Standardizing the Nomenclature for mtDNA Haplotypes with an Intuitive Hierarchal Execution Software Program

Bruce Budowle, Connie Fisher, Deborah Polanskey,
Bobi Den Hartog, Russel Kepler, and John Elling

ISFG, Copenhangen, Denmark, 2007

mtDNA Analysis

- Advantageous, compared with nuclear DNA
- For analyzing highly degraded or limited quantity forensic samples
- The technology and forensic interpretation for matching are robust and reliable

Nomenclature

- More refractory
- Inconsistencies occur
- Manual process
- A lot of debate for a minor issue

Nomenclature

- Displaying more than six hundred bases to describe results is unwieldy for communication
- So compare to a published reference sequence
- Cambridge Reference Sequence, or rCRS
- List only those sites that differ from the rCRS
- Provides a common language and a simple operational tool for describing the variation observed

Nomenclature Issues

- There are some situations where multiple alignments of the mtDNA sequence with the rCRS may be possible

- Therefore, identical sequences may be (and have been) aligned and listed differently and inconsistently within and among population databases

Misinformation

- Contrary to assertions by Bandelt and Parson (2007)
- Inconsistency in nomenclature typically does not lead to “unjustified exclusion of the culprit as the donor of the stain.”
- They do not provide any case for this
- Most comparisons in forensic casework are side-by-side such that the entire sequence information is directly accessible

Databases

- However, database entries and searching is a different issue
- The count may be underestimated if the same sequence has multiple alignments and is named differently or inconsistently
- On a practical level inconsistent nomenclature that has occurred has had nominal impact on estimates of the rarity of a mtDNA haplotype
- Because of the very high haplotype diversity of mtDNA

Goal

- Minimize the impact – Quality
- Develop a standard, stable nomenclature
- Three approaches:
- Operational
- Phylogenetic
- Full text searching
- Only Operational and Phylogenetic approaches discussed

Bifurcating Operational Approach
(Wilson et al)

1) Least differences - characterize the variant(s) using the least number of differences (i.e., substitutions, insertions, deletions) from the reference sequence;

2) Indels then substitutions - if there is more than one possible alignment, each having the same number of differences with respect to the reference sequence, a prioritization is made first for insertions/deletions (indels), then transitions, and lastly transversions;

3) 3' indel grouping - insertions and deletions should be placed 3' with respect to the light strand (when possible, insertions and/or deletions should be combined); and

4) 3' gap grouping - gaps are combined together only if they can all be placed in the most 3' position while maintaining the same number of differences from the reference.

Bifurcating Operational Approach

- Rules were not followed strictly for practical reasons
- Some listing of variants from the rCRS would be inconsistent with the traditional nomenclature
- For example, a profile with the variants from the rCRS --150T and 152C – would have been renamed -- 149.1T and 152d
- Nothing wrong with renaming, but…
- Traditional nomenclature was maintained to avoid confusion (ISFG recommendations)

Bifurcating Operational Approach

- If modify rule 2 so substitutions take precedent over indels
- Greater consistency with both historic traditional and current operational nomenclature
- Additional priority rules for those few situations where two or more alignments are still possible after the modified operational rules have been executed
- Outcome -- stable and consistent nomenclature
- Haplotypes once named and entered into the database never need to be renamed

Phylogenetic Approach

- Appears reasonable; scientifically apealling
- But issue here is operational
- Has a number of basic operational and practical limitations
- It is not stable
- Ignores limitations of evidence

Phylogenetic Approach Limitations

- Partially degraded or limited samples may not be possible to describe phylogenetically
- So evidence needs to be compared with all possible sequences that could match
- With phylogenetic approach - two regions could have different sequences
- But could possibly be aligned similarly
- Result is an underestimate of the “matching” sequences

Phylogenetic Approach Limitations

- Can not address all regions for nomenclature
- Private mutations and hotspots are not addressed
- Yet most of variation for forensic comparisons are the private mutations
- Still requires two nomenclature approaches

Phylogenetic Approach Limitations

- Most practitioners are not likely to become proficient sufficiently to understand the nuances of phylogenetics
- Current analyses are not readily informative
- EMPOP data are not available to the public to assess reliability

Algorithms for Hierarchal Approach

- WRV2.1 executes rules in the following order:

- Rule 1 - minimum number of differences with respect to the rCRS
- Rule - Anchor 310T (preferring that 310 T in the rCRS be aligned with a T when possible)
- Rule 2A - Substitutions then indels
- Rule AC - treat the “AC” motif as homopolymeric with respect to indels in the AC repeat region
- Rule 3B (*) - Combine indels whenever possible
- Rule 3A (*) - Place indels 3’ to a homopolymeric region
- Rule 2B - Transitions then transversions

Software Tool

- Under development
- Minimize human intervention when applying the modified formalized hierarchal rules
- Provide consistency and stability

Software Tool Data

- 4839 mtDNA haplotypes from SWGDAM data
- Identify regions - section of a sequence that could be accurately aligned and typed with its corresponding region in the rCRS independently of any other region
- Roughly half of the regions (41025 out of 91720) contained polymorphisms or differences from the rCRS
- 7420 (out of 41025) regions remained unresolved after the execution of Rule 1
- 99.9% of the regions in the SWGDAM database complied with full set of executable rules
- Additional rule handles last two ambiguities

Software Tool Data

- 40 regions where new rule set differed with SWGDAM data
- These differences were due to implementation of the new rules!
- The new rule set now takes precedence
- Demonstrates that there actually was a good deal of consistency
- New Hierarchy Rules are built around operational preferences

Conclusion

- A standard mtDNA nomenclature is described
- Enables consistency
- Provides stability
- Easier to understand and explain
- Will be executable using software tool

- Forensic scientist need not be distracted by this debate on nomenclature and instead focus on the more demanding aspects of casework