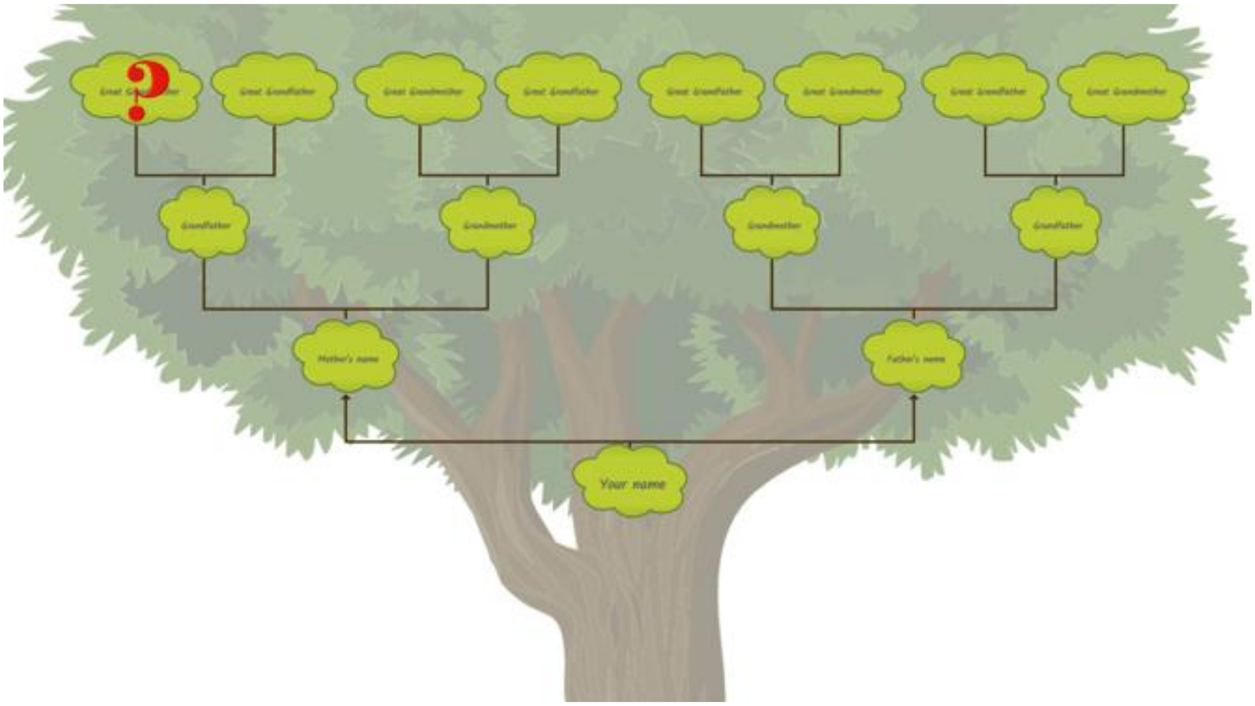
Navigating French-Canadian Endogamy and Solving Surname Mysteries Using Technology

Instructor:
Kate Penney Howard
katepenneyhoward.com
katepenneyhoward@gmail.com

Introduction



The advent of consumer DNA testing has revolutionized genealogical research, offering unprecedented opportunities to confirm family relationships and break through brick walls. However, not all DNA analysis is simple. Some of the most challenging cases involve complex inheritance patterns that can confound traditional analysis methods. This session explores advanced tools and methodologies for handling these challenging scenarios, with a particular focus on three interrelated phenomena: endogamy, pedigree collapse, and high runs of homozygosity (ROH).

Recent studies indicate that DNA testing complexity is more common than previously thought. A 2018 US study estimated that one in four people who take a DNA test discover an unexpected parentage event between themselves and their great-grandparents; approximately 25% between an individual and their closest fourteen ancestors. This high rate of complexity, combined with increasing rates of testing in endogamous populations, has created an urgent need for more sophisticated analysis tools and methodologies.

Researchers working with complex genetic patterns frequently encounter situations where standard relationship predictions fail. These difficulties often arise in cases involving:

- Endogamous populations
- Pedigree collapse
- Multiple relationship paths
- Unknown or misattributed parentage
- Historical populations with limited genetic diversity

In these contexts, traditional tools and automated relationship predictions may produce confusing or misleading results. Careful analysis, documentation, and methodological flexibility become essential.

Understanding how and why these patterns occur allows genealogists to interpret DNA evidence more responsibly and more effectively.

When DNA relationships are complex, we encounter additional difficulties:

- Traditional relationship predictions often fail when dealing with endogamous populations
- Standard analysis tools may produce confusing or misleading results
- Multiple relationship paths can create seemingly impossible amounts of shared DNA
- Privacy and emotional concerns may complicate research and reporting
- Documentation and proof standards require adaptation for complex scenarios

These challenges require genealogists to develop new skills and adapt existing methodologies to work effectively with complex DNA patterns.

Measuring DNA: Understanding CentiMorgans

Autosomal DNA testing compares genetic segments between individuals to determine whether they inherited DNA from a shared ancestor. The amount of shared DNA is measured in **centiMorgans (cM)**.

A centiMorgan is a standardized unit used to represent the likelihood that a segment of DNA will recombine in a generation. In genealogical analysis, it serves as a practical way to estimate the amount of inherited DNA shared between individuals.

In general:

- Close relatives share **larger amounts of DNA**
- More distant relatives share **smaller amounts**

DNA matches consist of **segments**, and the total amount of shared DNA is the sum of all shared segments. However, not all shared segments represent genealogically meaningful relationships.

In addition, when pedigree collapse or endogamy are a factor, individuals are often related to each other in multiple ways, which means the bit of DNA you are tracing may certainly be shared DNA from an ancestor, but it may not be the ancestor you are hoping it is.

Two different processes can produce matching DNA segments:

Identical by Descent (IBD)

IBD segments are inherited from a **common ancestor**.
These segments provide meaningful genealogical evidence.

Identical by State (IBS)

IBS segments match between two people **by coincidence (or through endogamy)**, not because they share a recent ancestor. Small segments are particularly prone to IBS matching. Because of this, researchers only consider segments above certain size thresholds.
Common guidelines include:

- Segments **under 10 cM** are often unreliable
- Segments **10–14 cM** are 90% likely to be genealogically meaningful
- Segments over 15cM are nearly 100% an indicator of a shared ancestor.

	A	B	C	D	E	F
1	Dorothy	395	🟦			
2	Sam	372		🟨		
3	Betsey	308			🟩	
4	Rudolph	279	🟦			
5	Paul	266			🟩	
6	Rudy	201			🟩	
7	Jonathan	180		🟨		
8	Elizabeth	164				🟥
9	Shelley	123	🟦			
10	Walter	98	🟦			
11	Franklin	97				🟥
12	Georgia	95			🟩	
13	Leonard	95				🟥
14	Elaine	92		🟨		
15	Hazel	91	🟦			

Ideally, when we use the Leeds Method (www.danaleeds.com) to sort our matches, we hope to see 4 distinct color clusters. I like to think of each color cluster representing a different family reunion. My parents are from different ethnic communities. My father's grandparents were immigrants, while my mother's side had been in North America for hundreds of years. So, if I were to go to a family reunion on my mother's side, I would not expect to see anyone from the Penney family. This chart represents the matches of a person from a similar situation. See, this person's matches have no overlapping colors. No one person is in two family groups.

	A	B	C	D	E
1	Dorothy	395	🟦		
2	Sam	372		🟨	
3	Betsey	308			🟩
4	Rudolph	279	🟦		
5	Paul	266			🟩
6	Rudy	201			🟩
7	Jonathan	180		🟨	
8	Elizabeth	164			
9	Shelley	123	🟦		
10	Walter	98	🟦		
11	Franklin	97			
12	Georgia	95			🟩
13	Leonard	95			
14	Elaine	92		🟨	
15	Hazel	91	🟦		

If you show fewer than 4 columns, but have no overlap, you may have no matches at the 2nd-3rd cousin level, or you may have some intermarriage or endogamy obscuring a lineage.

Endogamy: Marriage Within the Community

Endogamy occurs when individuals marry within the same cultural, ethnic, geographic, or religious community over many generations. This practice produces distinctive genetic patterns that complicate DNA interpretation.

Common endogamous populations include:

- Ashkenazi Jewish communities
- French Canadians & Acadians
- Mennonite and Amish groups
- Polynesian populations
- Some island or isolated populations

Because ancestors may appear repeatedly in family trees, individuals from endogamous populations often share DNA through **multiple ancestral pathways**.

This causes several analytical challenges:

- Individuals may share **more DNA than expected**
- Standard relationship predictions may be inaccurate
- Large numbers of matches appear related through many lines

Segments that appear to represent a single genealogical relationship may actually reflect multiple distant connections.

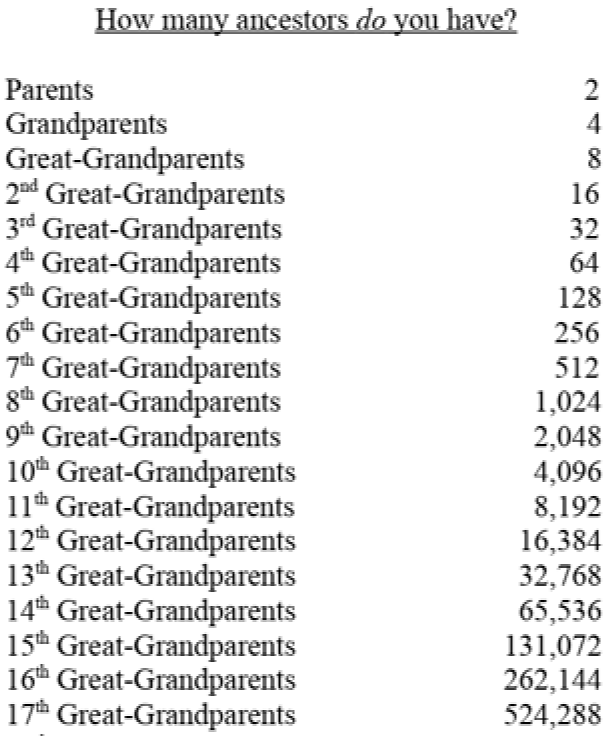
Pedigree Collapse

Pedigree collapse occurs when two individuals share a common ancestor and have children together.

This reduces the number of unique ancestors in a family tree.

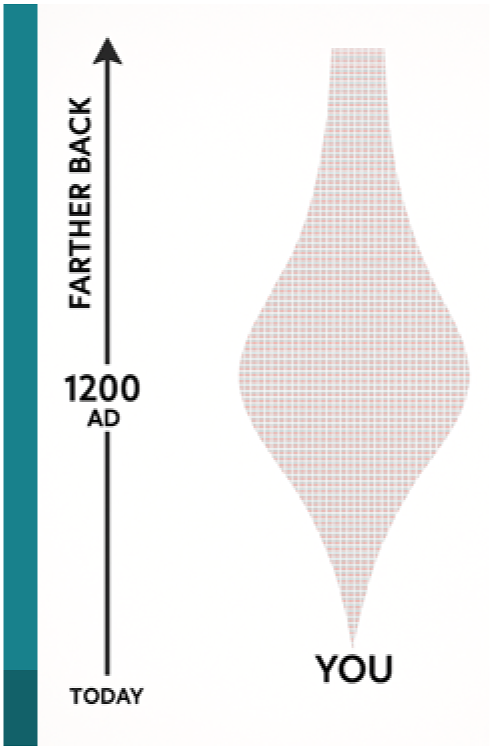
Without pedigree collapse, an individual's ancestor tree follows a binary progression: 2 parents, 4 grandparents, 8 great-grandparents, and so on.

By generation 10, you'd have 1,024 ancestors, and by generation 30, over 1 billion theoretical ancestors. This creates what genealogists call the "ancestor paradox" and by generation 30, you would theoretically have more ancestors than the world population at that time. Pedigree collapse resolves this paradox by having the same ancestors appear in multiple positions.

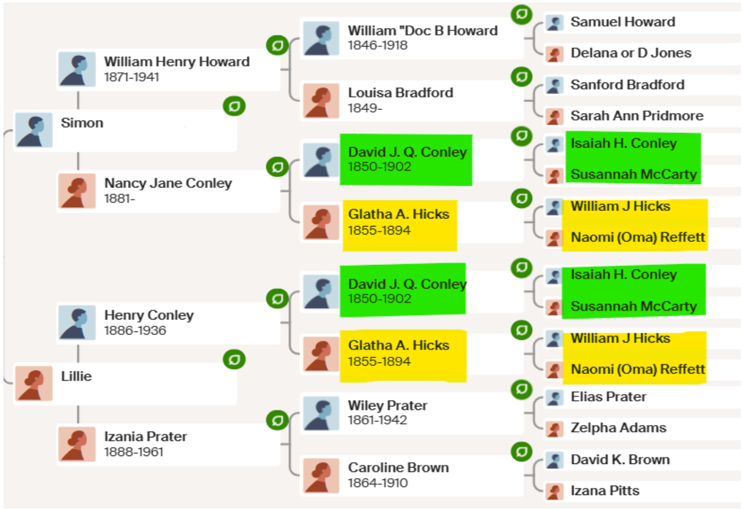


How many ancestors *do* you have?

Parents	2
Grandparents	4
Great-Grandparents	8
2nd Great-Grandparents	16
3rd Great-Grandparents	32
4th Great-Grandparents	64
5th Great-Grandparents	128
6th Great-Grandparents	256
7th Great-Grandparents	512
8th Great-Grandparents	1,024
9th Great-Grandparents	2,048
10th Great-Grandparents	4,096
11th Great-Grandparents	8,192
12th Great-Grandparents	16,384
13th Great-Grandparents	32,768
14th Great-Grandparents	65,536
15th Great-Grandparents	131,072
16th Great-Grandparents	262,144
17th Great-Grandparents	524,288

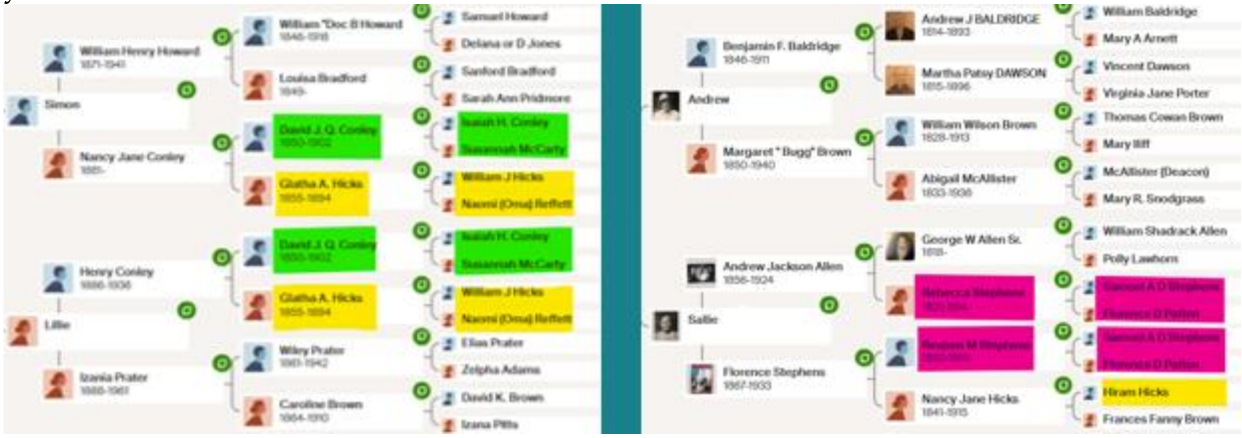


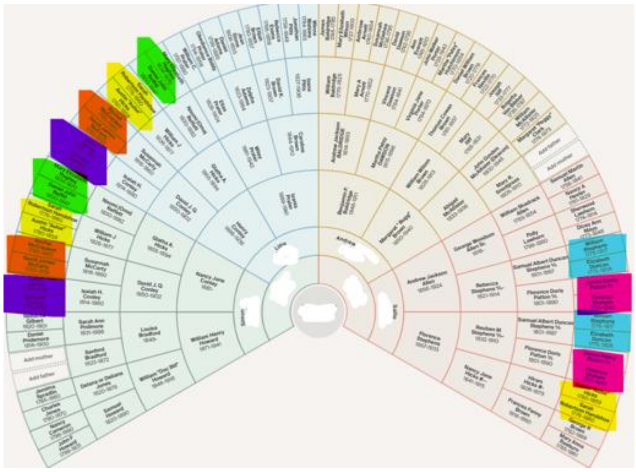
Pedigree collapse increases the likelihood that descendants will inherit identical DNA segments from both sides of their family tree. This can inflate the amount of DNA shared between relatives.



This is my husband's grandfather's tree. As you can see, his parents were first cousins. Significant cM inflation would be seen between the children of Simon & Lillie and their Conley or Hicks third cousins. However, the grandchildren of Simon and Lillie will not show significant inflation between their Conley & Hicks third cousins, because there will not be high runs of homozygosity.

Let's explore that, and add my husband's grandmother. Her mother's parents were first cousins. However, note that she also descends from the Hicks, highlighted in yellow, who are also his ancestor.

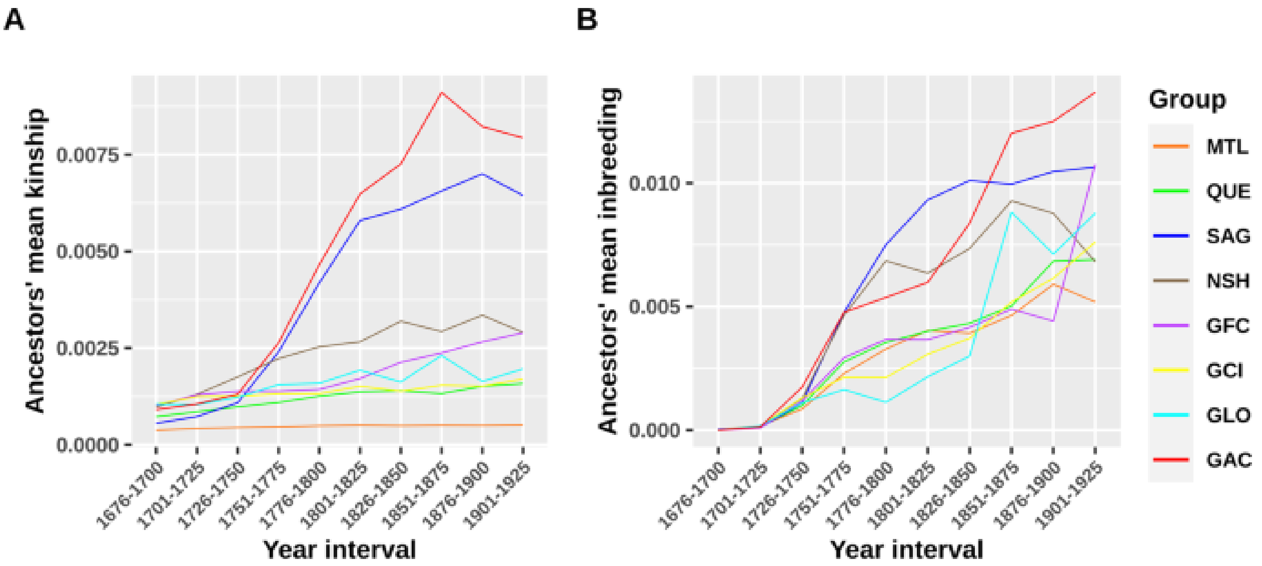




As a result, their children have several redundant ancestors on their tree. This is pedigree collapse.

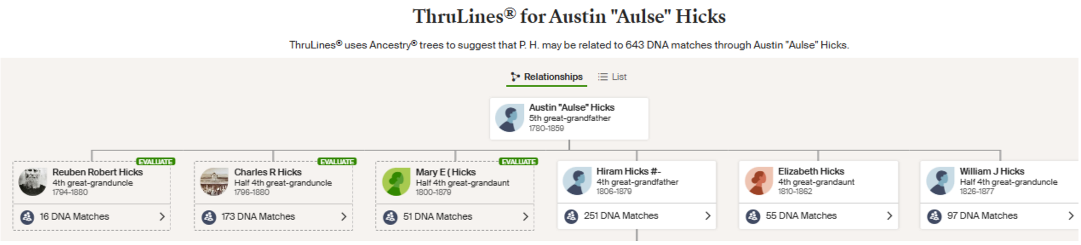
Endogamy

When Pedigree collapse happens over and over again within an isolated community, that is “endogamy” (Greek for ‘marrying within’).



Endogamy refers to the practice of marrying within a specific ethnic, cultural, social, religious, or tribal group over multiple generations. This practice creates a unique genetic signature that can significantly complicate DNA analysis. It's crucial to understand that endogamy exists on a spectrum, ranging from mild to severe, depending on the population's history and isolation level.

In endogamous communities, several genetic patterns emerge. DNA match patterns show extremely high numbers of matches, inflated amounts of shared DNA, multiple small matching segments, and complex relationship networks. Relationship predictions often fail as multiple relationship paths are common, making shared matches appear closer than actual relationships. Segment analysis reveals smaller average segment sizes, multiple small segments from distant ancestors, and difficulty determining identical by descent versus identical by state segments. AND, you get a LOT of DNA matches!


ThruLines® for Austin "Aulse" Hicks

ThruLines® uses Ancestry® trees to suggest that P. H. may be related to 643 DNA matches through Austin "Aulse" Hicks.

Austin "Aulse" Hicks
5th great-grandfather
1780-1859

Reuben Robert Hicks
4th great-granduncle
1794-1880
16 DNA Matches

Charles R Hicks
Half 4th great-granduncle
1796-1880
173 DNA Matches

Mary E (Hicks
Half 4th great-grandaunt
1800-1879
51 DNA Matches

Hiram Hicks #-
4th great-grandfather
1806-1879
251 DNA Matches

Elizabeth Hicks
4th great-grandaunt
1810-1862
55 DNA Matches

William J Hicks
Half 4th great-granduncle
1826-1877
97 DNA Matches

Runs of Homozygosity (RoH)

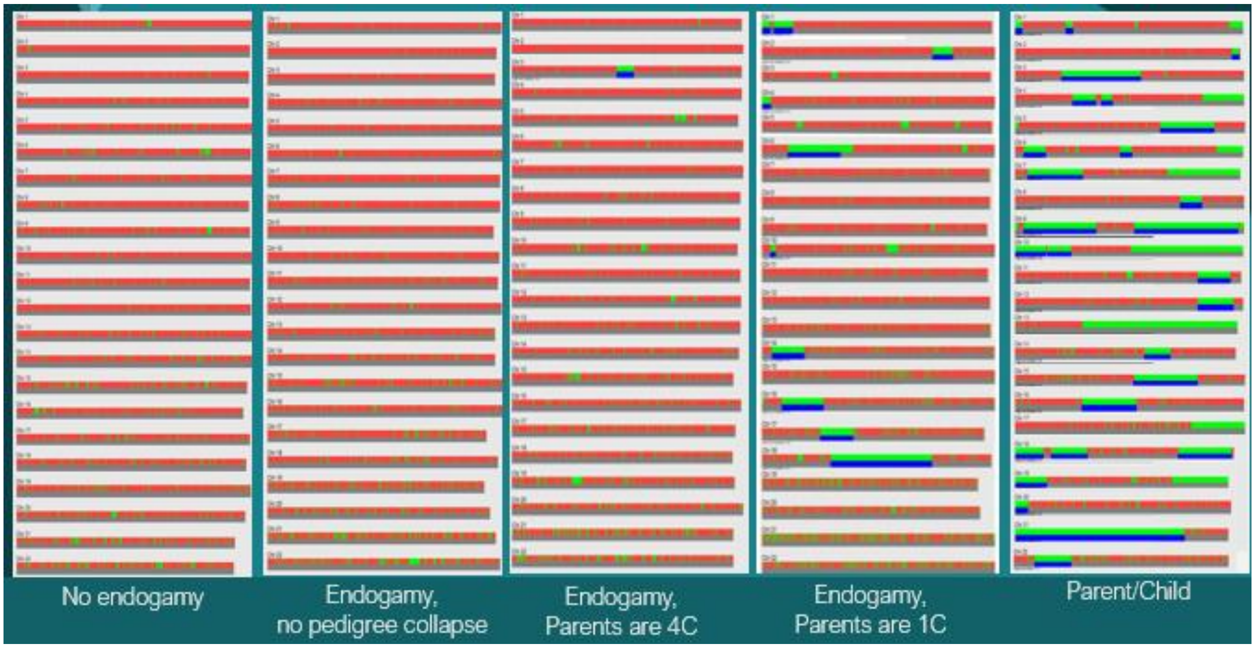
Runs of Homozygosity are long stretches of DNA where the genetic material inherited from both parents is identical.

These regions suggest that the parents share a common ancestor. High levels of RoH are often observed in populations with historical isolation, endogamy, or small founding populations.

These patterns can complicate relationship predictions because large segments may appear without indicating a close genealogical relationship.

This is the output for the GEDMATCH tool "Are Your Parents Related"

- On the left are my own results from the "Are Your Parents Related" tool.
- The second is a Jewish individual with no recent pedigree collapse.
- The third person is one of their cousins. His parents share some ancestry, and may be something like 4th cousins to each other.
- The fourth is from an endogamous community and his parents are first cousins.
- The fifth is illustrative of a parent/child incident. Note the large spans of green.



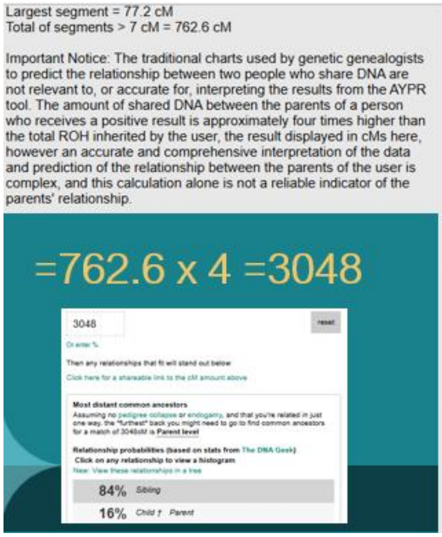
Segment Thresholds in Endogamous Populations

Standard thresholds used in genetic genealogy may not apply in endogamous populations. Because of the high number of distant shared ancestors, small segments may appear far more frequently.

Researchers often apply **higher thresholds** when working with endogamous DNA. For example:

- A **15–30 cM segment** in an endogamous population may have a similar probability of being IBD as a **7 cM segment** in a non-endogamous population.

No universal best practice exists yet. Ongoing research continues to refine these guidelines. Each community will have different 'pile up' regions and will need different strategies to find the right one for you.



Largest segment = 77.2 cM
Total of segments > 7 cM = 762.6 cM

Important Notice: The traditional charts used by genetic genealogists to predict the relationship between two people who share DNA are not relevant to, or accurate for, interpreting the results from the AYPR tool. The amount of shared DNA between the parents of a person who receives a positive result is approximately four times higher than the total ROH inherited by the user, the result displayed in cMs here, however an accurate and comprehensive interpretation of the data and prediction of the relationship between the parents of the user is complex, and this calculation alone is not a reliable indicator of the parents' relationship.

=762.6 x 4 =3048

3048

Or enter %

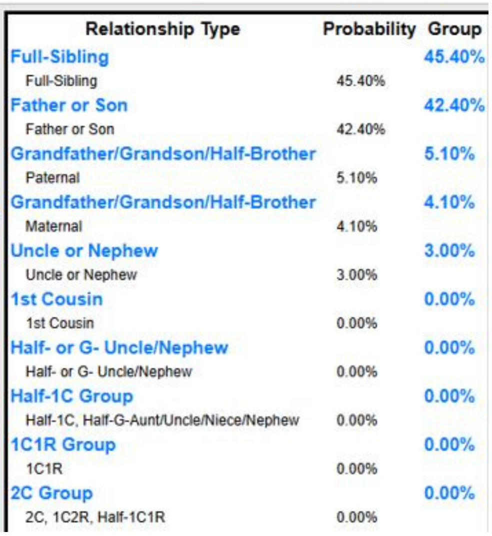
Then any relationships that fit will stand out below

Click here for a shareable link to the cM amount above

Most distant common ancestors
Assuming no pedigree collapse or endogamy, and that you're related in just one way, the "furthest" back you might need to go to find common ancestors for a match of 3048cM is Parent level

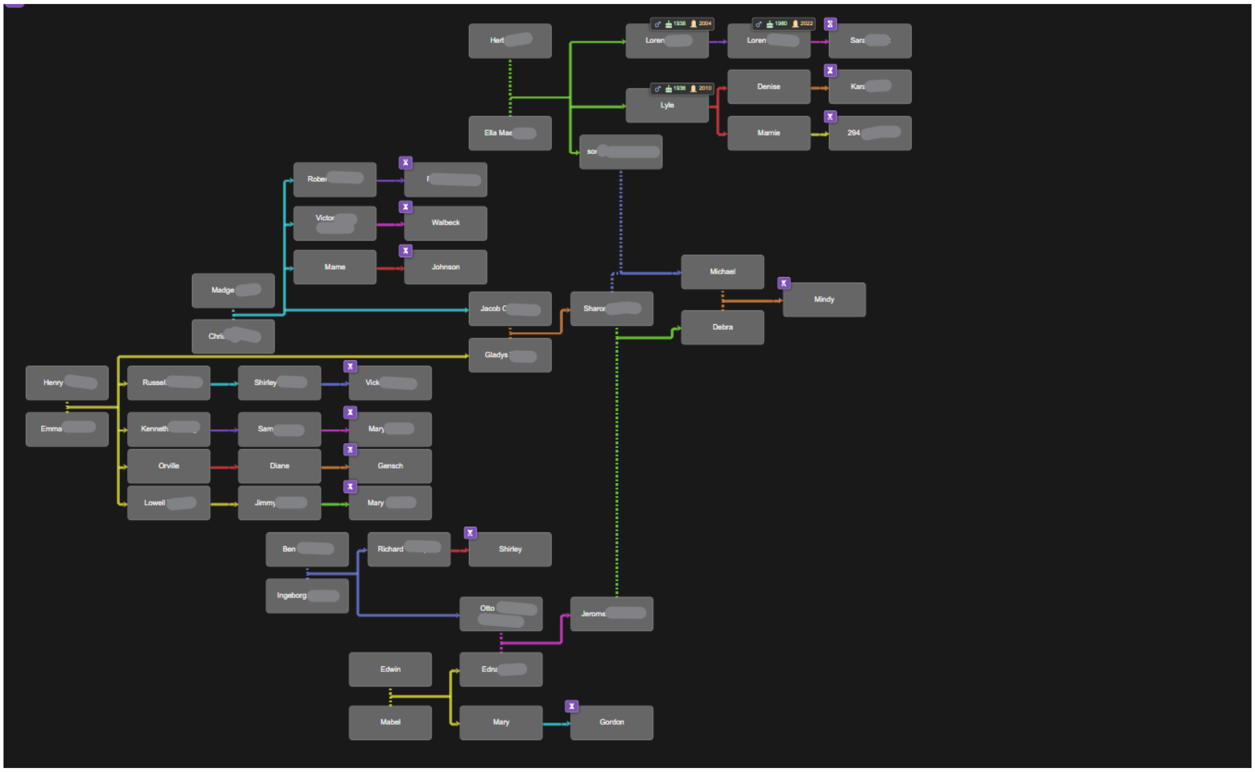
Relationship probabilities (based on stats from The DNA Geek)
Click on any relationship to view a histogram
New: View these relationships in a tree

84%	Sibling
16%	Child ↕ Parent



Relationship Type	Probability	Group
Full-Sibling		45.40%
Full-Sibling	45.40%	
Father or Son		42.40%
Father or Son	42.40%	
Grandfather/Grandson/Half-Brother		5.10%
Paternal	5.10%	
Grandfather/Grandson/Half-Brother		4.10%
Maternal	4.10%	
Uncle or Nephew		3.00%
Uncle or Nephew	3.00%	
1st Cousin		0.00%
1st Cousin	0.00%	
Half- or G- Uncle/Nephew		0.00%
Half- or G- Uncle/Nephew	0.00%	
Half-1C Group		0.00%
Half-1C, Half-G-Aunt/Uncle/Niece/Nephew	0.00%	
1C1R Group		0.00%
1C1R	0.00%	
2C Group		0.00%
2C, 1C2R, Half-1C1R	0.00%	

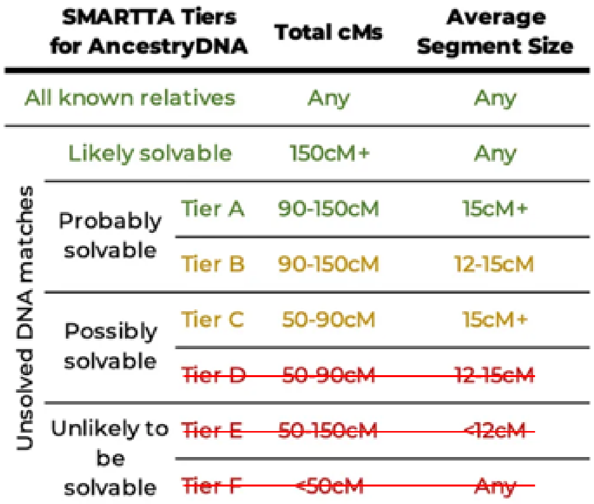
The first step when working a case where parents are closely related is establishing how closely related they are. The established method was to take the total amount of shared DNA and multiply by 4, which gives the appropriate amount. Now there is a calculator available on the site.
In this case, the individuals were father/daughter. But full sibling is more probable. Remember that "most probable" does not always mean "true."



For cases where parents are closely related, map the matches in the same way that you would if the parents were not related.

Once you have mapped them, use BANYAN DNA to estimate the relationship.

On your screen, you see a hypothesis where the tester is the child of two half siblings. Therefore, that person’s maternal and paternal grandmother are the same person, however, their grandfathers are two different men.



	SMARTTA Tiers for AncestryDNA		Total cMs	Average Segment Size
	All known relatives		Any	Any
Unsolved DNA matches	Likely solvable		150cM+	Any
	Probably solvable	Tier A	90-150cM	15cM+
		Tier B	90-150cM	12-15cM
	Possibly solvable	Tier C	50-90cM	15cM+
		~~Tier D~~	~~50-90cM~~	~~12-15cM~~
	Unlikely to be solvable	~~Tier E~~	~~50-150cM~~	~~<12cM~~
		~~Tier F~~	~~<50cM~~	~~Any~~

The SMARTTA Method

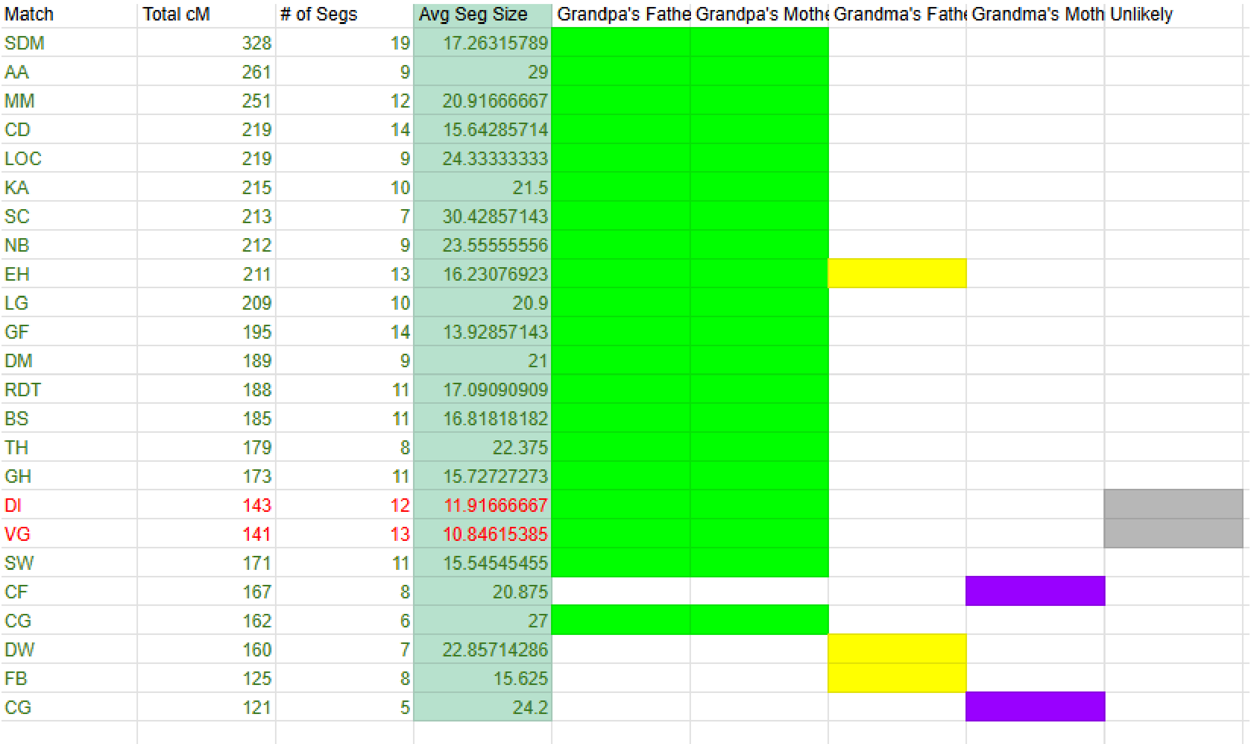
Designed by Steve Taub, the SMARTTA, or Shared Match Analysis on Rows with Top Tier Averages method offers a more refined approach for endogamous DNA analysis. Key components include focusing on matches with larger average segment sizes, performing detailed shared match analysis, and considering cluster patterns carefully. Implementation steps involve listing matches meeting specific criteria, analyzing known relatives first, assigning colors systematically, reviewing patterns thoroughly, and validating conclusions.

Core Principles

1. Focused Match Selection
    - Prioritizes matches with larger segment sizes
    - Concentrates on potentially solvable relationships
    - Eliminates noise from distant endogamous connections
    - Reduces analysis complexity
2. Systematic Analysis
    - Structured approach
    - Clear documentation
    - Consistent evaluation
    - Reproducible results
3. Tiered Classification
    - Tier A: Strong matches (≥15cM average segment)
    - Tier B: Moderate matches (12-15cM average)
    - Tier C: Weak matches (<12cM average)

Implementation Process

1. Initial Setup
    - Create spreadsheet with essential columns
    - Document match details systematically
    - Calculate average segment sizes
    - Establish color-coding system
2. Match Selection
    - Include all matches >150cM total
    - Add Tier A matches (90-150cM, ≥15cM average)
    - Include Tier B matches (90-150cM, 12-15cM average)
    - Optional: Consider Tier C matches



Match	Total cM	# of Segs	Avg Seg Size	Grandpa's Fathe	Grandpa's Mothe	Grandma's Fath	Grandma's Moth	Unlikely
SDM	328	19	17.26315789	🟩	🟩			
AA	261	9	29	🟩	🟩			
MM	251	12	20.91666667	🟩	🟩			
CD	219	14	15.64285714	🟩	🟩			
LOC	219	9	24.33333333	🟩	🟩			
KA	215	10	21.5	🟩	🟩			
SC	213	7	30.42857143	🟩	🟩			
NB	212	9	23.55555556	🟩	🟩			
EH	211	13	16.23076923	🟩	🟩	🟨		
LG	209	10	20.9	🟩	🟩			
GF	195	14	13.92857143	🟩	🟩			
DM	189	9	21	🟩	🟩			
RDT	188	11	17.09090909	🟩	🟩			
BS	185	11	16.81818182	🟩	🟩			
TH	179	8	22.375	🟩	🟩			
GH	173	11	15.72727273	🟩	🟩			
DI	143	12	11.91666667	🟩	🟩			▨
VG	141	13	10.84615385	🟩	🟩			▨
SW	171	11	15.54545455	🟩	🟩			
CF	167	8	20.875				🟪	
CG	162	6	27	🟩	🟩			
DW	160	7	22.85714286			🟨		
FB	125	8	15.625			🟨		
CG	121	5	24.2				🟪	

I applied the SMAARTA principles to a case with high endogamy. This one is from Virginia.

This was helpful. Basically, the system helps you qualify matches – prioritizing those who have fewer segments. Each segment could possibly be due to a different relationship, so this method is important for endogamy.

However, this particular case is not just endogamy, it's also recent intermarriage. This is why you have one match that is both yellow and green.

Tips:

Look for close matches that only match in one way

- Build out their tree to 3rd great-grandparents, if possible
- Qualify the match. Look for misattributed parentage, etc.
- Leverage Ancestry Pro Shared Matches to identify people who match closely in one way and match distantly in another way. Disregard the distant match.

Look for matches that move away

- Look for clues in the tree that grandparents, parents may have been born elsewhere

Look for matches that bring an ethnically unique parent or grandparent into their tree

Recommended Resources

Bettinger, Blaine. *The Genetic Genealogist.*
https://thegeneticgenealogist.com

Bettinger, Blaine and Debbie Parker Wayne. *Genetic Genealogy in Practice.*

Leeds, Dana. "The Leeds Method for Clustering DNA Matches."
https://www.danaleeds.com

Mills, Elizabeth Shown. *Evidence Explained.*

Diamond, Lara. "Endogamy in Practice."
https://larasgenealogy.blogspot.com

Southard, Diahan. *Your DNA Guide.*