Fishing in Your Match List with MyHeritage DNA Tools

Paul Woodbury ● Research Team Manager at Legacy Tree Genealogists
(https://www.legacytree.com/)

Introduction

MyHeritage's AutoClusters are a great way to start exploring the clusters of your genetic cousins and narrow in on potentially pertinent genetic cousins, but sometimes AutoClusters can only take you so far. In this webinar, learn to leverage MyHeritage's Theories of Family Relativity, shared matching, and DNA match labeling to isolate pertinent genetic cousins for a research question and solve your family history mysteries.

Review of AutoClusters

In February 2019, MyHeritage DNA in collaboration with Evert-Jan Bloom of Genetic Affairs, introduced a licensed version of Bloom's AutoClusters tool as part of the MyHeritage DNA experience.[1] The AutoClusters tool is one of many genetic network tools that assists users in grouping genetic cousins based on their relationships to each other. These groups can then be analyzed further to determine which ancestral lines they likely represent, and better isolate the genetic cousins whose relationships are most likely pertinent for a research question.

Shortly after the release of the AutoClusters tool at MyHeritage, Ran Snir offered an excellent webinar through Legacy Family Tree Webinars describing how to obtain an AutoClusters report, and how to interpret it.[2] A follow up webinar offered additional insights on network theory, reviewed the elements of AutoClusters tool at MyHeritage, explored tips and tricks for interpretation, and considered direct applications of how AutoClusters might be used.[3]

AutoClusters are an excellent tool for understanding and organizing DNA matches as part of a first pass or initial review on DNA test results. However, MyHeritage Auto Clusters are limited to a test taker's first hundred or so DNA matches, and MyHeritage automatically sets the threshold for what is defined as a shared relationship based on the context of the user's test results. In order to solve historic research questions using DNA, it is often necessary to consider more distant matches and it may be important to consider shared matches with pertinent cousins who share less than the automated thresholds calculated by MyHeritage. Further, clusters generated through AutoClusters could represent several types of groupings. The members of a

---

[1] Esther, "Introducing AutoClusters for DNA Matches," *MyHeritage Blog,* 28 February 2019, https://blog.myheritage.com/2019/02/introducing-autoclusters-for-dna-matches/, accessed July 2020; and, Evert-Jan Bloom, "AutoCluster," https://www.geneticaffairs.com/features-autocluster.html, accessed May 2023

[2] Ran Snir, "Visualizing Ancestral Lines with DNA AutoClusters," https://familytreewebinars.com/webinar/visualizing-ancestral-lines-with-dna-autoclusters/, accessed May 2023.

[3] Paul Woodbury, "Connecting the Dots – Introduction to Auto Clusters at MyHeritageDNA," https://familytreewebinars.com/webinar/connecting-the-dots-introduction-to-auto-clusters-at-myheritagedna/, accessed May 2023.

cluster might be descended from the same ancestral couple, or they could all be members of the same endogamous population. In order to successfully resolve historic research questions, it is often necessary to perform more targeted analysis of shared relationships with known independent descendants of a research subject rather than relying on clusters generated through automated tools. Here, we explore one approach to researching your genetic match lists with "fishing" network approaches using MyHeritage tools. This same approach can be applied at other DNA testing companies. Also, this approach is only one among many ways to address a historic research question.

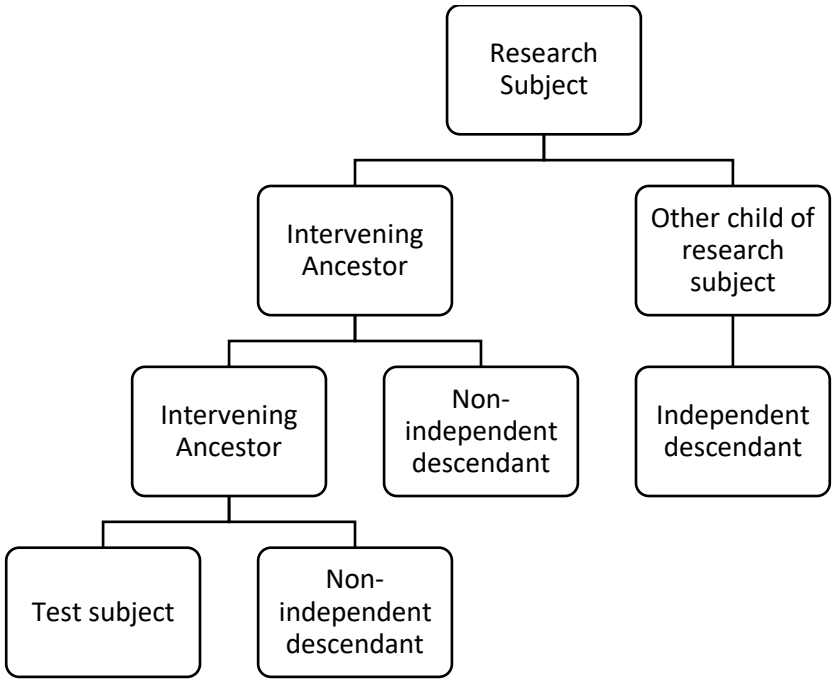
Fishing Networks Overview

1. Identify independent descendants (bait)
2. Identify shared matches to the independent descendants (catch)
3. Analyze shared matches and identify additional independent descendants (rebait)
4. Identify additional shared matches to all independent descendants (another catch)
5. Observe patterns of relationship with more distant shared matches to identify ancestors and relatives of a research subject.

What is an Independent Descendant

One key element of applying fishing clustering to a problem is first to identify independent descendants of a research subject. A research subject is the individual being researched and is often the earliest known ancestor on a test taker's ancestral line. The research question and objective is often centered around a genealogical relationship to the research subject.

Independent descendants are individuals who descend from a research subject through children other than intervening ancestors of the test subject.



Finding Independent Descendants of a Research Subject at MyHeritage

Some ways of finding an independent descendant of a research subject at MyHeritage include:

- Searching for the surname of the research subject in the match list.
- Applying the filter for "Has Theory of Family Relativity" or "Has Smart Matches"
- Searching for genetic cousins whose most recent common ancestor(s) are the research subject or the research subject and his wife.
- Verifying the proposed relationships between genetic cousins with Theories of Family Relativity or Smart Matches.
- Identifying the AutoCluster clusters with which those matches are associated.
- Building out the family trees of other members of the same cluster.

Who Are the Shared Genetic Cousins of a Known Independent Descendant?

Genetic genealogy is founded on the observation that when two people share DNA, they share ancestors as well. When three or more individuals are genetic cousins to each other, then they often (but not always) share at least one set of common ancestors. If you have several genetic cousins who are known relatives through a documented common ancestral couple (a research subject with unknown parentage and his wife) and several of those known relatives share a genetic cousin in common, there are several possibilities of how that genetic cousin might be related to the you and your known relatives.

1. They might also be a descendant of the research subject and his spouse in which case their shared matches in turn may or may not be pertinent to your question.

2. They might be a descendant of the research subject by another spouse in which case their shared matches in turn would likely be pertinent to your research question.

3. They might be related through a collateral relative of the spouse of the research subject in which case their shared matches would not be pertinent to your research question but would provide a useful filter group to identify other individuals not of interest.

4. They might be descended from a sibling or half-sibling of the research subject in which case their shared matches in turn would be pertinent to your research question.

5. They might be descended from a collateral relative of one of the research subject's parents in which case their shared matches in turn would be pertinent to your research question.

6. They might be related to one branch of descendants through an independent and unrelated ancestral line, and they might be related to another branch of descendants through another independent and unrelated ancestral line in which case their shared matches may not be pertinent to your research question.

After the family tree of the genetic cousin is analyzed, and after it is determined which category they fit in (if it can be determined), then the match and their more distant shared matches can be deemed either pertinent or not pertinent to the research question at hand. If pertinent, their shared matches might be added to a list of potentially pertinent genetic cousins. If not pertinent, they can be marked as such and their shared matches can be added to a list of non-pertinent genetic cousins.

How to Identify Potentially Pertinent Genetic Cousins at MyHeritage

- Review the match profile of an independent descendant of the research subject and assign them a label as an independent descendant.
- Sort shared matches based on amount of shared DNA with an independent descendant.
- Create a label for shared matches of independent descendants and assign all shared matches over a certain threshold of shared DNA with the match (20 cM is good, certainly higher than 10 cM) to the same group label. Assign matches who share more than the threshold amount with the genetic cousin even if they share very low amounts of DNA with the test subject.
- Return to the main match list and filter to genetic cousins with that group label.
- Analyze those matches to determine the nature of their relationship (Some may be additional independent descendants of the couple in which case their shared matches can

also be added to the same labeled group. Others may be collateral relatives of the research subject and their label might be changed after further exploration.

Filter Categories

In addition to identifying independent descendants of a research subject and shared matches with those independent descendants, it can sometimes be helpful to also label known relatives from other ancestral lines and shared matches with those individuals in order to clearly identify which genetic cousins are probably *not* pertinent to a research question.

Some Labels to Consider

- Independent descendant of the research subject
- Non-independent descendant of the research subject
- Shared matches of independent descendants of the research subject
- Confirmed collateral relatives of the spouse of the research subject
- Shared matches of collateral relatives of the spouse of the research subject
- Confirmed relatives from other ancestral lines
- Shared matches of confirmed relatives from other ancestral lines

A Note on Distant Matches

While true that the genetic cousins sharing less than about 10 cM with a test subject on a single segment of DNA are unreliable, the fact remains that if an individual (whether match to a test taker, unreliable match to a test taker or not even a match to a test taker) shares more than a particular threshold of DNA with at least two independent descendants of a research subject, they merit additional exploration. While they may not be a "true" match to one test taker (often the individual for whom a researcher has test result access), looking at a research question through the lens of a single test taker unnecessarily limits and narrows the body of potentially pertinent evidence available for exploration. Their shared relationship with members of the pertinent group at levels which are more reliable could help in identifying the parents or relatives of a research subject.

Conclusion

By considering network groups of autosomal DNA matches within the context of confirmed independent descendants of a research subject, as well as documented relatives from other ancestral lines, it is possible to narrow in on more distant pertinent matches who may be key in solving research questions in historic time periods.