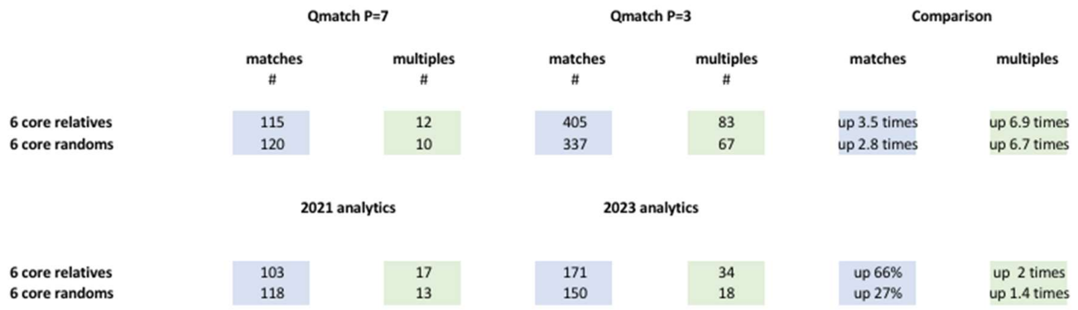
HOW Q-MATCH COMPARES WITH 2021 AND 2013 GEDMATCH ANALYTICS IN THE CONTEXT OF 3CM MATCHES

In our 8 by 8 study, we noted that between 2021 and 2023, the default GEDmatch analytics for autosomal matching at the 3cM threshold changed significantly. GEDmatch then suggested to us that its Q-match technique should be used for small matches. Here we report our experience of using Q-match for small matches.

Q-match has a parameter P (for 'Precision') which can be adjusted downward to get more matches, albeit of lower quality, than the default setting P = 7 provides. Accordingly, for purposes of comparison, we used the same core 6 relatives and randoms we identified in our 8 by 8 study, which is where the lower half of the following table came from. In its totality the table illustrates what we found thus allowing a comparison between Q-match and its precursors.



	Qmatch P=7		Qmatch P=3		Comparison	
	matches \#	multiples \#	matches \#	multiples \#	matches	multiples
6 core relatives	115	12	405	83	up 3.5 times	up 6.9 times
6 core randoms	120	10	337	67	up 2.8 times	up 6.7 times
	2021 analytics		2023 analytics			
6 core relatives	103	17	171	34	up 66%	up 2 times
6 core randoms	118	13	150	18	up 27%	up 1.4 times

The first pair of columns shows 2021 analytics beneath P = 7 results. We can see that while Q-match selects matches better in that there are more matches but fewer multiples – most of the multiples are still inconclusive with just 2 multiples - the Cr2 and Cr8 indicator segments - being the difference between the family and the random sets. The inconclusive multiples are mainly weak and strong triples for both relatives and randoms. However, this result still clashes with John Griffiths' intuition in that the family set continues to yield fewer matches than the random set.

Selecting lower Precision rectifies this, for in the second pair of columns, which show 2023 analytics beneath P = 3 results, we see that Q-match produces many more matches with the family set ahead in both matches and multiples. But here, the large number of multiples presents a challenge in sorting out the useful from the inconclusive. In fact, at P = 3 we produce so many matches that the number of multiples is constrained by their coverage of a large fraction of the genome – the third pair of columns shows this slowdown in rate of increase of frequency.

Our method for doing this sorting is to look for rare coincident segment boundaries (RCSBs). For a time we considered that abutting segment boundaries (ASBs) might also be useful. We no longer do, but leave them in the present discussion. During the present study we considered afresh the frequency at which RCSBs and ASBs occur at random. These events occur when a terminal SNP of one match coincides with a terminal SNP of another match.

On average there are 50,000 SNPs per chromosome. If there are N matches on a chromosome, there are N*(N-1)/2 pairs of matches. Thus, there are N*(N-1)/100,000 possible instances of coincident segment boundary SNPs on a chromosome. Using the observed average number of matches on each chromosome, we find the estimated frequency of random occurrence of boundary coincidences to be:

Group	Matches on a chromosome	Frequency on a chromosome	Frequency on the genome
6 core randoms	15.3	0.0022	0.049
6 core relatives	18.4	0.0032	0.070
12 relatives	89.2	0.0786	1.730

In this study we found 1 of the rhRSBCs and 2 of the 3-person ASBs in both the random set and the 6-relative set. But the relatives gave us 24 of the 3-person SBCs (13 lefthand and 11 righthand) while the randoms gave us 10 (6 lefthand and 4 righthand). Our initial interpretation here is that 3-person RBCs should not be dismissed, as here they indicate a clear difference between relatives and randoms. Indeed, the randoms appear to be showing a family relationship or two (previously unnoticed within the random group) in which case the difference between relatives and randoms is somewhat greater than implied by the 24 to 10 ratio just cited.

Perhaps our most important conclusion is just this - that 3-person SBCs should not be ignored. Because the great majority of SBCs are 3-person SBCs and not our 4-person RSBCs, the above probabilities – when applied to 3-person SBCs – have much greater significance in practice. In addition, the number of SBCs – when the 3-person version is included – greatly exceeds the number of ASBs – which category already includes the rarer 4-person variety of ASB as well the more common 3-person variety. The number of 4-person ASBs is only twice random. It may be that by re-examining the significance of the 3-person SBCs, we can dispense entirely with the counter-intuitive ASBs. Indeed we did.