

ENDOAMY & PEDIGREE COLLAPSE CHECKLIST

A systematic approach when your matches don't fit the normal rules

Before you begin — quick glossary:

Endogamy — a long history of marriage within a small, relatively closed community — common in isolated regions, some religious communities, and many other tight-knit populations worldwide.

Pedigree collapse — when the same ancestor appears more than once in your family tree, because relatives married each other at some point in the past. Endogamy over many generations causes heavy pedigree collapse.

Multiple relationships — when you and a match are related through more than one ancestral line at once, common in endogamous populations — this inflates your shared cM well beyond what a single relationship would produce.

Endogamy affects a significant share of researchers worldwide — Scandinavian, Jewish, Acadian, and many other tight-knit populations among them. It doesn't break DNA genealogy, but it does mean every standard rule of thumb needs a second look.

Phase 1 — Recognize the signs

- ☐ Notice an unusually high overall match count relative to your total number of tested cousins — *endogamous populations produce far more matches per person than expected, because everyone is distantly related to everyone else through multiple lines*
- ☐ Watch for matches sharing more cM than their documented relationship should produce — *e.g. a documented 3rd cousin sharing 2nd-cousin-level cM — this is a classic sign of multiple relationships, not a documentation error*
- ☐ Note any family or community history of intermarriage — *an isolated village, a small religious or ethnic community, or a region with historically limited outside migration are all common settings for endogamy*
- ☐ Check whether your Leeds Method clusters refuse to separate cleanly — *★ if clusters keep merging into one giant group instead of forming four distinct colours, endogamy is the most likely cause — see the Leeds Method Checklist*

Phase 2 — Adjust your expectations for cM interpretation

- ☐ Understand that standard cM charts assume no endogamy — *in an endogamous population, a match's TRUE relationship is very often more distant than the standard chart suggests for their cM total*
- ☐ Treat any single cM-based relationship estimate as a starting hypothesis, not a conclusion — *cross-check every estimate against shared matches and paper-trail research before trusting it*
- ☐ Use endogamy-aware tools where available — *some newer scoring tools include an explicit endogamy adjustment setting — check each tool's documentation before assuming its default output accounts for this*

Phase 3 — Adjust your clustering approach

- ☐ Expect more than four colour groups, or heavily overlapping ones, when running the Leeds Method — *this is normal in endogamous data and doesn't mean the method has failed*
- ☐ Re-run clustering using a higher cM floor, such as 90 cM instead of 40 cM — *this filters out the most distantly-related "noise" matches that are hardest to place correctly in endogamous data*
- ☐ Expect to need several passes at different cM thresholds rather than one clean result — *run the grid at 40 cM, then separately at 90 cM and 150 cM, and compare which grouping gives the clearest picture*

Phase 4 — Adjust your triangulation approach

- ☐ Require a larger minimum segment size than the usual 7 cM floor — *small segments are far more likely to match by chance (identical by state) in endogamous populations — consider raising your working minimum to 10–15 cM*
- ☐ Be skeptical of any single shared segment used to claim a specific ancestor — *★ in endogamous data, require multiple independent triangulated segments, ideally on different chromosomes, before treating an ancestor assignment as confirmed*

- ☐ Document every relationship you find between two matches, not just their relationship to you — *in heavily collapsed pedigrees, understanding how your matches relate to EACH OTHER is often more revealing than how each relates to you alone*

Phase 5 — Build a fuller picture rather than working line by line

- ☐ Research multiple ancestral lines simultaneously rather than one at a time — *in endogamous trees, progress on one line very often depends on information from a seemingly unrelated line*
- ☐ Actively look for "multiple relationship" cases among your closest matches — *a match related to you through more than one line will show inflated cM — once you spot this pattern, it usually explains several other confusing matches in the same cluster too*
- ☐ Keep detailed notes on every ancestor who appears more than once in your tree — *this is your working record of pedigree collapse — it will keep growing as you research further back*

Phase 6 — Tools and resources

- ☐ Check whether WATO or BanyanDNA offers an endogamy-adjusted mode before relying on their scores — *read each tool's current documentation, since default scoring in both tools can significantly misjudge relationships in heavily endogamous data*
- ☐ Look for genealogy resources specific to your community — *many long-established endogamous populations have dedicated genealogical societies, record projects, or DNA study groups with specialised expertise in exactly this problem*

Pro tip: In heavily endogamous populations, cM totals can overestimate closeness by a full relationship category or more. Always corroborate a DNA-based relationship estimate with paper-trail research before treating it as confirmed — the DNA alone is not enough to trust on its own here.