GENETIC VOYAGE

geneticvoyage.com

IDENTIFYING AN UNKNOWN FATHER

The complete beginner's checklist for cluster-building and probability scoring when you have no direct match

Before you begin — quick glossary:

cM (centimorgan) — the unit for measuring shared DNA — see the reference chart below.

Cluster — a group of matches who all share DNA with each other, usually because they descend from the same ancestor or couple.

ICW ("in common with") — the list of people who match both you and a specific other match — the raw material clusters are built from.

Leeds Method / AutoCluster — techniques (manual or automated, e.g. via MyHeritage AutoClusters, Genetic Affairs, or DNAGedcom) that sort your matches into colour-coded clusters.

Triangulation — confirming that three or more people share the exact same DNA segment because they inherited it from the same ancestor.

WATO (What Are the Odds?) — a free tool at dnapainter.com/tools/wato that scores different hypotheses about how a mystery person fits into a known family tree.

BanyanDNA — a probabilistic genealogy tool (banyandna.com) that models multiple matches and relationships at once using Bayesian statistics, useful for comparing several candidate scenarios side by side.

This checklist is for the hardest version of the problem: you have DNA matches, but none of them is the father himself, or anyone who obviously *is* him. You're working entirely from cousins, using clustering and probability scoring to triangulate a family — then a person. Go through it in order; each phase builds on the one before it.

1 LAY THE DATA FOUNDATION

- ☐ Test with AncestryDNA if you haven't already — *largest database, most likely to surface useful cousins*
- ☐ Upload your raw DNA to MyHeritage, FTDNA, and GEDmatch — *each has a different user base and different tools — all three uploads are free*
- ☐ Export your match list into a spreadsheet — *columns: name, shared cM, segments, largest segment, tree link, notes*
- ☐ Sort matches by shared cM, descending, and scan for anyone above 90 cM

2 ACCEPT AND ISOLATE THE NO-MATCH REALITY

- ☐ Confirm you truly have no match closer than roughly 2nd–3rd cousin (i.e. nothing above ~200 cM) on the unknown side — *use the chart below to double check this cutoff against your actual matches*
- ☐ Separate out matches you can already explain via your known/maternal side — *compare against your mother's match list if she has tested — anything matching her is not part of this search*
- ☐ Build a working list of only the unexplained matches — ★ *this shortlist is what everything else in this checklist works from*
- ☐ Estimate the likely generation of the unknown father from your closest unexplained match's cM — *use the chart below — e.g. ~200–450 cM often points to a 2nd cousin, implying a shared great-great-grandparent*

Quick reference — shared cM to likely relationship

Shared cM	Most likely relationship(s)
3,330–3,720 cM	Parent / child, identical twin
2,300–3,400 cM	Full sibling
1,300–2,300 cM	Grandparent/grandchild, aunt/uncle, half-sibling, niece/nephew
575–1,330 cM	1st cousin, great-grandparent, great-aunt/uncle, half-aunt/uncle

200–850 cM	1st cousin once removed, half-1st cousin
90–450 cM	2nd cousin, 1st cousin twice removed
25–200 cM	2nd cousin once removed, 3rd cousin
0–50 cM	3rd cousin once removed, 4th cousin, or more distant

Ranges overlap significantly between relationship types — always cross-check with shared matches, ages, and tree research rather than the cM figure alone.

3 BUILD CLUSTERS FROM YOUR UNEXPLAINED MATCHES

☐ Run an AutoCluster tool on your full match list — *MyHeritage: DNA tab → "AutoClusters"; for Ancestry data, use the free Genetic Affairs or DNAGedcom tools instead, since Ancestry has no built-in clustering feature*

☐ Alternatively, build a manual Leeds Method grid by hand for matches in the 40–400 cM range — *full instructions in the Leeds Method Checklist*

☐ Identify which resulting cluster(s) sit entirely on the unexplained/paternal side

☐ Pull the ICW ("in common with") list for your top unexplained match to sanity-check cluster membership — *Ancestry: "Shared Matches" tab; MyHeritage: "Shared DNA Matches"; FTDNA: "In Common With" filter*

☐ Note the approximate cM range within each cluster — *a wide spread of cM within one cluster often signals more than one generation is mixed in — consider splitting it further*

4 TURN CLUSTERS INTO FAMILY TREES

☐ Pick the 3–5 strongest matches in your target cluster

☐ Build a "quick and dirty" descendant tree for each, working forward from their earliest known ancestor — *names, approximate birth years, and locations are enough at this stage*

☐ Look for a point where two or more of these trees intersect at a shared ancestral couple

☐ Treat that couple as your anchor ancestor(s) for the unknown line

☐ Extend descendant lines forward from the anchor couple toward the generation you estimated in Phase 2

5 SCORE HYPOTHESES WITH WATO

☐ Go to dnapainter.com/tools/wato and start a new tree — *free account required — sign up if you haven't already*

☐ Enter your anchor ancestor couple as the root of the tree

☐ Add each match as a tester, entering their shared cM to you accurately — *double-check each figure against your spreadsheet before entering it — a single wrong number skews every hypothesis score*

☐ Place your mystery father as a hypothetical person in several different plausible tree positions — *add one hypothesis per plausible branch/generation*

☐ Let WATO score each hypothesis and compare the results — *higher scores indicate a better statistical fit given the cM data — not certainty*

☐ Discard hypotheses WATO scores at or near zero — ★ *don't discard everything else based on the top score alone — a leading hypothesis still needs independent confirmation in Phase 7–8*

6 CROSS-CHECK WITH BANYANDNA

☐ Go to banyandna.com and create an account — *check current pricing — BanyanDNA has historically offered a free tier for smaller trees, with paid tiers for larger analyses*

☐ Enter the same match set and cM values you used in WATO

☐ Compare its top-ranked scenarios against your WATO results

☐ Pay attention to cases where the two tools disagree — *this usually means more matches or more tree-building is needed before concluding anything — treat disagreement as a signal to gather more data, not as a reason to pick whichever tool you prefer*

☐ Factor in age gaps and any known endogamy in the population — *both tools handle these differently — read each tool's documentation on how it treats these factors before trusting a borderline score*

7 NARROW TO REAL, NAMED CANDIDATES

☐ List every person in the anchor family who fits the estimated age and generation

☐ Cross-check each candidate against known facts — *your mother's location and timeframe around conception, family stories, any names ever mentioned*

☐ Rule out candidates who don't fit the timeline or location — ★ *ruling out is often faster and more reliable than trying to rule in — actively look for disqualifying facts, not just confirming ones*

☐ Identify whether any candidate, or a close relative of theirs, might be willing to test

8 CONFIRM BEFORE YOU CONCLUDE

☐ Require at least two independent lines of DNA evidence pointing to the same person

☐ Pursue a targeted confirmation test with the candidate or a close relative if at all possible

☐ Re-run WATO or BanyanDNA once new data arrives to see if the score strengthens or weakens

☐ Only mark the case "confirmed" once independent evidence lines converge — *use three labels: hypothesis → probable → confirmed, and be explicit in your notes about which one currently applies*

9 MOVE FORWARD THOUGHTFULLY

☐ Decide, on your own timeline, whether and how to make contact

☐ Consider a search angel or intermediary for the first approach if the situation feels sensitive — *DNAAdoption.org and the "Search Squad" Facebook group both connect people with free volunteer search angels*

☐ Keep your research log updated so the full evidence trail is documented, not just the conclusion

☐ Give yourself permission to pause at any phase — *this process can take weeks or months, and that's normal — see the NPE Research Checklist if the emotional side of this needs its own attention*

Pro tip: Tools like WATO and BanyanDNA score hypotheses — they don't hand you an answer. Treat a high score as "worth investigating further," not as proof, and keep building your trees and gathering matches even after one hypothesis pulls ahead.