Genetics for Genealogy: Navigating the DNA Marketplace
Linn County Iowa Genealogical Society
Cedar Rapids, IA
January 26, 2019

Speaker:
Bryant McAllister, PhD
Associate Professor of Biology
University of Iowa
319-335-2604
bryant-mcallister@uiowa.edu

Website: https://wiki.uiowa.edu/display/2360159/G4G

Overview: With a single mouse-click sophisticated genetic tests for human ancestry analysis are available directly to consumers at a relatively low cost. Advances in human genetics over the past 30 years have enabled the emergence and growth of this "recreational genetics" marketplace. My goal is to introduce the science underlying these tests, compare the products delivered by the companies that dominate the US marketplace (e.g., 23andMe, AncestryDNA, MyHeritageDNA, and Family Tree DNA), and demonstrate their utilities and limitations. Personal genetic analysis provides a window into the past at different time scales, including genetic inferences of pre-genealogical ancestry along with the discovery of genetic relationships between cousins as a genealogical research tool.

Presentation Objectives

- Describe the direct-to-consumer genetic testing marketplace
- Consider differences between DNA testing options and testing strategies that support your interests
- Introduce principles for discovering biological relationships among individuals

DNA Tests in Genealogical Research

- Discover living relatives with common genealogical interest
- Support documented ancestral relationships
- Reveal inaccuracy in documented family history
- Interrogate a branch of poorly (or non) documented ancestors
- Discover relatives with previously unknown biological relationship
- Test a hypothesis of biological relationship

Major Testing Options (list price)

- **23andMe** –Includes Ancestry Estimates, DNA Matching, Carrier Status, and Trait Prediction reports ($199 or $99 options).
- **AncestryDNA** – Includes Ancestry Estimate, DNA Matching and other features ($99). Integrated with Ancestry.com family tree for an extra subscription fee.
- **MyHeritageDNA** - Includes Ancestry Estimate and DNA Matching ($79 list). DNA matches on same platform as tree-building infrastructure.
- **Family Tree DNA** – Various tests available including Family Finder SNP test ($79).

Potential Surprises from the Results of a DNA Test

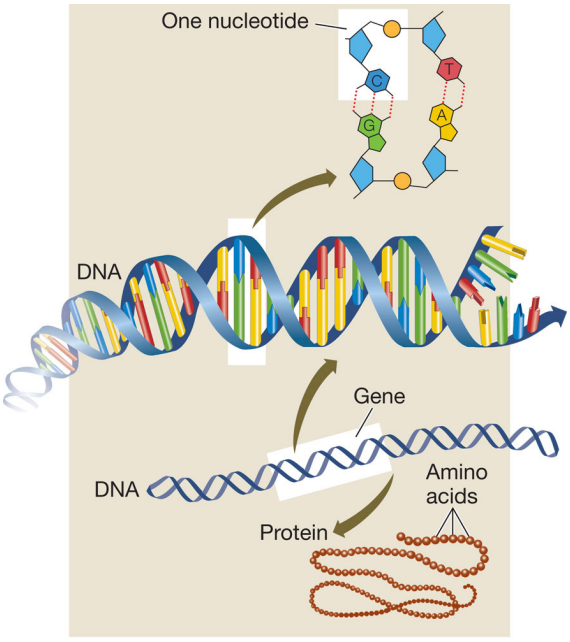
- Surprising findings about family relationships
  - *New relatives discovered by genetic matching*
  - *New ancestors discovered by genetic matching*
  - *Greater distance than expected with other tested family members*
- Surprising prediction of ancestry composition
  - *Individuals contain a genetic mixture representative of their ancestry through human history*
- Surprising discovery of genetic susceptibility to disease
  - *You and/or your children may have an increased chance of developing an incurable genetic disease*

Questions about the use of DNA tests

- What outcome(s) do you want in the test results?
- What does the DNA test assay?
- What interpretations are delivered in the results?
- How would you use these results?
- What is the cost, and how does it compare with competitors?
- Are there additional fees associated with access or use of test results?
- Are other customers satisfied with the service?
- Are the terms of service acceptable to you?

DNA is the cellular recipe for life

- A **Genome** is the entire set of **DNA** molecules present in a cell.
- Each DNA molecule contains a string of A=T pairs and G≡C pairs as a **DNA sequence**.
- DNA sequence contains the cellular recipe for life. Regions of DNA molecules (genes) encode products (mostly proteins) that perform cellular and organismal functions.
- Different cells of an individual contain the same genome sequence, but unless you have an identical twin, your DNA sequence is different from any other individual.



LIFE 10e, Figure 1.5

- The human genome is organized in 23 pairs of linear DNA molecules (chromosomes) inherited from both parents; 22 pairs of autosomes, and a single pair of sex chromosomes.
- Gametes (egg or sperm) contain only 1 member of each chromosome pair.
- An egg also has a circular DNA molecule in the mitochondrion (the mtDNA) inherited solely from the mother.

Methods for placement of an individual within the Human Family Tree

- Genetic similarity with reference populations sampled throughout the globe
  Prevalence of mtDNA and Y chromosome haplogroups
  Estimate of compositional mixture from different ancestry groups
- Segments of sequence identity along the DNA of autosomes and X chromosome
  Number and length of identical segments decreases with relationship distance
- Shared sequence in DNA of maternally-inherited mitochondrion or male-limited Y chromosome
  Placement on a branch of the human family tree
  Revelation of closest relatives by fewest mismatches

Genome-Wide SNP Analysis

- Most of the human genome is contained in 23 pairs of linear DNA molecules, where one member of each chromosome pair is inherited from each parent.
- Autosomes are inherited equally from each parent, whereas inheritance of the pair of sex chromosomes depends on the sex of the parent and child.
- Each company uses a common technology that assesses about 600,000 different known variable sites (SNPs) in the DNA of your genome (including mtDNA in some cases). Most of the same sites are used in the tests of different companies.
- Each **Single Nucleotide Polymorphism** (**SNP**) analyzed has a unique ID and has an identified position along a chromosome in the reference human genome sequence.

Ancestry Composition Estimates from Genome-Wide SNP Data

- *It's an estimate!!!*
  Each analysis is based on a set of computer algorithms for handling data from a broad diversity of individuals.
  The set of reference samples used in the analysis, and how the analysis is (or isn't) tailored to self-reported ancestry, differs between companies.
- Current political boundaries have little relevance
  Broadest categories capture genetic similarities present among peoples in different countries.
- Small contributions can represent a variety of things
  Application of an admixture model using many reference populations over-emphasizes effect of rare variants.

Autosomal SNP Matching

- The 22 pairs of autosomes are inherited equally from each parent, and are used by each of the companies to identify DNA relatives. The X chromosome can also be used to identify DNA relatives.
- Each Single Nucleotide Polymorphism (SNP) analyzed has a unique ID and has an identified position along a chromosome in the reference human genome sequence. Each customer's genetic profile is determined at each of the approximately 600,000 SNPs analyzed. The genetic profile may reveal that the pair of chromosomes contain the same base composition at a SNP (e.g., GG), or that the pair of chromosomes contain different bases (e.g., AG).

- Segments of DNA sequence half-identity, where an extended stretch of assessed SNPs indicates a common genetic profile on one member of the chromosome pair, are used to identify DNA relatives.
- Begin to lose DNA sequence identity between 3rd cousins and the loss increases between more distant relatives.

Features of each DNA Relative

- Identifying information (for non-anonymous DNA matches)
- Mechanism for messaging relatives through web interface or email
- Estimate of relationship, and the genetic basis for that inference
- User provided information on family history; surnames, family tree, etc.

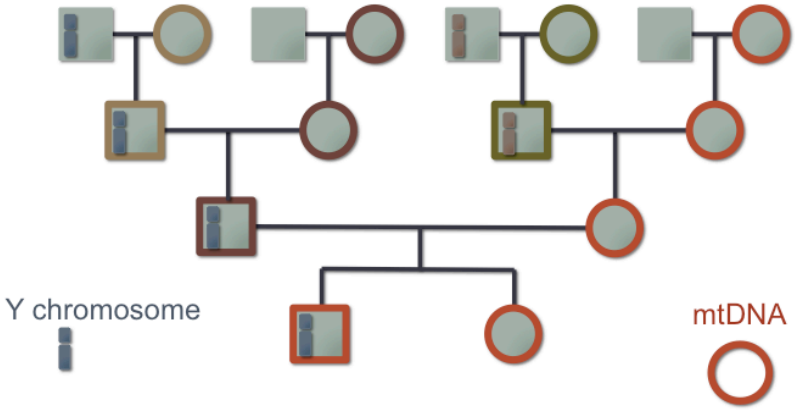
Mechanisms for Clustering Relatives

- "Shared Matches" tab in AncestryDNA
  Shared matches limited to 4th cousin and closer to tester
  Shared matches limited to 4th cousin and closer of 1° match
  Sharing results of a DNA test bypasses this limit
- "DNA Circles" and "New Ancestor Discoveries" in AncestryDNA
  DNA kit linked with shared family tree
  Dependent on both DNA and Ancestor Information matches
- "In Common With" sorting in Family Finder (FTDNA)
  Use "Family Finder Matrix" to compare multiple DNA relatives
- "Relatives In Common" listed with each 23andMe match
  With shared profiles, also identifies whether there is genome overlap among the matches

DNA Testing Strategies for Family History Research

- Full siblings share about 50% autosomal DNA segment identity, thus provide the ability to confirm DNA matches and to expand (primarily 4th cousin and beyond) match lists.
- Close relatives (parents, 1st and 2nd cousins) with known relationships enable methods for positioning matches within the branches of your family tree.
- DNA matching with known half relatives provides support for common ancestry from a single individual, whereas DNA matching with full relatives beyond 1st cousins only provides support for common ancestry from one person of the common ancestral couple.
- Tabular organization of historical and genetic relationships within family groups can be used to summarize the support of the historical record.

Uni-Parental DNA Markers



- Single DNA molecules inherited from one parent, and that trace a single ancestral line.
- **Mitochondrial DNA** (mtDNA) is a small circular molecule present in the mitochondria. Only transmitted from the maternal parent, and thus is inherited through the maternal line.
- The DNA molecule of the **Y chromosome** contains a gene that directs development toward the male pathway, thus it is present only in males. The Y chromosome is transmitted to sons only, and is thus inherited through the paternal line.
- Shared changes in the sequence of the mtDNA or the Y chromosome indicate common ancestry. Acquisition and comparison of DNA sequences from mtDNA or the Y chromosome from many individuals reveal deep relationships of maternal and paternal ancestry, respectively. Branches on reference trees depicting these relationships are designated with a specific **haplogroup** nomenclature that represents the underlying DNA sequence diversity.

Potential Outcomes of Y Chromosome Test

- Support for Patriline
  Close relationship with other men that trace their ancestry back through common patriline
- Association with a Different Surname
  Close relationship with other men that share a common surname
- No Clarity in Ancestral Surname
  Close matches exhibit a variety of different surnames
- No Matches
  No person in the database meets threshold criteria for matching

Genetics for Genealogy

- Results vary dramatically based on individual ancestry and database composition – *test the oldest generation.*
- Endogamy increases the background level of matching among individuals.
- Lists of DNA matches contain two types of errors:
  The genetic identity shared with the relative was not inherited from your most common ancestor(s), but rather a more distant ancestor.
  Relatives descended from a known common ancestor(s) do not appear in the list of DNA relatives, which is an increasingly likely outcome for 4th cousins and beyond.
- DNA relatives may be connected through multiple pathways of common ancestry.

Genetics for Genealogy: Major Testing Options

- **AncestryDNA** ($99) – Fully integrated with active Ancestry.com tree with powerful features; no ability to directly examine matching DNA segment(s)
- **MyHeritage DNA** ($79) – Detailed information on genetic basis of DNA matches; DNA matches exist on site for building family trees (accepts raw data from 23andMe or AncestryDNA)
- **Family Finder Test** at Family Tree DNA ($79) – Broad set of features to examine DNA matches; website navigation is challenging and database is relatively small (raw data from 23andMe or AncestryDNA can be transferred into Family Finder database)
- **23andMe** ($199 or $99) – Good for broad set of interpretations, entertainment, & validating relationships; most users are private and not motivated by genealogy

  This website contains a details comparison of the major testing companies https://wiki.uiowa.edu/display/2360159/DTC+Genetic+Testing+Companies+Compared

AncestryDNA

- Primary application is family history research through DNA matches, but includes an ancestry composition estimate and novel migrations inference.
- Largest database of tested customers. (over 10 million)
- AncestryDNA test can be linked with family tree in Ancestry.com for fully integrated research.
- Sophisticated DNA matching algorithm that reduces the number of false-positive matches of close relationship.
- Automated family clustering in DNA circles.
- Ability to manage DNA kits of multiple family members in a single account, and share results between accounts.

Family Tree DNA

- Primary application is family history research through DNA matches.
- Broad set of tests and tools for working with DNA matches for family history research, but their utility and integration in the website is not intuitive.
- Relatively small database, but participants are mostly engaged and knowledgeable about family history research.
- Each DNA sample is registered in an account, and different tests on that sample are managed in a single account.
- Participate and/or manage projects that include multiple participants.
- Accepts raw DNA data from AncestryDNA, The Genographic Project, and older 23andMe tests.

23andMe

- Broadest set of interpretations from a DNA test including maternal line, paternal line for males, ancestry composition, Neanderthal composition, DNA relatives, carrier status, and traits.
- Choice to maintain privacy or not.
- Interface for browsing raw DNA data.
- Multiple kits can be managed in a single account, and some results can be shared between accounts.