

Genetic Evidence Relative to the Native American Ancestry of Catharine, the Wife of Lt. John Young (1742-1812)

By

David K. Faux

While the present author has created a 50 plus page document outlining all of the specific details of the testing of descendants of Catharine, the wife of Lt. John Young, the use of this wealth of data at this juncture in time is questionable.

While about 19 descendants have participated in the large autosomal study of the Young family, the primary testing has focused on three individuals:

Larry Young: The reason being that he is the great great great grandson of Catharine and Lt. John Young making him the individual closest generation – wise to the couple and so most likely to show evidence of Native American ancestry.

Betty Yundt: Betty was included in the in depth investigation since she is a descendant of two of the sons of Catharine and Lt. John Young, making it more likely that segments of Catharine's Native American heritage would have passed down to her.

David Faux: The reason for the inclusion of the author is one of practical expedience, plus the fact that he is 8 generations removed from Catharine and Lt. John Young and would be a good example of how many generations segments from a Native American ancestor might be expected to remain detectable.

While huge volumes of data have been generated in this enterprise, it is largely inconsistent and vague. While all three individuals have been shown to possess three or more Native American segments, in none will all of these blocks of DNA be plainly visible across the board. Typically:

Larry has by a good margin the most Native American ancestry in all global tests and subtests where this ancestry is measured.

Betty has the most robust Native American segments.

David has the most informative segments thanks to the cutting edge and high powered testing that has been thrown at it.

That being said, the document that was included has been filed away until such time as there is more consistency seen and less confusion likely for any reader other than the present author.

In the opinion of the present author the weight of genetic evidence show clearly that Catharine was highly admixed. While this would be true for most Mohawks, it appears

to be particularly true for her. This means that in addition to having a European biological father, her only clear Native American line would appear to be the Clan line via the mother's mother's mother and so on. Catharine's great uncle Johannes Crine Anequendahonji ("Dark Belly") was known as "White Hans" and as "a whitish Indian living at the Mohawks". The son of Hans was known as John "Blue Eyed" Green Aronghyengtha. It goes without saying that the evidence here points strongly towards significant admixture in the family, and thus the questionable value of using present day genetic tests to validate Native American ancestry in this family.

The company which did the testing, providing the raw data which can be analyzed using in house features such as the "Native American Ancestry Finder" tool, as well as enabling the author to send away raw data files to be analyzed by third parties who offer sophisticated analyses beyond what is presently available through 23andMe (who uses the Han Chinese as a proxy for Native American in all their comparisons). The explanation that is included with the above tool is that Native American ancestry can at this point (since 2008) only be reliably detected for 5 generations in the past. Beyond that it is likely that there will be no genetic evidence of Native American ancestry. This puts a severe restriction on the use of DNA research to cross validate the genealogy. When all is said and done, the genealogy is excellent and does not profit from being compared and contrasted to measures that may or may not see evidence of Native American background in present day descendants of Catharine.

However, as an example of the types of data one would find in the more extensive file, here follows the most recent data analysis, which will illustrate the types of data being investigated:

Most Recent Examples of Native American Ancestry Testing

Recently many of those doing genetic research using the most up to date admixture tools have contributed their reference samples and algorithms to Gedmatch.com. From here users can upload their raw data and use any of a number of methods to analyze the data to determine percentage of one group or another in the genome, or a chromosome by chromosome analysis (including a colour coded painting of where the ancestral blocks are found).

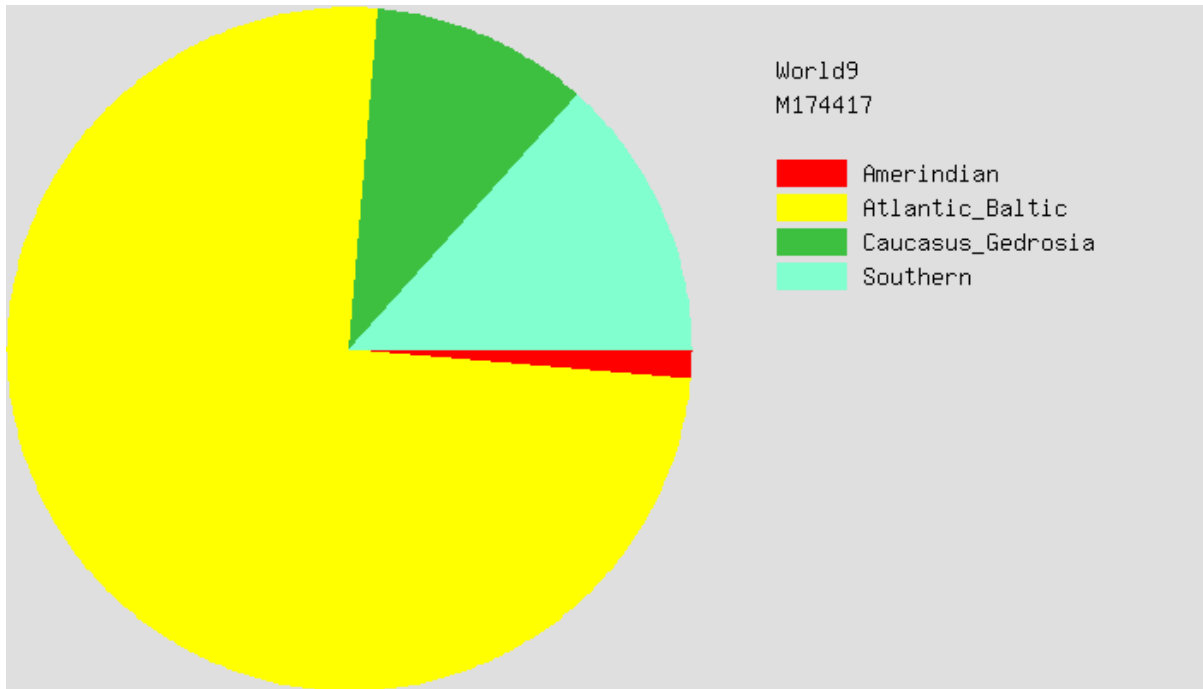
Dodecad World 9 Admixture via Gedmatch.com

Larry Young:

Population

Amerindian	1.28%
East_Asian	-
African	-

Atlantic_Baltic	75.04%
Australasian	0.66%
Siberian	0.55%
Caucasus_Gedrosia	10.32%
Southern	11.74%
South_Asian	0.40%



In the “Oracle” matching of populations using World9, Larry’s results are as follows:

Mixed Mode Population Sharing:

#	Primary Population (source)	Secondary Population (source)	Distance
1	98.7% British (Dodecad)	1.3% PEL30	0.6
2	98.8% British (Dodecad)	1.2% Maya	0.62
3	98.9% British (Dodecad)	1.1% Pima	0.62
4	97.8% British (Dodecad)	2.2% MEX30	0.62

5	98.9% British (Dodecad)	1.1% Karitiana	0.63
6	98.9% British (Dodecad)	1.1% Surui	0.63
7	98.9% British (Dodecad)	1.1% Colombians	0.63
8	98.1% British (Dodecad)	1.9% Ecuadorian	0.64
9	98.6% Kent (1000 Genomes)	1.4% Athabask	0.64
10	98.5% British (Dodecad)	1.5% Athabask	0.66
11	98.6% Cornwall (1000 Genomes)	1.4% Athabask	0.7
12	98.8% Kent (1000 Genomes)	1.2% PEL30	0.71
13	98.9% Kent (1000 Genomes)	1.1% Maya	0.71
14	98.2% Kent (1000 Genomes)	1.8% Ecuadorian	0.71
15	99% Kent (1000 Genomes)	1% Pima	0.71
16	98% Kent (1000 Genomes)	2% MEX30	0.72
17	99% Kent (1000 Genomes)	1% Colombians	0.72
18	99% Kent (1000 Genomes)	1% Karitiana	0.72
19	99% Kent (1000 Genomes)	1% Surui	0.72
20	97.5% British (Dodecad)	2.5% Colombian	0.76

Here it is crystal clear that without exception, Larry’s matchings are only between British and all Native American reference groups (admixed such as PEL30, and unadmixed such as Karitiana). The signal is so strong that the algorithm did not chose even one Northern European group such as French with which to match. This is perhaps the single most persuasive and conclusive DNA evidence of Native American ancestry seen to date. The above pattern is not seen in anyone who does not have Native American ancestry, above about 1%.

With a newer addition, Oracle X, we see two “best fit” predictions:

Pct. Calc. Option 1

0	Unable to determine	0.22%
1	Kent	93.89%
2	French_Basque	2.21%

3	Finnish	2.10%
4	Karitiana	0.41%
5	Papuan	0.41%
6	Athabask	0.38%
7	AthabaskHD4	0.25%
8	Aleut	0.12%
9	WestGreenland	0.01%
10	Koryak	0.00%

Pct. Calc. Option 2

1	Kent	88.11%
2	French_Basque	5.03%
3	Norwegian	4.05%
4	Mordovians	1.10%
5	Karitiana	1.10%
6	Chuvashs	0.23%
7	NAN_Melanesian	0.14%
8	Papuan	0.11%
9	Selkup	0.11%
10	Dolgan	0.01%

David Faux:

Population

Amerindian	0.59%
East_Asian	-
African	-
Atlantic_Baltic	73.55%
Australasian	0.33%
Siberian	0.25%
Caucasus_Gedrosia	13.59%
Southern	11.62%

South_Asian 0.07%

Employing the Oracle X option for a maximum likelihood ancestry estimate we find:

Pct. Calc. Option 1

1	Dutch	97.63%
2	Lezgins	1.12%
3	Chechens	0.81%
4	AthabaskHD4	0.41%
5	Balkars	0.02%
6	MEX30	0.01%
7	Adygei	0.00%
8	Nogais	0.00%
9	North_Ossetians	0.00%
10	Abhkasians	0.00%

Pct. Calc. Option 2

1	Dutch	95.70%
2	Aleut	1.46%
3	Lithuanians	0.76%
4	Chechens	0.71%
5	Lezgins	0.51%
6	AthabaskHD4	0.51%
7	Abhkasians	0.17%
8	CLM30	0.12%
9	Adygei	0.05%
10	Brazilian	0.01%

It is interesting to note that Larry has twice or more the percentage of David on Amerindian, Australasian, South Asian and Siberian. This ratio, or close to it, is what one would expect if this measure was tapping into valid Native American and surrogate DNA. It is also interesting that David, who shows two “Bering Strait” segments (as well as one Native American) segment with the most sophisticated admix program, LAMP+, is also seen here to have a strong Aleut affiliation (as in the above noted segments as well).

The Dodecad world9 Oracle results for others with documented Native American ancestry (including two individuals who are about 4% Cree, one who is 1/32 Oneida – Mohawk, and a fourth individual who is Mi'qmaq about 2.5% with a Native American maternal mitochondrial DNA (mtDNA) haplogroup - X2a1a) have a duplicate of what is seen above with Larry on this test.

The above “Oracle” pattern was not seen with David, most of whose primary matches were German, and all of whose secondary matches were West Asian, the likely explanation being that the signal was much weaker from Native American sources (Larry being 4 generations closer to Catharine Hill Young) – although the Amerindian plus Australasian and Siberian percentages seen in World9 are higher than those whose only ancestry is British.

Update:

- 1) Ancestry Composition: In December 2012 the testing company 23andMe released an excellent feature to replace the old Ancestry Painting which did not include any Native American reference samples – East Asians such as the Han Chinese were used as a proxy. None of the descendants of Catharine, including Larry, received a single Native American segment on this test (except those known via genealogy to have other Indian ancestry via the Cree). This Ancestry Composition Test (replacement to the Ancestry Painting) is particularly damaging to the attempt to match up the genealogical and DNA evidence and may have the power to dismiss the hypothesis that Catharine was Native American – or leaving us with the realization that she had so much admixture that it could not be detected with the newest of measures in even those most proximal to her generation – wise.
- 2) Chromopainter and Finestructure: Since it appears that the segments that are linked to Native America are best observed via the more sophisticated tests not routinely available, it is still possible that NA scores are low to not observed due to a complete lack of relevant reference groups. Previous work has identified segments in the author best described as “Bering Strait” and not typical Native American (via the use of South Americans as proxys). Recently Anders Palsen sent the author another cutting edge analysis which is in keeping with the need to have the groups most similar to Mohawk ancestry from the 1700s – and this may require reference samples from two regions being combined. The following report suggests precisely this:

I have been working for some time to be able to infer minority ancestry using Chromopainter and Finestructure in the autosomes. The included table is from a method I have been working with for some time now manipulating Chromopainter output data. I did actually attempt this earlier using another method in Chromopainter however then you appeared not different to these populations than other West- and Central-Europeans. The point of this analysis is the difference between received and donated. If the difference is 0 or very close to 0 it means that the connection is very recent. If its old the difference is larger. Just like the difference between received and donated for Africans vs non-

Africans is huge but between same ethnicities individuals very small. So it basically tells a "geneology".

It is done in two parts. The first table CLREC and CLDON shows the total cM shared to the different populations both as received and as donated. Here your difference is smallest to the Yugagir, East-Greenlander, Dolgan and a Pima. Sounds good given your geneology, however these total cM also include not only total identical segments but also total cM related or mutated segments so they can be misleading if they show large deviation in mutations. So we need to also take into consideration mutational deviation to narrow it further. So we then go to MUTREC and MOTDON difference. Here we see that the Pima and the Yugagir still stand out as absolutly closest to 0 if we count the 4 closest deviations.

So basically the Pima and the Yugagir reference both passed the divergence control in your case.

I did a similar analysis of Fre4 who is in the same cluster as you. Here you see that Fre4 do have Native American populations high on the list, on one or the other but not on both. Only one selkup show up around 0 both places and this individual do have considerable European admixture. So the Fre4 fails comming through both test.

I have also done a similar analysis of a Saami individual SA2 that appears as the most Siberian like. They are totally different and show no similarity to any populations and no similar pattern as you have or the Fre4 have. Whatever Siberian, East-Asian or Native American affiliation the Saamis have do not have these populations as a direct source. It is probably very ancient or not close to these reference populations.

However you appear in the bullseye for this analysis given your geneology and the analysis Davidski did for you so you may view this is a confirmation of his findings in the autosomes.

Here follows the tables showing the data:

Faux-U3						Faux-U3					
Recipient	CLREC	CLDON	SUM	DEV	REALDEV	Recipient	MUTREC	MUTDON	SUM	DEV	REALDEV
Yukagir9	10,98	11,34	22,32	-0,37	-0,02	Pima	2,67	2,77	5,44	-0,10	-0,02
eastGreenland7	12,15	12,94	25,10	-0,79	-0,03	Yukagir9	2,42	2,63	5,06	-0,21	-0,04
Dolgan6	10,66	11,49	22,15	-0,83	-0,04	buryat24	3,28	3,93	7,21	-0,65	-0,09
Pima	11,69	12,71	24,39	-1,02	-0,04	Karitiana	3,10	3,78	6,88	-0,68	-0,10
Maya	13,83	15,27	29,10	-1,45	-0,05	eastGreenland7	2,42	3,03	5,46	-0,61	-0,11
Yakut	10,14	11,20	21,34	-1,06	-0,05	Nganassan1	3,84	4,94	8,78	-1,11	-0,13
ket1	13,31	14,91	28,22	-1,60	-0,06	Maya	3,49	4,55	8,04	-1,06	-0,13
tuva2	12,52	14,03	26,55	-1,51	-0,06	Daur	4,25	5,65	9,90	-1,40	-0,14
Karitiana	11,14	12,60	23,74	-1,47	-0,06	evenk3	3,34	4,53	7,87	-1,20	-0,15
Daur	14,66	16,65	31,31	-2,00	-0,06	ket1	2,47	3,43	5,89	-0,96	-0,16
Colombian	12,55	14,39	26,94	-1,85	-0,07	Yakut	2,18	3,10	5,28	-0,92	-0,17
Surui	11,30	13,11	24,41	-1,81	-0,07	Tu	4,06	5,93	9,98	-1,87	-0,19
altai10	14,32	16,66	30,98	-2,33	-0,08	tuva2	1,84	3,12	4,96	-1,29	-0,26
selkup7	16,01	18,64	34,66	-2,63	-0,08	Hezhen	2,61	4,45	7,05	-1,84	-0,26
Nganassan1	11,20	13,12	24,31	-1,92	-0,08	chukchi119	2,27	3,94	6,21	-1,67	-0,27
Hezhen	9,82	11,62	21,45	-1,80	-0,08	Uygur	3,59	6,30	9,89	-2,72	-0,27
Uygur	13,83	16,42	30,26	-2,59	-0,09	She	3,82	6,76	10,59	-2,94	-0,28
koryakE78	9,97	11,85	21,82	-1,88	-0,09	Dolgan6	1,67	3,15	4,82	-1,48	-0,31
ket2	13,82	16,58	30,40	-2,76	-0,09	selkup7	2,27	4,28	6,55	-2,01	-0,31
evenk3	11,11	13,39	24,49	-2,28	-0,09	koryakE78	2,03	3,86	5,90	-1,83	-0,31
Han	10,59	12,87	23,46	-2,28	-0,10	Naxi	2,32	4,44	6,75	-2,12	-0,31

Conclusion: In the opinion of the author, the genealogy, which is extremely robust, extensive and consistent should be relied on here until such time as genetic testing allows further information that is both valid and reliable to be added to the mix.

While Larry shows some consistent genetic evidence of Native American ancestry in a number of admixture tests, its “no show” in 23andMe’s Ancestry Composition is puzzling. The fact is that he is between two and four generations closer to Catharine Hill Young than the rest of the family in the Project and should show evidence comparable to others with between 1/32 and 1/128 Native American – which is not generally the case. Of all the participants in this branch of the family, he has the strongest link (as expected by virtue of his generational proximity). However when more robust DNA test measures are applied to the genomes of descendants, output more in line with expected outcomes is seen. Confusion reigns supreme.

A sensible conclusion, largely based on data such as the above charts relating to Larry, and other data shown in the more comprehensive document, is that the DNA evidence at best only weakly supports the paper trail, but little is to be gained at this point to continue focusing on this line of research until more sophisticated technological advances offer the possibility of a conclusive answer. Hence we can leave the matter here until full genome sequencing, further advances in admixture programmes, better selection of ancestral informative markers, and the use of more appropriate Native American reference samples emerge. At that point we can resume the exploration of the genomes of various descendants of Catharine, the wife of Lt. John Young, to see what can be learned from this data source.

David K. Faux, Ph.D., C.Psych. (Retired)
Cypress California and Caledonia Ontario
16 March 2013
Copyright 2011-2013