DNA for Genealogy

Robert Casey

March 19, 2009

www.rcasey.net/present

Overview of DNA

- Some science behind DNA
- Many kinds of DNA tests
- Y-DNA - primary test for genealogists
- Best results for DNA
- When DNA does not do well
- Clusters are important
- Overlapping haplotypes & Non-Paternal Events
- Most Recent Common Ancestor (MCRA)
- DNA - as big as Internet and PCs

Some science behind DNA

- DNA strands are very long ladders (Y-DNA is one ladder)
- Each ladder step is connected by molecules (marker)
- The number of molecules of ladder step (marker value)
- The number of molecules change over time (mutation)
- Markers are selected for the rate of change
- Rate of change matches genealogical timeframes
- Tests include 12, 25, 37 and 67 markers (FTDNA)
- More submissions - more accuracy
- More markers - more accuracy

Some science behind DNA

- 95 % DNA tests for genealogists use Y-DNA
- Y-DNA can trace only all male ancestries
- Only junk DNA used - no known medical information
- Most analysis requires math rather than science
- Mutations randomly add or lose molecules
- Mutations randomly occur over generations
- Donor, father, grandfather, etc. all mutate the same
- Mutation may be genealogically significant (very old)
- Mutation may be recent and not important (very recent)

Many kinds of DNA tests

- Y-DNA (STR) - mutate at genealogical rates
- Y-DNA (SNP) - determines only deep male ancestry
- mtDNA - determines only deep female ancestry
- X-DNA - determines recent female relationships
- Heritage (ethnic) - determines race mixture
- Autosomal (paternity) - determines adoption issues
- Y-DNA (STR) is only real test for genealogical usage
- Y-DNA (SNP) is useful when 67 markers are not enough
- FTDNA primary tester - others gaining market share

Y-DNA - primary test for genealogists

- 95 % DNA tests for genealogists use Y-DNA (STR)
- Unfortunately female DNA does not change fast enough
- Other tests do not solve genealogical questions
- Must locate male descendant of surname of interest
- Multiple tests required to verify age of mutations
- You want to eliminate recent mutations
- You need enough tests to determine older mutations
- You want sons and grandsons of oldest proven ancestor
- Testing first, second and third cousins not that useful

Best results for DNA

- Genetic tests can prove / disprove possible connections
- Will not prove actual connection - just closely related
- Your random submission may fit into related cluster
- A new random submission may fit into your cluster
- Developing large cluster of lines unconnected by paper
- Focus research on these unconnected lines by paper
- Stop unproductive research on lines that are not related
- Can support (not prove) speculative connections
- Can disprove lines connected by paper

When DNA does not do well

- Rarely can be used to prove actual connections
- Mutation rates are average - not absolute
- Some lines beat the odds and mutate faster than normal
- Other lines mutate much slower than normal
- Number of mutations may mislead degree of closeness
- Proves interconnection of lines - not actual connections
- Concept of recent and older mutations misleads many
- 80 % of mutations are not important to genealogists
- Too high expectations to solve exact relationships

Clusters are important

- A collection of related DNA submissions is a cluster
- There are two types of clusters - genealogical & genetic
- Genealogical are submissions related in last 300 years
- Genetic are submissions in last 600 years
- You look for genealogical connections at 300 years
- Genetic is when our ancestors first used surnames
- Genetic clusters not used for genealogical connections
- Genetic cluster used primarily to determine real MRCA
- Books & web sites rarely address genetic clusters

Overlapping Haplotypes

- This issue is rarely covered by books & web sites
- Close relatives chose different surnames 600 years ago
- These relatives naturally overlap with our surname
- Due to only 67 markers, a lot of overlap in general
- You want to categorize non-surname matches
- Choices are “overlapping haplotypes” or “NPEs”
- Overlapping haplotypes are filtered out
- NPEs can be a new gold mine of genealogical treasures
- Significant confusion on distinction of two categories

Non Paternity Events (NPEs)

- NPEs are genetic relatives and are very significant
- NPEs are adoptions, out of wedlock, name changes, etc.
- NPEs have both genetic and geographic connections
- Many times NPEs are already known
- Early NPEs tend to have a cluster of submissions
- If closer relatives are 90 % your surname & 20 % other
- Then the 20 % other surname is probably NPE line
- If closer relatives are 90 % different surnames
- Then “overlapping haplotype” are most common

Most Recent Common Ancestor (MRCA)

- MRCA is a set of DNA markers that fits oldest ancestor
- Some donors could have same markers as MRCA
- Some donors will have a lot of mutations (yellow flag)
- Most donors will have average of 2 or 3 mutations
- “Majority Rules” usually determines MRCA
- Most common marker values must be MRCA is assumed
- Sometimes there are two related clusters in 600 years
- MRCA programs calculate percentage of relatedness
- Mutation rates of every marker varies dramatically

DNA - As big as Internet and PCs

- DNA is major “new” source of genealogical information
- Excellent for validating speculative connections
- Can also be used to disprove speculative connections
- New gold mines are NPE lines that can be added
- If you have 20 or 30 submissions at 67 plus markers
- Then you may be able to prove some connections
- Charges for DNA testing continues to fall
- Economical way to eliminate unproductive research
- Economical way to validate speculative connections