Using the tools at GEDmatch: a free third party site for DNA comparisons

Presentation by Kitty Cooper, blogging at http://blog.kittycooper.com

Slides can be seen at https://slides.com/kittycooper/gedmatch/live

Table of Contents

Introduction

GEDmatch is a free third party site to which you can upload your DNA test results from any of the three main companies and compare them to other results in the GEDmatch database. There are currently 80,000 kits uploaded to GEDmatch. You can also use a number of tools there.

Some of the most interesting tools at GEDmatch are:

- **Ancestry composition** (called Admixture)
- **Comparing your DNA results**: one to many, one to one, triangulation
- **Phasing**: separating out what you got from one parent
- **Analyzing your DNA**: Parents related? Problems in your data? Rare SNPs?
- **Comparing GEDcoms** – we may not get to this one today

Admixture

- There are many ancestry composition calculators at GEDmatch
- All created by citizen scientists
- None are guaranteed
- Which one to use depends on where your ancestors are from
- Click on Admixture (heritage) to see your choices

Admix Choices

- **MDLP** – world 22 – General (n.b. Pygmy should perhaps be Mbuti)
- **Dodecad –** General and Africa
- **Eurogenes** – best for Europeans, has some special extras (jewish, hunter gatherer)
- **Harappa World** – best for South Asia
- **Ethiohelix** – best for African, use Africa+French for European mixture

Once you choose a calculator you are presented with further options; note that the number at the end is the number of reference populations the result is divided into

You might start with the MDLP world 22 calculator to get an overview. It uses specific colors for each population group which makes comparisons easier than with some of the other calculators

The reference population areas for the MDLP calculator are here:
http://magnusducatus.blogspot.com/2012/09/behind-curtains-mdlp-world-22-showcase.html

EUROGENES calculator

- North Europe does best in Eurogenes K12
- K13, the default, is good for most Europeans
- Jtest will show you if you have any jewish ancestry
- The hunter gatherer graph is amusing

Eurogenes is described here
http://bga101.blogspot.com/2012/05/geographicethnic-peaks-in-eurogenes.html

For African American use the African + French in the Ethiohelix calculator described here:
http://ethiohelix.blogspot.com/2012/06/intra-african-genome-wide-analysis-v2.html

For South Asian ancestry use the HarappaWorld calculator described here:
http://www.harappadna.org/

All the diagrams we have looked at have been for the entire individual. You can also get the chromosome by chromosome breakdown or compare two kits on a single chromosome.

Do you know the definition of cM or SNP or base pair?
Use the ISOGG wiki to look up terms like these
http://www.isogg.org/

See this ISOGG wiki article for the relatedness percentages showed in the slide

http://www.isogg.org/wiki/Autosomal_DNA_statistics

and this one from the ISOGG Wiki for the expected number of cMs and segments

http://www.isogg.org/wiki/IBD#Ranges_of_total_centimorgans_of_IBD_segments_expected.2C_based_on_family_relationship

Parent/child:	3539-3748 cMs
1st cousins:	548-1034 cMs
1st cousins once removed:	248-638 cMs
2nd cousins:	101-378 cMs
2nd cousins once removed:	43-191 cMs
3rd cousins:	43-ca 150 cMs
3rd cousins once removed:	11.5-99 cMs
4th and more distant cousins:	5-ca 50 cMs

Comparing your DNA results:

- **One to Many**
- **One to One**
- **One to One on the X**
- **An ICW function: People who match one or both kits**

The One to Many:

Looking at the one to many display, these are the columns

Kit Nbr	A=ancestry, F=FamilytreeDNA, M=23andme
Type	Which chip for 23andme (V3 or V4)
Triangulate	Click this to run the GEDmatch triangulation utility (actually an ICW function)
GEDCOM	If this person has uploaded a GEDCOM there will be an underlined G here to click
List	Click the underlined L to see list of matches for this kit (hint: right click and open in new tab)
Select	Check these boxes when doing a multi person function (described later)
Sex	M or F hopefully!
Mt	Mitochondrial DNA haplogroup (if the user entered it)
Y	Y chromosome haplogroup (if the user entered it)
	AUTOSOMAL
Details	Has an underlined A to click for the autosomal comparison
Total cM	Total cM segments > 5cM across all chromosomes sharedfor this match
Largest cM	The single longest segment of DNA shared with this match
Gen	Estimated number of generations between kits (1= parent/child, 2= 1st cousin, etc)

	X-DNA
Details	Has an underlined X to click for the X comparison
Total cM	Total cM segments > 3 cM on just the X chromosome
Largest cM	The single longest segment of DNA on the X for this match
Name	Often a pseudonym, sometimes a real name
Email	This cannot be cut and pasted, only the name, not the domain

What to do with a likely 3rd-5th cousin Match?

- If they have a G, click on it to look at their pedigree for common ancestors
- If you collect segment data, use the A and the X to look at shared segments and add that to your master spreadsheet for this kit
- Send your possible cousin an email with a link to your tree and your findings from the above and some of your surnames

The One to One Comparison

If you have clicked the A from the one to many then the kit numbers will be filled in for you.

Sometimes you may wish to lower the thresholds to see the smaller segments:
- The *SNP count minimum threshold* can be taken down as low as 300
- It is sometimes useful to lower the *minimum segment CM size* to 5 or 6

While the graphics are pretty, it is easier to cut and paste the results from the page with no graphics

An ICW function: People who match one or both kits

You can find the people who you match in common with (ICW) another kit by using the function called “People who match one or both kits” on your home page in the analyze section

Advanced uses of the One to Many Grid

The chromosome browser

The GEDmatch version of a chromosome browser lets you select multiple people from the one to many matrix and display the overlapping segments

- N.B. the color coding indicates the size in cMs since the graph is in terms of base pairs.
- Each person is a different line rather than a different color
- Old branch means that it is match to a previous person, a white line shows where

See the manual for GEDmatch available in my downloads area for details on the other advanced uses of the one to many matrix form

Phasing

- If you have at least one parent tested then you can separate what you got from each parent – this is called phasing
- GEDmatch will make you two new kits that have what you inherited from each parent
- You can then use the phased data for your one to one comparisons or any other function on the web site that takes a kit number
- Your phased kit number will show in your DNA resources box but it will not turn up in other peoples matches

From the GEDmatch wiki

"The program will look at every data line in the childs raw data file and try to deduce the paternal and maternal value. If the position is homozygotic (that is AA, or CC, or GG or TT) split the data to each phase If the position is a nocall it will be ignored. If the position is not included in at lease one parents data it will be ignored. If the position shows a data discord it will be ignored (see data discord below) If the maternal or paternal phase can be determined the values will be assigned else the original unphased values will be kept

Analyzing your DNA data

There are lots of fun things you can look at in your DNA results.

- Are your parents related?
- What color are your eyes?
- Any errors in your DNA upload?

Thanks for joining me, send me feedback via my blog

http://blog.kittycooper.com