Journal of Genetic Genealogy

Journal: www.jogg.info
Originally Published: Volume 12, Number 1 (Spring 2024)
Reference Number: 121.002

RECOMBINATION AND PHASING FOR A GROUP OF THREE OR FOUR (OR MORE) SIBLINGS--TWO PRACTICAL APPROACHES

Author(s): T. Whit Athey, PhD and Kathryn Johnston, MD

© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

Journal of Genetic Genealogy

RECOMBINATION AND PHASING FOR A GROUP OF THREE OR FOUR (OR MORE) SIBLINGS--TWO PRACTICAL APPROACHES

By T. Whit Athey, PhD and Kathryn Johnston, MD

Abstract

A method is presented for determining the recombination patterns and phasing in a group of siblings. Two versions of the same basic method are described, one more numerical in nature and the other more visual or graphic. The method requires at least three siblings, preferably four or more, each with an autosomal SNP dataset available, such as those from companies such as Family Tree DNA, 23andMe, or Ancestry. The approach requires for each chromosome, the complete set of matching segments for each sibling pair--the endpoints of the segments along with the type of match (identical on one chromosome of a pair, or identical on both chromosomes of a pair, designated as a "single" or "double" match), plus matching segments for confirmed paternal and maternal second or third cousins. Some examples of recombination in real families will be described, and some general observations on recombination will be summarized.

Introduction

When a parent passes a set of 22 autosomal chromosomes to a child via a sperm cell or egg cell, each parent derives each chromosome from cutting and pasting together an amalgam of the two corresponding parental chromosomes. This process is known as *recombination*, and it brings with it almost unlimited possibilities for producing new and unique chromosomes. The points on the chromosome passed to the child where the DNA has been cut and pasted from the parental chromosome pair are called *crossovers*.

The term "recombination" technically applies to just one point on the chromosome where a crossover has occurred. In this article we also use the term to apply to the collection of crossovers that has occurred in a group of siblings. The term "phasing" technically applies to the separation of the bases of a chromosome into its paternal and maternal copies. In this article we also use the term to apply to the schematic separation of the regions of a chromosome pair into the constituent regions inherited from each grandparent, without regard to the underlying sequence of bases. Thus, we will use the term "visual phasing" to apply to the visual/graphical approach to phasing since that term has already been in general usage.

Determining just how recombination has occurred in a family group of siblings is difficult, but not impossible if the family group is sufficiently large--at least three siblings, preferably four or more, each with an autosomal SNP dataset available, such as those from the companies Family Tree DNA, 23andMe, or Ancestry. The same basic method can be implemented in two versions or approaches, one more numerical or computational in nature [Athey, 2010a] and the other more visual or graphic.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

The approach may yield more than one solution for each chromosome pair, and it is important to have autosomal match data from known cousins to eliminate all but the one correct solution. Having in addition, the data from one parent can also be very helpful, but in this case it would be possible to completely phase the data, and this would result in determining the recombination patterns as a by-product of the phasing analysis [Athey, 2010b]. Therefore, the present approach will normally be most useful when compatible autosomal SNP datasets on three or more siblings are available, but no dataset from a parent is available. Our present approach also avoids the need to examine or use the raw data directly.

Intuitively, we normally expect that recombination will produce two new sets of chromosomes in a child such that all four grandparents contribute DNA approximately equally. However, this is not necessarily the case, even when averaged over all chromosomes. When considering a single chromosome pair, it is even possible, though unusual, to see only two of the four grandparents contributing DNA.

A few principles are very important in regard to the analysis of recombination in a particular chromosome pair in a group of siblings:

1. In any particular region of a chromosome pair, two siblings can (a) match each other exactly on both chromosomes of the pair--said to be "fully identical," (b) match each other on just one of the pair of chromosomes--said to be "half identical", or (c) not match on either one of the pair.

2. Each crossover point in a family group is unique to one sibling and unique to just one of the chromosomes of the pair. We will say that this sibling "owns" that crossover. In practice, some of the crossover points in different siblings may be fairly close to the same location, but it is assumed in the present approach that crossover locations can be determined and are all different.

3. As our focus moves along a chromosome and passes over a crossover, the nature of the matches between the sibling who owns the crossover and the other siblings will change, while the nature of the matches between the siblings not owning the crossover will remain the same.

In the first part of the present article, we will first present the numerical approach to the analysis, while the second part will present the visual or graphical approach. For those who may be averse to numerical methods, it may be preferable to skip directly to the visual approach in the second section, because the visual approach is perhaps more intuitive and understandable. The disadvantage of only considering the visual approach is that there is a particular difficulty that often occurs in the analysis that prevents a unique solution from being possible. This difficulty appears random and unexplainable if only the visual approach is used, but the basis of this difficulty is readily explained in the numerical approach. Therefore, we have elected to present the numerical approach first.

Numerical Method--Case of Four Siblings

In Figure 1 consider an example of recombination in a group of four siblings where the recombination patterns, i.e. the inheritance pattern from the four grandparents, have already been determined. We begin this way, with the “solution” already in hand, in order to define several terms. The chromosomes shaded light blue and pink came from the paternal grandfather and paternal grandmother respectively (through the father), while the dark blue and the

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

green chromosomes came from the maternal grandfather and the maternal grandmother (through the mother). In Figure 1 the vertical lines represent the boundaries of two sliding windows--not crossover points--but in subsequent figures, vertical lines *will* denote crossovers. In the region just to the left of the crossover at 27 million, the types of matches between the pairs of siblings are shown at the bottom of Figure 1. After our sliding window passes the crossover at 27 million (located on the maternal side of Sib2), the types of matches in the window change, but only in the three match types involving Sib2, who "owns" that crossover. The match types between Sib1 and Sib3, Sib1 and Sib4, and Sib3 and Sib4 remain the same as single, double, and none respectively.

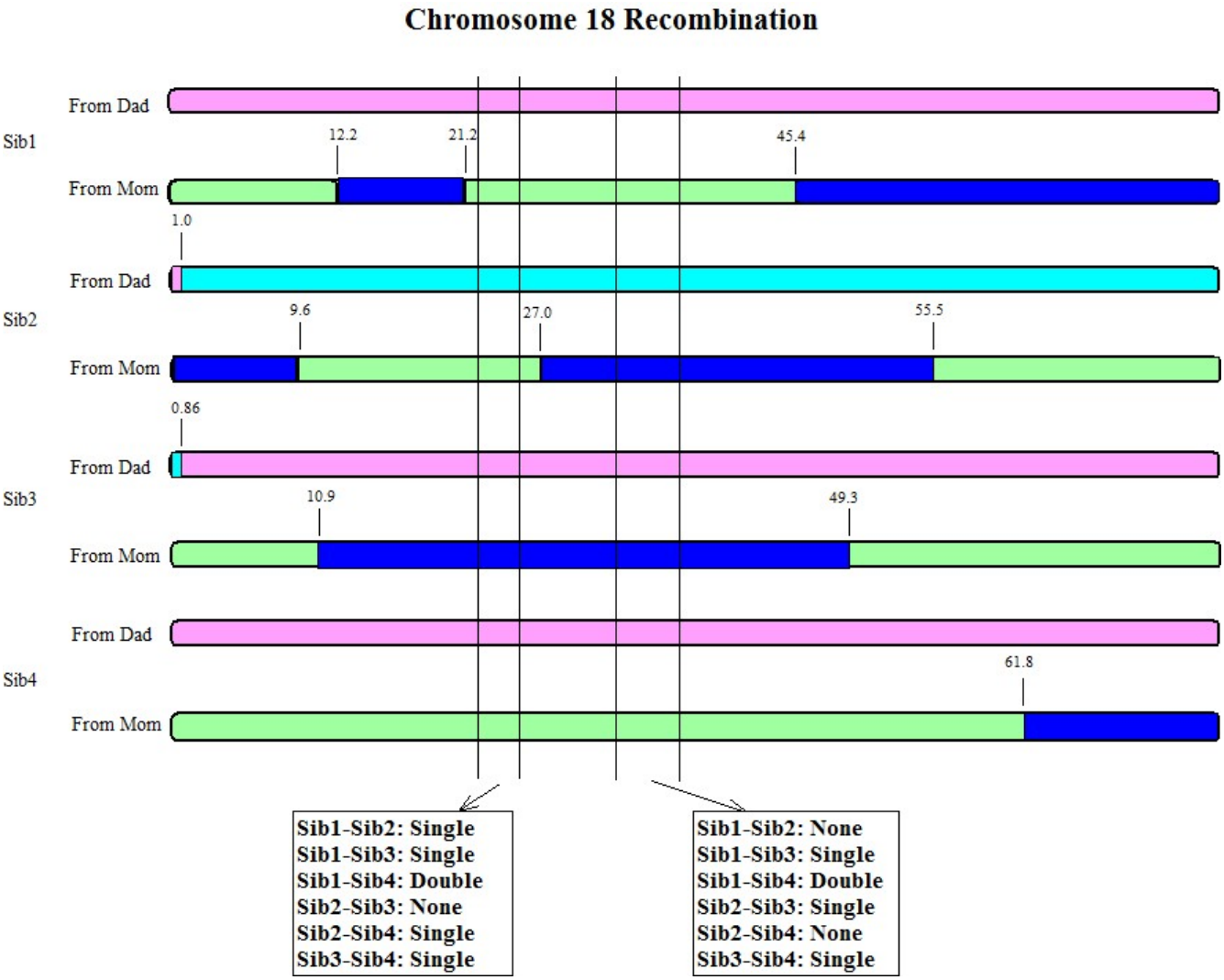


Figure 1. An Example of Four Siblings Where the DNA Inheritance is Already Known

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

In Figure 1 we also see that sometimes, especially in the small chromosomes such as 18, a father may pass along a chromosome whole with no recombination. Rarely, the mother may do the same. In general, more recombination will be found in the chromosomes passed down by the mother than those passed by the father, and this phenomenon may sometimes be useful in distinguishing the paternal and maternal chromosomes when no other information is available. Note also in the above example, there is nothing in the matching types alone that would indicate that the crossover at 27 million occurred on the maternal chromosome 18 of Sib2. We have independently confirmed through cousin matches that it did not occur on the paternal side.

In Figure 2 we see for the same example, the match types determined for every region of chromosome 18, and in this figure the vertical lines do represent crossover points. For simplicity in presenting this example, we consider the origin of our analysis to be just to the right of the crossover at 1 million. We ignore for the present the beginning of the chromosome where two crossovers lie very close together and begin our analysis at 2 million. We represent the match types with numerals: 2 means a double match (fully identical), 1 means a single match (half-identical), and 0 means no match. Note that the set of matching codes does not change in the region between crossover points (indicated by the vertical lines).

We will discuss two types of *patterns* in this article, and the patterns of codes shown at the bottom of Figure 2 will be called "segment matching patterns." Again, note that on either side of a crossover, only the code involving the sibling to whom that crossover belongs, will be different--the other three codes must remain the same.

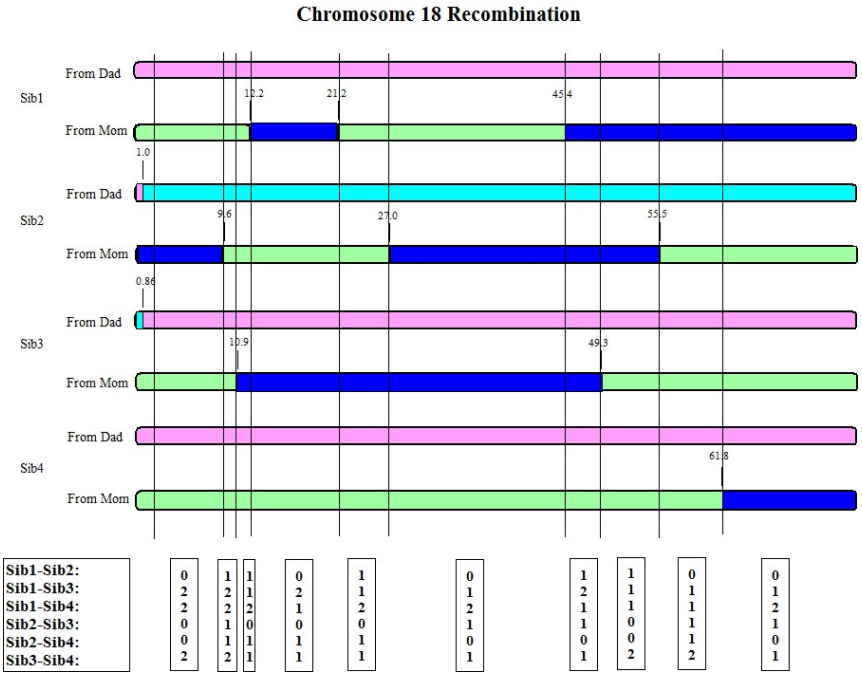


Figure 2. Matching Codes Shown for All the Regions of the Chromosome (except the beginning).

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

A second kind of *pattern* used in this presentation will be called "chromosome inheritance patterns." For example, in a particular region between crossovers, if Sib1 and Sib3 received DNA from one paternal grandparent and Sib2 and Sib4 received their DNA from the other paternal grandparent, then the chromosome inheritance pattern for the four siblings would be designated as "ABAB". The A's may sometimes represent DNA from the grandfather and sometimes from the grandmother--we will not know which it is before the analysis is complete. The chromosome inheritance patterns for the paternal and maternal sides will usually be different. Note that the pattern BABA is the same as ABAB, but we will, by convention, require that each of these patterns begin with an "A", and we take the complement of the pattern, if necessary, in order to begin it with an "A". For four siblings there are eight possible patterns for the four siblings' paternal chromosomes and eight for the four siblings' maternal chromosomes. These eight patterns are AAAA, AAAB, AABA, AABB, ABAA, ABAB, ABBA, and ABBB. In the actual case for the example above, in every segment the paternal chromosome inheritance pattern would be the same (ABAA), because Sib1, Sib3, and Sib4 received all of their paternal DNA from the pink chromosome of the paternal grandmother, whereas Sib2 received DNA from the light blue chromosome, and there were no crossovers anywhere on the paternal side (at least in the region beyond 0.2 million). The eight possible chromosome inheritance patterns (for the four-sibling case) are collected in Table 1.

Table 1. The Eight Chromosome Inheritance Patterns for Four Siblings

Pattern Number	Chromosome Inheritance Pattern			
	Sib1	Sib2	Sib3	Sib4
1	A	A	A	A
2	A	A	A	B
3	A	A	B	A
4	A	A	B	B
5	A	B	A	A
6	A	B	A	B
7	A	B	B	A
8	A	B	B	B

Note that there will be a chromosome inheritance pattern for the four siblings' four paternal chromosomes, and another, generally different, for the four siblings' maternal chromosomes. For example, in Figure 2 in the region between 27 million and 45.4 million, the paternal chromosome inheritance pattern, as already noted, is ABAA, while the maternal pattern is ABBA. In the latter case of the maternal chromosomes, the A represents the green chromosome and the B represents the dark blue chromosome, at least in this one segment. In general, for a group of siblings where the recombination has not yet been determined, we will not know which chromosome the A or B represents (paternal or maternal), and we may switch them anyway if we always start a pattern with an A. And, while we will know which sibling owns each crossover, we will not know whether the crossover has occurred on his/her paternal or maternal chromosome.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

It is important to note that there is a relationship between the two kinds of patterns that we have introduced. For each pair of *chromosome inheritance patterns*, regardless of which pattern is from the paternal side and which is from the maternal side, the pair taken together has one and only one corresponding *segment matching pattern*. However, looked at from the point of view of what we would actually know in the beginning from the sibling match comparisons, for each segment matching pattern, we would have just one pair of possible chromosome inheritance patterns, but we wouldn't know which chromosome matching pattern was from the paternal or maternal side. It is the task of this approach to resolve this ambiguity, and if successful, the chromosomes will be resolved into their component grandparent contributions. See Table S1 and Table S2 for a listing of all of the possible segment matching patterns for four siblings, plus the corresponding pair of chromosome inheritance patterns. Due to the length of these tables, they appear at the end of the article.

Table 2. The Two Types of Patterns and Their Relationship

Paternal Chromosome Inheritance Pattern				Maternal Chromosome Inheritance Pattern				Corresponding Segment Matching Pattern					
Sib1	Sib2	Sib3	Sib4	Sib1	Sib2	Sib3	Sib4	Sib1-Sib2	Sib1-Sib3	Sib1-Sib4	Sib2-Sib3	Sib2-Sib4	Sib3 Sib4
A	B	A	A	A	B **X**	A	A	0	2	2	0	0	2
A	B	A	A	A	A	A **X**	A	1	2	2	1	1	2
A	B	A	A	A **X**	A	B	A	1	1	2	0	1	1
A	B	A	A	A **X**	B	A	B	0	2	1	0	1	1
A	B	A	A	A	A **X**	B	A	1	1	2	0	1	1
A	B	A	A	A **X**	B	B	A	0	1	2	1	0	1
A	B	A	A	A	A	A **X**	B	1	2	1	1	0	1
A	B	A	A	A	A **X**	B	B	1	1	1	0	0	2
A	B	A	A	A	B	B	B **X**	0	1	1	1	1	2
A	B	A	A	A	B	B	A	0	1	2	1	0	1

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

We will be able to determine the segment matching pattern in each segment by using the chromosome browser tools at the company originating the data, and this set of segment matching patterns becomes the input for our analysis. We can then look up each pattern in Table S1 to find the unique pair of chromosome inheritance patterns. This will be explained in the following development.

In Table 2 we illustrate the relationship between the two types of patterns using the actual patterns from Figure 2. Note that the chromosome inheritance patterns are actually what we would be trying to determine in carrying out a recombination analysis, while the match types would be derived from the empirical segment matching information--the former represents the solution to our problem, while the latter is the input to the problem. Again, the important characteristic of the two kinds of patterns is that there is almost a one-to-one unique correspondence between them. For any given pair of chromosome inheritance patterns, there is just one segment matching pattern, but the two chromosome inheritance patterns may be switched between the paternal and maternal sides without affecting the segment matching pattern. Our task will be to determine the chromosome inheritance patterns from the segment matching patterns, while finding a way to resolve the ambiguity.

In Table 2 an "X" has been placed between the rows of patterns to indicate in which sibling that the crossover has occurred. That is, the "X" shows which sibling "owns" that crossover. In the line following the "X", note that only the segment match type (i.e., double, single, or none) involving that particular sibling will have changed. For example, in the last line of the table, since the crossover starting the last segment occurred in Sib4, only the third, fifth, and sixth match types have changed because those three positions correspond to the Sib1-Sib4, Sib2-Sib4, and Sib3-Sib4 comparisons. This example has crossovers only on the maternal side, but in general we would not know on which side a crossover has occurred.

A second important principle for the analysis is that in passing over a crossover, only one of the pair of chromosome inheritance patterns will change, either the one for the paternal side or the one for the maternal side. Also, only one of the letters in the chromosome inheritance pattern will change, as can be seen from the first to the second segment of the example (Table 2). However, in the case of the transition from the third to the fourth segments on the maternal side, the pattern AABA changes to BABA, but our convention requires that the pattern begin with an "A". Therefore, we write the pattern for the fourth segment as ABAB.

The chromosome inheritance patterns can only change from one to the next at a crossover in a restricted manner because of the constraints discussed in the last paragraph. Table 3 shows the permitted and prohibited transitions for chromosome inheritance patterns, with a "Y" in the right-hand part of the table indicating "permitted" and an "N" indicating prohibited.

Table 3 shows, for example, that a transition from Pattern 1 (AAAA) to Pattern 4 (AABB) is not permitted since it involves two changes.

Our starting point in a recombination analysis is to completely determine the matching segments between each pair of siblings. In the example above, we would obtain graphically, the segments shown in Table 4, with the blue color bar under each segment indicating a match, either half-identical match or a fully identical match, with the green color above the bar indicating fully identical matches. This is an example of the graphical output from the

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

comparison tool at GEDmatch, but similar tools are available at Family Tree DNA and 23andMe.

For our computational approach to recombination analysis, we need to have the endpoints of each of the matching segments, which are also the crossovers. These can be obtained, for example by using the utilities of David Pike, or by using GEDmatch. For the four siblings in our example, these segments are shown in Table 5. In practice, the segment boundaries obtained by any matching algorithm will be only approximate, and the same crossover point in two different comparisons may appear to be slightly different. In Table 5 the segment boundaries have been harmonized so that the same physical crossover is assigned the same location in all three segments affected by the crossover. These boundaries do not need to be precisely known, but the locations assigned must be consistent and properly ordered.

In Table 5, for a segment where there is a double match, it is considered by default also to be a single match. In the chromosome diagram from GEDmatch, the fully identical segments are indicated by those segments that are continuously colored green. Note again that we have chosen the starting point of the analysis at location 2 million for simplicity, even though we have information from the beginning of the chromosome. Note also that every segment boundary in Table 5, except for the start and end locations, represents a crossover point and appears three times in the table.

Table 3. Permitted and Prohibited Transitions from one Chromosome Inheritance Pattern to Another

FROM Pattern Number	Chromosome Inheritance Pattern for that FROM Pattern				Permitted or Prohibited Transitions TO Pattern Number							
					1	2	3	4	5	6	7	8
1	**A**	**A**	**A**	**A**	-	**Y**	**Y**	**N**	**Y**	**N**	**N**	**Y**
2	**A**	**A**	**A**	**B**	**Y**	-	**N**	**Y**	**N**	**Y**	**Y**	**N**
3	**A**	**A**	**B**	**A**	**Y**	**N**	-	**Y**	**N**	**Y**	**Y**	**N**
4	**A**	**A**	**B**	**B**	**N**	**Y**	**Y**	-	**Y**	**N**	**N**	**Y**
5	**A**	**B**	**A**	**A**	**Y**	**N**	**N**	**Y**	-	**Y**	**Y**	**N**
6	**A**	**B**	**A**	**B**	**N**	**Y**	**Y**	**N**	**Y**	-	**N**	**Y**
7	**A**	**B**	**B**	**A**	**N**	**Y**	**Y**	**N**	**Y**	**N**	-	**Y**
8	A	**B**	**B**	**B**	**Y**	**N**	**N**	**Y**	**N**	**Y**	**Y**	-

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 4. Graphical Illustration of the Matching Segments

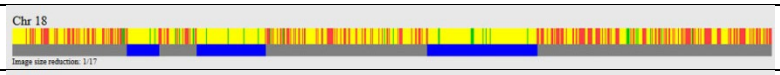
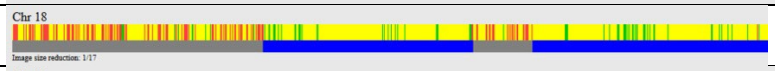
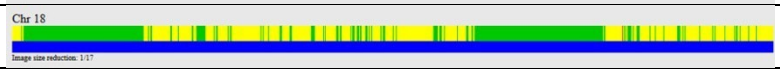
Comparison	
Sib1 to Sib2	
Sib1 to Sib3	
Sib1 to Sib4	
Sib2 to Sib3	
Sib2 to Sib4	
Sib3 to Sib4	

Table 5. Harmonized Matching Segments Between the Four Siblings

Chr	MatchType	#SNPs	Start	Stop	Length	Comparison	
18	Single	708	9.64	12.86	3.02	Sib1	Sib2
18	Single	1523	20.20	27.00	6.82	Sib1	Sib2
18	Single	2459	45.50	55.40	10.73	Sib1	Sib2
18	Double	2690	2	10.89	10.03	Sib1	Sib3
18	Double	1208	12.86	20.20	9.08	Sib1	Sib3
18	Double	924	45.50	49.40	3.92	Sib1	Sib3
18	Single	17010	2	76.12	58.24	Sib1	Sib3
18	Double	3193	2	12.86	12.09	Sib1	Sib4
18	Double	5369	20.20	45.50	25.32	Sib1	Sib4
18	Single	17012	2	76.12	76.06	Sib1	Sib4
18	Double	3821	61.80	76.12	14.31	Sib1	Sib4
18	Single	500	9.64	10.89	1.25	Sib2	Sib3
18	Single	4735	27.00	49.40	22.44	Sib2	Sib3
18	Single	5079	55.40	76.12	19.43	Sib2	Sib3
18	Single	3076	9.64	27.00	17.39	Sib2	Sib4
18	Single	1512	55.40	61.80	6.48	Sib2	Sib4
18	Double	2690	2	10.89	10.03	Sib3	Sib4
18	Double	2873	49.40	61.80	12.49	Sib3	Sib4
18	Single	17011	2	76.12	76.05	Sib3	Sib4

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 6. Segment Boundaries, Crossovers, and Match Types

Segment Boundaries			Segment Matching Patterns					
Start (millions)	Stop (millions)	Sib who Owns the Crossover at Stop	Sib1-Sib2	Sig1-Sib3	Sib1-Sib4	Sib2-Sib3	Sib2-Sib4	Sib3-Sib4
2.00	9.64	Sib2	0	2	2	0	0	2
9.64	10.89	Sib3	1	2	2	1	1	2
10.89	12.86	Sib1	1	1	2	0	1	1
12.86	20.20	Sib1	0	2	1	0	1	1
20.20	27.00	Sib2	1	1	2	0	1	1
27.00	45.50	Sib1	0	1	2	1	0	1
45.50	49.40	Sib3	1	2	1	1	0	1
49.40	55.40	Sib2	1	1	1	0	0	2
55.40	61.80	Sib4	0	1	1	1	1	2
61.80	76.12	Sib4	0	1	2	1	0	1

The first step in the analysis is to determine the set of unique crossovers and order the segments, resulting in the following segment definitions as shown in Table 6 (and illustrated in Figure 2). We then add the matching types for each sibling-pair comparison.

Corresponding to each segment matching pattern is just one pair of chromosome inheritance patterns. We have set up a look-up table to determine the pair of chromosome inheritance patterns corresponding to each segment matching pattern (see Table S1). We will not know which of the pair of chromosome inheritance patterns belong to the paternal and maternal sides, but we will have the possible patterns (Pattern numbers as in Table 3). These are shown in Table 7, with the pairs of Chromosome Inheritance Patterns just ordered numerically with the smaller value on the left.

For the first segment from 2.0 million to 9.64 million, the patterns are the same (pattern 5) for both the paternal and maternal sides, so there is no decision to be made regarding which pattern is on the paternal side and which is on the maternal side. However, we do not know if the crossover at 9.64 million is on the paternal or maternal side. Coming out of any segment where the chromosome inheritance patterns are the same on both sides, we will not know on which side the next crossover occurs. We must depend on the existence of a cousin match in the next region to make this determination. Alternatively, a cousin match further down the chromosome may be worked backward to this segment, or a guess may be made that any given crossover has a somewhat greater probability of being on the maternal side. In this example we will assume that we have information that shows that the first crossover at 9.64 million has occurred on the maternal side. If we have the wrong initial assignment of the crossover, we can simply switch the results at the end.

Another important principle needed for making the determination of which side the chromosome inheritance patterns belong on, is that only one of

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

the patterns, either that from the paternal chromosome or that from the maternal chromosome, will change at each crossover. Therefore, if we assume that the crossover at 9.64 million is on the maternal side, then the ordered set of pattern numbers in the second segment is 5-1. One of these patterns, either the 5 (ABAA) or the 1 (AAAA), will continue into the next segment since the crossover must be on either the paternal or the maternal chromosome. Therefore, since the pattern 5 appears again in the third segment, it must be on the same side as in the second segment--the paternal side. That leaves pattern 1, changing to pattern 3, to describe the maternal sides in the second and third segments. We can continue to step through the segments, assigning the pattern numbers as above. In this particular case, the pattern on the paternal side continues the same (pattern 5) through all of the segments, but this will not be the case in general. As long as we do not encounter a segment where the Inheritance pattern is the same on both paternal and maternal sides, we will have the rest of the chromosome determined. The final solution to the patterns will be as shown in Table 8, and we will have all the information required to construct a schematic recombination diagram. In fact, the last six columns of Table 8 represent such a schematic diagram.

Table 7. Segment Boundaries and Chromosome Inheritance Patterns

Start (millions)	Stop (millions)	Segment Matching Pattern	Only possible paternal/maternal Chromosome Inheritance Patterns	
2.00	9.64	022002	5	5
9.64	10.89	122112	1	5
10.89	12.86	112011	3	5
12.86	20.20	021011	5	6
20.20	27.00	112011	3	5
27.00	45.50	012101	5	7
45.50	49.40	121101	2	5
49.40	55.40	111002	4	5
55.40	61.80	011112	5	8
61.80	76.12	012101	5	7

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 8. Segment Boundaries and Resolved Chromosome Inheritance Patterns

Start (millions)	Stop (millions)		Only Possible Paternal/ Maternal Inheritance Patterns			Schematic Recombination Diagram (In the first segment, both paternal and maternal sides have a Chromosome Inheritance Pattern of 5 (ABAA), so we fill in the same colors for Sib1, Sib3, and Sib4, and a different color for Sib2. Then, if we assume that the first crossover is on the maternal side, then the rest of the diagram is specified by the Chromosome Inheritance Patterns and can be filled in)							
						Sib1		Sib2		Sib3		Sib4	
						Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat
2.00	9.64		5	5									
9.64	10.89		5	1									
10.89	12.86		5	3									
12.86	20.20		5	6									
20.20	27.00		5	3									
27.00	45.50		5	7									
45.50	49.40		5	2									
49.40	55.40		5	4									
55.40	61.80		5	8									
61.80	76.12		5	7									

As noted in the table heading, we cannot determine from just the paternal/maternal inheritance patterns, which of the four grandparents contributed each color in the schematic recombination diagram. That is, another valid recombination diagram can be made by interchanging the pink and blue along with the blue and green. The (tentative) paternal and maternal chromosomes could also be interchanged. The correct choice must usually be determined from second cousin matches. First cousin matches will not work because first cousins would have DNA from both of your common grandparents, just like you, though first cousin matches can help distinguish the paternal chromosomes from the maternal, and if the first cousin matching segments extend across one or more of the sibling crossovers, they may help extend the analysis and resolve ambiguities. Second cousins, on the other hand, would share DNA with only one of your grandparents. If data from a parent is available, this could correctly distinguish the paternal-maternal chromosome assignments. Obviously, third (or more distant) cousin matches would work also if you have any for whom the exact relationship is known, but usually a third cousin would match with you on only a small number of chromosomes. Any matches you have with known third (or more distant) cousins would still be useful in analyzing that small set of chromosomes where such matches occur.

Now we must look for some second cousin matches to complete the analysis. To illustrate this process, we will assume that we find the following matches:

> A match from 44 million to 74 million between Sib1, Sib3, and Sib4 (but not Sib2) and a second cousin who is related to the four siblings

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

through their paternal grandmother. This serves to identify the pink color with the paternal grandmother and confirms that we have made the correct initial assignment of the paternal and maternal sides. This also serves to identify the light blue color with the other paternal grandparent, the paternal grandfather.

A match from 2.2 million to 8 million between Sib1, Sib3 and Sib4 (but not Sib2) with a second cousin who is related to the four siblings through their maternal grandmother. This serves to identify the light green color with the maternal grandmother, and therefore the dark blue color would then be assigned to the maternal grandfather.

The resulting complete schematic diagram can be shown in Table 9.

Table 9. Segment Boundaries and Resolved Chromosome Inheritance Patterns

Start (millions)	Stop (millions)		Only Possible Paternal/ Maternal Inheritance Patterns			Schematic Recombination Diagram							
						Sib1		Sib2		Sib3		Sib4	
						Pat	Mat	Pat	Mat	Pat	Mat	Pat	Mat
2.00	9.64		5	5		PGM	MGM	PGF	MGF	PGM	MGM	PGM	MGM
9.64	10.89		5	1		PGM	MGM	PGF	MGM	PGM	MGM	PGM	MGM
10.89	12.86		5	3		PGM	MGM	PGF	MGM	PGM	MGF	PGM	MGM
12.86	20.20		5	6		PGM	MGF	PGF	MGM	PGM	MGF	PGM	MGM
20.20	27.00		5	3		PGM	MGM	PGF	MGM	PGM	MGF	PGM	MGM
27.00	45.50		5	7		PGM	MGM	PGF	MGF	PGM	MGF	PGM	MGM
45.50	49.40		5	2		PGM	MGF	PGF	MGF	PGM	MGF	PGM	MGM
49.40	55.40		5	4		PGM	MGF	PGF	MGF	PGM	MGM	PGM	MGM
55.40	61.80		5	8		PGM	MGF	PGF	MGM	PGM	MGM	PGM	MGM
61.80	76.12		5	7		PGM	MGF	PGF	MGM	PGM	MGM	PGM	MGF

PGF=paternal grandfather, PGM=paternal grandmother, MGF=maternal grandfather, MGM=maternal grandfather.

Note that the shaded columns represent a schematic set of chromosome diagrams that are identical to those shown in Figures 1 and 2, the only difference in the two tables being the chromosome orientations--horizontal in the figures and vertical in the table.

It is important to note that it is normally not necessary to have a second cousin match for every segment shown in the table. The matches with the two second cousins in our example serve not only to identify the segments where the matches occur, but also the remainder of the chromosome diagram since it is already fixed schematically from Table 9, so all the segments colored the same would be from that same grandparent.

However, the extension of the grandparent assignment throughout a chromosome, based upon just one second cousin matching segment, often runs into a fundamental difficulty. When a segment in the middle of a chromosome has the same

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

chromosome inheritance pattern on both the paternal and maternal sides, which we term an "ambiguous segment," then a second cousin match either downstream or upstream from this ambiguous segment cannot be extended through to the other side. Then it is necessary to have an additional second cousin match on the other side of the ambiguous segment to resolve the difficulty. Sometimes there may be more than one ambiguous segment in an analysis, and each one requires an additional cousin match on that chromosome.

At the end of this article under "Web Resources," a link is provided for downloading an Excel spreadsheet that will automatically do all of the pattern determinations and conversions discussed above for the four-sibling problem, with the schematic recombination diagram as final output. There is also a link to instructions for use with this spreadsheet.

There are a number of practical complications to using the approach described above (and any other approach).

1. When there are crossover points very close together, the segment boundaries being only approximate may cause difficulties in defining the unique segments--particularly in regard to which of two closely located crossovers comes first. Sometimes, approaching the crossovers from different directions, along with use of the rules on permitted and prohibited transitions, can help resolve which crossover comes first. It is also possible for cousin matches to aid in location of close crossovers.

2. As discussed above, whenever the chromosome inheritance patterns are the same within a segment on both the paternal and maternal chromosomes--an ambiguous segment, then the next crossover cannot immediately be assigned to the paternal or maternal side, and two possible solutions issue from that point. This equality of the chromosome matching patterns may occur more than once along the chromosome, each time doubling the number of possible solutions. Usually, by carefully examining cousin matches, the difficulty can be resolved. Even matches with distant cousins that are known to be on the paternal or maternal sides, if the match extends across the crossover in question, can serve to assign the crossover to the proper side. This problem also applies to the visual/graphical approach that is described next--there is no way around the difficulty except to resolve the ambiguity with cousin matches.

One advantage of using four or more siblings is that the more siblings in the analysis, the more possible inheritance patterns exist and the probability of finding inheritance patterns the same on both paternal and maternal sides becomes smaller. The probability of finding both sides with the same inheritance pattern in any given segment is approximately 1/4 for 3 siblings, 1/8 for 4 siblings, 1/16 for 5 siblings, etc. Adding more siblings complicates the analysis, but a complete (unique) solution becomes more probable.

3. in order to use the referenced Excel spreadsheet for the recombination analysis, the endpoints of the matching segments between siblings must be harmonized manually. The program will not work if the same crossover has slightly different locations in different matching segments. It is often the case that the most important part of the analysis is coming up with a clean set of sibling matching segments with the crossovers harmonized for use as input.

4. Sometimes a matching segment will be too small to be reported by the standard comparison

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

algorithms using the default match parameters. Since we know that siblings are definitely related, false positives are less of a concern and the match criteria can be significantly relaxed. But, it may sometimes be necessary just to infer that a segment is missing and then construct a suitable matching segment to complete the set. For example, in Table 5 the segment from 9.64 to 10.85 was not reported by the matching algorithm, so it was artificially constructed to complete the segment set.

5. The same method may be applied in the case of five or more siblings with data, but the number of possible patterns becomes rapidly much larger--the number doubles with each added sibling. It will be easier in general to just analyze four of the five (or six, etc) siblings, then analyze four more including those left out of the first analysis, though the choice of which siblings to include in the analysis should be made in a way that minimizes the number of ambiguous segments.

Numerical Method--Case of Three Siblings

The minimum number of siblings that can be analyzed using the approach described above is three. We present an example of a different set of three siblings, but also using data from their chromosomes 18. We begin with the match segments for the three siblings. The match segment endpoints must have been harmonized so that each crossover has a uniquely assigned location. The short double match from .1 to .96 has been inferred, so no number of SNPs has been entered into Table 10.

As in the four-sibling example, we determine the unique segments, each delineated by a crossover, as we "scroll" through the chromosome.

Similar to the four-sibling case, with three siblings we have four possible chromosome inheritance patterns and there are restrictions on the transitions from one to another. Table 12 shows the permitted transitions (indicated by a Y) and prohibited transitions (indicated by an N) for the four patterns.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 10. Segment Match Data for Three Siblings

Chr	MatchType	#SNPs	Start	Stop	Length	Comparison	
18	Single	1018	0.10	6.00	5.90	Sib1	Sib2
18	Double		0.10	0.96	0.86	Sib1	Sib2
18	Single	4013	47.00	76.09	29.09	Sib1	Sib2
18	Single	12550	6.00	76.09	70.09	Sib1	Sib3
18	Double	4559	52.81	76.09	23.28	Sib1	Sib3
18	Single	5470	0.96	47.00	46.04	Sib2	Sib3
18	Single	4878	52.81	76.09	23.28	Sib2	Sib3

Table 11. Segment Boundaries and Match Types

Start (millions)	Stop (millions)		Sib1-Sib2	Sig1-Sib3	Sib2-Sib3
0	1		2	0	0
1	6		1	0	1
6	47		0	1	1
47	53		1	1	0
53	76 (end)		1	2	1

Table 12 . Permitted or Prohibited Transitions From One Inheritance Pattern to Another

FROM Pattern Number	Chromosome Inheritance Pattern for that FROM Pattern			Transitions TO Pattern Number 1	2	3	4
1	**A**	**A**	**A**	-	**Y**	**Y**	**N**
2	**A**	**A**	**B**	**Y**	-	**N**	**Y**
3	**A**	B	A	**Y**	**N**	-	**Y**
4	**A**	**B**	**B**	**N**	**Y**	**Y**	-

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Also similar to the four-sibling case, we have a one-to-one correspondence between the segment matching pattern and the pair of chromosome inheritance patterns. The four-chromosome inheritance patterns are shown again by their number as a part of the complete correspondence table as follows.

Table 13. Look-Up Table--Converting Segment Matching Patterns to Pairs of Chromosome Inheritance Patterns

Segment Matching Patterns				Chromosome Inheritance Pattern	
Sib1-Sib2	Sib1-Sib3	Sib2-Sib3		(either order)	
2	2	2		1	1
2	1	1		1	2
1	2	1		1	3
1	1	2		1	4
2	1	1		2	1
2	0	0		2	2
1	1	0		2	3
1	0	1		2	4
1	2	1		3	1
1	1	0		3	2
0	2	0		3	3
0	1	1		3	4
1	1	2		4	1
1	0	1		4	2
0	1	1		4	3
0	0	2		4	4

Note: Some Segment Matching Patterns do not occur. For example, the pattern 2-1-2 does not occur, because if siblings 1 and 3 are fully matching, and sibling 2 and 3 are fully matching, it is impossible for the sibling1-sibling3 comparison to be anything other than fully matching.

Returning to our example, in each unique segment we can take the segment matching pattern and look it up in Table 13 to obtain the corresponding pair of chromosome inheritance patterns. This pair of chromosome inheritance patterns will initially be unordered--we will not know which of each pair is paternal and which is maternal. However, we can either (1) arbitrarily guess the assignment in the first segment and then propagate that assignment throughout the chromosome (subject to later correction), or (2) use a cousin match to make the initial assignment in one of the segments, and propagate that assignment to either side of the cousin-matching segment.

In Table 14 we make an initial guess that the first crossover takes place on the maternal side. If this turns out to be incorrect on the basis of cousin

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

matches, then we would simply switch the assignments. Even with the paternal-maternal assignments determined, we would still not know which grandparent should be assigned to each color.

Table 14. The Segment Matching Patterns Transformed Into a Schematic Recombination Diagram

Start	Stop		Segment Matching Pattern (from Table 10)				Only possible Paternal/ Maternal Chromo-some Inheri-tance patterns			Ordered Chromo-some Inheritance Patterns			Schematic Recombination Diagram (In the first segment, both paternal and maternal sides have a Chromosome Inheritance Pattern of 2 (AAB), so we fill in the same colors for Sib1 and Sib2, and different colors for Sib3. Then, if we assume that the first crossover is on the maternal side, then the rest of the diagram is specified by the Chromosome Inheritance Patterns and can be filled in)					
			1-2	1-3	2-3		Unordered			Pater-nal	Mater-nal		Sib1		Sib2		Sib3	
													Pat	Mat	Pat	Mat	Pat	Mat
0	1		2	0	0		2	2		2 (AAB)	2 (AAB)		. . .	. . .	. . .	. . .	. . .	. . .
1	6		1	0	1		2	4		2 (AAB)	4 (ABB)		. . .	. . .	. . .	. . .	. . .	. . .
6	47		0	1	1		3	4		3 (ABA)	4 (ABB)		. . .	. . .	. . .	. . .	. . .	. . .
47	53		1	1	0		2	3		3 (ABA)	2 (AAB)		. . .	. . .	. . .	. . .	. . .	. . .
53	76		1	2	1		1	3		3 (ABA)	1 (AAA)		. . .	. . .	. . .	. . .	. . .	. . .

Now we must look for cousin matches to complete the analysis. To illustrate this process, we will assume that we find the following cousin matches:

> A match from 40 million to 50 million between both Sib1 and Sib3 (but not Sib2) and a second cousin who is related to the three siblings through their paternal grandmother. This serves to identify the blue color with the paternal grandmother and confirms that we have made the correct initial assignment of the paternal and maternal sides. This also serves to identify the green color with the other paternal grandparent, the paternal grandfather.
>
> A match from 20 million to 40 million between both Sib2 and Sib3 (but not Sib1) with a second cousin who is related to the three siblings through their maternal grandfather. This serves to identify the yellow color with the maternal grandfather, and therefore the purple color would then be assigned to the maternal grandmother.

The resulting complete schematic diagram can be shown in Table 15.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 15. The Segment Matching Patterns Transformed Into a Schematic Recombination Diagram

Start	Stop		Segment Matching Pattern (from Table 10)				Only possible Paternal/ Maternal Chromo-some Inheri-tance (CI) patterns			Order-ed CI Pat-terns			Schematic Recombination Diagram (The grandparent assignment has been filled in here, with PGF = Paternal Grandfather, PGM = Paternal Grandmother, etc.)					
			1-2	1-3	2-3		Unordered			PAT	MAT		Sib1		Sib2		Sib3	
													Pat	Mat	Pat	Mat	Pat	Mat
0	1		2	0	0		2	2		2	2		PGF	MGM	PGF	MGM	PGM	MGF
1	6		1	0	1		2	4		2	4		PGF	MGM	PGF	MGF	PGM	MGF
6	47		0	1	1		3	4		3	4		PGM	MGM	PGF	MGF	PGM	MGF
47	53		1	1	0		2	3		3	2		PGM	MGM	PGF	MGM	PGM	MGF
53	76		1	2	1		1	3		3	1		PGM	MGM	PGF	MGM	PGM	MGM

Since we did not have any ambiguous segments in the middle of this chromosome, we could propagate our cousin matches throughout the chromosome. However, we will not usually be so lucky as will be illustrated in an example from a different family, but which is also from chromosome 18. Consider the following Table 16 that is similar to Table 10:

Table 16. Segment Boundaries and Match Types

Start (millions)	Stop (millions)		Sib1-Sib2	Sig1-Sib3	Sib2-Sib3
0.0	1.0		1	1	2
1.0	12		2	1	1
12	23		2	0	0
23	34		1	0	1
34	44		1	1	0
44	56		2	0	0
56	71		1	0	1
71	72		0	0	2
72	END		1	1	2

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table 17. The Segment Matching Patterns Transformed Into a Schematic Recombination Diagram

Start	Stop		Segment Matching Segment Matching Pattern (from Table 15)				Only possible Paternal/ Maternal Chromosome Inheritance (CI) patterns	
			1-2	1-3	2-3		Unordered	
0.0	1.0		1	1	2		1	4
1.0	12		2	1	1		1	2
12	23		2	0	0		2	2
23	34		1	0	1		2	4
34	44		1	1	0		2	3
44	56		2	0	0		2	2
56	71		1	0	1		2	4
71	72		0	0	2		4	4
72	79 End		1	1	2		1	4

These segment matching patterns can be looked up in Table 13 and the corresponding pairs of chromosome inheritance patterns can be determined as shown in Table 17.

In this example we now see that we have three ambiguous segments (highlighted in pink) in the chromosome. It will be necessary to have cousin matches in all four of the regions outside of the ambiguous segments, and this makes the analysis more difficult. However, it is not impossible because in this example from an actual case, the necessary cousin matches were found and the solution could be completed. That solution will be illustrated in the following section which uses the visual approach. It is important to note that the progress of the solution, using either the numerical approach or the visual approach, is temporarily stymied by the presence of the ambiguous segments. However, the nature of the problem would not be so obvious when using the visual approach alone if one did not understand how ambiguous segments arise. If there are no ambiguous segments to deal with, both the numerical approach and the visual approach can proceed in a straightforward manner to the final solution. If ambiguous segments are present, both approaches will depend on extra cousin matches to complete the analysis.

For the case of three siblings, we can see from Table 13 that the four segment matching patterns that produce ambiguous segments are: 222, 200, 020, 002, two of which are present in Table 17. It will usually be worthwhile when using the visual approach to use at least this much of the numerical approach--checking the segment matching patterns in each segment--to determine if the ambiguous segment problem will be present in a proposed analysis.

As was the case for the four-sibling problem, a link to an Excel spreadsheet that can do all of the analysis automatically for the three-sibling problem is included under "Web Resources." A link to instructions is also provided.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Visual/Graphical Method or *Visual Phasing*

We now start over with the last three-sibling example discussed above in Tables 16-17, but this time we take a more intuitive and visual approach, which is often termed "visual phasing" (Johnston 2015). We use GEDmatch for this example because of the chromosome diagrams that can be obtained along with the matching segment determinations. GEDmatch also has the advantage that fully identical matches may be identified in the diagrams. Microsoft PowerPoint was used in this example solution and it is recommended for this purpose, but other software (e.g., Excel) could potentially be used if it has similar functionality, or even paper and pencil could be used.

Because we already know that there are three ambiguous segments involved in the analysis of this example, we can anticipate that the analysis is going to stall and that more cousin matches than usual will be necessary to complete the analysis.

Following is a step-by-step approach to the visual phasing analysis based on recombination.

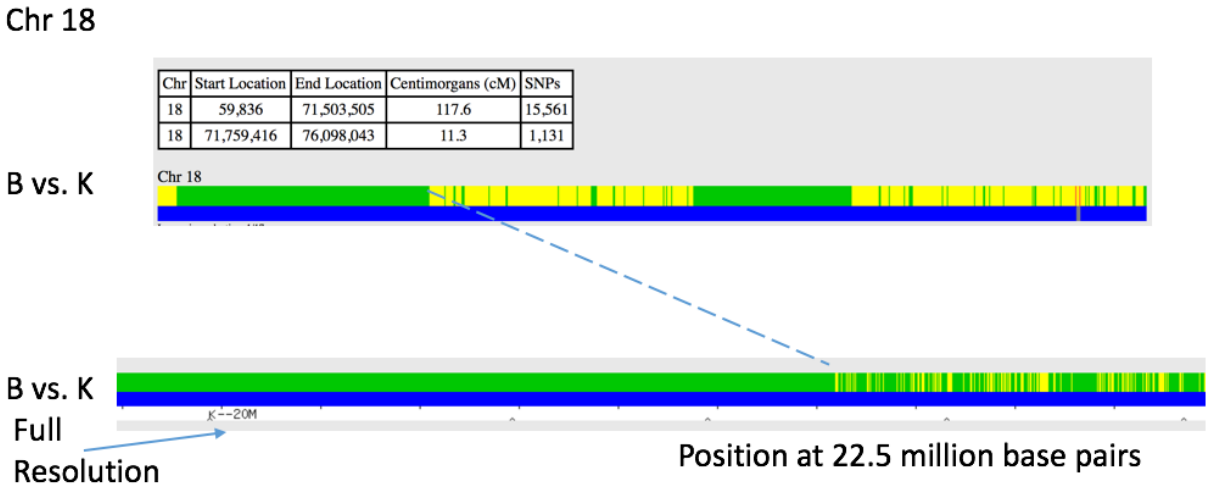
1. The goal is to use crossover lines among siblings in PowerPoint to phase the parents' chromosomes, then determine segment mapping to grandparents when no parents or grandparents are available for testing. Although this is an intuitive method geared toward genealogists, an understanding of recombination is required.

2. Choose a single chromosome to compare at GEDmatch in the one-to-one comparison between full siblings. The same version from the same company is recommended for the best alignment. The unique graphic that GEDmatch provides helps to distinguish the fully identical, half identical and non-identical segments which aids in the identification of crossover borders made during recombination. A half-identical "single match" is determined when one out of every two alleles over a segment are matching and a fully identical "double match" is determined when both pairs of alleles match for a minimum distance of 7 cM. The raw data includes selected single nucleotide variants and the matching segment measurement is not representative of an exact distance in the traditional sense. You can expect some indistinct borders in the match process even though the crossover border created by a parent is at a distinct physical site. It is therefore advisable to round off to the nearest million base pairs when identifying crosssover points.

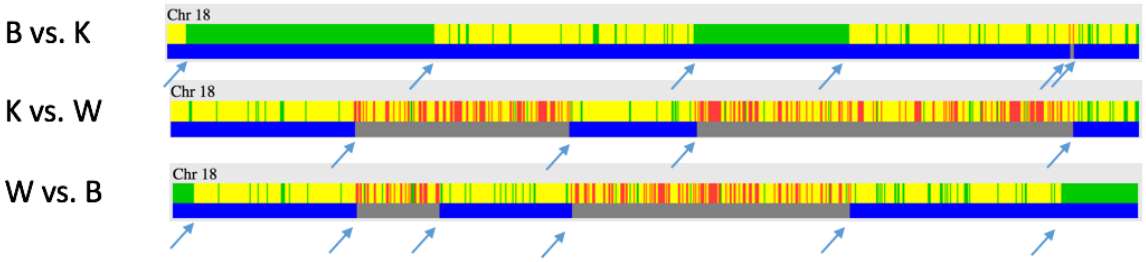
3. In our example, Chromosome 18 data from 23andMe is used to compare siblings at GEDmatch, using the one-to-one comparison feature. The present version of match tools at 23andMe also identifies full- and half-identical segments, which can be quite useful. We use the initials B, K and W to label each sibling, and we use the standard default match settings. In GEDmatch, take a screenshot of the graphical match diagram, then adjust the size (under Format) after copying it to PowerPoint. It should be emphasized that screenshots should be exactly cropped at the beginning and end. The image can also be resized by dragging its edge or corner. Stack each comparison between the three siblings as shown below.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

4. Line up the segments according to location:

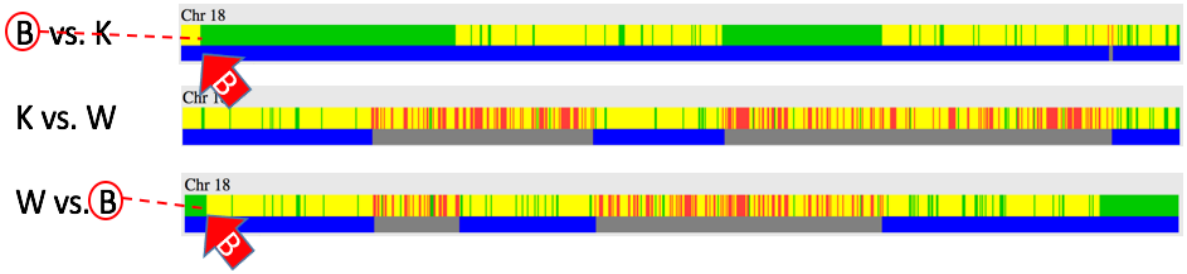


5. Identify the borders to the single matching and double matching segments. Skip the arrows if you already know how to do this.



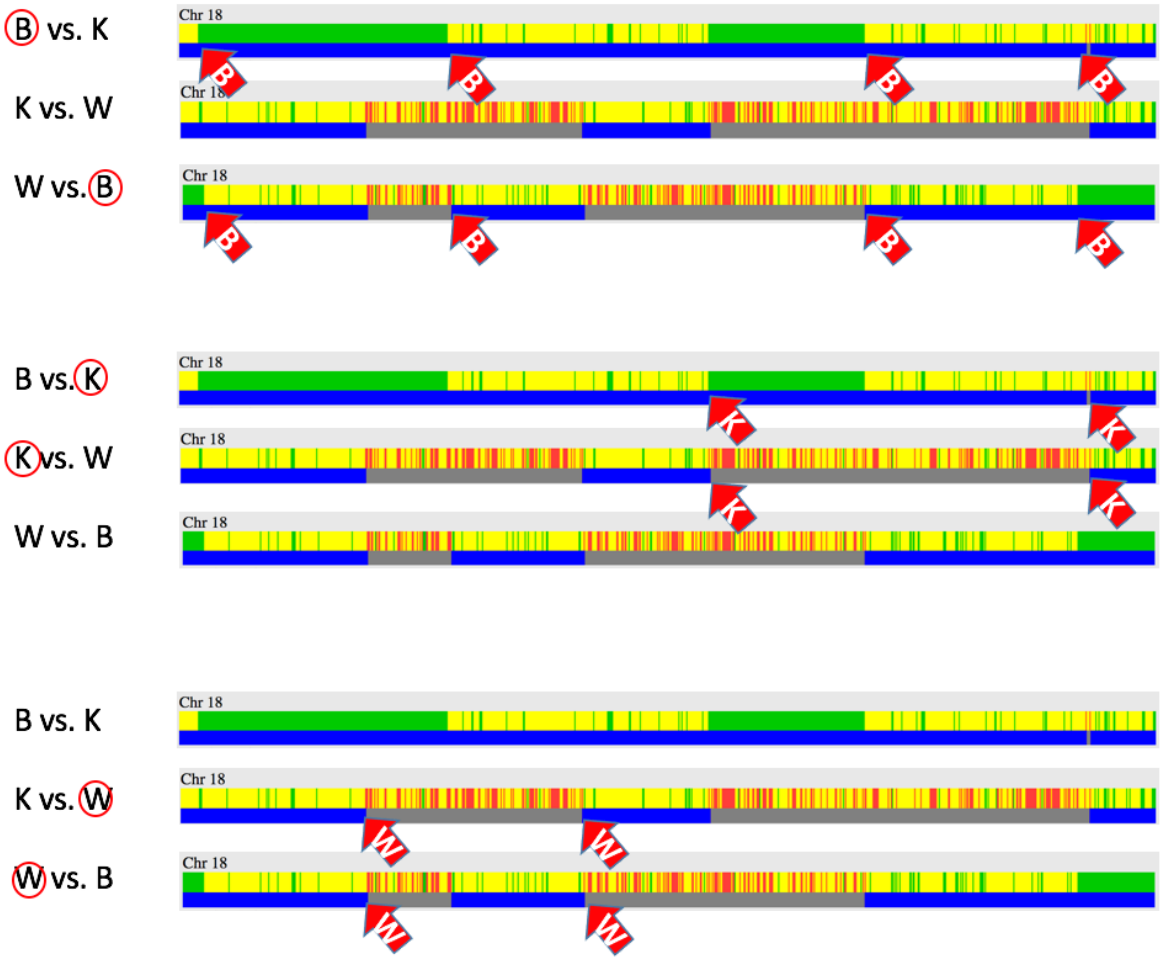
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

6. Any particular crossover point will occur in only one sibling; identify that sibling. The owner of the crossover is the one who is in-common with that point for two comparisons. For example, the first one belongs to B because the first and third comparisons both involve B. The second comparison (K vs. W) does not show any breakage or borders at this initial transition so therefore by process of elimination, K and W are not likely to own this crossover. There are exceptions to be discussed later.



© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

7. The following shows how to quickly identify each crossover with the most likely sibling.



A common challenge is the occurrence of two crossover points that are located very close as seen in the last crossover site. The last two crossovers may have looked like a single one made for W but on closer inspection, there was instead an interruption in the top blue B vs. K segment match indicative of nearby crossovers for B and K. Unless crossovers are made by the same parent for the same child there is no rule of thumb on how close the crossovers can be. Crossovers can be hidden from sight when parents use the same location for a crossover for two children or when two parents use the same location for one child. This author has seen crossovers as close as 6 cM made by the same parent on the same chromosome for the same child, but not closer than that. If the final configuration

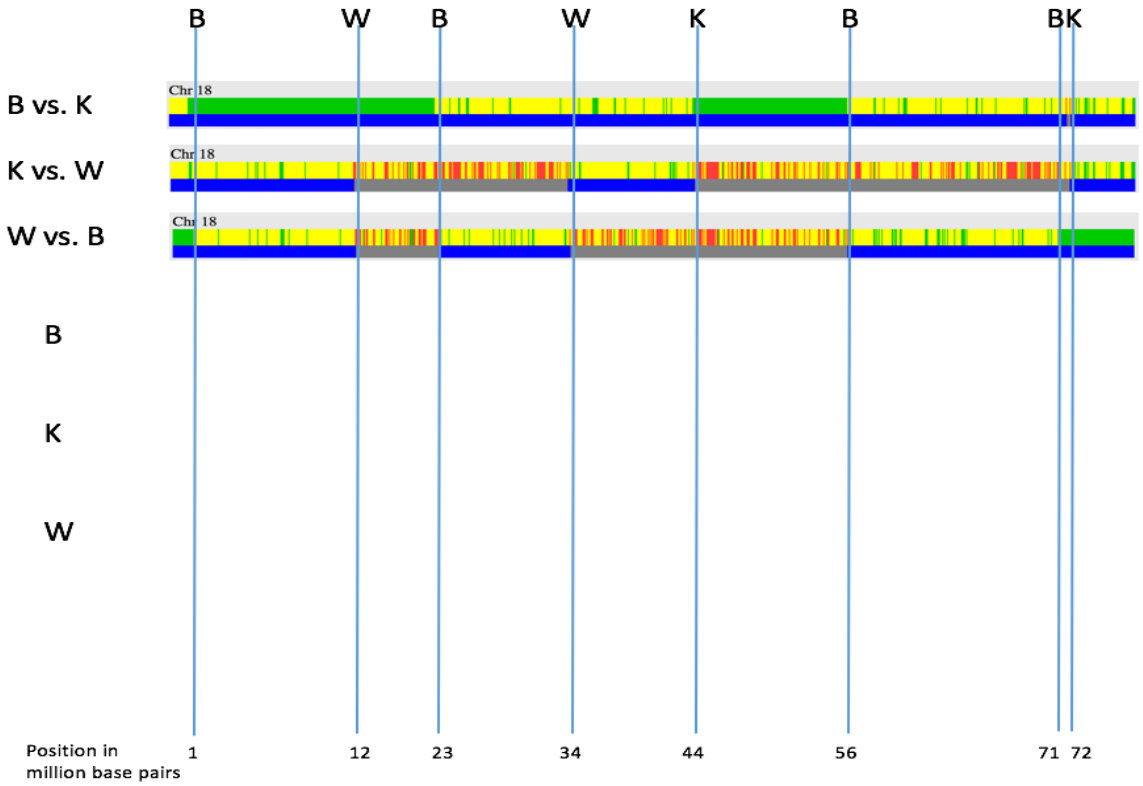
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

shows that the same parent made very close crossovers for the same child, that can be an indication that the identification of crossover ownership was incorrect. Interference makes this an unlikely event. More research needs to be done in real life situations to explore interference.

When each chromosome can be identified as either maternal or paternal, that is called phasing. When the segments on each phased chromosome can be identified from the crossover lines, that is what we are calling grandparent mapping. Through crossover recombination, each parent slices and splices two chromosomes together and passes a single one on to each offspring independently. A parent's two homologous chromosomes originated from the pair of chromosomes coming from his or her side of the family, the grandparent spouses. Ultimately this method will be used to map those alternating segments back to the grandparents' pairs and to fill in all the gaps through cousin matches. This is a first step in the DNA reconstruction of your ancestors from living descendants. Note that only the chromosomes coming from the parents will end up being phased, not the pairs coming from the grandparents.

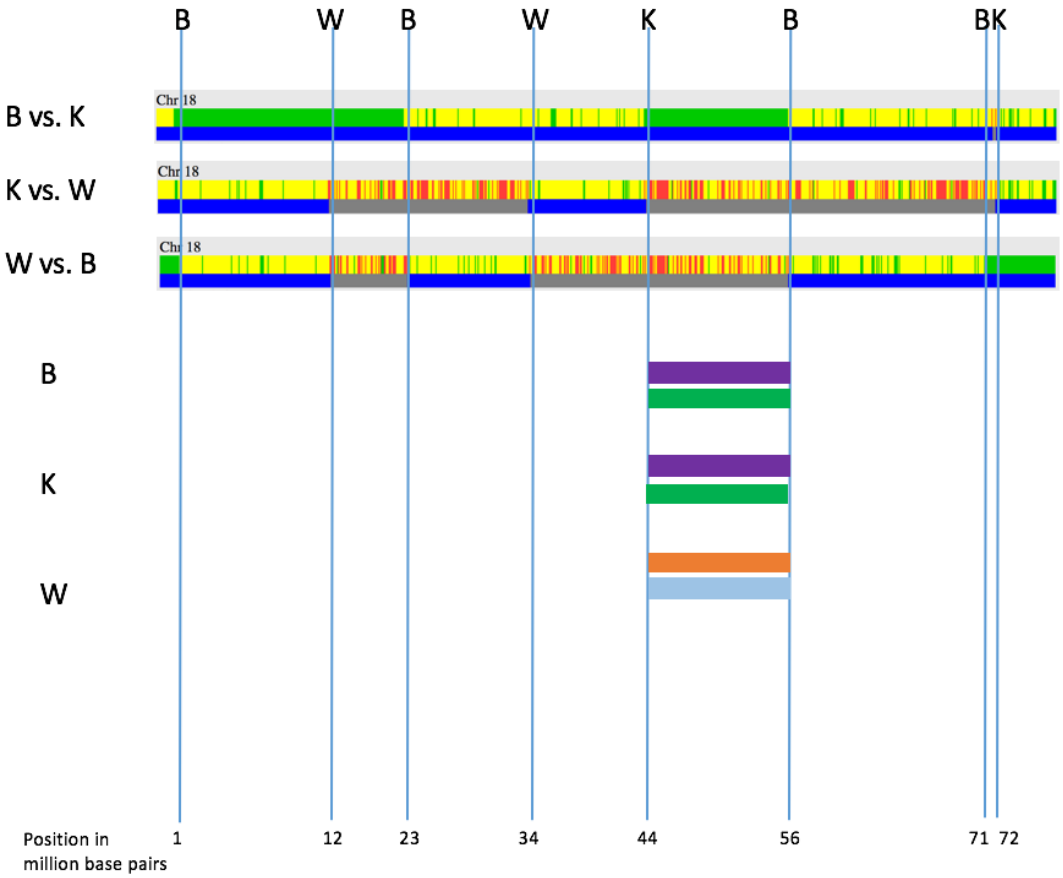
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

8. Click on “Insert” then “Shape” to add the crossover lines. Make sure the crossover lines are perpendicular to the comparisons. Move the crossover lines to run through approximate borders. Label each crossover line with the owner of that crossover. The parent who provided that crossover is still unknown. Later on that parent will be determined.



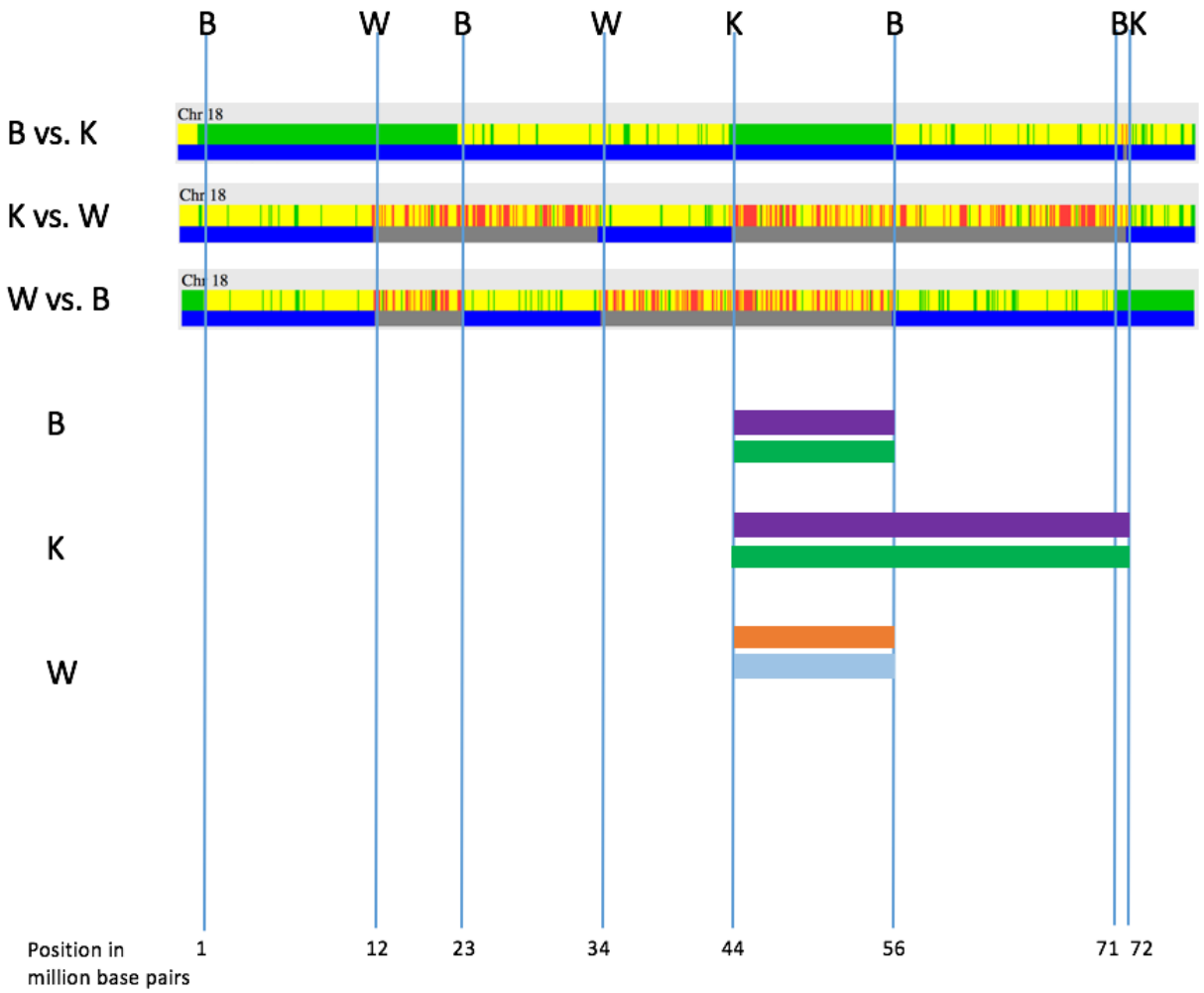
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

9. In PowerPoint, start with this skeleton of crossover lines. These will be the borders to the segments (these locations are also shown in Table 15). Leave room for three pairs of chromosome diagrams belonging to siblings B, K and W. Use a text box that can be filled in with four colors representing four grandparents. Double matching will show the same two colors. Single matching will show only one matched color. Non-identical regions will have four different colors, i.e. zero matching. Note how the number of color matches correlates to the numbers 2,1,0 used in the computational method. Pick any four colors but try not to add any extensions to segments that are not certain until you reach a point where you can map no further. To fill in the region with color, click on "Insert" then "Text Box" then click on "Format", "Shape" and "Fill". Click on the border of the segment until you see a solid cross, then move the colored chromosome segment to the crossover border line. Always refer back to the original stacked chromosome comparisons with the crossover lines. You may instead want to use colored pencils and skip the computer program for your comparisons or use another program such as Excel. Double matched, fully identical areas are a good place to start as shown below. B and K show an exact match between 44 and 56 but W does not share with the others in this location. We picked warm colors, purple and orange for the top set of grandparents and cool colors blue and green for the bottom set of grandparents.



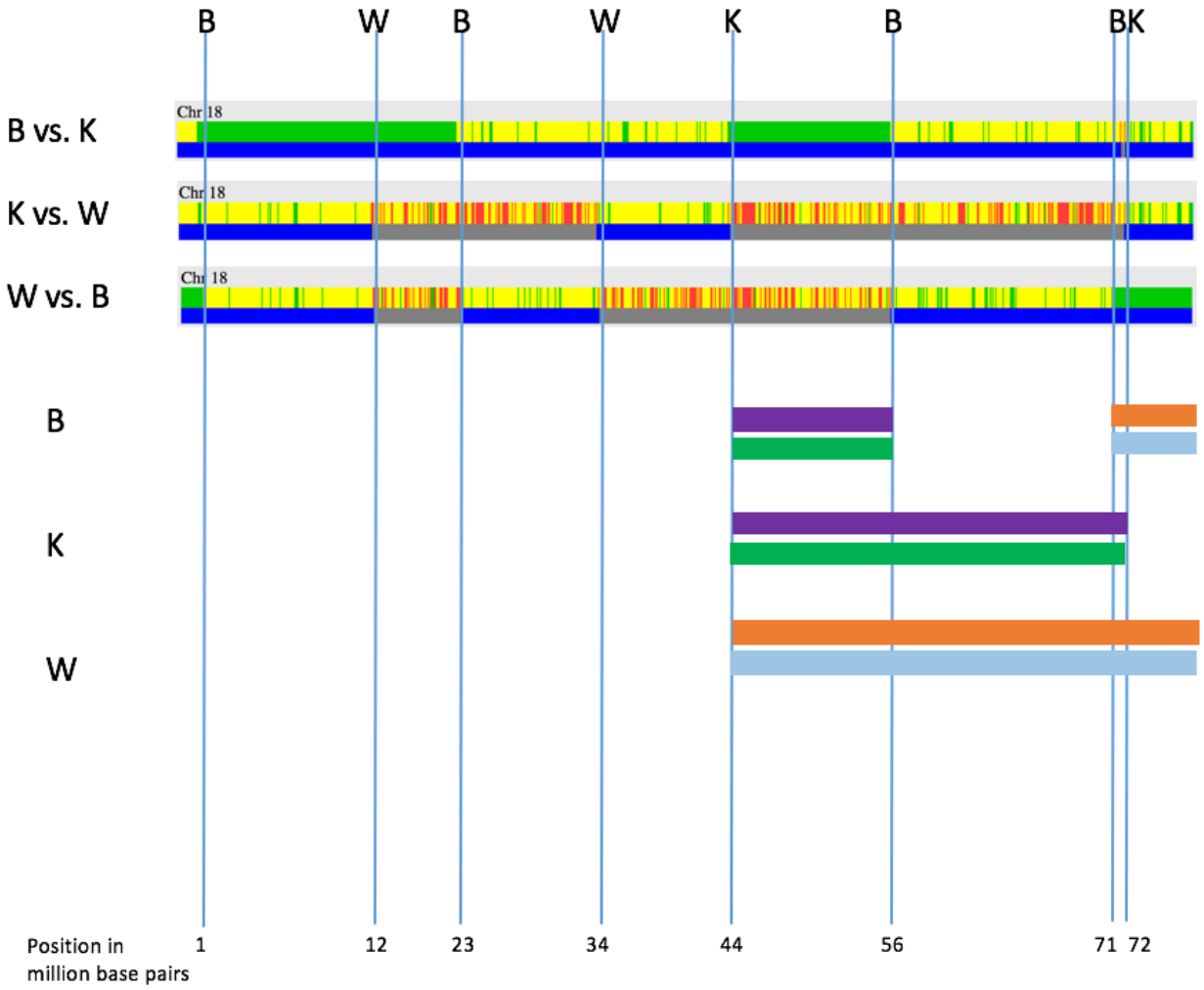
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

10. Extend the chromosome colors for K to the right to the crossover line K. Since K did not have any crossovers in the next two segments, then K maps to these same grandparents for the entire length until a K line signals a stop or until the chromosome ends. The top chromosome in the pair could come from either parent, but if the top chromosome is maternal, then the bottom chromosome has to be paternal and vice versa.



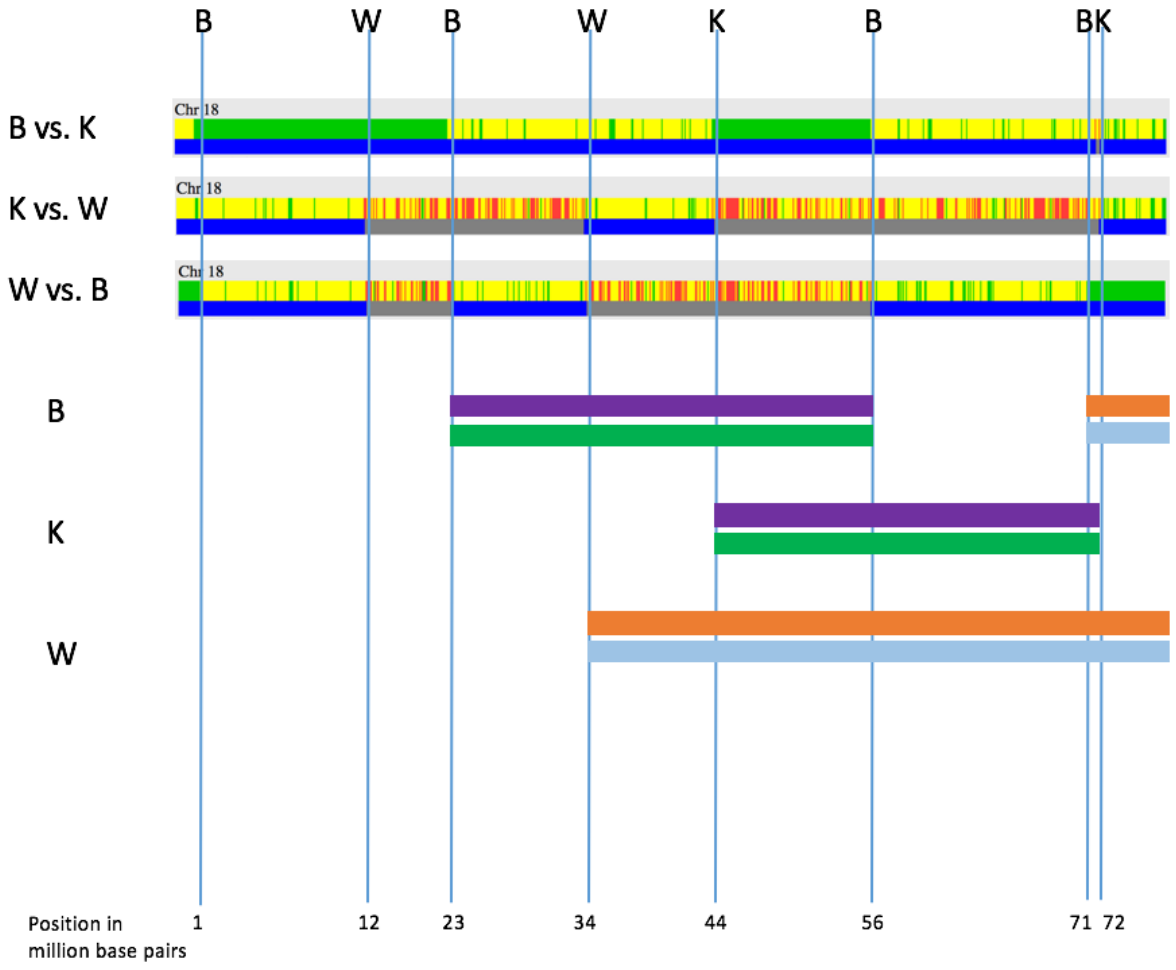
© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

11. The two segments for W can be extended out to the end since no more W crossovers are encountered in this region. In the last segment before the end, W completely matches B, so we can color B's chromosomes in this last section to be the same as W's. Note that these are just the preliminary assignments, and we will be able to extend all of the colored regions farther as we go along.



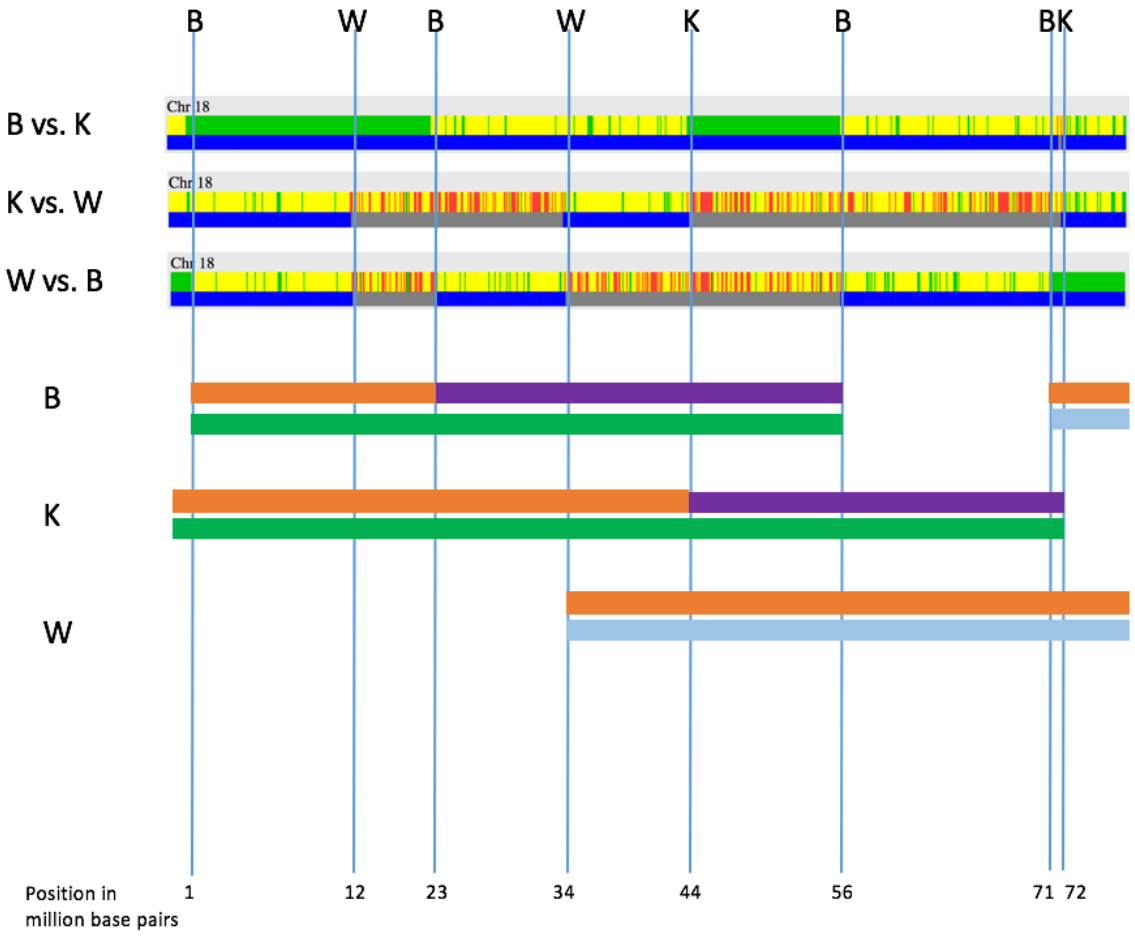
© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

12. Extend both segments for W to the left to the crossover W at position 34 where W and B no longer match. Extend both B chromosomes to the left until the B crossover line at position 23 signals a preliminary stopping point.



© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

13. Up until now the filled-in regions of the chromosomes on the top and bottom have been the same size i.e. symmetrical, covering the same locations on both chromosomes. At this point a decision must be made to break this symmetry. No other color matches are certain. One chromosome color can be arbitrarily chosen for K that extends all the way to the beginning. That means that the other chromosome for K would have to change colors at the border. As a general rule only one segment can change at a crossover line within a pair. Try not to make any more arbitrary decisions after the first one. You get one guess during the entire process. If the first guess does not go as far as you wish, you can return to this point, delete (or save elsewhere) your original extension and try another guess. We choose the upper chromosome for B to make the change at B's crossover at 23 million (from purple to orange). It is now obvious that the chromosomes for B and K chromosomes can be filled in between the B crossover at 23 million and the B crossover at 1 million because B matches K exactly between positions 1 and 23. Then we can fill in the gap in K's chromosome pair between 23 and 44 million so that the same colors match at 44 million.

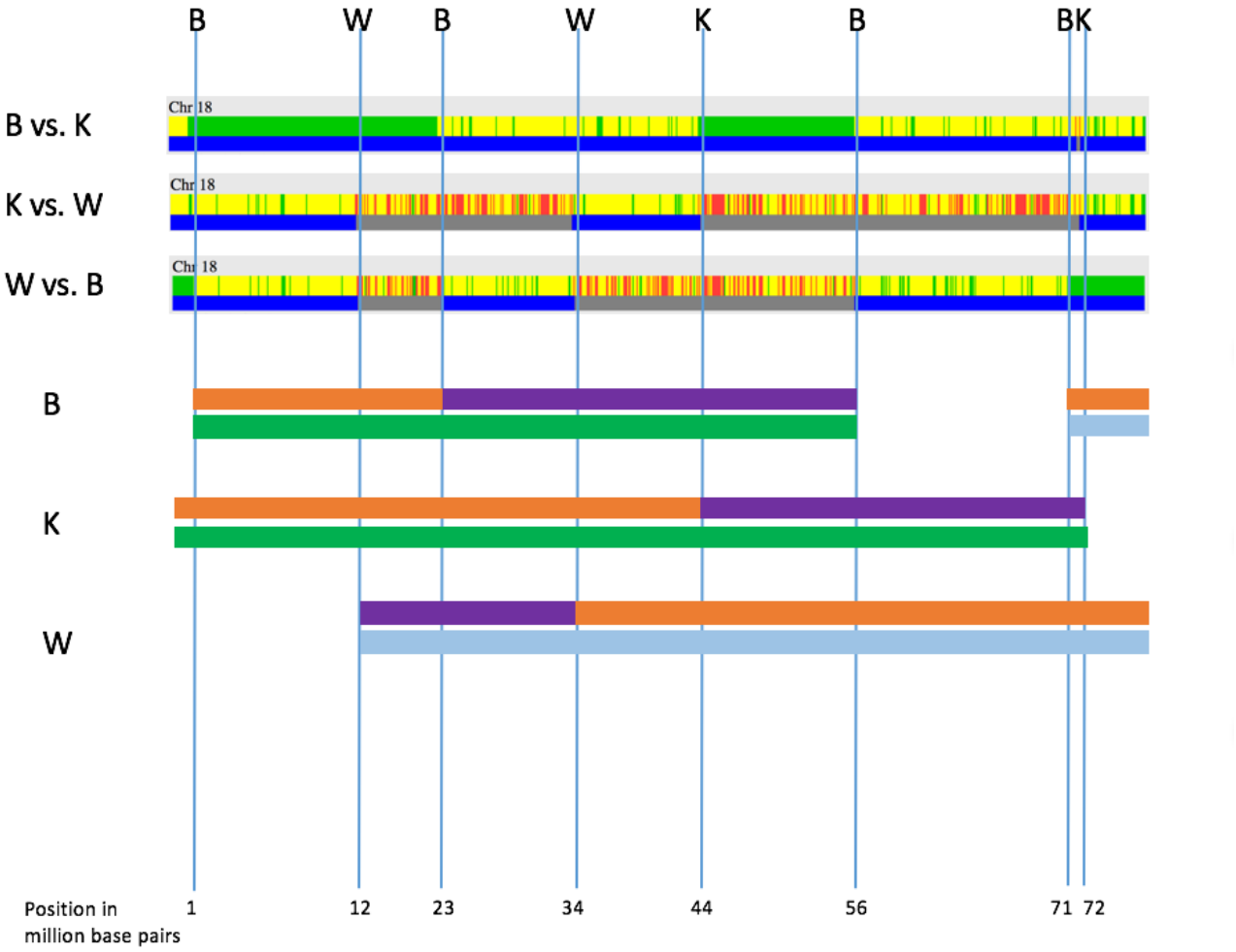


© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

14. The remaining colors are now ambiguous for B and K where more than one color could be chosen as a single match. It is the single matching areas for two comparisons that often cause this uncertainty. For example, between positions 56 and 71, B's purple could be extended from56 to 71, or orange could be extended to the left from 71 to 56. It is best not to continue if the configuration is unknown until other relatives are found to confirm the matching colors.

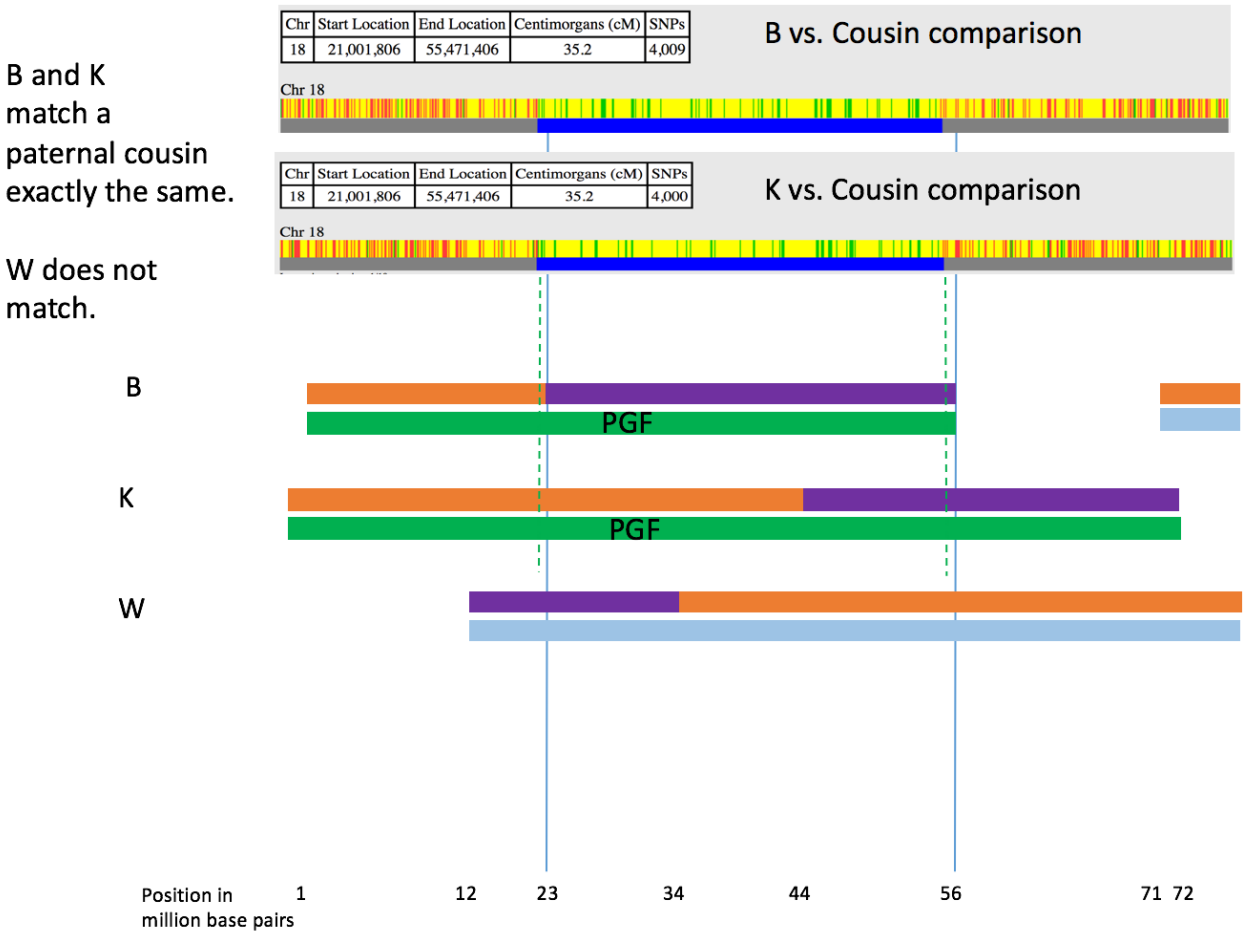
    However, we do know that the orange segment for W cannot be extended between 34 and 23 because K and W do not match between 12 and 34, so there is no uncertainty there. We can fit the W vs K matching by filling in the upper chromosome of W with the purple color while extending W's lower blue color from 34 down to 12.

    At this point we have taken the sibling matching data as far as we can without resorting to cousin matches. This reason for this roadblock is discussed above in the Numerical Approach. Without cousin matches we also can't assign the four colors to a specific grandparent, or even which of the colors are paternal or maternal. Four or five siblings are always preferred over three to minimize the ambiguities, but the demand has become great among genealogists to find a way to solve these puzzles with fewer siblings and more cousin matches.



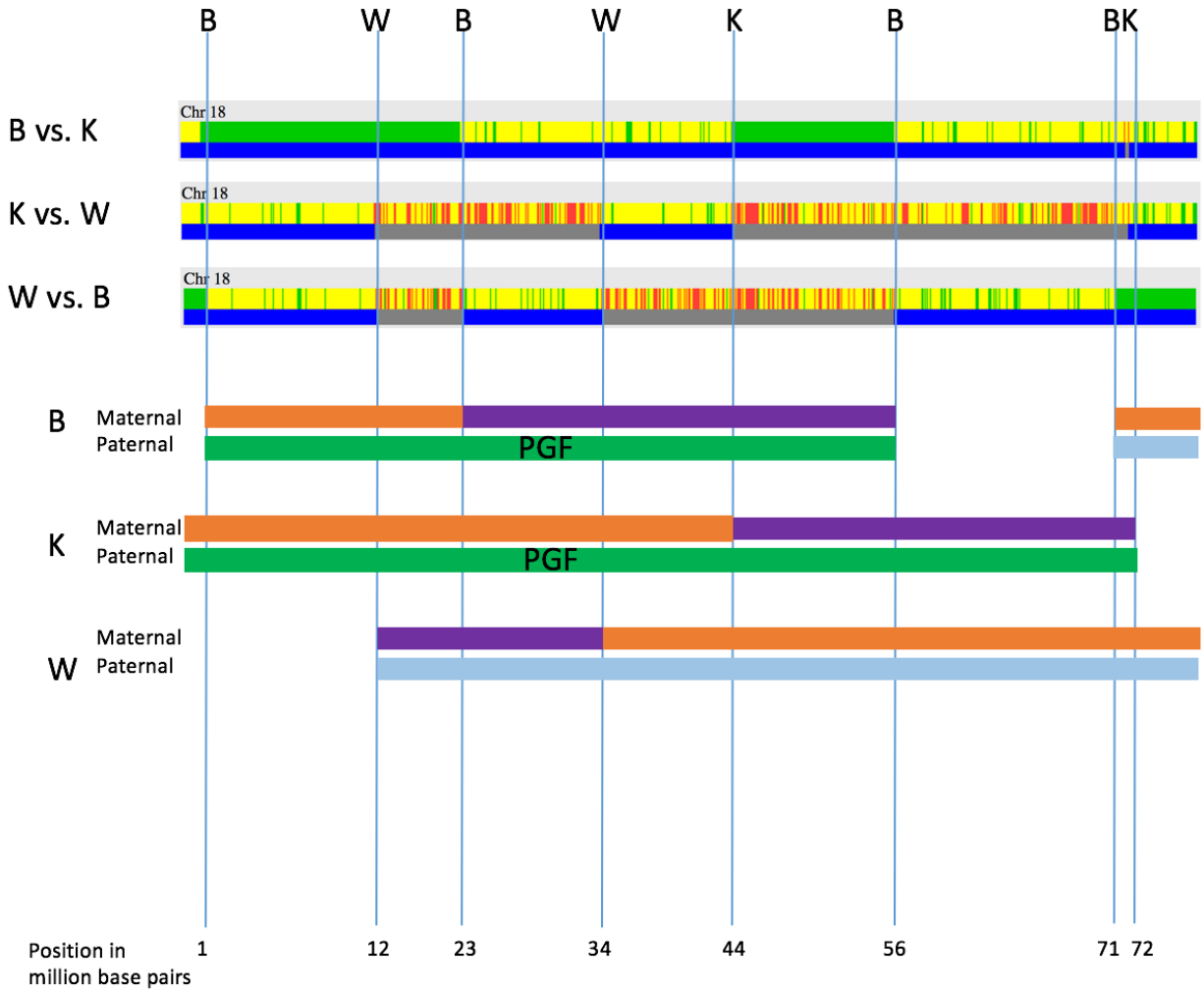
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

15. At GEDmatch, a paternal third cousin appeared who only matched B and K associated with the chromosome colored green above. The green single matching region can be labeled PGF (Paternal Grandfather) because out of the four grandparents, this cousin is only related to the PGF. This segment must fit within the borders passed down from the father. Each step is based on logic. The parents are not related to each other according to another GEDmatch tool, so only one parent can transmit the entire matching segment which must fit within the chromosome template. Think of a segment match with a cousin like a puzzle piece that must either be the same size or smaller than the match with a single color on a chromosome as shown below. The green regions can now be assigned to the PGF as shown below.



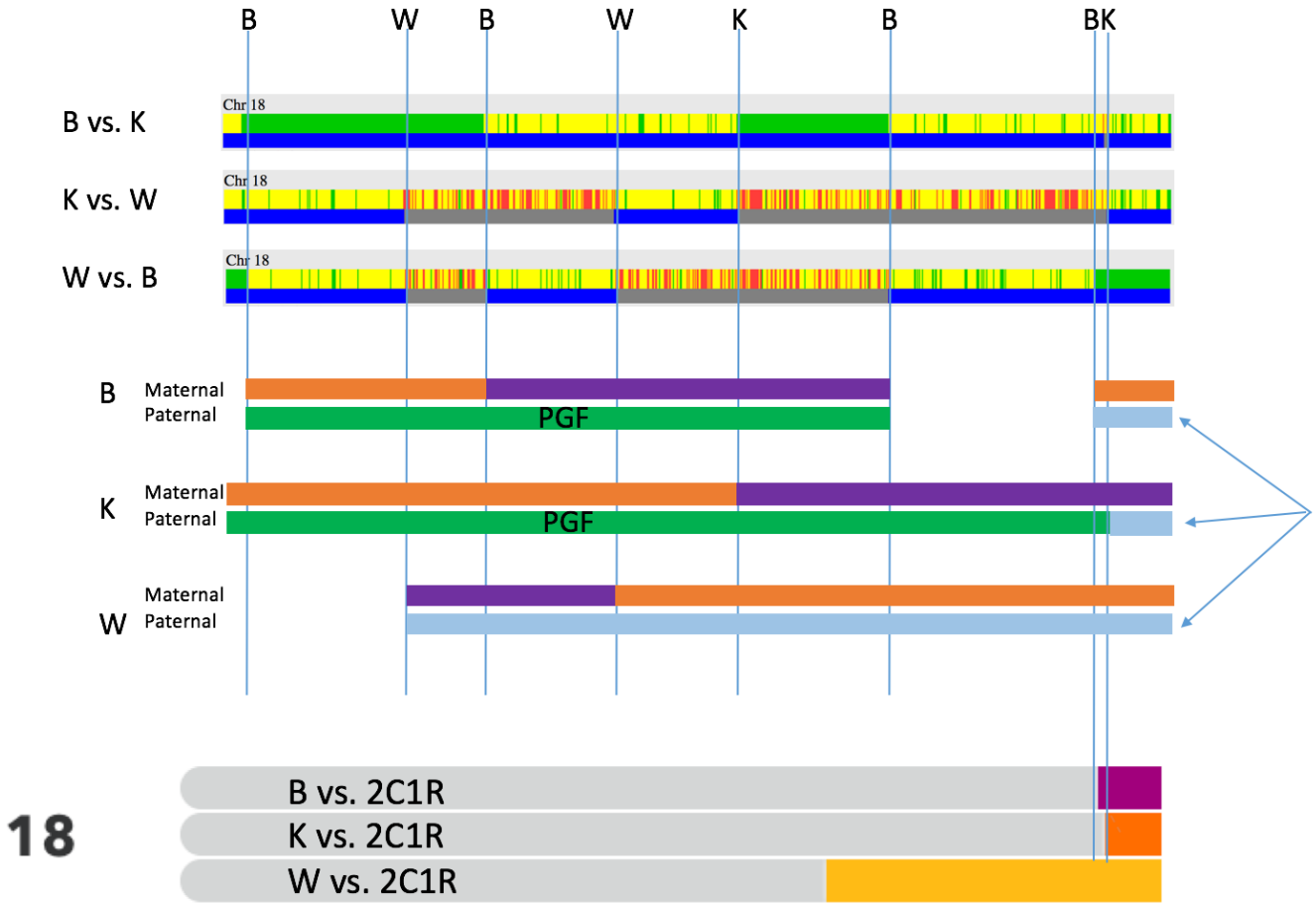
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

16. We also now know that the green-blue colors are paternal, while the orange-purple colors must be maternal as shown below:



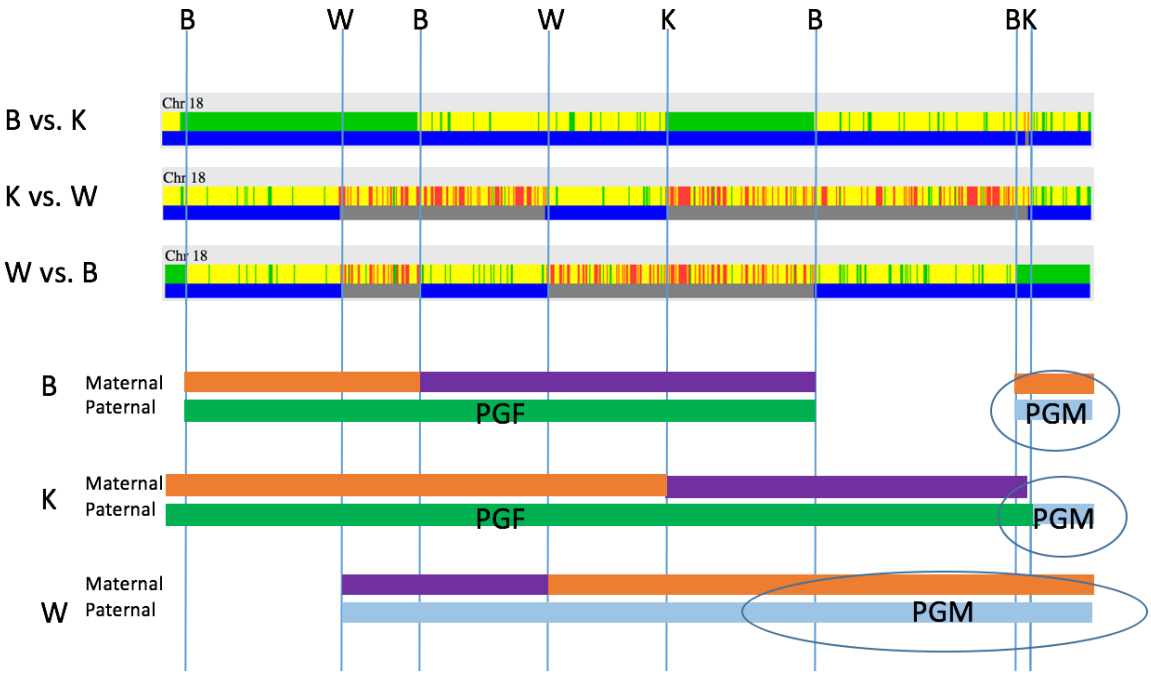
© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

17. A second cousin once removed related to the PGM (Paternal Grandmother) provided the needed proof for the rest of the paternal assignments. The start borders in the 23andMe comparison showed differences between the siblings that indicated B and K received close crossovers at 71-72 million, but not exactly in the same location.



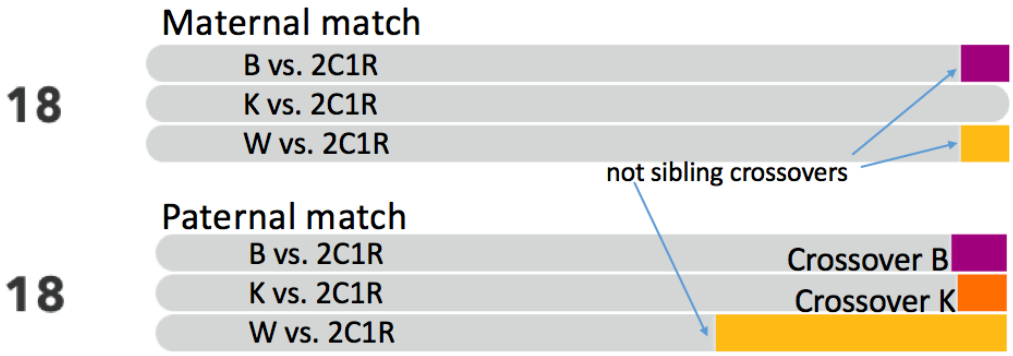
© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

18. Assign blue PGM to the last segment for K because the cousin not only relates to the paternal grandmother but also because the borders and segments must match within the paternal chromosome. Through process of elimination, the maternal line has no crossover point for K at the end and a maternal relative later confirmed the configuration. The last crossover for B also only involves the paternal blue chromosome and not the orange. Note that the Build 36 position numbers were no longer available at 23andMe so conversions may be necessary to match the GEDmatch start points. Currently the different builds are usually within a million base pairs of each other.



© 2024. This work is licensed under a
CC BY-NC-SA 4.0 license

19. Another cousin was identified who was only related to the MGF's side of the family and as expected did not match the cousin above. One group of chromosome segments fits with maternal orange and the other group fits with paternal blue. It is always helpful when cousins match your family on the same chromosome location but they do not match each other proving the maternal and paternal identity of the segments.



.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

20. A paternal uncle confirmed the first crossover at position 1 for B. All remaining maternal and paternal crossovers could be identified by either the process of elimination or by testing known relatives. Reconstruction of most of the grandparent contributions (represented by all four colors covering most of the chromosomes) was possible using no more than three siblings despite the fact that no direct line ancestors were still living.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Conclusions

Knowing the recombination patterns for a family group can be very useful in researching matches with unknown persons in the company and third-party databases. Using the approach described herein, we can map every part of the chromosomes of a group of siblings back to particular grandparents. A comparison of just one sibling with an unknown person will not tell us exactly which grandparent we are matching through, but by examining the pattern of matches of each sibling with the unknown person, and comparing with the phased chromosome diagrams, we can determine just which grandparent that the matching segment came through to the siblings.

The methods described here can be applied to each of the 22 autosomal chromosomes, and usually the X chromosome as well, to completely determine the way that recombination has occurred in a family group. Two approaches to recombination analysis, one more computational in nature, and the other more visual, have been described and illustrated.

Since these methods were developed, several third-party tools have appeared that make the process easier. For example, see below under Web Resources, Steven Fox. Also, the DNAPainter tool has gained popularity as an aid to chromosome mapping to more distant ancestors.

It has been somewhat surprising to us in analyzing several family groups that over several generations, the contributions, by great-great-grandparents for example, may vary widely, from 3% to 9% in one case, when an average of 6.25% would be expected. In the case of one particular small chromosome in the family of one of us (Athey), I received my father's maternal chromosome copy whole (which he got from his mother), and I passed it whole to my daughter, and she passed it whole to her son, all without any detectable recombination in four generational passages. There was another interesting finding in a grandson's recombination diagrams--there were many chromosomes where only three or four of my daughter's eight great-grandparents, contributed anything to the copy that she passed to her son.

In the analysis of several family groups using our method, we have confirmed the previous suggestions that males produce fewer crossovers than females in putting together the composite chromosome that is passed to offspring (Coop 2008). For example, in the four-sibling analysis presented above, there were two crossovers (total, in all four siblings) on the paternal side and nine crossovers on the maternal side. In the 22 pairs of chromosomes of these same four siblings, there was only one chromosome pair where there were more crossovers on the paternal side than the maternal side, and the excess was only one crossover. In the example of the visual approach, there were three crossovers on the paternal side and five on the maternal side.

Males also tend to put in crossovers near the beginning of chromosomes, say a location under 5 million, whereas females usually do not.

In summary, recombination is not a very regular or even process on any given chromosome in terms of the relative contributions of grandparents and great-grandparents. It only approximately approaches normality in the average of contributions of ancestors over all chromosomes.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Web Resources

Four-Sibling Excel Spreadsheet http://www.hprg.com/storage/Recomb-4-Siblings.xlsx

Instructions for using the 4-sibling spreadsheet http://www.hprg.com/storage/Instructions-4-Sib.docx

Three-Sibling Excel Spreadsheet http://www.hprg.com/storage/Recomb-3-Siblings.xlsx

Instructions for using the 3-sibling spreadsheet http://www.hprg.com/storage/Instructions-3-Sib.docx

GEDmatch http://www.gedmatch.com

David Pike's Utilities http://www.math.mun.ca/~dapike/FF23utils/

Blaine Bettinger's blog, The Genetic Genealogist, has a five-part series exploring the visual phasing technique, and this has been instrumental in bringing the technique to a wider audience. See: http://thegeneticgenealogist.com/2016/11/21/visual-phasing-an-example-part-1-of-5/

Steven Fox has developed an Excel-based program to do much of the time-consuming parts of the visual phasing process. A video presentation may be found at: https://vimeo.com/224877731/6ba212fa67. To access the Excel program you must join the Visual Phasing Working Group on Facebook and download it from that site. The Visual Phasing Working Group site provides many helpful suggestions, and feedback may be obtained on any problems that arise.

Conflicts of Interest

The author(s) declare no conflicts of interest and no commercial interests in the subjects covered by this study.

References

Athey TW (2010a) The numerical method was first presented by one of us (Athey) in September 2010 at the monthly meeting of the Northern Virginia Genetics Interest Group of the National Genealogical Society. A more developed and formalized version of this work was again presented to the same group in the Spring of 2016.

Athey TW (2010b) Phasing the chromosomes of a family group when one parent is missing. *Journal of Genetic Genealogy*, **6**(1).

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M (2008) High-Resolution Mapping of Crossovers Reveals Extensive Variation in Fine-Scale Recombination Patterns Among Humans. *Science*, **319**:1395-1398. DOI: 10.1126/science.1151851 (http://science.sciencemag.org/content/319/5868/1395.full)

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Johnston K (2015) "Visual Phasing of a Single Chromosome -- the Use of Crossover Lines." Presented on the Family Tree DNA Forum, "Family Finder Advanced Topics," 23 Jan 2015.

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table S1. Look-Up Table--Converting Segment Matching Patterns to Pairs of Chromosome Inheritance Patterns

Segment Matching Patterns							Chromosome Inheritance Patterns	
Sib1-Sib2	Sib1-Sib3	Sib1-Sib4	Sib2-Sib3	Sib2-Sib4	Sib3-Sib4		(either order)	
2	2	2	2	2	2		1	1
2	2	1	2	1	1		1	2
2	1	2	1	2	1		1	3
2	1	1	1	1	2		1	4
1	2	2	1	1	2		1	5
1	2	1	1	2	1		1	6
1	1	2	2	1	1		1	7
1	1	1	2	2	2		1	8
2	2	2	2	2	2		2	1
2	2	0	2	0	0		2	2
2	1	1	1	1	0		2	3
2	1	0	1	0	1		2	4
1	2	1	1	0	1		2	5
1	2	0	1	1	0		2	6
1	1	1	2	0	0		2	7
1	1	0	2	1	1		2	8
2	1	2	1	2	1		3	1
2	1	1	1	1	0		3	2
2	0	2	0	2	0		3	3
2	0	1	0	1	1		3	4
1	1	2	0	1	1		3	5
1	1	1	0	2	0		3	6
1	0	2	1	1	0		3	7
1	0	1	1	2	1		3	8
2	1	1	1	1	2		4	1
2	1	0	1	0	1		4	2
2	0	1	0	1	1		4	3
2	0	0	0	0	2		4	4
1	1	1	0	0	2		4	5
1	1	0	0	1	1		4	6
1	0	1	1	0	1		4	7
1	0	0	1	1	2		4	8
1	2	2	1	1	2		5	1
1	2	1	1	0	1		5	2
1	1	2	0	1	1		5	3
1	1	1	0	0	2		5	4
0	2	2	0	0	2		5	5
0	2	1	0	1	1		5	6
0	1	2	1	0	1		5	7

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

0	1	1	1	1	2		5	8
1	2	1	1	2	1		6	1
1	2	0	1	1	0		6	2
1	1	1	0	2	0		6	3
1	1	0	0	1	1		6	4
0	2	1	0	1	1		6	5
0	2	0	0	2	0		6	6
0	1	1	1	1	0		6	7
0	1	0	1	2	1		6	8
1	1	2	2	1	1		7	1
1	1	1	2	0	0		7	2
1	0	2	1	1	0		7	3
1	0	1	1	0	1		7	4
0	1	2	1	0	1		7	5
0	1	1	1	1	0		7	6
0	0	2	2	0	0		7	7
0	0	1	2	1	1		7	8
1	1	1	2	2	2		8	1
1	1	0	2	1	1		8	2
1	0	1	1	2	1		8	3
1	0	0	1	1	2		8	4
0	1	1	1	1	2		8	5
0	1	0	1	2	1		8	6
0	0	1	2	1	1		8	7
0	0	0	2	2	2		8	8

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

Table S2. Same as Table S1, but Sorted on Segment Matching Patterns for Ease of Look-Up

Segment Matching Patterns							Chromosome Inheritance Patterns	
Sib1-Sib2	Sib1-Sib3	Sib1-Sib4	Sib2-Sib3	Sib2-Sib4	Sib3-Sib4		(either order)	
0	0	0	2	2	2		8	8
0	0	1	2	1	1		7	8
0	0	1	2	1	1		8	7
0	0	2	2	0	0		7	7
0	1	0	1	2	1		6	8
0	1	0	1	2	1		8	6
0	1	1	1	1	0		6	7
0	1	1	1	1	0		7	6
0	1	1	1	1	2		5	8
0	1	1	1	1	2		8	5
0	1	2	1	0	1		5	7
0	1	2	1	0	1		7	5
0	2	0	0	2	0		6	6
0	2	1	0	1	1		5	6
0	2	1	0	1	1		6	5
0	2	2	0	0	2		5	5
1	0	0	1	1	2		4	8
1	0	0	1	1	2		8	4
1	0	1	1	0	1		4	7
1	0	1	1	0	1		7	4
1	0	1	1	2	1		3	8
1	0	1	1	2	1		8	3
1	0	2	1	1	0		3	7
1	0	2	1	1	0		7	3
1	1	0	0	1	1		4	6
1	1	0	0	1	1		6	4
1	1	0	2	1	1		2	8
1	1	0	2	1	1		8	2
1	1	1	0	0	2		4	5
1	1	1	0	0	2		5	4
1	1	1	0	2	0		3	6
1	1	1	0	2	0		6	3
1	1	1	2	0	0		2	7
1	1	1	2	0	0		7	2
1	1	1	2	2	2		1	8
1	1	1	2	2	2		8	1
1	1	2	0	1	1		3	5
1	1	2	0	1	1		5	3

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license

1	1	2	2	1	1		1	7
1	1	2	2	1	1		7	1
1	2	0	1	1	0		2	6
1	2	0	1	1	0		6	2
1	2	1	1	0	1		2	5
1	2	1	1	0	1		5	2
1	2	1	1	2	1		1	6
1	2	1	1	2	1		6	1
1	2	2	1	1	2		1	5
1	2	2	1	1	2		5	1
2	0	0	0	0	2		4	4
2	0	1	0	1	1		3	4
2	0	1	0	1	1		4	3
2	0	2	0	2	0		3	3
2	1	0	1	0	1		2	4
2	1	0	1	0	1		4	2
2	1	1	1	1	0		2	3
2	1	1	1	1	0		3	2
2	1	1	1	1	2		1	4
2	1	1	1	1	2		4	1
2	1	2	1	2	1		1	3
2	1	2	1	2	1		3	1
2	2	0	2	0	0		2	2
2	2	1	2	1	1		1	2
2	2	2	2	2	2		1	1

© 2024. This work is licensed under a CC BY-NC-SA 4.0 license