

Quantitative Traits

Preliminaries

If you are not already familiar with the structure of these exercises, read the Introduction first.

Note

Reminder: Save your work regularly.

Important

If you are using a Mac, we recommend that you use either Chrome or Firefox to complete these exercises. Some of the default settings in Safari prevent these exercises from running.

Contact information

If you have questions about these exercises, please contact Dr. Kevin Middleton (middletonk@missouri.edu) or drop by Tucker 224.

Learning objectives

The learning objectives for this exercise are:

- Contrast polygenic traits with Mendelian traits
- Demonstrate how quantitative traits with continuous-valued phenotypic measures result from the combined effects of many different genes
- Describe the process by which many genes can each contribute a small amount to a measurable phenotype

Contrasting Mendelian traits and polygenic traits

The first phenotypes that you learned about as well as those described in the first set of exercises (Transmission of Genetic Information) were Mendelian traits. In Mendelian traits, a single gene is responsible for a single trait. In this context, you also learned about dominant and recessive alleles (and their variations), which lead to different observable phenotypes.

This set of exercises focuses on phenotypes that are determined by multiple genes: **polygenic traits**. Polygenic traits often (but not always) can be measured on a numeric scale and are thus referred to as **quantitative traits**. The two main types of quantitative traits are:

- *Meristic traits*: Traits that take on integer values such as the number of peas in a pod. 3, 4, and 5 are all possible values, but a pod can't have 3.5 peas.
- *Continuous-valued traits*: Traits that can take on any number on the number line. For example lengths and weights can have any value depending on the scale (e.g., grams, millimeters, etc.)

One major difference between Mendelian and polygenic traits is that we have to alter our thinking about alleles. We will replace the dominant vs. recessive dichotomy that was useful for Mendelian traits. For polygenic traits, we will start to simply consider alternative alleles at a single locus in the genome. We'll come back to this idea later.

For now, let's think about all the different ways that alleles from different genes can be combined with one another when gametes are formed.

Counting the ways that alleles can combine

The 1:2:1 genotypic ratio and 3:1 phenotypic ratio in the heterozygote cross example for a dominant Mendelian trait (Figure 1) represent theoretical probabilities for the distributions of genotypes and phenotypes.

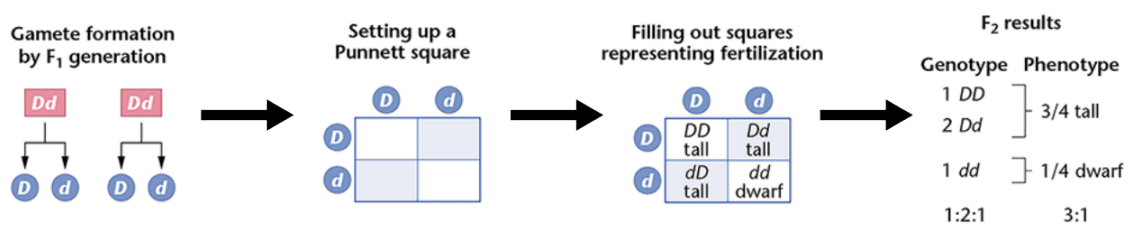


Figure 1: The process of gamete formation leads to predictable genotypic and phenotypic proportions in the F₂ generation. Image modified from Klug et al. (2019)

Higher level crosses, with two and three genes, can be done manually, but very quickly you will find that keeping track of all the combinations becomes very challenging (Figure 2).

	RYS	RYs	RyS	Rys	rYS	rYs	ryS	rys
RYS	RRYYSS	RRYYSs	RRYySS	RRYySs	RrYYSS	RrYYsS	RrYySS	RrYySs
RYs	RRYYSs	RRYYss	RRYySs	RRYyss	RrYYsS	RrYYss	RrYySs	RrYyss
RyS	RRYySS	RRYySs	RRyySS	RRyySs	RrYySS	RrYySs	RryySS	RryySs
Rys	RRYySs	RRYyss	RRyySs	RRyyss	RrYySs	RrYyss	RryySs	Rryyss
rYS	RrYYSS	RrYYsS	RrYySS	RrYySs	rrYYSS	rrYYsS	rrYySS	rrYySs
rYs	RrYYsS	RrYYss	RrYySs	RrYyss	rrYYsS	rrYYss	rrYySs	rrYyss
ryS	RrYySS	RrYySs	RryySS	RryySs	rrYySS	rrYySs	rryySS	rryySs
rys	RrYySs	RrYyss	RryySs	Rryyss	rrYySs	rrYyss	rryySs	rryyss

Figure 2: Punnett square for a trihybrid cross in peas. 64 combinations of alleles result in 6 possible phenotypes in a 1:6:15:20:15:6:1 ratio for three Mendelian traits.

Punnett squares with more than 3 genes are extremely difficult. For example, a cross with 4 genes requires 256 combinations, and 1,024 combinations are required for 5 genes (a 32 x 32 grid).

Fortunately, rather than keeping track of all these combinations manually, we can just calculate them directly using a little bit of math.

Flipping coins

Before we think about combining many alleles, let's consider something simpler: flipping coins. Imagine that you flip a coin twice. Each flip is *independent* of the other (i.e., heads or tails on one flip does not predispose the next flip to be heads or tails or the next).

Because each flip can result in either heads or tails, with two coin flips there are four possibilities:

1. Heads, Heads
2. Heads, Tails
3. Tails, Heads
4. Tails, Tails

To get two heads or two tails, both coin flips have to be the same. But to get one head and one tail, there are two possibilities:

- Heads, Tails
- Tails, Heads

Because the sequence doesn't matter, both result in one of each (head and tail). If we then think about adding up the possible sets of results, we have three possibilities but 4 ways to arrive at them:

1. 2 Heads (1 way)
2. 1 Head, 1 Tail (2 ways)
3. 2 Tails (1 way)

If you look back up at the monohybrid cross above, you will find that there is 1 DD, 2 Dd, and 1 dd. This 1:2:1 genotypic ratio is the same as for our coin flipping example.

All the different ways that you can arrive at a count of heads from a set of coin tosses are represented by a number called the *Binomial Coefficient*. The equation for the binomial coefficient for the number of heads from a set of coin flips is:

$$\frac{Flips!}{Heads! \times (Flips - Heads)!}$$

Those exclamation points (!) are the factorial function. For example, $3! = 3 \times 2 \times 1 = 6$.

If we plug in the numbers for 1 Head from 2 Flips:

$$\frac{2!}{1! \times (2 - 1)!}$$

which reduces to:

$$\frac{2 \times 1}{1 \times (1)!}$$

which is just $\frac{2}{1}$ or 2. Thus, there are 2 ways to get 1 head from 2 coin flips according to the binomial coefficient, just like we figured out manually.

At this point, hopefully you are convinced that we can use the binomial coefficient and some math to determine the number of ways that genes can combine and that they will match up with the counts that we would get by hand.

We can calculate the binomial coefficient directly using the `choose()` function. Run the code below to confirm that there are 1, 2, and 1 ways to get different genotypes. n represents the number of alleles (2 times the number of genes, because each individual receives 2 copies of each gene). k represents the number of D alleles that any one individual receives.

```
choose(n = 2, k = 0) # dd
choose(n = 2, k = 1) # Dd or dD
choose(n = 2, k = 2) # DD
```

```
[1] 1
[1] 2
[1] 1
```

Feel free to change the values for n (the number of alleles) and k (the number of D alleles) and try out some additional combinations. In general, n should be greater than or equal to k , or you will have zero ways to get a particular combination.¹

From coin flips to independent alleles

Recall that we stipulated that coin flips are independent of one another. The same process happens when gametes are formed: alleles are distributed independently of one another.

We can replace heads and tails, with different alleles like D and d above. One difference is that, for quantitative traits, we usually just consider alternate alleles at a certain position in the genome.

For this example, we will call these alternate alleles A and T. For a single gene, any individual could have AA, AT, or TT. Figure 3 shows these combinations.

¹ As a side note, the binomial coefficient calculation is also used to calculate probabilities for lotteries (the Powerball odds of winning is 1 in $\text{choose}(69, 5) * \text{choose}(26, 1)$) and for poker hands (the odds of a royal flush in 5-card draw is 4 in $\text{choose}(52, 5)$ or equivalently one in $\text{choose}(52, 5) / 4$), among other games of chance. In fact, much of the basics of probability were originally developed in the 17th and 18th centuries to try to understand (and cheat at) various forms of gambling.

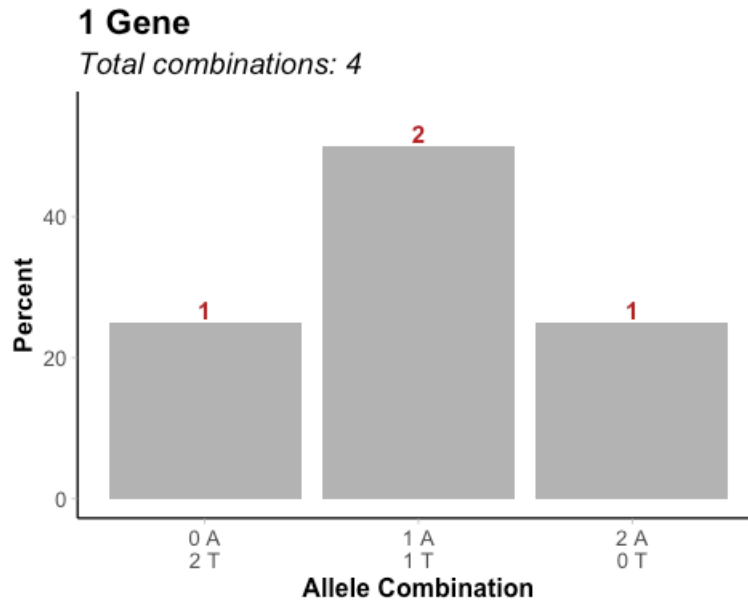
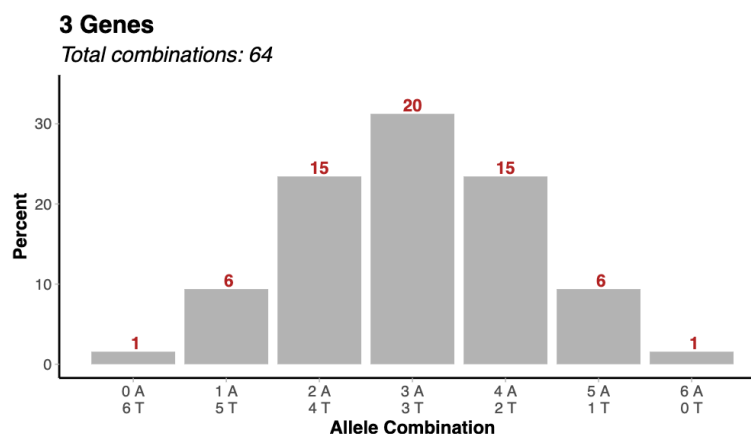


Figure 3: Allelic combination plot for 1 gene. There is a 1:2:1 (TT:AT:AA) ratio of two alternate alleles at a single position in the genome. There are 4 possible combinations (just like for two coin flips). The counts of each set of combinations are shown with the red number at the top of the bars. The y-axis shows the relative percentage of each.

If you run the code below, it will generate a figure like the one above, but for all the combinations for a trihybrid cross with 3 genes, like the Punnett square example shown above (Figure 2). Notice how the grid above is 8 by 8, giving 64 possible combinations. The total combinations below is also 64 (but with a lot less manual accounting for all the possible combinations).

```
genotypes_plot(n_genes = 3)
```



Because there are three genes, there are 6 possible alleles to be either A or T. Something to keep in mind is that when there is more than one gene, the A's and T's are in different genes, so each is independent of the others. In reality, the A's

and T's could be any of A, T, G, and C. The math works out the same, but the accounting is easier if we just think of them as all either A or T. Very often geneticists don't even worry about what the nucleotide is when thinking about quantitative traits. They will instead only think about which allele is more common at a given position in the genome (the "major allele") and which is less common (the "minor allele").

The `choose()` function below will calculate the number of way to get 3 A and 3 T.

```
choose(n = 6, k = 3)
```

```
[1] 20
```

Change the value of k to find the number of ways to have 0, 1, 2, 4, 5, and 6 A. Check them against the plot above.

For example, at these three loci, the 6 ways to get 1 A and 5 T are:

1. locus 1: AT, locus 2: TT, locus 3: TT
2. locus 1: TA, locus 2: TT, locus 3: TT
3. locus 1: TT, locus 2: AT, locus 3: TT
4. locus 1: TT, locus 2: TA, locus 3: TT
5. locus 1: TT, locus 2: TT, locus 3: AT
6. locus 1: TT, locus 2: TT, locus 3: TA

We can calculate all the ways for 0 through 6 A alleles (0:6) and then add them up with `sum()`:

```
(ways <- choose(n = 6, k = 0:6))  
sum(sum(ways))
```

```
[1] 1 6 15 20 15 6 1  
[1] 64
```

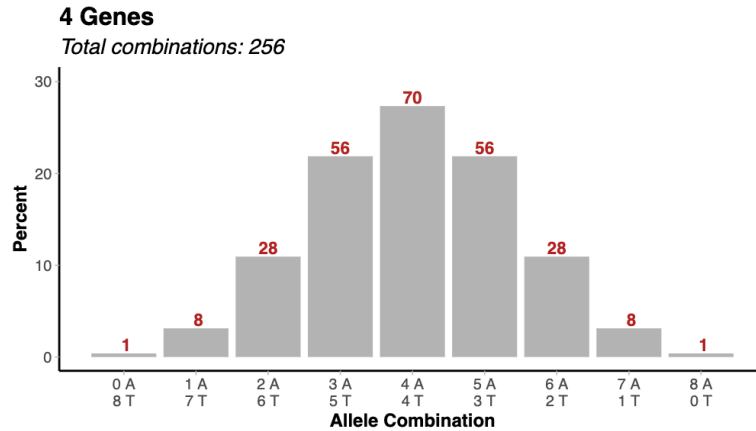
This is where the 64 total combinations in the figure come from. Now let's move to more genes.

Predict what the distribution of possible genotype combinations would be when four genes are involved. In general, what will the shape of the distribution look like? Will there be more or less total possible ways than for 3 genes? How many more or less?

The pattern will be similar, but with more combinations (more ways to get to 4 A and 4 T). Perhaps a lot more, because the number is increasing rapidly just from 1 gene to 3 genes.

Execute the code block below to generate the plot.

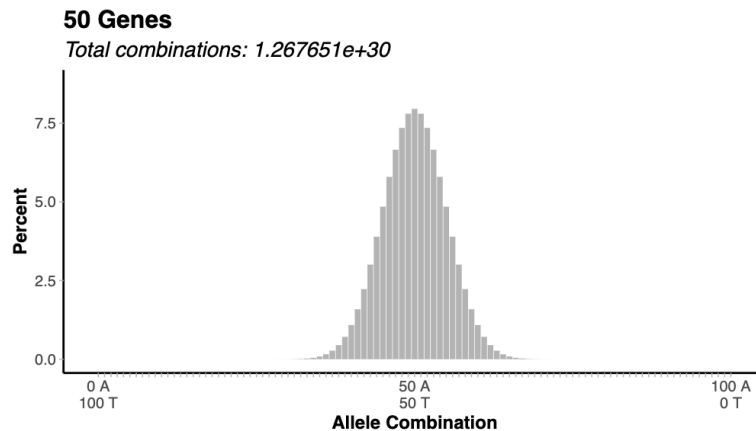
```
genotypes_plot(n_genes = 4)
```

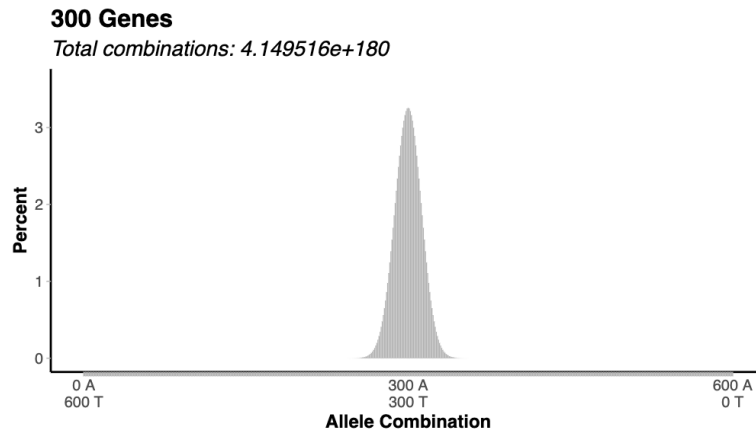


How does the plot compare to your prediction?

It goes from 64 to 256 combinations (over 4 times as many) for 4. The shape is pretty similar, with a single peak in the center. As the number increases most of the combinations center on an even split of A and T, centered on the number of genes.

Return to the code block above and continue increasing the number of genes: 5, 6, 7, 8, 10, 15, 20, 50, 100, 200, 300.² Run the code each time to regenerate the plot.





What happens to the number of combinations as the number of genes increases?

The number of combinations increases very quickly: 256 (4 genes) to over 1,000,000 with 10 genes. 300 genes have 1.4×10^{180} combinations (~1 followed by 180 zeros). For comparison, there are only about 10^{82} atoms in the observable universe. So, there are about 10^{100} more combinations than there are atoms.

For any number of genes, what is the most probable count of A's and T's?

The most probable is that there will be equal A and T and the number of each will be the same as the number of genes (4 genes: 4 T & 4 T)

What happen to the percent of rare combinations (e.g., all A or all T) as the number of genes increases?

They become extremely rare. So rare in fact that they are effectively 0 on the plot.

Combinations of alleles for a set of genes, where the data take the form of “one or the other” (heads or tails, major allele or minor allele) follow what is called a *binomial distribution*. Thinking back to the case of a single gene with 4 possible combinations and 3 different possible genotypes (TT, AT, and TT). One interesting feature of binomial distributions is that as the number of “chances” (i.e., the number of genes or alleles) increases, the distribution starts to take on a characteristic shape. When there are less than about 10 genes, the distribution appears stepped. But as the number approaches 20, it starts getting smoother and smoother.

The characteristic shape that a binomial distribution for many genes (a “bell curve”) is also the shape of one of the most common distributions in biology: the

normal distribution. A normal distribution has a single peak at the center and decreases down, gradually moving away from the peak, to very small probabilities far away from the center.

Additive nature of many alleles

In the crosses shown in the Punnett squares above (Figure 1, Figure 2), genes acted independently. In peas, for example, all combinations of the three traits were possible. In contrast, many genes act additively, wherein their effects combine to produce a phenotype. Genes involved in quantitative traits work additively, but Mendelian traits can as well. For example, crosses between plants with white flowers and those with red flowers can produce hybrids with pink flowers.

An extreme example of variation in flower color is the purple morning glory (*Ipomoea purpurea*), which is native to Central America and Mexico. Clegg and Durbin (2000) studied the genetic basis of flower color in *Ipomoea* and found that combinations of four genes were able to produce a wide range of phenotypes (Figure 4).

In this way, alternate alleles (e.g., an A vs. a T) at a single gene can combine (“additively”) with other to produce a vast array of phenotype.

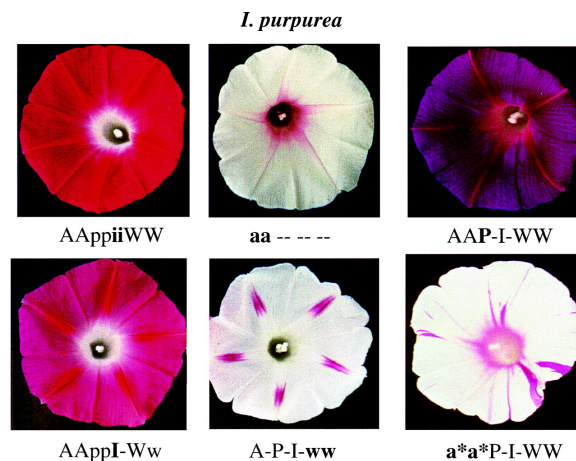


Figure 4: Variation in flower color in purple morning glory (*Ipomoea purpurea*) resulting from the additive effects of four genes. Image from Clegg and Durbin (2000).

Feedback

We would appreciate your anonymous feedback on this exercise. If you choose to, please fill out this optional 4-question survey to help us improve.

References

- Clegg, M. T., and M. L. Durbin. 2000. Flower Color Variation: A Model for the Experimental Study of Evolution. *Proc. Natl. Acad. Sci. U. S. A.* 97:7016–7023.
- Klug, W. S., M. R. Cummings, C. A. Spencer, M. A. Palladino, and D. Killian. 2019. *Concepts of Genetics*. 12th ed. Pearson.