1. Backcrossing is an important strategy in agriculture for introducing a new locus into the genome of an otherwise desirable strain. In the rice example in lecture, the authors introduced the submergence tolerance locus from a tolerant strain T into the crop strain Swarna (S). What proportion of the genome originated from the T strain in:

(a) The F1 generation?

50%.

(b) The B1 generation?

25%.

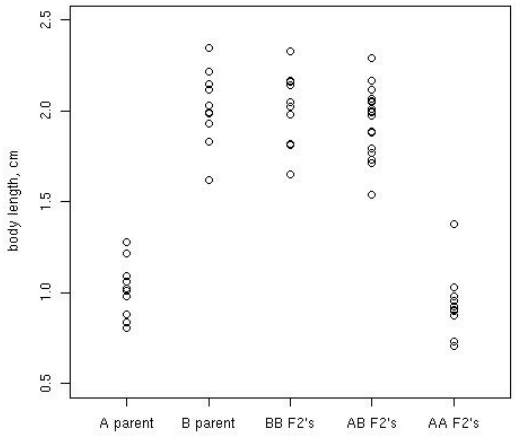
(c) The B2 generation?

12.5%.

(d) The B12 generation?

0.012%.

2. A study of body length in millipedes was launched to dissect the genetic basis of the difference in body length between two homozygous lines A and B, which are on average 1 and 2 cm in length, respectively. The lines were crossed to generate F2's, and F2's were genotyped at a number of markers; researchers mapped a single locus that was completely responsible for the variation in length. The phenotypes of 10 genetically identical lines of each parent strain, and 40 genetically distinct F2's split out by genotype at the marker, are shown below.



(a) What single-locus genetic model best explains the data?

Single-locus dominant model; the B allele is dominant to the A allele.

(b) What is the expected average phenotype across all F2's?

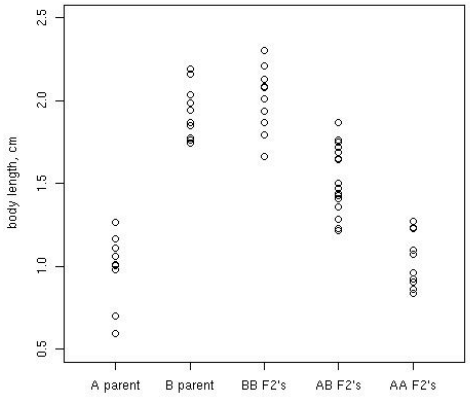
[10(2 cm) + 20*(2 cm) + 10*(1 cm)]/40 = 1.75 cm*

(c) What is the expected average phenotype across all measurements of BOTH parents?

[10(2 cm) + 10*(1 cm)]/20 = 1.5 cm*

(d) For what kind of single-locus genetic model would the answers to (b) and (c) be the same? Draw a picture to illustrate.

If the alleles were incompletely dominant (additive) then the average across all F2's would be the same as the average of the parents.



(e) Heritability is the ratio of the estimate of genetic variance across genetically diverse individuals (F2s) and the total variance across these individuals. To get genetic variance, we estimate the environmental/error variance from replicate measurements of genetically identical individuals, and subtract this from the total variance. For this calculation, when we have replicate measurements for BOTH parents, the usual tactic is to calculate the variance for each and then take the average over these two variances. Write down an expression for the heritability using all the information you have here, including means; you will not be able to calculate a numerical value.

$$h^2 = [\sigma_{F2}^2 - \sigma_P^2]/\sigma_{F2}^2$$

where σ_{F2}^2 is the variance across the F2's, and σ_P^2 is the average variance across all parent measurements.

$$\sigma_{F2}^2 = \Sigma(l_i - 1.75)^2/40$$

where l is length, i counts F2 lines, and the sum goes from 1 to 40.

$$\sigma_P^2 = [\Sigma(l_i - 2)^2/10 + \Sigma(l_i - 1)^2/10]/2$$

where l is length, i counts replicate lines, and the sums each go from 1 to 10, or

$$\sigma_P^2 = \Sigma(l_i - 1.5)^2/20$$

where the sum goes from 1 to 20.

3. You are studying the rate of blinking in humans and want to understand the genetic basis of the variation between people in how frequently they blink. You collect pairs of twins raised apart and use a Blink-O-Meter™ to measure the blinks per minute for each individual. Your data look like this:

	Twin 1	Twin 2	Mean
Twin pair 1	10.2	15.6	12.9
Twin pair 2	12.8	15.3	14.05
Twin pair 3	9.7	10.4	10.05
Twin pair 4	13.4	9.1	11.25

(a) Based on these data, the heritability for this trait is quite low. Explain how you would have a sense for this just from looking at the data, without doing a calculation.

The means do not look more different across twin pairs than does one twin in a pair from the other.

(b) Imagine that there is a genetic locus varying in the population with a modest, but real, effect on blink frequency. In other words, the heritability of this trait should be appreciable, yet in the data above it is not. (Imagine that the trends were recapitulated over many more twin pairs than what you see here!) Propose a hypothesis to explain the discrepancy, and describe what you would do to test your idea.

One explanation is high measurement error. If measurements are not reliable, heritability is low, because most of the variation in the data is not due to genetics. In this way, genetic effects can be obscured. To test this, one

could use another method beside the Blink-O-Meter™ to collect measurements and try the heritability calculation again. An alternative explanation is that environmental variability is unusually high. Perhaps these sets of twins live on different continents and have radically different diets or lifestyles; one could test this by restricting the study to twins in one country, town, socioeconomic stratum, etc.

4. Imagine that, by sequencing DNA from fossils, it is possible to take genetic snapshots of ginkgo species through evolutionary time. In single chromosomes isolated from unrelated samples from 150 million years ago, you see the following alleles at a region on chromosome 5:

	marker 1	marker 2	marker 3
Sample 1	G	C	C
Sample 2	G	T	G
Sample 3	A	A	A
Sample 4	G	T	G

From unrelated trees collected in modern Berkeley, you see:

	marker 1	marker 2	marker 3
Sample 5	A	A	A
Sample 6	G	A	A
Sample 7	G	A	A
Sample 8	T	A	C

(a) If these trends were borne out in much larger samples, what might you conclude about the evolutionary importance of the region around marker 2?

It looks like the A-A-A haplotype was at low prevalence early in the history of this species and the A at marker 2 has become fixed in modern isolates (along with some of the surrounding haplotype), possibly due to positive selection on something in the region.

(b) Which marker, 1 or 3, would you expect to be closer to marker 2 on the basis of these data, and why?

It looks like there have been fewer recombinations between markers 2 and 3 than between markers 2 and 1, so we would infer if this trend is borne out in larger samples that marker 3 is closer to marker 2.