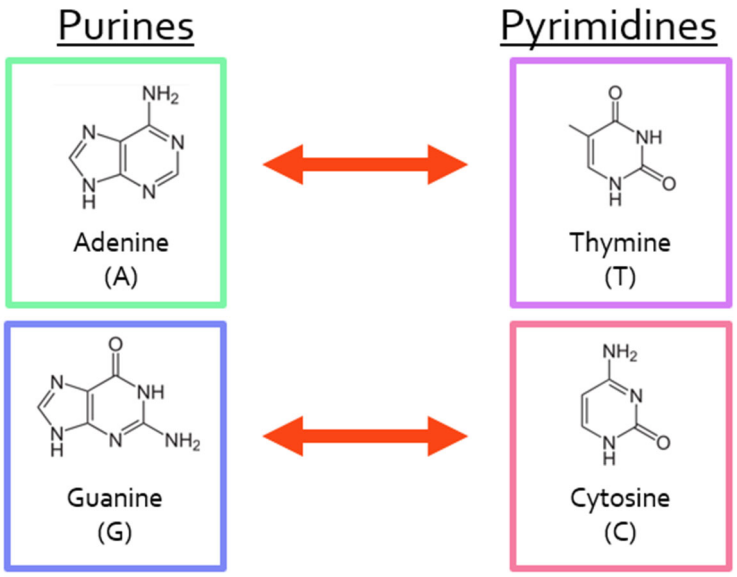
Using DNA Testing for Your Research

Jim Brewster
jim.brewster@genebygene.com

I. Shape of DNA
   a. DNA is formed by two curling strands connected with base pairs like rungs on a ladder.
      i. Referred to as the "forward strand" and the "reverse strand" for reporting purposes.
   b. **Nucleotide Bases** - The four chemicals that form the rungs of the ladder. Usually abbreviated as A, C, T, and G.



      i. **Purines** – From a chemical standpoint, these chemicals have two rings. They are A and G.
      ii. **Pyramidine** – From a chemical standpoint, these chemicals have one ring. They are T and C.
   c. **Base Pairs** – Each rung is made of two nucleotides, with one on each strand. Together, this is a "base pair."
   d. **Chargaff's Rule** – Nucleotides are <u>always</u> paired with A and T together and C and G together.
      i. Because of this constant, each pair is referred to only by the nucleotide on the forward strand. For example, AT is reported as A, while TA is reported as T.

II. Reference Sequence
   a. A DNA sequence used to compare all people in order to identify differences.
   b. **Location** – A numbered position on the chromosome containing a base pair.
   c. **Value** – The nucleotide found on the forward strand at a specific location.

d. **Build** – A standardized set of values for each location on the chromosome for us to compare changes. This varies depending on DNA type.
e. **Mutation** – A change in the value of a nucleotide at a specific location.

III. Mutations
a. Changes in the value of a position when compared to a reference sequence. For example, if Position 1 has a value of A in the reference sequence and you have a T, this is a mutation.
b. Mutations do not affect the function of the cell in most cases but are valuable to genealogists because they are passed on to children.
c. **Ancestral** – The original base state at a location on the chromosome. No mutation has occurred, and the tester is negative for the SNP.
d. **Derived** – A change from one base to another at a location on the chromosome.
e. Tracking mutations in different people allows us to determine how loosely related they are.

IV. Autosomal DNA
a. A random 50/50 mix of DNA from both of your parents.
b. **Random Recombination** - Picture your mom as a deck of cards, and your dad as a deck of cards. Shuffle the two decks together, take half, and that is your DNA. Shuffle it again and that is your siblings, so each child starts from the same deck, but gets a different shuffle.
c. **Centimorgan** – A unit of measurement for DNA. Equal to a length of base pairs that have a 1% chance of containing a mutation in a single generation. This can vary in size depending on location in the chromosome but averages about 1 million base pairs.
d. **Longest Block –** An unbroken block of centimorgans shared between two people. In the card analogy, think a line of Ace of spades, 2 of spades, 3 of spades, and so on.

V. Autosomal Matches
a. People who share enough centimorgans with you that they share a common ancestor within a genealogical timeframe.
b. More centimorgans shared between two people means a closer relationship.
c. Provides a range, and not an exact number of generations to common ancestor, because the same kinship can can vary depending on the shuffle. For example you may share more centimorgans with a first cousin than your full sibling does.
d. Matches can share ancestry on any of your family lines within approx. 5 or 6 generations. After 5 or 6 generations, the amount of DNA from any given ancestor is so small it is hard to distinguish one ancestor from another.

VI. Ethnicity report
a. **Reference Populations** – Groups of people that have lived in a region for many generations without admixture from other regions. Can be used as a reference population to identify mutations unique to that area.

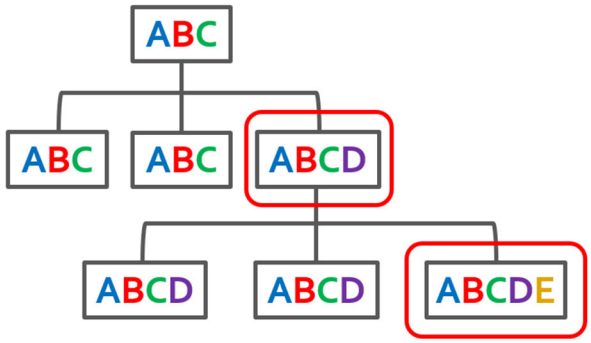
b. **Percentages** – Provides a breakdown of what portions of your DNA are shared with reference populations to provide a percentage chart of your ancestry from all lines.
c. **Caveats** –
    i. Because of random recombination, you may have an ancestor from a reference population whose DNA did not survive the many shuffles.
    ii. The reference population might not be inclusive of the entire region.
        1. Eg. Native American – Many groups of indigenous Americans existed throughout the continent and had variations from one tribe to another. It is hard to find reference populations from the entire continent that have not had any European admixture.
    iii. No matter where a group of people live, if you go back far enough, they came from somewhere else. This means every reference population shares some DNA with the neighboring populations. Science does its best to prevent overlap, but it can occur. It is always improving.

VII. Y-DNA
a. One of two sex chromosomes that determines a person's genetic sex.
b. Only males have the Y chromosome. Fathers pass it to their sons, but not daughters.
c. Two main types: Y-STRs and Y-SNPs

VIII. Y-STRS
a. Stands for Short Tandem Repeat – An area of the Y chromosome where the same pattern of base pairs repeats over and over. Eg. ATCGATCGATCG.
b. Each STR is designated with a letter/number combination such as DYS319. The number of times the sequence repeats is the value of that STR for a person.
c. A repeat can be gained or lost in an STR from one generation to another. This does not happen every generation, but changes in values accumulate over time.
d. **Genetic Distance (GD)** – The number of differences in STR values between two people.
e. **Used for matching** - Can compare different numbers of markers. The more markers compared, the more accurate the closeness of relationship can be inferred.

IX. SNPs
a. SNPs do not occur every generation, and when a SNP occurs, it is unlikely to revert to the ancestral value in later generations.
b. SNPs tend to be unique to specific populations or lineages.
c. By studying groups of people around the world, we can see which SNPs are shared by different groups of people or individuals and which are unique to a specific group.
d. Unique SNPs help us understand when the two groups split apart.

i. For example, if Groups A and B both share the values at positions A, B, and C, but differ at position D, then we know that the common ancestor to both groups had A, B, and C, but one of the groups developed mutation D after it split.



e. These groups are called **Haplogroups**.

X. Haplogroups
   a. SNPs shared by a group of people are called haploid groups or haplogroups for short.
   b. Major haplogroups define large groups of people and are assigned a letter, such as R, G, I, etc.
   c. Each haplogroup is defined by a unique SNP.
   d. When one person in a haplogroup is positive for a SNP while another is negative for it, a new branch is formed within that haplogroup.
   e. Each branch can have many sub-branches as more SNPs are discovered. Each subgroup is named after the SNP that identifies it. For example, R-M269, I-M253, etc.
   f. Some SNPs are found in multiple different haplogroups, and this indicates that they arose independently in different lines. These are not added to the tree as they cannot be put in a single branch.

XI. Mapping the time and place of haplogroups
   a. Archaeological remains can help place haplogroups in a time and place. The age of the remains can be determined using techniques such as radiocarbon dating to determine how old they are. Then if we sequence that DNA, we know that any SNPs we find must be at least that old.
   b. We compare this ancient DNA to modern-day people to determine the rate of change for SNPs. For example, if we find ten differences between a modern-day test taker and remains that are 1,000 years old, then a change occurred every 100 years on average.
   c. Studying differences in haplogroups can help us determine when they split apart, and we can trace these breadcrumbs all the way to our common ancestors.

d. Haplogroups can be thousands or tens of thousands of years old, and the oldest haplogroups and remains that we have found are in Eastern Africa.

XII. mtDNA
   a. Everyone has mtDNA, but only mothers pass it to children. This means that it traces the direct maternal line in the same way that Y-DNA traces the direct paternal line.

XIII. mtDNA Matching
   a. Connects to living people who share direct maternal ancestry. It mutates very slowly, so it can go back very far, up to 52 generations.

XIV. mtDNA Haplogroups
   a. Traces to a theoretical "Mitochondrial Eve," the original mtDNA that survived through many generations.

XV. Pros and cons of testing
   a. Autosomal
      i. Pros
         1. Can connect to living relatives from any ancestral line.
         2. Finds relatives in a genealogical time frame.
      ii. Cons
         1. Does not specify which ancestral line a match connects to you.
         2. Cannot trace ancient ancestry.
   b. Y-DNA
      i. Pros
         1. Every match is known to be from the paternal line.
         2. Can help with surname research.
         3. Provides haplogroups.
         4. Can stretch far back in time.
      ii. Cons
         1. Only males can take this test.
         2. May be difficult to find male relatives to test who can trace this line.
         3. Matches may be outside of a genealogical time frame.
   c. mtDNA
      i. Pros
         1. Every match is known to be on the direct maternal line.
         2. May be difficult to find male relatives to test who can trace this line.
         3. Matches may be outside of a genealogical time frame.

XVI. Testing options
   a. 23andMe
      i. Provides autosomal matching and health reporting.
      ii. Provides autosomal matching.

iii. Provides upper level mtDNA and Y-DNA (for males) haplogroups but not matching.

b. Ancestry DNA
   i. Worlds largest autosomal matching database.
   ii. Provides upper level mtDNA and Y-DNA (for males) haplogroups but not matching.

c. myHeritage –
   i. Large overseas matching database.
   ii. Testing provided by FamilyTreeDNA and provides similar results.

d. FamilyTreeDNA –
   i. Provides full range of testing options .
   ii. Only site with Y-DNA and mtDNA matching databases.
   iii. Provides mtDNA and Y-DNA haplogroups.

XVII. Transfer options

a. GED Match
   i. A free to use autosomal matching database. Does not test but accepts data uploads from any testing vendor.

b. mitoYDNA
   i. A free-to-use matching database where you can upload mtDNA or Y-DNA results from any vendor.

c. FamilyTreeDNA
   i. Accepts autosomal transfers from 23andMe, Ancestry DNA, and MyHeritage. Free to upload with additional tools for a one-time fee.

XVIII. Resources

a. FamilyTreeDNA Help Center – A variety of topics on genetic genealogy with a section devoted to mitochondria.
b. Group Projects – Free to join collaborative efforts that focus on a variety of genealogical topics. These are run by volunteers and can connect you to experts and other researchers with similar goals.
c. ISOGG wiki – A useful encyclopedia of terms and resources
d. Eupedia – The historical and anthropological history of all mtDNA and Y-DNA Haplogroups.
e. DNAeXplained – An blog on a variety of genetic genealogy topics.