STA 685 / CS 695 / PPA 784, Assignment 2 (Schardl). **Due Sept. 15 in class.**

For the following assignment, please answer the **questions in bold**.

In class we went over the Punnett square, illustrating Mendel's first law. In it, we started with two alleles of a gene, and each parent was heterozygous, meaning each parent had both alleles, one of each. We could then calculate the probability of each genotype. We can generalize that by the following notation: For a gene A, we have alleles A1 and A2.

Situation 1: We assume we have two parents, F and M (female and male), where F has both alleles, and M has both alleles. The probability of parent F contributing A1 to her progeny is 0.5, and the probability of F contributing A2 to her progeny is 1 - 0.5 = 0.5. The same holds for parent M. If F and M mate to give 1000 progeny, then we expect the ratios of all possible combinations of alleles received from the parents to be close to 1:1:1:1. Note, however, that genotype A1A2 (where A1 is from parent F) is indistinguishable from A2A1 (where A2 is from parent F), and we will call these one genotype designated A1A2 (i.e., alleles are unordered). Now, the expected ratio of the three possible genotypes A1A1:A1A2:A2A2 is 1:2:1. In other words, the *expected* numbers among our 1000 progeny are:

A1A1: 250
A1A2: 500
A2A2: 250

Dividing through by 1000, we get the frequencies:

A1A1: 0.25
A1A2: 0.5
A2A2: 0.25

Let us say we have done the cross, and get the following *observed* numbers of progeny:

A1A1: 200
A1A2: 600
A2A2: 200

Question 1, Test whether the *observed* numbers *significantly* differ from the *expected* numbers: (a) Define the null and alternative hypothesis, (b) compute the S-statistics, and (c) use the chi-square distribution to test for significance. Note that there are some nice online computational tools, such as http://stat.utilities.googlepages.com/tables.htm (and more nice applets posted on the STA 291 course website http://polytopes.net/courses/STA291F09/).

Situation 2: Now we want to generalize the problem from Situation 1. Let's start by allowing different probabilities of contributing the alleles, as follows: For each of parents F and M (both genotypically A1A2), the probability of contributing A1 to a progeny is p (where $0 \leq p \leq 1$), and the probability of contributing A2 to a progeny is $q = 1 - p$. Assuming independent segregation, We can compute the probabilities of obtaining the three genotypes in terms of p, as follows:

A1A1: $p*p = p^2$
A1A2: $pq + pq = 2p(1 – p) = 2p – 2p^2$
A2A2: $q*q = (1 – p)*(1 – p) = 1 – 2p + p^2$

Note that, if we know the frequency of genotype A1A1, f(A1A1) in the F2 progeny, then we can easily calculate p = sqrt(f(A1A1)); and if we know the frequency f(A2A2) we can calculate p = 1 – q = 1 – sqrt(f(A2A2)).

Situation 3: Let us generalize further. We have discovered some little green rodents from Mars, and find that they have a more complex system of reproduction where the generations alternate between diploid (each individual has two homologues for every chromosome) and triploid (each individual has three homologues for every chromosome). We will now set up a test mating of two triploid parents, F and M, each with genotype A1A2A3, and get 1000 diploid progeny. If we assume equal probability for a progeny to receive allele A1 or A2 or A3 from parent F, and also equal probabilities to receive A1 or A2 or A3 from parent M, then we can calculate the expected number of each of genotype (again, unordered). We get:

A1A1: 111
A1A2: 222
A1A3: 222
A2A2: 111
A2A3: 222
A3A3: 111

Situation 4: Let us also generalize Situation 3 for any distribution of probabilities, but (as in Situation 2) we assume that distribution to be identical for contributions from parents F and M. Hence, for parent F the probability to donate allele A1 is p, the probability to donate allele A2 is q, and the probability to donate allele A3 is r = 1 – (p + q), where p, q, and r are all real numbers ≥ 0; and exactly the same is true for parent M. **Question 2: Derive the expressions for the probabilities of all possible genotypes in terms of p and q:**

A1A1:

A1A2:

A1A3:

A2A2:

A2A3:

A3A3:

Question 3: Assume that A1 is dominant to A2 and A3, so that when we score progeny we cannot distinguish between the genotypes, A1A1, A1A2, and A1A3, how then can we calculate p, and how can we calculate q?

Question 4: Assume that A1 is dominant to A2, A1 is dominant to A3, and A2 is dominant to A3. This means that we cannot distinguish between A1A1, A1A2, and A1A3, and we also cannot distinguish between A2A2 and A2A3. Is it still possible to calculate p, q and r? How?

Situation 5: Now we will travel back to Earth, and consider the dihybrid cross shown on slide 8 of Lecture 2 (entitled "Linkage", cited as Figure 13.12 part 2, from Genomes 3). Here we have tall vs short plants, and round vs wrinkled seeds. We will designate the gene for plant size T, and designate the one for seed shape R. For the F2 progeny, we will define:

p_T = probability of donating the dominant (tall) plant size allele,
q_T = probability of donating the recessive (short) plant size allele,
p_R = probability of donating the dominant (round) seed shape allele, and
q_R = probability of donating the recessive (wrinkled) seed shape allele

Again, for each p_i and q_i, $0 \le p_i \le 1$, $0 \le q_i \le 1$, and $p_i + q_i = 1$.

Question 5: Derive the expanded polynomial in terms of p_T, q_T, p_R and q_R for the probabilities of each F1 genotype (TTRR, TTRr, TTrr, TtRR, TtRr, Ttrr, ttRR, ttRr, and ttrr). If we further assume that each $p_i = q_i = 0.5$, then estimate how many of each genotype (again, with alleles unordered) that you will get among 1000 of the F2 progeny.

Situation 6: Finally, my favorite: haploid organisms. For these, there is only one copy of each homologous chromosome — therefore only one allele of each gene — in an organism. In this case, we can do all our analysis in the F1 generation, and for each gene of each F1 progeny the chance that its allele came from the mother = the chance it came from the father = 0.5. We take as an example the fungus, *Aspergillus nidulans*, for which the wild type (non-mutant) grows Green and Fluffy. If we have an inactivating mutation in gene G, giving g, the color is yellow; If we have an inactivating mutation in gene F, giving f, the growth form is waxy.

Question 6: If we mate a GF parent with a gf parent, what are the probabilities for each of the four phenotypes in an F1 progeny:

Green, Fluffy
Yellow, Fluffy
Green, Waxy
Yellow, Waxy

And, if we mate Gf x gF, what are the probabilities of each of these four phenotypes in a progeny?