

Maximizing Your Use of GEDmatch

Tim Janzen MD tjanzen@comcast.net

GEDmatch (<https://www.gedmatch.com>) is DNA analysis company used by genetic genealogists that was founded in 2011 by Curtis Rogers and John Olson. The GEDmatch database now holds autosomal DNA data for approximately 1 million people. The database is growing quite rapidly with over 1500 raw data files being uploaded each day on average. GEDmatch has many features that the primary autosomal DNA testing companies do not offer. The GEDmatch website provides a means for people to compare autosomal SNP data among the four primary genetic genealogy companies that do autosomal DNA testing and provide match lists: 23andMe, Family Tree DNA, Ancestry.com, and MyHeritage. This option is crucial for genetic genealogists who would like to find additional matches above and beyond those in the databases of the companies that they have already tested with. This website is particularly important for Ancestry.com customers since Ancestry.com does not reveal the matching half-identical region (HIR) data to customers who test there. It is thus imperative for Ancestry.com customers to transfer their data to GEDmatch so that they can see which HIRs they share in common with other Ancestry.com customers.

The first thing to do when you start using GEDmatch is to create an account there. Once you have done that, you can start up uploading raw data files for your family members and for yourself. Uploading data is relatively simple to do. After logging in, go to https://www.gedmatch.com/g_upload1.phpnf?u=1795 and enter the appropriate information in each field. Then at the bottom of the screen choose the appropriate raw data file to link to the profile your creating and then click on “Upload”. Within a minute or so, GEDmatch will generate a kit number for the raw data file that you have uploaded.

Learning how to use GEDmatch can be somewhat challenging for neophytes. I suggest that new users become familiar with the home page first before starting to use the basic comparison tools at GEDmatch. On the home page you will find information about your profile, a helpful “Learn More” section which contains links to various different help sections, and “Your DNA Resources”, which is where you see a list of all of the people whose autosomal DNA files you have uploaded to GEDmatch as well as their GEDmatch kit numbers. At https://www.gedmatch.com/Using_GEDmatch.php you can view 3 excellent presentations about the basics of using GEDmatch as well as some additional more advanced features. GEDmatch also offers a wiki at https://www.gedmatch.com/gedwiki/index.php?title=Main_Page which provides additional background information about GEDmatch and autosomal DNA analysis.

At https://www.gedmatch.com/msg_list_forums.php you can learn more about the GEDmatch Forum. This forum provides users with a means to correspond with other GEDmatch users. The topics are broken down into 6 primary categories: general interest, genealogy, DNA utilities, Tier 1 utilities, admixture utilities, and ancestor projects. There are over 7700 threads dedicated to these topics with over 1.1 million views to the posts in the forum so far.

While a basic account at GEDmatch is free, GEDmatch users will want to become Tier 1 members if they start to use GEDmatch regularly. There is a monthly subscription fee of \$10 to be a Tier 1 member. The utilities only available to Tier 1 members include a versatile “one-to-many” match feature and the “Matching Segment Search” utility. GEDmatch also offers two triangulation tools and a phasing tool found only among the Tier 1 utilities.

The first and most basic feature that I suggest everyone become familiar with is the “User

Lookup” feature at https://www.gedmatch.com/user_lookup1.php. I use this feature on a regular basis. I sometimes search for specific email addresses to see if the person associated with that email address has uploaded a file to GEDmatch. I also use the feature to look up the email addresses for specific GEDmatch kit numbers. I also sometimes use it to see if there is a GEDCOM file linked to a specific kit.

The next feature that I suggest that GEDmatch users become familiar with is the “One-to-many” feature. When you access this feature at <https://www.gedmatch.com/r-list1z.php> you are asked to enter a GEDmatch kit number. This kit number can be the number for one of your family members or for yourself or it can be for any other person in the database who has a public GEDmatch kit. If you click on “Display Results”, GEDmatch will quickly generate a list of your closest 2000 genetic matches. Information provided about these matches includes their kit number, the type of raw data file that was uploaded to GEDmatch, a link to their GEDCOM file if they uploaded a GEDCOM file to GEDmatch, the total number of centimorgans (cMs) that they share with you on the autosomal and X chromosomes, as well as their name and email address. If you click on the “L” hyperlink for any of your matches you are taken to the “One-to-many” results for that match. Clicking on the “A” hyperlink takes you to the “One-to-one” comparison tool where you can see precisely which segments (half-identical regions) that you share with that particular match. Clicking on the “X” hyperlink takes you to a comparison screen where you can see which segments on the X chromosome that you share with that particular match. Matches in the “One-to-many” list can also be sorted by kit number, Y chromosome and mitochondrial DNA haplogroups, name, and email address.

After you have familiarized yourself with the “one-to-many” feature, I suggest that you then learn how to use the “One-to-one” comparison tool. This is the tool that I utilize the most frequently at GEDmatch. This tool allows you to compare any two public kits at GEDmatch to each other to see which segments (half-identical regions) that they share in common with each other. The output from these comparisons can be displayed as a chart showing the start and stop positions for the matching segments or the output can also show a graphic display of the matching segment data. I really like the fact that the SNP threshold can be adjusted to as few as 50 SNPs and the minimum segment size can be lowered to as little as one centimorgan. There are sometimes situations when I am doing chromosome mapping where I need a lower matching segment threshold than what the major genetic testing companies allow.

A feature at https://www.gedmatch.com/seg_srch1.php?u=1795 allows you to copy the matching segment data for all of your matches at GEDmatch and download it into a spreadsheet if you are a Tier 1 member. I generally suggest that typical users use the default SNP threshold of 700 SNPs and use a minimum of 7 cMs as the threshold for matching segments. The matching segment data can be integrated into a master match list spreadsheet that also includes the matching segment data from MyHeritage, Family Finder and 23andMe.

Another helpful tool at GEDmatch is the “People who match one or both of 2 kits” found at https://www.gedmatch.com/d_list1z.php. You can compare two kits to each other and find the other kits in GEDmatch who match one or both of these two kits. You can select a number of these kits for additional analysis. Then after clicking on the “Submit” button you are given the option to do additional comparisons using two different types of chromosome browsers, see matrices of the comparison data, generate CSV files that include the matching segment data or the match lists, find matching GEDCOM files or create tag groups from the matches.

Among the more sophisticated tools that GEDmatch provides are its two phasing tools.

The basic phasing feature may be found at <https://www.gedmatch.com/phase1.php>. This tool allows you to generate a phased kit for a child by entering the kit number for the child and one or both parents. Two new kit numbers are generated, one for the child's paternal DNA and the other for the child's maternal DNA. Another type of phased kit that can be generated using the tool at https://www.gedmatch.com/phase_alt1.php is what John Olson refers to as the "My Evil Twin Phasing" tool. John created this tool after discussing this topic with me several years ago. This type of phased kit includes the DNA that the parent did not pass down to a specific child. For example see PT786939M2, which is a phased kit which generates a list of matches that my mom has but that I don't have. Both types of phased kits are helpful to have for use by GEDmatch customers. The advantage of using phased kits is that you get fewer false matches when reviewing matches who share segments with you that are 12 cMs or less in length.

GEDmatch also offers some tools to help you compare a GEDCOM file that you have uploaded to GEDmatch to other GEDCOM files at GEDmatch. You can compare your GEDCOM file to all of the other GEDCOM files at GEDmatch or you can simply compare two GEDCOM files to each other. You can also search all of the GEDCOM files for a specific person. You may also search for GEDCOM files linked to your DNA matches. There are over 85,000 GEDCOM files on GEDmatch containing over 190 million people in aggregate.

GEDmatch and WikiTree have been cooperating since 2016. Anyone who has uploaded their DNA data to GEDmatch may now link to their personal profile on WikiTree. You can find all of your matches who have linked to WikiTree using GEDmatch's Tier 1 utility feature "One-to-many matches" or you can scan through your matches from the regular "One-to-many matches" comparison feature.

A very nice multiple-kit analysis feature is available at <https://www.gedmatch.com/MultiKitAnalysis.php>. You can enter multiple kit numbers for a cluster of relatives and generate interesting statistical data from the comparisons. A similar feature called the 3D Chromosome browser at https://www.gedmatch.com/3D_ChromosomeBrowse1.php allows you to do some similar comparisons and analysis.

The "Lazarus" feature at <https://www.gedmatch.com/lazarus1c.php?u=1795> is appreciated by power users. It allows you to create a kit for one of your close ancestors by using data from descendants of that ancestor as well as known cousins who are related to that ancestor. Use of this feature can more rapidly allow you to find your genetic cousins who are related to you on a specific ancestral line. Analysis of the pedigree charts for the people who match a kit generated by the Lazarus tool may allow you to find genealogical connections that would be more difficult to find by simply looking at the matches for your relatives one at a time.

GEDmatch also offers two triangulation tools. The "Segment Triangulation" tool at <https://www.gedmatch.com/triangulate1.php?u=1795> generates a list of triangulated matches who are among your closest 500 matches. I generally set the threshold limit for this tool at 250 cMs or 500 cMs. This feature can take as long as 45 minutes to run, but the information that it provides is helpful in that you can see multiple matches who all share the same segments. Reviewing the pedigree charts for the people who share the same segment may help you find a shared ancestor, surname, or geographic location. Another triangulation tool called the "Triangulation Groups" feature at <https://www.gedmatch.com/phastree1.php> allows you to group your triangulated groups together. You can include as many as 600 kits when using this triangulation feature. I generally set the upper cM limit of matches to use for this tool at 250

cMs or 500 cMs. This tool runs much faster than the “Segment Triangulation” feature. The results are displayed in a tree format or in a table format with an accompanying display for each chromosome. The output may also be downloaded as a CSV file for additional analysis.

GEDmatch also offers a number of biogeographical (ethnicity) analysis tools that help complement the ethnicity estimates offered by other genetic genealogy companies. These tools include the MDLP Project, Eurogenes, Dodecad, HarappaWorld, Ethiohelix, puntDNAL, and GedrosiaDNA. These tools can be helpful in analyzing one’s overall ethnic ancestry as well as for looking for smaller segments of DNA that originated from specific populations or ethnic groups. MDLP World 22 is a good general option. Dodecad is a nice general option and is particularly good for analysis of people with ancestry from Africa. Eurogenes is generally the best option for people with European ancestry. HarappaWorld is the best option for people with ancestry from South Asia. EthioHelix is probably the best option for people with African ancestry. It recommended that GEDmatch users use the K10+French option for African Americans or other people with a combination of African and European ancestry. Eurogenes K12 is the best option for people with ancestry from Northern Europe. Eurogenes K13 a good option for people who have ancestry anywhere in Europe. The Eurogenes EUtest V2 K15 admixture tool is the most recent and most sophisticated option for Europe. The Eurogenes Jtest is the best option for demonstrating Jewish ancestry.

Due to the fact that there is insufficient overlap in the shared SNPs found on the Illumina OmniExpress chip used by most of the major genetic genealogy companies and the Illumina GSA chip used by Living DNA and 23andMe for its version 5 files, GEDmatch created a separate project in 2017 called the Genesis Project for people who have tested at Living DNA or have tested at 23andMe after July 2017. There is a separate login for the Genesis Project at <https://genesis.gedmatch.com/login1.php>. You may also upload raw data files that were created from the Illumina OmniExpress chip into the Genesis Project, but you need to be aware that you will get more false matches when you run comparisons against people who were tested on the Illumina GSA chip. In September 2018 GEDmatch introduced its “Q-matching” one-to-one comparison tool to help reduce the number of false positives that one gets in the Genesis Project. The Genesis Project allows the user to do the same functions found on the primary GEDmatch website and also provides an updated Multiple Kit Analysis tool and a new feature on the “One-to-one” comparison tool that shows the start and stop locations for fully identical regions. GEDmatch is in the process of transferring all accounts on regular GEDmatch over to the Genesis Project. Sometime in 2019 the primary GEDmatch website will be completely merged with the Genesis Project website. Matching segment data is not imputed at this time so a lower threshold for the number of SNPs is being used (200 to 400) in order see matches who share relatively short segments.

In summary, GEDmatch offers a powerful palate of tools that are of significant help to both beginning and sophisticated autosomal genetic genealogists. GEDmatch is certainly worthy of our support since many of the features at GEDmatch cannot be found among the features of the major genetic genealogy testing companies. Hopefully, GEDmatch will continue to increase the number and sophistication of the tools that it offers as the database continues to grow and as it continues the development of the Genesis Project.