Int J Legal Med (2001) 115:29–39

ORIGINAL ARTICLE

Arne Röhl · Bernd Brinkmann · Lucy Forster
Peter Forster

An annotated mtDNA database

Received: 14 August 2000 / Accepted: 22 December 2000

Abstract We have compiled a database of mitochondrial DNA (mtDNA) control region, hypervariable regions 1 (HVR1) and 2 (HVR2) sequences of a total of 14,138 individuals compiled from 103 mtDNA publications before 1 January 2000, 13 data sets published in 2000 and 2001 and 2 unpublished data sets of Iraqi Kurds and Indians from Kerala. By contacting the authors and by other means, we have confirmed and corrected sequence errors, eliminated duplications and harmonised the sequence format. These changes affected all but 26 of the 116 publications. Furthermore, we have implemented a geographic information system ("mtradius") which searches for closest matches to a given mtDNA control region sequence and displays them on a geographic map. A potential application is to estimate a chance matching probability when a forensic stain and a suspect have an identical mtDNA sequence: we suggest that the geographic area with the highest frequency of closely related mtDNA sequence types may be used to define a reference population to give the suspect the maximum benefit of doubt in accordance with the ceiling principle.

Keywords Identification · DNA fingerprint · Ethnic · Population

Introduction

Mitochondrial DNA (mtDNA) typing is not the method of choice for DNA fingerprinting of individuals, as mtDNA represents a non-recombining genetic locus with alleles which are usually shared by many individuals with the same maternal ancestor, unless mutations have occurred. However, when a forensic or anthropological sample is highly degraded, mtDNA is often the only option due to its abundant copy number compared to nuclear DNA. Given an exact mtDNA sequence match between a (crime) stain and a suspect, it is then of interest to determine the frequency of the sequence in the population in order to calculate the probability of a chance match. The question we wish to address is, which population is the relevant one for determining the frequency? It is desirable to err on the safe side, i.e. in favour of the suspect, and choose that region of the world where the mtDNA type is likely to be found at the highest frequency ["ceiling principle" (National Research Council Committee 1992)].

A simple database search to identify geographic regions with high matching frequencies is not reliable, given that only a few thousand mtDNA sequences worldwide are available and thus many mtDNA types will be unique in the database although a very large number of non-sequenced humans may have precise matches in particular regions of the world. (See for example Pfeiffer et al. 2001 where a sample of 1200 German villagers yielded many unique mtDNA types even though the sequence was short and maternal relatives were included.). Instead, we advocate searching for the geographic region with the highest frequency of the most similar mtDNA sequences to a given sequence, regardless of whether the resulting best matches are exact matches or only close matches. The rationale is that the only geographic sorting mechanisms known for mtDNA types are mutation, migration and founding events during human history and prehistory [for example, geographical sorting of mtDNA types by selection has been rejected in the case of selection pressure at high altitude (see Torroni et al. 1994 and Comas et al. 1998), and preferential survival of centenarians with certain mtDNA types, claimed by De Benedictis et al. (1999), is unlikely to create a suspect pool of forensic relevance]. If, then, a new mtDNA type has evolved locally by mutation at some time in the past, we may not observe that particular type in a limited sample of say, 10,000 sequences, but a search of closest matches will always yield related mtDNA types, whose geographic location may approximate the location of the queried sequence (P. Forster et al. in preparation). We

A. Röhl · B. Brinkmann · L. Forster
Institut für Rechtsmedizin, University of Münster, Germany

P. Forster (✉)
The McDonald Institute for Archaeological Research,
University of Cambridge, UK
e-mail: pf223@cam.ac.uk

have therefore established an mtDNA database with a geographic information system ("mtradius") which searches the database with progressively increasing mutational distance until close matches are identified. The best matches (as well as second- and third-best matches) are displayed on a world map, which yields frequency values for every desired grid square or rectangle. An appropriate grid rectangle or rectangles can then be chosen as the reference population for maximal frequency estimates in court cases.

Methods

Database

We have compiled from the literature a database of all mtDNA control region sequences in the range np 15996–16569 and np 1–576, published until 1 January 2000, as well as a few published in 2000. There are three main reasons why raw published data cannot be used for frequency calculations: (1) sequence errors and duplicate publication of sequences from identical samples, (2) varying notation of length polymorphism around, e.g. np 16189, 310, and 249 and (3) intentional preselection of sequences according to RFLP typing or 9 bp-deletion typing at np 8272–8289.

1. We have tried to identify and eliminate as many errors and duplications as possible, as detailed below. In the case of ancient DNA, only ancient DNA sequences which were reproduced by independent laboratories were included. Nevertheless, we classified even independently confirmed ancient DNA sequences as non-representative, regardless of the quality of the particular ancient DNA study.

2. We have harmonised the database for length polymorphism notation according to forensic practice, namely notation of A to C transversions at np 16182/16183 rather than deletions; assignment of insertions in homopolymeric stretches at the highest possible nucleotide number (e.g. 309.1C instead of 303.1C) and rejection of C to T transversions at np 311, which are generally a reading artefact due to length heteroplasmy in the preceding C-tract. In spite of these corrections, we suspect that many authors have disregarded variation completely at these two C-stretches. We have therefore implemented a weighting option for each nucleotide position, which is set to zero by default for the np 16182, 16183, 309 and 315. This weighting option can incidentally also be used for enhancing the search for phylogenetically related sequences by weighting ancestral nucleotide states.

3. With respect to the intentional preselection of sequences, we have classified such sequences as non-representative. Since we distinguish between representative and dubious/non-representative sequences in the database, rather than simply omitting the latter, the user can decide whether to launch a qualitative search using all available sequences, or whether to obtain quantitative localisation of sequences using representative data. Geographical (maternal) origins of the samples were entered as precisely as possible, usually down to one arcminute. If the sample origin only specified a country/state rather than a location, then we assigned the sample to the capital of that country/state (e.g. French Canadians were assigned to Paris). Thus, the user of the geographical map display is warned to ascribe lesser weight to samples located on capitals. If the sample origin in a publication only specifies continent or (phenotypic) race, we assigned the sample to a standard location in the sea close to the continent to indicate the uncertainty for map users. The standard locations are 65°N 0°E (if the paper specifies "Europeans", "white Caucasians" etc.), 30°S 40°E ("Africans", "Black Americans", "Negroids" etc.), 20°N 130°E ("Asians", "Mongoloids" etc.), and 10°N 100°W ("Hispanics" etc.). In papers describing mixed samples which contain both indigenous and non-indigenous mtDNA (e.g. the Brazilians of Bortolini et al. 1997), we assigned the indigenous mtDNA types to the sample location, and placed the non-indigenous types at the standard continent locations. All samples placed at standard continent locations can be deactivated in mtradius to avoid lowering the precision of geographic assignment if required. It has been observed that inconsistencies exist between the number of sequences that have been submitted to GenBank and the number of sequences that appear in the corresponding publication, presumably because mtDNA types rather than individuals were submitted (Miller et al. 1996). For this reason we have usually based our entries on the literature or from files obtained from the authors rather than on databases. We made a few exceptions (detailed below) when the publication contained obvious errors.

The sequences are taken from the following publications and modified as annotated:

ALV (Alves-Silva et al. 1999) We classified these sequences as non-representative as they were preselected for the presence of the 9 bp deletion at np 8272–np 8289. The few European and African mtDNA types were assigned to the European and African standard locations, respectively. RFLP information is available.

ALV1 (Alves-Silva et al. 2000) We renamed deleted nucleotide positions as 249, 290 and 291. We corrected the systematic shift in sample names in the final 11 entries of HV2 sequences in their Table 6. Their text (p.453) states that only one southeast Brazilian lineage was omitted for HVR2 sequencing, while their Table 6 omits two individuals; we therefore interpreted "lineage" as "mtDNA HVR1-type". The European and African mtDNA types were assigned to the European and African standard locations, respectively. RFLP and 9 bp information is available.

AND (Anderson et al. 1981) The placenta sample providing the first complete mtDNA sequence (Cambridge reference sequence CRS) was resequenced, identifying 11 errors, all of which were outside the control region (Andrews et al. 1999). The placenta was obtained from the Cambridge maternity hospital and is presumably of local origin (F. Sanger, pers. comm.).

ARN (Arnason et al. 1996) The full mtDNA sequence for the Swedish individual is available.

BAA (Baasner et al. 1998) We entered the data as published.

BAM (Bamshad et al. 1996) We entered only the Indians, because evidently the Asian, African and European sequences had been previously published in Jorde et al. (1995). The frequencies of the mtDNA types are not given and we could deduce only some frequencies by comparing their Fig. 8 and Tables 1 and 3. Accordingly, we classed the sequences as non-representative pending information from the authors.

BAR (Barbujani et al. 1996) The original data file, which differs from the published table, was kindly provided by the authors. We harmonised their scoring with ours by deleting postulated transitions at np 311, and by reassigning insertions to np 309 and np 315.

BAT (Batista et al. 1995) We reassigned insertions to np 191, 309 and 315. RFLP and 9 bp information is available.

BEL (Belledi et al. 2000) We formatted the data from an electronic table with detailed geographic information kindly provided by the late Michele Belledi. The sequenced range is np 16024–16323 (M. Belledi and R. Casalotti, pers. comm.).

BER (Bertranpetit et al. 1995) We entered the data as published.

BET (Betty et al. 1996) We classified these sequences as non-representative as they were preselected for the presence of the 9 bp deletion at np 8272–np 8289.

BOR (Bortolini et al. 1997) We redefined single and double insertions in their published duplicate column np 16184 as transversions at np 16182 and 16183. Information for 9 bp is available.

BRI (C. Brinkmann et al., unpublished) unpublished Kurdish sequences.

BRO (Brown et al. 1998) We defined this dataset as non-representative as the samples had been preselected by RFLP analysis. We reassigned the deletion to np 16166.

CAL (Calí et al. 2001) RFLP information for the Sicilian sequences is available.

CHE (Chen et al. 2000) We corrected nucleotide variants accidentally scored by the authors as identical to CRS as transitions relative to CRS. We renamed 309C+ and 315C+ as 309.1C and 315.1C. It should be noted that the phylogenetic nomenclature used by the authors conflicts with the similar nomenclature by Watson et al. (1997). We relocated the geographic origin of the Kung and Khwe from Schmidtsdrift (South Africa) to Omega base camp (Caprivi at the Angolan-Namibian border) where they had been recruited as trackers. RFLP information is available.

CLF (Calafell et al. 1996) Bulgarians 30 and 31 appear to have exchanged their HVR2 sequences in the published table (Bandelt et al. 2000); we have accordingly classified these two HVR2 sequences as dubious.

COM (Comas et al. 1996) We entered the data as published.

COM1 (Comas et al. 1998) The sequence table contains one less Kirghiz (Sary-Tash) sequence than stated in the text.

COM2 (Comas et al. 2000) We placed the Georgians and Kurds, both sampled around Tbilisi, 1° apart to distinguish them in the map functions.

COR (Côrte-Real et al. 1996) Some of the Spanish, Portuguese and Basque sequences had been taken from Richards et al. (1996). We therefore omitted these when entering the sequences of Richards et al. (see below).

DEL (Delghandi et al. 1998) The sequence table contains one Saami sequence less than stated in the text.

DIM (Dimo-Simenon et al. 2000) We formatted the sequences from an electronic file kindly provided by the authors. We reassigned T insertions to np 310.

DUP (Dupuy and Olaison 1996) Their Tables 1 (Norwegians) and 2 (Saami) contain several inconsistencies. The first column in Table 1 (np 16113) contradicts Table 2 (np 16013). The Anderson sequence is given with six errors, three columns are empty and thus at first glance redundant. In the first table, the column for np 16390 is empty – we reassigned the A in the neighbouring np 16362 column to np 16390, assuming the affected sequence belongs to mtDNA group I (cf. Torroni et al. 1996). One Norwegian-Saami sequence match was overlooked, as were two Norwegian-Norwegian matches. The 30 Norwegian samples are a superset of the 28 Norwegians sequenced by Miller (1996), hence we were able to identify most sample locations within Norway using the information in Miller (1996). The comparison of Miller's Norwegian sequences with Dupuy's sequences reveals that Miller committed 9 sequence errors affecting 5 out of 28 sequences, while Dupuy and Olaison committed only 3 errors in 30 sequences. We therefore classified the corrected Norwegians of Dupuy and Olaison as reliable and representative, and the duplicate Norwegian sequences of Miller as non-representative. See also our comment on Opdal et al. (1998). We classified the Saami sequences, for which similar cross-checking was not available, as dubious pending information from the authors.

EAS (Easton et al. 1996) The published table is flawed, containing a high proportion of transversions (Merriwether et al. 2000). We have classified the sequences as dubious, pending publication of the erratum (A. Merriwether, pers. comm.). We reassigned a deletion to np 16007 and an insertion to np 16367. RFLP and 9 bp information is available.

FIN (Finnilä et al. 2000) We classified these sequences as non-representative due to preselection for mtDNA group status. We recoded positions with length polymorphism (np 248, 303, 311, 514, and 568) in their Table 2 using our nomenclature (np 249, 309, 315, 522 and 523, and 573). We defined the sequence end at np 576 as stated on their page 1019 (but cf. their Table 2). We corrected np 16115.C as np 16115.A in their Fig. 2b. Full mtDNA sequence information is available.

FOR (Forster 1996) Detailed geographical and genealogical information for the German and Danish sequences published in Richards et al. (1996) are given in Forster (1996). We also added unpublished partial sequences to the database.

FOR2 (L. Forster et al. unpublished) unpublished sequences from Kerala.

FRA (Francalacci et al. 1996) We omitted the 3 maternally related sequences, leaving a total of 49 individuals. Only 48 were displayed by Torroni et al. (1996), where RFLP information is available.

GIL (Gill et al. 1994) We entered maternal relatives into the database, but marked them as non-representative. We traced the maternal genealogies of the royal subjects to their thirteenth century geographical origins.

GIN (Ginther et al. 1993) Their Table 1 contains no entries in the column for np 195. We reassigned insertions at np 316 to np 315 and 9 bp information is available.

GRA (Graven et al. 1995) RFLP information is available.

GRE (Green et al. 2000) Assignments of sequences to the two Mexican cities was kindly provided by the authors. RFLP and 9 bp information is available, whereas the np 16390 information is unpublished (pers. comm.). The European/unknown sequences and the African sequences were assigned to the standard continent locations. An excess of B types was sequenced, part of which we therefore labelled as non-representative.

HAN (Handt et al. 1994) We classified the sequence of the Oetztal Ice Man as non-representative. We took the 16 Austrians from HvrBASE (Burckhardt et al. 1999).

HEL (Helgason et al. 2000) The published GenBank accession numbers are not current and should read AF236888-AF237281 for HVR1 and AF305969-AF306314 for HVR2. The submitted sequence lengths are generally shorter than published, and there are 402 not 401 individuals in total (A. Helgason, pers. comm.).

HOF (Hofmann et al. 1997) We reassigned insertions to np 309 and np 315. We corrected np 16290 relative to the CRS. We corrected individual 700 and annotated 722 as dubious following Fig. 2 of Macaulay et al. (1999). The author kindly confirmed that the subjects were native Germans drawn from the area around Munich (K.-D. Gerbitz, pers. comm.). Further mtDNA sequence information is available.

HOL (van Holst Pellekan et al. 1998) Australian aboriginal T27 is of European descent (S. van Holst Pellekan, pers. comm.). A corrigendum for the published analysis is available from the authors.

HOR (Horai and Hayasaka 1990) The published table contains several errors around np 16325, while the sequences submitted to DDBJ are correct. Note furthermore that the uncorrected sequence CJK5 was taken (see comment below on Greenberg et al. 1983). The full mtDNA sequence for Ugandan SB17 was published by Horai et al. (1995). We reassigned insertions to np 16188.1C and 16262.1C.

HOR1 (Horai et al. 1993) We reconstructed the provenance of the samples using their Figs. 2 and 3.

HOR2 (Horai et al. 1996) The sequences deposited in DDBJ are longer than stated on page 581, 9 bp information is available.

IVA (Ivanova 1993) The sequence table can be reconstructed using the accompanying graphs.

JOR (Jorde et al. 1995) The San and Chinese samples each contain one sequence less than stated in the text.

JUN (Jun et al. 1994) The full mtDNA sequence is available.

KIT (Kittles et al. 1999) The Swedish sequences contain numerous Saami sequences; Saami ancestry cannot be excluded (R. Kittles, pers. comm.). We therefore assigned the Swedes to the European standard location. The haplotype frequencies per population are misplaced in the published table, we therefore entered the correct values kindly provided by the author.

KIV (Kivisild et al. 1999) We amended Lamb 37 as having the transversion np 16240.C (T. Kivisild, pers. comm.). We entered RFLP information for np 73 as sequence information. Further

RFLP information is available. The publication supplement includes the data analysed but not published by Bamshad et al. (1998) (T. Kivisild, pers. comm.).

KOB (Kobayashi et al. 1991) According to the sequencing protocol, the sequence results were intentionally biased towards the CRS, potentially yielding a hybrid sequence. We classified this sequence as dubious. The full mtDNA sequence is available.

KOL (Kolman et al. 1995) We renamed np 16182.X and np 16183.X as np 16182.C and np 16183.C. We renamed np 106–111.X as a 6 bp deletion. RFLP and 9 bp information are available.

KOL1 (Kolman et al. 1996) Length polymorphisms at np 16180–np 16188 (scored by us as transversions at np 16182 and np 16183) were not recorded by the authors; we have therefore scored np 16182.N and np 16183.N in each sequence harbouring a np 16189 C-stretch. We reassigned insertions to np 16193 and 16227. RFLP and 9 bp information is available.

KRI (Krings et al. 1999) RFLP information is available.

LAH (Lahermo et al. 1996) Some corrections of the sequences generated in Turku were kindly provided by the authors via H.-J. Bandelt.

LAH1 (Lahermo et al. 2000) An electronic file of the data was kindly provided by the authors, from which we deduced the Csango and Budapest samples using the published mtDNA group frequencies. In the published table 77 individuals are given but 78 are mentioned in the text and given in the file. We scored np 309.1T as np 310.1T.

LEE (Lee et al. 1997) We recoded the insertions around np 191 for consistency; np 203 and np 16173 are incorrectly scored relative to CRS, we entered affected sequences as having N at these positions. We reassigned deletions to np 16193 and 16262.

LEV (Levin et al. 1999) We classified these sequences as non-representative, as they were derived from cell lines. The full mtDNA sequence is available.

LOR (Lorenz and Smith 1997) Several previously published mtDNA sequences were incorrectly reproduced by these authors (insertion of an artefactual np 16224 column). We therefore also classified the new sequences in their table as potentially dubious, pending information from the authors; np 16256 is incorrectly scored relative to CRS, so we entered type 85 as having 16256.N. Furthermore, the sequences are non-representative as mtDNA group D was omitted.

LUM (Lum et al. 1994) 9 bp information is available.

LUM1 (Lum et al. 1998) We shortened sequence KSR24 to eliminate a dubious flanking triple deletion. We realigned KIR13 by four nucleotides.

LUM2 (Lum and Cann 2000) We formatted the sequences from GenBank, with the accession numbers starting at AF285638 rather than AF85638.

LUT (Lutz et al. 1998) The published table and the two subsequent errata are flawed; we entered the data from an electronic table kindly provided by the author. The samples are specified as German Caucasians. We reassigned an insertion to np 16538.

MAC (Macaulay et al. 1999) RFLP information is available.

MAL (Malyarchuk et al. 1995) We entered the data as published.

MAT (Mateu et al. 1997) We entered the data as published.

MEL (Melton et al. 1998) 9 bp information is available. We recoded 248 as 249.

MIL (Miller 1996; Miller et al. 1996) The sequences were kindly provided in electronic format by J. Dawson, and compared with the printed thesis of Miller (1996), and with the Norwegian sequences of Dupuy and Olaison (1996), who used identical Norwegian samples. The discrepancies with the Dupuy sequences are listed above. See also our comment on Opdal et al. (1998). Furthermore, NIR0292 differed between the file and the thesis. We accordingly classified all sequences as dubious. We interpreted the electronic table notation via the thesis, because the file does not distinguish between HVR1 Anderson sequences and missing HVR1 sequences, and because the extent of N-stretches is not self-explanatory. We redefined C insertions at np 16184 as transversions at np 16182 and 16183.

MON (Monsalve and Hagelberg 1997). We omitted sequence information at np 16126, as this was found to be ambiguous (V. Monsalve, pers. comm.). Only one Black Carib has an Amerind type; the others were assigned to the African standard location. Nine bp information is available.

MOU (Mountain et al. 1995) The identity of the European and Indian control samples was kindly provided by the authors.

MUR (Murray-Macintosh et al. 1998) The published sequence table has suffered distortion (16365T instead of 16362T). Explanations were kindly provided by the authors. The new data comprise sequences for 31 Maori individuals (rather than 31 haplotypes as stated in the Materials and Methods section); 9 bp information is available.

NAT (Nata et al. 1999) 102 Chinese are given in the published table as opposed to 101 Chinese in the text.

NIS (Nishimaki et al. 1999) We reassigned insertions to np 16188.1C and 16188.2C, and deletions to np 16370. Furthermore, np 16095, 16298 and 16352 are incorrectly scored relative to CRS; we scored these positions in the affected sequences as N.

NSN (Nishino et al. 1996) The full mtDNA sequence is available. We classified the sequence as non-representative as the individual was preselected for a potentially mitochondrially transmitted disease.

OHL (Ohlenbusch et al. 1998) We interpreted np 310.1C as 309.1C and reassigned the CA insertion to np 523. The full mtDNA sequences are available.

OOT (Oota et al. 1995) We reassigned insertions and deletions to np 16262 and 249, respectively.

OPD (Opdal et al. 1998) Only 215 individuals are mentioned in the text, whereas 216 are presented in their Table 1. The sequence range in the abstract contradicts the sequence range in the results; we entered the shorter range, thus omitting a variant nucleotide in S56. We corrected C78 as a J sequence and classified C32, C52, C59, C62 and C78 as dubious after scoring erroneous nucleotides as "N". Opdal et al. evidently included the previously published Norwegians (Miller 1996; Dupuy and Olaison 1996) in their Table 1 (control samples); we omitted these duplicates here.

ORE (Orekhov et al. 1999) The assignment of each individual to one of the three Russian cities was kindly provided by the authors in electronic format. Their electronic file omits the published insertions np 16183.1C and 16189.1C (which we recoded as np 16188.1C and 16193.1C) in individual 76 (number 7 in the publication) for formatting reasons (pers. comm.).

OZA (Ozawa et al. 1991a) We reassigned insertions to np 16188, 16193, 309, 315, and 573. The sequence MELAS P-1 was corrected in the following paper as noted by the authors, but the sample name was inadvertently confused with MELAS P-2. The full mtDNA sequences are available. We classified the sequences as non-representative as the individuals were preselected for a potentially mitochondrially-transmitted disease.

OZA1 (Ozawa et al. 1991b) We reassigned insertions to np 311 and 315, and the CA deletion to np 522/523. We omitted four sequences (PD P-2, Melas P-1, FICM, Melas P-2) which had been previously published. The full mtDNA sequences are available. We classified the sequences as non-representative as the individuals were preselected for a potentially mitochondrially-transmitted disease.

PAR (Parson et al. 1998) We recoded the "T insertions" at np 309 as C insertions.

PFE (Pfeiffer et al. 1998) We entered the data as published.

PFE1 (Pfeiffer et al. 1999) This publication contains 109 sequences from around Münster, and reports on 700 other sequences. As the

700 sequences are a subset of those published subsequently Pfeiffer et al. (2001), we assigned the 700 sequences to the latter publication. We formatted the 109 sequences from the pre-publication file which differs in its numbering from the published table. The following errors were detected post-publication (V. Macaulay) and corrected in our database: published sample numbers 7 (295.T), 28 (16296.T instead of 16295.T), and 36 (16356.N).

PFE2 (Pfeiffer et al. 2001) These 1,200 sequences were sampled from a north German village near Braunschweig and include 10% or more non-Germans (e.g. Poles and Turks) as ascertained by surname or birthplace. We assigned the foreign-born individuals to their country of origin and classified them as representative, while we assigned the German names to Braunschweig but classified them as potentially non-representative (surname analysis not being diagnostic for second-generation maternal descent). We assigned foreign-sounding but non-localisable surnames to the European standard location.

PIE (Piercy et al. 1993) The samples are from Caucasoid persons arrested in England but potentially originating from any part of the United Kingdom (Miller 1996). We reassigned insertions to np 309 and np 315.

PIN (Pinto et al. 1996) The text contains a typing error on page 325 concerning the sequenced range. We reassigned the Tenerife samples to Rando et al. (1998) and the 18 Berber sequences to Rando et al. (1998), where they were accurately reread.

POL (Polyak et al. 1998) An electronic file of the sequences was kindly provided via T. Kivisild and A. Eyre-Walker. The omnipresent transversion at np 16220 was confirmed as a typographical error (K. Polyak, pers. comm.). The full mtDNA sequences are available.

PUL (Pult et al. 1994) A table with detailed geographic or linguistic assignment of each sequence was kindly provided by the author. Of the 77 sequences, 4 were determined from known brothers or sisters of other samples (I. Pult, pers. comm.), hence we classed these 4 as non-representative.

QUI (Quintana-Murci et al. 1999) We scored these data as non-representative as they were preselected for RFLP status.

RAN (Rando et al. 1998) See comment on Pinto et al. (1996). We corrected five errors discovered post-publication (H.-J. Bandelt, pers. comm.) and reassigned a deletion to np 16325. RFLP information is available.

RAN1 (Rando et al. 1998) See our comment on Pinto et al. (1996). We reassigned a deletion to np 16185.

RED (Redd et al. 1995) These samples were preselected for 9 bp deletion status at np 8272–8289, while the subsequent publication (Redd and Stoneking1999, see our comments below) preselected their sequences for the absence of the deletion. We classified a few "excess" 9bp-deletion sequences as non-representative according to the true proportions of 9 bp deletions measured in these population samples. We omitted sequences culled from Vigilant (1990), but included those stated to be from Stoneking et al. (1992), as the latter publication does not display sequence data. Redd et al. are co-authors of the earlier Vigilant publications and resequenced and rescored the frequent N-stretches of Vigilant et al. (1991) as C-stretches around np 16189C; we follow this interpretation when entering the Vigilant et al. sequences. However, comparison with the original Vigilant sequences reveals that Redd et al. furthermore omit the informative insertions at np 309 and np 315. We have therefore added N at these positions when entering the new data of Redd et al. We obtained an electronic file of the sequences published in 1995 and 1999 (see below) from M. Stoneking and used this to correct five sequences (F04, CP10, TInew, F1new, 1803), to obtain extended sequence ranges, and to obtain detailed geographic information.

RED1 (Redd and Stoneking 1999) This publication is complementary to the Redd et al. (1995) study in preselecting sequences not having the 9 bp deletion at np 8272–8289 (see above). We omitted the single aboriginal Australian and the Papuan sequences previously published by Vigilant (1990). We did not omit the Papuan and Indonesian sequences stated to have been taken from Stoneking et al. (1992), as the latter publication does not display sequence data. The entire published table is erroneously shifted by one nucleotide backwards at np 16290, 16291, 16292, 16293, 16294 and 16295. The published table furthermore omits deletions at np 249 and misscores np 16224 and 16335. We obtained geographical information and extended sequences from the electronic file mentioned in our comments on Redd et al. (1995) (see above), which contains most of the published sequences.

REI (Reid et al. 1994) Full sequence information is available. We classified this sequence as non-representative due to preselection for a potentially mitochondrially transmitted disease.

RIC (Richards et al. 1996) The Portuguese, Basque and Spanish samples are a subset of those published in Côrte-Real et al. (1996), which is why we omitted them when entering the Richards data. Most of the sequences were determined between np 16090 and 16365, and some sequences suspected to be mtDNA group J were additionally sequenced around np 16069; to avoid a bias, we entered the sequenced region as starting at np 16090 throughout.

RIE (Di Rienzo and Wilson 1991) The first author kindly provided information on the geographic origin of those sequences described as Mediterranean in the publication (one of them is in fact Persian). We reassigned the deletion to np 16183.

ROU (Rousselet and Mangin 1998) We reassigned insertions to np 309 and np 315. We corrected the typographical error in sequence 10 by moving the T at np 16380.1 to np 16380.

SAI (Saillard et al. 2000) RFLP information is available.

SAJ (Sajantila et al. 1995) The high frequency (11/34) of different deletions found in the Volga sample is unprecedented in humans. We therefore shortened sequences to eliminate flanking deletions, and classified sequences with internal deletions as dubious.

SAL (Salas et al. 1998) np 16195 is misprinted as np 16295 in their table. We interpreted Galician 51 as having a variant at np 16261. We reassigned a deletion to np 16351.

SAN (Santos M. et al. 1994) We reassigned insertions to np 309. RFLP and 9 bp information is available.

SAS (Santos S. et al. 1996) RFLP and 9 bp information is available.

SCH (Schurr et al. 1999) RFLP and 9 bp information is available.

SEO (Seo et al. 1998) We interpreted Japanese 92 as having a transition at np 312 and insertions at np 315.1C and np 315.2C: np 16221 and np 260 are incorrectly scored relative to CRS, so we entered types 80 and 94 as having N at these positions pending feedback from the authors.

SHI (Shields et al. 1993) We classed the sequences as dubious due to irreproducibility of hypervariability at np 16362 in Eskimos (Forster et al. 1996; Saillard et al. 2000). The nucleotide numbering is erroneous in the legend to their Fig. 1 and population assignment ambiguous (Miller et al. 1996).

SMI (Smith et al. 1999) These sequences were preselected by RFLP analysis, we thus classified them as non-representative.

SOO1 (Soodyall et al. 1996) We classified these sequences as non-representative because they were preselected for the 9 bp deletion at np 8272–np 8289. The Pygmy sequences were evidently taken from Vigilant (1990) so we omitted them. Comparison with the original Vigilant data shows that insertions around np 309 were omitted (compare also Redd et al. above). The final three nucleotide positions in the table heading are frameshifted in their Table 2, but this only affects the sequences we omitted.

SOO2 (Soodyall et al. 1997) Although the Tristan da Cunha mtDNA types are traceable to five historical colonists, we entered Tristan da Cunha as the geographical location, as no native population existed on the island.

STA (Starikovskaya et al. 1998) In Table 5, CR sequence 13 should be listed under SIB40, and CR sequence 16 should be listed under SIB48. RFLP and 9 bp information is available.

STE (Stenico et al. 1996) We entered the data as published.

STK (Stoneking et al. 1995) We took Pommern (Germany) as the geographic origin of Schanzkowska.

STO (Stone and Stoneking 1998) We classified these ancient DNA sequences as non-representative. We omitted two sequences, in accordance with the authors' statement that they are likely cases of contamination. RFLP and 9 bp information is available.

SYK (Sykes et al. 1995) Haplotypes 49, 104, and 125 are corrected as noted in the GenBank records. We classified all sequences as non-representative since they were preselected on the basis of 9 bp deletion status or np 16265 transversion status. This classification also solves the problem of potential sample overlap with Lum et al. (1998), who obtained some of their samples from the IMM in Oxford (K. Lum and J. Martinson pers. comm.), and sample overlap with the Taiwanese of Melton et al. (1998) (V. Macaulay, pers. comm.). Comparison of the duplicate Taiwanese sequencing efforts reveals four discrepancies in Sykes et al. (1995), which we corrected; 9 bp information is available.

TOR and TOR1 (Torroni et al. 1993a, 1993b). We classified these sequences as non-representative since they were preselected for RFLP status. One sequence contradicts the RFLP cleavage pattern at np 16051 and since original documentation is unavailable (A. Torroni, pers. comm.), we have adhered to the sequencing results in this case.

VIG (Vigilant 1990; Vigilant et al. 1991) Chinese UC37 has a West African L1b sequence; this individual is from San Francisco and African admixture cannot be excluded (Penny et al. 1995). Another African mtDNA, HRO 3, appears to be a recombination of an African L3b sequence preceding np 16189 and a Polynesian sequence after np 16189; omission of this sequence is recommended by M. Stoneking (pers. comm.). We have classified both these sequences as dubious. We have interpreted the sequences of N's from np 16184–np 16192 as np 16189.C, in accordance with later publications by the same co-authors (Redd et al. 1995). We corrected the baseshift error np 16107/np 16111 as in Stoneking et al. (1991). We corrected np 196 as np 195. We corrected AA5 as AA1 on page 26. We entered np 198 as N for individual ZH4. We reassigned insertions to np 44.

WAR (Ward et al. 1991) RFLP and 9 bp information are available.

WAR1 (Ward et al. 1993) We entered the data as published.

WAR2 (Ward et al. 1996) 9 bp information is available.

WAT (Watson et al. 1997) The sequences described in the publication differ from those deposited earlier in GenBank in that some duplicates were omitted (Watson et al. 1997). RFLP and 9 bp information is available.

WEI (Weichhold et al. 1998) We classified the sequences of Hauser and one of Stephanie's descendants as non-representative and entered their locations as Nuremberg and Paris, respectively.

WTK (Watkins et al. 1999) We interpreted sample xh21 as a Xhosa rather than as a Sotho-Tswana, according to Soodyall (1993) and we assumed the Pedi samples to correspond to the Sotho-Tswana samples described previously by Jorde et al. (1995). The samples were preselected for the 9 bp deletion and we thus classified them as non-representative.

YON (Yoneda et al. 1990) The full mtDNA sequence is available. We classified the sequence as non-representative as the individual was preselected for a potential mitochondrial disease.

We did not enter data from the following publications:

Bamshad et al. (1998). The analysed sequences were not presented but published subsequently in Kivisild et al. (1999) (T. Kivisild, pers. comm.).

Greenberg et al. (1983). The sequence for individual pCJK5 underwent two cloning errors as shown by later resequencing (Vigilant 1990; Kocher and Wilson 1991), we have therefore entered the corrected sequence as one of Vigilant's (NC3). Another sample, pCDK1 was evidently resequenced and listed by Vigilant (NC2). The remaining five sequences are either from cell lines or from samples with uncertain origins (e.g. the "Caucasoid" sample pBHK2 has an Asian F sequence), so we omitted entering the Greenberg sequences altogether.

Soodyall (1993) and Sherry et al. (1994). The nucleotide numbering is partially frameshifted by one nucleotide in both HVR1 (Bandelt and Forster 1997) and HVR2. We have reverted both frameshifts, and have omitted the sequences at the request of the author pending a revised table (H. Soodyall, pers. comm.). We recoded Dama 14 as having a deletion at np 309 and a C insertion at np 315.

Stoneking et al. (1991). These data are a subset of Vigilant (1990) and Vigilant et al. (1991).

Stoneking et al. (1992). New sequences analysed but not presented in their publication comprise 30 individuals from coastal and highland Papua New Guinea and 20 individuals from the Moluccas and Lesser Sunda islands in Indonesia. The Indonesian sequences are preselected for the 9 bp deletion at np 8272–8289. Sequences with the 9 bp deletion are displayed in Redd et al. (1995) and the sequences without the deletion are displayed in Redd and Stoneking (1999).

Torroni et al. (1996) The sequence data had been published by Francalacci et al. (1996).

Vigilant et al. (1989) These data are a subset of those in Vigilant (1990).

Proofreading

We screened the database for systematic typing errors (caused either by ourselves or by the original authors) with the help of our notation system, which registers nucleotide positions only if they are different from the Cambridge reference sequence (Anderson et al. 1981). The notation system implies that any error will have a chance of 1 in 4 (assuming a homogeneous AGCT composition of the mtDNA control region, which naturally is only an approximation) of yielding a variant nucleotide position which looks identical to the CRS. The variant nucleotide positions are listed in the weighting option of the search programme, enabling a rapid check. With this approach, we found two systematic errors in our database (one HVR2 dataset had been accidentally defined as an HVR1 dataset and another dataset had transversions coded as transitions), as well as 29 isolated errors. Of the isolated errors, 12 were due to the original authors, whereas the 17 remaining ones were our errors. Of these 17 errors, 8 were due to confusing the variant nucleotide with the CRS nucleotide and the remaining 9 errors were typing errors. This would imply about $3 \times 9 = 27$ typing errors still present in the database of 14,138 individuals, which possibly is a minimum estimate because erroneously added or omitted nucleotides would not be detected.

Implementation

The database is linked to "mtradius", a geographic information system programmed for Windows. The user may specify the sequence range he or she is interested in (e.g. np 16093–np 16323 and np 73–np 315, thereby omitting the sequence from np 16324–np 72), and whether geographically or biologically doubtful sequences should be considered. The programme accordingly deactivates sequences that do not meet the specified criteria. The user then en-

ters a sequence as variants relative to the CRS. Optionally, the user may change the weight of any nucleotide position (low or zero weights may be given to rapidly mutating or unreliably scored positions; for example, the programme weights np 16182 and np 16183 as zero by default). Another application of the weighting option is to weight ancestral nucleotide states in order to enhance the search for phylogenetic clades. mtradius displays two outputs: the most similar sequences, the second-best matches and the third-best matches are listed, including the sequence, the publication, the ethnic or linguistic affinity, the geographical coordinates and the absolute frequency of each match. The second output displays the sequence matches on a high-resolution world map which can be zoomed, with best, second-best, third-best matches and remaining sequences displayed as coloured pie charts. Optionally, the display of particular match classes can be deactivated, for example to enhance the display of best matches in the pie charts. Local relative and absolute match frequencies can be queried for any rectangle projected on the surface of the earth by defining a latitudinal and longitudinal grid and clicking a grid square with the computer mouse.

Discussion

For forensic and population genetic applications, an important conclusion is that the peer review system as well as proofreading procedures have failed to eliminate obvious data errors in the majority of publications. The database presented here resolves some of these errors but there are probably many more which remain undetected. From the forensic point of view, our database can be used for qualitative identification of relevant reference populations for a given mtDNA type, whereas the determination of a legally defensible frequency estimate of an mtDNA type within a population should be performed with higher-quality data yet to be produced. Sequence errors will typically lead to biologically non-existent sequences, so this database would only provide a lower bound for a frequency estimate. Changes are necessary at several levels to improve the quality of mtDNA data. The author of a manuscript should present mtDNA sequences in increasing or decreasing order of the first variable nucleotide position (e.g. Pfeiffer et al. 1999), so that erroneously omitted or "shifted" nucleotides and HVR1/HVR2 "recombination events" become visually apparent. The reviewer in turn should investigate potential omissions and shifts via a database: sequences with omitted nucleotides would be at distance 1 from real sequences, and sequences with shifted nucleotides would be found at distance 2. The reviewer should then routinely request copies of electropherograms. When potential mistakes have been published and identified by other researchers, it would be helpful if the authors and editors developed more enthusiasm for participating in errata – we received feedback from only about 50% of our queries. An increasing number of forensic institutes will establish their own mtDNA databases: ideally, contributors to such databases should provide not only sequence information, but also electropherograms and possibly even samples for future resolution of problems.

Acknowledgements We thank C. Brinkmann for contributing unpublished sequence data.

References

Alves-Silva J, Santos MS, Pena SDJ, Prado VF (1999) Multiple geographic sources of region V 9-bp deletion haplotypes in Brazilians. Hum Biol 71:245–259

Alves-Silva J, Silva Santos M da, 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

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC et al, (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Arnason U, Xu X, Gullberg A (1996) Comparison between the complete mitochondrial DNA sequences of Homo and the common chimpanzee based on nonchimeric sequences. J Mol Evol 42: 145–152

Baasner A, Schäfer C, Junge A, Madea B (1998) Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. Forensic Sci Int 98:169–178

Bamshad M, Fraley AE, Crawford MH, Cann RL, Busi BR, Naidu JM, Jorde LB (1996) MtDNA variation in caste populations of Andhra Pradesh, India. Hum Biol 68:1–28

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu M, Prasad BVR, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies Hindu castes. Nature 395:651–652

Bandelt H-J, Forster P (1997) The myth of bumpy hunter-gatherer mismatch distributions. Am J Hum Genet 61:980–983

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

Barbujani G, Stenico M, Excoffier L, Nigro L (1996) Mitochondrial DNA sequence variation across linguistic and geographic boundaries in Italy. Hum Biol 68:201–215

Batista O, Kolman CJ, Bermingham E (1995) Mitochondrial DNA diversity in the Kuna Amerinds of Panamá. Hum Mol Genet 4:921–929

Belledi M, Poloni ES, Casalotti R, Conterio F, Mikerezi I, Tagliavini J, Excoffier L (2000) Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. Eur J Hum Genet 8:480–486

Benedictis G de, Rose G, Carrieri G, Luca M de, Falcone E, Passarino G, Bonafe M, Monti D, Baggio G, Bertolini S, Mari D, Mattace R, Franceschi C (1999) Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. FASEB J 13:1532–1536

Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. Ann Hum Genet 59:63–81

Betty DJ, Chin-Atkins A, Croft L, Sraml M, Easteal S (1996) Multiple independent origins of the COII/tRNALys intergenic 9-bp mtDNA deletion in aboriginal Australians. Am J Hum Genet 58:428–433

Bortolini MC, Zago MA, Salzano FM, Silva-Junior WA, Bonatto SL, Da Silva MCBO, Weimer TA (1997) Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. Hum Biol 69:141–159

Brown MD, Hosseini SH, Torroni A, Bandelt H-J, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC (1998) MtDNA haplogroup X: an ancient link between Europe/western Asia and North America? Am J Hum Genet 63:1852–1861

Burckhardt F, Haeseler A von, Meyer S (1999) HvrBase: compilation of control region sequences from primates. Nucleic Acids Res 27:138–142

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

Calí F, Le Roux MG, D'Anna R, Flugy A, De Leo G, Ayala GF, Romano V (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med (in press)

Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000) MtDNA variation in the South African Kung and Khwe – and their genetic relationships to other African populations. Am J Hum Genet 66:1362–1383

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J (1996) Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. Mol Biol Evol 13: 1067–1077

Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Martinez-Arias R, Clarimon J, Facchini F, Fiori G, Luiselli D, Pettener D, Bertranpetit J (1998) Trading genes along the Silk Road: mtDNA sequences and origin of central Asian populations. Am J Hum Genet 63:1824–1838

Comas D, Calafell F, Bendukidze N, Fañanas L, Bertranpetit J (2000) Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. Am J Phys Anthropol 112:5–16

Côrte-Real HBSM, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, Bertranpetit J, Sykes BC (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. Ann Hum Genet 60: 331–350

Delghandi M, Utsi E, Krauss S (1998) Saami mitochondrial DNA reveals deep maternal lineage clusters. Hum Hered 48:108–114

Dimo-Simenon 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

Dupuy BM, Olaison B (1996) MtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics 6. Springer, Berlin Heidelberg New York, pp 23–25

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

Finnilä S, Hassinen IE, Ala-Kokko L, Majamaa K (2000) Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. Am J Hum Genet 66:1017–1026

Forster P (1996) Dispersal and differentiation of modern *Homo sapiens* analysed with mitochondrial DNA. PhD thesis, University of Hamburg, Germany

Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. Am J Phys Anthropol 100:443–460

Gill P, Pavel L, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nat Genet 6:130–135

Ginther C, Corach D, Penacino GA, Rey JA, Carnese FR, Hutz MH, Anderson A, Just J, Salzano FM, King M-C (1993) Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) DNA fingerprinting: state of the science. Birkhäuser, Basel, pp 211–219

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L (1995) Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345

Green LD, Derr JN, Knight A (2000) MtDNA affinities of the peoples of north-central Mexico. Am J Hum Genet 66:989–998

Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49

Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, Schaffner W, Utermann G, Sykes B, Pääbo S (1994) Molecular genetic analyses of the Tyrolean Ice Man. Science 264:1775–1778

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

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A, Gerbitz K-D (1997) Population genetics and disease susceptibility: characterisation of central European haplogroups by mtDNA gene mutations, correlation and association with disease. Hum Mol Genet 6:1835–1846

Holst Pellekan SM van, Frommer M, Sved JA, Boettcher B (1998) Mitochondrial control-region sequence variation in aboriginal Australians. Am J Hum Genet 62:435–449

Horai S, Hayasaka K (1990) Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am J Hum Genet 46:828–842

Horai S, Kondo R, Nakagawa-Hattori Y, Hayashi S, Sonoda S, Tajima K (1993) Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol Biol Evol 10:23–47

Horai S, Hayasaka K, Kondo M, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536

Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara S, Park KS, Omoto K, Pan I-H (1996) MtDNA polymorphism in east Asian populations, with special reference to the peopling of Japan. Am J Hum Genet 59:579–590

Ivanova AV (1993) Mitochondrial DNA polymorphism in native inhabitants of Chukotka. PhD thesis, University of Novosibirsk

Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, Soodyall H, Jenkins T, Rogers AR (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538

Jun AS, Brown MD, Wallace DC (1994) A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. Proc Natl Acad Sci USA 91:6206–6210

Kittles RA, Bergen AW, Urbanek M, Virkkunen M, Linnoila M, Goldman D, Long JC (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108:381–399

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

Kobayashi Y, Momoi MY, Tominaga K, Shimoizumi H, Nihei K, Yanagisawa M, Kagawa Y, Ohta S (1991) Respiration-deficient cells are caused by a single point mutation in the mitochondrial tRNA-Leu (UUR) gene in mitochondrial myopathy, encephalopathy, lactic acidosis, and strokelike episodes (MELAS). Am J Hum Genet 49:590–599

Kocher TD, Wilson AC (1991) Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In: Osawa S, Honjo T (eds) Evolution of life: fossils, molecules, and culture. Springer, Tokyo, pp 391–413

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F (1995) Reduced mtDNA diversity in the Ngöbé Amerinds of Panamá. Genetics 140:275–283

Kolman CJ, Sambuughin N, Bermingham E (1996) Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. Genetics 142:1321–1334

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, Stoneking M (1999) MtDNA analysis of Nile River Valley populations: a genetic corridor or a barrier to migration? Am J Hum Genet 64:1166–1176

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus M-L (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58:1309–1322

Lahermo P, Laitinen V, Sistonen P, Beres J, Karcagi V, Savontaus ML (2000) MtDNA polymorphism in the Hungarians: comparison to three other Finno-Ugric-speaking populations. Heriditas 132:35–42

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

Levin BC, Cheng H, Reeder DJ (1999) A human mitochondrial standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection. Genomics 55:135–146

Lorenz GJ, Smith DG (1997) Distribution of sequence variation in the mtDNA control region of native north Americans. Hum Biol 69:749–776

Lum JK, Cann RL (2000) MtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. Am J Phys Anthropol 113:151–168

Lum JK, Rickards O, Ching C, Cann RL (1994) Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. Hum Biol 66:567–590

Lum JK, Cann RL, Martinson JJ, Jorde LB (1998) Mitochondrial and nuclear genetic relationships among Pacific island and Asian populations. Am J Hum Genet 63:613–624

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

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

Malyarchuk BA, Derenko MV, Solovenchuk LL (1995) Types of mitochondrial DNA control region in eastern Slavs. Russ J Genet 31:723–725

Mateu E, Comas D, Calafell F, Pérez-Lezaun A, Abade A, Bertranpetit J (1997) A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. Ann Hum Genet 61:507–518

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. The McDonald Institute for Archaeological Research, Cambridge, p 108

Miller KWP (1996) Molecular genetic analysis of human populations in Orkney and the north Atlantic region. PhD thesis, University of Cambridge, England

Miller KWP, Dawson JL, Hagelberg E (1996) A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region. Int J Legal Med 109:107–113; http://shelob.bioanth.cam.ac.uk/mtDNA/

Monsalve MV, Hagelberg E (1997) Mitochondrial DNA polymorphisms in Carib people of Belize. Proc R Soc Lond B 264:1217–1224

Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottolenghi C, Gadgil M, Cavalli-Sforza LL (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992

Murray-McIntosh RP, Scrimshaw BJ, Hatfield PJ, Penny D (1998) Testing migration patterns and estimating founding population size in Polynesia by using human mtDNA sequences. Proc Natl Acad Sci USA 95:9047–9052

Nata M, Adachi N, Hashiyada M, Kanetake J, Ji G, Mimasaka S, Takahashi K, Sagisaka K, Funayama M (1999) Sequence polymorphism of the D-loop of mitochondrial DNA in Japanese and Chinese populations. Acta Crim Japon 65:244–254

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

Nishino I, Seki A, Maegaki Y, Takeshita K, Horai S, Nonaka I, Goto Y (1996) A novel mutation in the mitochondrial $tRNA^{Thr}$ gene associated with a mitochondrial encephalomyopathy. Biochem Biophys Res Comm 225:180–185

National Research Council Committee on DNA Technology in Forensic Science (1992) DNA technology in forensic science. National Academy Press, Washington, DC

Ohlenbusch A, Wilichowski E, Hanefeld F (1998) Characterisation of the miochondrial genome in childhood multiple sclerosis. I. Optic neuritis and LHON mutations. Neuropediatrics 29: 175–179

Oota H, Saitou N, Matsushita T, Ueda S (1995) A genetic study of 2000-year-old human remains from Japan using mitochondrial DNA sequences. Am J Phys Anthropol 98:133–145

Opdal SH, Rognum TO, Vege Å, Stave AK, Dupuy BM, Egeland T (1998) Increased number of substitutions in the D-loop of mitochondrial DNA in the sudden infant death syndrome. Acta Pædiatr 87:1039–1044

Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov P, Yankovsky N (1999) Mitochondrial DNA sequence diversity in Russians. FEBS Lett 445:197–201

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 (1991a) 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

Ozawa T, Tanaka M, Sugiyama S, Ino H, Ohno K, Hattori K, Ohbayashi T, Ito T, Dguchi H, Kawamura K, Nakane Y, Hashiba K (1991b) Patients with idiopathic cardiomyopathy belong to the same mitochondrial DNA gene family of Parkinson's disease and mitochondrial encephalomyopathy. Biochem Biophys Res Comm 177:518–525

Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. Int J Legal Med 111:124–132

Penny D, Steel M, Waddell PJ, Hendy MD (1995) Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. Mol Biol Evol 12:863–882

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

Pfeiffer H, Brinkmann B, Hühne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112: 291–298

Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. Int J Legal Med 114:169–172

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

Pinto F, González AM, Hernández M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60:321–330

Polyak K, Li Y, Zhu H, Lengauer C, Willson JKV, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B (1998) Somatic mutations of the mitochondrial genome in human colorectal cancers. Nat Genet 20:291–293

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe-Seyler 375:837–840

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

Rando JC, Pinto F, Gonzalez AM, Hernandez 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

Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoneking M (1995) Evolutionary history of the COII/tRNALys intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. Mol Biol Evol 12:604–615

Reid FM, Vernham GA, Jacobs HT (1994) Complete mtDNA sequence of a patient in a maternal pedigree with sensorineural deafness. Hum Mol Genet 3:1435–1436

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

Rienzo A Di, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Rousselet F, Mangin O (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 M-L, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver 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, Carracedo A (1998) MtDNA analysis of the Galician population: a genetic edge of European variation. Eur J Hum Genet 6:365–375

Santos M, Ward RH, Barrantes R (1994) MtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. Hum Biol 66: 963–977

Santos SEB, Ribeiro dos Santos AKC, Meyer D, Zago MA (1996) Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. Ann Hum Genet 60: 305–319

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

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

Sherry ST, Rogers AR, Harpending H, Soodyall H, Jenkins T, Stoneking M (1994) Mismatch distributions of mtDNA reveal recent human population expansions. Hum Biol 66:761–775

Shields GF, Schmiechen AM, Ward RH (1993) MtDNA sequences suggest a recent evolutionary divergence for Beringian and North American populations. Am J Hum Genet 53:549–562

Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999) Distribution of mtDNA haplogroup X among native North Americans. Am J Phys Anthropol 110:271–284

Soodyall H (1993) Mitochondrial DNA polymorphisms in southern African populations. PhD thesis, University of the Witwatersrand, Johannesburg

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) MtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

Soodyall H, Jenkins T, Mukherjee A, Du Toit E, Roberts DF, Stoneking M, et al, (1997) The founding mitochondrial DNA lineages of Tristan da Cunha Islanders. Am J Phys Anthropol 104:157–166

Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998) MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. Am J Hum Genet 63: 1473–1491

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

Stone AC, Stoneking M (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62:1153–1170

Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. Am J Hum Genet 48:370–382

Stoneking M, Sherry ST, Redd AJ, Vigilant L (1992) New approaches to dating suggest a recent age for the human mtDNA ancestor. Phil Trans R Soc Lond B 337:167–175

Stoneking M, Melton T, Nott J, Barritt S, Roby R, Holland M, Weedn V, Gill P, Kimpton C, Aliston-Greiner R, Sullivan K (1995) Establishing the identity of Anna Anderson Manahan. Nat Genet 9:9–10

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 57:1468–1475

Torroni A, Sukernik RI, Schurr TG Starikovskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993a) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with native Americans. Am J Hum Genet 53:591–608

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, Wallace DC (1993b) Asian affinities and continental radiation of the four founding native American mtDNAs. Am J Hum Genet 53:563–590

Torroni A, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T, Wallace DC (1994) Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaption to high altitude. Am J Phys Anthropol 93:189–199

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Vigilant LA (1990) Control region sequences from African populations and the evolution of human mitochondrial-DNA. PhD thesis, University of California, Berkeley

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. Science 253:1503–1507

Ward RH, Frazier BL, Dew-Jager K, Pääbo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci USA 88:8720–8724

Ward RH, Redd A, Valencia D, Frazier B, Pääbo S (1993) Genetic and linguistic differentiation in the Americas. Proc Natl Acad Sci USA 90:10663–10667

Ward RH, Salzano FM, Bonatto SL, Hutz MH, Coimbra CEA, Santos RV (1996) Mitochondrial DNA polymorphism in three Brazilian Indian tribes. Am J Hum Biol 8:317–323

Watkins WS, Bamshad M, Dixon ME, Bhaskara Rao B, Naidu JM, Reddy PC, Gai PB, Bhanu A, Kusuma YS, Lum JK, Fischer P, Jorde LB (1999) Multiple origins of the mtDNA 9bp-deletion in populations of south India. Am J Phys Anthropol 109:147–158

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

Weichhold GM, Bark JE, Korte W, Eisenmenger W, Sullivan KM (1998) DNA analysis in the case of Kaspar Hauser. Int J Legal Med 111:287–291

Yoneda M, Tanno Y, Horai S, Ozawa T, Miyatake T, Tsuji S (1990) A common mitochondrial DNA mutation in the t-RNA^{Lys} of patients with myoclonus epilepsy associated with ragged-red fibers. Biochem Int 21:789–796