Tools and techniques for predicting relationships with autosomal DNA

Jonny Perl

dnapainter.com

jonny@dnapainter.com / twitter: @dnapainter

When we look at our match list, we normally have a few burning questions, the main one being 'Who are all these people!'

This lecture will look at different ways to figure this out, discussing:

- What it means when you have a DNA match with someone
- Why the amount of DNA you share with a match can vary so much
- Tools that can help untangle mysterious DNA matches
- How these tools work, and their limitations
- The challenges of endogamy and multiple relationships

What is a DNA match?

This might seem like a strange question, but it might be useful to consider what it means when someone appears on your match list.

In general, this means that you:

- Share a common ancestor, or more often, one or more common ancestral couples
- Both inherited some of the same segments of DNA from these ancestors

However:

- Everyone's match list is different. Many of those with ancestry from a population where there has been less testing will have very few big matches
- Many of your smaller matches are going to share common ancestors with you that date from a period prior the "genealogical timeframe" for which we have paper records
- At the very low end of the scale (e.g. 6-15cM), the match itself might be false – the result of the testing site's algorithm mistakenly identifying a mixture of your maternal and paternal DNA as a segment that matches someone else.

What does a DNA match look like?

Your testing company will host your 'match list,' a list of all the people you share DNA with in their database, ordered with the closest relatives at the top.

For each relative, you will see some or all of the following:

- Name
- Number of centimorgans (cM) of DNA shared.
- Number of segments
- Estimated relationship
- Share matches
- If you're lucky, their family tree

If you're like me, your goal is then to do whatever it takes to identify this person so that you can figure out the connection. Why? Because you never know, this person may have information or materials that will help you in your research.

Occasionally, you might be able to see a connection immediately, but much more often, you'll have no idea. Hopefully, there will be shared matches and a tree that will help point you in the right direction, but the 'cM shared' amount will be key to figuring out how far back you need to be looking for the ancestors you have in common with any match.

The randomness of inheritance

When each of us is created, we receive one copy of each numbered chromosome from each parent. Each of these copies is a random mix of DNA from our parent's parents (our maternal or paternal grandparents).

Because of this randomness, the amount of DNA you share with your genetic relatives can vary significantly. For example, you might share 680cM with one first cousin and 1205cM with another; and you might share 180cM with one third cousin and nothing at all with another.

- You might happen to have inherited many of the same segments of DNA from the common ancestors.
- Or you might have inherited completely different segments, and thus share no detectable DNA at all with a genuine relative

- Current evidence and research suggests that relatives who are second cousins or closer will always share at least some DNA

Shared DNA ranges

It follows therefore that **ranges** exist, as opposed to specific amounts of DNA associated with specific relationships. There have been various initiatives to explore and define these ranges, both via computer simulations and via crowd sourcing.

The most prominent crowd-sourced initiative is the Shared cM Project, founded by American genealogist Blaine Bettinger in 2015.

Many ranges overlap, to the extent that you can never determine the relationship you have with a DNA match based on the amount of DNA you share alone; you'll always need some genealogical evidence to back it up. For example, a match of 900 centimorgans has many possibilities, including first cousin or half-aunt/uncle and niece/nephew.

Tools

My DNA Painter site hosts two popular tools that can help you narrow down the possible relationships based on the number of centimorgans (cM) shared:

Shared cM Project tool

The shared cM project tool is an online interface for Blaine Bettinger's Shared cM Project. (https://dnapainter.com/tools/sharedcmv4).

You can enter the number of cM you share with a match. All the relationships that are not possible for this amount will be faded out so you just see those that are possible.

In addition, the tool will present probabilities for each relationship. For example, sharing 70cM is possible for a 2nd cousin relationship, but very unlikely. A more distant relationship is much more likely.

WATO

What are the Odds? (WATO) is a more advanced tool that is useful for scenarios where:

- You have a mystery person (the “Target”) who shares DNA with a number of matches
- You are able to fit these matches into a family tree based on genealogical research.

WATO uses the same probabilities as the shared cM tool. These are based on the simulations cited in Ancestry’s White Paper on Matching.

In WATO, you build out a simple tree containing all DNA matches.

- For each person who has tested that you can fit into the tree, you enter the number of cMs of DNA they share with the “Target”.
- The target (the person you’re trying to locate in the tree) might be yourself, an adoptee you’re helping, or even just an unknown person who has appeared on your match list.
- You can then use the genealogical information you have (e.g. ages, locations of the testers and their ancestors) to make ‘hypotheses’ – different nodes in the tree where the target could feasibly fit.
- The tool will then produce a score for each hypothesis.
- Scores are relative to each other. The tool can help you make the most of the information you have about a match.

Limitations

While these tools can be very helpful, they are not the final word! Please bear in mind:

- Outliers exist; also, unknown multiple relationships and/or endogamy can inflate shared DNA amounts and make the results less accurate.
- Crowd-sourced data is extremely useful, but may have inherent biases (e.g. in favour of relationships where DNA is shared), and may contain errors.
- In the future, new simulations may provide additional insights and/or more precise predictions.

Endogamy and multiple relationships

At the time of writing, the impact that endogamy and multiple relationships can have on the amount of DNA shared is sufficiently variable that it has been hard to develop tools to help with prediction.

Those with endogamous ancestry will have noticed certain hallmarks in their matches, such as large numbers of relatively small segments. But based on crowd-sourced data, endogamy does not appear to automatically result in a higher amount of DNA shared. Different degrees of endogamy between the matches can present additional challenges to predictions.

For multiple relationships (e.g. double third cousins), the possible range of shared cM expands, with higher shared amounts possible. However, the small number of data points makes predictions more challenging.

Useful links

Shared cM Project

https://is.gd/0wHMiY

Shared cM Project Interactive tool with probabilities

https://dnapainter.com/tools/sharedcmv4

What are the odds? (WATO)

https://dnapainter.com/tools/probability

Frequently asked questions about WATO

https://dnapainter.com/help/wato-faq

Leah Larkin's blog posts about WATO

https://thednageek.com/science-the-heck-out-of-your-dna-part-1/

https://thednageek.com/science-the-heck-out-of-your-dna-part-7/

Andrew Millard's lecture about WATO

https://www.youtube.com/watch?v=Zf-Zeid6L2Y

Lara Diamond's Ashkenazic Shared DNA Survey

https://is.gd/5jcpm7