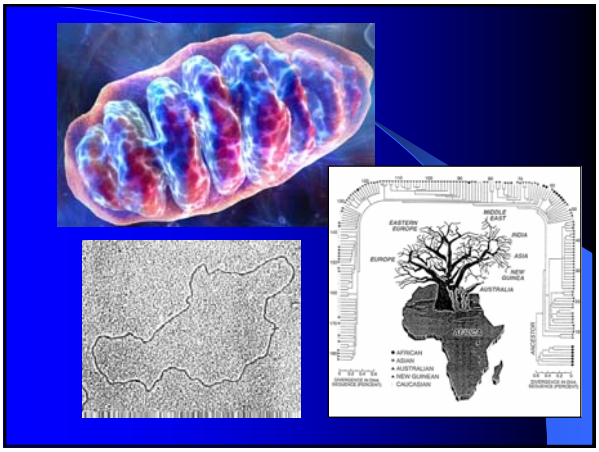
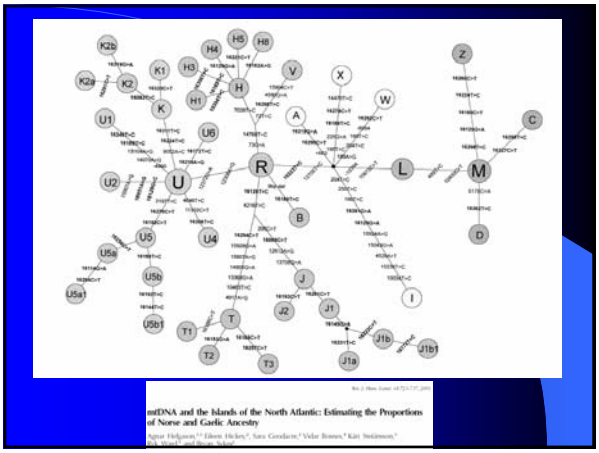
Evolutionary Rates in mtDNA Sequences: Forensic Applications and Implications

Thomas J. Parsons

Director of Forensic Sciences
International Commission on Missing Persons





Obligatory Reminders, The mtDNA Story

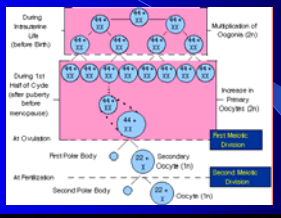
- Multi-copy DNA
  - ~3-10 per mitochondrion,
  - Hundreds or thousands per cell
- Maternally Inherited (yes, still true)
- No recombination (yes, still true)
- High Mutation Rate (~10-fold higher than nuclear)
  - Free radical environment of mitochondrion
  - Deficient Repair? (Beware of myth).

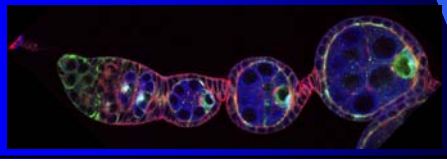
Obligatory Reminders, The mtDNA Story

High copy number and high mutation rate should be a recipe for a locus useless for purposes of identification. There should be huge population genetic variation within a single individual.

- Deficient Repair? (Beware of myth).

Genetic Bottleneck in Oogenesis



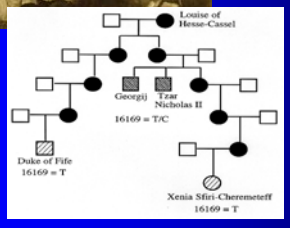


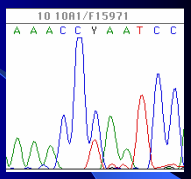
mtDNA "Mutation"

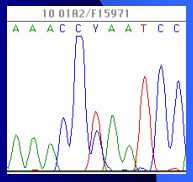
- Mutation occurs on single molecule
  - Among 3-10 mtDNAs within mitochondrion
- Mutation segregates within:
  - organelle
  - cell
  - Individual
  - Germ line: Bottleneck
- Mutation segregates between generations, becomes majority type= SUBSTITUTION
- "Mutation Rate" is often misused as the rate of sequence evolution on a population or species level.
  - Will continue in this noble tradition…

What is the "mutation rate?"

- Compared to Chimpanzee mtDNA sequences, early studies indicated that mtDNA mutations become established along mtDNA lineages about once in 300-600 generations.
- 1996-1997: Empirical studies comparing known maternal relatives showed differences more commonly than this.
  - **Howell, N., Kubacka, I., Mackey, D.A. 1996 *How Rapidly Does the Human Mitochondrial Genome Evolve?* Am. J. Hum. Genet. 59:501-509.**
  - **Parsons, T.J., Muniec D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P., Holland, M.M. 1997 *A High Observed Substitution Rate in the Human Mitochondrial DNA Control Region.* Nature Genetics 15:363-368.**

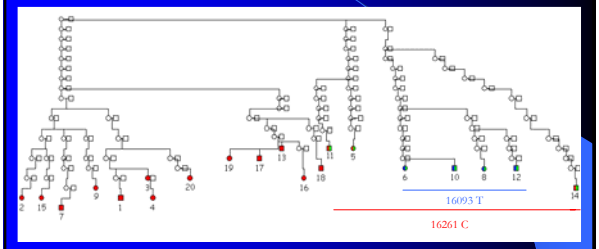






European Royalty Study
AFDIL

One of 49 Pedigree Lineages Studied: 686 generational events.



All individuals share an additional 10 distinctive polymorphisms, confirming matrilineal relationship.

Pedigree Studies

- Our results match with and add to other published studies.
  - Heyer et al, 2001, Siguoardottir et al, 2000,
  - Howell et al, 2003.
- Fixation Mutation Rate is approximately **1/100 generations**.
- ~5 times faster than the rate predicted by evolutionary studies.

The problem of defining an mtDNA exclusion.

- Rather high chance of intergenerational mutation.
- Cannot exclude on the basis of a single difference between two sequences.
  - Formally true: a mutation could cause the difference, just as you don't exclude paternity on the basis of a single STR allele inconsistency.
- Reporting: FBI. Single difference= "Inconclusive"

This can be very problematic.

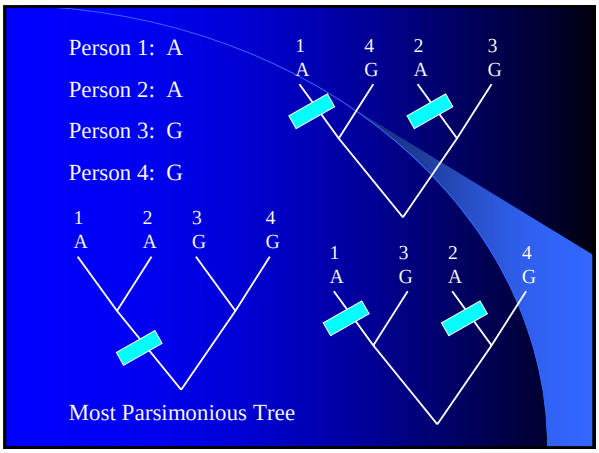
- If a sample has a common type.
- There are multiple individuals in an incident.
- Chances are good that someone else will match, or be within one base
  - Multiple families “cannot be excluded.”

How to Evaluate Single Base Differences?

- Not all sequence differences are created equal.
  - Extreme rate heterogeneity.
- To scientifically consider the significance of matches and mismatches, we need to evaluate:
  - The relative mutation rate at the site of difference
  - The overall mutation rate between generations
  - The population frequency of the background sequence type.

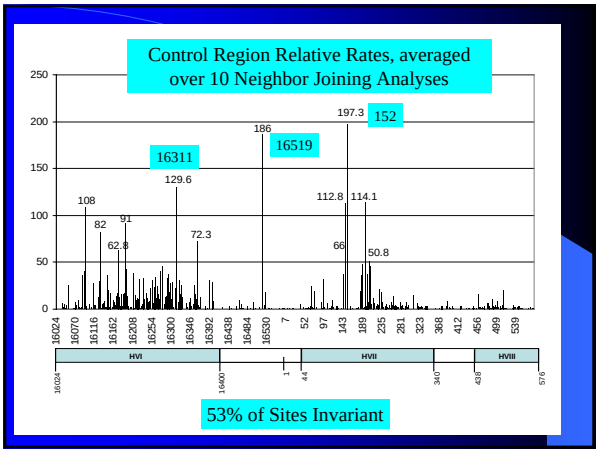
Comprehensive Analyses of Relative Rates over the entire mtDNA Genome

- Mike Coble, PhD Dissertation (2004):
  - Relative rates in the Coding Region
  - 646 Whole mtDNA Genome Sequences
  - Phylogenetic Analysis using Parsimony
- Katherine Strouss, Masters' Thesis (2006)
  - 2568 entire Control Region sequences
  - Phylogenetic Analysis Using Neighbor Joining Distance Methods

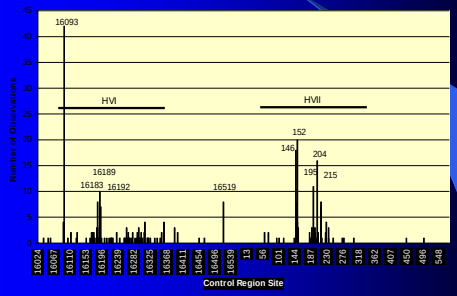


Phylogenetic Tree Reliability

- Phylogenetic Trees of mtDNA sequences are notably imperfect
  - Basically impossible to know "The True Tree"
  - Parsimony: there are thousands of equally parsimonious trees.
- Rate analyses demonstrated to be extremely robust to variation in tree topology.
  - Parsimony: comparison of different trees
  - Neighbor Joining: comparison of 10 algorithms
- Rate Spectrum, for all practical purposes, was identical.



Heteroplasmy Rates Match Very Well with Population Evolutionary Rates (Some Exceptions)
Data from AFDIL, ~5000 CR Sequences



LR approach

- We have reasonable estimates of intergenerational mutation rate and site specific relative rates.
- This can fit into a likelihood ratio approach.

$$LR = \frac{P(E|H1)}{P(E|H2)}$$

LR Treatment

Grandmother: 263G, 315.1C (MCT, p=.07)

Evidence Sample: MCT + 16311C (p=0.013)

Proportion of mutations at 16311= 0.028

2 Generations

$$\frac{P(E|H1)}{P(E|H2)} = \frac{[(freq)A]\,[1/100][0.028][2]}{[(freq)A]\,[(freq)sample]} = \frac{0.9875}{0.0121}$$

$$= 0.043$$

Discourages hypothesis of association by factor of 23

Range of results with mutations.

- Strongest evidence of relationship comes from a rare sequence with a common mutation.
  - Sequence not present in database, mutation at 152. ***LR= ~3***.
- Worst case: rare sequence mutating to commonest sequence, differing at very slow site
  - LR= 0.00006
  - ***16,000 fold discouragement***

AFDIL Case

- Skeletal remains: could be one of nine individuals.
- References for 6 individuals differ at multiple positions
  - Can **exclude** these.
- Reference sample from one family not available.

AFDIL Case

Sequence for remains: (16024-16365); (35-369)	**Family A: exact match**	**Family B: difference at 228**
16069T		
16126C		
73G		
185A	**Family C:**	
228A	**?????**	
263G		
295T		
315.1C		

Frequency of Remains' Sequence= 0.012
Frequency of Family B Sequence= 0.006

LR Treatment

- Remains belong to Family A with an exact match, versus they come from a random person:

$$\frac{P(E|H1)}{P(E|H2)} = \frac{[(freq)A]\,[1-(1/100)]}{[(freq)A]\,[(freq)sample]} = \frac{0.99}{0.012}$$

$$= 82.5$$

LR Treatment

- Remains belong to Family A (exact match) versus Family B (A-G difference at 228):

$$\frac{P(E|H1)}{P(E|H2)} = \frac{[(freq)A]\,[1-(1/100)]}{[\text{chance of a mutation at 228 in 2 gens}][(freq)B]}$$

[chance of a mutation at 228 in 2 gens]=
(1/100)(2)(0.0056)= 0.00012

(0.56% of CR mutations occur at 228)

LR Treatment

- Remains belong to Family A (exact match) versus Family B (A-G difference at 228):

LR= 8500

Summary

- DNA evidence is:
  - **8500 times** more likely if remains come from Family A (match), instead of Family B (difference at 228).
  - **80 times** more likely if remains are from Family A, than from the family with no reference.
  - If all other non-DNA evidence is equal, it is **80 times** more likely that remains are from Family A than not.

Caveats, Uncertainties

- Phylogenetic relative rate estimates may not mirror intergenerational mutation rates.
  - Additional filter of selection at the population level.
  - Fastest sites may still be underestimated.

Caveats, Uncertainties

- Phylogenetic rates do not reflect any "tissue specific" rates.
  - E.g. muscle specific hotspots
    - These are always "pretty fast" overall though.
  - When encountered in practice it is usually unknown if the mutation is somatic or germ line.
  - Somatic mutations will most often be heteroplasmic.
    - Situation of hair- genetic bottleneck in hair histogenesis.

Caveats, Uncertainties

- Assumes that propensity for mutation at a site is independent of other polymorphisms.
- Within neutral theory, this is normally considered true as first approximation.
  - No doubt some exceptions
  - 16193C, for example

In Conclusion…

- We have solid scientific information on the relative propensity of all sites in the mtDNA molecule for mutation and evolution.
- To a very good first approximation, at least, we have a framework for using this to assist in evidence interpretation.
- I recommend revising the category of "inconclusive" for single base mismatches.
  - In criminalistics, this is a bit loaded against the suspect
  - In missing persons, there is no conservative approach based on presumption of innocence- we need to know where the evidence takes us.

Acknowledgements

- AFDIL Research Section :
  - Odile Loreille, **Rebecca Just**, Jessica Saunier, Katie Strouss, Jennifer O'Callaghan, Carla Paintner, Toni Diegoli; Kim Sturk; Heather Williams, Kim Watson, **Mike Coble, Jodi Irwin**
- AFDIL:
  - Jim Ross, FTI: IT development
  - James Canik, Colonel Brion Smith, Scott Carroll (AFDIL)
  - Demris Lee and Suzie Barritt (Casework Section Chiefs)
- Antonio Salas, Universidad de Santiago de Compostela