

Introduction to Autosomal DNA Tools



Just as in the old joke about a new genealogist walking into the library and asking for “the book that covers my Johnson family,” some new DNA test-takers seem to expect the entire family tree as output after a DNA test. Genealogists soon learn that the book about their family does not exist until someone does the research and writes the book. Similarly, the first look at DNA test results makes a researcher realize that analysis is needed before those results can identify ancestors.

Many autosomal DNA tools are now available from testing companies and third parties. Although no list of tools in a limited space can be complete, the survey in this column describes some of the most useful tools available today.

Many of these tools are accessible on websites of the major testing companies (23andMe¹, AncestryDNA², and Family Tree DNA³) and the large DNA tool providers (DNAGedcom⁴ and GEDmatch⁵). Some tools are available on other websites.

In this column the term “focus person” identifies the person whose DNA account or profile is being used for the analysis. The term “test-takers” identifies the DNA matches of the focus person.

Using the match lists

Each testing company and tool displays a list of the test-takers in the database who share DNA with the focus person. This match list usually includes a prediction of how the focus person and the test-taker may be related (based on a mathematical prediction, not a family tree). The list, or its sub-pages, also includes information on the total amount of shared DNA, longest shared segment, and/or number of shared segments.

Debbie Parker Wayne, CG, CGL, is experienced using DNA analysis, as well as more traditional techniques, for genealogical research in Texas, the South, and the West. She coordinates the genetic genealogy courses at several genealogy institutes and is the Texas State Genealogical Society’s DNA project director. She wrote the NGS Continuing Genealogical Studies: Autosomal DNA course and co-authored the first genetic genealogy workbook, *Genetic Genealogy in Practice*, published by NGS. See <http://debbiewayne.com/> for more information.

Websites cited here were viewed on 15 January 2017.

1. 23andMe (<http://www.23andme.com>).
2. AncestryDNA (<http://www.ancestrydna.com>).
3. Family Tree DNA (<http://www.familytreedna.com>).
4. DNAGedcom (<http://dnagedcom.com>).
5. GEDmatch (<http://gedmatch.com>).

Matrix Matches									
	Steven	Willie	William	suzanne	Carolyn	Rob	Barry	Teresa	Beverly
Steven		✓	✓	✓					
Willie	✓		✓	✓					
William	✓	✓		✓					
suzanne	✓	✓	✓						
Carolyn									
Rob									
Barry									
Teresa									
Beverly									

☞ - This person is identified as a match.

Family Tree DNA Matrix.

The testing companies allow some of this data to be downloaded. DNAGEDCOM and GEDMATCH offer tools to download the data in formats useful as input to even more third-party tools.

For some matches, human eyeballs are still the best tools. The first step when looking at a new list of test-takers who share DNA should be to scan the user names. Where the test-takers use real names, a name may be known through documentary research. Perhaps a cousin known through correspondence back in the days of paper letters became interested in genetic genealogy and tested. The relationship to this person and the name of the common ancestor may be known.

The next steps may be to scan the match list for test-takers with (1) a surname of interest, (2) an ancestral surname that is familiar, or (3) the most DNA shared with the focus person. Some test-takers do not list ancestral surnames. Surnames listed with a date and place increase the likelihood of a researcher finding a common line.

Each testing company provides a match list of everyone in the company database who matches the focus person at the threshold levels enforced by that company. The GEDMATCH One-to-Many tool is the equivalent of a testing company match list.

Using family trees

Every person on a DNA match list is a cousin in some way whether the common ancestor can be identified or not.⁶ If both the focus person and test-taker have accurate and robust family trees, comparing the two trees should identify a common ancestor.

The researcher must always bear in mind the fact that two people may have more than one common ancestor. If either person has an incomplete tree for the timeframe of a genetic match, the common ancestor may not be

listed. The analysis must consider the possibility that the common ancestor may fit in one or more of the “holes” in the tree.⁷

Ideally, everyone has well-documented and individually researched family trees. An inaccurate family tree will erroneously appear to confirm a common ancestor, sending researchers down the wrong path. Care must be taken to confirm the accuracy of trees before reaching genealogical conclusions based on those trees. If four test-takers all copied the same inaccurate family tree, their trees will seem to confirm a common ancestor. The four test-takers may have a common ancestor who is not the one named in the inaccurate trees.

Each testing company offers some form of name and location search for the family trees provided by test-takers.

GEDMATCH offers several tree search options if a user has uploaded a GEDCOM file. Users can search GEDCOM files alone or with options that search only the GEDCOM files of DNA matches. DNAGEDCOM offers GWORKS to search GEDCOM files uploaded to the website.

DNARBORUM is a Chrome browser extension that displays a pedigree in Ahnentafel format for

6. A small number of persons on a match list may not be cousins (false positives) and some may share an ancestor so far back in time it cannot be proven with documentary research. The testing companies strive to provide as many real matches as possible with no false positives. The false positives may be the result of misread DNA data causing the comparison algorithm to see a DNA match where none exists. These false positives can sometimes be eliminated by testing parents, children, or siblings, and comparing their results to the suspected false match. False positives are more likely to appear in distant relationships than in close relationships.

7. Family trees are never truly complete. Every ancestor found has parents, and the genealogist must then search for the parents. Y-DNA and mtDNA matches may be many generations further back than can be documented.

some websites.⁸ An Ahnentafel to GEDCOM Converter converts an Ahnentafel file to a format that can be imported by a genealogy program or analyzed with a GEDCOM analyzer tool.⁹

Using chromosome browsers

Chromosome browsers offer a graphic display of shared DNA segments between a small number of test-takers. AncestryDNA does not provide a chromosome browser or the necessary data for segment analysis. Other testing companies and tools display the DNA segments shared by the focus person and those test-takers in the match list.

Researchers must understand that the chromosome browser only illustrates the DNA shared between the focus person and that person's DNA matches. Each person inherits one copy of each autosomal chromosome from the mother and one from the father. The chromosome browsers do not differentiate between those who match the chromosome 1 inherited from a person's father and the chromosome 1 inherited from the person's mother, even when the display indicates the segments overlap. There is no guarantee that the matches will match each other at all or in the same location where each matches the focus person as is required for segment triangulation.

In a triangulated group, the focus person (A) matches at least two others (B and C) on an overlapping segment **and** B and C match each other on the same overlapping segment. In order to confirm that B matches C on this same segment it is necessary to login to the account of B or C or use a third-party tool such as GEDmatch. The additional analysis of triangulation may help separate the maternal and paternal matches. An In Common With tool, described below, can also provide some clues to determine whether a match is on a maternal or paternal chromosome.



23andMe Chromosome Browser (New Experience).

Finding cousins who also match each other

Finding DNA matches who also match each other may provide clues to identify the common ancestor. For example, any test-taker who matches both the focus person and his or her paternal uncle may be related on the focus person's paternal lines. More analysis will be needed to ensure that the test-taker is not related on different lines, but starting with those who match a person with a known relationship is a good way to begin.

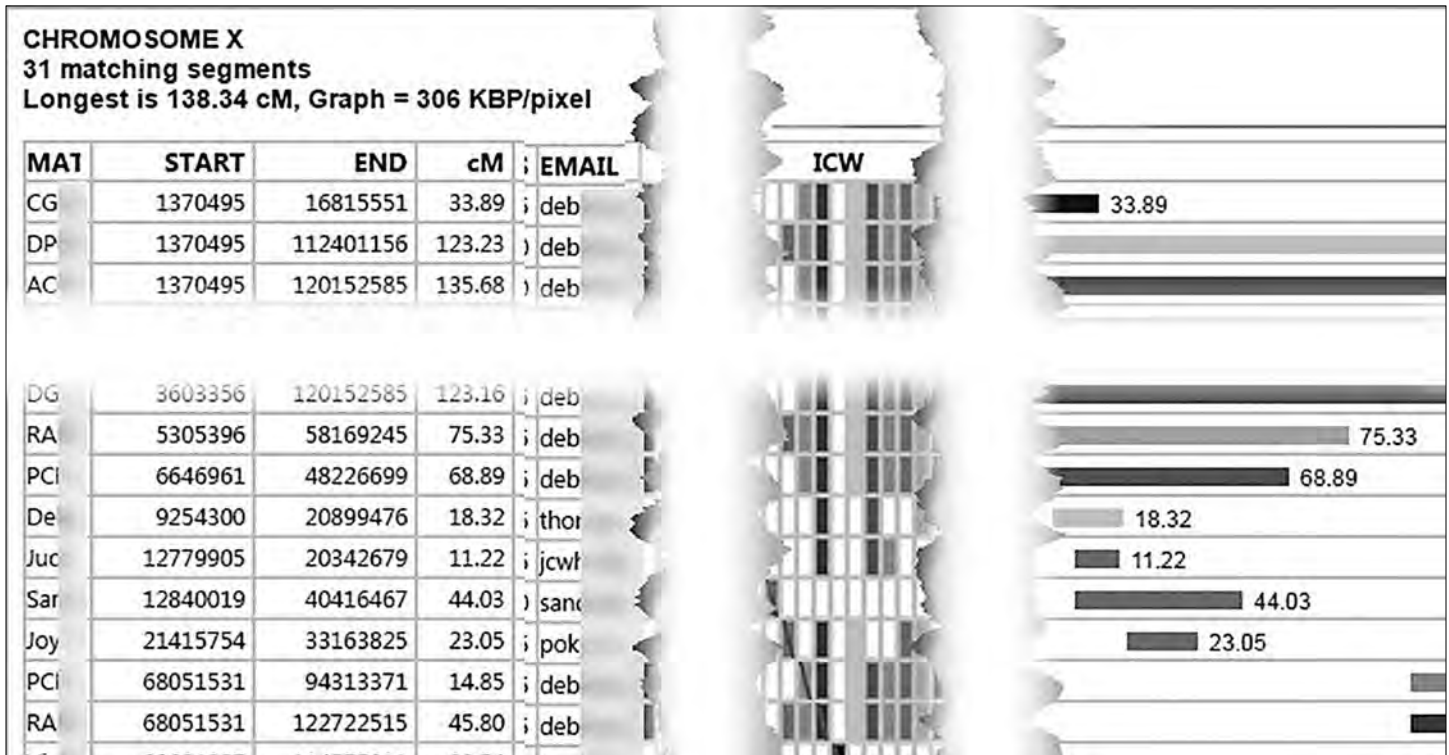
The tools of testing companies for finding cousins who match each other include In Common With (ICW), Matrix, Shared Matches, DNA Circles, and Matches of Matches. Some of these tools output a list and some generate a graphic format.

DNAGEDCOM's Autosomal DNA Segment Analyzer (ADSA) creates a graphic display showing ICW matches displayed side-by-side with the chromosome segments each test-taker shares with the focus person. ADSA allows more test-takers to be listed in one display than the company tools allow.

DNAGEDCOM also hosts JWorks, an Excel-based tool, and KWorks, a browser-based tool, to analyze match lists and ICW files. These tools eliminate small segment matches and group matches into ICW segments.

8. neanderling, "DNArboretum," *Google Play Store* (<https://chrome.google.com/webstore/detail/dnarboretum/oekcehcnbnfmeimggmkfliochkojkaj>).

9. Kitty Cooper, "Ahnentafel to GEDCOM Converter," *Kitty Cooper's Blog* (<http://blog.kittycooper.com/2015/12/announcing-my-ahnentafel-to-gedcom-converter>).



Autosomal DNA Segment Analyzer.

Finding cousins who share the same DNA segments

These tools help identify test-takers who share the same DNA segments:

- 23andMe: DNA Relatives tool
- GEDmatch: One-to-One, X One-to-One, Matching Segment Search, and Triangulation tools

The chromosome browser file lists the chromosome number with the start and end location of each matching DNA segment. The Double Match Triangulator¹⁰ uses the chromosome browser files of two focus persons to determine which DNA segments are shared by the two and by other test-takers in the database. The tool identifies segments that are triangulated and those that are not. The researcher must have access to the chromosome browser file for both focus persons.

Other useful tools

The Shared cM Project¹¹ is not a tool in the traditional sense but an extremely useful dataset listing the amount of DNA shared by two people with a known

relationship. Relationship predictions are based on statistical probabilities of the amount of DNA two test-takers are likely to share in a given relationship. Due to random recombination and inheritance factors, in real life the amount of shared DNA falls into a range as shown in the Shared cM Project.

The Shared cM Project is a collection of self-reported data showing the wide range of shared DNA amounts seen for relationships down to the level of fourth cousin, twice removed (4C2R). A few selected lower relationships are also listed. Caveats when using the Shared cM Project data include the possibilities of the user making an error when entering the data and of the true relationship being different than that reported. However, a statistician has reviewed the collected data and removed outliers, making the data more reliable.

A chromosome map can be useful to determine how other test-takers may be related to the focus person, once the common ancestors from whom specific DNA segments were inherited have been identified. For example, a focus person determines that the segment of maternal chromosome 2 starting

10. Louis Kessler, "Double Match Triangulator," *Behold Genealogy* (<http://www.beholdgenealogy.com/dmt.php>).
 11. Blaine T. Bettinger, "Update to the Shared cM Project," *The Genetic Genealogist* (<http://thegeneticgenealogist.com/2016/06/26/update-to-the-shared-cm-project>).

at location A and ending at location B was inherited from the maternal Anderson-McSpadden line, and that same segment of the paternal chromosome 2 was inherited from the paternal Parker-Black line. Then only the Anderson-McSpadden and Parker-Black lineages need to be searched when looking for a common ancestor with a new match on those same segments. Kitty Cooper's Chromosome Mapper¹² and Genome Mate Pro¹³ both provide chromosome mapper tools.

Genome Mate Pro also allows data from multiple testing companies and websites, and from multiple test-takers, to be combined in one database for analysis. Many of the functions described above are available. The database is contained on a personal computing device, not stored on a website. A disadvantage of this approach is that it takes some time to download the data from the testing company or tool website and import the data to a Genome Mate Pro database. The process must be repeated

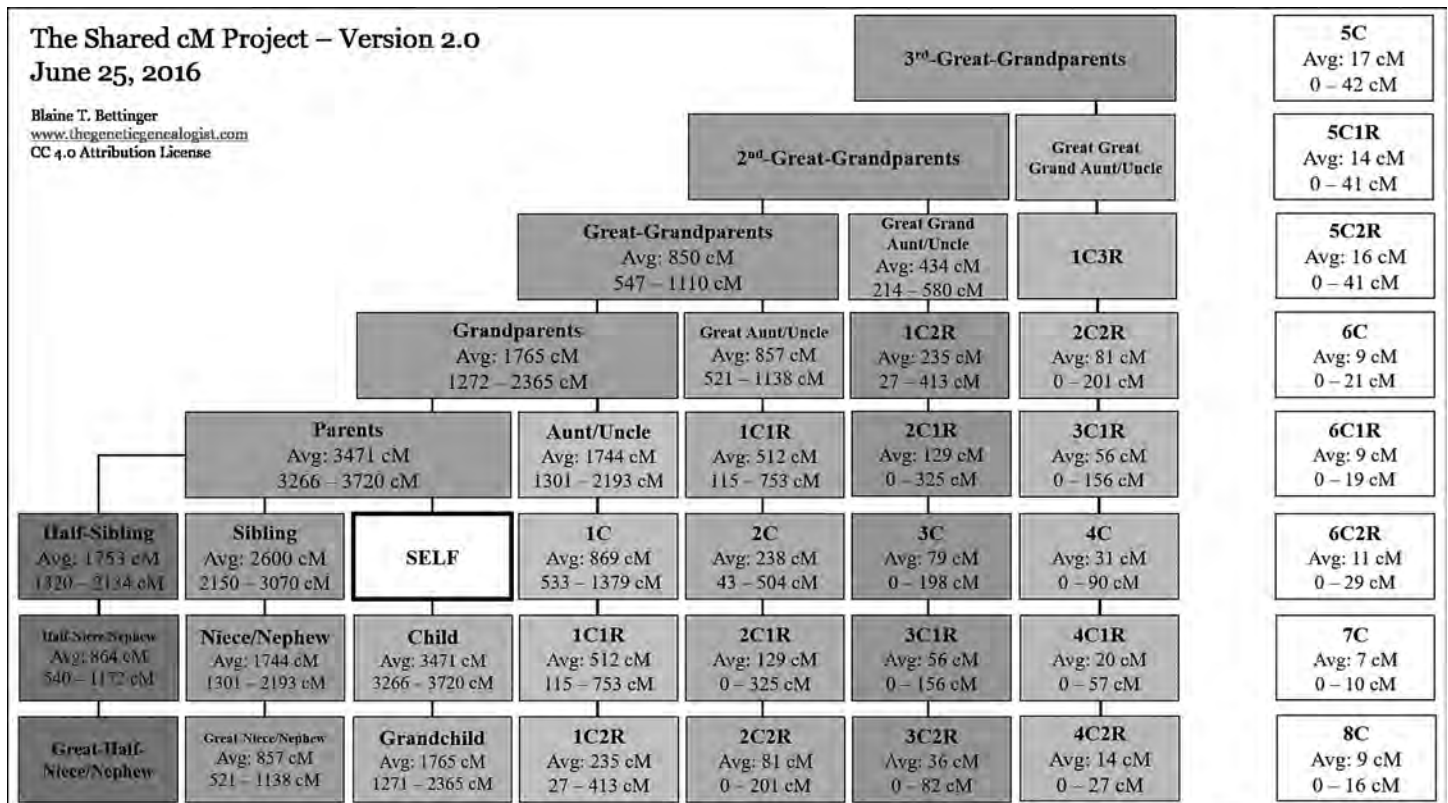
periodically to access new matches found after the initial import.

Conclusion

A wide range of autosomal DNA analysis tools is available to meet the needs of genealogists analyzing DNA test results. Some tools require the data to be available in a public database, and others allow the data to be held on a personal computing device. Genealogists should monitor online forums, blogs, and social media to learn of the availability and features of new tools. 🌳

12. Kitty Cooper, "Chromosome Mapper," *Kitty Cooper's Blog* (<http://blog.kittycooper.com/tools/my-graphing-or-mapping-tools/chromosome-mapper>).

13. Rebecca Walker, "A Tool for Managing DNA Comparisons," *Genome Mate Pro* (<http://www.getgmp.com/>).



Blaine T. Bettinger's Shared cM Project (CC 4.0 BY License).