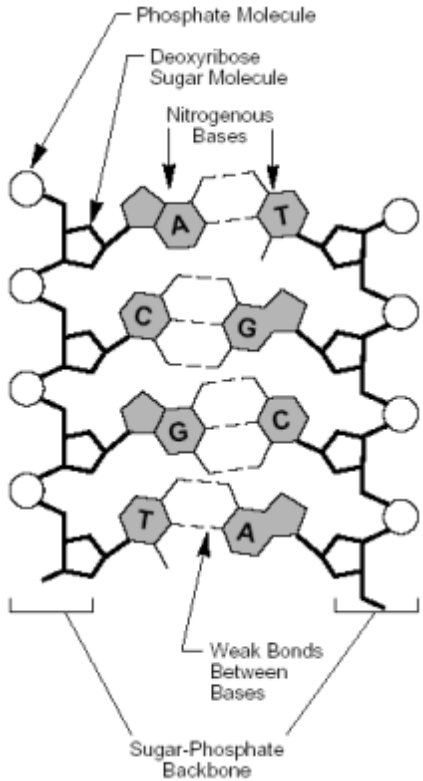
DNA and Genealogy

Charlottesville LDS FHC
2007 Family History Conference

Presenter: Matthew Orton

What is DNA?

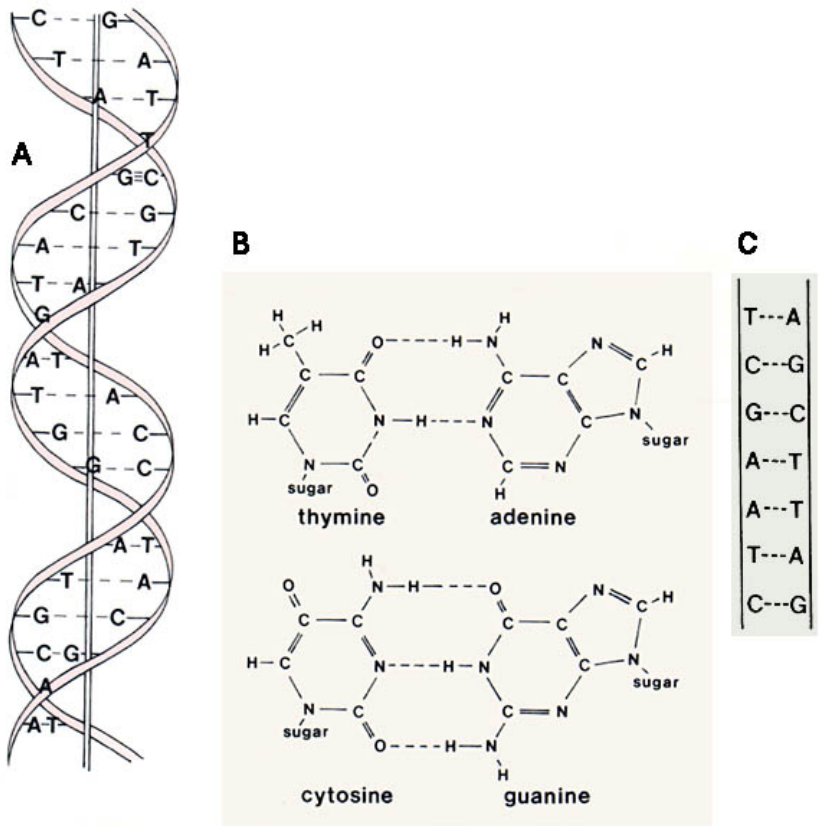
- DNA → **D**eoxyribo**N**ucleic **A**cid
- DNA contains all of the genetic instructions for the development and functioning of living organisms.
- DNA is the long-term storage of information in a cell and is often described as the set of blueprints for each cell.
- DNA is found in all “living organisms.”



DNA

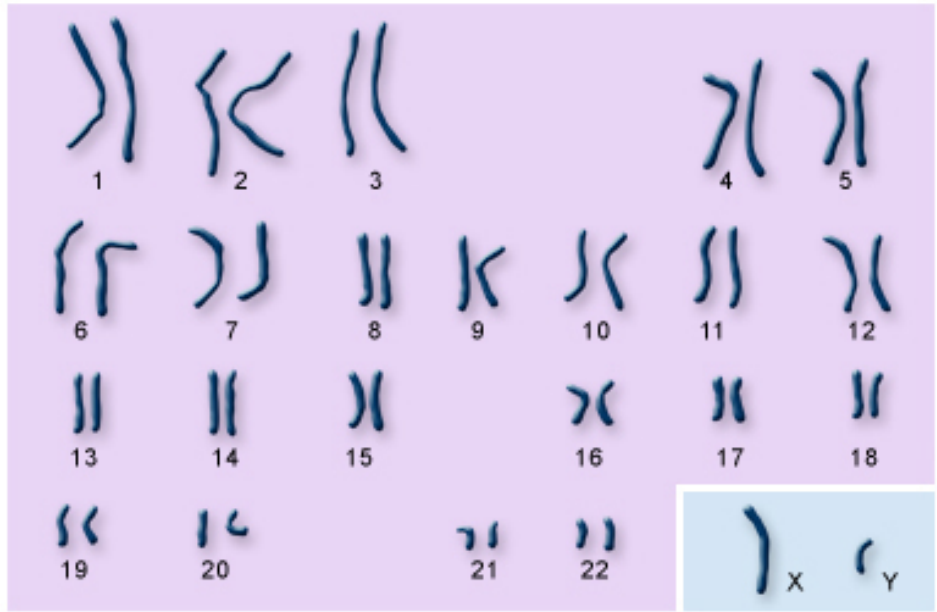
- DNA is made up of 4 nucleotide bases.
  - Adenine (A)
  - Cytosine (C)
  - Guanine (G)
  - Thymine (T)
- Nucleotides are paired together.
  - A – T
  - C – G
- How is DNA represented

D 5’ AGTCAGCGTGCAGT 3’



DNA

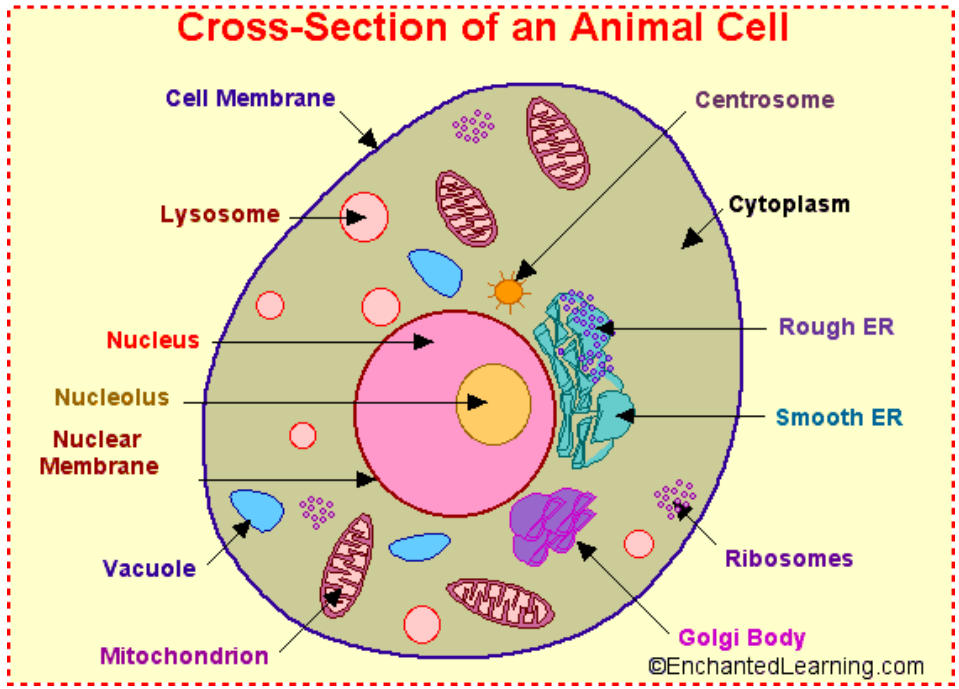
- Each human cell (aside from red blood cells and gametes) contains a full set of 46 chromosomes.
  - 44 Autosomal Chromosomes
  - 2 Sex Chromosomes
- 23 chromosomes are inherited from each parent.
  - 22 Autosomal
  - 1 Sex chromosome



U.S. National Library of Medicine

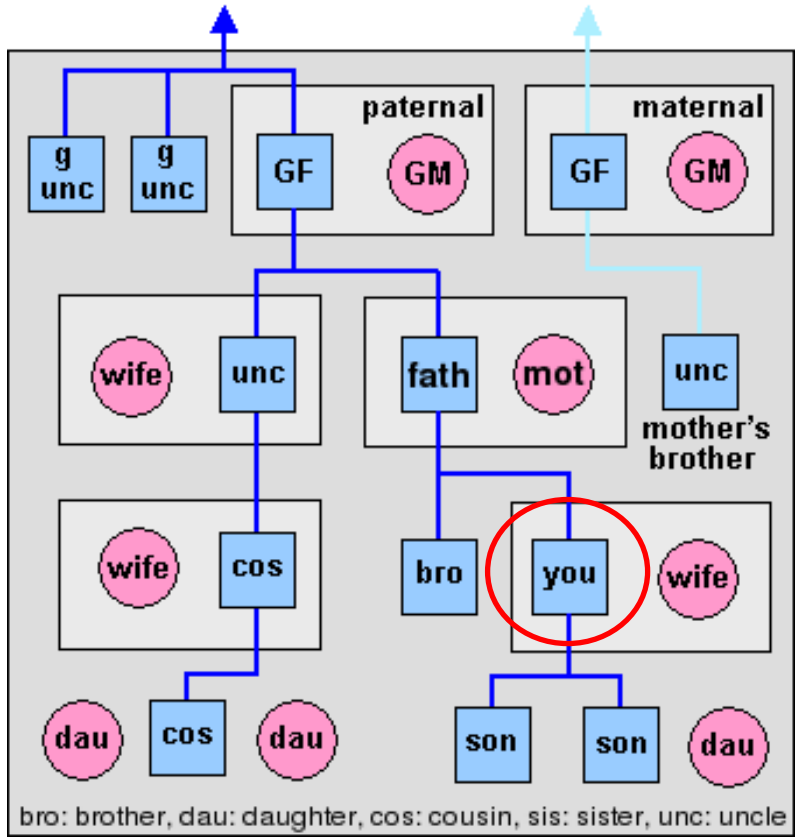
DNA

- DNA is found in most cells in the human body and for genealogy is classified into three types
  - Y-Chromosome DNA (Y-DNA)
  - Mitochondrial DNA (mtDNA)
  - Autosomal DNA



Y-DNA

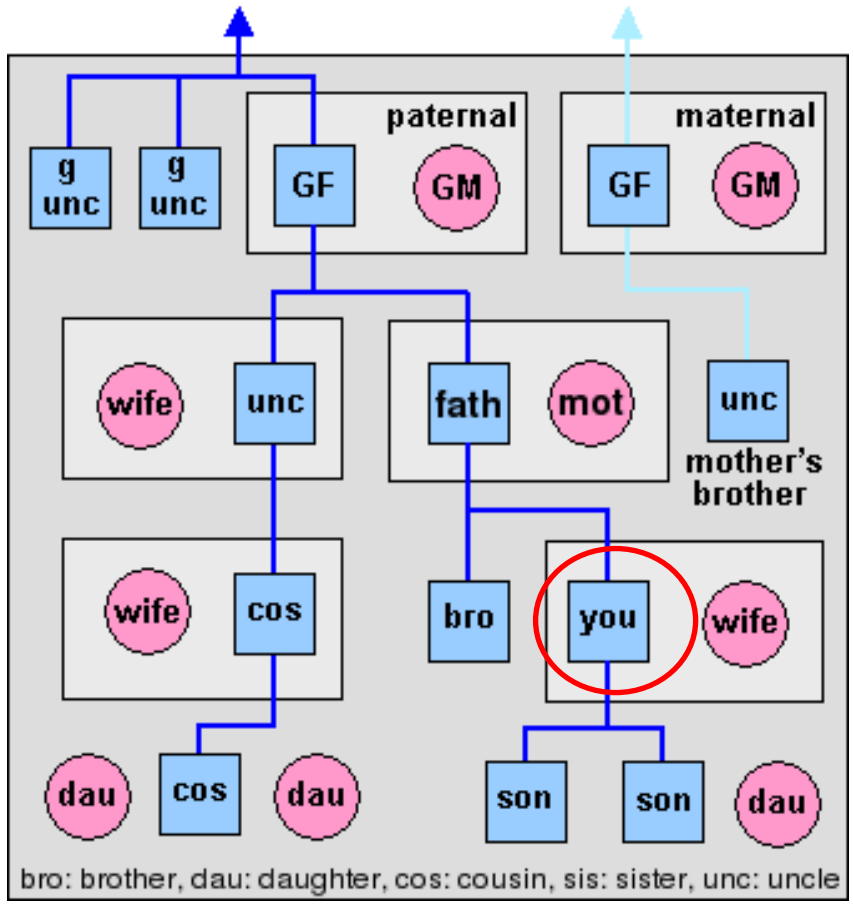
- **Y-DNA** is a type of DNA that is only carried by men, who inherit it from their fathers. This means that males with a common paternal ancestor have similar Y-DNA



All males connected by blue line have common Y-DNA

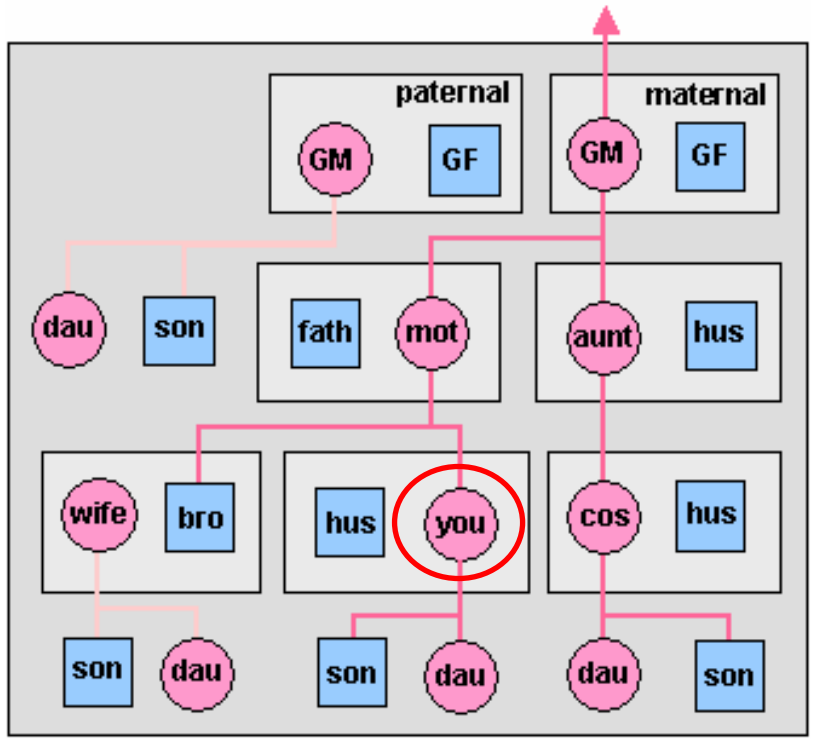
Y-DNA

- Y-DNA is particularly useful for tracing one's direct paternal line (father, paternal grandfather, etc.) because it changes slowly from generation to generation, and in most societies, the surname of the father is also inherited by his sons.

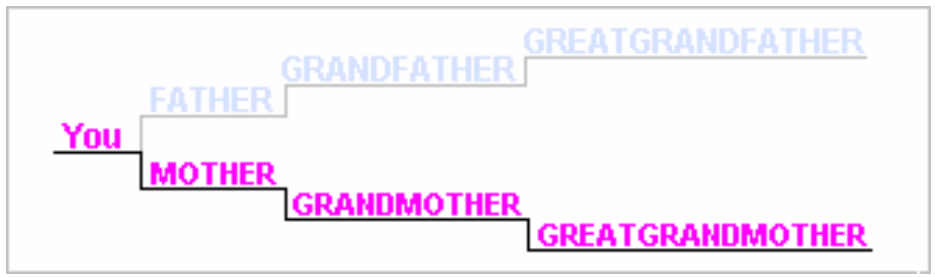


mtDNA

- **mtDNA** is a type of DNA carried by both males and females, but is only inherited from the mother. This makes mtDNA useful for tracing one's direct maternal line (mother, maternal grandmother, maternal great-grandmother, etc).



All persons connected by pink lines have common mtDNA



Autosomal DNA

- **Autosomal DNA** is the type of DNA responsible for most physical characteristics, such as height, eye color, etc. Autosomal DNA is inherited by sons and daughters from both parents (and from all four grandparents, etc). Currently Autosomal DNA is not used extensively for genealogy.

How Can DNA Identify Us?

- Mutations and the shuffle of maternal and paternal genes through sexual reproduction, ensure that each member of a species (except identical twins) has a unique DNA sequence.
- The ideal way to distinguish an individual from all the other people on Earth would be to describe the entire sequence of nucleotides in his or her DNA.
- However, since each human genome (all the DNA in a person's chromosomes) is made up of **more than 3 billion nucleotide basepairs**, describing an individual's complete DNA would be far to complicated and expensive to be practical!

How Can DNA Identify Us?

- Only about **five percent** of human DNA is actually thought to code for traits. Most of the rest is made of long, apparently nonfunctional, stretches of nucleotide basepairs (sometimes referred to a "junk" DNA.)
- Within these nonfunctional stretches are short, moderately repetitive base pair sequences. The number of repeats is inherited and is easily detectable making them ideal identifying markers.

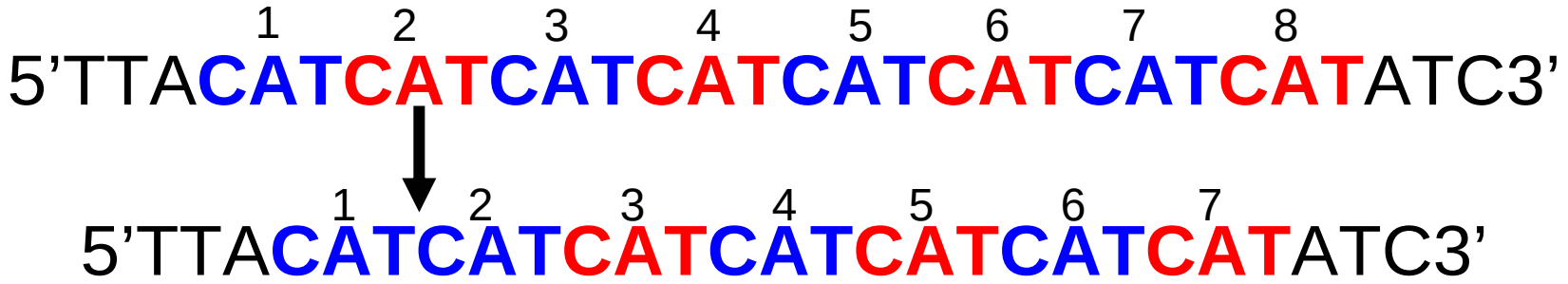
1 2 3 4 5 6 7 8

TTA**CATCATCATCATCATCATCATCAT**ATC

- In this example the nucleotides “CAT” are repeated 8 times.

How Can DNA Identify Us?

- The number of repeating units can occasionally change during evolution and descent.



- They are thus useful markers for familial relationships and have been used in paternity testing, forensic science and in the identification of human remains.

How Can DNA Identify Us?

- There are two types of these repetitive sequences. **VNTRs** (variable number tandem repeats) are repeated sequences that typically range from 10 to 80 bps. These occur fairly frequently in the human genome but there are relatively few different types.
- Short tandem repeat (**STR**) sequences (sometimes called microsatellites) are much shorter (2-10 bps) and may be repeated as many as 100 times at a given location on a chromosome. **The human genome contains hundreds of thousands of these STRs all evenly distributed on all the chromosomes.**

How Can DNA Identify Us?

- In 1997, the FBI announced the selection of 13 STR markers to be used in forensic investigations. If any two samples of DNA obtained from different sources (say a crime scene and a suspect) have matching numbers of repeats at all 13 markers, it is virtually certain they are from the same person.

So How Is This Used for Genealogy?

- The location of these STRs are known and the number of repeats can be analyzed.
- These unique locations are called "markers".
- Markers are identified with specific nomenclature.
  - Chromosomal markers are described by D (number of chromosome or X or Y) S (Numbers the signify a place on the chromosome).
    - DYS392 or D21S458
  - Mitochondrial DNA (mtDNA) is a continuous circle of 16569 genetic bases, each appearing at a distinct marker location. Locations are designated by number from 00001 to 16569.

So How Is This Used for Genealogy?

- Since each marker represents a specific location on a chromosome we can use this to report the number of repeats at a specific STR.

1 2 3 4 5 6 7 8

TTA**CATCATCATCATCATCATCATCAT**ATC

- Note there are eight repeats of the segment CAT. The number of repeats is the "value" that is shown on a DNA test report for the marker.
- In this example, a lab report would show DYS391=8 for this marker.

Haplotype

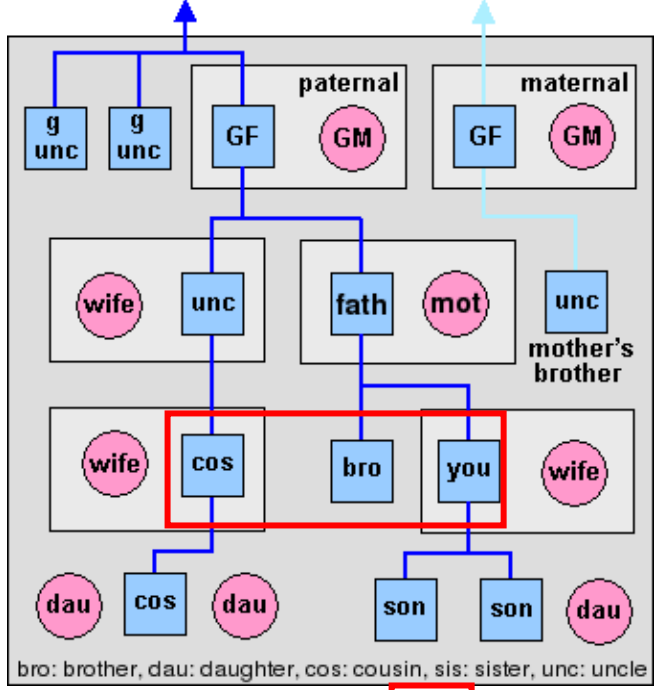
- DNA test reports typically show a series of markers and their corresponding values. These results are referred to as a "haplotype." For instance, the sequence would be called a 12-marker haplotype.

DYS391 **12**	DYS456 **15**	DYS876 **13**	DYS124 **12**
DYS234 **29**	DYS654 **22**	DYS938 **10**	DYS336 **11**
DYS765 **12**	DYS243 **16**	DYS569 **11**	DYS700 **15**

So How Is This Used for Genealogy?

- In some ways, DNA marker values are like telephone numbers.
- Consider that the same seven-digit telephone number, 428-1040, might appear in both Boston and Miami. However, adding more numbers ("area codes") allows us to distinguish between regions. The same thing is true about DNA results.
- If we compare a limited number of DNA markers then it's possible for two individuals to have the same marker values, yet not be closely related. The FBI uses 13 markers to identify an individual most molecular genealogy recommend using 36 markers for genealogy work.
- Testing for more markers helps avoid this possible ambiguity. In general, the more markers tested, the easier it is to distinguish individuals and family tree branches.

So How Is This Used for Genealogy?



Using our example, let look at “you”, bro”, and “cos” who have the following Y-DNA haplotype:

YOU: 11 14 12 13 29 13 12 15 12 12 13 12 12 12 14 25 19 **30**
BRO: 11 14 12 13 29 13 12 15 12 12 13 12 12 12 14 25 19 **30**
COS: 11 14 12 13 29 13 12 15 12 12 13 12 12 12 14 25 19 31

- Note that YOU and BRO have the same haplotype, but COS has a difference of one marker value (31 instead of 30).
- That difference would have been due to a mutation that occurred in his Y-DNA (or his father's), but not in the other cousins.
- In general, the greater the number of mutations we find between two individuals, the further in the past their common ancestor lived.

Review Slide

- Males who share a common paternal ancestor will have virtually the same Y-chromosome DNA.
- Everyone will have virtually the same mtDNA as their maternal ancestors.
- We use the word "virtually" since occasionally there are small changes or "copy errors" that might occur with each descendant.
- Those copy errors are called "mutations" and are generally harmless, but are useful for tracing one's direct paternal or maternal line.

What is Currently Available for DNA and Genealogy?

- There are many companies who will collect DNA samples and run your DNA for many different markers.
- However, **there are few databases** that will allow you to compare your DNA to DNA of other people.
- The largest and the one I am most familiar with is run by the **Sorenson Molecular Genealogy Foundation (www.smgf.org)**. The SMGF have both a Y-DNA and mt-DNA database open to the public.

DNA Databases

- Most databases are research databases that are used to help determine how to use DNA for genealogy as well as use the information collected to research population genetics.
- Population genetics is a field of research using DNA to determine how groups of people are related, where they came from and how they have moved over centuries.

DNA Databases

- All Databases, as mentioned are research oriented and love to have people submit their DNA and a extensive pedigree chart.
- At the **Sorenson Molecular Genealogy Foundation** it works by you submitting a DNA sample and a four generation pedigree chart which is entered into the database.
- This is all free but will not help you directly with your genealogy because you will not receive the haplotype of your DNA.

DNA Databases

- All databases have a privacy policy like this one:

  To protect privacy, we do not reveal any personal information of our participants, or the names of any ancestors born after 1906. You may find references to ancestors in SMGF pedigree charts on message boards and in collections of family trees.

- So you will not see any names of possible common ancestors born after 1906. But the names before 1906 are open to the pubic. This allows you to use the pedigree charts that others have submitted to augment your own.

DNA Databases

- To fully use the database you must know what your haplotype is.
- To do this you must pay to have your haplotype determined.
- There are many labs willing to do this at prices ranging from $95-$300 depending on the number of markers etc.
- Once you have your haplotype you may use the databases to search by surname and haplotype.

DNA Databases

- Currently the **Sorenson Molecular Genealogy Foundation** database will let you search by surname alone.
- It will bring up all the pedigree charts that people have submitted with a specific surname and will allow you to search through them.
- The database is more powerful when used with both haplotype and surname.

Not Perfect, But Promising Possibilities

- Molecular genealogy is based on probabilities, and like forecasting the weather, is not an exact science. It can provide important clues for family history research but traditional genealogy methods continue to be an important part of molecular genealogy.
- As databases become larger Molecular Genealogy will become a more powerful tool in genealogy work.

Links to Check Out

- www.smgf.org
  - This web site is the most user friendly and does a great job of explaining what is available and how to become involved.
  - It has links to all of the databases currently available.
  - It has links to all of the companies who will analyze your DNA and give you your haplotype and coupons for certain companies.