Who Am I?

I am Bryant McAllister, an evolutionary geneticist and Associate Professor of Biology at the University of Iowa. Although my research is primarily focused on the role of chromosomal organization in evolutionary processes, for a long time I have been interested in human genetic variation and the development of genetic tools and analytical methods that enable the reconstruction of human ancestry. In early 2014, I proposed to offer a first-year seminar, Who are you? Genetic insights on human ancestry, that focused on the personal genomics revolution and its applications to genetic genealogy. Through 2014 I actively explored the opportunities in personal genomics that enable individuals to investigate their relationships with others through DNA analysis. I taught the seminar course for the first time in Fall 2014 to a class of 17 first-year University of Iowa students representing a broad spectrum of majors. On this page I summarize the process and describe some of the things I’ve learned. A comparison of the different personal genome testing platforms that I have used in the discovery process is also presented in a table.

AncestryDNA

It all started with gifts of AncestryDNA kits to my parents. Since I inherited only half of each of their genomes, having their samples instead of mine empowers the relative matching with other AncestryDNA users just a bit more than doing a test myself. Furthermore, my mother is the genealogist in my immediate family. She has accumulated information and communicated with others about our ancestors and relatives, and she was already an Ancestry.com user. Using information mom had already collected, I joined Ancestry.com myself and started working to build the family tree in my account. The ability to share a single AncestryDNA test among users is a really nice feature, so my mother and I are each able to access the DNA results of each of my parents. The sharing feature in AncestryDNA was improved in August 2014. With the change, the individual that manages a DNA test can share the results with any other Ancestry.com user (however, there does appear to be a huge cultural difference in the enabling of the sharing feature between AncestryDNA and 23andMe users). Previously it was the case that the purchaser and the tester could both access the results, or the person on the help line could also set up sharing in cases where the kit was not registered properly (which happened to me when registering the first kit).

In my view AncestryDNA makes connecting with relatives plain and simple - probably too much so. I purchased a third kit when buying kits for my parents, because my college-aged son was interested in doing a personal genome test. Based on his low-level engagement with the AncestryDNA website, I didn’t think this was a good platform to use for a first-year college seminar. Other than connecting with relatives through DNA matches and supporting common ancestors within a family tree, there simply isn’t anything more someone can learn by working with the results. AncestryDNA is great for finding potential relatives when a person already has a good family tree, and even better if that person is a subscriber to Ancestry.com that uses the platform for genealogical research. I’ll describe below the connections I’ve made with my McAllister relatives through the AncestryDNA test of my father. Otherwise, the results form AncestryDNA provide rather limited access to the data underlying the matches.

Genographic Project

I also considered The Genographic Project for this course. It has the nice feature of being associated with National Geographic and a façade of scientific investigation of your ancestry. In reality, all of the services are based on the same technology and similar analyses. I was a bit concerned about the price; a test from the non-profit Genographic Project costs ($160 at the time; $200 now) double the price of a test from any of the for-profit companies ($79-$99). Each uses the same testing platform with some differences in the sites analyzed within the genome. What I got from the Geno 2.0 test is a slick summary report of my ancestry - it even comes with a nice poster summarizing the results. I could print out the poster and never need to return to the website, because the website does not provide much more information or interaction with the data beyond what is contained in the poster.

I learned that a bit more than a half-million people have also tested with The Genographic Project (as of Aug. 2014). The estimated percentage of Neanderthal ancestry in my genome is 1.4% (23andMe estimates 2.7%). The broad geographic sampling of native peoples over the globe is promoted as the key aspect of the science underlying The Genographic Project. However, as a descendant of Northern European immigrants to the USA, and belonging to a common European mitochondrial haplogroup, H27a, and belonging to a common Y-chromosome haplogroup, R-L21, this rich geographic sampling is unlikely to enhance the view of my genetic ancestry. Essentially, the test results from The Genographic Project consist of a report on your ancestry with lots of good information about the migrations throughout human history that have given rise to the patterns of genetic variation among modern humans. This is good stuff for a couple of class meetings, but not something I want to use over an entire semester of a class where the goal is for students to engage and interact with their genome data.
The Genographic Project does allow users to download their genome data; however, the set of markers used and formatting of the data is sufficiently unique that there is little opportunity to input the data into other analysis platforms. This is exemplified by the fact that only a small portion of the sites tested on the Y chromosome (and much smaller portion of overall sites tested) are reported in Family Tree DNA after the data are selected to be transferred to this partner site.

23andMe

I ended up also testing with 23andMe after purchasing kits for my wife and myself. Of all the testing options I've explored, 23andMe is the fastest - with a turnaround time of about 3 weeks for reporting the initial results. In contrast, AncestryDNA results appeared about 6 weeks after sending the sample (it was just after the Christmas holiday, so they probably were dealing with a high sample volume during this period). Results from the Genographic Project test, and various specialized tests I've done through Family Tree DNA (the testing lab for The Genographic Project), have generally appeared anywhere from 6 to 16 weeks after sending samples. The turnaround time for results is important to me when considering the use of DNA testing in a course, because I want to meet with students and know that they are well informed about potential surprises that may come from doing a genetic test. Since I'm enabling potentially life changing revelations, I think it is my responsibility to clearly inform students of potential outcomes (I'm shocked to read that someone in my position would not realize these outcomes). For a course that meets once a week, two weeks of the 15-week semester are needed to ensure students are well informed before consenting to test. I scheduled the course so that samples were shipped for processing after the 2nd class meeting, and I had planned to start working in class with the results in the 8th week allowing 6 weeks for shipping and processing. It turned out that students started getting results on the 23andMe website after about 2 weeks of processing time, so we jumped into the results during the 5th week - quicker than I had initially planned.

After evaluating the most popular options for personal genetic tests over a period of about 6 months, I concluded that 23andMe was the best option for my course. It has the most entertaining and engaging website features, reporting haplogroup identity, ancestry composition, and relatives. Furthermore, I wanted to use the test results to help students understand the science underlying the predictions, and the 23andMe website provides the ability to explore the underlying genetic similarities upon which relationship predictions are based. I used exercises constructed around the website to illustrate genetic principles. In some of these exercises students used their own data, or a student could use data provided by 23andMe. Given the availability of data in a 23andMe trial account, a student is not required to submit a sample to participate fully in the class. However, some of the students in the class already had an interest in genealogy, and unfortunately 23andMe is not a very good site for identifying the common ancestor(s) responsible for genetic matches between relatives (this hypothetical scenario discussed in the first class is often a reality). At least what a student learns from the course will be useful if the student decides to incorporate genetics into their genealogy research using a different platform (e.g., AncestryDNA, Family Tree DNA).

After using the 23andMe testing platform and website for the first offering of this seminar, I was very pleased with this choice. Because the sample processing time was faster than I had initially envisioned, this accelerated our engagement with the ancestry predictions from the data and allowed more class time for engagement with the genetic basis of physical differences than I had originally anticipated. Much has been said about the actions of the FDA toward the health reports provided by 23andMe. However, the demo version of the 23andMe account, enabled by adding the Example Profiles into any 23andMe account (we used the 'Mendel family' to illustrate the transmission of chromosomal regions), also provides access to the summaries of health-related and other trait information. Although the predictions are based on an example genetic profile, the 23andMe trait summaries are excellent and provide a useful guide for self exploration from your own genotype data accessed either through direct links from the pages or through the genome browser feature.

I'm offering the seminar again in Fall 2015. In this second iteration I'm changing the name to Who are you? Revelations from the personal genome which better reflects the broader engagement with ancestry and trait information through genetic data.

My Personal Genome

As I explained, I investigated each of the most popular direct-to-consumer genetic testing companies and their website presentation of the data. I also downloaded data from each of the tests and analyzed the data through other sites. One site I particularly like for the evaluation of DNA matches is GEDmatch. I uploaded all of the data from my immediate family, and thankfully the genetic relationships are consistent with my expectations. Half of my genome is shared with each of my parents, and my son contains half of mine and my wife's genomes. The image at right shows me within this pedigree and the estimated number of generations connecting each of my immediate family members.
Each parent-child comparison is 1 generation removed, or 50% genetic identity, and the grandparent-grandchild comparisons show 1.5 generations, or 25% genetic identity.

Interestingly, GEDmatch provides several tools for working with data downloaded from tests by 23andMe, AncestryDNA, or Family Tree DNA and uploaded to the site. The problems are that the site is frequently down due to the heavy traffic, and this even with the user database being not nearly as large as each of the databases maintained by the companies. When searching for relatives to identify a common ancestor(s), database size and composition are critical. Although I like the tools available in GEDmatch, I have not been very successful using the site to support my family's genealogy with identified genetic relationships.

Thus far, my most successful attempt (and most of my efforts) at genetic genealogy has been in the McAllister family. We already had pretty good records of the McAllister family history, so it's simply an effort to identify genetic similarities with relatives that share a common ancestor(s) to support the family tree. The connections among the relevant members of the family tree are illustrated with the extended pedigree to the right showing me and coloring in green individuals that have done genetic tests where DNA matching supports a genetic relationship. I'm a descendant of Patrick McAllister, born in Ireland in 1830. He departed from Belfast and arrived in Philadelphia in 1848 accompanied by his mother Mary, sister-in-law, a nephew, and a niece. Patrick's older brother Thomas McAllister apparently migrated from Ireland to the USA previously, but we have not been able to find record of his arrival. Patrick settled in Ohio. Thomas and family lived temporarily in Pennsylvania before also moving to Ohio.

Very soon after my father's results were available on AncestryDNA, I identified three matches between my father and descendants of Thomas McAllister. One of these is a mother-daughter (daughter excluded from the pedigree), so it is really two independent matches that are closely related (the aunt-nephew pair). Simply by entering 'McAllister' in the surname search feature of AncestryDNA, a list of matches is returned each containing McAllister ancestors in the tree linked to the DNA test. A bit of searching on Ancestry.com for missing relatives and the connections become clear. About a week after my parents results were reported in AncestryDNA, my dad had a really close match appear that was reported as a 1st cousin. This one was easy. It's my 1st cousin, an avid genealogist and history buff, that apparently also was gifted an AncestryDNA kit. Other matches in this McAllister family have appeared in AncestryDNA since my dad's test was completed. One of these is another descendant of Thomas McAllister. This individual descends from a different child of Thomas', so I have now connected with descendants of two of Thomas' children - my 4th cousins. Interestingly, the match with an estimated genetic relationship that is closest with my father (3rd-4th cousin) is one more generation removed from my father. Each of the other matches is estimated as a 4th-6th cousin.

AncestryDNA does not provide the ability to look at the underlying genetic similarity, e.g., percent of the genome that is identical, the number and length of matching segments, so I'm unable to investigate the matches further on the site.

The success in making connections within the McAllister family motivated me to do a specialized test on my Y chromosome to see if I can identify more distant matches specifically to my paternal, McAllister, lineage. None of the McAllister descendants that I have communicated with have made any progress finding records of our ancestors in Ireland. This has been a major brick wall in our genealogy for many years, so I thought a test of my Y chromosome may help to make a connection. The only information available is that Thomas' headstone reads 'Native of County Down.' Note that we do not know the maiden name of Mary, Patrick and Thomas' mother, and we do not have any record of their father either.
I opted to test 67 STR (short tandem repeat) markers on my Y chromosome using Family Tree DNA. As I waited for the results, I joined the McAllister surname project so that my results would be compared specifically with others having McAllister as a surname. When the results were returned, it was immediately clear that I'm not very similar to any of the other participants in the McAllister surname project. Currently I'm in Subgroup #13 with many differences compared to others in the group. In contrast to not having a close match with others in the McAllister project, I have lots of close matches among all individuals that have done Y-chromosome tests with Family Tree DNA, but none yet that are identical to my Y chromosome at all 67 markers tested. The predominant surname among my matches is also not McAllister, rather it is various forms of Irwin (Irvin, Urwin, Ervin, etc.). I included a screenshot of my matches that differ at a single marker out of the 67 tested. Of the 14 matches, 6 are individuals with some form of Irwin as a surname. Given the result, I quickly joined the Clan Irwin surname project. My Y chromosome falls within a large group descended from an ancestral family in the Scottish Borders region, from which there was migration to the Ulster Region of Ireland.

After getting the results from the analysis of my own Y chromosome, I began actively seeking a male patrilineal descendant of Thomas McAllister to solidify the link between the immigrant McAllister brothers and to confirm the findings from my test. I searched Ancestry.com for evidence of a surviving male lineage of Thomas, I also requested assistance to find a candidate from my newly discovered DNA cousins, when out of blue I was contacted by email from such a guy that belongs to a McAllister line for which I was not able to make much progress researching. He consented to providing a cheek swab sample that was tested at 37 STR markers on the Y chromosome. I wrote previously that I didn't have any identical matches in the Family Tree database. Well, this is because I have a fairly recent mutation at one of the
markers on my Y chromosome. There has been a change at one marker in one of the generations of my descendant lineage since his and my common ancestor - the unknown man that was the common father of Thomas and Patrick. Without this mutation, the results of Thomas’ descendant more strongly place our McAllister family within the Irwin family by his identity to the modal haplotype of the Irwin Borders Group. It seems pretty clear that at some point in the past, but prior to the migration to the USA, there was a transition in our surname (an NPE or non-paternal event) from Irwin to McAllister.

So, who am I? It appears that I’m Bryant Fulton McAllister Irwin. Each of my given names is a family name, Bryant and Fulton respectively being the family names of my paternal and maternal grandmothers. Now I have a new family name, Irwin, and I’m interested to learn more about this connection and the connections in other branches of my family tree that can be explored through genetic analysis of myself and my relatives.