

Automated alignment and nomenclature for consistent treatment of polymorphisms in the human mitochondrial DNA control region

Bruce Budowle, Ph.D.¹, Deborah Polanskey¹, Constance L Fisher, Ph.D.¹

Bobi K Den Hartog, Ph.D.², Russell B. Kepler², John W. Elling, Ph.D.²,

Corresponding Author:

Deborah Polanskey

Tel: 703-632-7585

Fax: 703-632-7573

Email: deborah.polanskey@ic.fbi.gov

¹ FBI Laboratory, 2501 Investigation Parkway, Quantico, VA.

² Mitotech, 147 Gonzales Road #4, Santa Fe NM, www.mitotech.com

1 Abstract

Naming mtDNA sequences by listing only those sites that differ from a reference sequence is the standard practice for describing the observed variations. The operational alignment and nomenclature rules, i.e., “Wilson Rules,” suggested for this purpose do not guarantee a single consistent sequence description for all observed polymorphisms. In this work the operational alignment/nomenclature rules were reconfigured to better reflect traditional user preferences, and a computer-facilitated method of aligning mtDNA sample sequences with a reference sequence was developed. A software package named Mitotyper™ was created to implement the nomenclature rules that allows for rapid sequence alignment, provides for absolute stability and consistency, and provides increased database accuracy at all regions of an mtDNA sequence.

Keywords: mtDNA, nomenclature, hierarchal rules, parsimony, software, Mitotyper Rules™, phylogenetics, validation

2 Introduction

Mitochondrial DNA (mtDNA) sequencing can be used effectively to analyze highly degraded or limited quantity forensic samples, particularly when nuclear DNA is not typeable [1]. The analysis of the non-coding control region or only the hypervariable regions – HV1 and HV2 - within the non-coding control region results in a string of approximately eleven hundred or six hundred bases, respectively (or less if the sample is severely compromised), that are sequenced. Displaying and conveying so many bases is cumbersome and impractical. Instead only those limited number of bases that differ from a published reference sequence (the revised Cambridge Reference Sequence, or rCRS [2, 3]) are described. Naming mtDNA sequences by listing only those sites that differ from the rCRS provides a common language and a simple operational tool for describing the variation observed. Identical or concordant sequences are identified by exhibiting the same differences from the rCRS.

This nomenclature concept of comparison to the rCRS has been well accepted but has a complication that needs to be addressed. This complication is the possibility of multiple rule-compliant alignments occurring against the rCRS for some sequences. A practical alignment scheme for naming the variants contained within the mtDNA sequences needs to be defined so practitioners can consistently select the same alignment option for a particular sequence. Such practice will provide nomenclature stability within and among forensic laboratories [4, 5]. Wilson et al [6, 7] proposed an alignment and nomenclature protocol (the “Wilson Rules”) to attempt to standardize the description of differences from the reference sequence. While the concept of the rules proffered by Wilson et al was generally accepted, the rules themselves were not strictly practiced for several reasons. First, the rule favoring indels over substitutions did not reflect some of the historically preferred alignments. [Note: we use historical to mean an alignment that was found in databases prior to implementation of any nomenclature rules]. Second, the nomenclature process was performed manually, which led to situations where the same sequence was named differently by different practitioners [8, 9, 10]. Third, the rules as described did not always guarantee a single consistent sequence description at all

observed polymorphic regions within an mtDNA sequence. Therefore, the alignment/nomenclature rules were reconfigured, and a computer-facilitated method of aligning mtDNA sample sequences with a reference sequence was developed. A software package named MitotyperTM was created to implement the nomenclature rules that allows for rapid sequence alignment, provides for absolute stability and consistency, and provides increased database accuracy.

In this paper, a hierarchical set of rules called the Mitotyper RulesTM are described. They were developed for greater consistency with both historical and current operational nomenclature. Additional rules were added to stabilize the operational nomenclature and to select among those few situations where two or more alignments are still possible after the modified operational rules have been executed. With this approach, absolute consistency and stability in nomenclature is attained. Those haplotypes, once named and entered into the database never need to be renamed; identical sequences will always be listed the same way and database searching for statistical inferences will be greatly facilitated.

2.1 Software Operation and Alignment Strategy

The mtDNA type generation software first roughly aligns the reference rCRS and the sample sequence in order to identify large regions in which the sequences “match” one another. The Smith-Waterman-Gotoh algorithm is used for this rough alignment [11]. Regions between matching regions are then, by our definition, polymorphic regions within which the sample sequence needs to be defined in relation to the rCRS. Polymorphic regions will contain one or a set of closely spaced sequence variants, the placement of which interact with one another according to the alignment and nomenclature rules.

The boundaries of the polymorphic regions are adjusted to include contextual information about neighboring bases for use when aligning and typing the region. This adjustment can include expansion or subdivision. Expansion can take place, for example, when maintaining a repeat motif in a contiguous fashion. Subdivision occurs, for example, around the HV2 C-stretch area in order to ensure alignment between specific landmarks of the polymorphic region.

In each polymorphic region, the set of every possible parsimonious alignment between the reference and the sample sequence is created. This set of possible alignments for each region is then processed by applying each rule in series to select the compliant alignments in the set. The standard execution orders of the rules are shown in Figure 1 and the additional modified procedure used for aligning to the rCRS 300 to 315 region is shown in Figure 2.

The software uses each rule to discard all of the alignments in the region’s set of possible alignments that do not comply with the rule. If only one alignment remains after discarding the noncompliant alignments, that alignment is used to create the type description of the polymorphism(s) in that region. However after running the rule if more than one valid alignment still remains in the set of alignments for that region, the software uses the next rule in the series for processing the set.

2.2 Typing Rules and Procedures

The software applies the rules to analyze sets of candidate alignments in the hierarchy described in Figures 1 and 2.

2.2.1 Rule 1 – Least Number of Differences

The general overarching principle for the nomenclature rules is to select an alignment with the least number of differences between the rCRS and the sample sequence; that is, the most parsimonious alignment [6, 7]. For each polymorphic region, the software only generates the set of all possible alignments of the sample sequence in the region that have the minimum number of differences from the rCRS.

An ambiguous or heteroplasmically equivalent substitution in the sample consensus sequence is not considered a difference from the rCRS. The International Union of

Pure and Applied Chemistry (IUPAC) nomenclature is used to determine heteroplasmic equivalence. For example, an A is considered equivalent to an R, W, M, D, H, V, or N but not a Y [12]. An inserted ambiguous base in the sample sequence is counted as a difference from the rCRS.

2.2.2 Rule 2 – Maintain the AC Repeat Motif

The rCRS HVIII region includes an AC repeat motif in the 515 to 525 range [13, 14]. If the polymorphic region includes the AC repeats, the software selects alignments that both place insertions or deletions (indels) at the 3' end of the region and preserve the “AC” motif from those that place indels in less preferred positions. Table 1 contains an example of a historical alignment and a MitoTyper alignment in which the two deletions are placed together to reflect the deletion of one AC motif in the sample sequence, generating a preferred type of 523 A -, 524 C – in this region.

Table 1. Maintain AC Motif

	Historical	Mitotyper Rules (Preferred)
rCRS 512 to 527	AGCACACACACACCGC	AGCACACACACACCGC
Sample consensus	AGCACACACAC-C-GC	AGCACACACAC--CGC

2.2.3 Rule 3 – Prefer Substitutions to Indels

Polymorphisms between sequences can be aligned by describing the difference in bases as a substitution(s) or an indel(s). A substitution occurs when sequences are aligned so that one base is replaced with another to describe the polymorphism. Indels can be used to describe the same polymorphism, by deleting the mismatched base and inserting the sequence base that is observed, or some variation thereof.

This rule dictates a preference for alignments with substitutions rather than indels. The software chooses between candidate alignments by selecting the alignment with the most substitutions.

Table 2 shows an example of a preference for substitutions over indels to describe a polymorphism and contrasts that to a historical alignment found in the Scientific Working Group on DNA Analysis Methods (*SWGAM*) database. This historical alignment was described by three indels while the Mitotyper alignment is achieved with two substitutions and one indel: 16183 A C 16187.1 – T, and 16189 T C

Table 2. Preference for Substitutions to Indels

	Historical	Mitotyper Rules (Preferred)
rCRS 16178 to 16193	TCAAAAACCCCTCCCC--	TCAAAAACCCC-CTCCCC
Sample consensus	TCAAA-CCCCCTCCCCC	TCAAACCCCTCCCCC

2.2.4 Rule 4 - Prefer Transitions to Transversions

Nucleotide substitutions can be classified as transitions or transversions. Generally, transitions are more biologically feasible because the substituted nucleotides belong to the same chemical group, purine or pyrimidine. In contrast, transversions involve nucleotide substitution between the purine and pyrimidine groups. This rule selects those alignments in which polymorphisms are described by transitions. Table 3 displays a historical alignment (that appears to favor transversions over transitions) in which the polymorphic region was described as an indel 56.1 - C, a transversion, 57 T G and a transition 73 A G. The software instead selected an alignment that described the polymorphisms with an indel and two transitions 57 T C, 57.1 - G, and 73 A G.

Table 3. Preference for Transitions to Transversions

	Historical	Mitotyper Rules (Preferred)
rCRS 55 - 75	TA-TTTTCGTCTGGGGGGTATG	TAT-TTTTCGTCTGGGGGGTATG
Sample consensus	TACGTTTTCGTCTGGGGGGTGTG	TACGTTTTCGTCTGGGGGGTGTG

2.2.5 Rule 5 - Place Indels Contiguously When Possible

Indels, at times, can be placed in several locations in order to align polymorphic sequences. Placement of indels at multiple locations is particularly possible in homopolymeric regions. The preferred alignment for nomenclature consistency places the indels in continuous groups as much as is possible. This rule is implemented by selecting between two candidate alignments the one with the fewest indel groups (i.e., groups of one or more contiguous indels). Table 4 shows an example of a historical type and a preferred type identified by the software where the same three indels can be placed contiguous to one another as 455.1 - T, 455.2 - T, and 455.3 - C.

Table 4. Example of the Place Indels Contiguously When Possible Rule

	Historical	Mitotyper Rules (Preferred)
rCRS 451 to 460	ATTTT--CCCC-T	ATTTT---CCCCT
Sample consensus	ATTTTTTCCCCT	ATTTTTTCCCCT

2.2.6 Rule 6 - Place Indels 3' to Homopolymeric Stretches

The Wilson Rules describe basic alignment procedures for length variants such as placing indels and gaps at the 3' end of the region when variation occurs in homopolymeric regions [6, 7]. This preference also is implemented in the Mitotyper software by selecting alignments that preserve homopolymeric stretches of base sequence and place the indels at the 3' end of the homopolymeric stretch. Table 5 and Table 6 illustrate examples of two historic alignments found in the SWGDAM data in which indels were not placed 3' to homopolymeric stretches (i.e., incorrectly placed) along with the preferred alignments that comply with this rule. In the historical example in Table 5, the indel is at the wrong end of the poly-A stretch (i.e. the 5'

end). In the example in Table 6, the preferred alignment creates a two-T homopolymeric stretch.

Table 5. Placing Indels 3' to a Homopolymeric Stretch, Example 1

	Historical	Mitotyper Rules (Preferred)	Preferred Type
rCRS 16170 to 16183	AATCCACATCAAAA	AATCCACATCAAAA	
Sample consensus	AACCCACATC-AAA	AACCCACATCAAA-	16172 T C 16183 A -

Table 6. Placing Indels 3' to a Homopolymeric Stretch, Example 2

	Historical	Mitotyper Rules (Preferred)	Preferred Type
rCRS 111 to 114	CCCT	CCCT	
Sample consensus	CT-T	C-TT	113 C - 114 C T

2.2.7 Rule 7 - Three Prime Preference Tiebreaker

It is possible for the set of alignments for each region to have been processed by all of the previous rules and still have multiple valid candidate alignments remaining. Rule 7 implements the general 3' preference [15] as a tiebreaker to select the single preferred alignment for each region. For a set of alignments to reach this point, each alignment will have the same number of indel groups but not have both insertions and deletions in exactly the same positions in any two alignments.

The Mitotyper software first selects between two candidate alignments the one with the insertion or insertions at the 3' end of the sequence. If a tie remains because two or more alignments have an insertion at the same 3'-most position, the software selects between these by preferring the alignment that has the 3' most deletion.

Table 7 shows two alignments of a sample sequence that are both compliant with all previous rules. The insertion tiebreaker rule leads to the selection of the alignment with an insertion at 16193 over the alternative alignment that placed an insertion at position 16188.

Table 7. Example of the Tiebreaker Rule using Insertions

	Alignment 1	Alignment 2 (Preferred)
rCRS 16180 to 16194	AAAACCCCC-TCCCC	AAAACCCCCCTCCCC-
Sample consensus	AAACCCCCCCCCCCC	AAACCCCCCCCCCCC

Table 8 displays another example of two alignments that are both compliant with all previous rules. The Rule 7 Tiebreaker leads to the selection of the alignment with a deletion at 16189 over the historical alignment that placed the deletion at position 16188.

Table 8. Example of the Tiebreaker Rule using Deletions

	Historical	Mitotyper
--	------------	-----------

		(Preferred)
rCRS 16180 to 16194	AAAACCCCCCTCCCCA	AAAACCCCCCTCCCCA
Sample consensus	AAAACCCC-ACCCA	AAAACCCCA-CCCA

2.2.8 HV2 C-stretch (HV2CS) Region Typing

The five cytosine residues in the rCRS in the HV2 region from positions 311 to 315 represent a rare configuration from the majority of the population that has six cytosines in this C-stretch region [2]. As a result, almost 99% of the sequences in the SWGDAM database (4834 of the 4839 sequences tested) have a 315.1 C type in this region.

Analysis of the SWGDAM and European mtDNA Population (EMPOP) databases reveals a preference for aligning the thymine at the 310 position (the 310T) in the rCRS with a thymine residue in the sample sequence and preserving the 315.1 C type description for polymorphisms in this region. Table 9 shows the sequence found in four SWGDAM samples (CHN.ASN.000091, CHN.ASN.000451, THA.ASN.000021, and USA.335.000132) and two EMPOP samples (GRC0500109 and GRC0500120) and their corresponding historical type in this region. This sequence is described with three indels, 308 C -, 309 C -, and 315.1 - C, which preserves the 310 T alignment. However application of the Mitotyper Rules that follow the hierarchy shown in Figure 1 would produce an alternate non preferred alignment which would have two differences from the rCRS (“308 C T” and “310 T -”), thus meeting the “least number of differences” requirement.

Table 9. Historical preference for aligning the 310T in the rCRS 300 to 315 region.

	Historical and Mitotyper HV2CS Results (Preferred)	Mitotyper Standard Rules
rCRS 295 to 317	CCACCAAACCCCCCTCCCC-GC	CCACCAAACCCCCCTCCCCCG
Sample consensus	CCACCAAACCCC--TCCCCCGC	CCACCAAACCCCTC-CCCCCGC

Analysis of the historical data also reveals a preference for aligning the 300 to 302 poly-A region and the 303 to 309 poly- C region in the rCRS with matching regions in the sample sequence, also at the expense of a strict Rule 1 compliance. Table 10 shows the sequence from CHN.ASN.000451 with the preferred historical alignment (299 C -, 309.1 - C, 315.1 - C) type and the non preferred alignment produced by application of the Mitotyper Standard Rules in Figure 1.

Table 10. Historical preference for aligning the poly-A and poly-C regions in the rCRS

	Historical and Mitotyper HV2CS Results (Preferred)	Mitotyper Standard Rules
rCRS 295 to 317	CCACCAAACCCCCCTCCCC-GC	CCACCAAACCCCCCTCCCC-GC
Sample consensus	CCAC-AAACCCCCCTCCCCCGC	CCACCAAACCCCCCTCCCCCGC

In order to correspond with historical preferences, typing polymorphisms in the HV2 300 to 315 region is carried out with a slightly different procedure shown in Figure 2.

The HV2CS typing protocol begins by isolating this region from the rest of the sequence and generating a set of alignments between the sample and the reference sequence in the region that aligns the rCRS 310T with the candidate thymine(s) in the sample sequence (if it exists). It also aligns the 5' 300 and 3' 302 positions of the AAA subregion and the ending 3' 315 boundary in the reference sequence with the corresponding candidate subregions in the sample sequence. In the case where there are ambiguous bases on the boundary, the software creates several alternatives that group the ambiguous bases in one subregion or the other.

With this procedure, each alignment in the set for this region will reflect the traditional practice of aligning T residues and the boundaries of the poly-A and poly-C subregions. This set of alignments is then processed by sequential application of the typing rules applied in the order illustrated in Figure 2. As before, each rule is responsible for discarding all of the alignments in the regions' set of possible alignments that do not comply with the rule.

When typing the HV2CS, the order in which the rules “place indels contiguously” and “place indels 3' to homopolymeric stretches” are applied is reversed from the standard Mitotyper rules (Figure 1). This change preserves the preferred placement of the sixth C after the 310T as 315.1C in the event that there is a second insertion in the sample in the 310 to 315 subregion of the HV2CS. Table 11 illustrates an example of the effect of this change. In the standard prioritization, “Place Indels Contiguously When Possible,” first discards any alignments in which the desired 315.1C indel is separate from another indel in this subregion. Instead, prioritizing the rule “Place Indels 3' to Homopolymeric Stretches” first in the HV2CS typing procedure results in the selection of the historically-preferred alignment of the 315.1C insertion at the end of the poly-C region with the additional 310.1 – T type.

Table 11 Historical preference for aligning the poly-A and poly-C regions in the rCRS

	Mitotyper HV2CS Rules (Preferred)	Mitotyper Standard Rules
rCRS 300 to 317	AAACCCCCCT-CCCC-GC	AAACCCCCCT--CCCCGC
Sample consensus	AAACCCCCCTTCCCCCGC	AAACCCCCCTTCCCCCGC

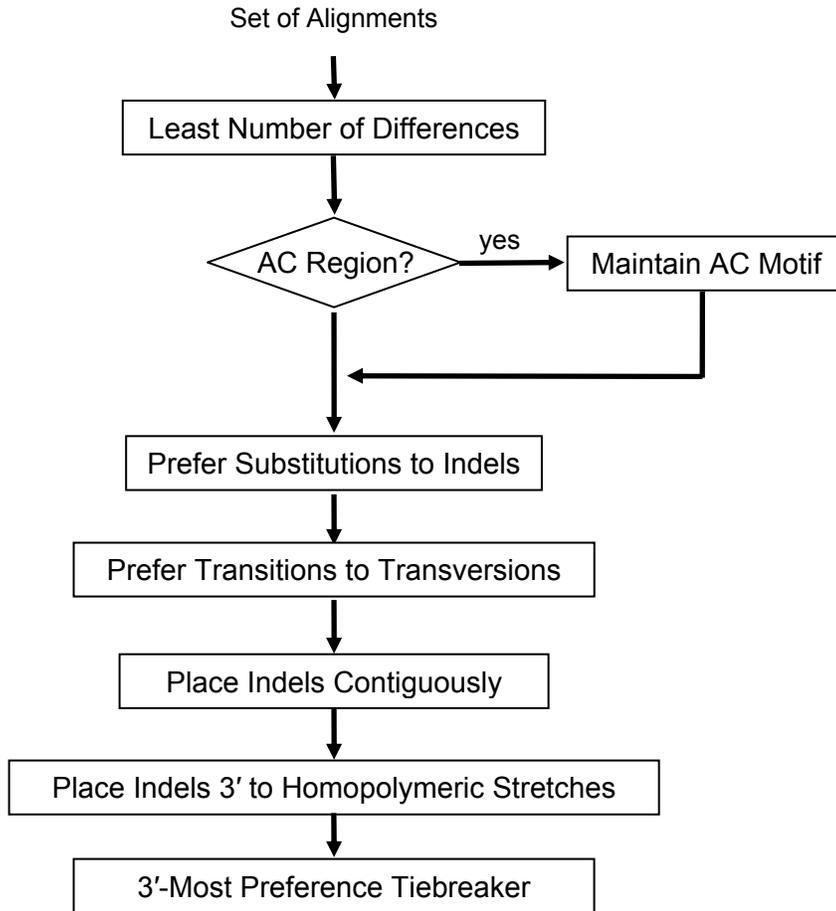


Figure 1. Standard application of the rules to the non-HV2CS regions containing polymorphisms

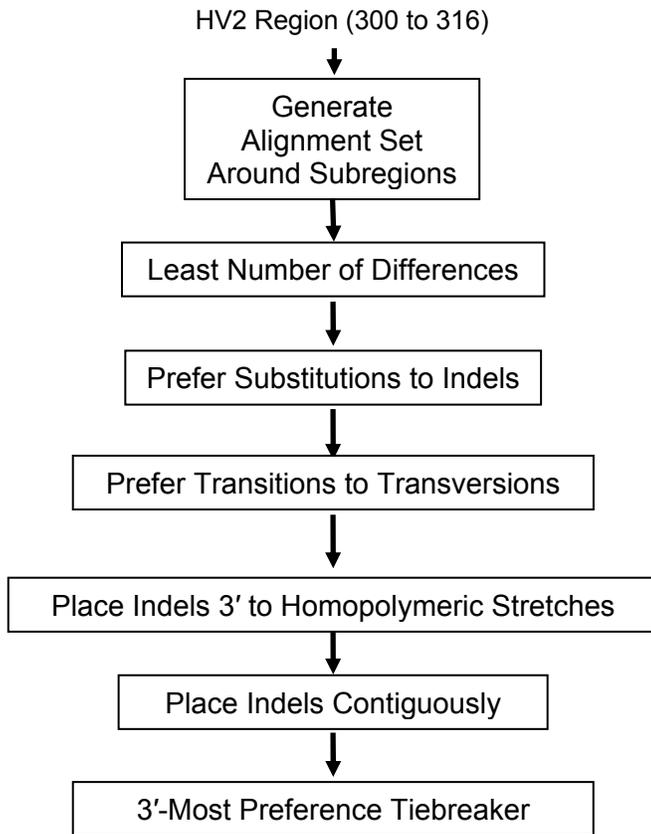


Figure 2. The HV2CS Typing Protocol, used for alignments in the 300 to 315 region, that anchors a T at position 310, if possible, and anchors the boundaries of subregions.

3 Results

The Mitotyper Software™ was produced to automate the alignment and nomenclature described by the rules herein for selecting consistent alignments of mtDNA sequences with the rCRS and generating type descriptions of polymorphisms for forensic and statistical use. The Mitotyper Rules were derived from the principle of a hierarchical operational alignment approach proposed by Wilson et al [6,7]. This development process resulted in nomenclature rules that are familiar to the forensic analysts and maximally congruent with historic practices.

3.1 *Testing and Validation*

The SWGDAM mtDNA population database was used as the historical reference data set when developing the Mitotyper Rules [16]. In the development process, these historical types were compared with the Mitotyper rule-compliant type generated for each sample.

In the 4839 SWGDAM sequences, the software identified 40,357 polymorphic regions that contained one or more differences from the rCRS. After typing with the software, only 33 of the 40,357 regions in 33 samples received types that did not agree with the historical types – a notably high 99.92% consistency considering that a manual nomenclature typing procedure was used to generate the database.

Of the 40,357 polymorphic regions, 32,915 had only one alignment between the sample and the rCRS with the minimum number of differences and so the type was determined solely by the Rule 1 parsimony requirement. For the remaining 7442 polymorphic regions (distributed across 4802 samples) the software generated a set of candidate alignments all with the same minimum number of differences which needed to be processed by subsequent rules.

An additional 1254 samples served as an independent validation data set. The typing software found 11,303 polymorphisms in this set of haplotypes; only 20 different sequences (with 56 occurrences) were discordant with the Mitotyper software types. In 11 of the 20 cases, the software produced a more parsimonious type. Manual examination of the 56 discrepancies by qualified analysts resolved these differences in favor of the software type. The low 0.18% discrepancy rate, similar to the rate achieved in the SWGDAM data set, indicates that the Mitotyper rules were not over fit to the SWGDAM data.

The EMPOP [17] version 1.0 database contains 5173 mtDNA control region haplotypes and was used as another independent validation data set. In total, the Mitotyper software found 103,460 polymorphic regions in the 5173 samples. Fifty two regions occurring in 24 sequence types differed between Mitotyper and EMPOP. Four examples were found in the EMPOP data in which the same polymorphic region was typed differently in different samples [18].

Mitochondrial DNA sequences of HV1 and HV2 of 1204 unrelated Japanese individuals were used as an additional independent validation data set [19]. The Mitotyper software found 8859 polymorphic regions in the 741 different haplotypes in the data set. Five polymorphic subsequences that occurred in 11 samples were typed differently between the historical type and the Mitotyper result. The 99.88% congruence illustrates the utility of the Mitotyper rules and automated software to support the currently-practiced nomenclature for describing mtDNA sequence polymorphisms.

3.2 Software Testing – Robustness

The robustness and reliability of the Mitotyper software was tested on over six million computer-generated sample sequences. Each test sample sequence was generated by introducing a random number of variants (between 1 and 30) to the reference sequence control region from nucleotide positions 16025 to 525. The modification location was random with an equal distribution across the sequence, and the modification was randomly chosen with an 80% chance of being a substitution and 10% chance each of being an insertion or a deletion. An additional heteroplasmic or ambiguous modification was inserted at a random location by replacing the base with an IUPAC symbol in the set "K", "M", "R", "D", "S", "H", "W", "V", "Y", "B", or "N". In every case the type generation software returned a type description that accurately described the computer-generated polymorphisms that had been randomly generated.

4 Conclusions

The Mitotyper software encodes a set of formalized hierarchical rules to describe sequence differences in relation to the rCRS. These rules are stable, well-defined and maximally consistent with both historical and current operational nomenclature. Implementing these hierarchical rules using robust software removes human involvement in selecting alignments, enhances consistency, and offers substantial savings in time and effort to describe mtDNA sequence polymorphisms.

The 33 differences between the software results and the historical types that were observed in the 4839 samples in the public SWGDAM database are resolved in favor of the software type. This updated version of the SWGDAM database will be available at <http://www.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm>. All future sample types in SWGDAM, including the additional 1254 samples from AFDIL used in the validation study, will be compliant with the Mitotyper rules.

It is important for the forensic community to adopt a strategy that resolves ambiguities and selects consistently a particular alignment for employing mtDNA nomenclature. These recommendations herein stabilize nomenclature, provide for higher quality databases, and enable better, more reliable profile search results for evaluating the weight of evidence. Conceivably, as more samples are typed, a novel polymorphism may require additional rules; these can be added to the software tool, but previously typed mtDNA profiles will not have to be renamed [15].

Acknowledgements

This is publication number 08-17 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

We would like to thank K. Sekiguchi, K. Imaizumi, K. Fujii, N. Mizuno, Y. Ogawa, T. Akutsu, H. Nakahara, T. Kitayama, and K. Kasai of National Research Institute of Police Science, Kashiwa-shi, Chiba, Japan, for kindly providing their population sequence data for our studies. We would also like to thank Patricia Aagaard, Dr. Susan Cropp and Dr. Alice Isenberg for their editorial contributions.

-
1. B. Budowle, M.R. Wilson, J.A. DiZinno, C. Stauffer, M.A. Fasano, M.M. Holland, K.L. Monson, Mitochondrial DNA regions HVI and HVII population data, *Forensic Sci. Int.* 103 (1999) 23–35.
 2. R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.
 3. S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, *Nature* 290 (1981) 457-465.
 4. K. Bender, P.M. Schneider, C. Rittner, Application of MtDNA sequence analysis in forensic casework for the identification of human remains, *Forensic Sci. Int.* 113 (2000) 103-107.
 5. R. Szibor, M. Michael, I. Plate, D. Krause, Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces, *Forensic Sci. Int.* 113(2000) 71-78.
 6. M.R. Wilson, M.W. Allard, K.L. Monson, K.W.P. Miller, B. Budowle, Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region, *Forensic Sci. Int.* 129 (2002) 35-42.
 7. M.R. Wilson, M.W. Allard, K.L. Monson, K.W.P. Miller, B. Budowle, Further Discussion of the consistent treatment of length variants in the human mitochondrial DNA control region, *Forensic Sci. Comm.* October; 4(4) (2002) <http://www.fbi.gov/hq/lab/fsc/current/index.htm>.
 8. A. Carrecedo, E. D'Aloja, B. Dupuy, A. Jangblad, M. Karjalainen, C. Lambert, W. Parson, H. Pfeiffer, H. Pfitzinger, M. Sabatier, D. Syndercombe Court, C. Vide,

Reproducibility of mtDNA analysis between laboratories: a report of the European DNA Profiling Group (EDNAP), *Forensic Sci. Int.* 97(1998) 165-170.

9. Y.G. Yao, C.M. Bravi, H.-J. Bandelt, A call for mtDNA data quality control in forensic science, *Forensic Sci. Int.* 141(2004) 1-6.
10. B. Budowle, D. Polanskey, M.W. Allard, R. Chakraborty, Addressing the use of phylogenetics for identification of sequences in error in the SWGDAM mitochondrial DNA database. *J. Forensic Sci.* 49 (2004) 1256-1261.
11. A. Moustafa, *JAligner: Open source Java implementation of Smith-Waterman*, <http://jaligner.sourceforge.net>
12. <http://www.bioinformatics.org/sms/iupac.html>
13. A. Bodenteich, L.G. Mitchell, M.H. Polymeropoulos, C.R. Merril, Dinucleotide Repeat in the human mitochondrial d-loop, *Hum. Mol. Genet.* 1 (1992) 140.
14. A. Bodenteich, L.G. Mitchell, C.R. Merril, A lifetime of retinal light exposure does not appear to increase mitochondrial mutations, *Gene* 108 (1991) 305-310.
15. B. Budowle, C.L. Fisher, D. Polanskey, B.K. Den Hartog, R.B. Kepler, J.W. Elling, Stabilizing mtDNA sequence nomenclature with an operationally efficient approach, *Progress in Forensic Genetics* 12.
16. K.L. Monson, K.W.P. Miller, M.R. Wilson, J.A. DiZinno, B. Budowle, The mtDNA population database: an integrated software and database resource for forensic comparison, *Forensic Sci. Comm.* 4(2) (2002).
<http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm>
17. European DNA Profiling Group (EDNAP) mtDNA population database project, <http://www.empop.org>.
18. D. Polanskey, B. Den Hartog, J. Elling, C. Fisher, R. Kepler, B. Budowle, Comparison of Mitotyper and phylogentic based nomenclature using the EMPOP database, *Int. J. Leg. Med.* (submitted).
19. K. Sekiguchi, K. Imaizumi, K. Fujii, N. Mizuno, Y. Ogawa, T. Akutsu, H. Nakahara, T. Kitayama, K. Kasai, Mitochondrial DNA population data of HV1 and HV2 sequences from Japanese individuals, *Leg. Med.* 10 (2008) 284-286.