B129 How Well Can Race be Predicted From mtDNA?

Mark D. Leney, PhD, Joint POW/MIA Accounting Command, Central Identification Laboratory, 310 Worchester, Hickam Air Force Base, HI 96853*

After attending this presentation, attendees will learn a technique to predict likely self-identified race from an unknown mtDNA profile and to attach statistical weight to such an attribution

This presentation will impact the forensic community and/or humanity by allowing explicit calculation of the likely race of an unknown from mtDNA. While not as accurate as nuclear DNA based techniques designed to estimate biological ancestry this will work to provide additional information for a biological profile of a unknown person where nuclear DNA cannot be recovered and could augment nuclear DNA based techniques by flagging the likely race of the mother of an unknown contributor of mixed ancestry.

While mitochondrial DNA (mtDNA) in forensic contexts has generally been used to support or refute identification hypotheses, it also contains information that permits inference of the race of an unknown person contributing the mtDNA profile. The mtDNA population database was compiled to provide the forensic community with frequency estimates for mtDNA hypervariable region (mtDNA HVR) profiles observed in evidence-to-reference matches for a specific population of interest. However, the database can also be used to find the frequencies of matching or closely related mtDNA HVR profiles of unknown origin in several different ethnic/racial sub-samples. As the majority of all profiles in the mtDNA population database are themselves unique within the database, it is no surprise that most evidentiary profiles are also observed zero or one times in the database. While such a low frequency is adequate for establishing the relative rarity of the profile in question, it is not informative of probable racial/ethnic origin of the contributor. Nevertheless, as most individuals are only distinguished by a few polymorphisms in the mtDNA HVR from multiple individuals in the database, the pool of closely related individuals is a much richer source of information for estimation of probable ancestry.

Attendees will learn how a simple likelihood ratio can be constructed from the probability of observing closely related profiles to an unknown evidence profile in a specific racial or ethnic sub-group and the probability of observing equally related profiles in the rest of the database. Attendees will then observe that by calculating such a ratio for each racial or ethnic sub-group of interest for all possible degrees of profile relatedness it is possible to determine the peak likelihood ratio for racial or ethnic sub-group attribution. This peak represents the balance between idiosyncratic polymorphic differences that are a result of recent within-lineage mutations and older polymorphic differences that are a result of shared common ancestry. The height of the peak indicates the statistical weight of the racial or ethnic attribution. The attendees will also be able to review validation studies conducted on profiles drawn at random from the mtDNA population database and from unrelated but known mtDNA HVR profiles.

Neither race nor ethnicity is an absolute biological category but each does contain a biological component. This shared genetic heritage is what allows this technique to work. However, the mtDNA population database is based on self-identified race and ethnicity and thus various other factors act to degrade the power and accuracy of the technique. The main factors are cultural biases in the reporting or recording of single racial or ethnic categories (when many people are actually of mixed ancestry) coupled with the solely matrilineal nature of mtDNA inheritance. These factors will be reviewed and their impact on careful and proper interpretation of results will be discussed.

Additionally the applicability of the technique depends on adequate sampling of the populations in question by the database used. The presentation will briefly cover the diversity in the published database, indicating where the database is adequate and where it falls short, giving examples from casework to indicate potential pitfalls in interpretation when a profile from an underrepresented population is analyzed.

This technique offers a novel method for placing an explicit statistical value on inference of racial or ethnic group from an mtDNA profile. This could provide a useful lead in building a biological profile of an unknown contributor where only mtDNA was recoverable (as is often the case where conventional nuclear DNA testing fails). Furthermore it could be used in conjunction with the recently developed techniques that estimate biological ancestry from nuclear DNA markers; for example in a person of mixed ancestry, the use of this technique could provide a likely estimate of the self-identified race of the unknown contributor's mother. At the Joint POW/MIA Accounting Command it is hoped that the technique will assist case resolution for long-term cold cases by highlighting those fragmentary remains that are likely to be unidentified American casualties and those that are likely to derive from indigenous Asian combatants and civilians from the battlefields of Korea and Vietnam.

Race, Ancestry, Mitochondrial DNA

 ** Presenting Author*