

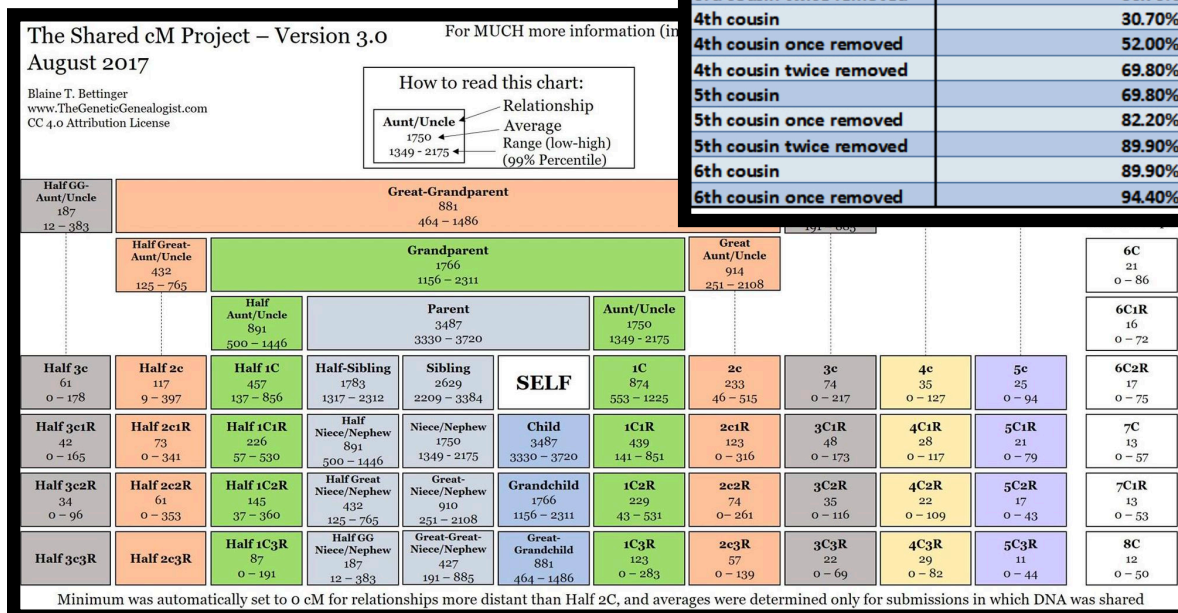
# Strategies to Analyze Endogamous DNA

Alec Ferretti

AncestorArchivist@gmail.com

- We measure DNA in centiMorgans
  - A standardized unit that measures an “amount of inherited DNA”
- People with closer relationships typically share more DNA
- DNA segments are added together to get a sum total
- Segments are only to be counted if they are longer than 7cM
- Shorter segments are likely to be “identical-by-state”, instead of “identical-by-descent”
- Segments longer than 7cM could still be IBS, but it is less likely
- For most ethnic groups, any segment longer than 10-15cM is very likely to be IBD
- See these charts from the ISOGG Wiki / TheGeneticGenealogist.com:

Genealogical Relationship	Probability of No Detectable DNA Relationship
1st cousin	0.00%
1st cousin once removed	0.00%
1st cousin twice removed	0.00%
2nd cousin	0.00%
2nd cousin once removed	0.10%
2nd cousin twice removed	2.30%
3rd cousin	2.30%
3rd cousin once removed	12.10%
3rd cousin twice removed	30.70%
4th cousin	30.70%
4th cousin once removed	52.00%
4th cousin twice removed	69.80%
5th cousin	69.80%
5th cousin once removed	82.20%
5th cousin twice removed	89.90%
6th cousin	89.90%
6th cousin once removed	94.40%

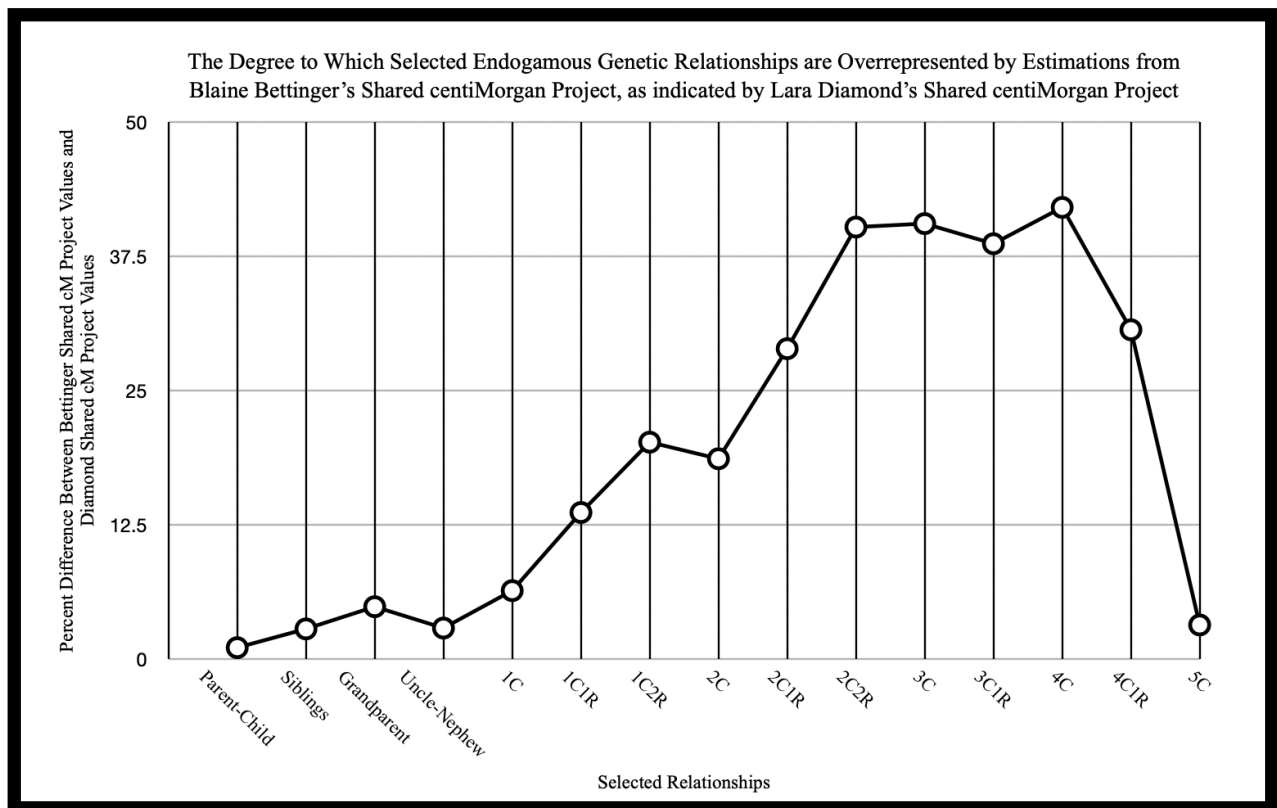


# Endogamy

- Some populations often intermarried over the course of centuries
- This created oddities that manifest themselves in DNA relationships today
- Relationships that would normally appear to be identical by descent are in fact NOT
- People share more DNA than would be expected
- The normal threshold of 7cM does not apply
- A 15-30 cM segment for an endogamous person has a similar probability of being IBD as does a 7cM segment for a non-endogamous person
- Inspired by Blaine Bettinger's Shared centiMorgan project, Lara Diamond published a similar dataset using crowdsourced match information for Jews:

Relationship	Number of submissions	Min total shared cM	Max total shared cM	Average largest segment	Average total shared cM	SCP Average	% Difference
Parent/Child	136	3379	3730	268.07	3521.02	3485	1.02
Siblings	154	2246	3037	185.26	2687.44	2613	2.77
Grandparent/Grandchild	15	1426	2215.1	177.83	1843.21	1754	4.84
Uncle or Aunt / Niece or Nephew	173	671	2303	122.87	1792.05	1741	2.85
First cousins	256	434.1	1473.2	82.31	924.58	866	6.34
First cousins once removed	389	61.2	1878	60.85	501.25	433	13.62
First cousins twice removed	37	167	439.4	45.08	276.83	221	20.17
Second cousins	562	42.4	698	47.63	281.41	229	18.62
Second cousins once removed	508	0	446	36.62	171.54	122	28.88
Second cousins twice removed	65	17.6	285.7	29.21	118.75	71	40.21
Third cousins	285	0	315	26.28	122.76	73	40.53
Third cousins once removed	236	0	199	18.9	78.25	48	38.66
Fourth cousins	162	0	664	13.97	60.39	35	42.04
Fourth cousins once removed	72	0	146	10.21	40.37	28	30.64
Fifth cousins	28	0	116	4.01	25.81	25	3.14

- One can cross reference this (very limited) data to graph how the shared DNA of different endogamous relationships is over-represented
- Further work is needed to ensure that this data is accurate, and that it is applicable to other populations



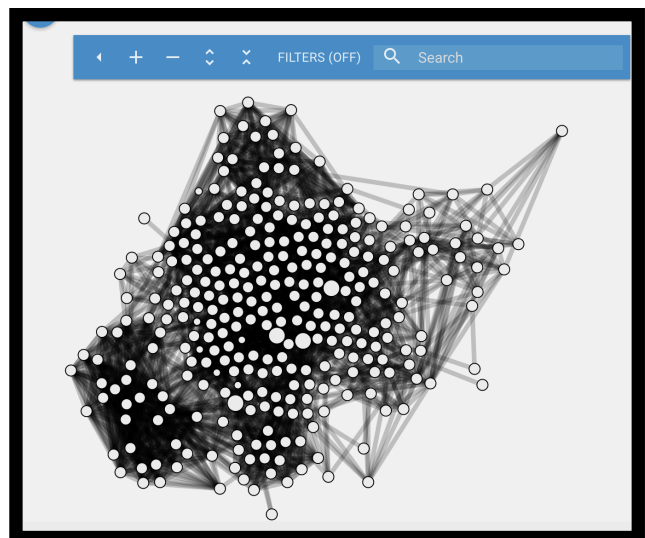
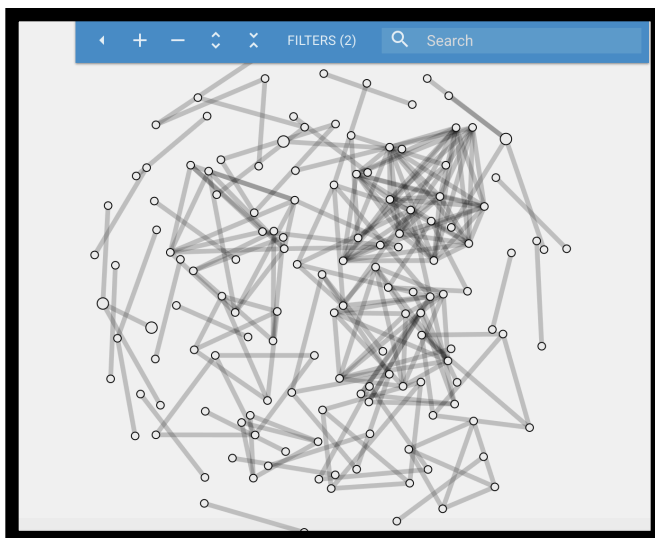
## Testing Companies

- Testing with AncestryDNA
  - Uses proprietary methods to try to filter out false matches
  - Does not give any segment data
  - Largest Database
- Testing with 23andMe
  - Given segment & triangulation data for matches
  - Can manually exclude smaller segments
  - Database is half the size of Ancestry's

- Testing with MyHeritage
  - Cheaper kit
  - Smallest sample size
  - Provides segment data, and allows to sort by largest segment
  - Nifty features - automatic triangulation detection
- Testing with Family Tree DNA
  - Smaller sample size than Ancestry or 23&me
  - Their interface includes segments that are smaller than 7cM
  - This is bad for all populations but especially misleading for those who deal with endogamy
  - It is possible to remove these small segments by going into their chromosome browser
  - Provides segment data, and allows to sort by largest segment

## Visualizing Endogamy

- Creating visualizations with RootsFinder
  - Show how matches relate with each other
  - Endogamous testers are related to a larger share of a given database
  - Compare non-endogamous person (left), with endogamous person (right)



- DNA Painter Bulk Import
  - Map all matches above a certain segment threshold onto chromosomes
  - View pile-up regions to imply IBS segments
  - Seek out those matches who match on unique parts of chromosome



## Strategies

- Organize matches by the largest segment shared AND the number of segments shared.
- Look for matches that share segments in unique areas of the chromosome.
  - While this does not guarantee a match, it suggests that the match may not be matching in a pile-up region.
- The tools provided in nearly all the commercial databases are inadequate for analyzing endogamous segments.
  - One must manipulate data in a program such as Excel to look for the strongest matches.
  - There is no magic threshold: a match who only shares one 35cM segment might be IBD while a match who shares six 10cM segments might be IBS.
  - Start with the strongest matches and work downward.
  - Test as many first, second, and third cousins as possible, so as to find triangulated segments.

# Links

ISOGG Wiki: <https://isogg.org/wiki/Endogamy>

Lara Diamond's Blog: <https://larasgenealogy.blogspot.com/>

Shared CentiMorgan Project: <https://thegeneticgenealogist.com/2020/03/27/version-4-0-march-2020-update-to-the-shared-cm-project/>

Roots Finder: <https://www.rootsfinder.com/>

DNA Painter: <https://dnainter.com/>