The Värmland-Hedmark Cluster

No. 1 (July 2019)

A newsletter with updates about the genetic genealogy project "The Värmland-Hedmark Cluster," which studies the Y chromosome haplogroup R1b → M269 → U106 → Z18 → S11601 → Y112538 → Y130179 and surrounding clades.

Contact information: The newsletter is written by Johan N. M. Lagerlöf (email: johan.lagerlof@econ.ku.dk; website: www.johanlagerlof.com).

Welcome

Welcome to the first issue of this newsletter. This is an experiment. I hope that a newsletter will be a useful way of communicating news about the project to my fellow testers and others who might be interested in what we do. By using the form of a newsletter instead of putting all the information in a group email (which is what I have done before), I think it will be easier for myself and others to find the information again and to refer to it whenever there is a need. It should also be useful for us to have old issues of the newsletter available when we want to recruit new testers and try to explain to them what we are doing, what we have learned so far, and what we hope to learn in the future.

My intention is to, in the newsletter, refer to testers only by anonymous labels, such as Tester 1, Tester 2, and so on. The only exception will be if a tester has explicitly told me that he does not mind being named and naming him makes for a more readable or informative text. Occasionally I will need to state names for the purpose of, for example, explaining to other testers that this new tester who just joined the project is the person NN, who several of us have listed as a match on our Big Y accounts at Family Tree DNA; however, instead of using the newsletter for that, I will provide the relevant information in an email to a select group of fellow testers. By keeping the newsletter immune against privacy concerns, we will be able to post it online and circulate it widely, whenever we find this useful.

A Recap: What Are We Trying to Do?

Figure 1 depicts the latest version of our part of the haplotree. In that tree, I am shown as Tester 1. The figure also indicates that my (i.e., Tester 1's) oldest known patrilineal ancestor was a man called Per Månsson, who lived in Skived in the late 1500s and early 1600s. Similarly, Tester 3 and 4's (common) oldest known patrilineal ancestor was called Christopher Andersson, and he lived in Väse in the 1600s. Both Skived and Väse are located in southern Värmland in Sweden, about 40 kilometers apart. The fact that Månsson and Andersson are the patrilineal ancestors of Tester 1 and Testers 3/4, respectively, has been established with the help of traditional genealogical methods (i.e., by using church records, censuses, etc.). In addition, Y-DNA tests reveal that Tester 1 and Testers 3/4 must have a (reasonably recent) common patrilineal ancestor—labeled A2 in Figure 1. Likewise, Testers 1 and 2 have a common patrilineal ancestor A1 (who is younger than A2); Testers 7-9 have a common patrilineal ancestor A4; and all nine testers have a common patrilineal ancestor A3 (who is older than all of A1, A2, and A4).

Our goal in the project (at least mine) is to, with the help of Y-DNA tests, learn about where A1, A2, A3, A4, and other ancestors lived and about migration paths.[1] Admittedly, we are very unlikely to ever learn about these individuals' names or to acquire other more specific biographical information about them, but it is reasonable to hope that we will be able to make some inferences about roughly when and where they lived (and hence also about migration paths). Obviously, being able to make reliable inferences about the whereabouts of ancestors living a very long time ago will be much harder than doing this for ancestors who lived relatively recently. One insight that we arguably already have gained is that A1 and A2 most likely lived in, or in the proximity of, southern Värmland, as all five of the known branches originating from A2 ended up in that geographical area.[2] Given the data that we currently have available, the question where A3 and A4 lived is much more open. For example, A3 has known descendants not only in Värmland and in nearby Hedmark, but also in Levanger in Norway (a municipality in Trøndelag county, located about 650 km from Väse), in England, and in Ireland. So, did A3 live in Sweden, in Norway, or on the British Isles? Or did he live somewhere else, perhaps on the European continent? We may also wonder, given that at least one of A3's patrilineal descendants must have migrated, roughly when did the migration occur? And in what direction—to or from Scandinavia? Although we cannot answer these questions today, we should have a good chance of finding reasonably convincing answers to some of them in the future, when we have access to richer data. A natural way forward is thus to try to recruit more testers. Currently, in most populations only a very tiny fraction of men have done ambitious (or, for that matter, *any*) Y-DNA tests.

When we seek answers to the above questions, the Y-DNA tests help us not only by telling us about the structure of the tree, but also by enabling us to compute estimates of the birth dates of A1-A4. The idea behind the computation of those estimates is to count the number of mutations (so-called SNPs)[3] that a group of testers with a common patrilineal ancestor do *not* share. Each one of those SNPs must have occurred in one of the generations *after* the common ancestor, because if it were present in him, all his male descendants would have inherited it. We also have access to (at least approximate) knowledge about the rate at which the SNPs occur.[4] This means that we can, so to speak, reason backwards in order to work out the likely birth date t of the common patrilineal ancestor. That is, we can ask: Given our knowledge about the rate at which the SNPs occur, which value of t makes it most likely to observe the number of non-shared (or "novel") SNPs that we actually observed in the test results? We can then take that number as our estimate of the common patrilineal ancestor's birth date.

It goes without saying that the age estimates that we can compute will indeed only be estimates. Their precision will be hampered by uncertainty from at least three different sources:

(i) For a given mutation rate, the realized number of SNPs is stochastic (just like the number of sixes we get when we roll a die one hundred times is random).

(ii) The mutation rate is not perfectly known and could itself behave stochastically (the die we roll might have been tampered with and we do not know exactly how, meaning that the probability of getting a six is not necessarily $\frac{1}{6}$ but could be something else).

(iii) The length of a typical generation in historical times is typically not known exactly and it may

[1] Another goal is to make use of the tests, whenever it is possible, to make further progress with our traditional genealogical research. For example, sometimes church records are available but it is hard for the researcher to confidently link an individual with his or her father, due to ambiguities in the records. In such situations it may be possible to confirm or reject a hypothesis about a particular link by DNA testing the appropriate (now-living) individuals.

[2] Väse, Skived, Nedre Ulleud, and Frykerud are located within a distance of 60 km from each other. Sør-Odal is also not very far away—it lies near Kongsvinger in Hedmark in Norway, immediately west of Värmland and about 200 km from Väse.

[3] SNP (often pronounced "snip") is an acronym for *single nucleotide polymorphism*. This refers to a change of base (A, C, G, or T) in a particular location on a chromosome. See e.g. Jobling et al. (2014, pp. 47-48 and 637)

[4] One way of obtaining this knowledge is to make use of ancient samples of human remains. The age of the remains can, with a high degree of precision, be estimated by radiocarbon dating. Moreover, the Y chromosome of the sample can be sequenced, and by then comparing the SNPs of the ancient individual with those of now-living testers, one can work out roughly how many new SNPs that have occurred since the ancient individual lived. This number can be compared to the age obtained by radiocarbon dating, which yields an estimate of the mutation rate. One example of a sample that has been used for this purpose is "Anzick-1," who was a boy who lived about 12.6 thousand years ago in (what is now) Montana, USA. See the discussions in Rasmussen et al. (2014), Adamov et al. (2015), and Jobling et al. (2014, pp. 53-54 and Ch. 6.6).

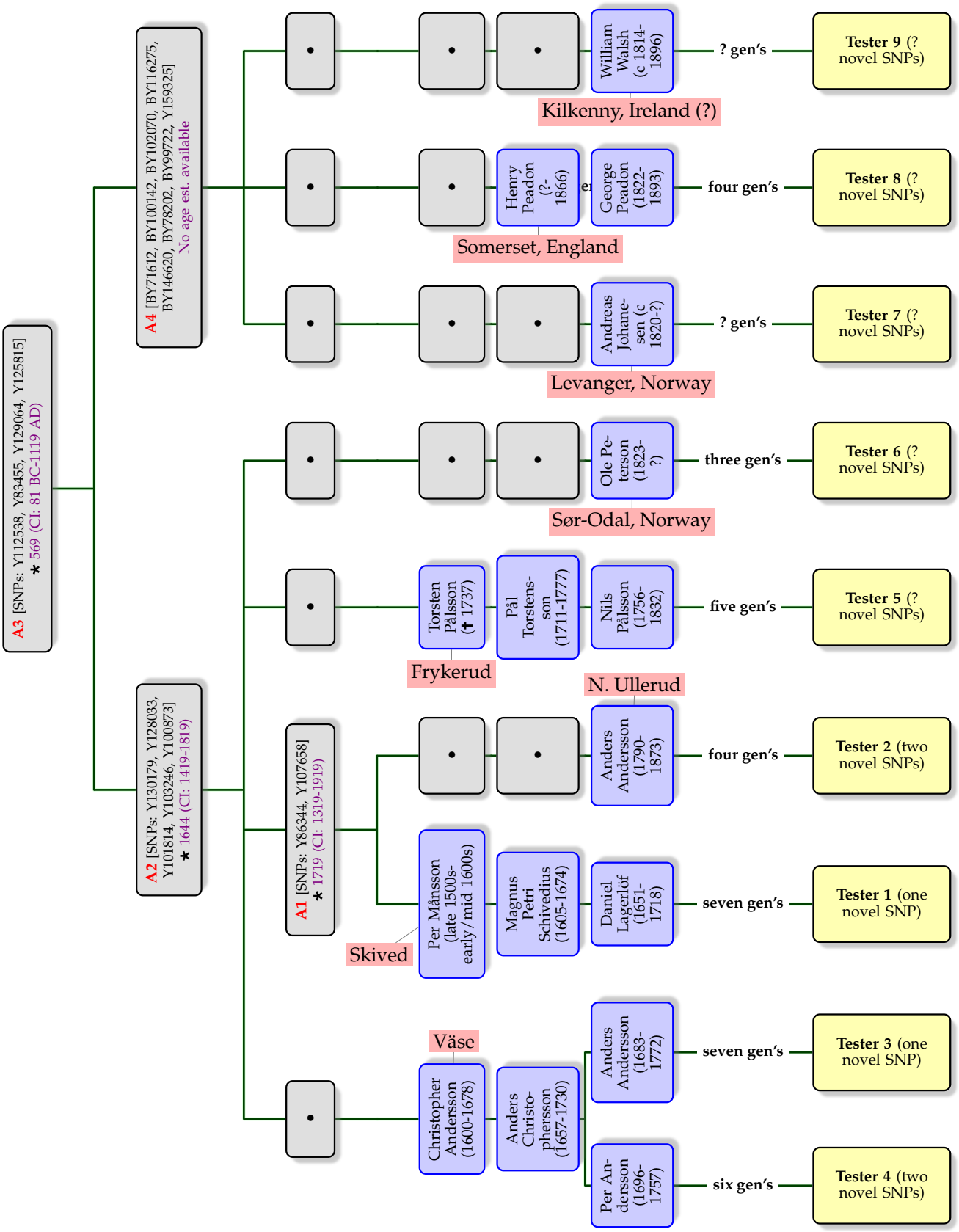

Figure 1: The Värmland-Hedmark cluster (R1b → M269 → U106 → Z18 → S11601 → Y112538 → Y100873) and a neighboring branch (BY71612), as of July 2019.

have varied over time and space.

However, by using standard statistical techniques we can try to measure the magnitude of the uncertainty. Using this information, we can do our best to formally and rigorously test carefully stated hypotheses. Moreover, we should expect some of the uncertainty to be reduced as we increase the number of observations. This is another good reason to recruit new testers.

The age estimates shown in Figure 1 are from YFull (YTree v7.06.00). This is a company that does not itself sell any tests. Instead it offers an analysis of advanced Y chromosome tests (nowadays also mitochondrial tests) that its customers have purchased at other companies. In particular, YFull builds a very useful and easily-navigated haplotree and it provides age estimates of the branches in the tree (the company's website: `www.yfull.com`). Inspecting Figure 1, however, it is evident that some of the point estimates from YFull must be wrong. In particular, the most recent possible birth year of A1 that is consistent with our genealogical data would be in the early 1600s (A1 could in principle be a brother of Schivedius, shown in the chart, but not younger). Still, the point estimate says 1719. It is not very surprising that some of the point estimates turn out to be unreliable, as they are computed using a very small sample and they do not make use of any genealogical information that the testers have access to.[5] I therefore urge everyone to, instead of focusing only on the point estimates, consider the whole confidence intervals ("CI" in the figure). The ones that are shown, also from YFull, are at the 95% level. The confidence intervals shown for A1 and A2 are fully consistent with these ancestors having lived some time either in the 1600s or as early as a few centuries before that. The point estimate for A3 is the year 569, but the confidence interval suggests that he may have been born any time in the first millennium A.D. Notice that if we want to understand how early in time the lineage leading from A3 to A2 and A1 was present in Värmland/Hedmark, then precise estimates of A2's and A3's ages are very helpful. It would be even more useful if we could find a new node (i.e., a new patrilineal ancestor) in the tree, somewhere between A2 and A3, such that all branches originating from that node still ended up in Värmland or in Hedmark. There is no age estimate available for A4, as Testers 8 and 9 have not submitted their results to YFull. (They are very much encouraged to do so!)

Recent and Future Developments in the Tree

New Testers in Our Part of the Tree

Since last time I showed you a version of the tree, the following two testers have been added:

- **Tester 8**. Oldest known patrilineal ancestor: Harry Peadon (?-1866), who lived in Somerset, England.
- **Tester 9**. Oldest known patrilineal ancestor: William Walsh (c. 1814-1896), who I *believe* lived in Kilkenny, Ireland.[6]

As one can see from the tree in Figure 1, these two testers are located on the same branch as Tester 7, whose oldest known patrilineal ancestor is from Levanger in Norway. I have tried to contact both Tester 8 and Tester 9, but so far failed to get a response (perhaps I do not have the right contact information or there is some other issue). I will try again later. Neither one of Tester 8 and Tester 9 is so far on YFull. Instead I found out about them through FTDNA's Big Y Block Tree.

[5] My ambition is to later (at some stage, when time allows…) compute my own age estimates, as a complement to those offered by YFull (using a maximum likelihood estimator, given the tree structure). I then also hope to be able to impose constraints that are implied by our genealogical data on the estimates. However, if and when I can do that, I will have to rely on information from YFull about which testers are positive for which SNPs. I will also use YFull's mutation rate for SNPs. Thus, I cannot do without YFull's analysis.

[6] The tester has not himself provided the information about Kilkenny and/or Ireland. Instead I found it in a public tree at Ancestry. It seems to be the right person that I have found there, but this is not confirmed.

Testers That Are in the Pipeline

The following testers are in the pipeline to be added to the tree, as they have recently ordered a test and are waiting for their results:

- **Tester P1**. Oldest known patrilineal ancestor: Botolf Magnussen (born in 1885), who was from Grue/Kirkenær in Hedmark, Norway. This is a 30X Whole Genome Sequencing (WGS) test from Dante Labs. The test sample was sent by me (from Sweden) to Dante Labs on March 3. This tester has previously done a Y37 STR test at FTDNA. Once we get the full results, I strongly expect him to end up somewhere in the Värmland-Hedmark cluster (in terms of Figure 1, I define this cluster as the branches originating from A2).

- **Tester P2**. Oldest known patrilineal ancestor: Axel Unger (1865-1959), who was born in Bro in Bohuslän, Sweden. Axel's father is unknown, so at the moment we cannot get further back in time. However, we will make an attempt to identify the father with the help of autosomal DNA tests (in combination with traditional genealogy). It could of course prove very hard to make progress, but we will have a go. Also this is a 30X WGS test from Dante Labs. The test sample was sent to Dante Labs by me (from Sweden) on June 8 and confirmed as received by Dante Labs on June 20.[7] This tester has previously done a Y37 STR test at FTDNA. Those results suggest that Tester P2 might not necessarily end up in the Värmland-Hedmark cluster; however, it seems very likely that he will be positive for the SNP S11601 and reasonably likely that he will be positive for the SNP ZP144 (both those SNPs are somewhat upstream of the branches shown in Figure 1).

- **Tester P3**. Oldest known patrilineal ancestor: Jon Pädhersson (1679-1733), who was from Kroppa parish in the (south-) eastern part of Värmland. This is a Big Y-700 test from Family Tree DNA. In contrast to P1 and P2, I have not been involved in the testing of this person. Instead, the tester showed up as an STR match to me on my FTDNA account. I have been in contact with the person who manages the test (a relative of the test person). The test, which is an upgrade from Y111, was ordered on March 26. The tester is not yet member of any FTDNA project, so I have so far not been able to study the STR results. However, I know that I myself match this person on the 111 level (the genetic distance is 10). That circumstance and the ancestral link to Värmland suggest to me that, most likely, this tester will end up in the Värmland-Hedmark cluster.

- **Tester P4**. Oldest known patrilineal ancestor: Nils Nilsson (1742-?), who was from the farm Rosensjö in Karlskoga parish in the south-eastern corner of Värmland.[8] Also this is a Big Y-700 test from Family Tree DNA. Both I and a few other people in the tree have had P4 as an STR match for a very long time, and I am very excited that he has now ordered a Big Y. Earlier, when I had not made as much progress with P4's family tree, the information I had was that his oldest known patrilineal ancestor was not from Värmland but from the neighboring county of Närke (in particular, from Knista parish). For that reason, I actually used the label "the Värmland-Närke cluster" for the cluster that I now call Värmland-Hedmark. I also speculated—some of you may remember this—about a possible link between our patrilineal lines and Riseberga kloster (in English, Riseberga Abbey), which is located in Närke and quite close to Knista. That was speculative already when I first suggested it, and now we should probably completely give up this idea. Tester P4 has kindly agreed to become a member of FTDNA's Sweden DNA project, which means that I have access to his STR results. However, I have not yet had time to study these in any detail (beyond the analysis presented in Figure 2, discussed below). I will do that later at some point.

[7] The submission of the test sample to Dante Labs was very much delayed due to a mail problem that arose. The tester lives in Australia and I first sent him the test kit from where I live in Sweden; this was delivered reasonably quickly. However, when he sent the sample back to me, the delivery took about two months. Probably the letter was literally shipped, as opposed to being airmailed. (We had deliberately chosen to let the tester send the sample to me in Sweden, and then I would send it to Dante Labs' address in Italy. This was very fortunate, I believe, because otherwise we would not have known whether the lack of any confirmation about reception from Dante Labs was due to Dante or to the mail service.)

[8] I have done some traditional genealogical research on this family line myself. It seems likely that one can get further back in time (probably, also Nils Nilsson's father and grandfather lived at the farm Rosensjö). However, I have not seen this being confidently confirmed, so for the time being I write Nils Nilsson as the oldest known patrilineal ancestor.

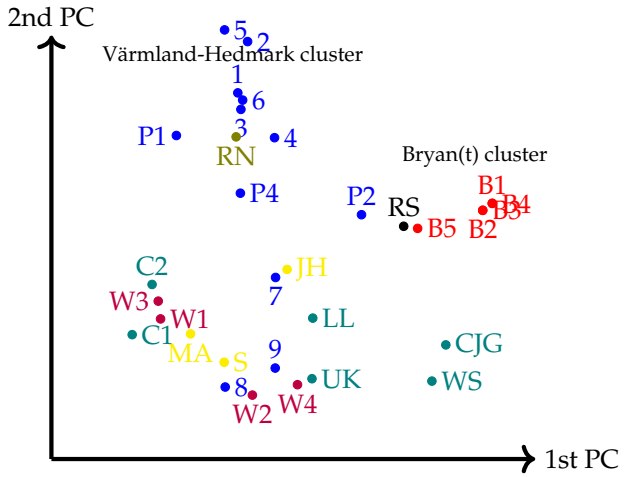

Figure 2: Principal component analysis of the "East-Anglia cluster" (R1b → M269 → U106 → Z18 → S11601). The plot is based on FTDNA's first 37 STR markers. The numbers 1-9 and P1, P2, and P4 correspond to the ones in Figure 1 and in the discussion of tests in the pipeline, respectively.

> Nevertheless, we know that P4 matches me (and I believe a couple of others) at the Y111 STR level, and there is an ancestral link to Värmland. I therefore strongly expect that this tester will end up in the Värmland-Hedmark cluster.

Figure 2 shows a "map" over the haplogroup R1b-S11601 that is created with a statistical technique called principal component analysis (PCA).[9] It is based on the STR values of 32 testers for FTDNA's first 37 markers. The idea behind this technique is to summarize as much as possible of the variation across testers in only two dimensions. This makes it possible to represent each tester as a point in a two-dimensional plot (so it is like a "map," although not in the geographical space). As you can see, the plots show very clearly that the testers that are close to each other in the tree form distinct clusters in the PCA plot. Thus, those of us who belong to the Värmland-Hedmark cluster are located close to each other also in Figure 2 (see the dots numbered 1-6—the same numbering as in Figure 1). Likewise, the Bryan(t)s form their own cluster (the dots labeled B) in Figure 2.[10] Testers P1, P2, and P4 are included in the plot and, as you see, P1 and P4 are located in or near the Värmland-Hedmark cluster, whereas P2 is a bit farther away, halfway toward the Bryan(t)s. (As already mentioned, I do not yet have access to P3's STR values.) The tester RN is an STR match of mine at the Y111 level, but who is not yet in the pipeline for a Big Y or other similar test.[11]

Upgraded or More Ambitious Tests

Some of the existing testers in the tree in Figure 1 have upgraded or ordered other, more ambitious tests. In particular, I have upgraded Tester 2's test (which I manage) from Big Y-500 to Big Y-700 (I placed the order on March 31). I know that Tester 3 and possibly also Tester 6 have done that as well, at about the same time.[12] As far as I know, none of us has yet received the results for this upgrade. The

[9] The technique is standard and is described in most textbooks on multivariate analysis. One example, which I have used myself, is Everitt and Dunn (2001, Ch. 3).

[10] The Bryan(t)s are members of a surname project (see this website: `www.familytreedna.com/groups/bryan/about`). Several of these testers can be found in YFull's tree, on a branch close to ours (see `www.yfull.com/tree/R-FGC54210`).

[11] There are about four or five further testers who are STR matches of mine at the Y37 level, or higher, and who I strongly suspect believe belong to the Värmland-Hedmark cluster (four of them have an ancestral Värmland connection and the fifth has an ancestral Sweden connection). These testers are not shown in Figure 2, as I do not have access to their STR values (they are not members of any FTDNA project).

[12] I am not sure about whether this is true for Tester 6, but in an email correspondence with some of us he wrote something that suggests that he might have done this.

important and valuable thing with the new test Big Y-700 is *not* the 200 additional STR markers but the fact that it has significantly better coverage than Big Y-500. Having "better coverage" means that the test's read of the Y chromosome is more thorough and therefore has a greater capacity to identify new SNPs. This, in turn, makes it more likely that the test will help us expand the tree with new branches. For example, in our current version of the tree, there are four branches that shoot out from the node A2 (see Figure 1), suggesting that they are equally old. Of course, this might be historically correct—perhaps the four branches represent four brothers. However, it seems more likely that one or two of these branches are in fact older than the others, but that our tests have failed to discern that difference. A test that is able to find more SNPs has a better chance to discern such a difference, if there is one, and to achieve a greater degree of tree granularity.

In order to form a new branch, one must of course find at least *two* testers with the same new SNP. This might create a difficulty for us if we find a new SNP in, say, Tester 3's Big Y-700 test and we do not know if his closest neighbor Tester 4 (who still "only" has a Big Y-500) is positive for the same SNP or not. In such a situation, we might be able to let Tester 4 test that individual SNP at the company YSEQ. This company has a very useful service called "Wish a SNP": The customer pays $1 to make a particular SNP available for testing, and then actually testing the SNP costs $18. However, one potential problem with such a strategy is that YSEQ's service cannot be used for all possible SNPs (it depends on exactly where on the Y chromosome the SNP is located). I think we should wait and see how to solve this issue when we get there. I received the question before whether I recommended others with a Big Y-500 to upgrade to 700. I was quite cautious in my response. I do think there is a clear potential value with upgrading (as I have tried to explain above). But I also think that (i) we do not yet know how big the value will be for us in our particular part of the tree; and (ii) I am cautious about asking other people to spend even more money on new tests, before we have been able to show them that the first tests (which were quite expensive) really helped us make progress and learn interesting things. The opportunity to upgrade the Big Y-500 to 700 will be there also in the future. Now a few of us have ordered the upgrade, so to me it makes sense to first look at what that will achieve for us.

I have for myself (i.e., Tester 1) not upgraded my Big Y-500 to Big Y-700; instead I have ordered a 30X WGS test from Dante Labs, just like the new testers P1 and P2 discussed above (I sent my sample to Dante Labs on January 20, and it was confirmed as received by the company on February 12). In terms of its coverage of the Y chromosome and its ability to find new SNPs, this test appears to be roughly at par with Big Y-700 or possibly somewhat better. There is an interesting and useful comparison between different kinds of tests carried out by James Kane and presented at the ydna-warehouse website (`https://ydna-warehouse.org/statistics.html`). The perhaps most useful column in the table on that page is the one showing the estimated number of years, on average, between each SNP. This number should of course be low: We want a small number of years to pass between the SNPs, so that we more easily can discern branches in the tree. The results in the table tell us that (as of July 11, 2019), for the Big Y-500 test, this number is 131.6 years; for the Big Y-700 test, it is 82.5 years; and for the Dante Labs 30X WGS test, it is 81.9 years. The test that performs best is "Chromium LR" (i.e., a so-called long read WGS test using Chromium technology), sold by the company Full Genomes Corporation (FGC). The number for this test is 71.4 years per SNP, on average. However, to get hold of that test you have to shell out the not very modest amount of $2900.

Conclusion

For quite a long time (at least since last summer), there was not very much development in our part of the tree. Now recently, however, we have been able to add two new testers (although they are not yet on YFull). Moreover, there are four new testers in the pipeline, three of which are very likely to belong to the Värmland-Hedmark cluster and the fourth one should at least end up on a neighboring branch. A few existing testers have also ordered either the new test Big Y-700 or a 30X WGS test from Dante Labs, which should be helpful in discerning hidden branches in the tree. All in all, I think we have good reasons to expect the next few months to be exciting and rich in insights and learning.

References

Adamov, Dmitry, Vladimir Guryanov, Sergey Karzhavin, Vladimir Tagankin, and Vadim Urasin. 2015. "Defining a New Rate Constant for Y-Chromosome SNPs Based on Full Sequencing Data." *The Russian Journal of Genetic Genealogy* 68–89.

Everitt, Brian, and Graham Dunn. 2001. *Applied Multivariate Data Analysis*.: John Wiley & Sons.

Jobling, MA, E Hollox, M Hurles, T Kivisild, and C Tyler-Smith. 2014. *Human Evolutionary Genetics. 2nd*.: Garland Science.

Rasmussen, Morten, Sarah L Anzick, Michael R Waters, Pontus Skoglund, Michael DeGiorgio, Thomas W Stafford Jr, Simon Rasmussen, Ida Moltke, Anders Albrechtsen, Shane M Doyle et al. 2014. "The Genome of a Late Pleistocene Human from a Clovis Burial Site in Western Montana." *Nature*, 506(7487): 225–229.