); and (iii) the original nucleotide numbering should be retained. The last suggestion represents a compromise between accuracy (correcting the numbering to account for the single C residue at nt 3,106 and 3,107) and consistency with the previous literature. We believe that renumbering all of the previously identified sequence changes beyond nt 3,106 would create an unacceptable level of confusion.

Acknowledgements

We thank A. Coulson and F. Sanger for their generous donation of the original placental mtDNA sample. This work was supported by the Medical Research Council (R.M.A. and D.M.T.), Wellcome Trust (D.M.T., P.F.C., R.N.L. and N.H.) and the John Sealy Memorial Endowment Fund (N.H.).

Richard M. Andrews¹, Iwona Kubacka², Patrick F. Chinnery³, Robert N. Lightowlers³, Douglass M. Turnbull³ & Neil Howell²

¹Departments of Ophthalmology and Neurology, The Medical School, University of Newcastle upon Tyne, NE2 4HH, UK. ²Biology Division 0656, Department of Radiation Oncology, The University of Texas Medical Branch, Galveston, Texas 77555-0656, USA.

³Department of Neurology, The Medical School, University of Newcastle upon Tyne, NE2 4HH, UK. Correspondence should be addressed to D.M.T. (e-mail: d.m.turnbull@newcastle.ac.uk).

1. Anderson, S. et al. *Nature* **290**, 457–465 (1981).
2. Brown, M.D. et al. *Genetics* **130**, 163–173 (1992).
3. Howell, N., McCullough, D.A., Kubacka, I., Halvorson, S. & Mackey, D. *Am. J. Hum. Genet.* **50**, 1333–1337 (1992).
4. Torroni, A. et al. *Genetics* **144**, 1835–1850 (1996).
5. Lammiman, T. et al. *Eur. J. Hum. Genet.* **5**, 271–279 (1997).
6. Torroni, A. et al. *Am. J. Hum. Genet.* **62**, 1137–1152 (1998).
7. Howell, N. & McCullough, D.A. *Am. J. Hum. Genet.* **47**, 629–634 (1990).
8. Howell, N. et al. *Genetics* **140**, 285–302 (1995).
9. Biggin, M.D., Gibson, J.J. & Hong, G.F. *Proc. Natl Acad. Sci. USA* **80**, 3963–3965 (1983).

Table 1 • Reanalysis of the Cambridge reference sequence

Nucleotide position	Previous sequence	Reanalysed sequence	Remarks
263	A	A	rare polymorphism
311–315	CCCCC	CCCCC	rare polymorphism
750	A	A	rare polymorphism
1,438	A	A	rare polymorphism
3,106–3,107	CC	C	error
3,423	G	T	error
4,769	A	A	rare polymorphism
4,985	G	A	error
8,860	A	A	rare polymorphism
9,559	G	C	error
11,335	T	C	error
13,702	G	C	error
14,199	G	T	error
14,272	G	C	error (bovine)
14,365	G	C	error (bovine)
14,368	G	C	error
14,766	T	C	error (HeLa)
15,326	A	A	rare polymorphism

We determined the complete mtDNA sequence using a series of 28 overlapping PCR-amplified fragments spanning the entire length of the mitochondrial genome. Both strands of each PCR fragment were sequenced with BigDye terminator cycle sequencing chemistry on an Applied Biosystems model 377 automated nucleotide sequencer. In addition, several regions of the CRS mtDNA were sequenced independently using a different approach^{7,8}. In brief, small PCR-amplified mtDNA fragments of ~300 bp were cloned into M13 sequencing vectors and the mtDNA sequence determined using a dideoxy chain termination method⁹. The two approaches yielded identical results.