DNA and its Place in Family History Research -Introduction-

Sheldon Jones

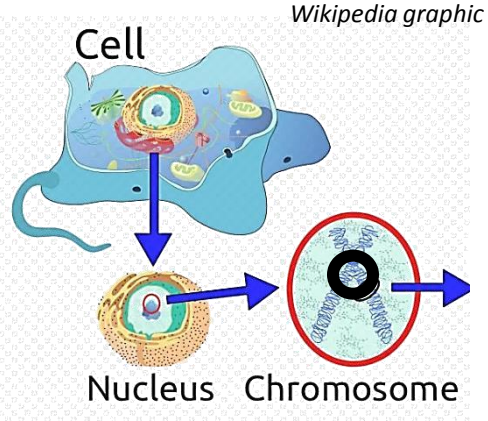
(srjones07@gmail.com)

07 May 2022

Outline

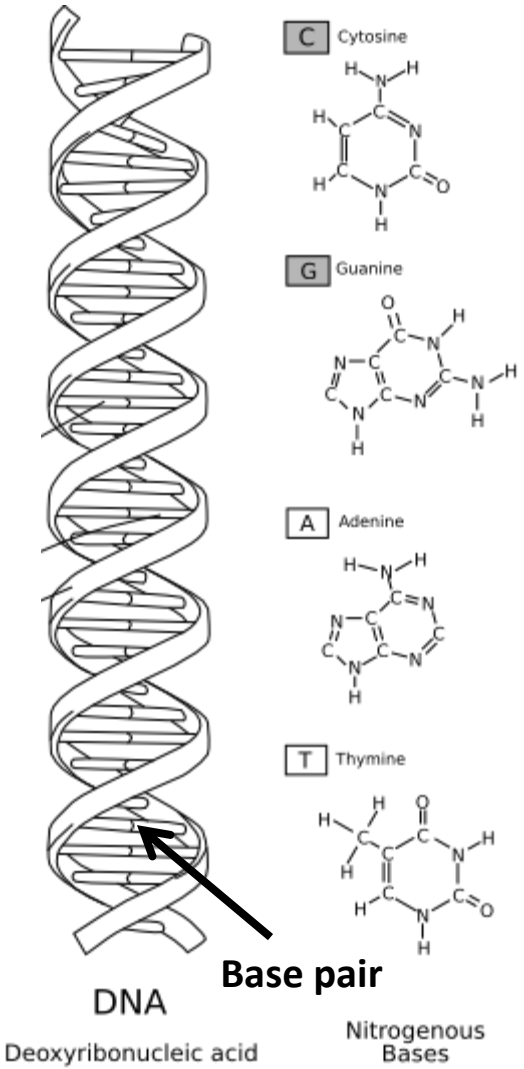
- **What is DNA?**
- **DNA Terms (test result reporting and social)**
- **DNA in Genealogy Research**
- **DNA Testing**
  - SNP mapping examples
  - Test types in use today
  - DNA testing companies & comparison
  - Can company databases be compared?
  - Consanguinity & Correspondence with DNA
- **Conclusion**
- **Triangulation Examples (recent)**

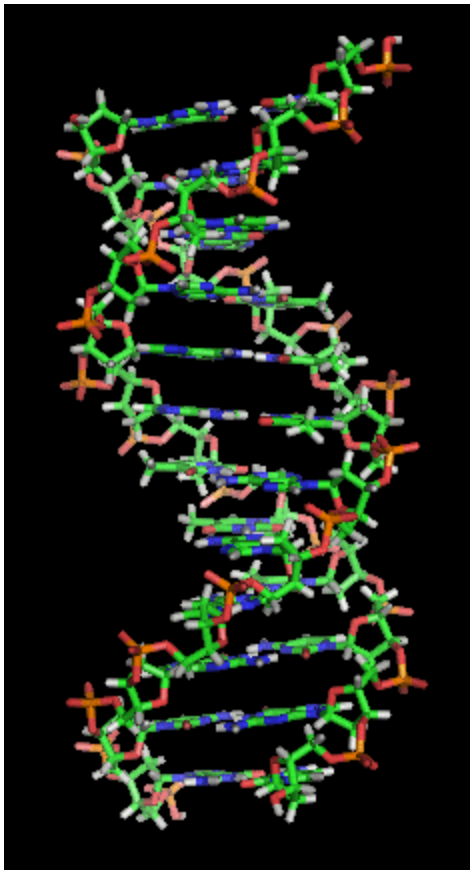
What is DNA? (1)



Wikipedia graphic

- DNA is what makes up each of the 23 Chromosome pairs within each nucleus in each cell in our bodies
- A pairs with T,
  C pairs with G

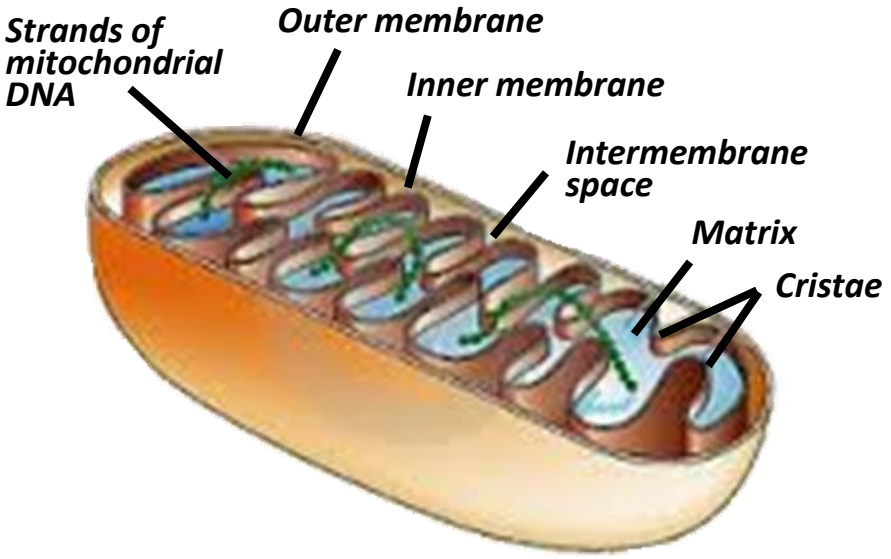




What is DNA? (2)

- Within the cell nucleus:
  - 22 Chromosome pairs (autosomes)
  - 1 sex chromosome pair: ("normally" XY = male, XX = female)
    - ✓ Males get their Y from their father; X from mother
    - ✓ Females get an X from mother and father (crossover can occur)

- We also have mitochondria in our cells which have DNA (mtDNA)
  - we get all our mtDNA from our mothers



A Mitochondrion within a cell

http://www.rajaha.com/description-structure-mitochondria-function

DNA Terms – Testing (1)

- MRCA – Most Recent Common Ancestor
  - "Estimated number of generations to MRCA"
- Y-STR – Standard Marker test for short tandem repeats (STRs) of a sequence in a region on the Y-Chromosome; "value" stated is the allele count for the marker (usually # repeats for simple markers)
  - *Two men matching 111 Y-STR markers may still be 8th to 9th cousins*
- SNP - Single Nucleotide Polymorphism (pronounced "snip") are single mutations in a short sequences of base pair values within a section on a chromosome; e.g.: AAGCCTA to AAGCTTA
  - *SNP testing is goal of large scale Genome testing*

DNA Terms – Testing (2)

- **Haplotype** - A collection of SNPs per chromosome or a collection of Y-STR values (for Y-DNA analysis) that tend to be inherited together
- **Haplogroup** – a group of similar haplotypes with a common ancestor determined by specific SNPs; Primarily defined for paternal (Y-DNA) and maternal (mtDNA) descendancy, e.g., "*Adam*" starts the paternal "A" haplogroup and "*Eve*" starts the maternal "L" haplogroup.
- **Illumina DNA sequencing/genotyping chips**:
  - AncestryDNA: Omniexpress v2 to 2019? Now GSA?
  - 23andMe: Global Screen Array (GSA) v5 Aug 2017+…
  - FamilyTreeDNA: changed to GSA April 2019 (they process MyHeritage kits as well)

DNA Terms – Testing (3)

- **cM (centiMorgan)** – in Autosomal DNA research, a measure of how frequently a DNA segment is exchanged , or *recombined*, when a sperm/egg is created. Humans have 46 chromosomes in 23 pairs. When males produce sperm and women produce eggs, only 23 chromosomes (half) are passed to the sperm/egg by each parent. The child produced will receive 23 ***recombined*** chromosome segments from each parent. The length of each segment passed from the parent to the sperm/egg may be complete, half, etc…, or zeroed out completely.

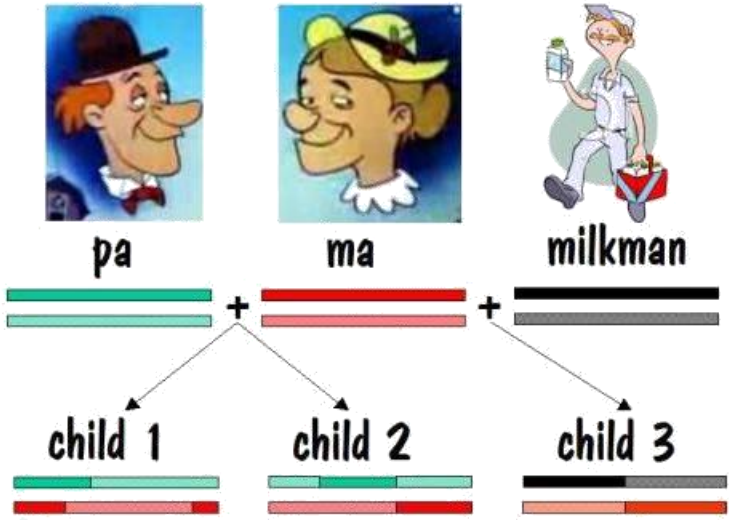
In practical terms, using current technology, there are ~7500 cM available to be shared (ancestryDNA says to use ~7000 cM for theirs) (e.g., full siblings share about 50% or 3750 cM)

DNA Terms - Social

- **NPE** – Non-Paternal Event; surname of child does not match Y-Chromosome biology – adoption, unmarried mother passing maiden name, extra-marital affair, etc… Estimate is 5-10% likelihood each generation
- **OOW** – Out Of Wedlock; child born to unmarried mother
  - ➢ 18% of US births in 1980, 30% 1990, ~40% 2007-2015

https://www.cdc.gov/nchs/data/databriefs/db162.htm

(NOTE: DNA Testing companies have a disproportionate number of NPE/OOW individuals in their databases!! – how much more than ~5-10% ???)



http://www.stevemorse.org/genetealogy/beyond.htm

DNA in Genealogy Research

- DNA directly stores information about our ancestry
  - Our chromosomes are a combination of our parents' chromosomes; their chromosomes are combinations of their parents, etc.
  - The combination is random (see cM description), however certain regions *'usually'* get passed down as a unit (genes are an example); Note: if a segment gets passed down exactly over multiple generations – MRCA wrong!
  - Mutations can occur during the copying/replicating process producing sperm/egg– this is critical to DNA testing!
    - ✓ Y Chromosome mutations: STR and SNP tests
    - ✓ Autosome and X chromosome mutations: SNP tests

Chromosome SNPs *(pronounced "snips")*

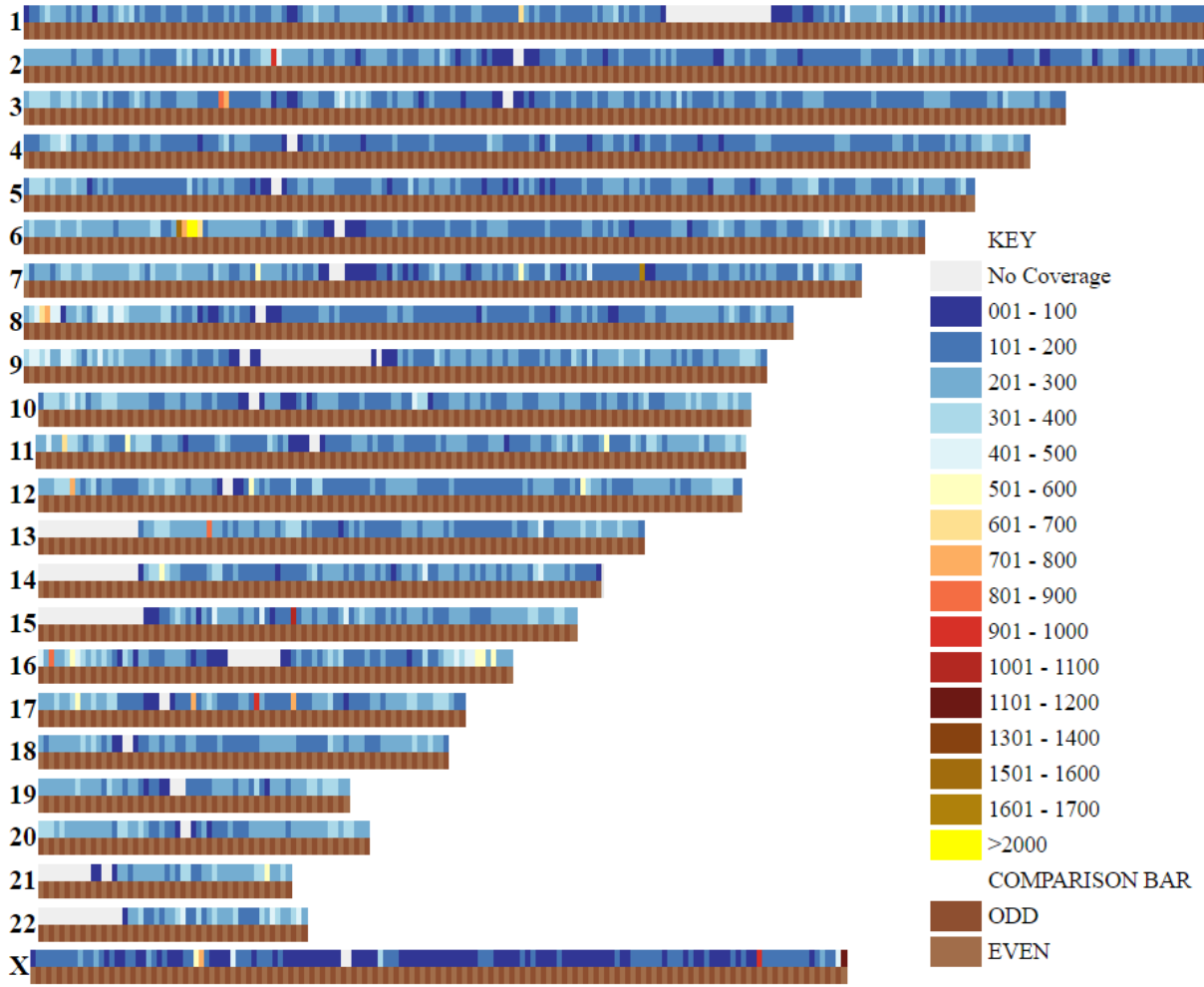
- Single Nucleotide Polymorphisms (SNPs) are short sequences of base pair values within a section on a chromosome

Source: 23andme.com (note: #SNPs is for older V3 chip)

Total Human Genome: ~3 Billion base pairs

DNA Terms – SNP example

- Single Nucleotide Polymorphisms (SNPs) : 23andMe V4 chip (~602K SNPs)



Source: http://haplogroup.org/23andme-v4-chip/

Ranges are SNP counts per 1M base pairs; Total Human Genome: ~3B base pairs

DNA Testing Types - Current

- **Y-DNA STR and SNP testing** can help determine ***paternal*** ancestry over many generations *for males only*
  - *STR testing usually done first with a standard set of markers; specific Y-SNP testing usually done next*
  - *Big-Y, Y Elite, …*
- **mtDNA SNP testing** can determine ***maternal*** ancestry over many generations for *both males and females*
  - *due to low mutation rate, not very useful within 10+ generations*
- **Autosomal testing** (and actually X sex chromosome too), allows *both males and females* to determine ***relationship with ~6-7 generations***
  - *Current technology, i.e., number of DNA locations sampled and min segment length considered significant, limits the ability to match others to ~4^{th}-5^{th} cousinship at best*
- **Whole/Full Genome testing** *(how to handle 100 to 300 Gb data!)*
  - *Below $1K, Nebula Genomics will do 30x sequencing depth for $199!*

Major DNA Testing Companies

- **FamilyTreeDNA.com** (FTDNA) - Y-STR (37/~~67~~/111), Y-SNP (Big-Y/Individual SNPs), mtDNA (~~HVR1+HVR2~~/Full), autosomal V3 chip (Family Finder – no Y-SNP), National Geographic Genographic Project
- **23andme.com** – autosomal "customized" GSA V5 chip (includes X, Y, and mtDNA SNP testing)
- **dna.ancestry.com** – autosomal; Blackstone private equity. Note: now, each kit must be registered under its own unique account (e-mail).
- **myheritagedna.com** – autosomal, Israel based (not available in Israel(!),France,Poland, Alaska); V3 chip

DNA Company Comparison- Autosomal

Company	23andMe	Family Tree DNA's Family Finder test	Ancestry.com's AncestryDNA test
Website	www.23andme.com *(56 countries)*	www.familytreedna.com *(worldwide)*	http://dna.ancestry.com *(~35 countries)*
Price (as of May 2022)	US: $99 ancestry;$199 +health; xtras $30/yr	$79	$99 (routinely reduced to $59 now) in the U.S. (subscription required to access some features)
Upload of raw data file allowed from other companies	No	23andMe V3, V4 & V5; ancestryDNA Omniexpress V2 &? newer GSA? ; MyHeritage	No
SNP chip used for testing	Illumina GSA "V5"	Illumina Omniexpress V3, now GSA?	Illumina GSA (as of late 2019?)
Number of autosomal SNPs tested	630,132 V5 (pre-Dec 2013 ~967K V3, pre-Aug 2017 ~577K V4)	708,092 ← *V3 chip – ~612K filtered*	682,549 ← *V2 chip; ~637K filtered*
Number of Y chromosome SNPS	3733 V5 (pre-Dec 2013 3089 V3, pre-Aug 2017 2329 V4)	None	1691 (labeled chromosome 24)
Number of X chromosome SNPS	16,530 V5 (pre-Dec 2013 26,087 V3, pre-Aug 2017 19,487 V4)	18,091	28,892 (labeled as chromosome 23), plus 440 SNPs labeled as chromosome 25 that are either from chromosome X or from the pseudoautosomal regions of the Y chromosome
Number of mitochondrial DNA SNPS	4318 V5 (pre-Dec 2013 2737 V3, pre-Aug 2017 3154 V4)	None	263 (2022), 195(2018), used to be none
Number of people in the database (as of Feb 2019)	12,200,000+	~1,800,000	21,000,000+

www.myheritagedna.com also does testing (~5.6M people!!) – based in Israel, but not allowed in Israel ?

From: http://www.isogg.org/wiki/Autosomal_DNA_testing_comparison_chart

DNA Testing - Comparing Databases

- Comparison web sites – Autosomal:
  - www.GEDmatch.com – upload and compare autosomal data (T# (randomized #) = FTDNA FF; M# = 23andme (Note: V5 chip new testing requires a different algorithm); A# = AncestryDNA)
  - FTDNA will accept the older 23andme.com V3 (Nov 2010 - Nov 2013) raw data and they will accept AncestryDNA – into their database. Also MyHeritage because they process it.

- Comparison web sites – Y Chromosome:
  - Y-SNP & whole genome: www.yfull.com – accepts www.familytreedna.com Big-Y500 and Big-Y700; www.fullgenomes.com Y-Elite; Nebula www.nebula.org
- Other helpful site:
  - www.isogg.org – International Society of Genetic Genealogy
  - (many, many others)

Consanguinity (Blood Relationship)

Common Ancestor	Child	Grandchild	Great Grandchild	2nd Great Grandchild	3rd Great Grandchild	4th Great Grandchild	5th Great Grandchild	6th Great Grandchild	7th Great Grandchild
Child	Sibling	Niece/Nephew	Grandniece / Grandnephew	Great Grandniece / Grandnephew	2nd Great Grandniece / Grandnephew	3rd Great Grandniece / Grandnephew	4th Great Grandniece / Grandnephew	5th Great Grandniece / Grandnephew	6th Great Grandniece / Grandnephew
Grandchild	Niece/Nephew	First Cousin	First Cousin, 1r	First Cousin, 2r	First Cousin, 3r	First Cousin, 4r	First Cousin, 5r	First Cousin, 6r	First Cousin, 7r
Great Grandchild	Grandniece / Grandnephew	First Cousin, 1r	Second Cousin	Second Cousin, 1r	Second Cousin, 2r	Second Cousin, 3r	Second Cousin, 4r	Second Cousin, 5r	Second Cousin, 6r
2nd Great Grandchild	Great Grandniece / Grandnephew	First Cousin, 2r	Second Cousin, 1r	Third Cousin	Third Cousin, 1r	Third Cousin, 2r	Third Cousin, 3r	Third Cousin, 4r	Third Cousin, 5r
3rd Great Grandchild	2nd Great Grandniece / Grandnephew	First Cousin, 3r	Second Cousin, 2r	Third Cousin, 1r	Fourth Cousin	Fourth Cousin, 1r	Fourth Cousin, 2r	Fourth Cousin, 3r	Fourth Cousin, 4r
4th Great Grandchild	3rd Great Grandniece / Grandnephew	First Cousin, 4r	Second Cousin, 3r	Third Cousin, 2r	Fourth Cousin, 1r	Fifth Cousin	Fifth Cousin, 1r	Fifth Cousin, 2r	Fifth Cousin, 3r
5th Great Grandchild	4th Great Grandniece / Grandnephew	First Cousin, 5r	Second Cousin, 4r	Third Cousin, 3r	Fourth Cousin, 2r	Fifth Cousin, 1r	Sixth Cousin	Sixth Cousin, 1r	Sixth Cousin, 2r
6th Great Grandchild	5th Great Grandniece / Grandnephew	First Cousin, 6r	Second Cousin, 5r	Third Cousin, 4r	Fourth Cousin, 3r	Fifth Cousin, 2r	Sixth Cousin, 1r	Seventh Cousin	Seventh Cousin, 1r
7th Great Grandchild	6th Great Grandniece / Grandnephew	First Cousin, 7r	Second Cousin, 6r	Third Cousin, 5r	Fourth Cousin, 4r	Fifth Cousin, 3r	Sixth Cousin, 2r	Seventh Cousin, 1r	Eighth Cousin

Consanguinity vs. % Autosomal DNA

- % Autosomal DNA in common between two people:
  - *NOTE: statistics are +/- these values; e.g., Siblings share ~40-~60% DNA (half-siblings are half this); 1st Cousins share ~7% to ~18%; 2C are ~1% to ~5%; 3C ~0% to ~2% (10-15% of 3rd cousins <u>don't share any DNA</u>!)*

Start	parent	grandparent	1st great-grandparent	2nd great-grandparent	3rd great-grandparent	4th great-grandparent	5th great-grandparent	6th great-grandparent	7th great-grandparent
parent	Sibling 50.000%	25.000%	12.500%	6.250%	3.125%	1.563%	0.781%	0.391%	0.195%
grandparent	25.000%	1st Cousin 12.500%	6.250%	3.125%	1.563%	0.781%	0.391%	0.195%	0.098%
1st great-grandparent	12.500%	6.250%	2nd Cousin 3.125%	1.563%	0.781%	0.391%	0.195%	0.098%	0.049%
2nd great-grandparent	6.250%	3.125%	1.563%	3rd Cousin 0.781%	0.391%	0.195%	0.098%	0.049%	0.024%
3rd great-grandparent	3.125%	1.563%	0.781%	0.391%	4th Cousin 0.195%	0.098%	0.049%	0.024%	0.012%
4th great-grandparent	1.563%	0.781%	0.391%	0.195%	0.098%	5th Cousin 0.049%	0.024%	0.012%	0.006%
5th great-grandparent	0.781%	0.391%	0.195%	0.098%	0.049%	0.024%	6th Cousin 0.012%	0.006%	0.003%
6th great-grandparent	0.391%	0.195%	0.098%	0.049%	0.024%	0.012%	0.006%	7th Cousin 0.003%	0.002%
7th great-grandparent	0.195%	0.098%	0.049%	0.024%	0.012%	0.006%	0.003%	0.002%	8th Cousin 0.001%

Conclusion

- Y-DNA can used to determine paternal ancestry via comparison with other males
  - watch-out for NPEs! – surname may be different!
- Autosomal DNA can be used to verify/validate close (~5th cousin or better) relationships; can be used for paternity/maternity/sibling verification
- Records based Genealogy still very important!
  - *DNA cannot really stand on its own; would need many, many samples to do so*
- DNA may be only hope for adoption/illegitimacy/etc. to obtain biological ancestry

NPE Search Examples Using Triangulation

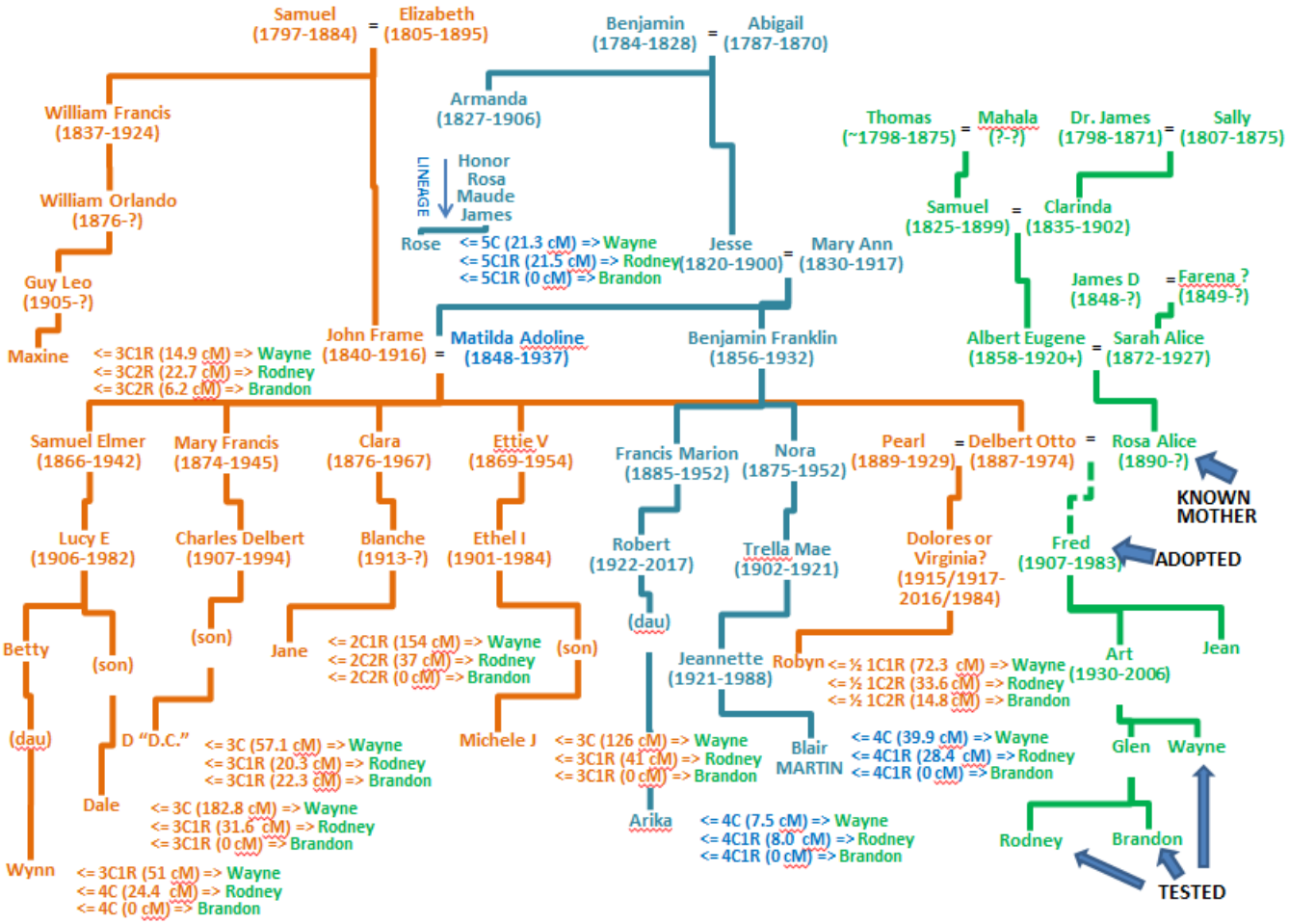
- If *paternal* line is unknown: Is *tester :*
  - Male? : 1st: Y-DNA to hopefully determine surname; 2nd Autosomal
  - Female? (or Y-DNA inconclusive) : Autosomal test
- If *maternal* line is unknown: Autosomal test
- Goal of triangulation is to separate matches into lines known (if any) and unknown – Ancestry finally makes this "easy"

AncestryDNA cM mapping: (Ancestry claims their test results in ~7000 cM per person)
Group A: 2800-4200 cM (40%-60%; 50% ctr) -> full siblings, parent-child
Group B: 1400-2100 cM (20%-30%; 25% ctr) -> half-sibling, Aunt/Uncle-Niece/Nephew; double 1st cousin; Grandparent-Grandchild
Group C: 490-1050 cM (~7% - ~18%; 12.5% ctr) -> 1C, half Aunt/Uncle-Niece/Nephew; GG-greatgrandchild; GAunt/GUncle-GNiece/GNephew
Group D: 210-630 cM (~3%-~9%; 6.25% ctr) -> 1C1R, 2GGrandparent-2GGrandchild, half 1st cousin; Half GAunt/GUncle-GNiece/Gnephew
Group E: 70 – 350 cM (~1%-~5%; 3.125% ctr) -> 2C, 1C2R, ½ 1C1R, 3GGrandparent-3GGrandchild
Group F: 35 – 210 cM (~0.5%-~3%; 1.563% ctr) -> 2C1R, ½ 2C, 1C3R, ½ 1C2R, …
Group G: 0 – 140 cM (0% - ~2%; 0.781% ctr) -> 3C, 2C2R, 1C4R, … (10-15% of 3rd cousins do not share any DNA)
Group H: 0 – 28 cM (0% - ~0.4%; 0.391% ctr) -> 3C1R, … (~50% of 4th cousins do not share any DNA)

Example: Paternal Search A (1)

- Fred was adopted in 1910 in Kansas – bio mother "known" (DNA matches not shown for her family)
- Rodney did Y-DNA and Autosomal; Brandon and Wayne did Autosomal tests
- *FTDNA Y-DNA test proved "surname" (very common name!)*
- AncestryDNA autosomal tests with quite a bit of work, proves who bio father "was" -> DNA matches on father's father's and mother's sides…

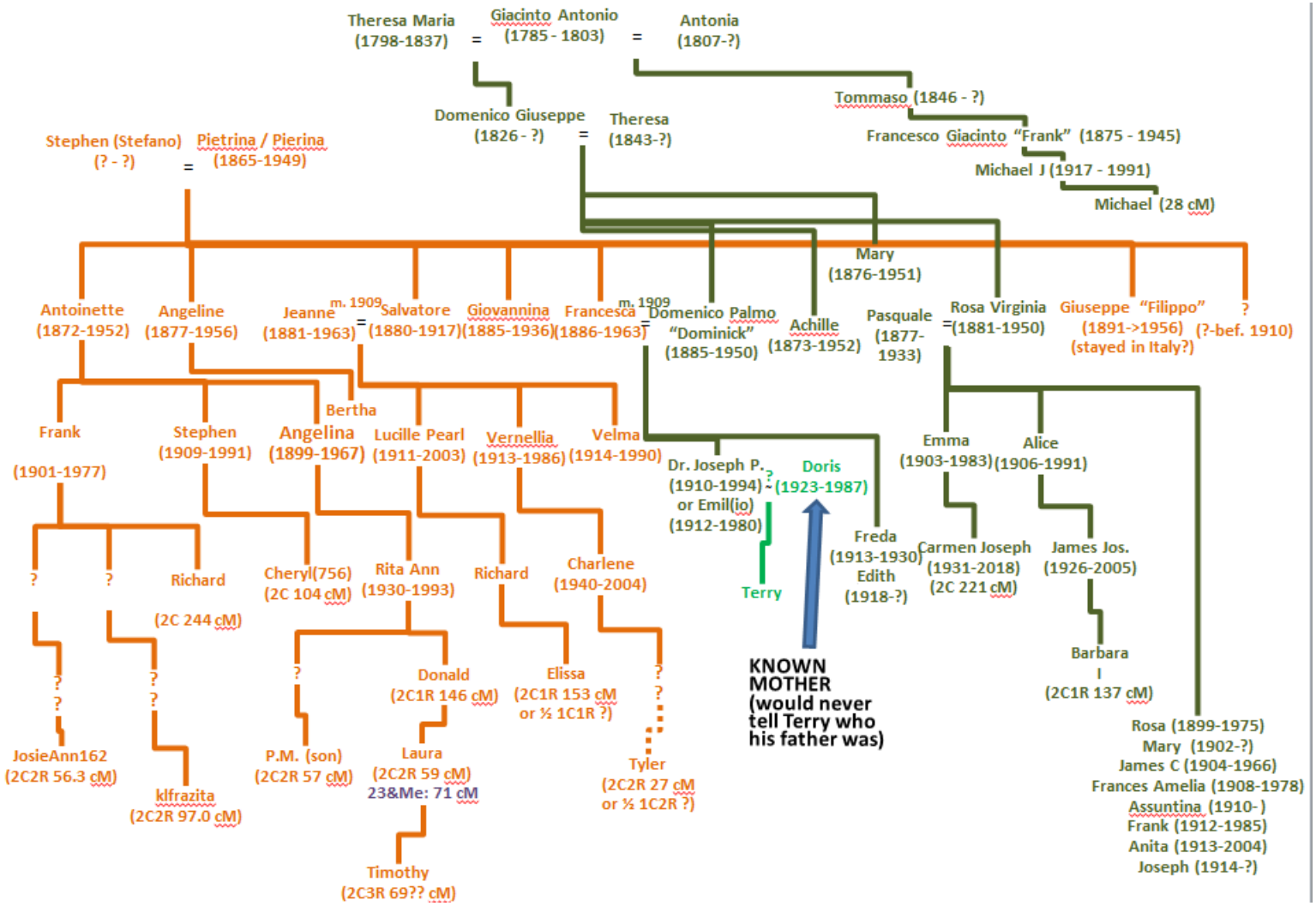
Example: Paternal Search A (2)



Example: Paternal Search B (1)

- Terry did Y-DNA and had his “brother” do Y-DNA and found out he wasn’t his dad’s son… (no match with “brother” and also family knowledge that he wasn’t surname)
- *Y-DNA totally inconclusive – closest matches were also NPEs. (Surname is Italian and “no male” has been tested… trying to get one to get tested!)*
- 23andMe – very confusing results
- AncestryDNA finally helped solve the mystery…at least down to one of two brothers was his father. Apparently “neither” had other children?

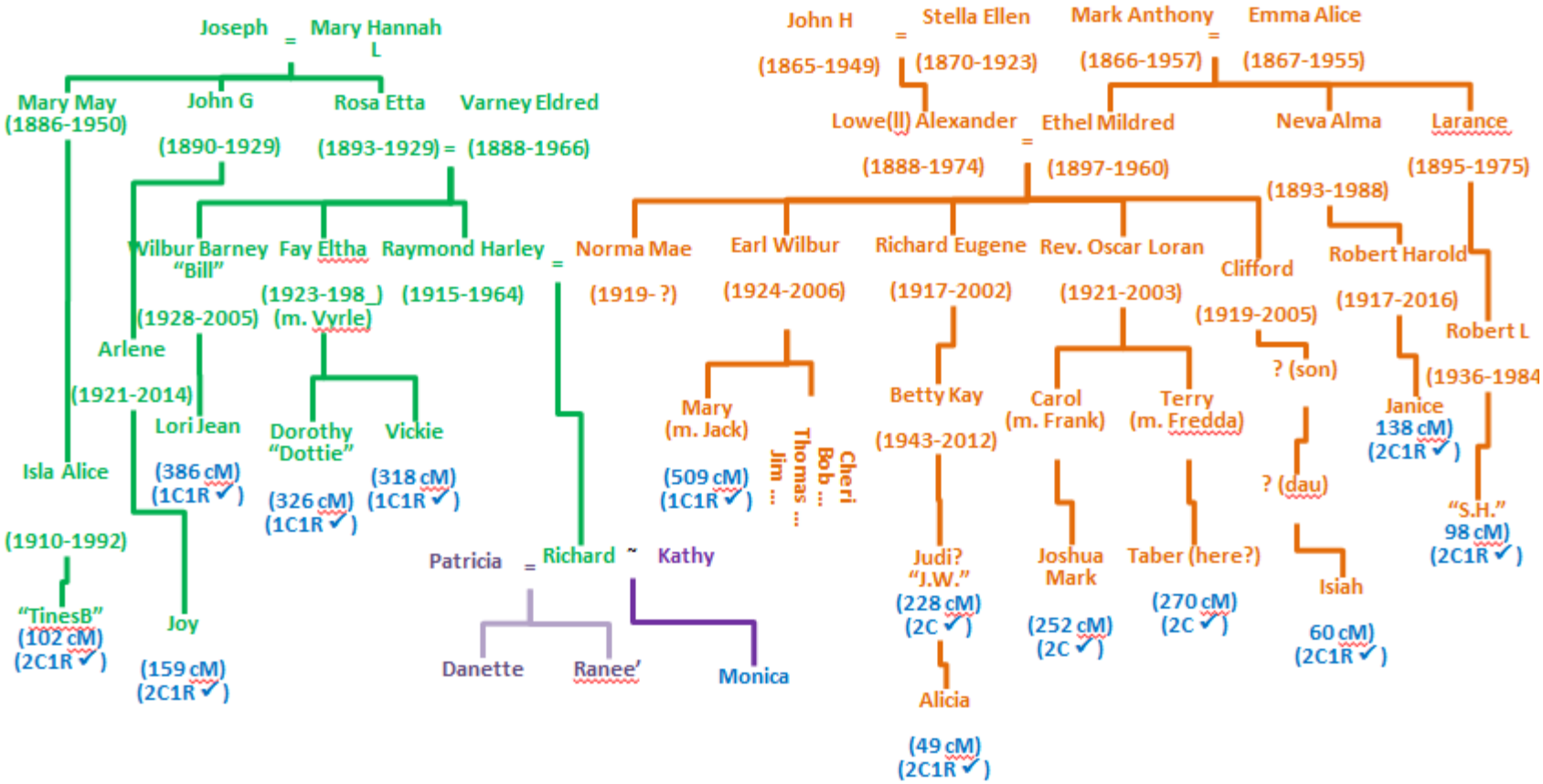
Example: Paternal Search B (2)



Example: Paternal Search C (1)

- “We” have always known my niece, Monica, was not my brother’s… Finally did AncestryDNA to prove it and figure out who her bio father was. Her maternal Aunt and Uncle “know” – but family still won’t acknowledge that “we” now know who her father is..
- She is in touch with the two half-sisters we know about now… sadly, father has Parkinson’s
- Ranee’ asked how we could figure this out – without her father’s DNA !!

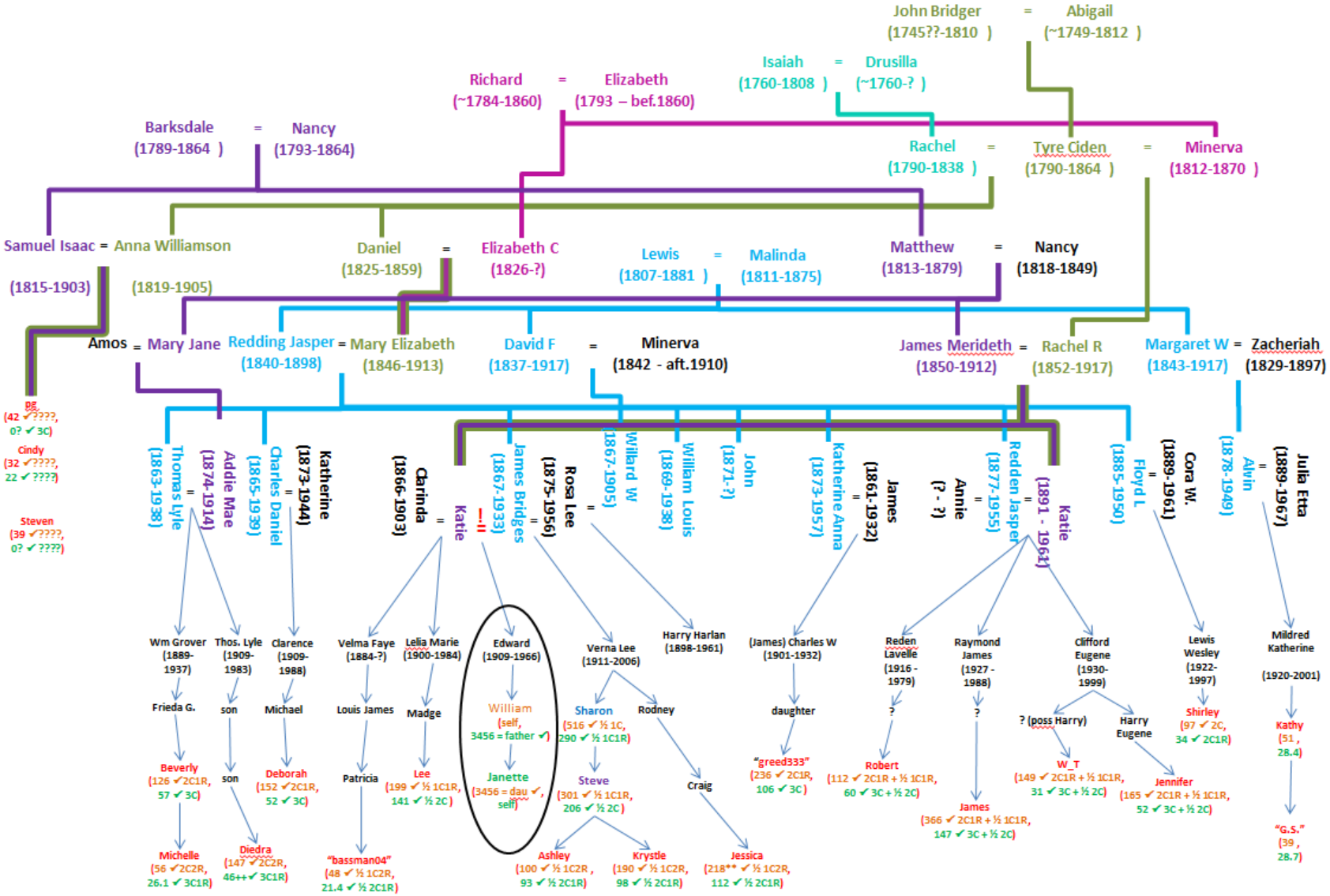
Example: Paternal Search C (2)



Example : Full Bio Parent Search (1)

- Very complicated – due to people related to each other multiple ways => you “add” up each cM range!
- Edward was Adopted in northern Illinois; his son killed him in 1966
- Another son, William, has a daughter, Janette, who wanted to figure out who Edward’s parents were – they were “told” never to look because if it ever came out – it would be big trouble…
- Turns out their roots are in southern Illinois…
- They have now met some of their cousins – all is well

Example : Full Bio Parent Search (2)



Example : Full Bio Parent Search (3)