The ABCs of DNA

Julie Roberts Szczepankiewicz
Polish Genealogical Society of New York State
14 March 2019

What is DNA?

- Deoxyribonucleic acid, a molecule made up of 4 chemical units called nucleotides, whose sequence contains instructions for growth, development, function, and reproduction of all living things.
- The 4 nucleotides which make up DNA are Adenine (A), Thymine (T), Cystosine (C) and Guanine (G), and these 4 units repeat in a particular sequence, forming a long string.
- DNA exists as 2 paired strands of nucleotides, coiled around each other in a double helix. A on one strand is always paired with T on the second strand, and C is always paired with G.
- If you think of the two DNA strands as the long uprights of a ladder, the pairs of nucleotides make the rungs

Where can DNA be found?

In the cell nucleus of every cell in the body.

- The nucleus contains 46 chromosomes, containing about 6 billion DNA base pairs (bps).
- There are 22 pairs of autosomes plus one pair of sex chromosomes – XX or XY – for a total of 23 pairs of chromosomes.
- A chromosome is a single strand of DNA plus associated proteins.

Eggs and sperm contain only one autosome from each pair, plus an X (eggs or sperm) OR a Y (sperm only). So each egg or sperm cell contains half the normal amount of DNA, and when combined, they create a new individual with the normal amount.

In the mitochondria:

- Mitochondria are organelles that serve as the powerhouse of the cell.
- Mitochondria have their own DNA in a single circular chromosome, consisting of 16,569 base pairs.
- In human fertilization, the sperm cell contributes *only* nuclear DNA, not mitochondrial. Mitochondrial DNA is therefore inherited *only* from the mother.

Some definitions:

- Gene: a unit of heredity, consisting of a distinct sequence of nucleotides in a chromosome, which is transferred from a parent to the offspring and can determine some characteristics of the offspring. All genes are made of DNA, not all DNA is part of a gene.
- Mutation: A permanent, heritable change in the sequence of nucleotides in the DNA

- Allele: Each unique form of a single gene, which arises by mutation and is found at the same place on both chromosomes of a pair. Can also refer to unique forms of parts of DNA which are in matching locations on paired chromosomes, but not within genes.
- Single Nucleotide Polymorphism (SNP): The existence of multiple versions of a DNA sequence that vary by only one nucleotide in one location. (Example: AACT**G**CATTGAG vs. AACT**A**CATTGAG
- Centimorgan (cM): The unit in which genetic distance is measured (not a physical distance). Genetic distance concerns crossing over.

Types of DNA tests:

Y-DNA test:

- Since only men have a Y chromosome, this is useful for tracing the paternal line.
- Y-DNA does not recombine, so it passes down virtually unchanged from father to son.
- Y-DNA testing will yield a particular *haplotype*, a kind of genetic profile based on existence of particular mutations at specific points in the DNA along the length of the Y chromosome. Individuals with the same haplotype form a *haplogroup*, which means they share a common ancestor at some point.

Mitochondrial DNA test (mtDNA test):

- Since mitochondria only come from the egg cell at fertilization, this can be used to trace the maternal line.
- Like Y-DNA, mitochondrial DNA does not recombine, so it passes down virtually unchanged from a mother to her children of both sexes.
- Mitochondrial DNA testing will also yield a haplotype.
- Most useful for studying population migrations. Because mitochondrial mutations happen very rarely, it's impossible to know whether two people with identical mitochondrial DNA share a common ancestor relatively recently, or thousands of years ago.

Autosomal DNA testing (atDNA test):

- Both men and women can test.
- Most useful for identifying relatively close relationships (5th cousins or closer), therefore results are relevant in genealogical time frame.
- **The real value of DNA testing lies not in the ethnicity estimates, but in the DNA matches, which allow you to confirm and extend your documentary research!**

Which tests are offered by which companies?

Ancestry DNA:

- Offers atDNA testing only.
- No health information provided.

Family Tree DNA (FTDNA):

- Offers Y-DNA, mtDNA, and atDNA testing.
- No health information provided.

23 & Me:

- Offers two packages: "Ancestry," which includes Y-DNA, mtDNA, and atDNA testing, and "Health + Ancestry" which includes all of the above plus health reports (higher price).

My Heritage:

- Offers atDNA testing only.
- No health information provided.

Living DNA:

- Offers atDNA, mtDNA and Y-DNA testing all in the same price.
- No health information provided.
- Offers more precise ethnicity estimates for those with UK ancestry.
- Recently began to offer DNA matches.

A closer look at atDNA testing:

- SNPs occur once every 300 base pairs on average, so there are approximately 10 million SNPs in the human genome.
- DNA tests sample a small percentage of these (about 6-7%) rather than sequencing the whole sample.
- When two people share the same alleles with a number of consecutive SNPs in common, it suggests that they share a segment of DNA in common.
- If this segment is longer than a threshold set by the testing company, the test algorithm reports it as a match.
- Specific threshold criteria for defining a match are usually expressed in number of shared centimorgans and shared SNPs.

Types of Matches:

Identical by Descent (IBD):

- A segment of DNA shared by two (or more people) which has been passed down from a common ancestor.
- Generally speaking, the larger the number of centimorgans shared, the closer the relationship. These segments are what you want to focus on in your research.
- Beginners should focus on identifying relationships to closer relatives (4^{th} cousins or better) and not worry about more distant relationships. Start with matches with whom you share one or more segments that are at least 20 cM or larger.

Identical by Chance (IBC):

- Matching sequences of DNA which appear as an artifact of testing process.
- "False positives."
- These can be eliminated if a parent is available for testing, since all your DNA must come from one parent or the other. If someone matches you, but does not match either parent, the match must be IBC.
- Typically, these are small segments, less than 7 cM.

Identical by Population (IBP):

- Happens when a large portion of a population group shares a particular segment of DNA.
- Segments involved can be greater than 7 cM and **are** identical by descent, but common ancestor may be too distant for identification through documentary research.
- Often happens with endogamous populations (Ashkenazi Jews, Acadians, etc.)

Genetic Recombination: The process by which offspring are produced with combinations of traits that differ from those found in either parent.

Important concepts regarding recombination ("crossing over"):

- It occurs during meiosis (the process by which egg and sperm cells are created which have half the normal number of chromosomes).
- It's not uniform throughout the genome; some areas of chromosomes exhibit more recombination than others. More crossing over is exhibited in females during meiosis than in males.
- It's why we don't inherit DNA in equal proportions from each of our ancestors.

A word about those ethnicity estimates:

- Generally considered accurate to continental level and not much more.
- Inaccuracy stems from comparison of individual samples to modern-day populations, rather than historical ones. DNA doesn't have nationality.
- Will vary between test companies.
- Be especially wary of estimates in the "trace" range.

Tips for evaluating your DNA matches:

- Build up your family tree as much as possible through documentary research!
- Evaluate closest matches first, seeking common surnames or places of origin.
- Attempt to determine how you must be related to a match by examining matches shared "in common with."
- Testing multiple relatives (oldest first) will allow you to use "in common with" more effectively to narrow the focus in determining a relationship.
- Use the Shared cM tool with Probabilities at DNA Painter (link below) to predict possible relationships between you and your match.
- Contacting matches and collaboration is the key to determining your relationships!

Other useful links:

Ethics, guidelines and standards for genetic genealogy:
https://isogg.org/wiki/Ethics,_guidelines_and_standards

How to do an Ancestry DNA test without spit:

http://anotherteenmom.com/reviews/how-to-do-an-ancestrydna-test-without-spit/ (Author shares her tested method for collecting DNA via cheek swab and combining with homemade artificial saliva solution.)

Autosomal DNA Testing Comparison Chart:
https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart

Autosomal DNA statistics: https://isogg.org/wiki/Autosomal_DNA_statistics This article and the preceding one come from the ISOGG Wiki. ISOGG is the International Society of Genetic Genealogy and their Wiki (https://isogg.org/wiki/Wiki_Welcome_Page) is replete with articles that are helpful for gaining a conceptual foundation for genetic genealogy. You can search by topic, and they have a whole page dedicated resources for beginners in genetic genealogy:
https://isogg.org/wiki/Beginners%27_guides_to_genetic_genealogy

GEDmatch: https://www.gedmatch.com/login1.php GEDmatch offers tools for analyzing your data (such as a chromosome browser, which Ancestry lacks) and "toys" (e.g. a tool for predicting your eye color based on your DNA), and allows you to compare your DNA results with results from people who have tested with other companies.

Promethease: https://promethease.com/ Promethease offers medical information based on analysis of your SNPs.

DNA Painter: https://dnapainter.com Allows you to quickly and easily create color-coded chromosome maps showing which segments were inherited from different ancestors, based on data imported from matches in GEDmatch, MyHeritage, 23&Me, and FTDNA, but not Ancestry. Also offers interactive shared cM tool with probabilities. https://dnapainter.com/tools/sharedcmv4

Some helpful articles from top genetic genealogy blogs:
Good explanation of atDNA testing coverage using multiple relatives:
https://www.legacytree.com/blog/introduction-autosomal-dna-coverage

On ethnicity estimates: https://dna-explained.com/2013/10/04/ethnicity-results-true-or-not/
http://www.legalgenealogist.com/2017/04/16/still-not-soup/

Roberta Estes' article on phasing from her blog "DNA eXplained": https://dna-explained.com/2016/04/06/concepts-parental-phasing/

Roberta Estes' article explaining IBD, IBC and IBP matches: https://dna-explained.com/2016/03/10/concepts-identical-bydescent-state-population-and-chance/

Blaine Bettinger's summary of sites to which you can upload your DNA results:
https://thegeneticgenealogist.com/2013/09/22/what-else-can-i-do-with-my-dna-test-results/

Leah Larkin's summary of which companies accept autosomal DNA transfers from which other companies: http://thednageek.com/whats-new-in-autosomal-dna-transfers/

One of many articles explaining privacy issues and use of DNA by law enforcement in the wake of the Golden State Killer case: https://dna-explained.com/2018/04/30/the-golden-state-killer-and-dna/

Explanation of GEDmatch Genesis: https://blog.kittycooper.com/2018/12/time-to-move-to-genesis/

And a couple articles from my own genealogy blog:

https://fromshepherdsandshoemakers.com/2019/01/15/dna-testing-for-the-scientifically-challenged/
https://fromshepherdsandshoemakers.com/2019/02/22/using-your-dna-match-list-at-ancestry/
https://fromshepherdsandshoemakers.com/2019/03/03/ancestrys-new-thrulines-utility-needs-more-work/