Homework Set #1 Key

1. (a) $p' = P_{AA} + \left(\frac{1}{2}\right)P_{Aa} = p_m p_f + \left(\frac{1}{2}\right)\left[p_m\left(1-p_f\right) + p_f\left(1-p_m\right)\right] = \left(\frac{1}{2}\right)\left(p_m + p_f\right)$, the even average of the parental allele frequencies. This makes sense because mothers and fathers contribute equally to autosomal genotypes of their offspring.
   (b) No. For example, if $p_f = 0$ and $p_m = 1$, then $P_{AA} = P_{aa} = 0$ and $P_{Aa} = 1$, which are not Hardy-Weinberg proportions. (Why?)
   (c) Because all the Hardy-Weinberg conditions hold for the offspring generation, the frequency of A will remain at $p = (p_f + p_m)/2$ in their descendants with genotype frequencies in Hardy-Weinberg proportions AA p^2: Aa $2p(1\text{-}p)$: aa $(1\text{-}p)^2$.

2. (a) Since 215 diploid people were sampled, the total alleles per locus samples was 2•215 = 430.
   At D2S44: $p_{A_7} = 15/430 = .035$, $p_{A_8} = 18/430 = .042$, $p_{A_9} = 60/430 = .140$.
   At D4S139: $p_{B_{21}} = 11/430 = .026$, $p_{B_{26}} = 81/430 = .188$.
   (b) Freq(A_7A_7) = $\left(p_{A_7}\right)^2 = (.035)^2 = .0012$; Freq($B_{21}B_{26}$) = $2p_{B_{21}}p_{B_{26}} = 2(.026)(.188) = .0098$.
   (c) Freq(A_7B_{21}/A_7B_{26}) = $2P_{A_7B_{21}}P_{A_7B_{26}} \underset{\substack{\text{assuming}\\ \text{LE}}}{=} 2\left(p_{A_7}p_{B_{21}}\right)\left(p_{A_7}p_{B_{26}}\right) = 2(.035)(.026)(.035)(.188) = 1.2 \times 10^{-5}$. Freq($A_7B_{21}/A_8B_{26}$) = $2P_{A_7B_{21}}P_{A_8B_{26}} \underset{\substack{\text{assuming}\\ \text{LE}}}{=} 2\left(p_{A_7}p_{B_{21}}\right)\left(p_{A_8}p_{B_{26}}\right) = 2(.035)(.026)(.042)(.188) = 1.4 \times 10^{-5}$.
   (d) Departures from H-W equilibrium and LE could lead to on- and two-locus frequencies that are higher or lower than those estimated above. For example, if all heterozygous genotypes with A_7 were lethal, the Freq(A_7A_7) = $p_{A_7} = .035 > \left(p_{A_7}\right)^2 = .0012$, the latter of which would be estimated using H-W equilibrium.
   (e) Freq(A_7B_{21}/A_7B_{26}) = $2(.106)(.069)(.106)(.138) = 2.1 \times 10^{-4}$.
   (f) Comparing (e) and (c), we can see that subgroups can differ in STRP genotype frequencies (in this case, by more than an order of magnitude). If a suspect were compared against the wrong subgroup, the "rareness" of a DNA fingerprint might be substantially over- or underestimated.

3. p. 107 #3: $p_m = .8$, $p_f = .4$

$$P' = p_m p_f = (.8)(.4) = .32$$
$$H' = p_m q_f + q_m p_f = (.8)(.6) + (.2)(.4) = .48 + .08 = .56$$
$$Q' = 1 - P' - H' = 1 - .32 - .56 = .12$$
$$\Rightarrow\ p_m' = p_f' = p' = (.32) + \tfrac{1}{2}(.56) = .6\left(= \frac{p_m + p_f}{2}\right).$$

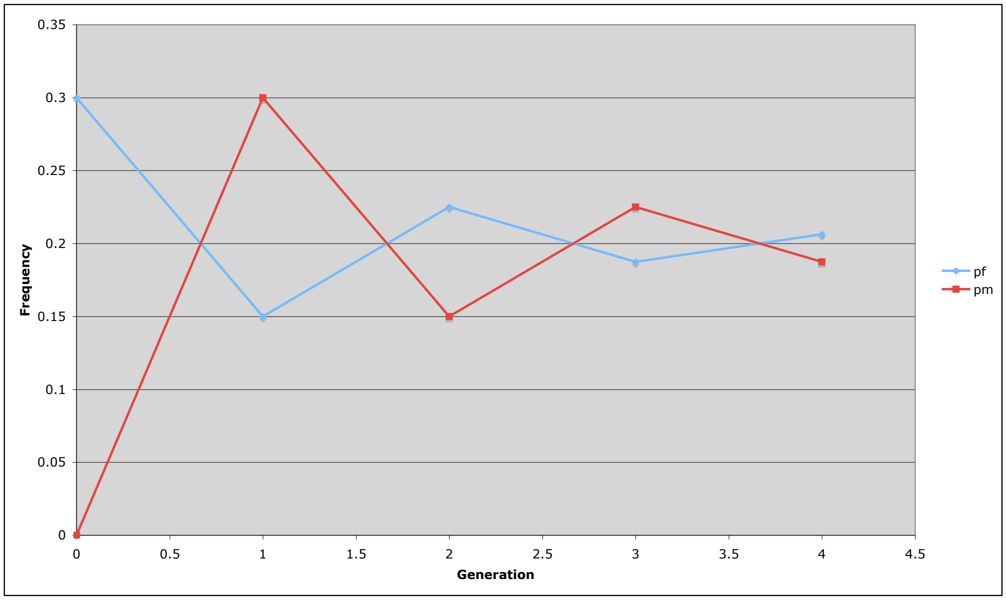
$$P'' = (p')^2 = (.6)^2 = .36$$
$$H'' = 2p'q' = 2(.6)(.4) = .48$$
$$Q'' = 1 - .36 - .48 = .16$$

***

$$H - 2\overline{p}\overline{q} = (p_m q_f + p_f q_m) - 2\left(\frac{p_f + p_m}{2}\right)\left(\frac{q_f + q_m}{2}\right)$$
$$= p_m q_f + p_f q_m - \tfrac{1}{2}(p_f q_f + p_f q_m + p_m q_f + p_m q_m)$$
$$= \tfrac{1}{2}(p_f q_m + p_m q_f - p_m q_m - p_f q_f) = \tfrac{1}{2}\left[p_m(q_f - q_m) - p_f(q_f - q_m)\right]$$
$$= \tfrac{1}{2}\left[p_m(p_m - p_f) - p_f(p_m - p_f)\right] = \tfrac{1}{2}(p_m - p_f)^2$$
$$Q - \overline{q}^2 = q_m q_f - \left(\frac{q_f + q_m}{2}\right)^2 = q_m q_f - \tfrac{1}{4}(q_f^2 + 2q_f q_m + q_m^2)$$
$$= -\tfrac{1}{4}(q_m^2 - 2q_m q_f + q_f^2) = -\tfrac{1}{4}(q_m - q_f)^2 = -\tfrac{1}{4}(p_f - p_m)^2$$

4. p. 107 #4:

Generation	0	1	2	3	4
p_f	.3	(.3 + 0)/2 = .15	(.15 + .3)/2 = .225	(.225 + .15)/2 = .1875	(.1875+.225)/2 = .20625
p_m	0	.3	.15	.225	.1875



$|d_t| = \left(\tfrac{1}{2}\right)^t |d_0|$. Find t such that $\left(\tfrac{1}{2}\right)^t \left| .3 - \dfrac{1\cdot(0) + 2\cdot(.3)}{3} \right| \le .001 \Rightarrow \left(\tfrac{1}{2}\right)^t (.1) \le .001 \Rightarrow \left(\tfrac{1}{2}\right)^t \le .01 \Rightarrow$

$t \ln(1/2) \le \ln(.01) \;\Rightarrow\; t \ge \dfrac{\ln(.01)}{\ln(1/2)} = 6.64$ generations. Thus, p_f will deviate less than .001 from $\bar{p} = \left(\tfrac{1}{3}\right)(0) + \left(\tfrac{2}{3}\right)(.3) = .2$ by generation 7.

5. p. 108 #8: $q^2 = 1/2500 \Rightarrow \hat{q} = 1/50 = .02$. Heterozygote frequency = 2 (.98)(.02) = .0392. The total frequency of carriers = $q^2 + 2pq$ so the frequency of matings between carriers is $\left(q^2 + 2pq\right)^2 = \left(q^2\right)^2 + 2(2pq)\left(q^2\right) + (2pq)^2 = (1/2500)^2 + 2(.0392)(1/2500) + (.0392)^2 = .00157$. [Note: this is approximately equal to $(.0392)^2 = .00154$. That is, almost all (> 98%) matings among carriers are between heterozygotes.]

6. p. 108 #13:

$$\hat{\pi} = \frac{N}{N-1}\sum_{ij} \hat{p}_i \hat{p}_j \pi_{ij} = \frac{5}{4}\left[\hat{p}_1\hat{p}_2\pi_{12} + \hat{p}_1\hat{p}_3\pi_{13} + \hat{p}_2\hat{p}_1\pi_{21} + \hat{p}_2\hat{p}_3\pi_{23} + \hat{p}_3\hat{p}_1\pi_{31} + \hat{p}_3\hat{p}_2\pi_{32}\right]$$

$$= \frac{5}{4}\left[2\hat{p}_1\hat{p}_2\pi_{12} + 2\hat{p}_1\hat{p}_3\pi_{13} + 2\hat{p}_2\hat{p}_3\pi_{23}\right] \quad \text{since } \pi_{ij} = \pi_{ji}$$

$$= \frac{5}{4}\left[2\left(\frac{2}{5}\right)\left(\frac{2}{5}\right)\left(\frac{4}{15}\right) + 2\left(\frac{2}{5}\right)\left(\frac{1}{5}\right)\left(\frac{4}{15}\right) + 2\left(\frac{2}{5}\right)\left(\frac{1}{5}\right)\left(\frac{8}{15}\right)\right] = .267$$

$\hat{p}_S = 8/15$.

7. p. 109 #16: $\hat{p}_1 = \hat{p}_2 = \hat{p}_3 = \hat{p}_4 = 1/4$. $\pi_{12} = 1/900$, $\pi_{13} = 4/900$, $\pi_{14} = 2/900$, $\pi_{23} = 5/900$, $\pi_{24} = 6/900$, and $\pi_{34} = 10/900$.
$\hat{\pi} = \dfrac{N}{N-1}\sum_{ij} \hat{p}_i \hat{p}_j \pi_{ij} = \dfrac{4}{3}\sum_{ij}\left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right)\pi_{ij} = \dfrac{4\cdot 2}{3\cdot 4\cdot 4\cdot 900}(1+4+2+5+6+10) = .00519$. [Note: the factor 2 is needed because each difference is counted twice since, in the summation $\pi_{12} = \pi_{21},\ \pi_{13} = \pi_{31},\ \ldots,\ \pi_{34} = \pi_{43}$.

8. (a) Only Ab and aB gametes are present and every individual carries one of each. So D = $P_{AB}\,P_{ab} - P_{Ab}\,P_{aB}$ = 0*0 – (1/2)(1/2) = –1/4.
(b) Every individual produces (1–1/4)/2 Ab, (1–1/4)/2 aB, (1/4)/2 AB, and (1/4)/2 ab gametes. So $P'_{AB} = 1/8$, $P'_{ab} = 1/8$, $P'_{Ab} = 3/8$, and $P'_{aB} = 3/8$, which implies $D' = (1/8)(1/8) - (3/8)(3/8) = -1/8$.

(c) The F_1 generation is formed by random mating so, as shown in class,

$$P''_{AB} = P'_{AB} - rD = (1/8) - (1/4)(-1/8) = 5/32$$

$$P''_{ab} = P'_{ab} - rD = (1/8) - (1/4)(-1/8) = 5/32$$

$$P''_{Ab} = P'_{Ab} + rD = (3/8) + (1/4)(-1/8) = 11/32$$

and

$$P''_{aB} = P'_{aB} + rD = (3/8) + (1/4)(-1/8) = 11/32$$

$\Rightarrow$

$$D'' = P''_{AB}P''_{ab} - P''_{Ab}P''_{aB} = (5/32)(5/32) - (11/32)(11/32) = -3/32$$

(d) The F_1 generation was formed by random mating so the reduction of disequilibrium by the factor $(1 - r) = (1 - ¼)$ applies. By contrast, the original (parental) generation could not have been formed by random mating since, e.g., there are no Ab/Ab homozygotes. The reduction of disequilibrium by $(1 - r)$ does not apply in this case. This exercise illustrates that the recursion $D' = (1 - r)D$ applies only if the population for which D is computed was formed by random mating.