Int J Legal Med (2001) 115 : 64–69

ORIGINAL ARTICLE

H.-J. Bandelt · P. Lahermo · M. Richards · V. Macaulay

Detecting errors in mtDNA data by phylogenetic analysis

Received: 30 January 2001 / Accepted: 2 April 2001

Abstract Sequencing and documenting a sample of homologous DNA stretches is prone to copying errors in a way rather analogous to the biological replication process. Previous attempts at obtaining representative mtDNA sequences, typically of the control region, for evolutionary studies or forensic purposes have yielded rather unsatisfactory results in many cases. The key ingredient in pinpointing problems with given data is the phylogenetic analysis of closely related mtDNAs within the framework of an established worldwide phylogeny that is supported by coding region information. We develop some general rules by which likely errors in data tables can readily be detected without rereading whole sequences repeatedly. Following these guidelines, one can expect to lower the error rate by at least an order of magnitude, although it will still be hard to beat the mitochondrial gamma polymerase in precision.

Keywords Mitochondrial DNA · Error detection · Sequence data · Phylogenetic analysis · Types of error

Introduction

Sequencing a segment of mitochondrial DNA (commonly within the control region) and documenting the result in a database or on the printed page, involves numerous cycles of in vitro and virtual replication of the sampled DNA. As with any process of replication, the copying is hardly ever perfect and some errors may be introduced. PCR does not necessarily produce unanimous results, especially with degraded (ancient) DNA (Krings et al. 1997), but replication does not stop with the final PCR product. DNA has its scribes: the graduate student or technical assistant reading sequencing outputs and formatting them into a data table plays the same role as formerly did the mediaeval monk who copied a piece of poetry. Fortunately, the scholar of genetics may write down his AGCT text at a considerably lower error rate than the quondam monk who did not always completely understand what he was writing and would even introduce changes (intended as re-correction) according to the perceived context, but still the gamma polymerase in the living cell will outperform the scholar by 1–3 orders of magnitude in accuracy.

Errors can arise at all stages: samples may be contaminated or incorrectly recorded and exchanged, biochemical problems with the sequencing kit may lead to false nucleotide signals, misreading of the sequencing output may occur, data tables may be distorted and the final printing process may go wrong. How can we tell from a single entry of an mtDNA data table whether it is authentic? Well, of course, we cannot with absolute certainty, but three circumstances allow good predictions: first, thousands of homologous mtDNA segments already exist in the database for comparison, second mtDNA evolves along a tree and third the virtual replication process seems to obey different rules from those used by the cell during replication, so that natural mutations show a pattern different from artefactual mutations. Phylogenetic analysis that reconstructs ancestral sequences (such as by parsimony methods) is the key tool in understanding the structure of the mtDNA data under study. Assigning new mtDNA types to a place in the global mtDNA tree has become more straightforward in recent years as knowledge of the tree has improved: consider the sequence of papers by Richards et al. (1996, 1998, 2000). This is based on information from combined high-resolution RFLP/control region studies

H.-J. Bandelt
Fachbereich Mathematik, Universität Hamburg,
20146 Hamburg, Germany

P. Lahermo
Finnish Genome Center, University of Helsinki,
00014 Helsinki, Finland

M. Richards
Department of Chemical and Biological Sciences,
University of Huddersfield, Huddersfield, UK

V. Macaulay (✉)
Department of Statistics, University of Oxford,
1 South Parks Road, Oxford, OX1 3TG, UK
e-mail: macaulay@stats.ox.ac.uk,
Fax: +44-1865-272595

and complete sequencing studies; control region motifs have been identified for many of the major haplogroups and their sub-haplogroups (Macaulay et al. 1999a; Quintana-Murci et al. 1999; Alves-Silva et al. 2000; Richards et al. 2000; Kivisild et al. unpublished data). Haplogroups are (major) monophyletic clades in the mtDNA tree (Torroni et al. 2000) which are highlighted for convenience. Be aware that not all nested clades have been given names yet. On the other hand, some so-called haplogroups named in the early literature have turned out to be paraphyletic groupings, due to insufficient coding region information. The ideal approach is therefore to confirm diagnostic coding region sites directly (e.g. by restriction analysis). A short cut that may be effective in most cases would be to identify the motif present in the new sequence and proceed from the portion of the database that it defines. If this fails, a neighbourhood search for sequences in the database may identify a subset of potentially close relatives, pinpointing one or more relevant haplogroups. In any case, it becomes possible to compare the sequence in question with a limited subset of the total database.

Suppose we have allocated a sequence to its putative position in the worldwide phylogeny of homologous sequences, then we can envision three cases:

(1) The sequence matches a type already found
(2) It is located at a previously unsampled interior node of the tree
(3) It creates a new tip of the tree.

In case (2), this might actually constitute a back-mutation, especially if it is a 1-step neighbour of a frequently sampled type. In case (3) the mutations at the corresponding terminal link are apparently private. If artefacts occurred in the data set, then they are normally manifest in those seeming back and private mutations.

Of course, real back-mutations and unfortunate parallelisms do occur. For example, the sequence type 16189-16278 (Macaulay et al. 1999a), which displays part of the HVS-I motif of haplogroup X, would immediately prompt one to check the site 16223 which is present in the X motif. (Throughout, we use "HVS-I" and "HVS-II" to refer to the first and second hypervariable segments of the control region, respectively and employ the numbering scheme of the reference sequence, CRS; Anderson et al. 1981). In this case, the sequence was confirmed: position 16223 had indeed experienced a back-mutation. The sequence type 16129-16223-16292 (Richards et al. 2000) harbours the HVS-I motif of haplogroup W, but also two-thirds of the motif of haplogroup I. In this case, one would check the remaining site in the haplogroup I motif (16391). It is confirmed absent and an assignment to haplogroup W (with a postulated parallel mutation at 16129) becomes the most convincing classification.

Classification of errors

There are five major types of errors which can be distinguished in a taxonomy of artefacts, prefigured by Bandelt et al. (1995). We illustrate the various kinds of errors with examples from the literature. Further instances are reported by Röhl et al. (2001).

Base shift (Type I)

This can occur as follows:

a. By a single position e.g. the single haplogroup A sequence in Moraga et al. (2000) has been scored at 234 in HVS-II rather than at 235 which is characteristic for haplogroup A (Alves-Silva et al. 2000)
b. By several positions (often 10 or 100 nucleotides) e.g. the infamous 16076 in place of 16086 etc. by Hagelberg et al. (1999, 2000)
c. By manipulations in the set-up of the data table, which may result in the shifting of a variant nucleotide from one column to the next e.g. Salas et al. (1998) shifted 16261 to 16270, an event testified to by a superfluous dot in their table. If the unique variant base in some column is shifted to a neighbouring column, then an empty column arises, which signposts the fatal base shift e.g. column 16171 in Table 2 of Nishimaki et al. (1999).

Reference bias (Type II)

This can be caused by:

a. Overlooking a single variant base relative to the CRS e.g. at 16362 omitted in several cases in the data of Sykes et al. (1995) according to a resequencing by Melton et al. (1998)
b. Missing the variation in a short stretch because of reading difficulties, as discussed by Saillard et al. (2000).

Phantom mutations (Type III)

These can be due to:

a. Random mutations, often transversions to G e.g. in the Yanomami data of Easton et al. (1996), as recognised by Merriwether et al. (2000)
b. Systematic appearances of spurious mutations, which may be incurred by biochemical problems (R. Villems, Estonian Biocentre, pers. comm.) e.g. 16106, 16221 and further transitions in Stenico et al. (1996), which are distributed over lineages belonging to at least three different (sub)haplogroups and which are rare in other studies.

Base misscoring (Type IV)

Due to:

a. Repeating the CRS base instead of the variant base in a "dot" table as at 203 in mtDNA H93 and at 16173 in P81 of Lee et al. (1997)

b. Scoring a transversion as a transition, especially in the case when the data format is not a dot table but a listing of variant sites, with transversions indicated by suffixes e.g. Watson et al. (1997) dropped the suffix T when citing the sequence type 16093-16223-16265T
c. Scoring a base as deleted e.g. 16208/9 by Sajantila et al. (1995)
d. Neglecting to record an insertion/deletion event (indel) e.g. 248del or 249del in Fig. 2 of Redd and Stoneking (1999)
e. Confounding the letters C and G as at 16362 in mtDNA P122 of Lee et al. (1997).

Artefactual recombination (Type V)

This can be caused:

a. By contamination
b. By combining different stretches of DNA from different individuals e.g. as in sequence #105 from the Swiss data of Dimo-Simonin et al. (2000), which combines a typical European haplogroup T HVS-I with a west African haplogroup L1b HVS-II (which can be identified by the motif 182-185T-247-357)
c. By shifting a single variant base into a neighbouring row e.g. from sequence #32 to #33 of the British data of Piercy et al. (1993), where the haplogroup J1a marker mutation 215 (cf. #12 and #19) has dropped into an innocent sequence from haplogroup H (in which this mutation has never been observed otherwise)
d. Row crossover of longer sequence stretches in the data table e.g. #30 and #31 in the Bulgarian HVS-I/HVS-II data of Calafell et al. (1996) (as shown by Bandelt et al. 2000, p 23)
e. Recombination with the CRS or another sequence by copying parts of rows in the assembly of a "dot" table e.g. in the paper of Rousselet and Mangin (1998), #38 matches the CRS in HVS-I but bears the HVS-II motif 242-295 for haplogroup J1b1, just as does #34
f. Enumeration shift in haplotype numbering in part of the data tables (Alves-Silva et al. 2000).

The assignment of a particular error to its place in the taxonomy may be ambiguous, as with every attempt at reconstructing the past. For instance, in the data table of Salas et al. (1998), the otherwise extremely rare transition at 16094 appears on lineages of three different haplogroups: this could constitute a phantom mutation (type IIIb), or represent base shifts (types Ia or Ic) affecting two haplotypes (in haplogroups W and pre-HV), whereas the instance within haplogroup H (exactly matching a type in Helgason et al. 2000) could be real.

Case studies

We will illustrate the natural occurrence of the above artefacts with two data sets of HVS-I and HVS-II sequences from the forensic community (Lee et al. 1997; Seo et al. 1998), which have their own specific error profiles. The data set of Lee et al. (1997), comprising 306 Korean mtDNA sequences, is conspicuous for its extreme excess of transversions between 16330 and 16401: the 11 positions with transitions are exceeded by 13 positions in this stretch that show transversions, all except 2 of which are unique in the worldwide database. One mtDNA, H169, even bears three such transversions. This suggests that almost all of these transversions are artefacts, perhaps constituting phantom mutations (type IIIa errors). Base shifts (type Ia and Ib errors) also occur: mtDNA S21 has the T at 16223 (where it would be expected given the combined HVS-I/HVS-II motif 16294-16295-146-199) shifted to 16213; further, mtDNA H99 has the T at 16245, where it would be expected given the unmistakable HVS-II motif 191+A-194-199 (see Ozawa et al. 1991 and Oota et al. 1995) shifted to 16244. There are numerous dubious back-mutations (mainly reference bias: type IIa errors) that one would certainly wish to check, although some of them may be real. For example, the typical haplogroup F sequences with HVS-I motif 16232A-16304 would be expected to have one of the As at 248/9 deleted (compare S36, S41, H362, S73), but mtDNA P30 does not (base misscoring: type IVd error).

A prominent feature of the HVS-I/HVS-II data of 100 Japanese obtained by Seo et al. (1998) is the appearance of the transition pair 16242-16248 and the transversion 16239G (phantom mutation: type IIIb error). The former event, which has never been reported previously, occurs in haplogroups B (#38, #72) and D (#85), in another specific subhaplogroup of M (#18, #22, #31, #46), as well as in yet another (probable M) lineage (#63). The transversion, which has appeared previously only as a probable unique mutation in haplogroup H (Richards et al. 2000), occurs in a particular branch of haplogroup F and in haplogroup M, accompanied by transitions 16242 or 16321. Another striking feature is the repeated crossing-over between HVS-I and HVS-II sequences (artefactual recombination: type Vb error). This is exemplified by the sample pair #27, #59 where an mtDNA with motif 16223-16257A-16261-150 (Lee et al. 1997; Pfeiffer et al. 1998) has "recombined" with one with motif 16245-16362-191+A-194-199. The mtDNAs with HVS-II motif 93-210 normally have the HVS-I motif 16140-16189-16266R and thus belong to haplogroup B (Melton et al. 1998), but this correlation is destroyed with the HVS-I motif in #90 and the HVS-II motif in #63. This seems to be only the tip of the iceberg of a general random association of HVS-I and HVS-II sequences in part of this data set.

Quality control

Here is a self-help guide to avoid some of the worst blunders at any stage of the data generation and proof-reading processes.

– Search the database for neighbours (up to a certain distance), attempt to localise the sequence to a part of the

phylogeny, at least to the level of the haplogroup. On this basis, if the haplogroup motif is not fully represented in the sequence, recheck the relevant positions in the sequence.
– Perform a phylogenetic analysis at the level of subhaplogroups (i.e. on rather closely related sequences) and check whether the private mutations are real.
– Always have in mind the relative mutability of sites as estimated for example by Hasegawa et al. (1993); in particular be aware of hypervariable sites (e.g. Bandelt et al. 2000; Stoneking 2000). Be sensitive to rare mutations occurring on different sequence backgrounds in one batch of sequencing: there may be sequencing artefacts of this type. Be extremely suspicious of numerous distinct transversions and indels that do not often appear in the database. If misreading can be excluded, resequencing will be necessary.
– Look out for incongruences between parts of the sequence which have been obtained in different PCR or sequencing reactions: this process is a rich source of "recombinants".

These rules have been applied so far to a number of data sets prior to as well as post-publication. In our experience of our own and others' data, a list of queries of suspicious bases will typically pinpoint actual errors in about 50% of the queries. Many oversights happen in zones which are difficult to read. In the case of the Rando et al. (1998, Table 1) data, a subsequent round of queries succeeded in pinpointing the following errors in the published sequences:

– 16094-16177-16181-16223-16278-16311-16390: erase 16094
– 16111-16172-16189-16219-16278-16293: add 16362
– 16093G-16287A-16301-16311: add 16362
– 16069-16126-16225: erase 16225
– 16169-16172: add 16189.

A partial re-reading of the (Inari, Skolt, Karasjok) Saami HVS-I sequences from Sajantila et al. (1995) led to the following modifications in the published sequences (and to the detection of the error source):

– 16144-16189-16298: omit sequence (contamination)
– 16189-16270: add 16183C (misread autoradiograph)
– 16129-16185-16223-16224-16260: add 16298 (misread autoradiograph)
– 16126-16136-16189-16223-16360-16362: add 16182C, 16183C (miscompilation)
– 16144-16298: erase 16144 (flaw in manuscript preparation)
– 16189-16270: add 16144 (flaw in manuscript preparation)
– 16144-16148-16189-16335: add 16270 (flaw in manuscript preparation).

Moreover, one Saami sequence which differs from the CRS only by a transition at 16129 was simply overlooked when the manuscript was prepared for publication (in a laboratory different from that where the sequences were analysed).

The latter study indicates that the final stage of manuscript preparation and data submission may be more deleterious to data quality than the early reading stage. In extreme cases the printing process may introduce the bulk of errors: compare the published data tables from Lutz et al. (1998/1999) with the final table available from Sabine Lutz. The discrepancies between Fig. 2 of Redd and Stoneking (1999) and the (seemingly reliable) Genbank records can also be seen in this light. Their table displays numerous base shifts by 1–19 positions.

Conclusions

The impact of the above listed blunders on forensic work is potentially significant. Firstly, if perpetrated on samples from crime scenes or from suspects, they are likely to result in false exclusions. Secondly, in the case of surveys designed to enlarge the reference database of the general population, they will tend to generate new types that do not in fact exist and in the process, reduce the estimate of the frequency of real types, albeit marginally, except in the worst cases. This reduces the probability that a crime scene sample could have come from the general population, hence making a match between a crime scene and a suspect more impressive than it really is. There is clearly a need for reliable databases. The motto "good data quality is ensured by using original sequences only" (Wittig et al. 2000) is, as we have seen above, not sufficient.

The impact on evolutionary studies might be thought less significant, but nevertheless problems can arise. In the first place, the introduction of errors into sequence data has led to serious and unnecessary concerns about in vivo recombination in non-recombining systems such as mtDNA: Hagelberg et al. (1999) vs. Hagelberg et al. (2000) and Eyre-Walker et al. (1999) vs. Macaulay et al. (1999b). Data recombination would clearly affect the accuracy of genetic dating. Rogue transversions, insertions and deletions are relatively less harmful for dating, if transitions alone are used for this purpose (as a response to the higher error rate for transversions and indels); furthermore the effect on the reconstructed phylogeny would normally be rather minor. Nevertheless, important effects can occur in extreme instances. For example, the very high time-depths for haplogroup T lineages reported in the Ladins (Stenico et al. 1996), interpreted as a severe and ancient bottleneck or founder effect in the Alpine region of Europe, must be regarded as doubtful given the problems associated with these data, as must the demonstration of mutational instability in haplogroup T mtDNAs adduced by Malyarchuk and Derenko (1999) based on the same data (although the effect on position 16296 is real; Richards et al. 2000). More often, base shifts, phantom transitions and the like will generate minor inflations of diversity and therefore time estimates and will have a potentially significant effect on phylogeographic analyses that rely upon identifying exact matching types, such as founder analysis (Richards et al. 2000).

In order to meet high-quality standards in forensics, sequencing should be performed in both directions (Bär et al. 2000). It is then important to read the two series of outputs separately (against the CRS) and to transform either series into a data table independently, preferably of different formats (motif vs dot table); finally, the two tables should be compared by computer.

Given the ease with which mistakes can be made and the frequency with which they seem to occur in the published literature, we suggest that it would be helpful for workers to be relaxed and open to the possibility of error and to maintain database errata for their work (see for example http://www.stats.ox.ac.uk/~macaulay/founder2000).

Acknowledgements We are indebted to Peter Robinson for his stimulating lecture ("Do manuscripts have DNA?" Bielefeld, June 8, 2000), to José M. Larruga and Vicente M. Cabrera for their careful rereading of previously published sequences and to Peter Forster and Toomas Kivisild for useful hints. This work has been supported by a travel grant from the Deutscher Akademischer Austauschdienst (DAAD) to H-J B and by a Wellcome Trust Research Career Development Fellowship to VM.

References

Alves-Silva J, Santos MDS, Guimarães PEM, Ferreira ACS, Bandelt H-J, Pena SDJ, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461, 775 (erratum)

Anderson S, Bankier AT, Barrell BG, Bruijn MHL de, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Bandelt H-J, Macaulay V, Richards M (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. Mol Phylogenet Evol 16:8–28

Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, Lincoln PJ, Mayr W, Morling N, Olaisen B, Schneider PM, Tully G, Wilson M (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Int J Legal Med 113:193–196

Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. Ann Hum Genet 60:35–49

Dimo-Simonin N, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. Int J Legal Med 113:89–97

Easton RD, Merriwether DA, Crews DE, Ferrell RE (1996) mtDNA variation in the Yanomami: evidence for additional New World founding lineages. Am J Hum Genet 59:213–225

Eyre-Walker A, Smith NH, Maynard Smith J (1999) How clonal are human mitochondria? Proc R Soc Lond B 266:477–483

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc Lond B 266:485–492

Hagelberg E, Goldman N, Lió P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. Proc R Soc Lond B 267:1595–1596

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347–354

Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) MtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66: 999–1016

Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origin of modern humans. Cell 90:19–30

Lee SD, Shin CH, Kim KB, Lee YS, Lee JB (1997) Sequence variation of mitochondrial DNA control region in Koreans. Forensic Sci Int 87:99–116

Lutz S, Weisser H-J, Heizmann J, Pollak S (1998/1999) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int J Legal Med 111: 67–77, 112:145–150 (erratum)

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, Torroni A (1999a) The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249

Macaulay V, Richards M, Sykes B (1999b) Mitochondrial DNA recombination – no need to panic. Proc R Soc Lond B 266: 2037–2039

Malyarchuk BA, Derenko MV (1999) Molecular instability of the mitochondrial haplogroup T sequences at nucleotide positions 16292 and 16296. Ann Hum Genet 63:489–497

Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823

Merriwether DA, Kemp BM, Crews DE, Neel JV (2000) Gene flow and genetic variation in the Yanomama as revealed by mitochondrial DNA. In: Renfrew C (ed) America past, America present: genes and languages in the Americas and beyond. McDonald Institute for Archaeological Research, Cambridge, pp 89–124

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: implications for the peopling of the southern cone of the continent. Am J Phys Anthropol 113:19–29

Nishimaki Y, Sato K, Fang L, Ma M, Hasekura H, Boettcher B (1999) Sequence polymorphism in the mtDNA HV1 region in Japanese and Chinese. Legal Med 1:238–249

Oota H, Saitou N, Matsushita T, Ueda S (1995) A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. Am J Phys Anthropol 98:133–145

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 Comm 176:938–946

Pfeiffer H, Steighner R, Fisher R, Mörnstad H, Yoon C-L, Holland MM (1998) Mitochondrial DNA extraction and typing from isolated dentin-experimental evaluation in a Korean population. Int J Legal Med 111:309–313

Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white caucasian genetic identification. Int J Legal Med 106:85–90

Quintana-Murci L, Semino O, Bandelt H-J, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence for an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Rando JC, Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM, Bandelt H-J (1998) Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Ann Hum Genet 62:531–550

Redd AJ, Stoneking M (1999) Peopling of Sahul: mtDNA variation in aboriginal Australian and Papua New Guinean populations. Am J Hum Genet 65:808–828

Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203

Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62:241–260

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H-J (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med (in press)

Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. Int J Legal Med 111:292–298

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

Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Geddedahl T, Isseltarver L, Di Rienzo A, Pääbo S (1995) Genes and languages in Europe – an analysis of mitochondrial lineages. Genome Res 5:42–52

Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375

Seo Y, Stradmann-Bellinghausen B, Rittner C, Takahama K, Schneider PM (1998) Sequence polymorphism of mitochondrial DNA control region in Japanese. Forensic Sci Int 97: 155–164

Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C, Barbujani G (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. Am J Hum Genet 59:1363–1375

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67:1029–1032

Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians – an interpretation from mitochondrial lineage analysis. Am J Hum Genet 57:1463–1475

Torroni A, Richards M, Macaulay V, Forster P, Villems R, Nørby S, Savontaus M-L, Huoponen K, Scozzari R, Bandelt H-J (2000) mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 66: 1173–1177

Watson E, Forster P, Richards M, Bandelt H-J (1997) Mitochondrial footprints of human expansions in Africa. Am J Hum Genet 61:691–704

Wittig H, Augustin C, Baasner A, Bulnheim U, Dimo-Simonin N, Edelmann J, Hering S, Jung S, Lutz S, Michael M, Parson W, Poetsch M, Schneider PM, Weichhold G, Krause D (2000) Mitochondrial DNA in the central European population. Human identification with the help of the forensic mt-DNA D-loop-base database. Forensic Sci Int 113:113–118