Int J Legal Med (2004) 118 : 267–273
DOI 10.1007/s00414-004-0455-2

ORIGINAL ARTICLE

H.-J. Bandelt · A. Salas · S. Lutz-Bonengel

Artificial recombination in forensic mtDNA population databases

Received: 11 December 2003 / Accepted: 7 April 2004 / Published online: 15 July 2004

Abstract Artificial recombination of two or more mitochondrial DNA fragments from different samples would constitute a serious cause of error in forensic DNA typing, and yet one can demonstrate that such events have happened in the preparation of several published mtDNA databases. Focussed database searches, phylogenetic analysis, and network representations can highlight mosaic patterns and thus pinpoint sample mix-up. Therefore, we suggest that this approach should be applied to data prior to publication in order to uncover such errors in time.

Keywords Mitochondrial DNA · Error detection · Recombination · Networks

Introduction

Sequencing mitochondrial DNA (mtDNA) for forensic purposes involves several independent amplifications of different fragments, such as the hypervariable regions (HVR-I, HVR-II, and occasionally HVR-III) and, most recently, also mitochondrial genes from the coding region. It is mandatory that for each of the regions screened the same sample is targeted. There are four main sources of error that lead to artificial recombination: first, the mtDNA under examination may be fairly degraded, so that in one or several instances some contaminant mtDNA predominates after amplification (sample mixture), or second, the primers for amplifications are contaminated or do not bind properly (sequencing artifact), or third, the person handling the DNA samples confuses sample numbers and thus creates artificial mosaic haplotypes (sample mix-up), or fourth, partial sequences are interchanged during the process of transferring the data in tables and databases.

H.-J. Bandelt (✉)
Fachbereich Mathematik, Universität Hamburg,
Bundesstrasse 55, 20146 Hamburg, Germany
Tel.: +49-40-428385141, Fax: +49-40-428385190,
e-mail: bandelt@math.uni-hamburg.de

A. Salas
Unidad de Genética Instituto de Medicina Legal,
Facultad de Medicina,
c/San Francisco s/n, 15705 Santiago de Compostela,
A Coruña, Galicia, Spain

S. Lutz-Bonengel
Institut für Rechtsmedizin, Klinikum der Universität Freiburg,
Albertstrasse 9, 79104 Freiburg, Germany

HVR-I versus HVR-II

To give a first illustration, let us recall the following prime instance of recombination between HVR-I and HVR-II (briefly mentioned by Bandelt et al. 2001). Sample #105 from Dimo-Simonin et al. (2000) is recorded as:

– 16126-16294-16296-73-152-182-185T-195-247-263-315.1C-357

where numbers refer to the corresponding sites in the revised Cambridge reference sequence (rCRS, Andrews et al. 1999); all changes are transitions except for one (suffixed by T) that is a transversion and one insertion of C (scored at the end of a C-stretch). The HVR-I part of this sequence points to a typical member of haplogroup T (Richards et al. 2000), namely, more specifically, to the branch T2 (Finnilä et al. 2001):

– T2 (ancestral motif) 16126-16294-16296-73-263-315.1C.

The ancestral T2 sequence matches e.g. sequence #80 from Piercy et al. (1993). On the other hand, the HVR-II part of the above sequence #105 is identical to the consensus/ancestral motif of the sub-Saharan mtDNA haplogroup L1b:

– L1b (ancestral motif) 16126-16187-16189-16223-16264-16270-16278-16311-73-152-182-185T-195-247-263-315.1C-357

(Graven et al. 1995; Watson et al. 1997). The mosaic origin of sample #105, with its mixed HVR-I/HVR-II haplo-

group association, which we denote by T2×L1b, thus only becomes apparent with mtDNA lineages of African ancestry. Singular back mutations at the characteristic HVR-I sites have been reported in haplogroup L1b, but two joint back mutations are rare and three joint back mutations have not been found so far (Salas et al. 2002). Therefore it is not realistic to assume that 16187-16189-16223-16264-16270-16278-16311 turned into 16294-16296, or vice versa. As to the second segment, even each of the extremely rare mutations 185T and 357 are virtually exclusive to mtDNA lineages belonging to L1b, with the sole exception of two sequences having the 357 transition, one presumably from haplogroup H (Malyarchuk et al. 2002) and another one from haplogroup pre-HV (Monson et al. 2002). Moreover, 247 is an excellent marker separating super haplogroup L2'3 (which includes haplogroup T) from the African haplogroups L1 and L0 (*sensu* Mishmar et al. 2003). Hence, there is practically no way for an innocent haplogroup T molecule to acquire in parallel those three rare plus three further HVR-II mutations *en bloque*. Note that a phylogenetic analysis of the data set of Dimo-Simonin et al. (2000) alone would not necessarily have led to the discovery of the trans-continental sample mix-up since otherwise all lineages are apparently of European ancestry. Since haplogroup T as a whole does not possess any characteristic HVR-II mutations, the recombinant T2×L1b type would have looked like a genuine haplogroup T sequence with a somewhat diverged HVR-II part. The reticulate pattern becomes most pronounced when we compare the T2×L1b type with genuine sequences from haplogroups T2 (Piercy et al. 1993, #80), L1b (Graven et al. 1995, #45), and L0a (Ingman et al. 2000, #Mbu-1057/1058): the median network (Bandelt et al. 2000) representing these four sequences includes a large rectangle where one side is exclusively supported by mutations from HVR-I (16187, 16189, 16223, 16294, 16296, and 16311) and the other by HVR-II mutations (152, 182, 185T, 195, and 357).

Poetsch et al. (2003) reported the following compound HVR-I×HVR-II sequence (#2 in their table), which they allocate to haplogroup C:

– 16223-16288-16298-16327-73-143-189-194-195-204-207-263-309.1C-315.1C.

This sequence should be compared with the reconstructed ancestral sequences of haplogroup C (Kong et al. 2003b) and the European haplogroup W1 (which was referred to as haplogroup W in Finland by Finnilä et al. 2001):

– C (ancestral motif): 16223-16298-16327-73-249d-263-315.1C
– W1 (ancestral motif): 16223-16292-73-189-195-204-207-263-315.1C.

In haplogroup C, the transition 16288 is not infrequent (e.g. Comas et al. 1998), and as for haplogroup W1, the mutations 143 and 194 occur (on top of the ancestral W1 HVR-II motif) in a minor branch of haplogroup W1 (Finnilä et al. 2001). The mosaic haplotype #2 from Poetsch et al. (2003) therefore definitely constitutes the result of a recombination event in the laboratory.

Further instances of recombination between the two hypervariable regions can be retrieved from the SWGDAM (Scientific Working Group on DNA Analysis Methods) database (Monson et al. 2002). In particular, the "African-American" database, contributed by the Federal Bureau of Investigation (FBI), offers a number of clear-cut recombinants; see Table 1. The first two entries constitute mixed L1b×L2a types. Indeed, the common HVR-II part nearly matches the corresponding part of a particular haplogroup L2a sequence from the "Hispanic" database:

– 16189-16223-16278-16294-16309-16390-16519-73-143-146-152-195-263-264-309.1C-315.1C-534 (L2a; USA.HIS. 000727).

The third entry also has a haplogroup L1b HVR-I part, but now combined with a HVR-II sequence from haplogroup C1, which has the characteristic double deletion 290d-291d (Bandelt et al. 2003). The fourth type in the table is clearly identified as a recombinant L2b×L0a type, (cf. Graven et al. 1995; Alves-Silva et al. 2000; Torroni et al. 2001). The fifth instance from Table 1 bears a rather unspecific HVR-II part, which would best match the consensus HVR-II motif of the African haplogroup L3f (in the nomenclature of Salas et al. 2002). But inasmuch as the Native American haplogroup A2 has rather unmistakeable HVR-I and HVR-II motifs (despite singular back mutations; Bandelt et al. 2003), one can safely infer recombination also in this case, especially when one compares genuine A2 and L3f sequences from the FBI database:

– 16051-16111-16209-16223-16290-16319-16362-73-146-153-235-263-309.1C-315.1C (A2; USA.HIS. 000122)
– 16129-16209-16223-16292-16295-16311-16519-73-189 200-263-315.1C (L3f; USA.AFR.000193).

Table 1 will actually not exhaust all instances of sample mix-up in the "African-American" database of the FBI,

Table 1 Recombinant HVR-I×HVR-II types in the SWGDAM database (Monson et al. 2002)

Code	HVR-I (16000+)	HVR-II	Haplogroup status
USA.AFR.000063	067-126-187-189-204-223-264-270-278-311	73-143-146-152-195-263-264-315.1C	L1b×L2a
USA.AFR.000074	126-187-189-223-264-270-278-311	73-143-146-152-195-263-264-315.1C	L1b×L2a
USA.AFR.000942	126-187-189-223-264-270-278-293-311-519	73-249d-263-290d-291d-309.1C-315.1C-489	L1b×C1
USA.AFR.000790	114A-129-213-223-248N-278-354	93-95C-185-189-236-247-263-315.1C	L2b×L0a
USA.AFR.000890	111-209-223-290-319-362	73-189-200-263-315.1C	A2×L3f?
FRA.CAU.000084	298	73-185-188-228-263-295-315.1C	pre-V×J2

since we would highlight only clear-cut recombination instances. The other constituents of the SWGDAM database may also include dubious sequences. For example, the "French Caucasian" series of claimed size 109 (Miller and Budowle 2001) but actually having only 107 entries (as of September 2003) essentially reappears as the mtDNA data set of 111 French individuals published by Cali et al. (2001) in slightly modified (corrected) form. The latter data set has been analyzed for 16008–16392 and 72–340, whereas the former series was generally truncated to 16024–16365 and 73–340. In particular, the truncated sequence FRA.CAU.000084 (last row of Table 1) appears as #1064 among the sequences of Cali et al. (2001). The HVR-I parts of these sequences are, however, not yet (as of September 2003) included in Table S1 from the journal website (but are available from the authors upon request). Inasmuch as the HVR-II part (typical of haplogroup J1 sequences bearing the coding region mutation 14798 in addition; Finnilä et al. 2001) would be linked to an expected HVR-I part 16069–16126, whereas the HVR-I part 16298 (typical of haplogroup pre-V) would be linked to an expected HVR-II part 72-263-315.1C, we can safely infer the mosaic status pre-V×J1 of this type.

In any case, one should be prepared to see databases changing over time (for the better). A good case in question constitutes the dynamic data tables of Tagliabracci et al. (2001). In their originally published Table S1 (downloaded as of April 2001), types #19, 42, and 44 were evidently of mosaic haplogroup status, viz. K×X, X×H, and J×H, respectively. In a recent (September 2003) download of the same Table S1 from the journal website the corresponding HVR-II parts of these three haplotypes appear to have been corrected; moreover, haplotypes #40 and 53 received a slightly modified HVR-I part as well and haplotypes #1, 45, 50, and 63 were corrected for sites 309.1/315.1, 324, 198, and 295 (thus restoring the HVR-II marker for haplogroup J), respectively. It is not surprising that no more than about one-half of the sequencing/documentation errors could actually be predicted through database comparisons.

The former examples all constitute clear instances of recombination, as chunks of mutations are recombined. However, there are other situations where only a single, rare mutation could signal a potential recombination event. For instance, the transition at site 235 in HVR-II is characteristic of haplogroup A, namely:

– A2 (ancestral motif): 16111-16223-16290-16319-16362-64-73-146-153-235-263-315.1C.

The 235 transition is also found in the following mtDNAs, for which the corresponding haplogroup status (G2, M10, and F2, respectively) can easily be inferred from the phylogeny of complete mtDNAs (Kong et al. 2003b):

– 16223-16278-16362-73-235-263-309.1C-315.1C (G2; Pfeiffer et al. 1998, #44),
– 16066-16223-16311-73-152-235-263-315.1C (M10; Imaizumi et al. 2002, #JP003),
– 16261-73-194-235-249d-309.1C-315.1C (F2; Yao et al. 2002, #GD7810).

At least the latter F2 sequence could not have been generated by sample mix-up because the HVR-II part has the deletion of nucleotide A at site 249 expected for haplogroup F. The 235 transition is, however, also found in mtDNA lineages of European ancestry in rare cases: for example, a minor side branch of haplogroup X2b is defined by 235 (Reidla et al. 2003; cf. also sample #103 from Crespillo et al. 2000). So, it is rather conceivable that a 235 transition may appear in parallel, although rather infrequently, so that it would not necessarily pinpoint recombination on its own (*contra* Maruyama et al. 2003), unless additional evidence would suggest sample mix-up. Since site 235 is hence not extremely conservative, one could also expect some rare back mutations from G to A at this site. Indeed, of the 193 haplogroup A2 sequences sampled from Apaches and Navajos (Budowle et al. 2002), one single case of back mutation is documented. The "Hispanic" database (as part of the SWGDAM database, contributed by the FBI) includes as many as eight cases of this kind; see Table 2. The pattern here, however, suggests that at least three instances must constitute clerical errors, which turned the expected 235G into 236G, or 235C, or 253G, respectively. But inasmuch as one would (most parsimoniously) reconstruct more than one independent mutational event, we could suspect further oversights rather than exclusively natural back mutations among the last five entries of Table 2. At least, recombination is rather unlikely to have acted here because, except for USA.HIS. 000274, both 146C and 153G are present in these instances.

Table 2 Haplogroup A2 types in the SWGDAM database (Monson et al. 2002) lacking the 235 mutation

Code	HVR-I (16000+)	HVR-II
USA.HIS. 000110	111-223-290-319-335	73-146-153-236G-263-309.1C-315.1C
USA.HIS. 000204	111-182C-183C-189-223-290-319-362	73-146-153-235C-263-310-315d-316d-317d
USA.HIS. 000100	111-223-290-319-362	73-146-153-253G-263-315.1C
USA.HIS. 000093	111-187-223-290-319-362	73-146-153-263-309.1C-315.1C
USA.HIS. 000267	111-223-290-319-335-362	73-146-153-263-310
USA.HIS. 000274	111-223-290-319-362	73-146-152-263-309.1C-315.1C
USA.HIS. 000552	111-223-290-319-362-524	64-73-143-146-152-153-185-195-263-315.1C-523d-524d
USA.HIS. 000770	111-223-290-319-362-524	64-73-143-146-152-153-185-263-315.1C-523d-524d

Coding region versus HVR-I and HVR-II

When additional coding region information is combined with HVR-I and HVR-II sequences, the chances for artificial recombination become higher since additional sequencing steps are involved. The Korean mtDNA data presented by Lee et al. (2002) bear the imprint of several erroneous combinations of the HVR-I and HVR-II sequences with the cytochrome B sequences. Some of the HVR-I and HVR-II sequences were published earlier (Lee et al. 1997) but without citation in Lee et al. (2002). Alarmingly, some of those recycled HVR-I and HVR-II sequences do not completely match the corresponding ones of the earlier study. With the present knowledge of the East Asian mtDNA phylogeny (Kivisild et al. 2002; Kong et al. 2003a, 2003b), one can allocate almost two-thirds of the cytochrome B mutations recorded by Lee et al. (2002) to specific branches of the mtDNA phylogeny. When comparing the HVR-I and HVR-II motifs with the cytochrome B motifs, we find the following mosaic types (where the first part refers to the control region and the second part to cytochrome B gene): sample F531.2 is an F1b×D4a type, sample H81 an A5×D4a type, sample H84 an M7a1×D4a type, sample H98 an A5×M type (Kong et al. 2003a; Bandelt 2004), and sample H75 is an M10×Y1 type. Indeed, as to the latter, the recorded cytochrome B mutations for H75 include the transitions at sites 15221 and 15460, which were found in sample XJ8426 of Kong et al. (2003b) belonging to haplogroup Y1. The M10×Y1 mix-up is quite astonishing because the whole data set does not include any other (partial) Y1 sequence! One should, however, keep in mind that many samples were recycled from the earlier paper by Lee et al. (1997), which contain 9 Y1 sequences recognizable by the motif 16126-16231-16266-73-146-263-315.1C. In particular, sample H76 from Lee et al. (1997) is actually from haplogroup Y1, which thus may have been mixed up with H75. Since only those recombination events in the data of Lee et al. (2002) can be identified that have hit samples of different haplogroup status as inferred from cytochrome B, a straightforward calculation shows that we should expect 3 unidentifiable instances per 7 recognizable instances on average. This means one has to reckon with an expected total number of about 6–7 recombination events here.

There are probably several artificial back mutations in the HVR-I, HVR-II, and cytochrome B sequences in these data as well (Kong et al. 2003a). In any case, their Table 2 contains some clerical errors: sample codes H17 (appearing twice) and P135 are mistyped, in one instance 73G was obviously mistyped as 173G. Moreover, the first part of the table refers to an incorrect CRS; namely, for all samples with code prefix "F" the states at 16362 in the reference sequence were interchanged! That is, each 16362T listed among the "F" samples means no mutation relative to CRS at this site, whereas the absence of 16362 means a mutation relative to CRS. We conclude that these Korean mtDNA data (Lee et al. 1997, 2002) are not appropriate for construction of a forensic database.

Mosaic coding region sequences

A prime example of mosaic compound sequences from (part of) the coding region is given by the data set of Silva et al. (2002), which contains a variety of errors (Yao et al. 2003). In other data sets, with fewer fragments of the coding region analyzed and a lower error rate, it may be much harder to identify sample mix-up or documentation errors by posterior data analysis. Nevertheless, a focussed search may quickly identify problematic haplotypes. Lutz-Bonengel et al. (2003) have sequenced six fragments of the coding region (see our Table 3) that were independently amplified and read. Since the compound haplotypes constitute binary data in this case, that is, each polymorphic site has no more than two different nucleotides across all samples, the variation can be represented by a median network (Bandelt et al. 2000). In order to highlight potentially recurrent mutations, we focus only on the parsimoniously informative sites, that is, those sites where at least two samples deviate from the other samples. The corresponding network representing the thus truncated haplotypes is displayed in Fig. 1, where numbers inscribed in the nodes refer to the original haplotypes of Table 3 of Lutz-Bonengel et al. (2003) represented by the first sequence in each row. The network is surprisingly complex and thus testifies to quite an amount

Table 3 Correction of the erroneous haplotypes #*78*, *85*, *98* from Lutz-Bonengel et al. (2003)

Haplotype	8354–8804	8735–9202	10709–11070	11063–11478	11422–11834	11783–11910*
#*78*	rCRS	8860	rCRS	11251	rCRS	rCRS
#78n	rCRS	8860	rCRS	11251	11719	rCRS
#*85*	rCRS	8860	rCRS	11332, 11467	rCRS**	rCRS
#85f	rCRS	8860	10907,11009	11332, 11467	11719	rCRS
#85m	rCRS	8860	rCRS	rCRS	rCRS	rCRS
#*98*	8697	8860	rCRS	11161, 11251	rCRS	rCRS
#98f	8697	8860	rCRS	11151, 11251	11719, 11812	11812
#98m	rCRS	8860	rCRS	11404	rCRS	rCRS

*Analyzed only in forward direction from two independent amplifications.
**Using primer L11474 instead of L11421.
Numbers in the first row denote reading frames of amplified fragments.
Numbers below the first row refer to transitions relative to rCRS.

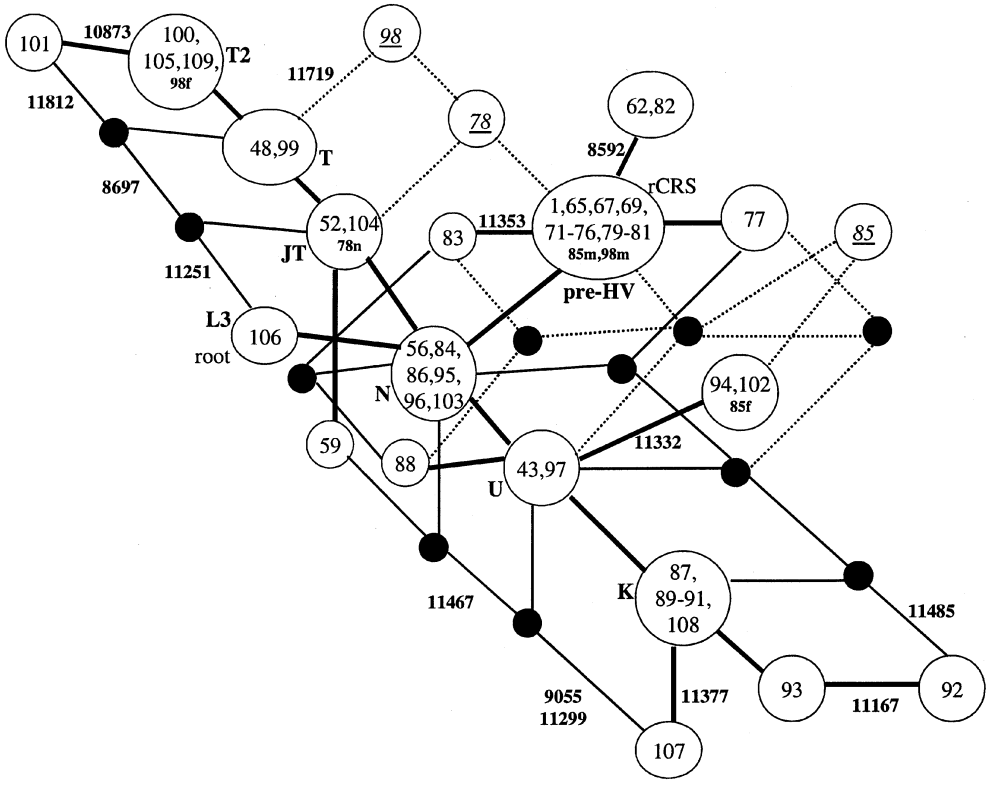

Fig. 1 The median network representing the (truncated) variation at the 14 parsimoniously informative sites of the mtDNA data set of Lutz-Bonengel et al. (2003). The node representing the revised Cambridge reference sequence (Andrews et al. 1999) is labelled by rCRS, and the root of the network is given by (truncated) haplotype #106. The nodes that constitute roots of (nested) major haplogroups are labelled accordingly (L3, N, pre-HV, U, K, JT, T, and T2). Substituting the three artificial types #*78*, *85*, and *98* (*underlined*) by the five correct haplotypes #78n, 85f, 85m, 98f, and 98m (*highlighted in small boldface*) results in the network with non-broken lines. The *bold lines* indicate a plausible (most parsimonious) tree for the corrected (truncated) data

of homoplasy (incurred by recurrent mutations) that warrants closer examination.

The subnetwork comprising all those haplotypes of this data set which are sampled at least twice is connected and constitutes a perfect tree. This tree has eight links corresponding to the mutations 11719 (characteristic of haplogroup pre-HV), 8592 (a novel mutation not observed before), 11251 (defining haplogroup JT), 8697 (defining haplogroup T), 11812 (defining haplogroup T2), 11467 (defining haplogroup U), 11332 (characteristic of one branch in haplogroup U4), and the mutation pair 9055, 11299 (defining haplogroup K); cf. Finnilä et al. (2001). Further, we can see that the truncated haplotype #107 matches a major haplotype from haplogroup K. When consulting the coding-region data from Herrnstadt et al. (2002), we find that the truncated haplotypes #93, 83, 59 are matched in that database. This, in particular, confirms the parallel mutations at site 11377. Truncated haplotype #106, which represents the root of the network, is found matched outside macro-haplogroup N. In fact, for sample #106 it was possible to read beyond the common reading frame and thus identify the two mutations 11914 and 11944, which point to the African haplogroup L2a. This leaves the truncated haplotypes #101, *98*, *78*, 77, *85*, 92, and 88 with private mutations for closer examination (haplotypes that turned out to be artificial are highlighted in italics). Since the mutations at sites 10873 (macro-haplogroup N), 11251 (haplogroup JT), and 11719 (haplogroup pre-HV, with only two potential back mutations) have not arisen on other haplogroup backgrounds, according to the data of Finnilä et al. (2001) and Herrnstadt et al. (2002), we must regard all those haplotypes (except possibly #88) as suspicious. Re-sequencing of samples #77, 92, 88, and 101 for the corresponding private variation confirmed the earlier findings. As to the latter haplotype #101, this is remarkable insofar as this is the first confirmed back mutation at site 10873, which is characteristic of macro-haplogroup N. But re-sequencing of samples #*78*, *85*, and *98* yielded surprising new results. The nucleotide at sample #*78* was simply incorrectly recorded earlier, so that the correct reading is A at 11719, yielding the new haplotype #78n. Haplotypes #*98* and *85* actually turned out to be mixed-up sequenced PCR-fragments of two couples! When, years ago, these two samples were originally collected from female staff members, mtDNA samples were also taken from their husbands and denoted with the same code number, only distinguished by the suffix ".1" for female and ".2" for male. Due to staff change later on, the information about this special meaning of the suffix got lost. For amplifying different fragments the template DNA originating from male and female were thus mixed up unintentionally. The PCRs were executed twice but prepared at the same time with the same template sample. The mistake was not detectable then since the short overlapping regions of the PCR fragments rarely show mutational events. When denoting the male and female donors of the corresponding samples by suffix m and f, respectively, the mosaic type #*98* arose as 98f×98f/m×98f/m×98f×98m×98m (listed in the order of the six fragments), and similarly type #*85* as 85f/m×85f/m×85m×85f×85m×85f/m. In sample #*98*, a clerical (base shift) error happened in addition to the couple mixture: at site 11151 a transition was

observed but erroneously recorded as 11161. The performed correction has a notable effect on the network representation: substituting the three artificial haplotypes by the five correct haplotypes reduces the reticulation in the corresponding median network considerably, thereby eliminating the two 3-dimensional cubes and three further squares (4-cycles).

Conclusion

The number of recombination instances that can be discovered in forensic databases is alarming (Bandelt et al. 2001): the risk for this kind of artefact seems to be omnipresent and can never be totally eliminated. The data reanalyzed in detail here demonstrate that sample mix-up can also involve mtDNAs that are not supposed to belong to the analyzed population sample. Even fragments of mtDNAs of different continental ancestry can slip into the data table. In order to discover such instances, the total worldwide mtDNA database (as taken from all publications in forensics, molecular anthropology, and medical genetics) needs to be taken into consideration. To avoid sample mix-up, it is recommended to perform two amplifications from each sample, albeit at different times so that one wrong sample extraction could not influence two sequencing reactions. Each amplification and sequencing round should cover all fragments to be screened (if possible), with the final result put in a table by an automated process with subsequent manual/visual control. The second round should end with a data table that is independently constructed. Then the final step is the comparison by computer of both tables. Even then one should not bypass a final comparison of the data table with the worldwide database in order to detect alarming idiosyncrasies that would warrant re-reading and re-sequencing.

The fact that even the publicly accessible portion of the FBI database, as included in the SWGDAM database, testifies to several obvious instances of sample mix-up is most worrying and may very well fuel attempts to question the admissibility of mtDNA in court in general. For instance, in the petitioner's brief on jurisdiction to The Supreme Court of Florida of this year (2003, Joseph Magaletti, petitioner, vs. State of Florida, respondent), one reads "... the opinion relied upon the assertion that the FBI database was accurate – there was no proof whatsoever that the database used in this case is accepted/used outside the FBI". Indeed, one of the "four Daubert factors" applied by the US federal court and several states in admissibility hearings is about "whether or not the technique has an acceptable rate of error, and whether or not there are standards that control the technique's operation" (Budowle et al. 2003).

For forensic casework there is thus an urgent need for more stringent laboratory routines that avoid sample mix-up ("Präparateverwechslung": Banaschak et al. 1999) and for a new generation of reliable forensic databases. The mix-up/contamination rate as documented in publications should at least be two orders of magnitude smaller than what one has to reckon with presently (~0.2–2.0% per amplified fragment). The EMPOP database project initiated by Parson et al. (2004) seems most promising in this respect.

Acknowledgements We thank Francesco Calì, Valentino Romano, and Peter Forster for sequence information and Claudio Bravi for database searches.

References

Alves-Silva J, Silva Santos M da, Guimarães PE, Ferreira AC, Bandelt H-J, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461, 775 (erratum)

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

Banaschak S, Witting C, Brinkmann B (1999) Präparateverwechslungen: Ursachen, Auswirkung, Prävention. Pathologe 20:155–158

Bandelt H-J (2004) Exploring reticulate patterns in DNA sequence data. In: Bakker FT, Chatrou LW, Gravendeel B, Pelser PB (eds) Plant species-level systematics: new perspectives on pattern & process. Regnum Vegetabile 142, Koeltz, Königstein, in press

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

Bandelt H-J, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. Int J Legal Med 115:64–69

Bandelt H-J, Herrnstadt C, Yao Y-G et al. (2003) Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. Ann Hum Genet 67:512–524

Budowle B, Allard MW, Fisher CL et al. (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. Int J Legal Med 116:212–215

Budowle B, Allard MW, Wilson MR, Chakraborty R (2003) Forensics and mitochondrial DNA: applications, debates, and foundations. Annu Rev Genomics Hum Genet 4:119–141

Cali F, Le Roux MG, D'Anna R et al. (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114:229–231

Comas D, Calafell F, Mateu E et al. (1998) Trading genes along the silk road: mtDNA sequences and the origin of Central Asian populations. Am J Hum Genet 63:1824–1838

Crespillo M, Luque JA, Paredes M, Fernández R, Ramírez E, Valverde JL (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. Int J Legal Med 114:130–132

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

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

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

Herrnstadt C, Elson JL, Fahy E et al. (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70:1152–1171, 71:448–449 (erratum)

Imaizumi K, Parsons TJ, Yoshino M, Holland MM (2002) A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. Int J Legal Med 116:68–73

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

Kivisild T, Tolk H-V, Parik J, Wang Y, Papiha SS, Bandelt H-J, Villems R (2002) The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751

Kong Q-P, Yao Y-G, Liu M et al. (2003a) Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. Hum Genet 113:391–405

Kong Q-P, Yao Y-G, Sun C, Bandelt H-J, Zhu C-L, Zhang Y-P (2003b) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676

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

Lee SD, Lee YS, Lee JB (2002) Polymorphism in the mitochondrial cytochrome B gene in Koreans. An additional marker for individual identification. Int J Legal Med 116:74–78

Lutz-Bonengel S, Schmidt U, Schmitt T, Pollak S (2003) Sequence polymorphisms within the human mitochondrial genes MTATP6, MTATP8 and MTND4. Int J Legal Med 117:133–142

Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Woźniak M, Miścicka-Śliwka D (2002) Mitochondrial DNA variability in Poles and Russians. Ann Hum Genet 66:261–283

Maruyama S, Minaguchi K, Saitou N (2003) Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. Int J Legal Med 117:218–225

Miller KWP, Budowle B (2001) A compendium of human mitochondrial DNA control region: development of an international standard forensic database. Croat Med J 42:315–327

Mishmar D, Ruiz-Pesini E, Golik P et al. (2003) Natural selection shaped regional mtDNA variation in humans. Proc Natl Acad Sci U S A 100:171–176

Monson KL, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. Forensic Sci Comm 4, #2

Parson W, Brandstätter A, Pircher M, Steinlechner M, Scheithauer R (2004) EMPOP – the EDNAP mtDNA population database-concept for a new generation, high-quality mtDNA database. Proceedings of the 20th International Congress of the International Society for Forensic Genetics, International Congress Series 1261, Forensic Genetics 10:106–108

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

Poetsch M, Wittig H, Krause D, Lignitz E (2003) Mitochondrial diversity of a northeast German population sample. Forensic Sci Int 137:125–132

Reidla M, Kivisild T, Metspalu E et al. (2003) Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73:1178–1190

Richards M, Macaulay V, Hickey E et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Salas A, Richards M, De la Fé T et al. (2002) The making of the African mtDNA landscape. Am J Hum Genet 71:1082–1111

Silva WA Jr, Bonatto SL, Holanda AJ et al. (2002) Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America. Am J Hum Genet 71:187–192

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114:224–228

Torroni A, Rengo C, Guida V et al. (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? Am J Hum Genet 69:1348–1356

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

Yao Y-G, Kong Q-P, Bandelt H-J, Kivisild T, Zhang Y-P (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Yao Y-G, Macaulay V, Kivisild T, Zhang Y-P, Bandelt H-J (2003) To trust or not to trust an idiosyncratic mitochondrial data set. Am J Hum Genet 72:1341–1346