

African DNA Segments in White Americans: An Exploration of Similarities with African-Americans in the Eurogenes Project

Introduction: Initial ancestry test results from 23andMe, as reflected in their Ancestry Painting feature obtained with a genome wide scan of 550,000 single nucleotide (SNP) markers is shown below. Here both an “Old Stock” (Colonial American - MF) wife (from Missouri USA) and her husband (from Ontario Canada - DF) appear to be 100% European. It would be “end of story” if one wished to accept the data as presented by 23andMe’s **Ancestry Painting feature** “as is”, “as of” April of 2008. At that time, they had only 3 ancestral groupings for reference populations – European, Asian, and African. Frequent additions of other reference groups changed the complexion of the painting (now called Ancestry Composition) and many, especially Colonial North Americans, have seen more variety in their ancestral percentages. One iteration after 2008 saw the paintings of both the above include, for example, include small red coloured segments of Sub-Saharan African ancestry on chromosomes 8 (MF) and 16 (DF). However later these segments were merely left as “Unassigned” and “greyed in”. In the most recent iteration (2020), 23andMe have added, for example, six “East Asian and Indigenous American” segments to the display of MF; and a large “North African” segment to chromosome 7 of DF. The earlier Sub-Saharan African segments have never reappeared since, presumably, they are now below the size (in length of segment expressed as Mb or cM) threshold used by 23andMe.

What follows is a deeper exploration of this initial finding. We will see how more “in depth” techniques provide strong evidence of a small percentage of African ancestry in both of the above individuals, which was inherited through their respective mothers (e.g., the husband’s maternal uncle has the identical African motif at the same chromosomal location). So, there was evidence with 23andMe that neither US72 nor CA4 (code numbers for MF and DF from the Eurogenes Project) have any minority African ancestry. Not only will it become clear that the above is a false negative for African, but can be described in the same manner for Native American. However here our focus is only on the findings in relation to African ancestry.

More sophisticated work using techniques that would not be commercially viable (due to costs and computing time) has called into question the accuracy of the findings of Ancestry Painting for those whose minority ancestry relates to admixture many generations back and is relatively small. Valid ancestry may not be detected due to the algorithm used here, and / or the inappropriateness of the reference samples. To remedy problems of this nature, it is necessary to use the raw data (over a half million SNP markers) provided by 23andMe, but send it for analysis by two entirely different methods (“aftermarket products”). In this case each method gave the same findings for both customers – US72 and CA4 have small but valid African segments “not seen” by Ancestry Painting (except in one earlier iteration). What follows is the result of supplementary testing by Dr. Doug McDonald, and DavidW with the Eurogenes Project.

The goal of the present study is to challenge the finding of zero percent African in some but not all iterations of 23andMe’s Ancestry Painting. In order to accomplish the goals here:

- 1) The findings of Dr. McDonald’s segment analysis will be presented via an ancestral chart of the chromosomes showing one of the two (maternal and paternal) chromosomes

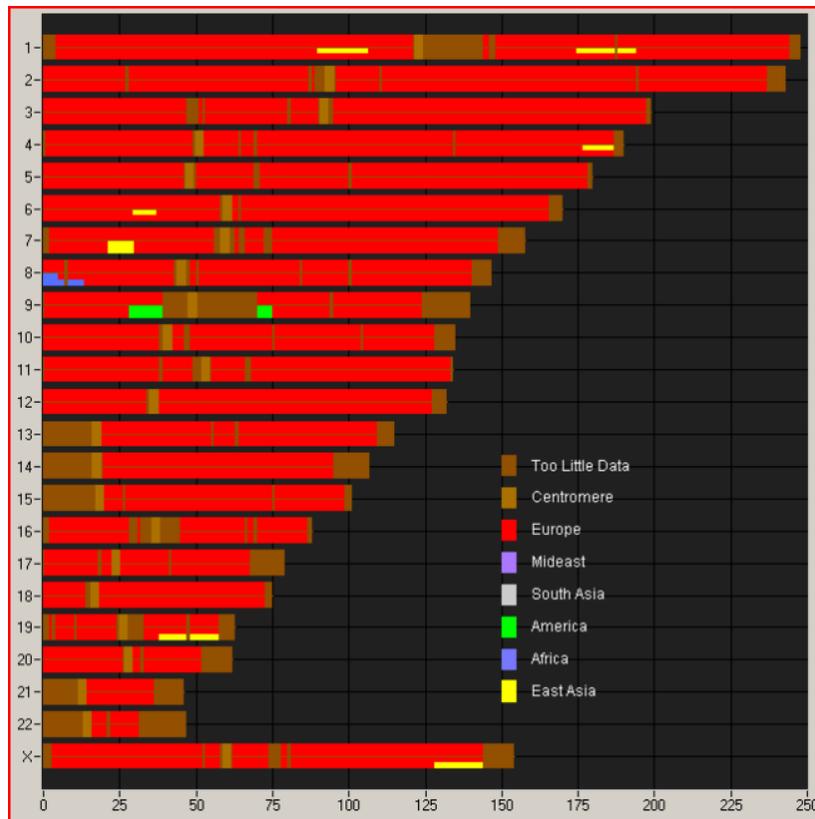
having an African segment. This finding is shown specifically at the left tip (P versus Q side) of chromosome 8 for US72, and that for chromosome 16 for CA4.

2) The HRR mosaic and associated PC-MDS plot (an explanation of this terminology will be provided later) conducted by David W of the Eurogenes Project will be provided as cross validation of the above.

3) The output of the above two tests will be compared to data from the Ancestry Painting feature of 23andMe that pertain to five African – Americans who are also participants in the Eurogenes Project. The object is to see if those who cluster with US72 and / or CA4 have one European and one African segment at the “African locations” shown on chromosome 8 and 16 respectively, as is predicted based on the rationale for the positioning of each of the above individuals on a PC-MDS (principal components-multidimensional scaling) plot (to visualize where each of them clusters).

Chromosome 8 African Segment in US72

US72 - Segment Analysis (McDonald):



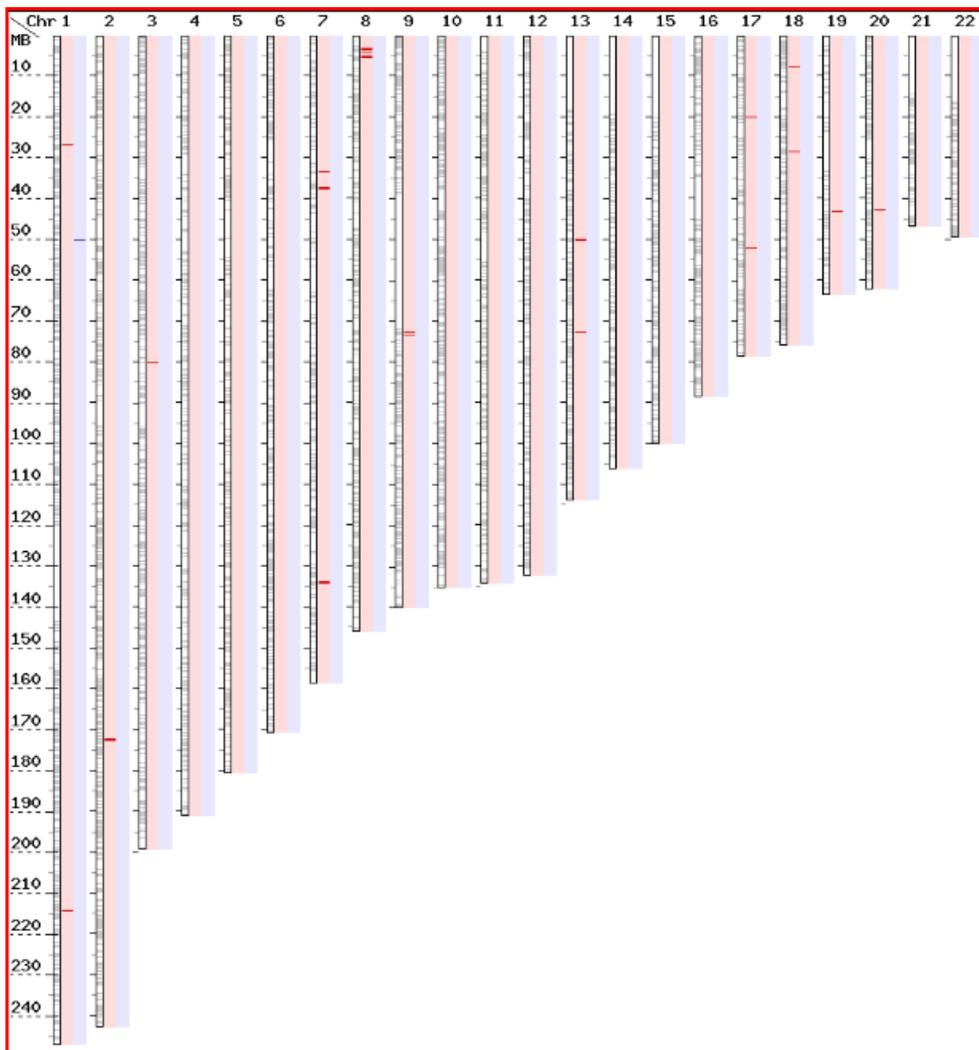
The above diagram shows that US72 has an African segment near the “p end” (smaller of the two parts separated by a “pinch” at the centromere) of chromosome 8 extending to approximately 10 Mb, at which point there are only European segments through the rest of the chromosome. This is the only African segment located in the genome of US72 using this method.

It is generally accepted that for a finding to be considered valid, it must show up in instances where two or more methodologies are employed.

The RHH mosaic work of DavidW, exploring rare alleles in European genomes, was available for use in matters of this nature (especially in relation to African segments).

US72 – RHH Mosaic:

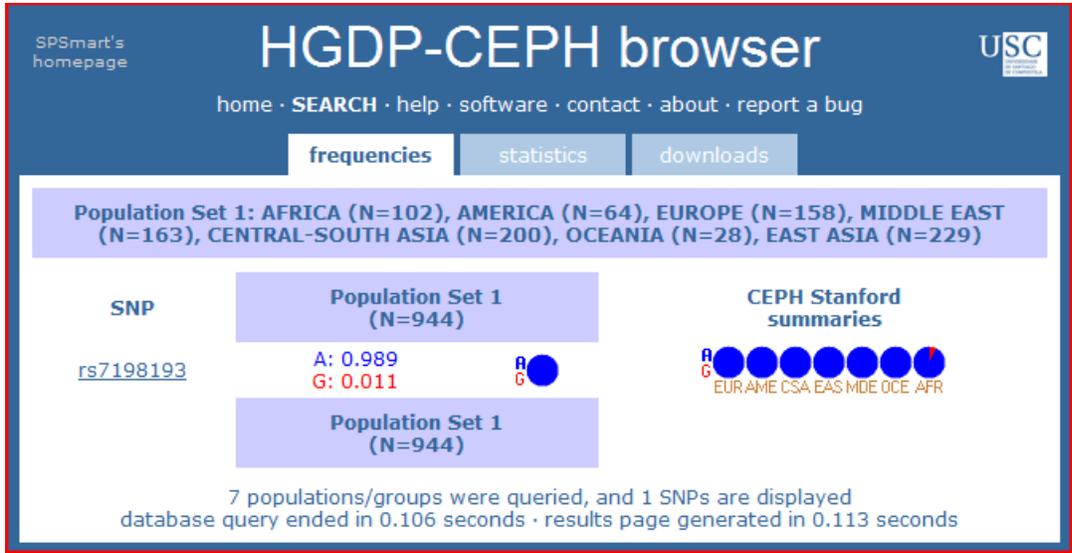
In order to interpret the following mosaic chart, it will be necessary to refer to the [Eurogenes](#) blog of DavidW, including the research article by McGinnis et al., 2010 noted therein. Here follows the mosaic of US72 at the 0.01 resolution level:



The cluster of HETs (heterozygous SNP markers that are extremely rare in Western Europeans) clearly shows near the tip of chromosome 8. One way to explore the specifics of this finding is to identify the HETs in relation to how common the alleles are in the population at large. In other words how frequently the participant's version of the SNP(s), such as a C or a T (among the A, C,

T, G possibilities) nucleotide base, found in say Africans and Europeans. This can be accomplished by using a tool such as dbSNP to portray this information in pie chart form as seen below.

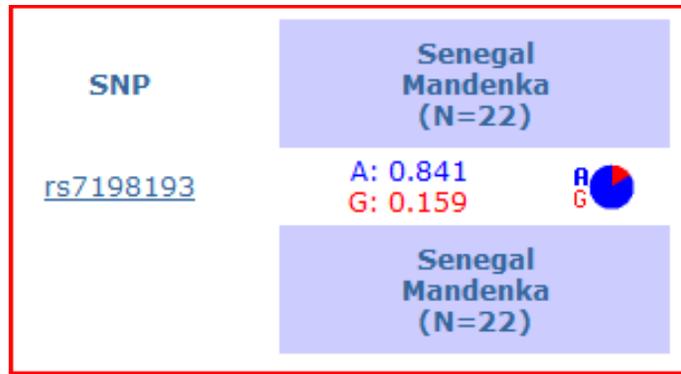
Ancestral Informative Markers from dbSNP:



As seen above, for the allele G for SNP rs7198193, only Africans have been shown to have this version. The following African groups were tested for this SNP:

- AFRICA**
 - C. African Republic - Biaka Pygmy
 - D. R. of Congo - Mbuti Pygmy
 - Kenya - Bantu
 - Namibia - San
 - Nigeria - Yoruba
 - Senegal - Mandenka
 - South Africa - Bantu

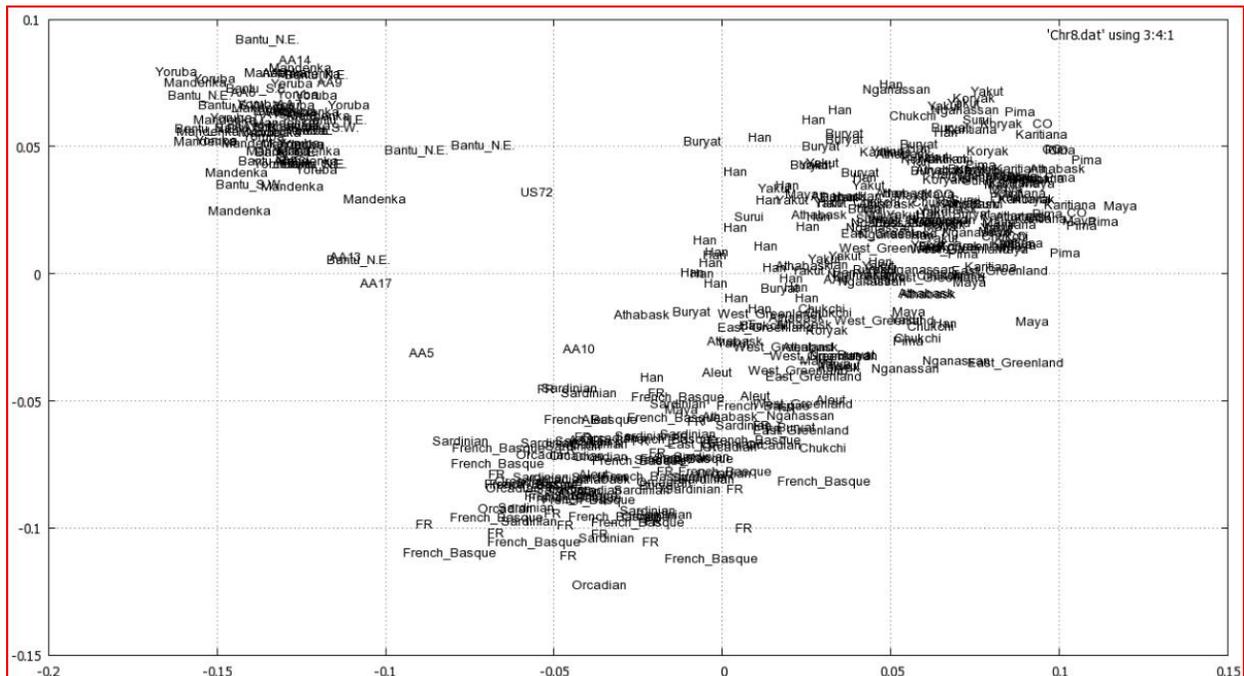
The results for one of these groups is shown below:



This can be combined with the PC-MDS plotting below to examine a specific segment, which appears to have an African affiliation. The MDS (Multi Dimensional Scaling) plot is more or less the same as the Advanced Global Similarity feature at 23andMe. Using this method DavidW can determine who within a reference or participant group is closest to whom, at the identified location. With the awareness that US72 appears to have one European and one African segment at this single location, it is expected that the icon of US72 should be found between the tight clusters of the two ancestral groups (European and African). Thus, those who are physically closest on the plot should likely have the same motif with one European segment and one African segment across this 10 Mb span shown in the McDonald segment painting and the RHH mosaic.

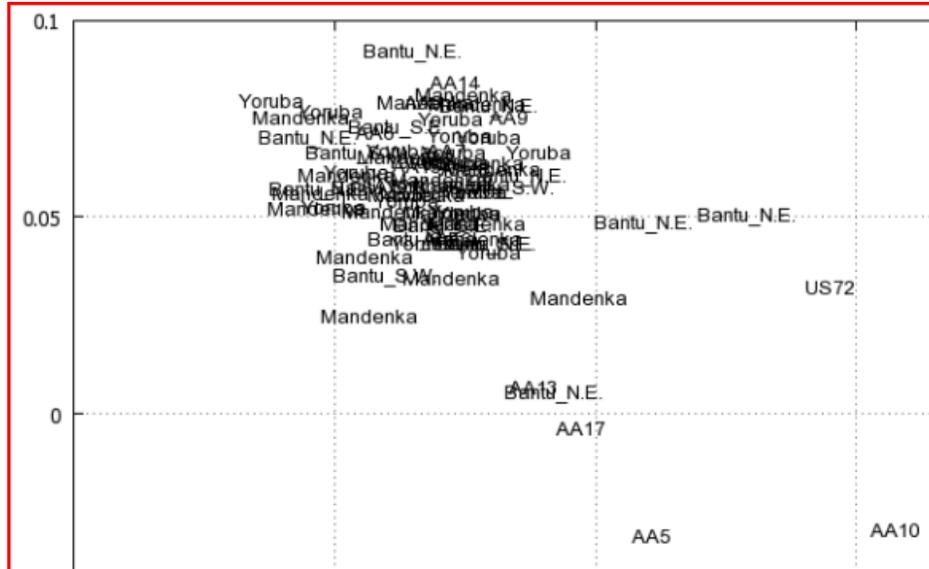
Using this data, supported by that provided by Dr. McDonald, DavidW created the following plot of the data:

US72 - MDS Plot (DavidW – Eurogenes):

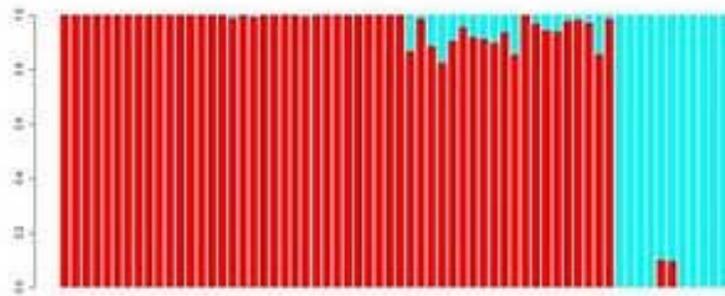


A full view in the MDS plot showing all the reference groups with US72 near the middle of the plot, and AA5 and AA17 to the lower left in the above diagram.

A close - up view follows below:



US72 is very close to a Bantu_North East who is undoubtedly one of the two in the reference group who have been determined to have about 10% European ancestry. Other Bantu_North East individuals are well to the back of the cluster of Africans since they have two African segments on this part of chromosome 8.



The red blocks (European) to the right among the entirely blue color bars (African) show that two of the HGDP-CEPF Bantu from Kenya samples have considerable Caucasian admixture ([Dienekes Blog](#)). It seems likely that one of them appears near the icon of MF.

The prediction (which cannot yet be tested) is that if the European segment was removed from the group of participants (via a process called phasing), they would be “pulled” back into the main area where the Africans cluster. If this is correct, then those African – Americans within the general cluster region should have more or less the same motif as their “cluster associates” as seen for example on the Ancestry Painting feature of 23andMe.

Ancestry Painting for chromosomes 8 (US72) in a sample of the African – Americans seen in the above MDS plots:

Two of those who cluster rather closely to US72 are AA5 and AA17. An exploration of the comparable segments (first 10 Mb) for 5 African – Americans on chromosome 8 follows. Here green = African, brown = Asian (but in An African – American is generally an African variant, blue = European.

Comparision of AA1, AA3, AA5, AA10, AA14, AA17 and US72:

AA1 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

\“8\“: [[154984, 10212547, \“CC\“], [10212902, 13918524, \“AA\“]

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA3 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

\“8\“: [[154984, 4829418, \“YY\“], [4838321, 5937551, \“CC\“], [5938053, 10161507, \“YY\“], [10162945, 10453427, \“AY\“],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA10 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

\“8\“: [[154984, 3089323, \“CY\“], [3089614, 3566820, \“CC\“], [3567490, 5213929, \“AC\“], [5217548, 6010430, \“CC\“], [6011583, 8985879, \“CY\“], [8987194, 10619292, \“AY\“],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA14 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

\“8\“: [[154984, 1612531, \“YY\“], [1617876, 2796904, \“AA\“], [2799115, 3750871, \“AC\“], [3751612, 4380342, \“AY\“], [4383093, 5159602, \“AC\“], [5159862, 5533383, \“CC\“], [5534186, 11537376, \“YY\“], [11538674, 12920929, \“AC\“],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA17 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

\“8\“: [[154984, 1942219, \“YY\“], [1946743, 5179603, \“CY\“], [5181456, 6118677, \“CC\“], [6136538, 10553595, \“YY\“],

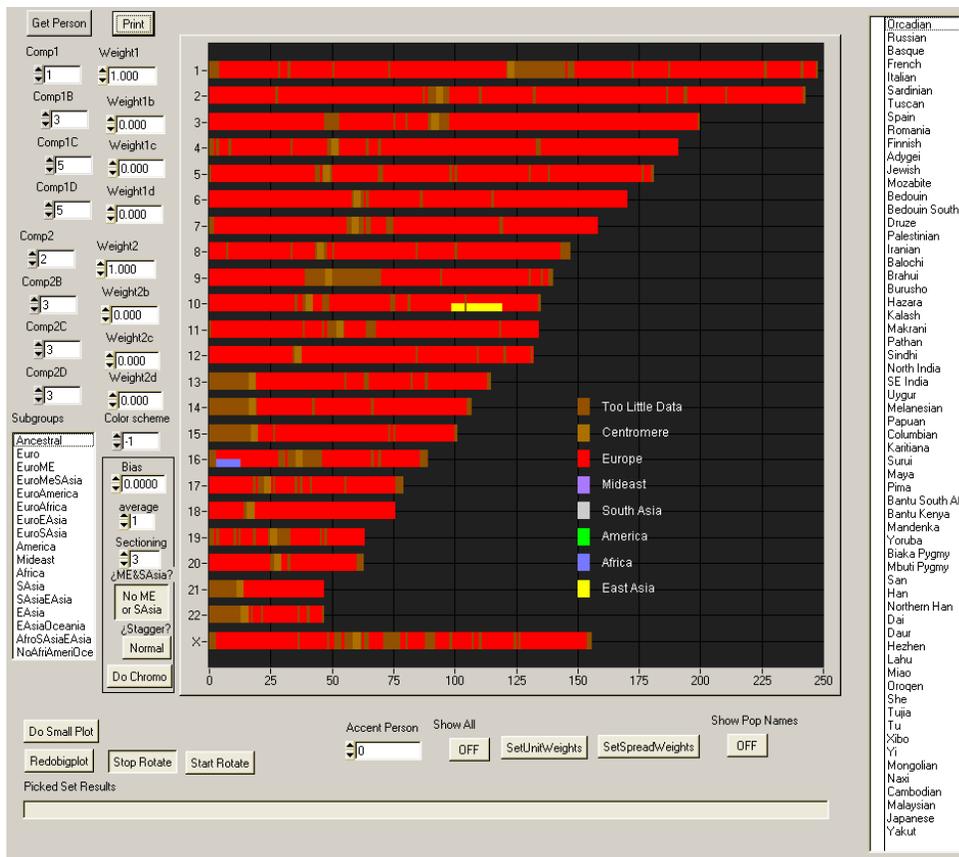
Key to the above: A=Asian, Y=African, C=European. CY=half European half African



Conclusion re US72: Clearly, according to these diagrams and the data used to compose them, AA1 does not cluster at all with US72 since the relevant segment here is fully European, and the icon is presumably to be found within the knot of Europeans. Unfortunately individual entries are impossible to read in places. AA5 and AA17 each have an approximation (although not as clear cut as the situation with CA4 seen later). They have close to a half European and half African segment in the same general area and hence display similar clustering on the MDS plot. The relevant segment of AA14 is very “complicated” with an interspersing of African, Asian and European all in this one small region although the African appears to predominate. This latter depiction appears correct in that reference to the MDS plot above shows AA14 at the top of the African cluster. Hence it is not surprising that AA14 does not cluster with US72 who has “uncomplicated” African and European strands. AA3 is largely African on both strands so is lodged in among the tight cluster of Africans (presumably since the icon cannot be seen). AA10 has a complex motif with a bit more European in total at this segment, but close enough to half European and half African that they are clustered closer to US72 than many others.

Chromosome 16 African Segment in CA4:

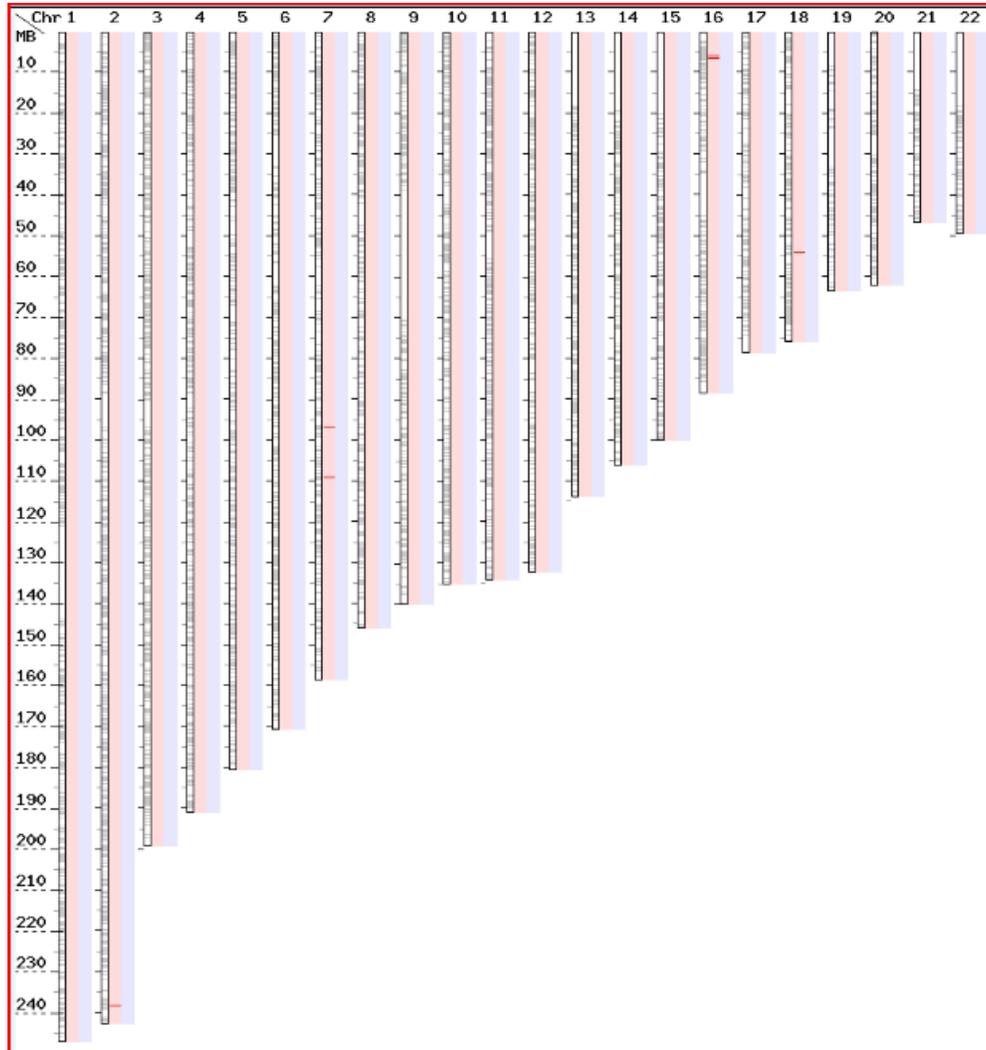
CA4 – Segment Analysis (McDonald):



The above figure shows an African segment for CA4 which extends from the tip (p end) of chromosome 16 (at the location where there is sufficient data, sampled SNPs, to about 10 Mb). It is important, as with US72, to have an independent source for confirmation of this data since it is

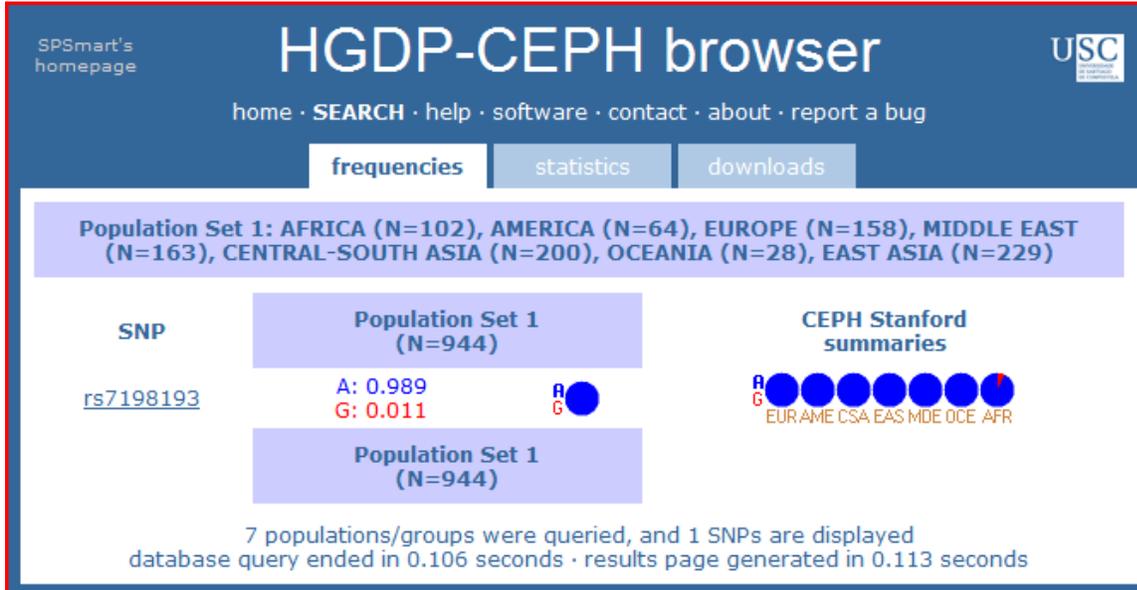
not consistent with the findings of 23andMe's Ancestry Painting. Again, we will turn to the RHH mosaic data and PC-MDS plotting to verify the validity of the above findings.

CA4 – RHH Mosaic:



As noted for US72 earlier, the above mosaic is the output of the RHH technique for CA4 showing regions of statistically rare SNP markers in anyone who is Western European. At present the method only appears to work well for African markers, at least at the resolution levels noted in the literature and chosen by DavidW – in this case the very “tight” 0.005.

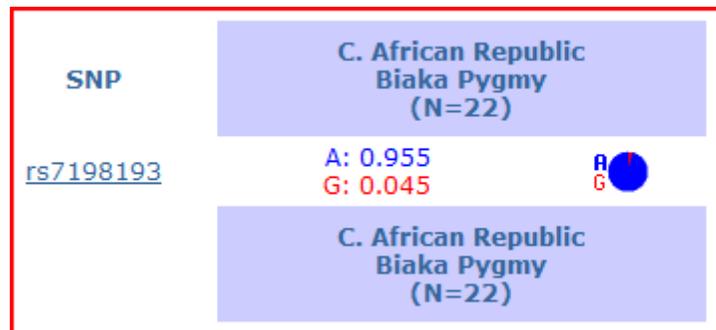
Ancestral Informative Markers from dbSNP:

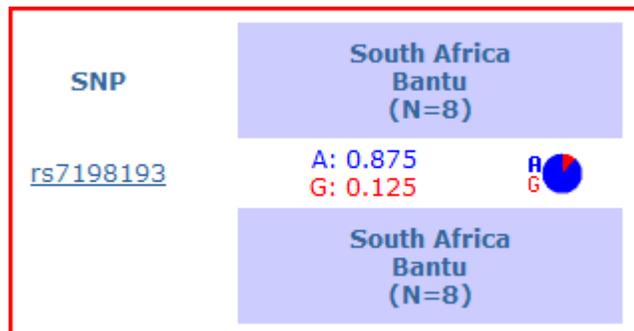
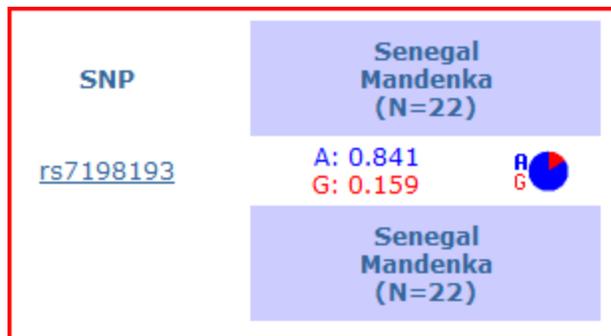
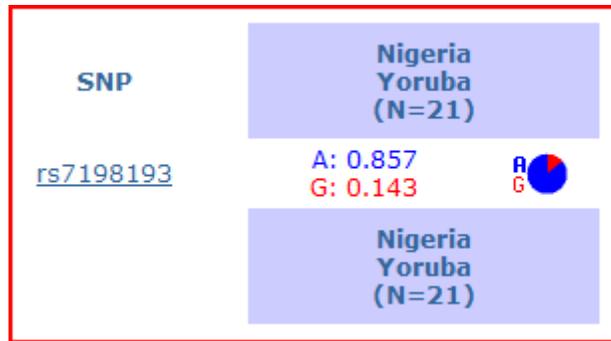


As seen above, for the allele G for SNP rs7198193, only Africans have been shown to have this version. The following African groups were tested for this SNP:

- AFRICA**
 - C. African Republic - Biaka Pygmy
 - D. R. of Congo - Mbuti Pygmy
 - Kenya - Bantu
 - Namibia - San
 - Nigeria - Yoruba
 - Senegal - Mandenka
 - South Africa - Bantu

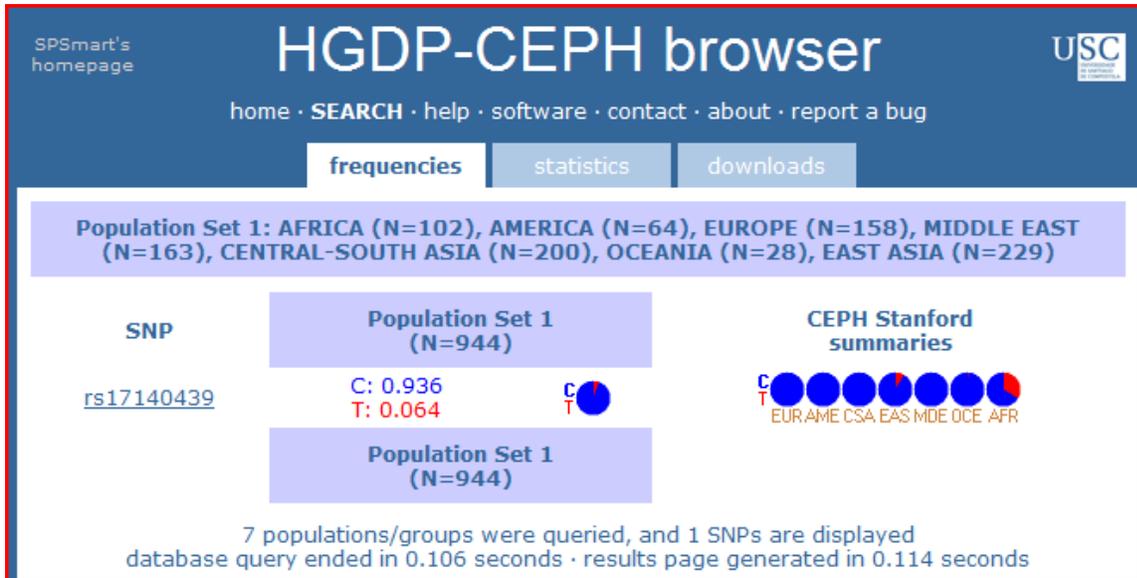
The following groups showed above zero percentages:



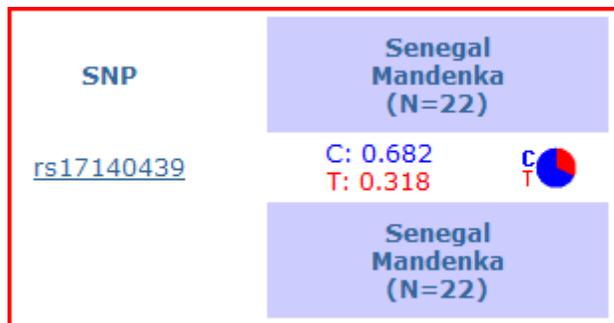


It is clear that the G allele is found only among the Bantu peoples of Sub – Saharan African (note, many more groups have been tested since this finding), in the western part of the Continent (e.g., not among the Kenyans of East Africa). This finding is somewhat unusual in that generally a marker will be seen at low levels in more than one location (Continent) such as African and a small percentage in Asia (but for example not in Europe). Here the sole location across the world sampled to date where the G allele is seen in Africa. This in effect makes [rs7198193](#) an AIM or ancestral informative marker akin to the Duffy-Null allele found almost exclusively in Africa and areas immediately adjacent to the tropical parts of the Continent. However the G allele here seems to be even more ethno-geographically specific, being confined to the Bantu peoples of Sub-Saharan Africa. This makes it a somewhat unique marker. However, this description also characterizes rs1721896 and rs11971652, separated by about 12 Mb on chromosome 7. This is a segment that 23andMe in their most recent (2020) iteration of Ancestry Composition, designate as “North African” – long after the initial version of this manuscript was composed.

The second SNP is also almost exclusively African (a much more typical scenario), with the exception of a small percentage of, in this case, East Asians, but the African percentages are much higher for this allele not found in other populations as seen below:



Most African groups have about the same percentage, so the Mankenka example will be shown here:



The following information may hint at the origin of the haploblock seen in David Faux and his uncle Dale Williamson:

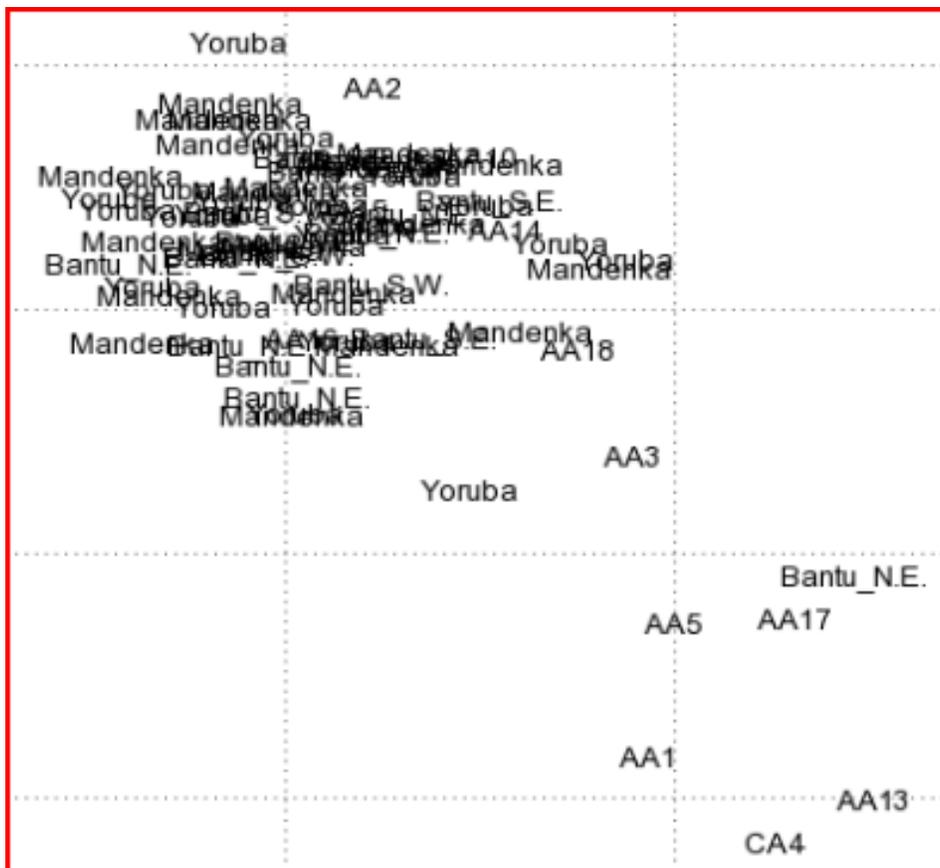
Mandenka of West Africa, Senegal where this sample was obtained, and 11 surrounding countries. *A significant part of the African-Americans in North America descended from Mandinka people During the 16th, 17th and 18th century as many as third were sold into slavery to Americas, through capture in conflict* (Wikipedia).

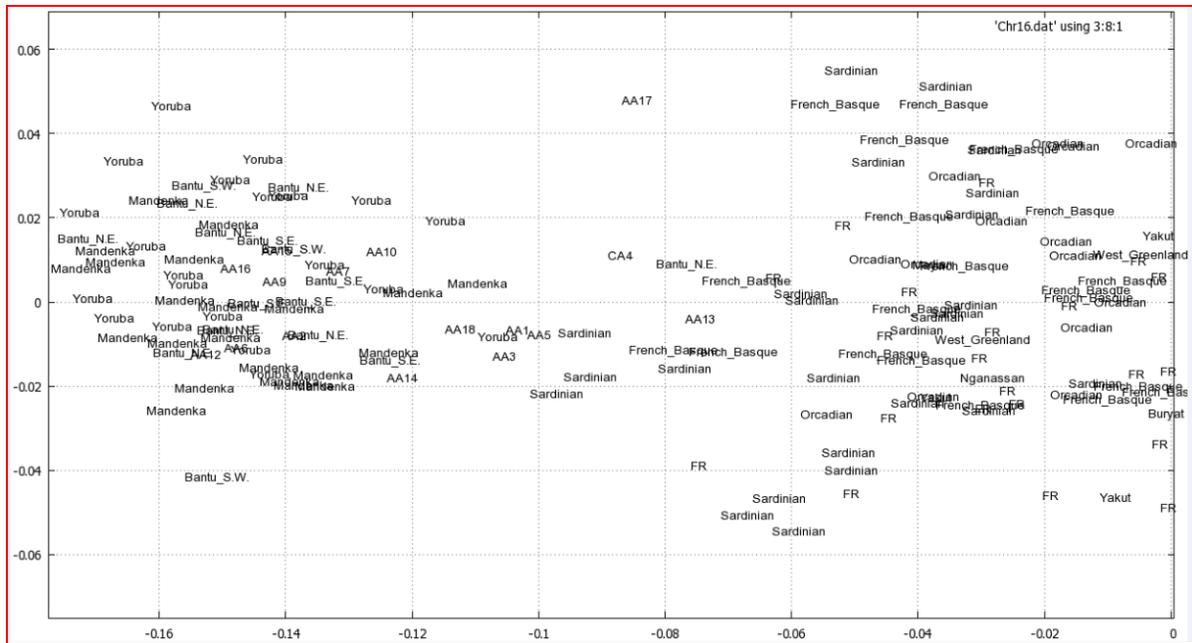
Subsequent to the above RHH analysis, DavidW provided a broader mosaic with a slightly reduced criteria as to how “rare” is defined (0.005% in the European sample). Here an array of other informative SNPs emerged, all with a very high probability of being African. Although some regions “seemed” to suggest African, once again the above area at the telomere of chromosome 16

was the best candidate. DavidW included the SNPs between position 5,575,804 and 7,675,034 Mb. This constitutes a block of 2,099,239 Mb in length. This amounts to only 0.0007% of the author’s genome. If Eva (Claasz) Pickert was his ancestor, then the amount of African contribution genealogical would be 1/1024 (taking into account the fact that she was half African and that she would be represented in two ancestral lines. This amounts to 0.00098% which is extraordinarily similar to the genetic finding.

Based on the identification of the small cluster of “HETs” on chromosome 16, DavidW “visualized” the area with a MDS plot as follows:

CA4 – MDS Plot (DavidW – Eurogenes):





The close up view above is quite similar to what is seen with US72 and once again a Bantu_NE reference individual is positioned nearby suggesting that they are one of the two admixed individuals from this group is seen above – with the admixture showing specifically on this segment of chromosome 16.

As can be seen above, CA4 and AA17 are equidistant from the Africans and the Europeans here – likely reflecting the straightforward 50% European and 50% African aspect of the chromosomes of both individuals at this location shown below. This can be verified by exploring the Ancestry Painting data for each of the individuals who were participants in the US72 analysis. AA3 also seems to be fairly close but the dimension that is chosen can slightly distort relationships (the configuration would vary to some degree as a function of the viewing point - dimension selected).

Comparison of AA1, AA3, AA5, AA10, AA14, AA17 and CA4:

AA1 - Biogeographical Genotype of Chromosome 16 from 23andMe's Ancestry Painting Using "Firefly" and Associated Chromosome Painting

"16": [[37354, 6300563, "CC"], [6300618, 13846342, "CY"]],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA3 - Biogeographical Genotype of Chromosome 16 from 23andMe's Ancestry Painting Using "Firefly" and Associated Chromosome Painting

"16": [[37354, 7078917, "CY"], [7079174, 26987807, "YY"]],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA5 - Biogeographical Genotype of Chromosome 16 from 23andMe's Ancestry Painting

Using “Firefly” and Associated Chromosome Painting

"16": [[37354, 2424784, "AA"], [2452524, 4583453, "AC"], [4600248, 7554014, "CY"], [7555007, 8799515, "CC"], [8799972, 11863979, "AA"], [11869359, 19624006, "YY"],
Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA5 shows a half European and a half African segment here. This is in the middle of the area shown for CA4 below.

AA10 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

"16": [[37354, 9629138, "YY"]

14

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA14 - Biogeographical Genotype of Chromosome 8 from 23andMe’s Ancestry Painting Using “FireBug” and Associated Chromosome Painting

"16": [[37354, 49584143, "YY"],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



AA17 - Biogeographical Genotype of Chromosome 16 from 23andMe’s Ancestry Painting Using “Firefly” and Associated Chromosome Painting

"16": [[37354, 10005342, "CY"], [10007044, 11198628, "CC"], [11199167, 24080450, "CY"],

Key to the above: A=Asian, Y=African, C=European. CY=half European half African



Conclusion re CA4: AA10 and AA14 have two African segments (strands) in the area we are focusing on, and hence are nowhere to be found in the clustering near CA4. Instead AA14 can be seen in the midst of those with two African strands in the main African cluster, and AA10 is easily seen in the second diagram nestled between a number of Yoruba Africans. AA1 has an all European as well as a half European – half African segment overlapping in the area where CA4 has the (half) African segment. However the half European and half African segments encompass most of the “area of interest”. Hence the close clustering between the two that is observed.

AA1, AA3, AA5 and AA17 do have very clear half European and half African segments in the area of interest in relation to CA4. Therefore, it is no surprise that they all tend to cluster together, stretching out between the African and European clusters seen in the MDS plots, when conducting an analysis of this particular segment. A problem is that it is not entirely clear where the African segment of CA4 begins and ends – so close estimates are all we have to work with. Of the participants to date, due to the shifting positions that can emerge with the selection of slightly different dimensions, it is difficult to say who is “closest” to CA4. At a glance, however, it would seem that the Bantu_NE stands shoulder to shoulder with the icon of CA4. Considering the history of Kenya, a European segment is likely to have come from an English ancestor – but this is conjecture at this point. the configuration of AA17 may be most similar to that of CA4 in this

specific region. AA17 has a clearly defined half European and half African segment across the entire span of that found in CA4 and so there is no “fuzziness” due complexity such as the prime segment starting and stopping in slightly different areas for CA4 and the others.

Thus, precisely as the theory (from the McGinnis study and the work of DavidW) predicts, both are equidistant from the African and European groups at least along the Y axis, but at different locations on the vertical axis (representing some differences in either or both of their European and African heritages here).

General Conclusion: The evidence, taken as a whole, makes it abundantly clear that both CA4 and US72 are non – African except for the distinct segments noted above. However in the search for ancestry, this data is consistent with each having an African ancestor who was likely born during the time of the transportation of millions of Africans to the New World. Here the time frame would likely be from the late 1500s to the early 1800s. CA4 has evidence as to the identity of the African ancestor, Eve Claessen Pickard born about 1680 almost certainly in New York City, Albany or Schenectady, and died 1767 Canajoharie, New York. Since Eve was born in New Amsterdam, historical records point strongly to the origin of her ancestors in the regions where the Portuguese (and later the Dutch) obtained “human cargo” for transport to their colonies. Here the likelihood shifts to Angola via the Port of Luanda, and the Mbundu people.

Based on genealogy, it is likely that the African ancestor of US72 resided from Virginia or points southward. Work will continue to see if it is possible to identify this individual and their specific homeland.

There have been a number of recent genome wide studies illustrating the percentage of European ancestry in African Americans – amounting to about 20% to 30% in some studies (with a range from zero in the Gullah of South Carolina to above 50%). In their 2003 study, Schriver et al. found that African Americans in their study had 18.6% European ancestry, and a sample of Pennsylvania White College students on average had less than 5% non-European admixture (including Native American). Schriver has stated that about 10% of African Americans have over 50% European ancestry.

It is unknown at present what percentage of “Old Stock” White North Americans have African ancestors that could be “seen” via genetic genealogy testing. None of the tests Schriver administered used the recently available chip technology, or the powerful techniques now available to expand on the general findings. It is very likely that the figure will vary as a function of geography (e.g., more in Americans from the Southern States than Canadians or those from say Upstate New York will have African ancestry), as well as the general percentage of Colonial American ancestry in White Americans. In more recent work, Schriver found that self-identified White Americans had an average of 0.7% African in their genomes (1/128 or a great great great grandparent). The upshot is that 70% of White Americans may have no African ancestors, but of those who do, Black admixture is in the order of 2.3% (from a trace amount to about 20%).

More recent work, conducted in part by scientists from 23andMe, has provided more data upon which to understand early admixture in the United States. In an article by Bryc et al., entitled, “*The*

Genetic Ancestry of African Americans, Latinos, and European Americans across the United States in the American Journal of Human Genetics, 2015, which can be seen by clicking [here](#). The article reports that, “Consistent with previous anecdotal results,³² the frequency of European American individuals who carry African ancestry varies strongly by state and region of the US. We estimate that a substantial fraction, at least 1.4%, of self-reported European Americans in the US carry at least 2% African ancestry. Using a less conservative threshold, approximately 3.5% of European Americans have 1% or more African ancestry (Figure S8). Individuals with African ancestry are found at much higher frequencies in states in the South than in other parts of the US: about 5% of self-reported European Americans living in South Carolina and Louisiana have at least 2% African ancestry. Lowering the threshold to at least 1% African ancestry (potentially arising from one African genealogical ancestor within the last 11 generations), European Americans with African ancestry comprise as much as 12% of European Americans from Louisiana and South Carolina and about 1 in 10 individuals in other parts of the South.” (p.44) The maps below offer a visual representation of the results:

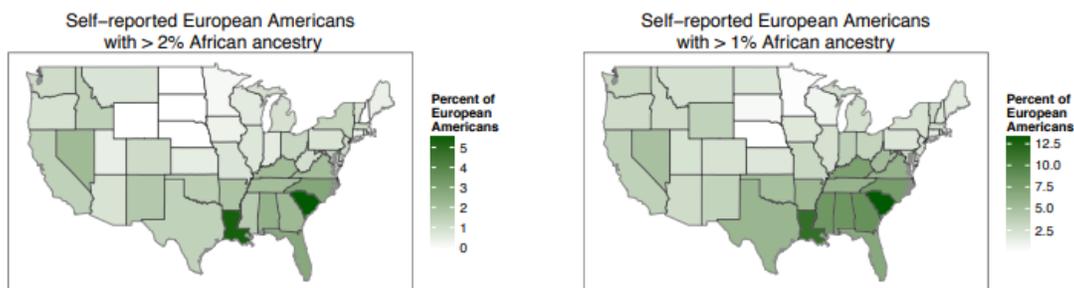


Figure S8: Frequency of self-reported European Americans with at least 2% African ancestry (left) and 1% African ancestry (right). The geographic distribution of self-reported European Americans with African ancestry. States with fewer than 20 individuals are excluded and shaded in gray. The proportion of individuals with African ancestry, out of the total number of European Americans per state, is shown by shade of green.

An important observation from this study is that, “Contrary to expectations under a social one-drop rule, or “Rule of Hypodescent,” which would mandate that individuals who knowingly carry African ancestry identify as African American, the probability of self-reporting as African American given a proportion of African ancestry follows a logistic probability curve, suggesting that individuals identify roughly with the majority of their genetic ancestry.” (p. 50)

Furthermore, the finding of significant African segments in two of the HapMap Utah sample, used as the reference group of Northwest Europeans (see McGinnis et al., 2010), should alert us to the possibility of “unseen” and “unknown” African ancestors in many White North Americans. Hence if one was to look for any evidence of African ancestry, however Small (below 1% African), the percentage would doubtless increase, perhaps dramatically. For some who are still influenced by the “one drop rule” this would be an interesting statistic.

With thanks to AA1, AA3, AA5, AA10, AA14 and AA17 of the Eurogenes Project for their kind assistance with this work.

Dr. David K. Faux
12 January 2011; 18 May 2023