Analysis of a Native American Segment on Chromosome 18

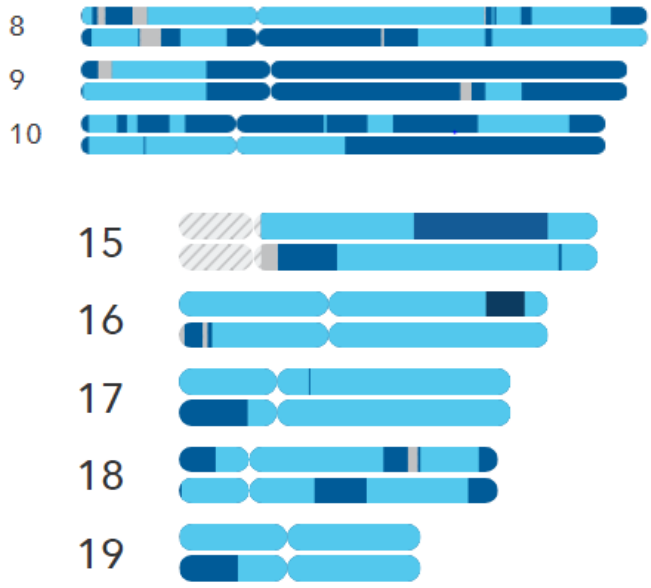
Introduction: For years now the author has maintained that each DNA segment on each chromosome has a "story to tell". It can perhaps be linked to a recent (last 300 years) ancestor, and an exploration of the specifics of a segment may have the potential to reveal its ancient history back (optimistically) to the Paleolithic (Ice Age) Era. Since the author has accumulated considerable information on a segment situated on chromosome 18, which was inherited from his mother, this will be the focus of our "story".

If the genealogy is correct, then this slice of a few million Mb of DNA was inherited by the author from his 6th great grandmother, Catharine Kayakhon (Hill – Brant) Young (1747-1792). If in fact she is the source of the segment, then it is linked to the Wyandot branch of the Bear Clan Astawenserontha Mohawk family of the Six Nations of the Grand River, Haldimand County, Ontario, Canada.

DNA Testing: We will now examine this segment from as many vantagepoints as possible with the hope that it will also reveal its ancient history.

1) **23andMe**:

a) 2018:



Above is part of the 23andMe chromosome array of 2018 for David Faux shown at the 90% confidence level – the gray areas are "Unassigned" by their methods – note in particular

chromosomes 16 (African using various calculators), as well as 8 and 18 (Beringian - Native American) as we will see later.

b) 2019:

Below is one of the recent 80% "Conservative" unassigned segments which (although begin and end points are not specified) appear to align with the more sophisticated analysis describing Native American / Beringian segments on chromosomes 8 and 18; and an African segment on chromosome 16. Perhaps at some point 23andMe will upgrade their algorithm and specify the "unknowns" above. Ironically, one of their early iterations did "recognize" the small African segment and painted it red with the Continent of Africa also being shown in red to reflect this finding.

c) 2020:

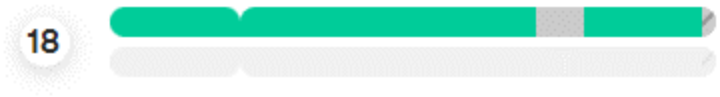
The adjacent segments in blue below represent most of the 0.3% of the genome that at 90% confidence were assigned to "French and German". There is only one other such segment, smaller, and on chromosome 7. It would be reasonable to conclude that this segment and the area in between came from Hannah Adelia (Young) Dawson, the only ancestor who was primarily Colonial German.

The small segment on chromosome 18 below is part of the 2.8% of the genome that is "Unassigned". It is in the same position as the gap above between the two blue - coloured segments.

2) **Ancestry dot Com**:

2020:

Chromosome 18 is from "Parent 1" which in this case is the maternal chromosome since there is a considerable amount of the green colour representing "Scotland" – which is consistent with the mother's genealogy, whereas the paternal side has zero Scottish in the genealogy and none in Parent 2 here. In addition, "Parent 1" includes maternal first cousins and other known relatives on the maternal side. The author's sister also has the same configuration as the author, seen below:



The only way to resolve the "Unassigned" segments would be to have more comprehensive reference samples. For example, there are zero reference samples from Pre-Columbian Eastern North America. If perchance samples such as the skeletal material from the Huron – Wendat ossuaries from Southern Ontario were tested and published, this would offer the opportunity to see how closely these samples match those from South America presently used to represent Native American ancestry – and perhaps transform some "Unassigned" to an "Eastern Native North America" category.

3) **Eurogenes - Haplotype Matching, Dating the Admixture Plus MDS Analysis**:

DavidW has developed a methodology which uses two entirely independent types of approaches, and compares the two. Here he runs a haplotype comparison between participants such as the present author, and individuals in a large set of reference groups. When a segment of interest is detected, DavidW then uses multidimensional scaling to look "microscopically" at the flagged segment. He noted that generally the two methods are in high agreement, but one assesses whether the admixture was ancient or recent (e.g., last 350 years). DavidW used a number of procedures to check on the validity of his findings. He used a program called GEDI-ADMX which is, in his words, *what I used to find the haplotypes, plus PLINK MDS and Eigenstrat PCA to double check them. I also used LAMP 2.4 (+ WINPOP) on a friend's computer to scan your genome* (DavidW, personal communication, 28 May 2011).

When running some segments with the above highly enhanced and upgraded software and algorithm, DavidW located four relevant segments, all of which reflected "recent" admixture (plus a fourth on chromosome 2 which showed "ancient" admixture. This was later shown to be Finnish, where some segments are often highly similar to Native American. This misattribution was only resolved after Finnish reference samples were included in the mix.

Here is a screenshot of the near neighbours on **chromosome 8**:

The author's icon is again closest to the Mayans and Athabaskan reference samples, with the European group far off to the right hand side, and Native American groups to the left hand side. Again the author's icon would not amongst either group presumably due to possessing one European and one Native American segment at that location. The specific location is between positions 18,831,286 and 21,947,885 on Chromosome 8.

Below is a screenshot of the near neighbours on **chromosome 18**:

The above segment extends from position 39,176,819 to 44,963,452, which is a segment 5,786,633 or almost 6 Mb in length. DavidW told the author that the segment is likely somewhat larger but he only explored the most clear cut part. The "near neighbours" here seem to include considerably more East Siberian samples than seen above for chromosome 8.

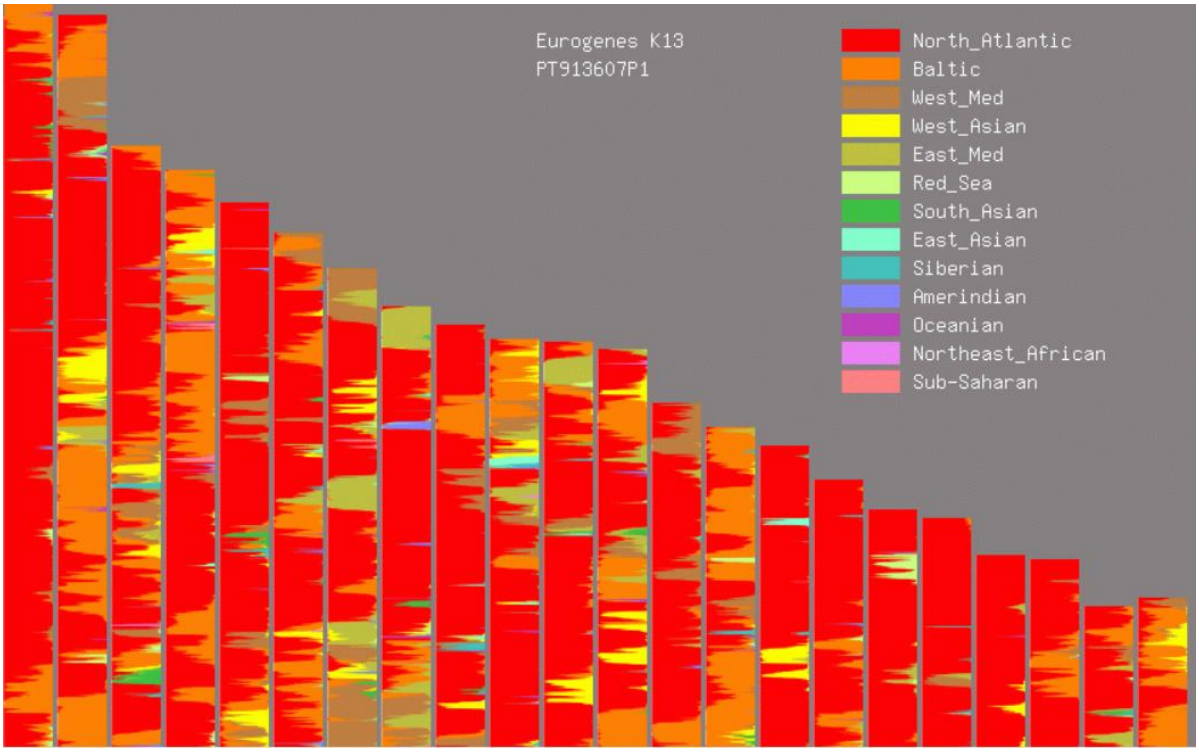
He labeled two of the findings, those on chromosome 8 and chromosome18, as "**Bering Strait segments**", and chr 10 as "**Amerindian"**, and the testing revealed that each was a recent admixture event that occurred less than 350 years ago.

It will be important to determine whether these segments were inherited from the mother or father. This can be accomplished in a number of ways, and over the years the author has determined which is which. The end result can be seen quite clearly, for example, in the chromosome paintings which are available with various calculators available at Gedmatch.com.

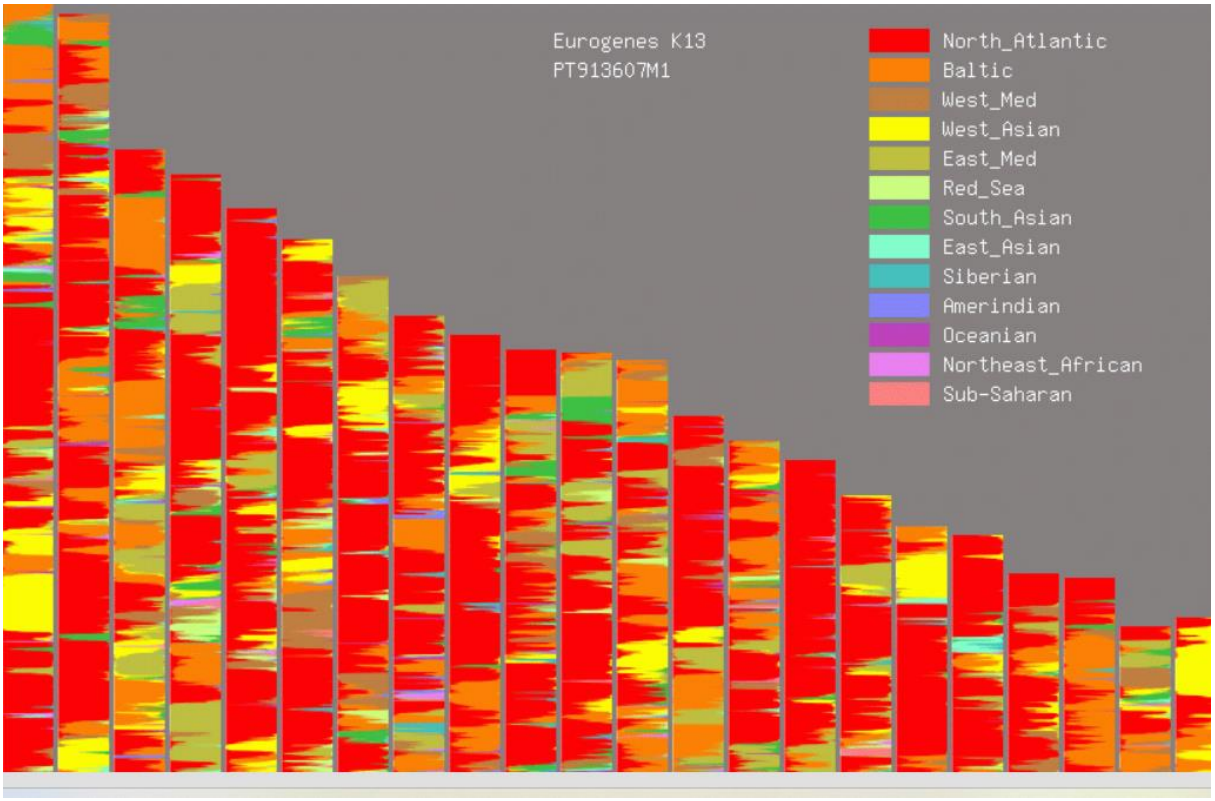
The two paintings from the Gedmatch calculator associated with Eurogenes K13 are shown below, pertaining to the parents of the author.

The first illustration below, pertaining to the phased contribution of the father to the genome of the author shows does not show any Native American or proxy such as Siberian or East Asian on chromosome 8 or 18. The subsequent painting, however, shows rather clearly the parental origin of both segments.

Paternal - Father:



Maternal Mother:



The above diagram clearly shows the Native American segments on chromosomes 8 and 18 (plus scattered smaller NA segments), and so was inherited by the author from his maternal side.

In summary, as to general classification of the above segment, DavidW said that the segment on chromosomes 8 and 18 are "Bering Strait" or "Beringian" with a mix of Northeastern Eurasian and North Amerindian. According to the software he used, as noted, the two segments reflect recent admixture (e.g., last 350 years). Later in this work we will examine the study by Wang et al. and Perego et al., as well as more recent studies, re the "Beringian" aspect of Native American DNA in North America. For the sake of clarity, only the more robust segment on chromosome 18 will be discussed further in this study.

It is important to note that DavidW found that unadmixed English and Irish did not show these "Bering Strait" segments. Any facsimile tended to be "surrounded" by Europeans and a scattered, indecipherable, group of Eurasians. So there was no pattern extant. In David's words, there is the occasional European with some anomalies, but these are *blips that usually don't even get them past the last Frenchmen on the MDS plot*.

DavidW is very confident about his approach in that, *this haplotype matching software doesn't seem to make mistakes* (DavidW, personal communication, 16 March 2011). Hence what is likely is that at some point, as more reference samples are added, that these plots will tell us more about each segment's migrational history, and hence the migration patterns of for example the Mohawk and those who they absorbed into their genomic mix over the centuries.

4) **yourDNAportal**:

Chromosome Analysis – Exploring Only Chromosome 18

Modern Native American Groups - Results:

Chromosynthesis

Native American Modern Chromosynthesis

Explore your DNA
via the power of Chromosynthesis!

By utilising chromosomal analysis, you can dig deeper into the more obscure areas of your ancestry. The fine detail of a chromosomal test, looks at each chromosome individually, allowing you to capture ancestry that may be missed by other tests.

Chromosynthesis is an excellent tool for capturing even the most distant ancestry. Each chromosome contains unique information and can have a very different autosomal profile. Usually the first 10 chromosomes are the most relevant, while the last 2 are less informative. A small percentage (1% of a chromosome) equals approximately 0.045% of the entire autosome. By analysing each chromosome, we can reveal both a deeper and more detailed ancestry than other tests.

Chromosome 18

Speculative results

Population	Value
Basque	75%
South Asian	18.75%
Shipibo Peru	6.25%

Conservative results

Population	Value
Basque	75%
South Asian	25%

Analysis using Genomewide Modern Native American Groups - Results:

Chromosome 18

Speculative results

Population	Value
Basque	75%
South Asian	18.75%
Native American	6.25%

Conservative results

Population	Value
Basque	75%
South Asian	25%

Ancient NA Reference Groups - Results:

Chromosome 18

Speculative results

Population	Value
Ancient West Mediterranean	75%
Ancient South Asian	18.75%
Ancient West Amazonian	6.25%

Conservative results

Population	Value
Ancient West Mediterranean	75%
Ancient South Asian	25%

5) **Archaic Samples**:

Gedmatch has obtained raw data from some of the archaeological samples from across the world which have been published in the academic literature. It is possible to compare the genomes of individuals alive today with each of these samples – when the SNP coverage is high enough. Excellent results are available from an ancient North American, the Clovis sample from Montana. The comparisons will be anthropological not genealogical so much lower criteria for matching would be used. What is observed in comparing the Clovis sample with that of the author, while there are many "matches" around the telomeres (ends) of various chromosomes (although not seen below), there are few in the main body of any chromosome. The author's most robust segment (size of cM and number of SNPs) match is on chromosome 18. The data is as follows:

Comparing Kit A265430 (David K Faux) and F999919 (Clovis, Montana, 12.5ky)

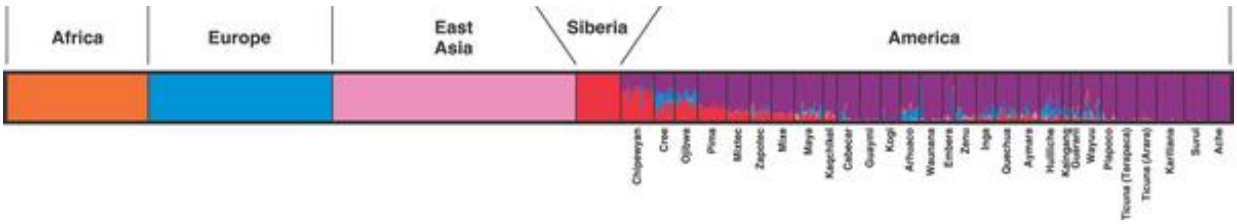
Minimum threshold size to be included in total = 200 SNPs
Mismatch-bunching Limit = 100 SNPs
Minimum segment cM to be included in total = 2.0 cM

Chr	Start Location	End Location	Centimorgans (cM)	SNPs
18	44,175,212	45,508,007	2.3	362

It is perhaps significant that this is the location that is also the most "robust" Native American segment based on multiple other analyses. The fact that it is only part of the whole segment noted earlier is still amazing in that modern Native Americans who have taken these DNA tests only have up to 5 cM matches with this sample, and only "spotty" – across the genome in general.

Academic Studies to Help Interpret the Data – "Beringian" Ancestry in Early Native American Genomes:

1) The above "Beringian" findings are also consistent with the results of the Wang et al., 2007 study diagrammed below:



Under the America category on the right of the diagram, the tribe / nation listed first at the bottom, the Chipewyan = Athabaskan is one of the reference samples seen in the three "PCA" charts above.

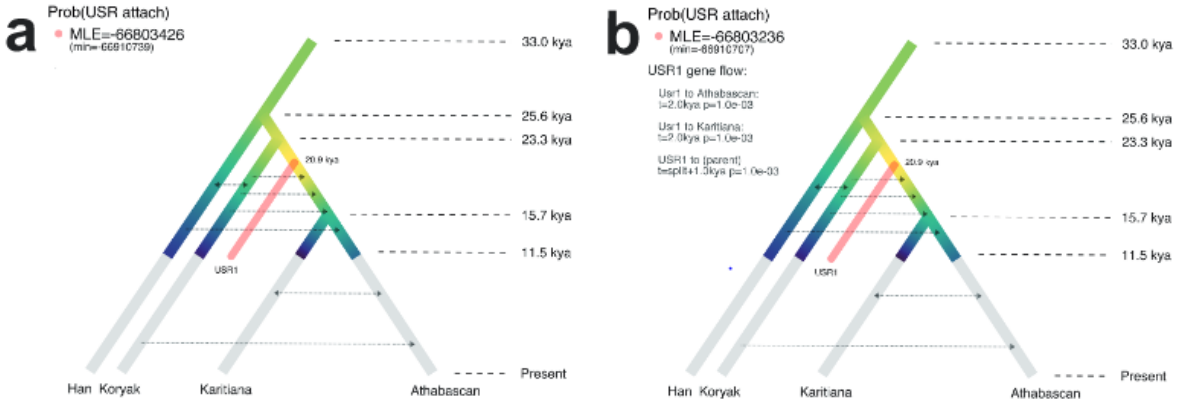
This Northern Canadian nation is one of the "Beringian" groups. They have a significant Siberian component, in contrast to the South American group at the end of the list such as the Karitiana who have none.

2) A paper by Perego et al., 2010 reports that nuclear and morphometric data such as the above diagram shows, *that some native groups from northern North America harbor stronger genetic similarities with some eastern Siberian groups than with Native American groups located more in the South* (p.1). This would highlight the possible importance of the above segments on chromosomes 8 and 18 as of potentially valid Native American origin – although having zero reference samples from anywhere in Eastern North America will always leave questions about how much true NA might be missed and in effect become false negative assignments by the various algorithms used by the testing companies.

3) Azevedo et al. (2011) modelled various migration scenarios using morphometrics (skull shape) integrated with genetic data. They concluded that, *the current genetic diversity of Amerindian populations is best explained by a model involving recurrent gene flow between Asia and the Americas, after initial colonization.* They also highlight the "Beringian" aspect of the present - day population as showing, *persistence of some DNA lineages that can be defined as 'Beringian' instead of ancestral Asian or autochthonous American.* This emphasizes, *the importance of gene flow among Circum-Arctic groups in order to explain recent genetic, skeletal, linguistic and archaeological data* (p.12).

4) The Flagontov et al., 2016 study (BioRxiv), was entitled, "*Na-Dene populations descend from the Paleo-Eskimo migration into America.*" They reported that, "*All methods detected Central and West Siberian ancestry exclusively in a fraction of modern day Na-Dene individuals, but not in other Native Americans. Our results are consistent with gene flow from Paleo-Eskimos into the First American ancestors of Na-Dene*." Furthermore, "*Substantial admixture of 22.3 – 23.8% from Siberians (22 genomes) into Northern Athabaskans was revealed in our model*."

5) In January 2018 a key article by Moreno – Mayar et al., was published in Nature. It is entitled, "*Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans"*. In summary, via complete genome analysis of an Alaskan Native dating to about 11.5 thousand years ago, they detected a third grouping among those who crossed the Bering Strait about 20 thousand years ago (this figure continues to be pushed back in time) and diverged not into two but three groups from whom all Native Americans are descended. For many years now it has been recognized that there is a Southern Native American branch from whom most are descended, but also a distinctive Northern Native American branch represented by the Athabaskan / Na – Dene people. The third group most recently identified has been termed "Ancient Beringians", a root stock group who spawned the other two but remained as a distinct entity for many thousands of years in their Alaskan homeland and then disappeared about 7,000 years ago. The sample carried 56% Native American and 44% Eurasian components. Their apparent range (based on the archaeological evidence) was from northern Alaska to northern Alberta. It is thought that perhaps the Athabaskan peoples killed them, merged with them – but also perhaps dispersed them to locations far to the east.



The above is Figure S29 from the above article showing the position of the sample USR1 (pink) in relation to other groups. The junction at 23.3 thousand years is where the Mal'ta sample fits. This individual was shown to be Northeastern Asian, but with no East Asian admixture – and this group is thought to be ancestral to all Native American people – and is where the European – like aspect to early Native Americans derives.

6) In June 2018 an article entitled, "*Ancient human parallel lineages within North American contributed to a coastal expansion"* (Scheib et al., Science 360, 2018, pp.1024 – 1027) was published. It appears clear via an analysis of 91 ancient Native American genomes "*from California to Southwestern Ontario*" that early on there were two groups of ancestral populations in the Americas, and they represent distinct groups that later merged (at least in the areas tested) primarily in Central and South American populations.

Although there is an "*Ancient Southwestern Ontario (ASO)*" grouping, the details are far from clear. There are only two samples, one each from two Huron-Wendat ossuaries (one near Vaughn and the other near Lake Couchiching). These samples were explored for mtDNA and Y chromosome DNA but only one sample included the latter. Furthermore, a grand total of only two, one from each ossuary were tested for autosomal DNA and hence the usefulness of the data to addressing the hypothesis of the article is questionable due to small sample size.

The other group is from a cemetery in Windsor, Ontario, with dates 1200 to 1450 AD. These would likely be Algonquin ancestors (but this is little more than an assumption based on who is residing today closest to the location). The authors report that 14 individuals were sampled, and Supplementary Table 1 shows that these were included in the autosomal analysis – but no Y chromosome data is included. They also note that these samples were from "*near modern Algonquian-speaking populations*", and that they cluster, "*with modern Algonquian speaking populations*" – but who these modern groups is not noted (although Tables S8 shows N=5). Combining the Windsor and Central Ontario samples conflates potential Algonquin and Iroquoian groups. Some of their Siberian samples are put into N=2 groups.

It begs the question that, since there is NO Native American population of today living in the eastern tier of North America that has been tested by standard chip testing – what are we to

expect – what is a DNA motif characteristic of say the Senecas? It must be emphasized that at this point we cannot simply assume that say Mohawk people are "close enough" to say Mayan for purposes of population genetics. Instead they may share affinities with the Maya, but their Native American DNA may also share affinities to say the Thai (Southeast Asia) and the Tlingit (West Coast Canada) and some of the Siberian groups.

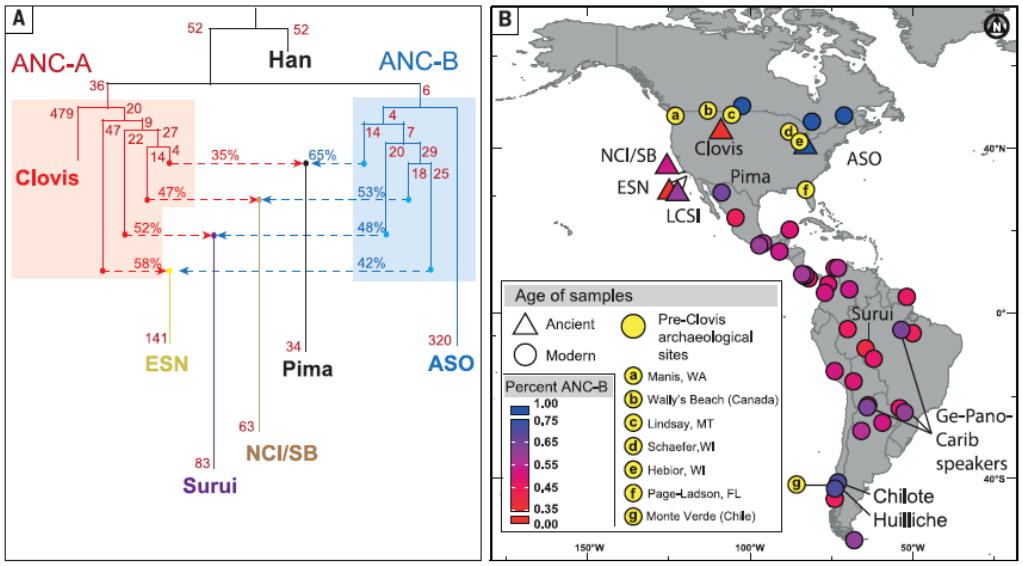


Fig. 2. Visual model of ancestry components and distribution of proportions in the Americas. (**A**) A model with four admixture events that offers a good fit to the data ($Z = 0.888$) (*15*). (**B**) Scale of ANC-B ancestry from 0% in Anzick-1 to 100% in the ASO and modern Algonquian-speaking populations.

Clearly the Southwestern Ontario group (ANC-B, ASO) are significantly different from the other groups of Native Americans – so the question becomes how much can we rely of the use of the Maya and similar groups to represent say the Mohawk for the purposes of ancestry testing. What is probably true, however, is that an unknown percentage of the Mohawk genome is outside the limits of the Mayan motif such that we use any of these groups to represent others such as the Mohawk – at our own peril.

7) In November 2018 Posth et al. published an article entitled, "*Reconstructing the Deep Population History of Central and South America*" (Cell, 175, pp. 1-13). They re-analyzed the data above by Scheib and found that there was no apparent contribution of the Eastern North American Group ("Ancient Southern Ontario") to the present - day genomic spectrum of South Americans. They in fact give the Southern Ontario group (apparent ancestors of the Wendat – Huron or Algonquins) their own migration (Ancestral – B) which split from the group that made up the early migrations to Western North America and South America (Ancestral – A) sometime between 17,500 and 14,600 years ago. Doubtless there was some later mixture in North America between the A and B groups but to what extent is unknown. The Clovis sample is classified with the A group, which was also the predominant group seen to date in early South America, but this Clovis input appears to have largely disappeared there. This is of some importance when we later integrate this academic finding with the data pertaining to the author. The A and B "split" is represented in the article in the following illustration.

Graphical Abstract

Summary and Conclusions:

Overview Description of Segment on Chromosome 18: The combined data above, relating to the genome of the author, point strongly to a **Native American** (NA) segment on the Q arm of chromosome 18 which is over 6 Mb in length. Note that when Native American reference samples have not been used (e.g., Dodecad 7b) the programme reverts to East Asian or Siberian or the closest proxy, but will be displayed as Native American with calculators using appropriate NA reference samples (e.g., Dodecad World9). It is also possible that some NA segments may not be labeled as such potentially due to the use of "missing" reference samples (e.g., none from Eastern Canada); or it may be a "phantom segment", merely noise in the system – especially with very small segments.

The segment on chromosome 18 is not what might be termed "*large and robust*" however it appears consistently on diverse calculators. The Native American segment on chromosome 18 can be considered as cross validated by various DNA methods. However, it is relatively small, and would be missed by many calculators that are not "fine - tuned" with methods that are far too expensive and time consuming for commercial DNA testing companies. There are also a

considerable number of smaller "slivers" of NA and East Asian coloured regions as seen in the Gedmatch Chromosome Paintings above, but these are below the bar set by commercial entities so they are either ignored, lumped in with another Eurasian group, or put in the "Unassigned" category.

Difficulties with Commercial DNA Test Companies: The size of the segment under consideration in the current study exposes a "genetic problem" facing genealogists looking to validate the ancestry of an ancestor born in say the 1600s or 1700s. While it is almost certain that you inherit something from all 16 of your great great grandparents, it becomes something of a 50/50 proposition for all 3rd great grandparents, and about 1% likelihood that you will inherit (detectible) DNA from all 64 of your 4th great grandparents. So even though ancestors correctly appear in your genealogical tree, some are not found in your genetic tree (or if they are, the segments are too small to rule out a false positive finding). It is easier to identify a segment explored for biogeographical purposes (ancestry / ethnicity) than genealogical. However even with ancestry testing Northwestern Europe is notoriously difficult to accurately parse, and DNA testing companies provide inconsistent (sometimes extremely so) estimates of ancestry from England, Holland, France and even Scandinavia. A non – European ancestry segment is generally easier for the test companies to flag due to its relative uniqueness. Genealogy testing is more problematic than ancestry / ethnicity testing such that different criteria need to be employed. The genealogical problem of determining whether a segment is identical by descent (IBD); or identical by state (IBS), meaning a match that is a false positive and a function of belonging to a broad category such as Northwestern European where the observed configuration is common.

Detailed Analysis Using Specialized Methods: The detailed work by DavidW using methods (multiple ancestry tests) generally only applied in academic studies can identify those segments which are "just under the bar or radar" for 23andMe's Ancestry Composition (which classify these specific segments as "Unassigned" as does Ancestry). He has provided a more refined analysis by showing the individual reference samples who "stand" closest to the author's icon in a Principal Components Analysis. In 2011 DavidW identified two segments in the author's genome (chromosomes 8 and 18) as "Bering Strait" or **Beringian** since they were not European, but appeared to be different from typical Native American segments (such as the one he found on chromosome 10).

Academic Studies: The recent finding using ancient DNA analysis of an 11.5 - thousand - year - old sample from Alaska seems to link up to the description of the segment on chromosome 18 (and chromosome 8). This third group of Ancient Americans has been dubbed as "Ancient Beringians". This in turn calls into question whether the segments descend from this group that is assumed to have "disappeared". However in fact no one knows whether they may have migrated east and later mixed with Southern Native Americans coming north via the Mississippi Valley corridor (for example). Their relationship to the Group B in Scheib's study (2018) is tantalizing but not entirely clear.

Source of Segment on Chromosome 18: The most recent Ancestry dot com Chromosome Painting shows that the "Unassigned" segment on chromosome 18 is wedged between two larger segments identified as German or French. The only ancestor who fits this description is the author's great great grandmother Hannah Adelia (Young) Dawson whose own great great grandmother is Catharine, the candidate "donor" of the mystery segment. It is important to note that both the

author and his sister have the same configuration such that the segment is less likely to be a false positive. In addition, the segment under consideration does not fall within any described area of the genome with "bunch up", identical by state, unreliable regions.

Statistical Aspects of Inheritance: There could be many of what this author terms "slivers" of DNA that may be only say 2 Mb in length (say 3 cM) peppered about the genome. However currently none of the commercial firms are able to match up these segments with those of other descendants of a particular ancestor since they typically set the size bar at 8 cM (it was 5 cM when 23andMe first started operations) since the likelihood of a false positive would be high. The same situation emerges in genealogy with the attempt to find a genetic match for say 4th cousins (who often share zero detectable DNA).

Since the documented NA ancestor in the present case is on the mother's side, and since the NA ancestor was the author's 6th great grandmother (one of 128 maternal sixth great grandparents, plus 128 paternal sixth great grandparents). The average inheritance would be an average of about 0.39% of the genome or 27.5 cM (centrimorgan, a measure of genetic distance) and 2.04 segments. The comparable value in Mb (megabytes, a measure of physical distance) is about 12 Mb from each 6th great grandparent. However simply through random factors it is possible that the author has inherited say 30 cM and 3 segments from this ancestor, or only "slivers" at best. The probability of no detectible sharing with a 6th great grandparent is 17.8%. However, unexpected apparent "oversharing" is occasionally seen such as the observed contribution from the author's 6th great grandfather Johan Georg Windecker born 1715 Schoharie, NY – whose contribution is doubtless enhanced by the fact that he is found in two ancestral lines due to a cousin marriage – and had many descendants.

The Role of European Admixture in a Native American Ancestor: A likely major issue here (based on the genealogy) is admixture. It is possible that Catharine was only ¼ biologically Native American – although 100% culturally a Mohawk. Therefore, all the figures relating to finding a NA segment, and the expected size, would have to be pushed back to say an 8th great grandparent level. It would appear rather fortunate to have identified two NA "Beringian" segments considering these circumstances. In retrospect it would have been more advantageous to test the author's sister with the more sophisticated test procedures since she has consistently more "Native American" (or proxy such as Siberian or East Asian) relative to the author.

In the manuscript studying the evidence pertaining to Catharine, wife of Lt. John Young, it was proposed that she was the daughter of Sir William Johnson. This hypothesis has received some support from a 10 cM DNA match at Ancestry dot com between the author's sister and "KB". The latter is a descendant of Christopher Johnson and Anne Warren of Smithstown, County Meath, Ireland via their daughter Anne Johnson. Christopher and Anne were also the parents of Sir William Johnson. See here for further information on the proposed parentage of Catharine.

Bottom Line: The point of this study was to explore the story that could be told by one single segment on one single chromosome. This seems to have been, to a degree, accomplished.

Dr. David K. Faux

14 May 2023. Revised: 22 May 2023