



Whole Genome Sequencing (WGS)

by Scott E. Lee

Scott@ScottLee.com

GenealogyTechTalks.com



This article covers the state of genealogical DNA testing as of September 2026.

Until recently, three types of DNA tests have been used by genealogists: autosomal (the most common), Y-DNA, and mitochondrial DNA. However, a new test is gaining traction: whole-genome sequencing (WGS). As its name implies, it examines the entire genome rather than just parts of it. Traditional genealogy DNA tests examine selected markers or specific regions of the genome rather than sequencing the entire nuclear genome. Autosomal tests typically examine hundreds of thousands of selected SNPs, while Y-DNA and mitochondrial tests concentrate on the Y chromosome or mitochondrial genome. WGS attempts to sequence nearly the entire genome, including the Y chromosome and mitochondrial DNA, over 3 billion nucleotides (or 6 billion if counting both copies).

On the surface, it appears this completely replaces all the other tests, but that's not necessarily true. Even though it collects the raw data on the Y chromosome, it does not do the analysis provided by Y chromosome testing. So the raw data is there, but the analysis is missing in the current WGS testing. The same is true for mitochondrial DNA testing. Although WGS samples the entire mitochondrial DNA, it does not include the analysis provided by a mitochondrial DNA test. That may change in the future, but that's the state today.

Another issue is that WGS is more expensive than other DNA tests, ranging from hundreds of dollars to over a thousand. At least one genealogy company (MyHeritage, as of September 2026) is offering WGS at a lower cost but with low read depth. The gold standard for WGS is a read depth of 30X, meaning that, on average, every position in the genome is sampled 30 times. Some spots will be sampled more often, and some less. However, currently, MyHeritage uses 2X. That means each spot of the genome is sampled only twice on average. As a result, some places are sampled more often, while others aren't sampled at all. Low-pass WGS works for autosomal genealogy because common variants occur in characteristic combinations called haplotypes. By comparing the observed DNA with large reference panels, statistical imputation can infer many variants that were not directly observed. This works particularly well for common autosomal variants but is much less suitable when the goal is to discover a rare or previously unknown mutation.

When choosing a DNA test, three use cases should be considered:

1. **Autosomal matching and ethnicity** – 2X read depth, combined with statistical reconstruction, appears adequate for company-provided autosomal matching and ethnicity estimates. This is why MyHeritage now uses this in place of traditional autosomal testing. The statistically inferred variants appear to be accurate enough for these company-provided results.
2. **Y-DNA and mtDNA lineage analysis** – 2X read depth is weak for this use case.
3. **Archival/future-proof raw data** – 2X read depth is poor for this use case. 30X is strongly recommended instead.

The reason 2X WGS is adequate for autosomal matching and ethnicity, but not for other uses, is because autosomal DNA is passed down in large, blocky segments, making it easy for an algorithm to use statistics to fill in missing data based on closely related populations. In contrast, Y-DNA and mtDNA lineage analysis often hinges on identifying individual, single-nucleotide mutations. If a 2X low-pass scan misses that exact spot, a critical branch of the family tree remains invisible.

Perhaps the most significant factor in choosing a DNA test is database size. Testing with a company that has a large user database will yield many individuals with matching DNA. For autosomal DNA, these are the approximate publicly reported or estimated database sizes (number of people taking that test):

- Ancestry.com is the largest, with over 30 million tests.
- 23AndMe is second with about 14 to 15 million tests.
- MyHeritage DNA has about 9.6 million tests.
- GEDmatch has about 2.2 million tests.
- FamilyTreeDNA has about 1.8 million tests.
- LivingDNA has about 1 million tests.

But that doesn't mean you should only test with the largest – test with all that you can afford. Some allow uploading a test from another company for free or at a reduced cost.

For traditional Y-DNA and mtDNA testing, FamilyTreeDNA is the only major, publicly accessible database.

FamilyTreeDNA has taken a slightly different approach. Its Family Finder test now uses targeted next-generation sequencing (NGS) rather than whole-genome sequencing. Instead of reading the entire genome at low depth, it reads selected genealogically useful regions at higher depth. This may prove to be another path forward for autosomal genealogy testing.

Whole Genome Sequencing (WGS) is the best choice for archival/future-proof raw data. Because it records DNA across the nuclear genome, including the sex chromosomes, and also includes mitochondrial DNA, it is the most complete genetic record available to genealogists. 30X WGS is not perfect (some regions are hard to sequence or align, and coverage varies), but it is the closest practical consumer test today to an archival record of a person's DNA.

Since genealogy DNA companies are not yet offering high read depth testing, the main alternative for high quality WGS is to use a health-oriented company such as Nebula Genomics, SelfDecode, or Sequencing.com. These companies will allow you to download the raw data, which may be 20 to 100 gigabytes, as a BAM or CRAM file (standard formats for storing DNA sequence data). This raw data is not directly useful for genealogy because there is no provision to identify individuals with matching DNA. To make it useful for genealogy, this raw file must be converted to another format and stored in a database alongside other tests for matching. Today, the options are:

- **YFull.com** – The BAM/CRAM file can be uploaded to YFull.com, which, for a fee, extracts your Y-DNA and mitochondrial data, assigns you a highly specific haplogroup, and places you on its Y-DNA and/or mitochondrial trees, where you can compare your results with others on nearby branches.
- **mitoYDNA.org** – There is also mitoYDNA.org, a free, crowdsourced database where users can upload extracted mitochondrial and Y-DNA data to match with others tracing the same maternal or paternal lines. Although, it does not currently process Y-SNP data.
- **WGSExtract** – Another possibility is to install WGSExtract on your computer (which is technically challenging) to read the BAM/CRAM file and produce a file that resembles raw data from AncestryDNA or 23andMe, which you can then upload to GEDmatch, FamilyTreeDNA, and LivingDNA.

Another reason to take a 30X WGS DNA test today is future-proofing. 30X WGS may be the best archival DNA test currently available. Many of the results produced by other DNA tests can potentially be derived or approximated later from a high-quality WGS BAM/CRAM file. This is especially useful for testing older individuals who might not live to see the genealogy DNA companies catch up. It may be expensive, but testing your parent or grandparent with 30X WGS today may be very valuable someday when they are no longer with us. Processing that raw data today is not necessary. It will likely become easier and cheaper later. But when your loved one is gone, it is too late.

The bottom line is that today's 2X WGS tests are limited, but that will likely change in the future. For reliable direct variant calling without heavy reliance on imputation, roughly 15X or greater is much more useful, while 30X remains the common standard for

high-quality WGS. Eventually, as costs come down, the genealogy companies will offer a much deeper test, such as 15X, or 30X, and will likely also offer the same analyses and database matching as Y-DNA and mitochondrial testing today. And because WGS samples every nucleotide, the analysis could potentially go much deeper than the current tests, offering far greater detail. For now, genealogists willing to navigate the technical steps can already benefit — and those who test at 30X today will be well positioned as the tools catch up.

It's worth noting that WGS is much more revealing than a standard genealogy test. It may expose medically relevant variants and carrier status. Users should understand each company's policies on data sharing, research use, and law-enforcement access before uploading raw DNA files.

Copyright © Scott E. Lee. *This article is licensed under Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).*

You may copy, share, redistribute, adapt, and build upon this article for noncommercial purposes, provided that appropriate credit is given, a link to the license is included, and any changes are indicated. Notification of redistribution is not required, but it is appreciated. If this article is reprinted in a newsletter, journal, syllabus, website, or other publication, the author would appreciate receiving notification and a copy or link at Scott@ScottLee.com.

License details: <https://creativecommons.org/licenses/by-nc/4.0/>