

[TABLE OF CONTENTS](#)

[BACK ISSUES](#)

[SEARCH](#)

[EMPLOYMENT](#)

[MEETINGS AND CONFERENCES](#)

[FBI PUBLICATIONS](#)

[EDITORS](#)

[ABOUT FSC](#)

[SUBMITTING MANUSCRIPTS](#)

[HANDBOOK OF FORENSIC SERVICES](#)

[FBI LABORATORY](#)

Forensic Science Communications

October 2002 — Volume 4 — Number 4

Research and Technology

Further Discussion of the Consistent Treatment of Length Variants in the Human Mitochondrial DNA Control Region

Mark R. Wilson

Supervisory Special Agent
Counterterrorism and Forensic Science Research Unit
Federal Bureau of Investigation
Quantico, Virginia

Marc W. Allard

Associate Professor of Biology
Department of Biological Science
George Washington University
Washington, DC

Keith L. Monson

Research Chemist
Counterterrorism and Forensic Science Research Unit
Federal Bureau of Investigation
Quantico, Virginia

Kevin W. P. Miller

Biologist-Forensic Examiner
DNA Analysis Unit 2
Federal Bureau of Investigation
Washington, DC

Bruce Budowle

Senior Biological Sciences Program Advisor
Forensic Analysis Branch
Federal Bureau of Investigation
Washington, DC

[Introduction.....Examples of Alternative Alignments.....HV II C Stretch](#)

[Discussion.....References](#)

Introduction

Alignments are used when generating mtDNA sequence profiles for comparison purposes. An alignment is made between a sample of interest and a generally recognized reference, such as the Cambridge Reference Sequence (CRS) (Anderson et al. 1981; Andrews et al. 1999). In the majority of situations, the alignment and naming of differences from the reference is straightforward. However, the treatment of insertions and deletions (gaps) may vary, causing some laboratories to code mtDNA sequences differently (Bortolini et al. 1997; Ginther et al. 1993; Kolman et al. 1996; Ribeiro-Dos Santos et al. 1996; Salas et al. 2001). Several authors have already provided rules for nomenclature issues (Carracedo et al. 2000; Tully et al. 2001); therefore, this paper will expand on these ideas.

Wilson et al. (2002) have defined a number of situations that may have been problematic in this respect. This manuscript discusses examples of alternative alignments that were not included in the Wilson et al. (2002) paper because of space limitations.

The general recommendations are as follows:

1. Profiles should be characterized so that the least number of differences from the reference sequence are present.
2. If there is more than one way to maintain the same number of differences with respect to the reference sequence, differences should be prioritized as follows:
 - A. insertions/deletions (indels)
 - B. transitions (purine-to-purine or pyrimidine-to-pyrimidine changes)
 - C. transversions (purine-to-pyrimidine or pyrimidine-to-purine changes)
3. Because all genes have a 5' to 3' direction of transcription and mtDNA genes are encoded on both the heavy and light strands of the closed circular molecule, this paper explicitly states that insertions and deletions be placed 3' with respect to the light strand of human mtDNA. Insertions and deletions should be combined in situations where the same number of differences from the reference sequence is maintained.

Examples of Alternative Alignments

A number of examples have been identified where alternative alignment strategies result in slightly different characterizations of mtDNA profiles. Some of these examples are discussed below. The first line is the sequence to be compared with the CRS; the second line is the CRS. The nucleotide position is referenced in the space below the position. All of these alignment examples were obtained from an expanded version of the Scientific Working Group on DNA Analysis Methods (SWGDM) forensic mtDNA database (Budowle et al. 1999; Miller and Budowle 2001). A summary of the examples can be found in [Table 1](#). It includes the CRS positions of the sequence under examination, the example sequence, the corresponding CRS sequence, the recommended alignment, and the recorded nucleotide positions of the difference(s). A complete discussion of all the examples shown in Table 1 can be obtained from a combination of this manuscript and the Wilson et al. (2002) publication.

Example 1

Both length and sequence changes are observed in and around nucleotide positions 498 and 499 in the human mtDNA control region. One such change is a transition at nucleotide position 499. Example 1 shown in [Table 1](#) contains sequence information from nucleotide positions 488–504. A simple sequence difference is found between the example and the reference; hence no decision regarding alignment is needed. A one-base difference is found between the profile and the reference at nucleotide position 499. No insertions or deletions are present, thus there are no alternative alignments other than alignment 1. The difference from the reference is coded as 499A.

Example 2

A sequence similar to that found in Example 1 has been observed, however, a one-base pair deletion is found next to the transition. Thus, the alignment requires a decision as to where to place the gap between these sequences, as shown below.

ATACAACCCCACCCAT	
ATACAACCCCCGCCCAT CRS	
490	500

Alignment 2A results from a deletion and a transition and is recorded as 498D, 499A.

Alignment 2A

ATACAACCCC- ACCCAT	
ATACAACCCCCGCCCAT CRS	
490	500

However, another possible alignment is 2B.

Alignment 2B

ATACAACCCCA-CCCAT	
ATACAACCCCGCCCAT CRS	
490	500

Alignment 2B can be described as two changes: a transversion at nucleotide position 498 and a deletion at nucleotide position 499, and would be recorded as 498A, 499D.

According to the recommendations listed in Wilson et al. (2002), alignment 2A is preferred because Recommendation 2 states that transitions have priority over transversions.

The deletion placed at nucleotide position 498 in alignment 2A could have been placed in a number of different positions within a continuous run of cytosine residues. Each of these alternative alignments results in two differences between the profile and the reference. However, Recommendation 3 states that insertions and deletions should be placed 3' with respect to the light strand. Thus again, alignment 2A is recommended, due to the 3' placement of the gap compared to the other alternative alignments. Recommendations 2 and 3 both agree that alignment 2A is the preferred alignment, and the differences from the CRS should be recorded as 498D, 499A.

Example 3

An example found in the hypervariable region II is shown in Example 3. The sequences of the profile and the reference, from nucleotide positions 244–253, are shown below.

ATTGATGTC	
ATTGAATGTC CRS	
	250

Alignment 3A places a deletion at nucleotide position 248.

Alignment 3A

ATTG-ATGTC	
ATTGAATGTC CRS	
	250

However, the deleted base could also be placed at the adjacent A residue, as shown in alignment 3B.

Alignment 3B

ATTGA-TGTC	
ATTGAATGTC	CRS
	250

Because both alignments result in a single deletion, Recommendations 1 and 2 cannot resolve the choice of alignments, and Recommendation 3 is applied. A deletion should be placed at the 3' end with respect to the light strand in such cases. Hence, alignment 3B is preferred over alignment 3A. The difference is coded as 249D.

Example 5

A short dinucleotide repeat is found in the human mtDNA control region near the tRNA-Phenylalanine gene (Bodenteich et al. 1992). The CRS lists five AC repeats, but individuals have been identified who have as few as three or as many as seven copies of the repeat. A common observation in many populations is the presence of six copies of the repeat, as shown below. This example illustrates positions 508-529 in both the six-repeat sample and the CRS reference sequence.

ACCCAGCACACACACACACCGCTG	
ACCCAGCACACACACACCGCTG	CRS
510	520

Conforming to Recommendation 3, the inserted bases are listed at the 3' end of the repeat, as shown in alignment 5A.

Alignment 5A

ACCCAGCACACACACACACCGCTG	
ACCCAGCACACACACAC--CGCTG	CRS
510	520

This alignment results in the addition of two bases at nucleotide position 524. This profile is, therefore, coded as 524.1A, 524.2C.

However, designation of the repeat in this example may result in some inconsistency. If the 5' end is used to determine the beginning of the dinucleotide repeat, the repeat is classified as a CA repeat. In contrast, if the repeat is moved to the 3' end to maintain the same number of differences from the reference, it is classified as an AC repeat. Alignment 5B illustrates this alternative.

Alignment 5B

ACCCAGCACACACACACACCGCTG	
ACCCAGCACACACACA--CCGCTG CRS	
510	520

To be consistent with Recommendation 3, alignment 5A is preferred because the inserted bases are shifted one base in the 3' direction with respect to the CRS. The insertion is thereby classified as an AC insertion, and the differences from the CRS are listed as 524.1A, 524.2C.

Example 6

The recommended treatment of differences from profiles with fewer repeat units than the CRS is shown in Example 6. This example has three copies of the repeat, rather than the five copies found in the CRS.

ACCCAGCACACACCGCTG	
ACCCAGCACACACACACCGCTG CRS	
510	520

Alignment 6A places the deleted bases on the 3' end of the dinucleotide repeat. The deleted bases are, therefore, coded as 521D, 522D, 523D, and 524D.

Alignment 6A

ACCCAGCACACAC-----CGCTG	
ACCCAGCACACACACACCGCTG CRS	
510	520

Example 13

Generally, a total of 14 residues are found between nucleotide positions 16180 and 16193 (Bendall and Sykes 1995; Casteels et al. 1999). However, Example 13 illustrates a situation where this is not the case. In Example 13, the T residue is found at nucleotide position 16186 rather than nucleotide position 16189. Also, the total number of residues between 16180 and 16193 is one fewer than the usual 14. Nucleotide pairs 16180-16198 are shown below.

AAAACCTCCCCCATGCT	
AAAACCCCTCCCCATGCT CRS	
	16190

Alignment 13A is coded as 16186T and 16189D.

Alignment 13A

AAAACCTCC-CCCCATGCT
AAAACCCCCTCCCCATGCT CRS
16190

Another possible alignment is shown as alignment 13B.

Alignment 13B

AAAACC-TCCCCCATGCT
AAAACCCCCTCCCCATGCT CRS
16190

Alignment 13B results in three changes, two transitions and a deletion. A third possible alignment with three differences is shown as alignment 13C.

Alignment 13C

AAAACCTCCCCC-ATGCT
AAAACCCCCTCCCCATGCT CRS
16190

Alignment 13A is the preferred alignment because it has the fewest differences from the CRS.

Example 14

In some cases, the number of C residues preceding and following the T residue at 16189 differs from what is found in the CRS. Example 14 is one example of this observation.

AAACCCCCCTCCCCCATGCT
AAAACCCCCTCCCCATGCT CRS
16190

Rather than the typical five cytosine residues observed preceding the T at 16189, this profile contains seven C residues. In addition, five cytosine

residues follow the T rather than four. Also, there are three A residues preceding the run of Cs rather than four. One possible alignment of this sequence to the CRS is shown as alignment 14A.

Alignment 14A

AAACCCCCCTCCCCATGCT
AAAACCCCC-TCCCC-ATGCT CRS
16190

This alignment yields three differences when compared to the CRS, a transversion and two insertions, and is coded as 16183C, 16188.1C, 16193.1C.

Because the insertions can be placed at any position within the series of C residues, there are many possible alignments that result in a total of three differences from the CRS, all of which have one transversion and two insertions (not shown). Again, the use of Recommendation 3 would place the insertions at the 3' end with respect to the CRS. Therefore, alignment 14A is preferred.

Example 15

Length-related variants are often complicated and warrant careful consideration, as shown in Example 15. Positions 16178-16198 are shown in this example.

T'TAAACCCCCCTCCCATGCT
TCAAACCCCCCTCCCCATGCT CRS
16190

As expected, there are many different ways to align this sequence to the CRS. One possible alignment is shown as alignment 15A.

Alignment 15A

T'TAAACCCCCCTCCCATGCT
TCAAACCCCC-TCCCCATGCT CRS
16190

This alignment yields five total changes, three transitions, one transversion, and one insertion. Alignment 15B results in four changes and therefore, is preferred.

Alignment 15B

TTAAACCCCCCCTCCCATGCT
TCAAACCCCCTC-CCCATGCT CRS
16190

The coded variants from the reference are 16179T, 16183C, 16189C, 16190.1T.

Example 16

Some of the other length variants in this region may involve other combinations of A-C transversions and insertions. One variant is shown below.

AAACCCCCTCCCCCATGCT
AAAACCCCCTCCCCATGCT CRS
16190

Alignment 16A

AAACCCC-TCCCCCATGCT
AAAACCCCCTCCCC--ATGCT CRS
16190

A total of four changes result from alignment 16A, a transversion, a deletion, and two insertions. However, other alignments with three total changes are possible, as shown in alignments 16B and 16C.

Alignment 16B

AAACCCCCTCCCCCATGCT
AAAACCCC-CTCCCCATGCT CRS
16190

Alignment 16C

AAA-CCCCCTCCCCCATGCT
AAAACCCCCTCCCC--ATGCT CRS
16190

Alignment 16B results in one transversion, one transition, and one insertion. These three changes in alignment 16C are all indels. Thus, 16C is preferred over alignment 16B. Alignment 16C is coded as 16183D, 16193.1C, 16193.2C.

HV II C Stretch

The HV II region also contains a C stretch region similar to the HV I region; however, some important differences have been reported (Greenberg et al. 1983; Hauswirth and Clayton 1985; Stewart et al. 2001). Whereas the T residue at position 16189 in the HV I region is often observed to be absent, the T residue in the HV II region is less frequently absent. More often, the T residue found at nucleotide position 310 is shifted as a result of length variants directly upstream (i.e., in the 5' direction). The CRS, beginning at nucleotide position 300 and ending at nucleotide position 317, is shown below with the T at nucleotide position 310 underlined:

AAACCCCCCTTCCCCCGC

Length variants in this region are illustrated below.

AAACCCCCCTCCCCCGC (7 Cs upstream from 310T, CRS)

AAACCCCCCTCCCCCGC (8 Cs upstream from 310T)

AAACCCCCCTCCCCCGC (9 Cs upstream from 310T)

Example 18

An example of length variation in this region results in alternative ways to align the sequence to the CRS. One such example is Example 18, shown below.

AAACCCCCCTCCCCCGC
AAACCCCCCTCCCCCGC CRS
310

Two transitions observed in alignment 18A can explain the differences with respect to the CRS and are coded as 309T, 310C.

Alignment 18A

AAACCCCC T CCCCCGC
AAACCCCCCTCCCCCGC CRS
310

In contrast, alignment 18B, which results in a deletion and an insertion, still maintains the same number of differences.

Alignment 18B

AAACCCCCC-TCCCCCGC
AAACCCCCCTCCCC-GC CRS
310

Recommendation 2 states that insertions and deletions should take precedence over substitutions. Therefore, alignment 18B is the preferred alignment, and the differences from the CRS are 309D, 315.1C.

Example 19

In this example, two additional bases are present with respect to the CRS, both of which may be considered as occurring within homopolymeric regions.

AAACCCCCCTTCCCCCGCT
AAACCCCCCTCCCCGCT CRS
310

As expected, there are many ways to align the sample with the reference sequence. One possibility is shown as alignment 19A.

Alignment 19A

AAACCCCCCTTCCCCCGCT
AAACCCCCC-T-CCCCGCT CRS
310

Alignment 19A results in a T insertion at nucleotide position 309 and a C insertion at position 310. Hence, it would be recorded as 309.1T, 310.1C. However, both insertions fall within homopolymeric regions. In this case, the sample has two Ts followed by six Cs. In the case of the extra C residue, it could be placed in a number of different positions within the homopolymeric region while maintaining the same number of differences to the CRS.

Because many different options exist, Recommendation 3 applies, and the insertion is placed at the 3' end of both homopolymeric regions as shown in alignment 19B. The differences are coded as 310.1T, 315.1C and are shown in alignment 19B.

Alignment 19B

AAACCCCCCTTCCCCCGCT

AAACCCCCCT-CCCC-GCT	CRS
---------------------	-----

310

Discussion

The recommendations and examples provided in this paper are offered in an effort to standardize the treatment of length variants in human mtDNA within the forensic community. It could be suggested that biological mechanisms should underlie any method of coding differences to a reference sequence. However, these mechanisms may be complex and may be explained differently by investigators who may argue that there are alternative biological processes. Thus, issues of inconsistency may still persist. It could also be suggested that different rules be applied to different regions of the mtDNA molecule. However, this approach may also result in discrepancies as consistency in defining the boundaries of the regions becomes an issue.

The current method of recording differences from a reference is preferred and should be continued because it facilitates communication. However, for database searches, an alternative approach would be to file the entire sequence of nucleotides in a database, then query a long string of bases rather than a set of differences from a reference. Such an alternative to the current method might be explored in an effort to avoid inconsistencies caused by optional alignments when applied to forensic applications.

Some investigators may disagree with these proposed rules, but it is important to adopt a set of rules for consistency. These rules as described herein may be accepted, or other proposed approaches may be considered. At least the issues are raised, and discussion can begin.

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, I. C., Eperon, I. C., Nierlick, D. P., Roe, B. A., Sanger, F., Schreier, P. M., Smith, A. J. H., Staden, R., and Young, I. G. Sequence and organization of the mitochondrial genome, *Nature* (1981) 290:457-465.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowers, R. N., Turnbull, D. M., and Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature Genetics* (1999) 23:147.

Bendall, K. F. and Sykes, B. C. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region, *American Journal of Human Genetics* (1995) 57:248-256.

Bodenteich, A., Mitchell, L. G., Polymeropolous, M. M., and Merrill, C. R. Dinucleotide repeat in the human mitochondrial D-loop, *Human Molecular Genetics* (1992) 1:140.

Bortolini, M. C., Zago, M. A., Salzano, F. M., Silva-Junior, W. A., Bonatto, S. L., da Silva, M. C., and Weimer, T. A. Evolutionary and anthropological implications of mitochondrial DNA variation in African Brazilian populations. In: *Human Biology, An International Record of Research*. Wayne State University Press, Detroit, Michigan, 1997, vol. 69, no. 2, pp. 141-159.

Budowle, B., Wilson, M. R., DiZinno, J. A., Stauffer, C., Fasano, M. A., and Holland, M. M. Mitochondrial DNA regions HVI and HVII population data, *Forensic Science International* (1999) 103:23-35.

Carracedo, A., Bar, W., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G., and Wilson, M. DNA commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing, *Forensic Science International* (2000) 110:79-85.

Casteels, K., Ong, K., Phillips, D., Bendall, H., Pembrey, M., Poulton, J., and Dunger, D. Mitochondrial 16189 variant, thinness at birth, and type-2 diabetes, *Lancet* (1999) 353:1499-1500.

Ginther, C., Corach, D., Penacino, G. A., Rey, J. A., Carnese, F. R., Hutz, M. M., Anderson, A., Just, J., Salzano, F. M., and King, M. C. Genetic variation among the Mapuche Indians from the Patagonian region of Argentina: Mitochondrial DNA sequence variation and allele frequencies of several nuclear genes. In: *DNA Fingerprinting: State of the Science*. S. D. J. Pena, R. Chakraborty, J. T. Eppelen, and A. J. Jeffreys, eds. Birkhauser Verlag, Basel, Switzerland, 1993, pp. 211-219.

Greenberg, B. D., Newbold, J. E., and Sugino, A. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA, *Gene* (1983) 21:33-49.

Hauswirth, W. W. and Clayton, D. A. Length heterogeneity of a conserved displacement loop sequence in human mitochondrial DNA, *Nucleic Acids Research* (1985) 13:8093-8104.

Kolman, C. J., Sambuughin, N., and Bermingham, F. Mitochondrial DNA analysis of Mongolian populations and implications for the origin of new world founders, *Genetics* (1996) 142:1321-1334.

Miller, K. W. P. and Budowle, B. A compendium of human mitochondrial DNA control region: Development of an international standard forensic database, *Croatian Medical Journal* (2001) 42(3):315-327.

Ribeiro-Dos Santos, A. K. C., Santos, S. E. B., Machado, A. L., Guapindaia, V., and Zago, M. A. Heterogeneity of mitochondrial DNA haplotypes in pre-Columbian natives of the Amazon region, *American Journal of Physical Anthropology* (1996) 101:29-37.

Salas, A., Lareu, M. V., and Carracedo, A. Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: A case report, *International Journal of Legal Medicine* (2001) 114:186-190.

Stewart, J. E. B., Fisher, C. L., Aagaard, P. J., Wilson, M. R., Isenberg, A. R., Polanskey, D., Pokorak, E., DiZinno, J. A., and Budowle, B. Length variation in HV2 of the human mitochondrial DNA control region, *Journal of Forensic Sciences* (2001) 46(4):862-870.

Tully, G., Bar, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Parson, W., and Schneider, P. Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretations of mitochondrial DNA profiles, *Forensic Science International* (2001) 124:83-91.

Wilson, M. R., Allard, M. W., Monson, K. L., Miller, K. W. P., and Budowle, B. Recommendations for consistent treatment of length variants in the human mtDNA control region, *Forensic Science International* (2002) Volume 129/1: 35-42.

[Top of the page](#)

[TABLE OF CONTENTS](#) — [BACK ISSUES](#) — [SEARCH](#) — [EMPLOYMENT](#) — [MEETINGS AND CONFERENCES](#)

[FBI PUBLICATIONS](#) — [EDITORS](#) — [ABOUT FSC](#) — [SUBMITTING MANUSCRIPTS](#)

[HANDBOOK OF FORENSIC SERVICES](#) — [FBI LABORATORY](#)