Solution to Rosalind problem IPRB: Mendel's First Law*

Susannah Go

November 15, 2015

Suppose that a population contains k homozygous dominant organisms, m heterozygous organisms, and n homozygous recessive organisms. Assume that any two organisms can mate. Let p be the size of the population, so $p = k + m + n$. Let A denote the event that two randomly selected mating organisms yield an offspring with the dominant phenotype. Let $B_1, B_2, \ldots B_6$ denote the following events:

B_1 : both parents are homozygous dominant
B_2 : one parent is homozygous dominant, the other is heterozygous
B_3 : one parent is homozygous dominant, the other is homozygous recessive
B_4 : both parents are heterozygous
B_5 : one parent is heterozygous, the other is homozygous recessive
B_6 : both parents are homozygous recessive

Then $P(A)$ may be computed by conditioning on the alleles of the parents:

$$P(A) = \sum_{i=1}^{6} P(A \mid B_i)P(B_i) \tag{1}$$

Punnett squares are used to find the quantities $P(A \mid B_i)$. Using Y and y as alleles of the parents, and using a checkmark to mark offspring which will have the dominant phenotype, the Punnett squares are shown in Table 1. Based on the squares,

$$\begin{aligned} P(A \mid B_1) = P(A \mid B_2) = P(A \mid B_3) &= 1 \\ P(A \mid B_4) &= 0.75 \\ P(A \mid B_5) &= 0.5 \\ P(A \mid B_6) &= 0 \end{aligned}$$

*http://rosalind.info/problems/iprb/

	Y	Y
Y	✓	✓
Y	✓	✓

(a) B_1

	Y	y
Y	✓	✓
Y	✓	✓

(b) B_2

	y	y
Y	✓	✓
Y	✓	✓

(c) B_3

	Y	y
Y	✓	✓
y	✓	

(d) B_4

	y	y
Y	✓	✓
y		

(e) B

	y	y
y		
y		

(f) B_6

Table 1: Punnett squares for $B_1, B_2, \ldots B_6$; ✓denotes the dominant phenotype

Now we must calculate $P(B_i)$ for $i = 1, 2, \ldots, 6$. Recall that the parent organisms are randomly chosen from the population, so the probability of choosing two homozygous dominant organisms is $P(B_1) = \left(\frac{k}{p}\right)\left(\frac{k-1}{p-1}\right)$. Similar calculations are used for $P(B_4)$ and $P(B_6)$:

$$P(B_4) = \left(\frac{m}{p}\right)\left(\frac{m-1}{p-1}\right)$$
$$P(B_6) = \left(\frac{n}{p}\right)\left(\frac{n-1}{p-1}\right)$$

Now let us consider $P(B_2)$. There are two ways for event B_2 to occur: either the first parent is homozygous dominant and the second parent is heterozygous, or vice versa. So $P(B_2) = \left(\frac{k}{p}\right)\left(\frac{m}{p-1}\right) + \left(\frac{m}{p}\right)\left(\frac{k}{p-1}\right)$. Similar calculations are used for $P(B_3)$ and $P(B_5)$:

$$P(B_3) = \left(\frac{k}{p}\right)\left(\frac{n}{p-1}\right) + \left(\frac{n}{p}\right)\left(\frac{k}{p-1}\right)$$
$$P(B_5) = \left(\frac{m}{p}\right)\left(\frac{n}{p-1}\right) + \left(\frac{n}{p}\right)\left(\frac{m}{p-1}\right)$$

We have all the quantities we need, and substituting them all into Equation 1 yields a rather lengthy equation:

$$\begin{aligned} P(A) = 1\left(\frac{k}{p}\right)\left(\frac{k-1}{p-1}\right) + 1\left[\left(\frac{k}{p}\right)\left(\frac{m}{p-1}\right) + \left(\frac{m}{p}\right)\left(\frac{k}{p-1}\right)\right] + \\ 1\left[\left(\frac{k}{p}\right)\left(\frac{n}{p-1}\right) + \left(\frac{n}{p}\right)\left(\frac{k}{p-1}\right)\right] + 0.75\left(\frac{m}{p}\right)\left(\frac{m-1}{p-1}\right) + \\ 0.5\left[\left(\frac{m}{p}\right)\left(\frac{n}{p-1}\right) + \left(\frac{n}{p}\right)\left(\frac{m}{p-1}\right)\right] + 0\left(\frac{n}{p}\right)\left(\frac{n-1}{p-1}\right) \end{aligned}$$

This can be simplified, of course:

$$\begin{aligned} P(A) &= \frac{k(k-1)}{p(p-1)} + \frac{2km}{p(p-1)} + \frac{2kn}{p(p-1)} + \frac{3m(m-1)}{4p(p-1)} + \frac{mn}{p(p-1)} \\ &= \frac{4k(k-1) + 3m(m-1) + 4(2km + 2kn + mn)}{4p(p-1)} \end{aligned}$$