CHAPTER 20 Answers to Problems

Problem 20.2. (a) North Pacific minke whale.

(b) The bootstrap value changes from 100 to approximately 75 when the 106 bp are removed. Note that you may get slightly different bootstrap values because bootstrapping is a random sampling process, and thus different outcomes can result from independent bootstrap sampling analyses. Also, the removal of 106 bp removes information and thus the bootstrap values should decrease in magnitude.

In addition, the position in which this unknown sample clusters also changes. With the full sequence, **Unknown5** clusters only with a Northern Pacific minke whale sample. With the shortened sequence, **Unkown5** clusters with a group containing the Northern Pacific minke whale plus samples of an Atlantic minke whale and an Antarctic minke whale.

(c) Two samples have the greatest match in a blast search. An Antarctic minke whale and an Atlantic minke whale sampled in the Mediterranean.

Problem 20.3.

The expected two-locus genotype frequency is 0.1944 in Indian, and 0.0384 in China. Thus, the bone appears approximately five times more likely to have originated in India than in China. More loci, highly polymorphic loci, and more formal statistical tests (e.g., assignment tests) would allow for a more powerful and appropriate assessment of the origins of the bones.

India:

P(*Aa*) = (0.40 x 0.60) = 0.24
P(*BB*) = (0.90 x 0.90) = 0.81
P(*AaBB*) = (0.24 x 0.81) = 0.1944

China:

P(*Aa*) = (0.60 x 0.40) = 0.24
P(*BB*) = (0.40 x 0.40) = 0.16
P(*AaBB*) = (0.24 x 0.16) = 0.0384

Problem 20.4. Assignment tests assume Hardy-Weinberg proportions when computing the expected frequency of genotypes in each candidate population of origin of an individual. If a population is not in H-W proportions, then the actual frequency of genotypes may not match the estimated frequency of genotypes, which could lead to

erroneous inference about the likelihood of an individual genotype originating from a population. To quantify the possible consequences of violating the assumption of H-W proportions, we might simulate data sets where a population is not in H-W proportions and conduct assignment tests to determine the proportion of individual correctly assigned when assuming H-W proportions; this proportion of correctly assigned individuals could then be compared to the proportion correctly assigned when we use the true genotype frequencies (not estimated by assuming H-W proportions).

Problem 20.6. A *PI* of 1/300 or 1/500 would be reasonably low because if only 300-500 individuals exist than all individuals are likely to be resolved (as having unique genotypes). The *PI* generally should be equal to or less than approximately $1/N_C$ (where N_C is the population census size).

Problem 20.8. A risk of harvesting from a mixed population (i.e., a population with individuals originating from different breeding subpopulations) is that we might over harvest one of the subpopulations. For example a harvest of 30% might be allowed from the mixed population but most individuals harvested could originate from only one subpopulation. Thus the percent harvested from the one subpopulation might be far higher than 30%.

Genetic methods that can help monitor harvest include individual-based assignment tests (Section 20.5) to determine the subpopulation of origin of individuals in a mixed population, and population composition analysis (Section 20.6) that determines the extent of contribution of different subpopulations to a mixed population. Advantages of using molecular genetic approaches are that all individuals have a tag (genetic tag) they are born with, which makes it unnecessary to capture and tag individuals. Also, genetic tags cannot fall off like traditional tags or marks. A potential limitation is that if subpopulations are not genetically-differentiated then assignment is difficult or impossible (even when using many loci).

Problem 20.9 For each locus separately we compute *PE* using expression 20.2, below (expression 2 from Jamieson and Taylor (1997) [hotlink]). Then we use the multiplication rule (assuming independence among loci) and multiply each single-locus *PE* (actually 1 – *PE*, see equation below) to compute the multi-locus *PE* (see Jamieson and Taylor (1997), expression 4). Expression 20.2 is a general formula for the case when one parent's genotype is unknown as is often the case in wildlife paternity studies.

$$PE = \sum_{i=1}^{n} p_i^{\,2}(1-p_i)^2 + \sum_{i>j=1}^{n} 2\,p_i\,p_j(1-p_i-p_j)^2$$

$PE_{Locus\ I} = [(0.5^2)(1 - 0.5) + (0.5^2)(1 - 0.5)] + [(2 \times 0.5 \times 0.5(1 - 0.5 - 0.5)^2 + (2 \times 0.5 \times 0.5(1 - 0.5 - 0.5)^2]$

$= [(0.25)(0.5) + (0.25)(0.5)] + [(0.5(1 - 0.5 - 0.5)^2 + (2 \times 0.5 \times 0.5(1 - 0.5 - 0.5)^2]$

$= 0.0625$

$PE_{Locus\ 2} = 0.0613$

$PE_{Locus\ 3} = 0.0518$

$PE_{Locus\ 4} = 0.0441$

$PE_{Multi\text{-}locus} = 1 - [(1 - 0.0625) \times (1 - 0.061265) \times (1 - 0.051756) \times (1 - 0.0441)] =$ **0.202**

Thus, these four loci give a probability of excluding a candidate parent (who is not the true parent) of 0.202. This probability is low. More markers would be needed to achieve high power. Markers that are more polymorphic (e.g., microsatellites) would also be helpful.

Problem 20.11.

Launch the program *GeneClass* by clicking on the program icon (geneclass2.exe).

Under the tab "Parameters", to the right of "Reference populations" click on "Open" to load the file titled "Prob 20.11 data.txt" .

Under the tab "(1) Computational goal" click on the circle beside "1) Assign / Exclude population as origin of individuals" and on the circle beside "Individuals"

Under the tab "(2) Criteria for Computation" click the circle beside "Rannala & Mountain" (to conduct the assignment method of Rannala and Mountain 1997).

Click on "Start".

The "Results" window appears and shows "Correctly assigned: 88.7%". This means that 88.7% of individuals in the data set were correctly assigned to their population of origin.

The Results window also shows that two individuals sampled from Yellowstone (YS) were assigned to the Montana population. For example, row 6 of the "Assigned samples" shows that sample number "YS6" from Yellowstone was most likely to have originated in Montana ("Mont"). That is to say that the (posterior) probability of this individual's genotype was highest in the Montana population (68.8%) not in the Yellowstone population.