Genome 560
Homework 2
Due April 17, 2008

1. [4 pts] I am interested in functionally characterizing offspring between inter-specific crosses. I suspect the IRB office at UW would frown upon such research using humans, but fortunately the Saccharomyces Genome Researching Project (SGRP) has completely resequenced 37 strains of *S. cerevisiae* (25 natural isolates, 7 laboratory strains, and 5 clinical isolates) and 27 strains of the closely related species *S. paradoxus* (22 natural isolates and 5 laboratory strains).

   A. If I want to perform all pair-wise crosses between the SGRP *S. cerevisiae* and *S. paradoxus* strains, how many matings will I have to do?

   B. Let's assume that none of the crosses performed above worked so instead I decide to perform all pair-wise matings within each species. How many matings will I now have to do for *S. cerevisiae* and *S. paradoxus*?

   C. More bad news. In setting up the crosses in B, I somehow managed to switch the labels on the plates for the *S. cerevisiae* strains and I have no idea which one is which. What I'm really interested in are crosses between the natural and clinical isolates. Rather than doing the obvious and reordering the strains, I decide to gamble and randomly pick two *S. cerevisiae* strains to cross. What is the probability that I guess correctly and mate a natural and clinical isolate?

2. [2 pts] In a gene expression study I identify 100 differentially expressed genes. I perform a Gene Ontology analysis and find that 15 out of the 100 are annotated as "sensory perception of sound". My array contains 10,000 genes in total, of which 1,000 are annotated as "sensory perception of sound". What is the probability that I would observe 15 or more genes out of 100 annotated with this term by chance?

3. [4 pts] In Thursday's lecture, we discussed the Wright-Fisher model of binomial sampling alleles from one generation to the next (slide 15). Assume a population size of 2N = 100. In the current generation there are i = 10 copies of the A allele (p_A = 0.10). The number of A alleles in the next generation is j. What is the probability that:

   A. j = 10
   B. j = 0
   C. j ≥ 10
   D. 5 ≤ j ≤ 15

4. [5 pts] There has been considerable recent interest in mapping loci that influence inter-individual variation in gene expression levels. In these experiments, gene expression levels are treated as a quantitative trait and linkage analysis is performed to find positions in the genome that contribute to variation in transcript abundance. Let's say you have performed a linkage analysis experiment to map gene expression quantitative trait loci (eQTL). In total, you have detected significant linkage for 1,013 expression traits. Next, you want to test for the presence of linkage "hotspots", which are regions in the genome showing linkage to multiple gene expression traits. To detect "hotspots", you divide the genome into 579 bins of equal size (this corresponds to

20 kb bins in yeast). Define the random variable X_i to be the number of linkages observed in the ith bin.

    A. What distribution does X follow? Briefly explain your choice.

    B. Assuming the eQTLs are randomly distributed, what is the probability that a bin contains no eQTLs.

    C. What is the probability that a bin contains 40 or more eQTLs?

    D. Find the number of eQTLs in a bin, x, such that $P(X \geq x) = 0.05$.

5. [3 pts] An important quantity in quantitative genetics is the population mean of a trait. Let's assume a very simple model, where the genetic basis of a quantitative phenotype is due to a single bi-allelic locus with three genotypes. As illustrated in the diagram below, the mean phenotypic value for genotypes AA, Aa, and aa are a, d, and –a, respectively. Similarly the frequencies of genotypes AA, Aa, and aa in the population are p^2, 2pq, and q^2, respectively.

Genotype	AA	Aa	aa
Frequency	p^2	2pq	q^2
Phenotypic Value	a	d	-a

    A. What is the expected (i.e., mean) phenotypic value in the population?

    B. What is the variance of the phenotypic values in the population (note, this quantity is referred to as V_G, the total genetic variation of a trait as we are ignoring environmental effects)?

    C. If p = q = 0.5 and a = 1, d = 0, and –a = -1, use the results from A and B above to calculate the mean and variance of phenotypic values.

6. [2 pts] Among the set of all families with two children, a family is selected at random and found to have a girl. What is the probability that the other child is a girl? Briefly explain your result.

7. [5 pts] Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that, when an individual actually has a disease, a positive result will occur 99% of the time, while an individual without the disease will show a positive test result only 2% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?