SMALL SEGMENT ANALYSIS - JULIAN LAND'S PROCEDURE

After Andrew Millard partially corrected my lack of knowledge of reproductive biology in May 2025, my only way forward was to complete the analysis of all 1963 3cM segments generated from my 12-relative set by GEDmatch (at P = 3). From this my procedure for using small segments to prove family branch connections can now be documented. The following steps will help others do the same thing.

STEP 1

Learn the key - we are looking for rare segment boundary coincidences (RSBCs). These occur when 2 independent pairs of relatives have a coincident boundary (left or right). I found 46 of them - 25 on the left boundary, 21 on the right.

Because the number of these occurrences greatly exceeds the number expected at random (see **Appendix 1**), each RSBC is significant and thus conveys some genetic information. If pair A&B forms a RSBC with pair C&D, then either A matches C or D, or B matches C or D. Of course, A&B and C&D might also be genuine matches. Better genetic information can be obtained as follows.

STEP 2

Put all matches in order, using the left boundary, with each chromosome on its own (Excel) worksheet. Find each RSBC and colour it the same colour whichever chromosome it sits in.

STEP 3

Inspect the matches around each RSBC, looking for evidence that the RSBC will yield more genetic information. In my case I found that 5 RSBCs could not be improved leaving 4 possible interpretations, as explained above. This was after – sometimes after an iteration through the entire set of connections

proven to that point – I found that one of the 4 possible interpretations had already been proven in which case no conclusion was possible, even if one of the 2 matches had been proven. So just 5 of the 1 in 4 situations remained.

STEP 4

Some of the remaining RSBCs show just 2 possibilities eg A matches C or D. This can be useful where C and D have a known genetic or genealogical link. I found one such case.

But much more progress can be made with the RSBCs which are amenable to logical deduction.

STEP 5

Here strict logic is required to reject any match around a RBSC which could be a false positive, whilst using any evidence to the contrary, usually in the form of linked 3-person Segment Boundary Coincidences (3pSBCs).

In my case, I was able to prove 21 genetic matches, quite enough to confirm connection between my family branches.

STEP 6

Display the results as in **Diagram 1,** so that the level of family connection between the 12 (in my case) relatives can be seen. The Diagram shows 11 are related via 21 proven family links. Missing are 45 links for there are 66 possible pairings. **Diagram 1** can also show the Step 3 and Step 4 cases.

Most pairs can be linked via 2 connections. Some relatives play a key role with more connections than others – Je, TP and RF have 6 connections each and Ju, SW and RM have 5. Those with the fewest connections are A, E and J having 1 connection each and MD none.

Diagram 2 shows how the uncertain connections can be resolved through genealogical knowledge (within a family branch). **Diagram 3** shows the known matches in a branch not picked up by our method and **Diagram 4** adds the 10 matches proven by Qmatch set at 7cM and P=7.

CONCLUDING COMMENT

It is certainly true that small segments are a challenge for those who have to work with them. We describe a method of dealing with this challenge, starting with the set of Rare Segment Boundary Coincidences (RSBCs) which occurs amid any large set of small 3cM matches generated from a set of relatives.

Previous work indicated that a sample of relatives generated 5% more 3cM matches than the same-sized random sample. Here our 46 RSBCs comprise 2-3% of our 1963 matches. The number of proven family matches was found to be about 1% of 1963.

Here, our procedure finds 21 matches, 7 of which were known through standard commercial DNA testing so our procedure misses 3 of these matches. Our procedure also misses 7 intra-branch matches known through genealogy. However, our procedure complements both genealogy and known autosomal matching, thus amply confirming inter-branch connection.

Julian Land

16 October 2025

APPENDIX 1 - PROBABILITY

The total number of coincidences expected at random involving 4 people from 12 in a set of N matches over each of 22 chromosomes would be 22*N*(N-1)100,000 = 1.7. We found 25 RSBCs (left) and 21 RSBCs (right). The 3-person Segment Boundary Coincidences (SBCs) occur more frequently. We suggest the excess incidence in both categories reflects the impact of family.

In this estimate of random occurrences, there were about (N=) 89 segment matches found per chromosome at the settings used (Qmatch 3cM P=3), and there are about 50,000 SNPs per chromosome.

The 3-person version involves one person occurring twice but not in the same match. The number of 4 person matches from M is:

$$M*(M - 1)*(M - 2)*(M - 3)/24 = 495 \text{ when } M = 12$$

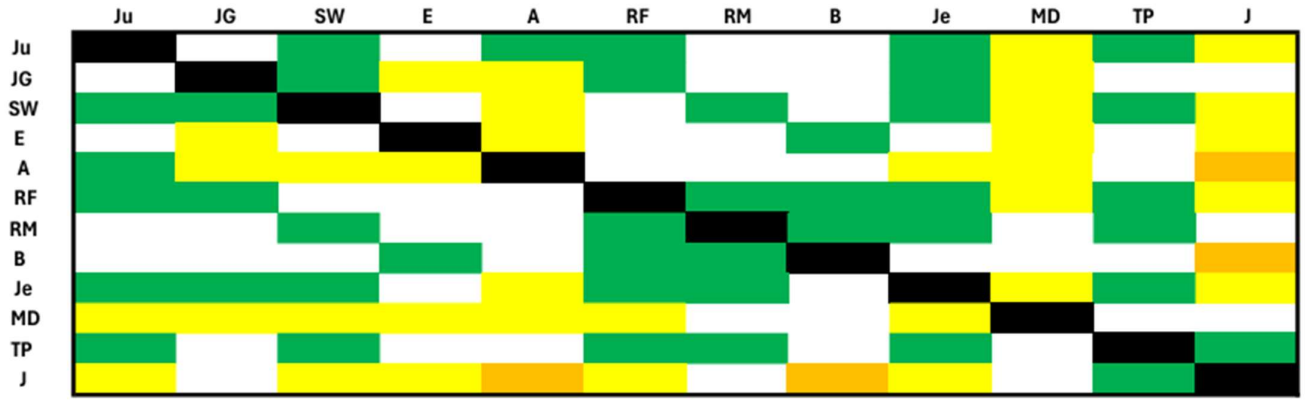
The number of 3-person matches from M is:

$$2/3*3*M*(M - 1)*(M - 2)/6 = 440 \text{ when } M = 12$$

Of course, this does not mean that the number of 3-person SBCs should be smaller than the number of RSBCs. In fact it is much larger, due to the fact that the same person across a pair of matches brings in that person's genetic history. The large number of SBCs thus contains family data but of course not all of it from the target family. One's initial temptation to dismiss 3pSBCs should be resisted, because they play a role in distilling the connections lying in the RSBCs.

DIAGRAM 1

Here we show the family matches generated by 46 RSBCs.

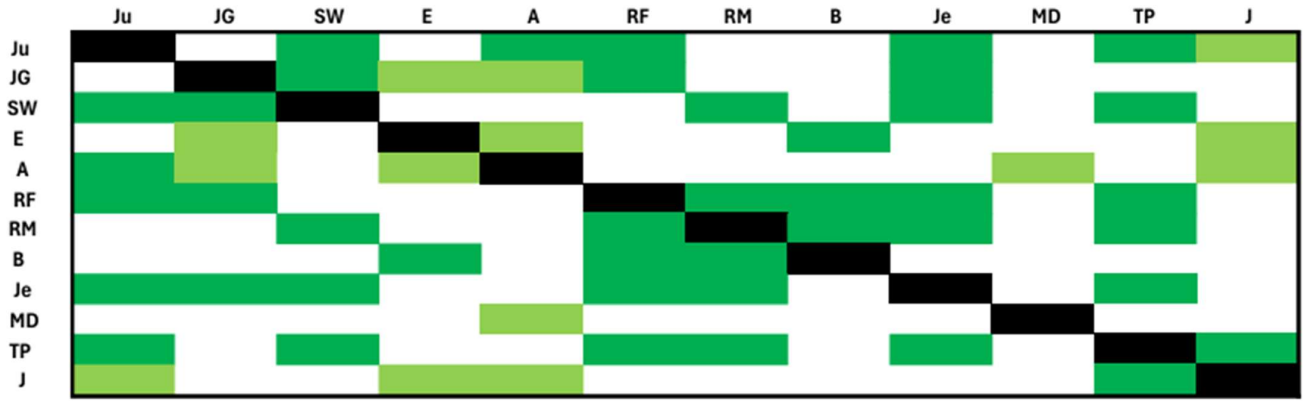


The 21 proven matches are shown in green. J matches A or B as shown in orange, each case having an a priori probability of 50%. In 5 cases we were left with 4 possibilities (shown in yellow) - but with 3 of the 20 possibilities duplicating themselves, we are left with 17 possible connections with a probability of 25%.

Those uncertain matches can be resolved through genealogy as shown in Diagram 2.

DIAGRAM 2

Here, all the matches shown as uncertain in Diagram 1 can be upgraded due to genealogical certainty. With one (or sometimes 2) of each set of possibilities confirmed, the others fall away, hence no yellow or orange entries.

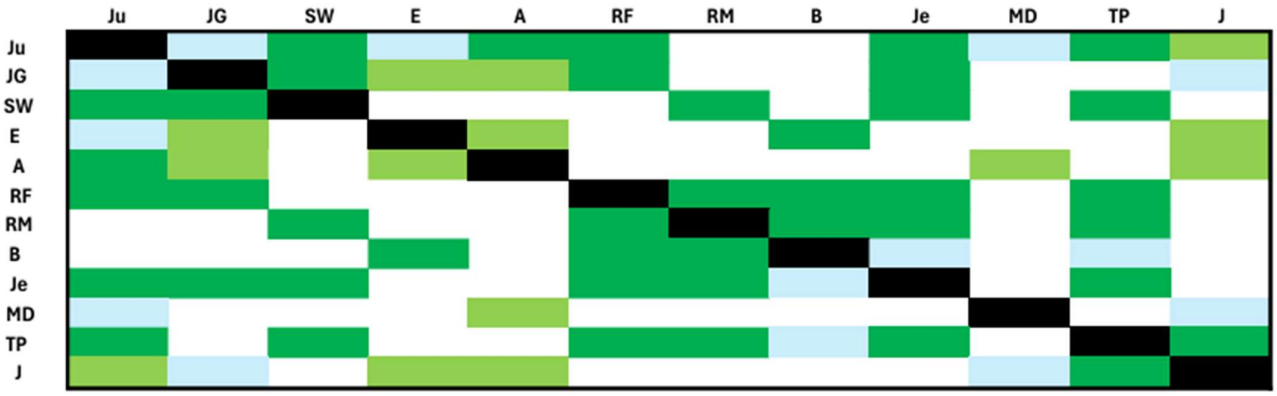


The lighter shade of green shows where we can upgrade the certainty of the match due to genealogy.

Some of the matches not found with our small match procedure are in fact known through genealogy. These are shown in Diagram 3.

DIAGRAM 3

Here we can show where matches not found by our small match procedure are in fact known.

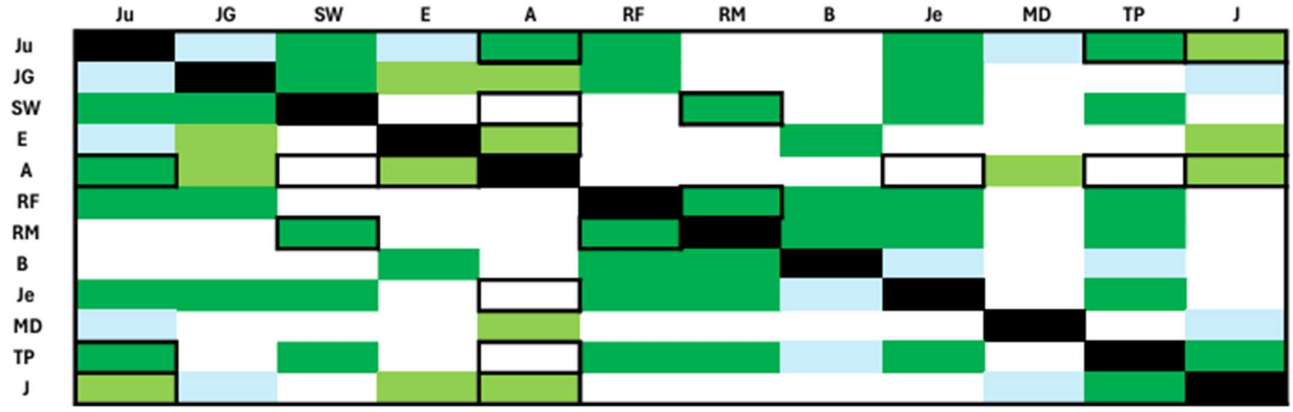


The blue colour marks those 7 connections which are known by genealogy alone, and not picked up by our small match procedure.

The 31 white entries perhaps indicate that we are near the limits of technology, for there are 66 possible pairings. We may have been somewhat fortunate to have found 6 of our 12 relatives having cross-branch proven connections.

DIAGRAM 4

Here we can show where 10 matches known through commercial autosomal testing are located. These matches were found by Qmatch set at 7cm and P=7. These conditions are strict enough to disallow B's 7cM matches with Je and TP found by Ancestry.com.



It can be seen that 3 of these matches – A with SW, Je and TP – are not picked up by our small match procedure but 7 are. Actually, A/SW and A/Je did appear in 1 in 4 situations but were eliminated because the relevant RSBC was deemed certain through 1 of its 4 possibilities being explained by genealogy. Thus, only A/TP was totally missed, an omission covered by the Ju/TP match, Ju and A being 1st cousins.

Of course, those 10 commercial matches already strongly suggest a Barrow Lousada connection with the Fischls and the New England Lousadas, a suggestion now validated by our small match procedure.