# What are the Odds?

June 3rd 2020

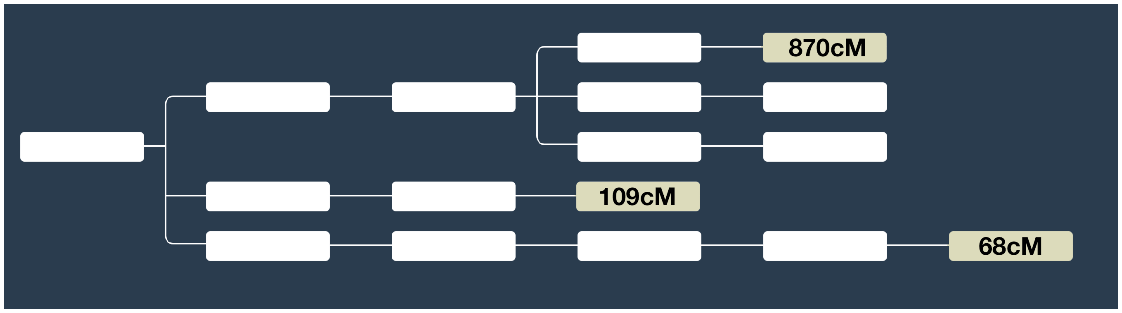
## An online tool that can help solve DNA puzzles

Jonny Perl / dnapainter.com / jonny@dnapainter.com / twitter: @dnapainter

## Introduction

What are the Odds? (WATO) is a web application for people who want to use DNA to solve a genealogical mystery. You can find it at https://dnapainter.com/tools

WATO works as follows:
- Once you are able to connect two or more matches in a family, WATO can help you figure out the position in this tree where you're most likely to fit.
- It does this with probabilities, making use of the amounts of DNA that you share with each match.
- These are measured in centimorgans, abbreviated as "cM".



### Who made WATO?

WATO was conceived by Leah Larkin, and the underlying model to calculate the odds was created by Andrew Millard. The online user interface was built by Jonny Perl (that's me). Leah and Andrew also both have webinars about WATO available, linked in the resources section below.

### Starting to make sense of your match list

When you first come to it, a DNA match list can feel overwhelming, but it contains many clues that can help you begin to solve your mystery.

These include:

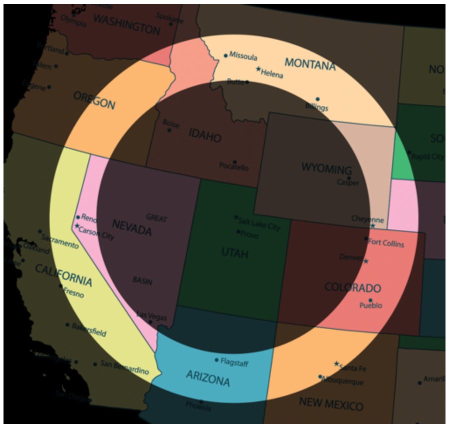
- The amount of DNA you share with each person
- Each person's 'Shared Matches' – which reveal networks of people who share DNA with each other

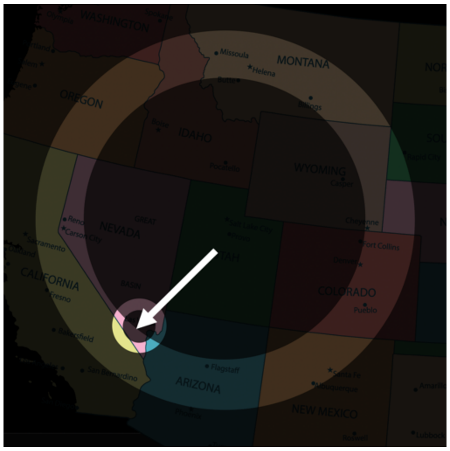
WATO aims to help you get the most out of this information.

## An analogy

To try and clarify the concept of WATO, I'll start with an analogy.

- I'm lost in the dark.
- I fumble for my torch and find a road sign and it says I'm 493 miles from **Salt Lake City**
- I have no idea in which direction those 493 miles are
- It's therefore very hard for me to locate myself using this information.
- I could be anywhere within this ring around Salt Lake City





- But if I can find out where I am relative to another city, this will help a great deal.
- For example, I know where Las Vegas is relative to Salt Lake City, so if I also discover I'm 73 miles from Vegas, I can use these two pieces of information together to narrow down where I am.
- From this information, I can't say **exactly** where I am, but I can certainly rule out several places.
- For example, I can't be in Montana because it's clearly far more than 73 miles by road from Las Vegas.

## Applying this to our DNA matches

Let's say I have a DNA match, **Sally Lake**.

- I share 109cM of DNA with Sally.
- This 109cM is an indication of how far I am from Sally in genealogical terms.
- Just as the distance on a road sign implies a range of 'as the crow flies' distances, this 109cM represents a range of possible actual relationships.
- The testing companies provide an approximated range.

Sharing 109cM, Sally and I could be 3rd cousins once removed

- If each generation is a 'step', that's 9 steps away from each other.

Or we could be closer, more like 1st cousins twice removed

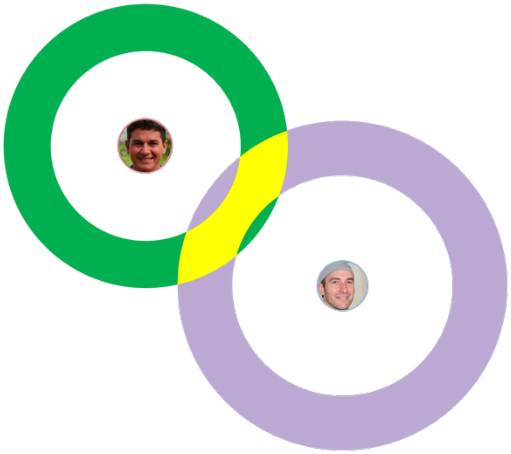
- That's 6 steps away from each other.

This means in order to find out how I'm related to Sally I need to go about 6 to 9 steps. But without more clues, I don't know the direction in which I need to take those steps – I don't know which branch of Sally's family tree I'm on.

## Finding more clues

When I look in ‘Shared Matches’ – a list of the people in the database who share DNA with both me and Sally – I find **Les Vegas**. I share 82.3cM with Les.
If I can find out how Sally and Les are connected to each other, then I can narrow down significantly the possible places on their tree where I might fit in.

The sleuthing work of placing people in trees is great fun for genealogists. I’ll discuss how to do this below. But if you’re very lucky, your match will have published a tree, making it easier. In this case, I’m able to figure out quickly that Les a third cousin to Sally on her maternal side.

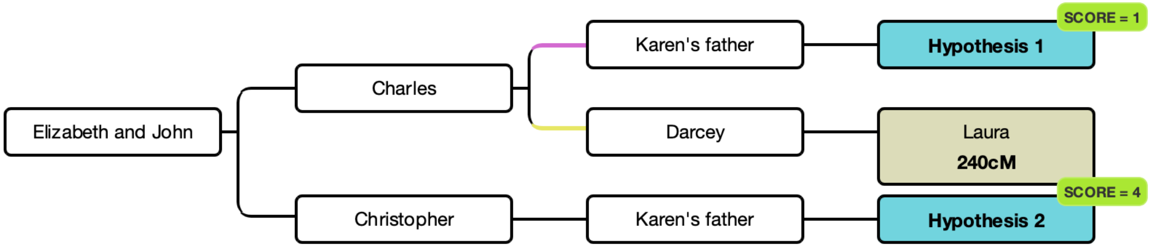


Fabulous. If I share 109 cM with Sally and 82cM with Les, and I know how they relate to each other, then I can narrow down the possible places where I might fit into their tree, and I can focus my research on these areas.

The green ring represents the possible ancestors via which I might connect to Sally, and the purple represents where I might connect to Les. The yellow overlap areas are therefore the branches of their tree I now need to focus on.

# How WATO is used

- WATO can be used to visualize and share a tree showing connections.
- It can help with genealogical mysteries such as unknown parentage cases.
- You can also use it to find an unknown grandparent/great-grandparent etc.
- But since you’ll be using the DNA matches of a descendant, you’ll need to duplicate the line from the mystery person down to the DNA testers in your WATO tree as shown below.



- It’s also possible to use WATO to research a mystery match – but unless you have access to the mystery person’s own matches, you will be limited to using shared matches on MyHeritage, 23andme and Gedmatch, all of which allow you to see how much DNA one of your matches shares with others. If you have multiple kits that match this person at Ancestry you could also use these.

## Who uses WATO?

WATO is used by a broad mix of amateur and professional genealogists, search angels, and more casual users who are curious about a family mystery.

## What are some of the challenges of using WATO?

Whatever the experience level of the user, it's not uncommon to feel overwhelmed.

The power of DNA brings the expectation that it can offer immediate solutions. But the reality is that this type of research can be very difficult and confusing.

However, there are some obvious routes to success.

## Before you start

- Define your goal: be completely clear about what it is you're trying to achieve
- This will also help you if you reach out to others to provide feedback; the sooner they can get their head around your puzzle, the sooner they can help
- WATO has a specific field: 'What is your research question' in order to help you keep this in the forefront of your mind. I try to keep this concise – e.g. "Who was Barry's father?"

Besides your research question, two other things are essential before you can use WATO:

1. **You need the amount of DNA you share (in centimorgans) with at least two DNA matches.** Remember that for WATO to work, you need to use matches from just one person.
2. **You need to know how these matches are related to each other.** Just as you know from a map where Salt Lake City is relative to Las Vegas, you need to figure out where the matches are on a family tree relative to each other.

## Placing people on a tree

Depending on how serious you are about genealogy, the prospect of building trees to connect your matches might feel like a delicious challenge, or it might seem like a time-consuming nuisance.

It would be possible to make an entire webinar on just this subject, so I'll be brief. I love the challenge of having to build a tree from very little information, being motivated by the possibility that this will somehow prove to be a big breakthrough!

- **Trees**: the person may have a tree linked to their DNA kit. Even if they have a tree with very little information, look for something 'searchable' – something distinctive in that tree, whether it's a name or a place
- **Other shared matches**: if they don't have a tree, perhaps another shared match does. If you're on MyHeritage or 23andme, look first at the people who also share significant DNA with the match as well as you.

- **Distinctive names**: if they have an unusual surname or even a merely distinctive username, this can be something you can search for.
- **Automated tools**: third party tools can potentially help speed up this search. These include DNAGedcom, Genetic Affairs, Shared Clustering and Genome Mate Pro.
- **Good old genealogical research**: Always needed. If the match feels like one that could really help you, chase lost causes and try to make that breakthrough.

### Glossary

Some terminology you'll come across in WATO:

- **Target**: This is the person you're trying to place in the tree. All the match cM amounts you enter should be from this person's match lists.
- **Hypothesis**: This is a position in the tree where you believe the target might fit.
- **Score**: This is an odds ratio that indicates how likely this hypothesis is compared to others.

## Case studies & Caveats

The webinar includes two brief case studies, each with different data and outcomes:

1. **Brian was born in 1954 to an unknown father**
    - Brian had a match of 870cM, indicative of very close family.
    - I used WATO to visualize and explore the possibilities and narrowed down his father to two brothers.
    - Based on genealogical information, we were able to determine that his father was James.
2. **Mary was born in 1940 and recently found out her biological father was not the man who had brought her up**
    - In Mary's case, the largest match was 114cM
    - We figured out a set of likely 2$^{nd}$ great grandparents and then worked each line forward, using genealogical facts to guide us
    - At each generation, we checked Mary's DNA matches for any sign of links to the spouse's family
    - We found a reasonable link and formed a hypothesis that Mary's father was one of two sons
    - We identified two living people who we could 'target test'. If our hypothesis was correct, they would be first cousins once removed to Mary.
    - One tested, and did indeed share an appropriate amount of DNA for a 1C1R
    - Since neither son had any other known descendants, we were unable to narrow things any further

### Limitations and caveats

#### Endogamy and multiple relationships are not accounted for

The probabilities do not account for endogamy, pedigree collapse or multiple relationships. If these exist on the lines you're exploring, WATO's predictions are far less likely to be useful or reliable.

This is because the range of possible cMs of DNA shared for a specific relationship is broader in cases where there's endogamy and/or multiple relationships.

#### Try not to use matches who are siblings

The mathematical model assumes that the probabilities for each relationship are independent of one another. If the matches are too closely related, this is no longer the case. It's therefore best to avoid adding siblings of the same family as matches, and you should aim for at least a 1st cousin level of separation between matches

## Introducing the new beta version

In the run up to this webinar I've been hard at work developing a long-overdue second version of WATO. I've finally had a chance to implement some of the features we talked about around the time of the tool's initial build in late 2017.

### Features of the beta version

- You can import the descendants of a selected individual or couple from a GEDCOM file
- WATO now uses birth/death information to evaluate hypotheses
- There's the option to suggest hypotheses based on the entered cM and the target's birth year
- You can enter spouse names for clarity within families with half-siblings
- The beta version uses updated probabilities based on newer information
- Improved usability and navigation

I'm looking forward to making further refinements based on user feedback.

## Summary

I hope this has helped you understand how WATO can help you
- Disprove hypotheses
- Visualize and share connections
- Help you identify branches of a tree that will help you solve your family mystery
- Clarify your thoughts

# Resources

## At DNA Painter

WATO v1
https://dnapainter.com/tools/probability

The Shared cM tool
https://dnapainter.com/tools/sharedcmv4

WATO beta v2
https://dnapainter.com/tools/wato

The Shared cM tool with new probabilities
https://dnapainter.com/tools/sharedcmv4-beta

## WATO: written resources

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

Leah Larkin's blog series
https://thednageek.com/science-the-heck-out-of-your-dna-part-1/

Facebook discussion group
https://www.facebook.com/groups/WhatAretheOdds/

## WATO: Video resources

Leah Larkin's lecture at Rootstech 2020
https://www.rootstech.org/video/introduction-to-what-are-the-odds

Andrew Millard's presentation at Genetic Genealogy Ireland in 2018
https://www.youtube.com/watch?v=Zf-Zeid6L2Y

Video tutorial by Andy Lee
https://www.youtube.com/watch?v=pFHLBBZbCWc

## Other resources

Blaine Bettinger's article 'Are You Doing Everything to Identify Your Matches?'
https://thegeneticgenealogist.com/2017/03/11/are-you-doing-everything-to-identify-your-matches/

Diahan Southard's book 'Your DNA Guide'
https://www.yourdnaguide.com/your-dna-guide-the-book