FamilySearch

Using Color to Effectively Cluster AncestryDNA™

Suzanne “Suzi” Trotter-Martin, AG®
FamilySearch International

This class focuses on using color to organize your analysis as you evaluate, sort, and cluster your AncestryDNA matches. The process explained here is one way to use your matches to solve research problems using clues from your DNA, which is called **genetic genealogy.**[i]

ESSENTIAL BACKGROUND

1. We inherit our **DNA**, our genetic code, from our biological ancestors. We do not equally inherit DNA from all our ancestors. (Some will be high; others will be low.)
2. DNA alone can only prove a parent-child relationship; all other genetic relationships require DNA combined with traditional, documentary research.
3. **Autosomal DNA**: [ii]
    a. The Genetic code carried in our non-gender chromosomes can be used to verify relatives back between 5 and 7 generations.
    b. Sections of our DNA can be identified as matching exactly with other relatives.
    c. **Recombination** is responsible for changes in segment size from generation to generation. It is the reason your DNA differs from your siblings.
4. **Y-DNA:** [iii]
    a. The Y-chromosome is passed from father to son entirely intact with changes only occurring by **random mutation**.
    b. Because the Y-chromosome is passed from father to son, it can be used to follow the direct patrilineal line of inheritance beyond 7 generations.
5. **Mitochondrial DNA (Mt-DNA):** [iv]
    a. The Mitochondria, an organelle inside of each of your cells, contains its own DNA.
    b. Because the Mitochondria is passed from mother to child, its DNA can be used to follow the direct matrilineal line of inheritance beyond 7 generations.
6. **AncestryDNA** is an Autosomal DNA test.

There are several types of DNA tests that are helpful for genealogical research. This presentation only addresses how to use the tools in AncestryDNA to analyze DNA results.

ABOUT CLUSTERING

- cluster: group of matches who share a subset of DNA.
- centiMorgans (cM): a measure of DNA.[v]
- segments: sections of DNA shared by individuals; the length of a segment is important to understanding how to interpret the cM.
- Most Recent Common Ancestor (MRCA): the genetic contributor (ancestor) shared (common) by matches at the closest location in their family tree (most recent).[vi]
- interrelatedness: when matches have more than one ancestral line in common. It is more general and more neutral than pedigree collapse or endogamy.[vii]

The number of shared **centiMorgans** combined with the number of **segments** shared between two matches provides context for interpreting the relationship between you and your matches. Many people attempt to **cluster** their DNA matches based on common surnames, but this is not effective. Clustering your matches by **MRCAs** creates specific intersections that allow us to determine how you are related to specific matches even when they have not linked their DNA to a tree. Clustering by **MRCA couples** adds context to half vs. full relationships.

Using color to represent specific relationships within your tree helps organize your findings and give meaning to your clusters. In a **full sort**, you seek matches who connect to each of your great grandparent couples and each of your great-great grandparent couples. A full sort is most helpful when the answer being sought is in the great grandparent generation or further. This helps when trying to identify an unknown cluster by employing the process of elimination.

SAMPLE COLOR-CODED PATTERN

Below is a chart showing one possible color-coded pattern to represent each MRCA couple in your tree gives meaning to a specific color. Related clusters can be represented by related colors. The sample color pattern provided here uses AncestryDNA dot colors aligned with the FamilySearch fan chart colors, so each dot color points to a specific relationship in your FamilySearch Family Tree. Paternal dots being teal and all blues and greens. Maternal dots being salmon and all pinks and yellows.

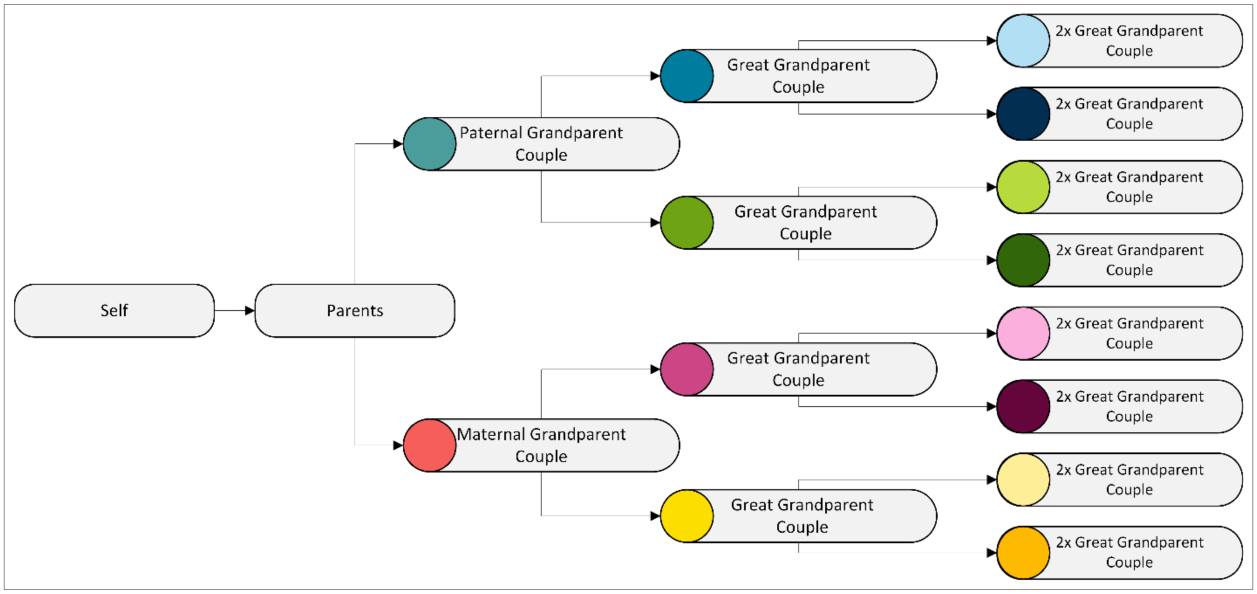

Figure 1: Color pattern based on colors in the FamilySearch fan chart.

In combination with this:

- Use dark and light teal when needed for additional generations on the paternal side.
- Use dark and light salmon when needed for additional generations on the maternal side.
- Use purple, orange, and gray for unknown clusters.

These dots are assigned at a specific step in the process that is the focus of this presentation. Establishing the meaning of the dots was necessary before taking explaining the process.

MATERIALS NEEDED TO UTILIZE THE PROCESS

- Access to AncestryDNA™ results—if you are not the **owner** of the DNA, you need **collaborator** status as someone who the DNA has been shared with.
- Access to a tree for what is known about the heritage of the person whose DNA is being studied.
- Access to DNA Painter's "The Shared cM Project 4.0 tool v4."[viii]

Note: Chrome has discontinued support for their Browser Extension "Click All Checkboxes." This tool often helps save time when applying dots to shared matches. As new tools become available, using them may be beneficial.

THE PROCESS FOR CREATING A CLUSTER

In your AncestryDNA results, select the "Matches" item from the dropdown menu. Your closest matches are at the top of your match list. If you have a parent-child match, you can use that to do your first sort.

1. Click on the name or tree hyperlinks for your match to open the DNA match.
2. Study the match's attached tree.
    a. Identify the Most Recent Common Ancestral (MRCA) couple in their tree. If the match IS your direct ancestor (For example: parent, grandparent), they are half of your MRCA couple, but through their siblings, the ancestors' parents are the MRCA couple
    b. Check that **only** this couple is shared to screen for interrelatedness.
3. Estimate the relationship between the provider of the DNA sample and their match based on this MRCA couple.
    a. Evaluate if the proposed relationship is possible based on amount of shared cM "The Shared cM Project 4.0 tool v4."[ix]
4. Create the note in AncestryDNA
    - identifying the MRCA couple
    - and the estimated relationship between the sample and the match
    - include observations about if the amount of shared cM is high or low based on the average and accepted range for the relationship
5. Create/Assign the appropriate dots
    - When a dot has already been created, you do not need to create it; however, you always need to make sure the dot has been assigned correctly to the match.
6. Go to the Shared Matches tab and add the same dots to all of the shared matches.

Repeat this process to account for each great grandparent and great-great grandparent couple. Continue further on lines beyond this generation where you are seeking to solve a research problem.

As you work through your match list, scroll to the relationship range you seek and prioritize matches who have linked trees and Ancestry™ identified Common Ancestors.

EXAMPLES OF USING THE PROCESS

Please watch the recording of this presentation to see the following examples:

1. Using the process to define the teal dot (paternal grandparents) cluster.
2. Using the process to define the green dot (paternal great grandparents) cluster.

BEYOND SECOND COUSINS

At the third cousin level (which includes 2nd cousin 1x removed, 1st cousin 2x removed, etc.), approximately 10% of your third cousins will not come up as matches in your AncestryDNA results. This creates a third cousin barrier, where not all third cousin level matches are shared matches with each other. In terms of clustering, you should use multiple matches to sort shared matches into a cluster representing 2x great grandparents and beyond.

MARGIN OF ERROR AND INTERRELATEDNESS

A margin of error is defined as the point at which false positive matches may appear. AncestryDNA specifies that their margin of error in terms of their ethnicity is 2%. Thus, you should ignore ethnicities identified as 2% or lower. In terms of matches, their cut-off for margin of error is 20 cM. This is why your initial match list will cut itself off at 20 cM. As you dig deeper, you will see shared matches below 20 cM.

When evaluated matches who share lower amounts of shared cM, it is vital that you also consider:

- How many segments those shared cM are spread across.
- If interrelatedness is causing a match to share more cMs than aligns with the closest relationship.

For example:

- If someone shares 35 cM over 35 segments, they are likely a false match. If someone shares 35 cM over 2 segments, they are likely to be a fifth cousin level match.
- If your 3x great grandparents had a brother and sister who also married, and their grandchildren or great grandchildren married, the descendant of all four parties looks like your 3rd cousin when they are really a double 4th cousin.

---

[i] "Genetic Genealogy," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[ii] "Autosomal DNA," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[iii] "Y Chromosome," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[iv] "Mitochondrial DNA," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[v] "centiMorgan," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[vi] "Most Recent Common Ancestor," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[vii] "Pedigree Collapse," ISOGG Wiki, *International Society of Genetic Genealogy* (https://isogg.org), accessed April 2025.

[viii] "The Shared cM Project 4.0 tool v4" Tools, *DNA Painter* (https://dnapainter.com), accessed April 2025.

[ix] "The Shared cM Project 4.0 tool v4" Tools, *DNA Painter* (https://dnapainter.com), accessed April 2025.