ETHNICITY AND GENES

What Do the ethnicity assessments by genealogical genetic test companies really mean?

Major testing companies and Gedmatch provide ethnicity assessments based on autosomal DNA

- Family Tree has “My Origins” and “ancient Origins”
- Ancestry has Ethnicity Estimates for “Thousands of Years Ago” and “Genetic Communities”
- 23andMe provides “Ancestry Composition” and “Neanderthal Ancestry”
- GedMatch (which accepts data from any of the above) provides 7 separate admixture utilities to compare to one’s DNA plus “Archaic DNA Matches” and a check for Neanderthal ancestry

What is behind these “estimates?”

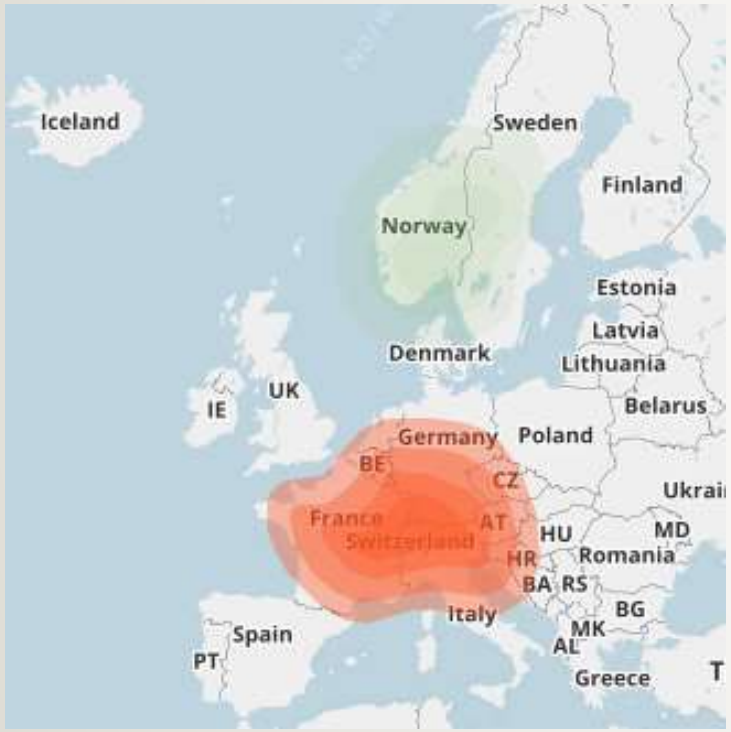
- All of these depend on the same principles:

  - *They are comparisons to autosomal or recombinant genetic samples from defined populations, either modern or ancient.*
  - *Each company uses different comparative samples from different areas*
  - *Each company attempts to control for and eliminate data from recent immigrants that would distort the sample, but none can accurately control for past migrations*

So What do ethnicity matches really mean?

- **A matching ethnicity or geographic distinction isn't necessarily from a location that was occupied by your ancestors.**
- Matching populations are groups whose autosomal genes show that they share some part of your ancestry – even if they migrated to their current location only 200 years ago.
- Your ancestors migrated to Orangeburgh Township almost 300 years ago. A substantial group of their relatives might have migrated into another area as well.
- A location to which you show no connection may simply be an area in which the company has no comparative sample.

Family Tree myOrigins



Family Tree DNA tells me that my ethnicity is:

European	98%
West and Central Europe	86%
Scandinavia	12%
Trace Results	
South America	<1%
East Europe	<2%

Is this accurate? As a measure of comparison to genetic samples, probably so. However, it does not provide a sound assessment of where my ancestors migrated from to come to South Carolina.

Family Tree Ancient European Origins

The Ancient Origins map helps to clarify. The shovel icons show locations of archaeological specimens that my genetics match. According to this, I am a fairly common:

0% Non-European (trace matches in South America probably reflect European immigrants there).

44% hunter-gatherer (pre-Neolithic migrants)

45% farmer (Neolithic migrants from the Near and Middle East)

12% metal age invader (both the paternal Shuler and Dukes lines appear to be post-Neolithic Bronze Age migrants to Europe, from the Levant and the Caucusus, respectively)

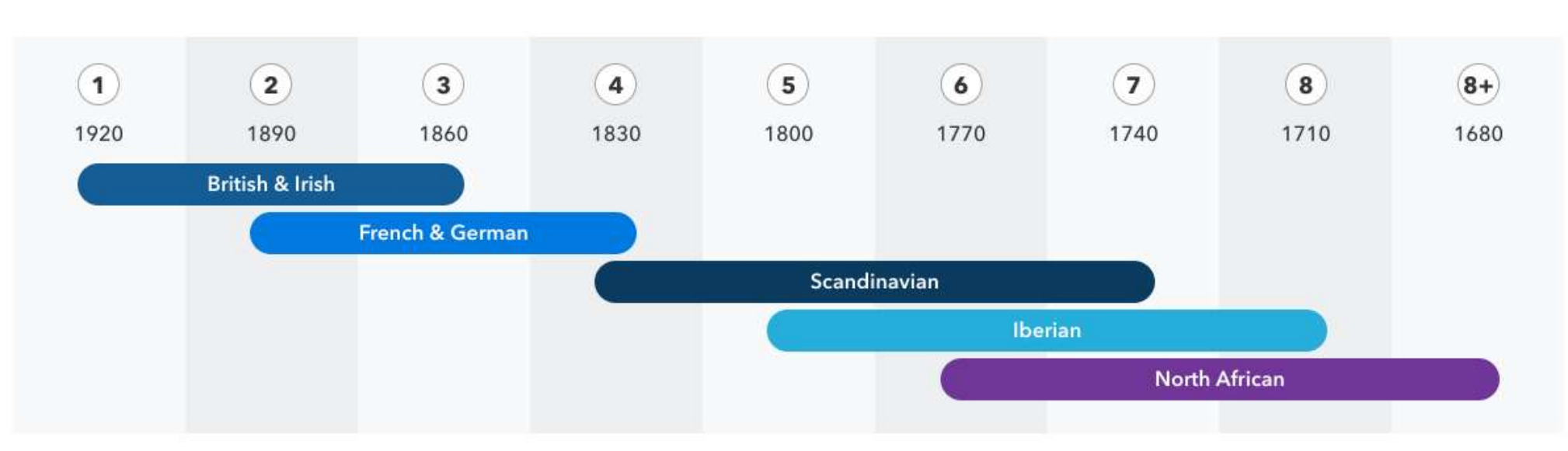
23andMe: Ancestry Composition

- In this version, I'm:
- European 99.7%
- Middle Eastern & North African 0.1%
- Sub-Saharan African < 0.1%
- Unassigned < 0.1%

23andme European components

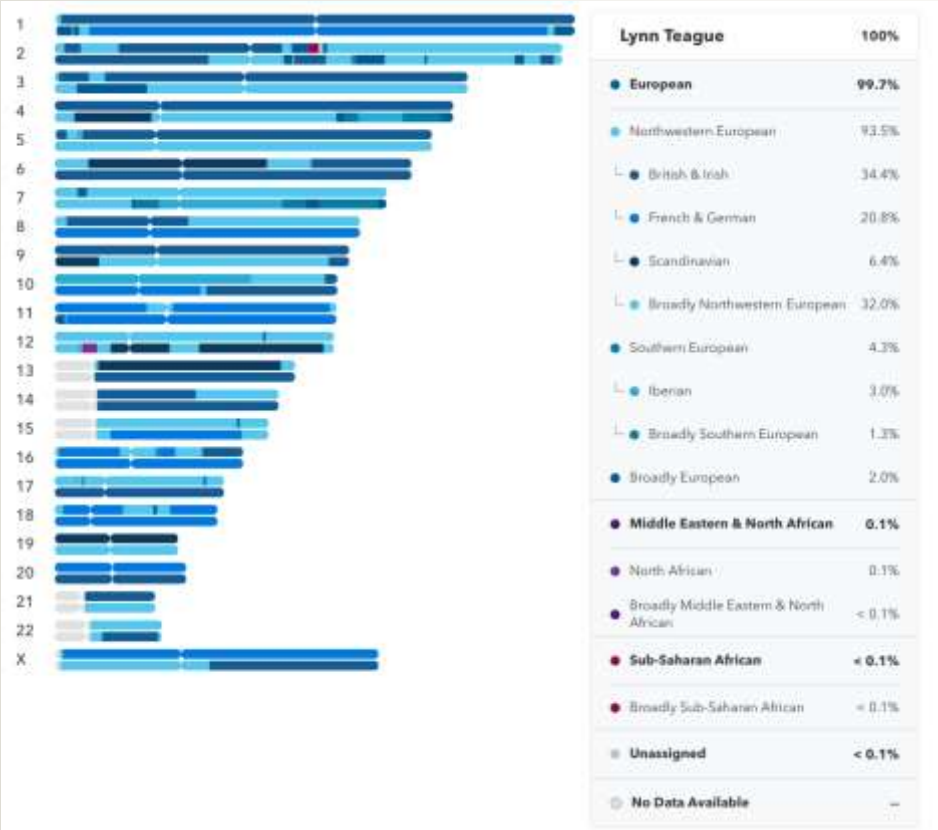
■ Northwestern European		93.5%
– *British & Irish*	*34.4%*	
– *French & German*	*20.8%*	
– *Scandinavian*	*6.4%*	
– *Broadly Northwestern European*	*32.0%*	
■ Southern European		4.3%
– *Iberian*	*3.0%*	
– *Broadly Southern European*		*1.3%*
■ Broadly European		2.0%

23andMe Ancestry Timeline



This supposedly points to an ancestor having 100% the indicated ancestry during that time frame. Some of this is verifiable through documents, but some of it is very unlikely, especially the pure Scandinavian and Iberian ancestry during the indicated time periods.

23andme Chromosome painting



This is the foundation of the 23andMe ethnicity estimates, broken down by chromosome segment. Each segment is represented by two lines, the paternal and maternal DNA contributions for that segment.

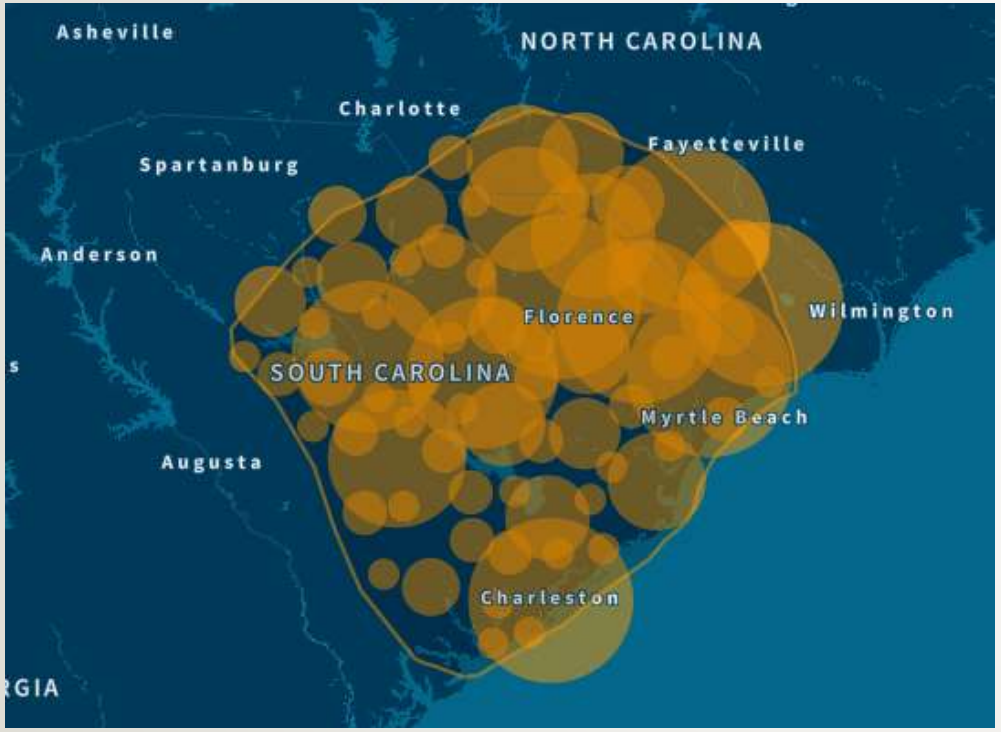
Most of the colors represent subgroupings of European ancestry.

The small red segment on chromosome 2 is what I think of as Parthena's genes.

Ancestry Ethnicity Estimate

Europe	97%
Great Britain	59%
Europe West	12%
Iberian Peninsula	7%
Scandinavia	7%
Ireland	5%
Low Confidence Region, West Asia	3%

Ancestry Genetic Communities: SC's Midlands and Coastal Plain

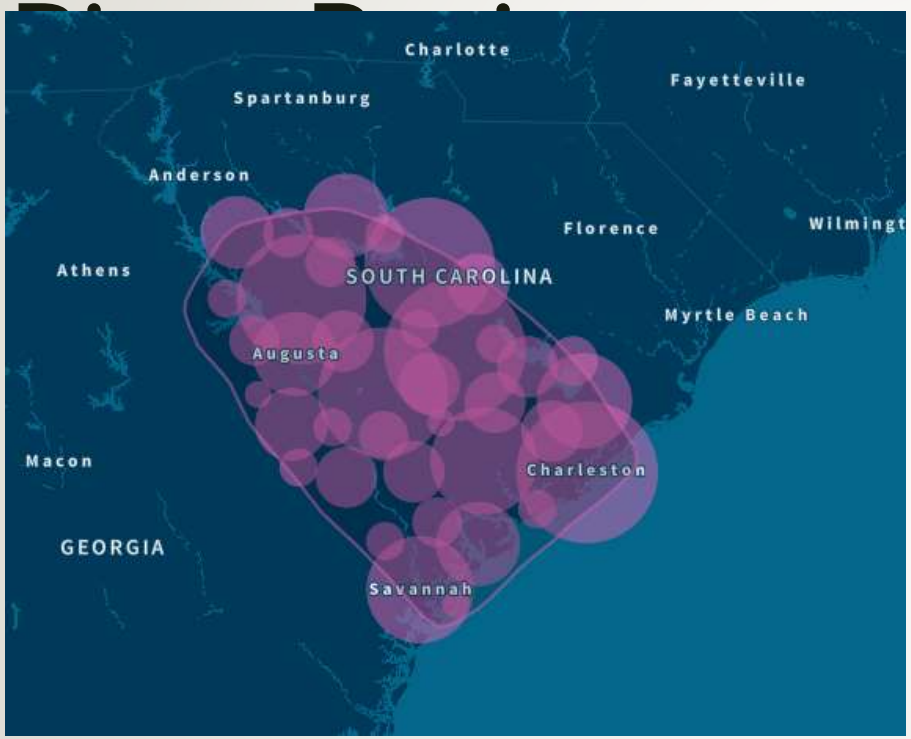


Genetic communities are unique to Ancestry, and an interesting contribution. This grouping is basically the one that my Orangeburg (and Dorchester and Charleston counties) ancestry belongs to. Ancestry hasn't picked up on the Swiss immigration in SC, but identifies this with German immigration, along with Scottish Highlanders joining the Welsh, English, and Scots-Irish who were already here.

Ancestry Genetic Communities: Settlers of the NC-SC Border

Ancestry also accurately identified my descent from the Scot—Irish and Northern English who settled the NC-SC borderlands. This is my maternal grandmother's family, the Cauthens, Hortons, Wests, Bells, Truesdales and related families of Lancaster County, SC.

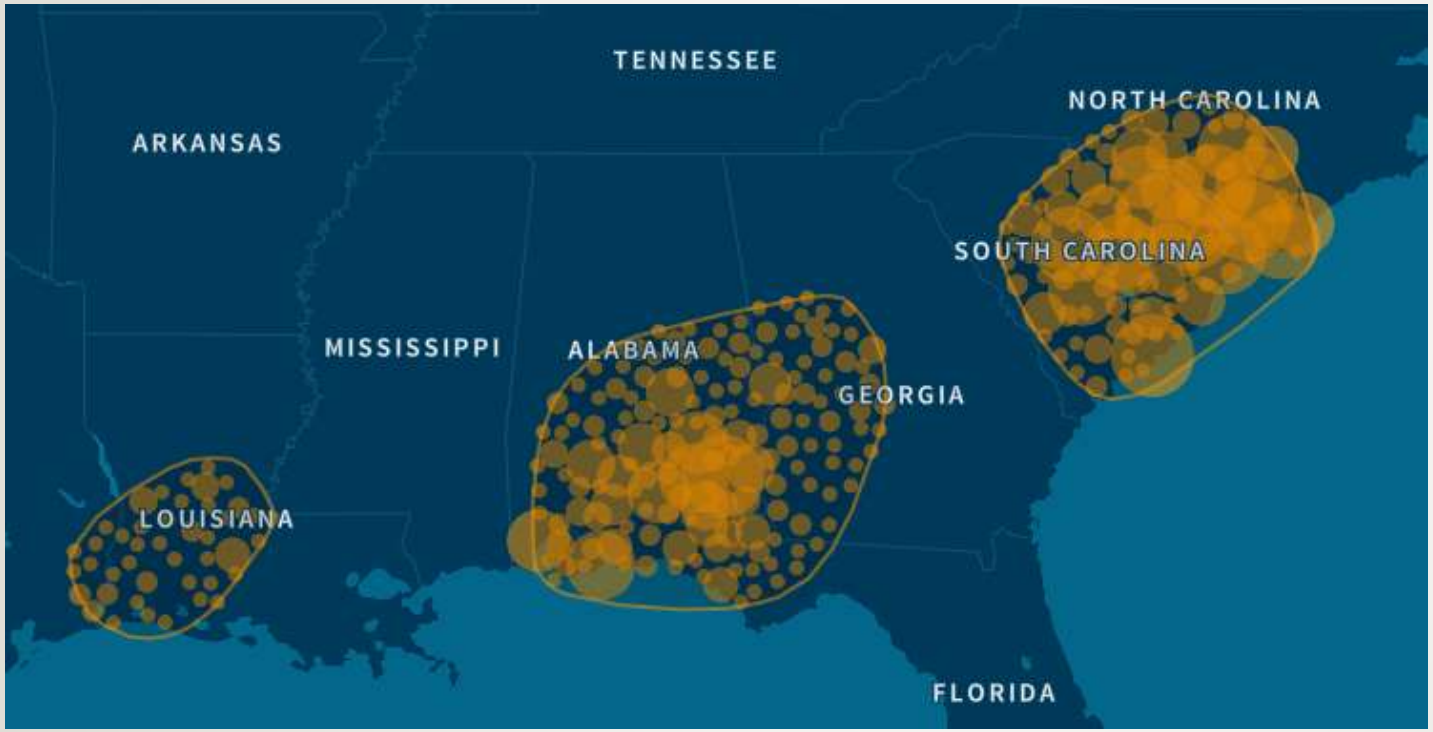
Ancestry Genetic Communities: African Americans in the Savannah



This last community requires some interpretation. Parthenia, mother of Joseph Pendarvis' six children, represents much less than 1% of my DNA and can't adequately account for my association with a community of African descent.

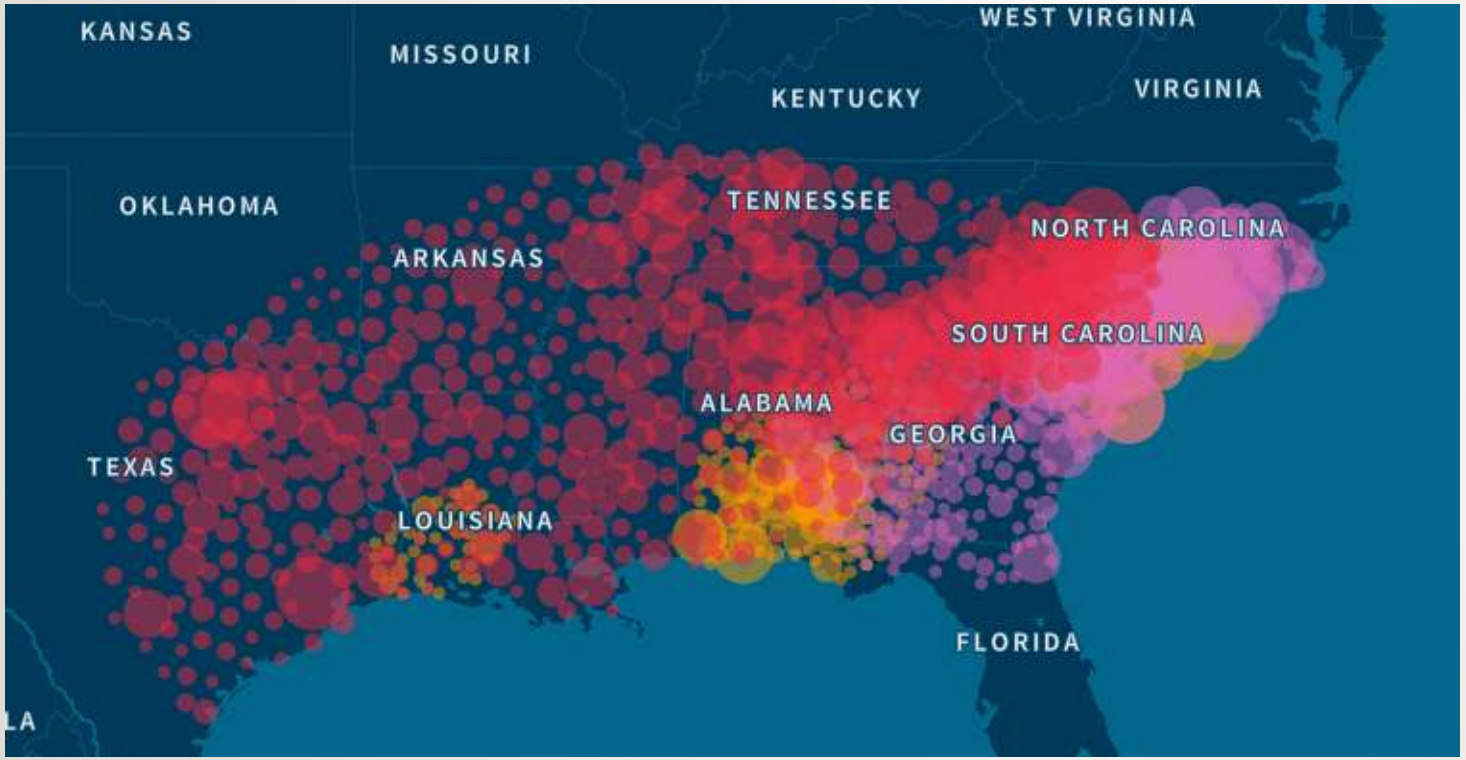
I believe this identification reflects the reverse process: the frequency with which our European colonist ancestors had children with African Americans in South Carolina. This is supported by the significant numbers of African American matches I see in my Family Tree DNA report, and in the reports of other Orangeburg DNA study members.

Migrations of our Communities



This interesting slide shows the Midlands-Coastal community as it spread out. The intense concentration in Alabama should be no surprise to anyone who has studied Orangeburg family movement westward in the 1800's.

South Carolina Shapes the Southeast



Genetic Communities™
Hundreds of years ago

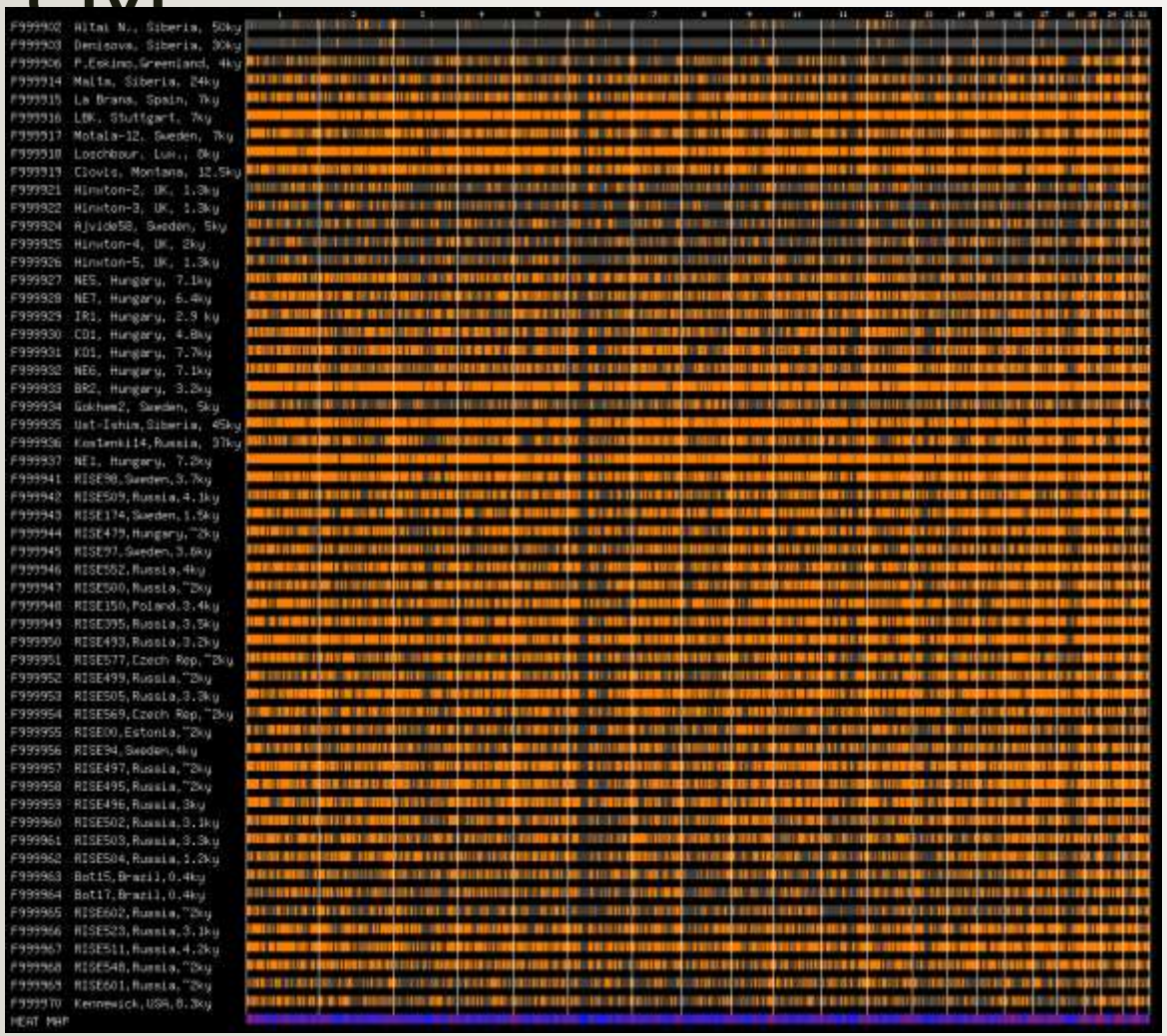
- Settlers of South Carolina's Midlands and Coastal Plains
- Settlers of the North Carolina-South Carolina Border
- African Americans in the Savannah River Basin

GedMatch: Multiple Comparative data Bases

- MDLP Project (Magnus Ducatus Lituaniae Project, focus on the former Grand Duchy)
- Eurogenes (23andMe and ftDNA data sources)
- Dodecad (targeted populations; http://dodecad.blogspot.com/)
- Harrapa World (focuses on India and adjacent areas)
- Ethiohelix (Africa only)
- Puntdena (ancient ancestry by group, with Oracle feature by current ethnicity)
- Gedrosia DNA (most accurate for those with South or West Asian ancestry)
- Archaic DNA Matches (compares to archaeological specimens at varying levels of match)

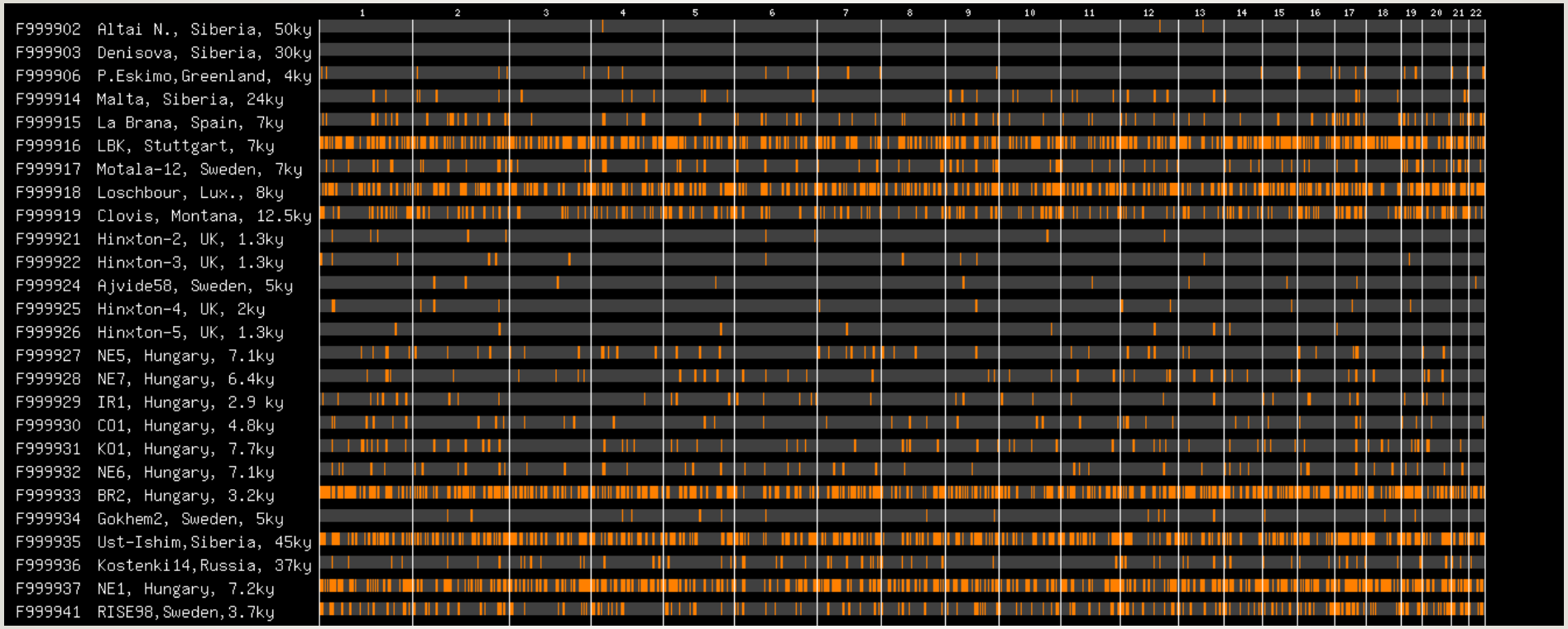
Each database can be used in admixture proportions (with link to Oracle), admixture proportions by chromosome, chromosome painting at large and small size, painting differences between 2 kits for 1 chromosome or 2 kits for 22 chromosomes at reduced size.

Archaic DNA Matches at GedMatch, 0.5 cM



This tool includes many archaeological specimens of varying ages.

Archaic Matches at 1.0 cM Match



At 1.0 cM, the strongest matches are Stuttgart (SW Germany) 7K BP; Loschbour Luxembourg 8K BP; BR2 Hungary 3.2k BP, Ust-Ishim Siberia 45k BP' and NE1 Hungary, 7.2k BP. What does this mean? It means that when you have Swiss and German farmer ancestors, the further back you go, the more Swiss and German farmers you find. The Siberian match is exceedingly old, and represents an era when there was great homogeneity across Europe and Asia because migration out of Africa was relatively recent.

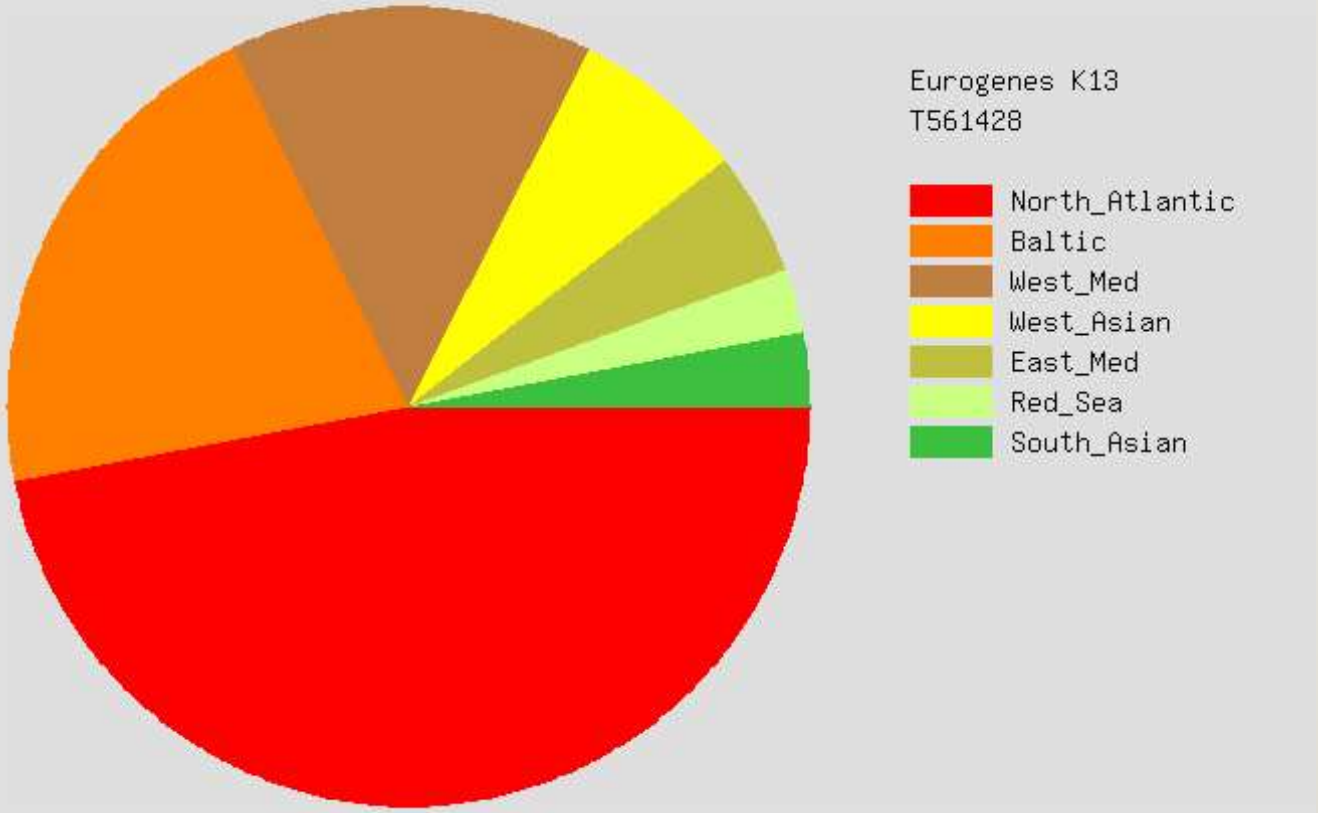
The Eurogenes model at GedMatch

Population	
North_Atlantic	47.05
Baltic	20.82
West_Med	14.70
West_Asian	6.75
East_Med	5.14
Red_Sea	2.59
South_Asian	1.29
East_Asian	0.45
Siberian	-
Amerindian	0.65
Oceanian	0.56
Northeast_African	-
Sub-Saharan	-

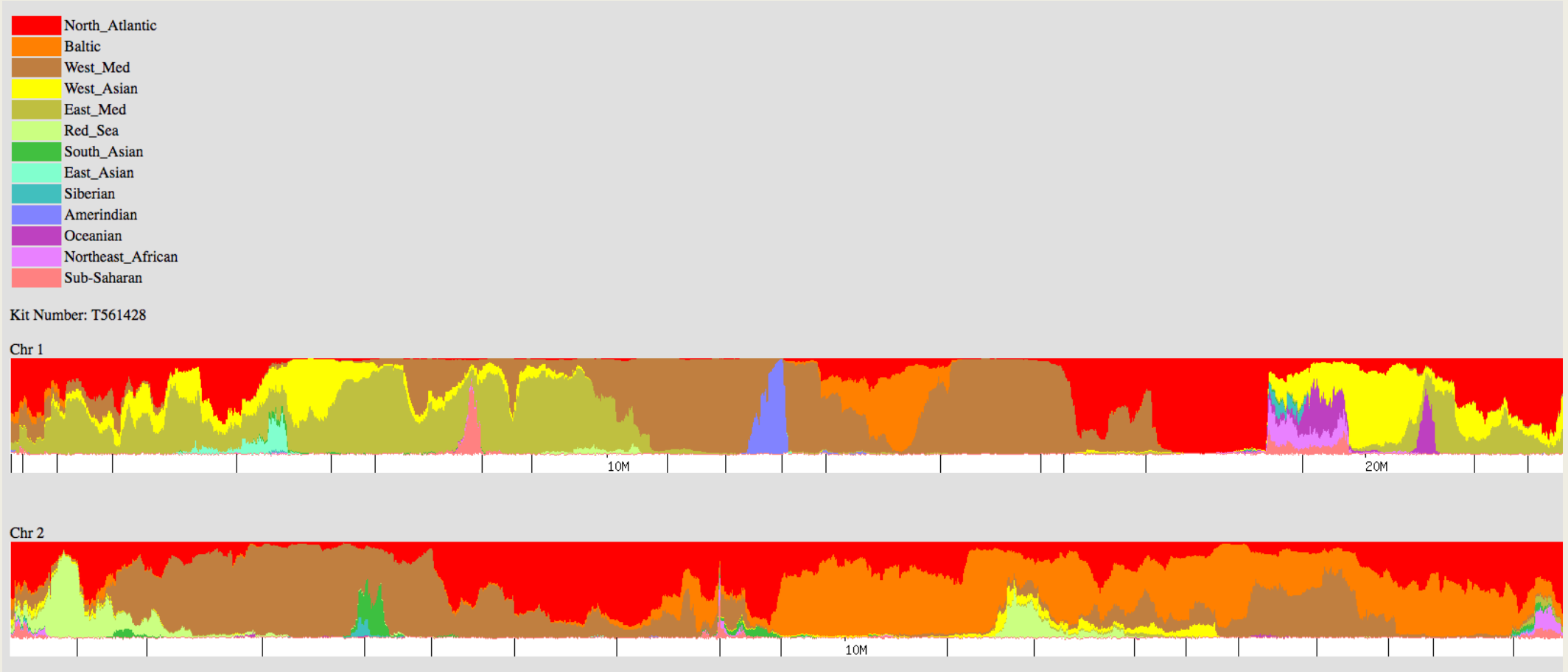
Oracle

Oracle-4

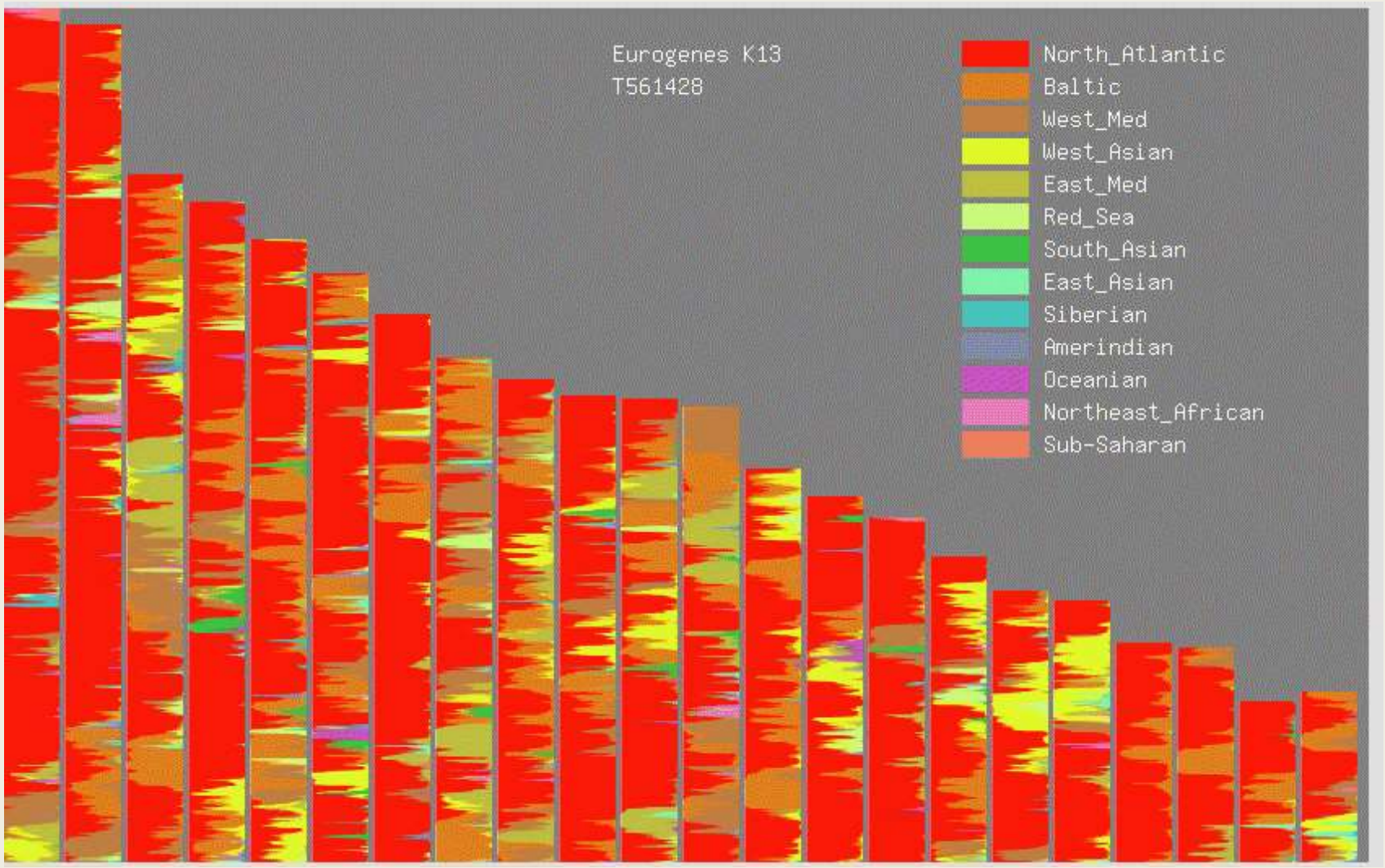
Spreadsheet



Chromosome Painting at GedMatch: Eurogenes



Chromosome Painting, Small, Eurogenes



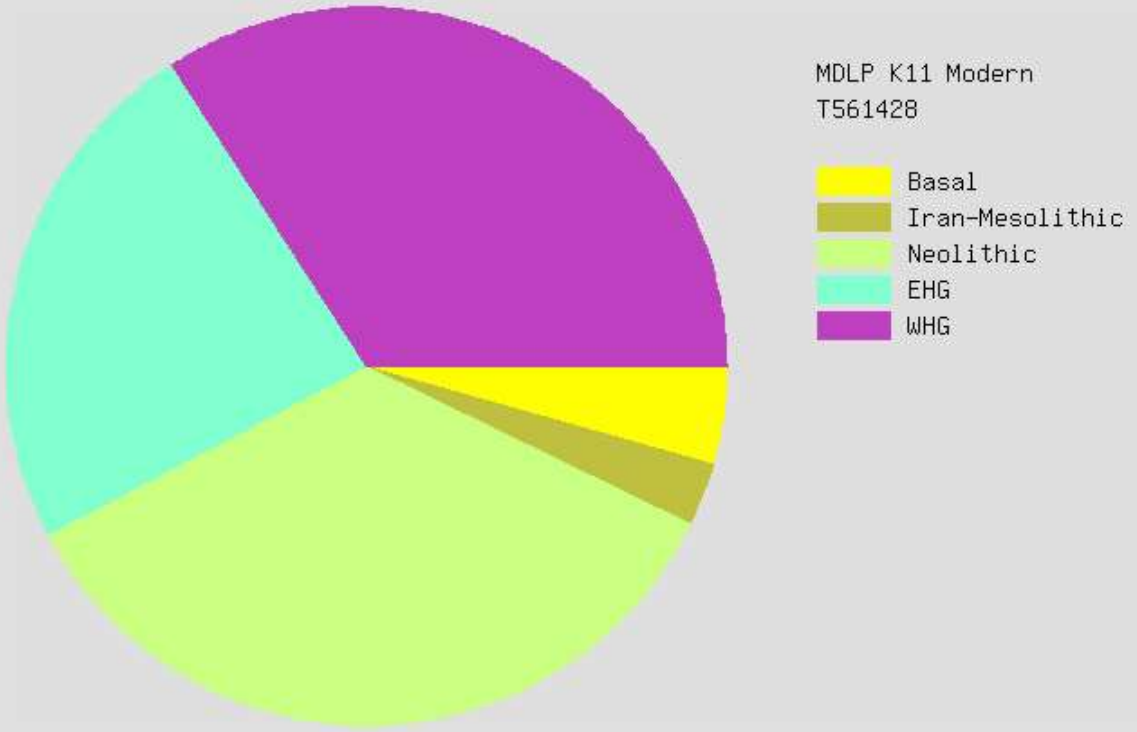
MDLP K11 Modern Admixture

Population	
African	0.13
Amerindian	0.21
ASI	-
Basal	4.47
Iran-Mesolithic	2.69
Neolithic	35.01
Oceanic	-
EHG	23.66
SEA	-
Siberian	0.76
WHG	33.06

Oracle

Oracle-4

Spreadsheet



MDLP picked up on an African component, and threw in Amerindian (if true, confirming my suspicions about the maternal origins of my Drayton lineage).

This comparison doesn't show me having quite as much Neolithic (early farmer) ancestry as ftDNA did.

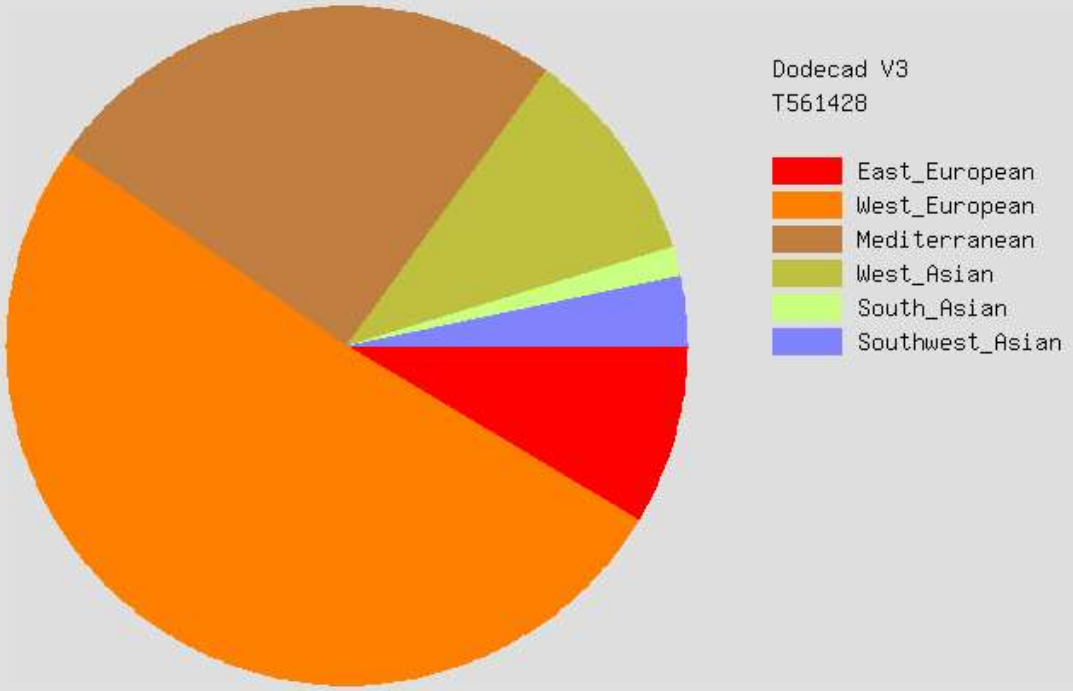
Dodecad V3

Population	
East_European	8.73
West_European	50.89
Mediterranean	25.34
Neo_African	0.22
West_Asian	10.47
South_Asian	1.37
Northeast_Asian	0.26
Southeast_Asian	-
East_African	0.16
Southwest_Asian	2.58
Northwest_African	-
Palaeo_African	-

Oracle

Oracle-4

Spreadsheet



Dodecad gives a more specific African breakdown, associating my <0.1% descent from Parthena with the Neo-African group – Yoruba, Mandenka, Bantu.

Using Oracle with Dodecad

Oracle approximates which ethnic origin would most readily result in the tested person's autosomal genetic profile. In my case . . .

Using 1 population approximation:
1 CEU_HapMap [Germanic + Slav] @ 5.427127
2 N._European_Xing @ 5.475606
3 Argyll_1000 Genomes @ 6.787333
4 Orcadian_HGDP @ 7.229597
5 Orkney_1000 Genomes @ 7.499057
6 German_Dodecad @ 9.916516
7 Mixed_Germanic_Dodecad @ 10.431575
8 French_HGDP @ 10.481844
9 French_Dodecad @ 11.370853
10 Dutch_Dodecad @ 11.968355

- Using 2 populations approximation:
  1 50% Norwegian_Dodecad +50% Tuscan_Xing @ 3.611132

Using 3 populations approximation:
1 50% Norwegian_Dodecad +25% S_Italian_Dodecad +25% Argyll_1000 Genomes @ 1.310395

Using Oracle with Eurasia K9

#	Population	Percent
1	WHG	40.47
2	Early_Neolithic_Farmers	23.71
3	Eastern_Hunter_Gatherer	13.65
4	Caucausus_Hunter_Gatherer	12.85
5	SW_Asian	7.57
6	Ancestral_South_Indian	1.23

Eurasia K9 Oracle Again

Eurasia K9 finds me closer to French than English ancestry, closest of all to Hungarian, but not much more so than Croatian or French.

- Using 1 population approximation:
  1 Hungarian @ 3.546062
  2 Croatian @ 3.660563
  3 French @ 3.795167
  4 English @ 5.848977
  5 Scottish @ 6.810075
  6 Czech @ 7.916175
  7 Norwegian @ 9.837579
  8 Ukrainian @ 9.939985
  9 Bulgarian @ 11.350293
  10 French_South @ 11.374978

Using 2 populations approximation:
1 50% Bulgarian +50% Icelandic @ 2.144954

Using 3 populations approximation:
1 50% English +25% Norwegian +25% Tuscan @ 1.965945

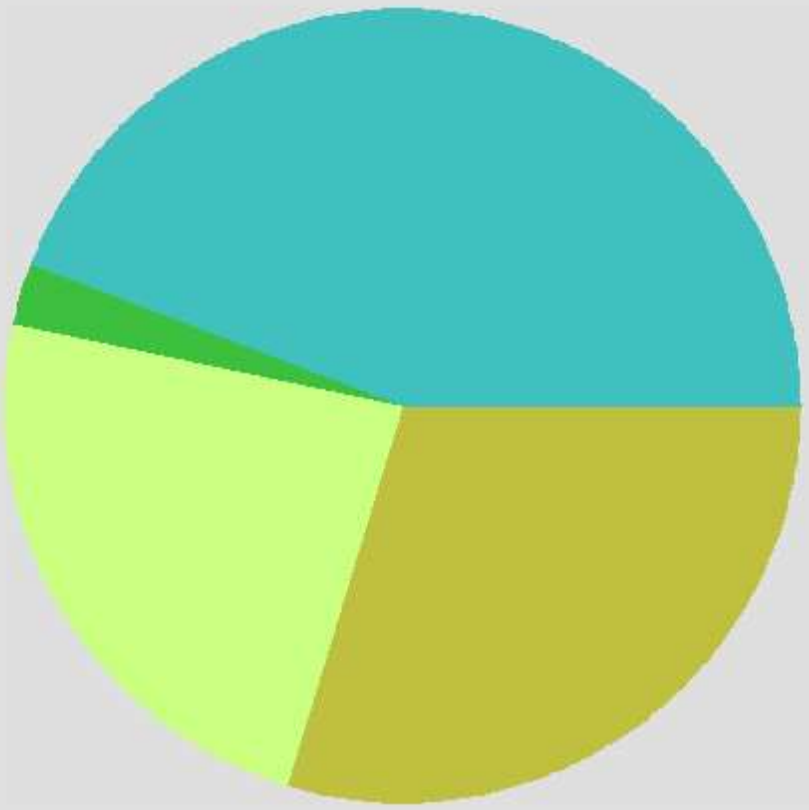
puntDNAL K10 Ancient Admixture

Population	
ASI	-
Sub-Saharan	0.12
Oceanian	0.26
Beringian	-
ENF	29.77
CHG	23.45
Siberian	2.49
E_Asian	-
WHG	42.97
Amerindian	0.94

Oracle

Oracle-4

Spreadsheet



puntDNAL K10 Ancient
T561428

ENF
CHG
Siberian
WHG

Estimates of Ancient DNA

Source	Western Hunter Gatherer	Other Hunter Gatherer	Neolithic Farmer	Metals Age Migrant	Other	Notes
puntDNAL K10	42.97	23.45	29.77	-	3.81	OHG is Caucasus HG. Other is Oceania, Siberia, Amerindian, Sub-Saharan Africa
Eurasia K9	40.47	26.71	23.71	-	8.8	
MDLP K11	33.06	26.35	35.01	-	5.58	OHG is Iran Mesolithic, Eastern Hunter Gatherer. Other is Basal, Siberian, African, Amerindian
ftDNA Ancient	44.00	-	45.00	12%	0	

The Sampling Effect

Source	Western Hunter Gatherer	Other Hunter Gatherer	Neolithic Farmer	Other
MDLP K11 LST	33.06	26.35	35.01	5.58
MDLP K11 GAT	35.56	25.05	33.85	5.54

After a few thousand years of European and British ancestors and a few centuries of Southeastern North American ancestry, our overall genetic proportions tend to converge. My husband's ancestry is much more heavily British than mine, his family isn't even from South Carolina, but his deep ancestry is very similar to mine.

Comparing Estimates of Modern DNA

Source	Europe	Western Europe & Britain except Scandinavia & Southern Europe when breakdown available	Southern Europe	Scandinavia	Other
Ancestry	97	76	7	7	3
Family Tree	98	86	-	12	2
23andMe	97.7	-	-	-	2.3
Eurogenes	87.71	47.05	19.84	20.82	12.29
Dodecad	84.96	50.89	25.34	-	15.04

Which one is right?

All of these measures of ethnicity are probably correct, as data comparisons.

Each represents a statistical comparison with a sample population, either modern or ancient. As long as we remember that, they are interesting and informative.

They are not, however, a way to identify the specific locations of one's ancestors 300 or 500 years ago. To the extent that testing companies present them as such, they are misleading.

Considered properly, each provides a different perspective, a new understanding. So use them and enjoy – with caution.