



Laboratory Protocol: Recombination Calculations by Branch Diagram

William D. Gilliland* and Michael G. Rosenbaum
DePaul University Department of Biological Sciences
2325 N. Clifton Ave.
Chicago, IL
60614

*Corresponding Author: wgillila@depaul.edu

Accepted for publication June 20, 2013

Citation:

Gilliland, William D. and Rosenbaum, Michael G. (2013). Recombination Calculations by Branch Diagrams. *Genetics Society of America Peer-Reviewed Education Portal (GSA PREP)*: 2013.002; [doi: 10.1534/gsaprep.2013.002](https://doi.org/10.1534/gsaprep.2013.002)

Abstract:

Genetic recombination can be difficult to understand because of the abstract nature of the experiments used to measure it. We have developed a method to use a branch diagram for analyzing recombination events. We show how to use this method to list all possible phenotypes for the progeny and identify which classes are recombinant, calculate the expected phenotype frequencies from known map distances, and use progeny count data to calculate map distances and interference. The branch diagram method provides visual cues for which classes are recombinant, as well as a structural guide to assist students with the mathematical analysis of recombination data.

Introduction and Background:

It has been 100 years since Alfred Sturtevant first showed that crossing over could be used to determine the order of genes on chromosomes, which he used to build the first genetic map in any organism (Sturtevant 1913). These experiments also revealed the existence of interference, a phenomenon where one recombination event inhibits other recombination events from occurring nearby, the mechanism of which is still being studied (for review, see Berchowitz and Copenhaver 2010). While recombination mapping of mutants has become somewhat obsolete in modern research, as whole genome sequencing is becoming a faster and more cost effective way to identify mutations of interest (Blumenstiel *et al* 2009), recombination is still a part of many genetics curricula. Analogously to how the study of springs and falling objects in physics are both historically important topics as well as tractable ways to demonstrate physical laws, recombination is a historically important part of genetics, as well as an accessible way for students to directly observe genetic processes in action.

One common experiment in genetics lab classes is to cross an organism (such as the fruit fly *Drosophila melanogaster*) that is heterozygous for multiple linked markers that can reveal recombination events. Just by counting a number of progeny, it is possible to estimate recombination rates between markers. We have found that many students struggle with the textbook method of analyzing the progeny from this experiment. Part of the problem is that in the textbook method (e.g. Pierce 2012), students must first exhaustively list all possible gametes, and then deduce which classes are recombinant and non-recombinant. This approach is rather abstract, and many students have difficulty with this first step as they lack a good conceptual framework to guide them. Without a solid understanding of how the recombinant classes arise, these students then have difficulty understanding how to calculate the expected number of recombinant progeny from published map distances, or how to calculate map distances from their recombination data.

To develop a more accessible method, we have taken the branch diagram concept (already widely used to analyze Mendelian crosses) and applied it to the analysis of recombination crosses. The key to this method is that each marker in the cross adds one

fork to the diagram, and a recombination event is represented by a *change of direction* at a fork. If movement through a fork changes direction (e.g. enter moving up, exit moving down) then there was a recombination in that interval. Conversely, traveling straight through a fork indicates that it did not recombine. This simple change-of-direction rule provides a structure for analyzing a recombination cross, which a number of students have found helpful in completing this type of analysis.

Method:

There are three tasks that are commonly carried out in analyzing a recombination experiment: (1) listing all possible phenotypes for the progeny and identifying which classes are recombinant, (2) calculating the expected phenotype frequencies from known map distances, or (3) using progeny counts to calculate map distances and interference. The branch diagram provides structural organization for all of these tasks. We developed this in a lab that uses a three-point cross, which is the minimum number of markers needed to determine their relative order. Although extension of this method to four or more markers is straightforward, we have focused on the three-point cross for this article.

The simplest scenario in a recombination experiment is when the linkage and order of the markers is known. To use a *Drosophila* example, if two fly stocks are crossed, one of which is wildtype while the other is homozygous for the recessive X-linked mutants *yellow* (*y*), *crossveinless* (*cv*) and *forked* (*f*), then in the female progeny all three mutant alleles will be linked on the same chromosome. The order of these markers is also known, with *cv* in the middle, so there are two intervals where recombination can occur: interval 1 between *y* and *cv*, and interval 2 between *cv* and *f*. (This does assume that only one parent is recombining, while the other parent is homozygous for all recessive alleles, so the complete genotype of the progeny can be determined.)

$$\frac{y \quad cv \quad f}{y^+ \quad cv^+ \quad f^+}$$

To begin, draw the branch diagram with the same number of forks as there are markers in the experiment; this experiment has three markers, so the diagram will have three forks, for a total of eight (2^3) tips to the diagram. Next, write down the parental chromosomes at the root of the diagram, one on top of the other. The first fork in the diagram represents the choice of the left-most marker; in this case, *y* versus *y*⁺. Then, each branch forks into two additional branches for every subsequent marker. In this case, the second fork represents the middle marker (*cv* versus *cv*⁺) and the third fork represents the third marker (*f* versus *f*⁺). For each fork, the two alternate alleles are written on the two straight segments after that fork. It is important to be consistent while filling out this diagram, as the upward branches must always get alleles from the

top chromosome ($y\ cv\ f$) while the downward branches must always get alleles from the bottom chromosome ($y^+\ cv^+\ f^+$) (Figure 1.)

With the branch diagram set up, the first task of enumerating all possible gametes can be completed by traveling from the first node of the diagram to each tip, noting the alleles on each segment along the way, and counting how many times a change of direction occurred. For example, the topmost tip is reached by going up at all three nodes, for a phenotype of $y\ cv\ f$. Since the direction never changed, this class is therefore non-recombinant. The second tip is reached by going up-up-down, which results in a $y\ cv\ f^+$ gamete, with a single recombination in the second interval. By following the same procedure for the remaining tips, all possible recombinant classes are enumerated, and which classes underwent 0, 1 or 2 recombination events are indicated (Figure 1.)

The second task in analyzing a recombination experiment is calculating the expected frequencies for the various classes using published map distances. A map distance between two markers is the proportion of all progeny that are expected to recombine between those markers. For example, the y - cv interval is 13.7 map units long (Marygold *et al.* 2013). As each map unit corresponds to a 1% recombination rate, the expected proportion of recombinant progeny for the first interval is 0.137, while the expected proportion of non-recombinant progeny is $1-0.137$, or 0.873. To use the branch diagram, these recombinant and non-recombinant proportions are written on each segment after each node, according to whether or not they represent a recombination event. The first fork represents which reciprocal product was randomly chosen, and therefore the two branches each have a proportion of 0.5. Then, each subsequent branch is assigned a value for the recombination proportion in that interval. In this example, at the second fork the two branches that changed direction are assigned 0.137, while the two branches that did not change direction are assigned 0.873. The cv - f interval is 43.0 map units long, so the branches after the third fork that did change direction are assigned 0.43 while the branches that did not change direction are assigned 0.57. Finally, to calculate the expected proportion of each gamete, follow the line from the base to each tip, multiplying the numbers encountered along the way (Figure 2). Expected progeny counts can then be obtained by multiplying these proportions by the total number of progeny observed.

The third task, calculating map distances from progeny counts, can be done by using the branch diagram to determine which phenotypes are non-recombinant (NR), single-recombinant in the first interval (SR1), single-recombinant in the second interval (SR2) and which are double-recombinant (DR). First, set up the branch diagram to get the correct order of classes. Next, write the observed progeny counts next to the phenotypes in the diagram. Finally, the recombinant proportions for the two intervals are calculated with the following formulae:

$$\text{Recombinant Proportion 1 (RP1)} = (SR1 + DR) / \text{Total}$$

$$\text{Recombinant Proportion 2 (RP2)} = (SR2 + DR) / \text{Total}$$

If desired, these proportions can be multiplied by 100 to convert them to map units.

Once the recombinant proportions the two intervals are known, the genetic interference can be calculated. Interference is a departure from independence of recombination events, and results in a reduction in the number of double recombinant progeny recovered. Interference can be calculated by the following formulae:

$$\text{Expected Doubles (ED)} = RP1 * RP2 * \text{Total}$$

$$\text{Coefficient of Coincidence (CoC)} = \text{Observed Doubles} / \text{Expected Doubles}$$

$$\text{Interference} = 1 - \text{CoC}$$

To work through an example, Table 1 presents data collected by undergraduates in a genetics laboratory course, which are in the same order as the branch diagram (Figure 1). The proportions calculated from the published map distances (Figure 2) are multiplied by the total, to generate the expected number of each phenotype (Table 1). The actual map distances in this experiment are calculated as:

$$RP1 = (10+9+2+4) / 224 = 0.112$$

$$RP2 = (15+37+2+4) / 224 = 0.286$$

Finally, the interference is be calculated as:

$$ED = 0.112 * 0.286 * 224 = 7.2$$

$$\text{CoC} = 6 / 7.2 = 0.83$$

$$\text{Interference} = 1 - 0.83 = 0.17$$

Why Does this Method Work?

The branch diagram helps analyze a recombination experiment in three ways. The first way is by providing a structured method to list all possible gametes. With the traditional method, students need to manually list all 2^N possible outcomes for their N markers, and they often have trouble seeing how each additional marker will double the number of possible outcomes, as well as methodically listing all of the permutations without missing any. The branch diagram helps this process by both providing a visual demonstration of the doubling, and by providing a method that will generate all possible permutations.

The second way the branch diagram can help is in visualizing the process of switching over from one chromosome to another. By relating recombination events to changes in the direction through the tree, visual learners can see what gametes arose from

recombination events, and the reciprocal recombination classes can be seen reflected across the midline.

The third way the branch diagram helps in studying recombination is it lets shows how the effects of tighter or weaker linkage relate to how the markers segregate. If there is no recombination, then the only gametes that will be produced are the ones along the completely straight outside branches, which are the original two chromosomes in the parent. Conversely, if there is no linkage, then the two branches at a node will have equal probability ($p = 0.5$). In this case, the branch diagram will be exactly equal to a Mendelian branch diagram, and the gametes will be produced in the same ratios predicted by Mendel's law of independent assortment. The fact that recombination rates max out at 50% also demonstrates why nonrecombinant classes of gametes should be the most abundant, while the double-recombinant classes should be the least abundant.

Other Applications:

A more difficult version of the three-point cross arises when the linkage or the order of the markers is not known, and is to be determined from experimental data. The branch diagram can also be used in this analysis. First, identify the two most abundant progeny classes, which should be reciprocal for all markers. These are the two non-recombinant gametes, which are written in the NR blanks. Next, identify the two least abundant phenotype classes, which should be the double recombinants. These two classes should both have the same marker change its association from the two non-recombinant classes; that marker is the one located in the middle. Write these two classes in the DR positions that match the first marker of the non-recombinant chromosome on that side of the diagram (e.g. $y\ cv^+ f$ would be the double that matches the $y\ cv\ f$ side, while $y^+ cv\ f^+$ would match the $y^+ cv^+ f^+$ side.) Finally, identify the two pairs of single recombinants, which should be about equally abundant as well as reciprocal products, and write those in the appropriate SR1 and SR2 slots.

Conclusion:

To summarize the branch diagram method:

1. Draw the diagram with as many forks as there are markers in the cross.
2. Write the two parental chromosomes at the root of the diagram, one on top of the other.
3. On the two branches after each fork, write the two alternate alleles for that fork. The allele from the top parental chromosome goes on the top branch, and the allele from the bottom chromosome goes on the bottom branch.
4. To enumerate all possible gametes, follow the branches from the root to each tip, recording the alleles and the number of direction changes on the path. Each change of direction is a recombination event.

5. To calculate the expected number of progeny from published map proportions, write the recombination proportions at each fork. The first fork is 0.5 , while each subsequent fork is (map proportion) on the change-of-direction branch and $(1-\text{map proportion})$ on the same-direction branch. Finally, multiply the proportions from the root to each tip to get the expected proportion of each gamete, and multiply those proportions by the total to get the expected progeny counts.
6. To estimate parameters from progeny counts, write the progeny counts in the same order as the branch diagram. Then use the formulae to calculate the recombination proportions for each interval, the expected number of doubles, and finally interference.

Genetic recombination is a challenging topic, in part because this process can be difficult to visualize. We have found that the branch diagram method gives students a more structured guide for analyzing this type of experiment, and provides visual cues as to which gametes have undergone recombination. We believe that integrating this approach into a genetics course should improve student learning outcomes.

Supplemental Material:

We have included blank branch diagrams for 2, 3 and 4 marker recombination crosses as supplemental material. These worksheets have step-by-step instructions for applying the method when the order and linkage of the markers is known.

Works Cited:

Sturtevant, A.H., 1913 *The Linear Arrangement of Six Sex-Linked Factors in Drosophila, as Shown by Their Mode of Association*. J. of Experimental Zoology 14:43-59.

Berchowitz, L.E., G.P. Copenhaver, 2010 *Genetic interference: Don't Stand So Close to Me*. Curr. Genomics 11(2):91-102

Blumenstiel, J.P., A.C. Noll, J.A. Griffiths, A.G. Perera, K.N. Walton, W.D. Gilliland, R.S. Hawley and K. Staehling-Hampton, 2009 *Identification of EMS-induced Mutations in Drosophila melanogaster by Whole-Genome Sequencing*. Genetics 182(1):25-32.

Pierce, B., 2012 *Genetics: A Conceptual Approach*. W.H. Freeman and Co., New York.

Marygold, S.J., P.C. Leyland, R.L. Seal, J.L. Goodman, J.R. Thurmond, V.B. Strelets, R.J. Wilson and the FlyBase Consortium, 2013 *Flybase: Improvements to the Bibliography*. Nucleic Acids Res. 41(D1):D751-757.

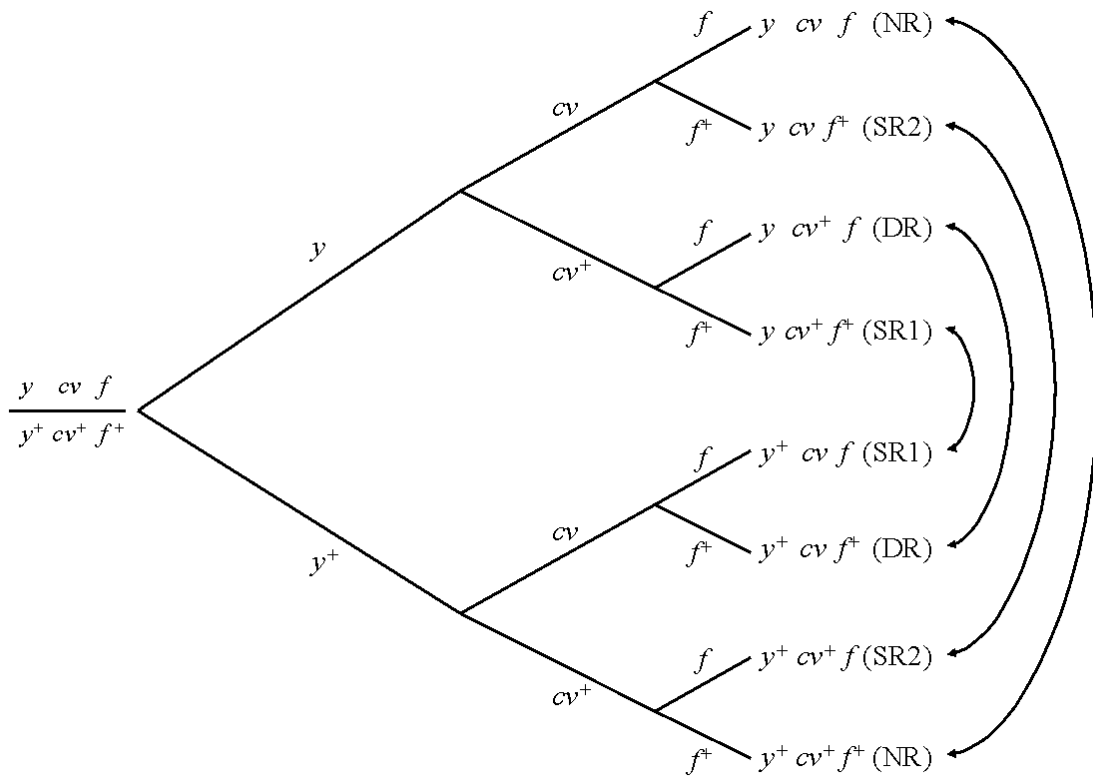


Figure 1: Setting Up the Branch Diagram. First the two parental chromosomes are written at the root. There is one fork in the diagram for each of the three markers, resulting in 8 tips. Next, the alternate alleles are written on the two branches after each fork, writing the top chromosome allele on upward branches, and the bottom chromosome allele on downward ones. The possible gametes are found by tracing from root to tip, noting the alleles on each branch. Each change in direction is a recombination event, which defines the classes. NR = Non-recombinant. SR1 = Single Recombinant, Interval 1. SR2 = Single Recombinant, Interval 2. DR = Double Recombinant. Note that the reciprocal products of recombination are reflected across the diagram (curved arrows.)

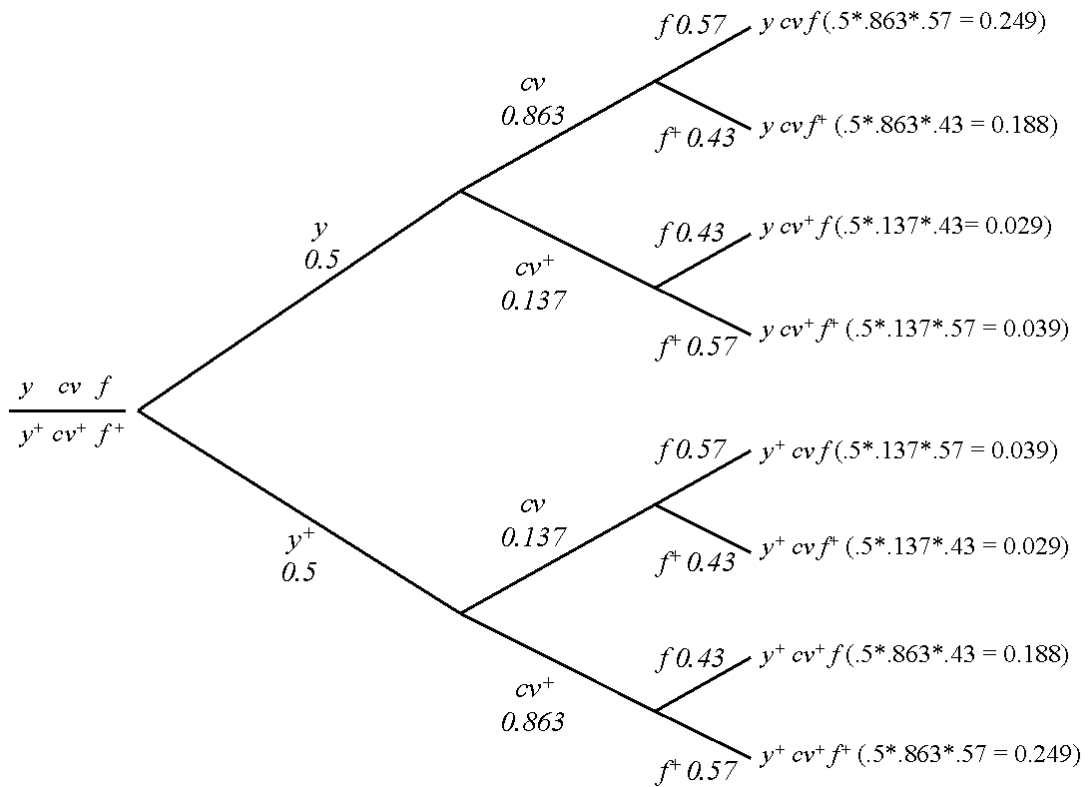


Figure 2: Calculating Expected Values. Each branch is assigned a proportion based on known map distances and whether they recombined or not. For the first fork, the two chromosomes are expected to be equally abundant, so are assigned a proportion of 0.5. Next, the proportion based known map distances is assigned to branches that change direction, and one minus that is assigned to branches that did not change direction. (e.g. the y - cv interval is 13.7 map units, so the change-of-direction branches get 0.137, while the other branches get 0.863). Finally, the numbers along the path from base to tip are multiplied, to get the expected proportion of that gamete. Note the reciprocal products of recombination have the same expected values.

Phenotype	Class	Number Observed	Number Expected
<i>y cv f</i>	NR	60	$0.249 \cdot 224 = 55.8$
<i>y cv f^+</i>	SR2	21	$0.188 \cdot 224 = 42.1$
<i>y cv^+ f</i>	DR	2	$0.029 \cdot 224 = 6.5$
<i>y cv^+ f^+</i>	SR1	10	$0.039 \cdot 224 = 8.7$
<i>y^+ cv f</i>	SR1	9	$0.039 \cdot 224 = 8.7$
<i>y^+ cv f^+</i>	DR	4	$0.029 \cdot 224 = 6.5$
<i>y^+ cv^+ f</i>	SR2	37	$0.188 \cdot 224 = 42.1$
<i>y^+ cv^+ f^+</i>	NR	81	$0.249 \cdot 224 = 55.8$

Table 1. Class data. Students in a genetics class phenotyped 224 progeny from a cross of *y cv f*/*y^+ cv^+ f^+* females crossed to *y cv f*/Y males, which have been written in the order from the branch diagram. This indicates the counts of non-recombinants (NR), single recombinants in the two intervals (SR1, SR2) and double recombinants (DR). The proportion of each gamete from Figure 2 is multiplied by the total to get the number expected. The map distances for each interval can also be calculated by summing the SR and DR flies for that interval, divided by the total (e.g. the *y-cv* interval is $(10+9+2+4)/224 = 0.112$ or 11.2 map units.)

Supplemental Material

To provide sample data sets for use with this method, we reproduce some of the two and three marker crosses from Table 8 of A.H. Sturtevant's original 1913 publication of the first genetic maps in any organism. We have translated the genotypes into modern nomenclature, and present the progeny in the same order used by the branch diagram method. In Sturtevant's crosses, only male progeny were informative for these X-linked traits, as recombining females were not mated to males that carried all recessive markers. Therefore, we have omitted the male parents from these crosses. Note that most of these data show a slight viability effect that reduces the number of mutant alleles. This tends to be more pronounced when a genotype carries a larger number of mutant alleles.

Known map positions, from <http://flybase.org> (Marygold *et al* 2013):

<i>yellow (y)</i>	0
<i>white (w)</i>	1.5
<i>crossveinless (cv)</i>	13.7
<i>vermillion (v)</i>	33.0
<i>miniature (m)</i>	36.2
<i>forked (f)</i>	56.7

Cross 1: *yellow* and *white* (*white-eosin allele*¹, *w^e*)

Recombining Parent: $\frac{y^+ w^e}{y^- w^+}$

Progeny:	NR	$y^+ w^e$	176
	SR	$y^+ w^+$	0
	SR	$y^- w^e$	2
	NR	$y^- w^+$	195

Cross 2: *yellow* and *vermillion* (*cis* configuration)

Recombining Parent: $\frac{y^+ v^+}{y^- v^-}$

Progeny:	NR	$y^+ v^+$	59
	SR	$y^+ v^-$	16

¹ When the original work was published, it was thought that *white* and *eosin* were different genes.

SR	$y^- v^+$	24
NR	$y^- v^-$	33

Cross 3: yellow and vermilion (trans configuration)

Recombining Parent: ♀ $\frac{y^+ v^-}{y^- v^+}$

Progeny:	NR	$y^+ v^-$	149
	SR	$y^+ v^+$	54
	SR	$y^- v^-$	41
	NR	$y^- v^+$	119

Cross 4: yellow and miniature

Recombining Parent: ♀ $\frac{y^- m^+}{y^+ m^-}$

Progeny:	NR	$y^- m^+$	82
	SR	$y^- m^-$	48
	SR	$y^+ m^+$	51
	NR	$y^+ m^-$	89

Cross 5: white-eosin, vermilion and miniature

Recombining Parent: ♀ $\frac{w^+ v^- m^+}{w^e v^+ m^-}$

Progeny:	NR	$w^+ v^- m^+$	109
	SR2	$w^+ v^- m^-$	3
	DR	$w^+ v^+ m^+$	1
	SR1	$w^+ v^+ m^-$	49
	SR1	$w^e v^- m^+$	53
	DR	$w^e v^- m^-$	0
	SR2	$w^e v^+ m^+$	8
	NR	$w^e v^+ m^-$	85

Cross 6: yellow, crossveinless, vermilion, and forked

The final cross was conducted by crossing mutant females (a stock with an X chromosome marked with $y\ cv\ v\ f$) to wildtype males (Oregon-R), then allowing the F1 males and females to mate in bottles for 5 days. A large sample was collected to attempt to get triple recombinants, a single one of which was observed. Adults were then discarded, and progeny were counted until day 18. F2 males and females were both scored, but not recorded separately.

$$\text{Parents: } \frac{y^+ \ cv^+ \ v^+ \ f^+}{y^- \ cv^- \ v^- \ f^-} \times \frac{y^- \ cv^- \ v^- \ f^-}{Y} \text{ } \begin{matrix} \text{♀} & & \text{♂} \end{matrix}$$

Progeny:

NR	++++	1174
SR3	+++f	343
DR23	++v+	56
SR2	++vf	317
DR12	+cv++	12
TR	+cv+f	0
DR13	+cvv+	32
SR1	+cvvf	120
SR1	y+++	187
DR13	y++f	28
TR	y+v+	1
DR12	y+vf	13
SR2	ycv++	270
DR23	ycv+f	19
SR3	ycvv+	240
NR	ycv vf	540

(For clarity, superscripts have been omitted. Instead, + denotes wildtype phenotype for that trait, while the gene symbol denotes a mutant phenotype.)