PBL MODULE 2 - Population Structure, Genetic Variation, and Conservation   
   
Sub-Module B: Population Structure   
   
Data Files:   
• File 5: Microsat\_walrus\_structure\_nopopassign.txt   
• This is a tab delimited file formatted for the program STRUCTURE (Pritchard, Stephens,   
& Donnelly, 2000). Note that the populations are NOT assigned, that’s what structure   
will figure out. However, the 5 breeding areas are in order of populations, as in the   
other GenaLex files   
• File 6: Walrus\_females.xlxs – only females, formatted for GenaLex   
• File 7: Walrus\_males.xlxs - only males, formatted for GenaLex   
   
Analysis 1: STRUCTURE

New project: name it, set the directory, choose data file is microstat walrus no pop assigned. Its trying to determine how many populations this data structure has, see if there is sub structuring going on, more than 5. Row of marker names and click the top special format, and individual id. To run parameter set

Mixture of color shows genetic mixture

Do the STRUCTURE analysis first as it may take 10-15 minutes to run.   
Note on STRUCTURE:   
• A model-based clustering method (Pritchard et al. 2000)   
• Bayesian approach (MCMC: Markov Chain Monte Carlo)   
• Detects the underlying genetic population among a set of individuals genotyped at multiple   
markers   
• Computes the proportion of the genome of an individual originating from each inferred   
population (quantitative clustering method)   
How to Run STRUCTURE:   
Open structure (or download if needed with interface for windows -   
https://web.stanford.edu/group/pritchardlab/structure.html   
1. Start a project   
FILE –NEW PROJECT   
• Name the project, select the directory where File 5 is located, and then select the file,   
NEXT   
o If you get errors try and create a new folder on your desktop and work from   
there or run as administrator   
• Put in the number of individuals (172), No. Loci (17), and put a dash (-) for missing data   
value, NEXT

• Click Row of marker names, and under SPECIAL FORMAT, click the box that indicates   
data for individuals is on one row, NEXT   
• Click Individual ID for each individual, FINISH   
• PROCEED   
2. Set Parameters   
PARAMETER SET –NEW   
• Run Length: Burnin – 100,

Burnin want all samples to meet stationary

# of MCMC reps – 50000 (Normally it would be a longer run,   
maybe 1 million)   
• All other tabs can be left at default settings   
• Click OK   
• Name Parameter set E.g. Parameter1   
   
 which populations are the most in danger due to inbreeding, make decisions on how to breed them   
2. Start a Job   
PROJECT –START A JOB   
• Click your named parameter set   
• For Set K, choose 1 to 5. Note: The K value is something that needs to be determined . you can even do 10.  
empirically. Usually you set from 1 to the number of expected populations, in this case   
5, for the 5 main breeding areas. You can do more than 5 too.   
• Number of iterations – 5 (normally this would be longer (e.g. 10))   
• START, Click OK to the pop up box   
• Your job should be running, and it will tell you when its complete – should take about   
10-15 minutes   
While structure is running, work on other aspects where possible   
   
2. Determine how many populations (K) are indicated by the data, (i.e. most likely/probable)   
First, you need to create a zip file of all files within the results folder (all iterations for all values   
of K). The results folder will be in which ever directory you put the datafile (name of the   
project)   
Go to STRUCTURE HARVESTER: http://taylor0.biology.ucla.edu/structureHarvester/# (Earl,   
2012)   
• Open the zipped file you created in structure Harvester   
• Click Harvest   
• Examine the top graph output.   
• Note: When K is approaching a true value, L(K) plateaus (or continues increasing slightly) and   
has high variance between runs (Rosenberg et al., 2001)

• The plot should asymptote at the most likely value of K – what is it?   
o Note if there are confusing results, examine the plots for the two possible k-  
values that might be right – which one seems reasonable?   
3. Get your structure diagram   
Go back to STRUCTURE. Click on one (of the 5) results runs (left hand area), for the K-value that is most   
probable based on STURCTURE HARVESTER.   
   
• On the right hand side, Go to Bar Plot – SHOW   
• Then plot on multiple lines. Compare with the plot on one line   
• Save your preferred version as a jpeg   
• You can include this in your report   
• You can also look at the tree plots for any of the iterations for the appropriate k-value.   
Which populations seem to be the closest genetically?   
   
Questions to answer from STRUCTURE ANALYSIS

As you go through the variance should go up (ex 5 shouuld be high) and the 5 should asymptote out? Its meets a curve kinda plateaus  
   
1. Based on the result, how many populations can you infer? In other words, how many   
populations demonstrate distinct clustering from the other populations?   
• What does this suggest about gene flow between each of the 5 main breeding   
areas? Hint: Look at the map   
   
Analysis 1: Examining sex-biased dispersal   
Note: the following analyses should be based off the results of the STRUCTURE analysis, such   
that the number of populations equals your most probable k value. Thus you may need to re-  
format Files 6 and 7 for Genalex - Please discuss with me when you reach this point as I will   
provide the analyses for you from GenAlex.   
   
For both the male and female datasets (Files 6 and 7)   
Click “add-ins” tab   
 - From GenAlEx dropdown menu, select “frequency”   
 - Verify number of loci, samples, pops, and codominant data format.   
 - click OK   
 - Check only the following:   
- Het, Fstat & Poly by Pop   
- Pairwise FST   
- NEi Distance

- Step by Step   
You will only need the HFP and FST tabs   
   
There are a number of ways to determine sex-biased dispersal, as indicated in the reading   
(Prugnolle & De Meeûs, 2002)   
“Thus, ... one can expect that the sex with higher dispersal will   
have a lower between-subpopulations Fst value compared with   
the sex that is dispersing less.”   
Questions to Answer:   
1. Look for the average number of migrants between males and females, under the HFP   
tab. What does this suggest about dispersal among the sexes?   
i. Are they statistically different (look at the SE and determine the 95%   
confidence intervals to determine significance).   
ii. Use the Confidence.norm function and use an alpha of 0.05, the Standard   
Error here, and the sample size of 17 (since the mean is determined by   
the Nm determined for each locus).   
   
b. Look at the Fst value for the total population between males and females under   
the HFP tab. What does this suggest about dispersal among the sexes?   
i. Are they statistically different (look at the SE and determine the 95%   
confidence intervals).   
   
2. Look at the FST tab. For ease, paste the male pairwise Fst data into the female one for   
side by side comparison.   
a. How do the Fst values compare for each subpopulation between males and   
females. What does this indicate about sex-biased dispersal in these walruses?   
(note we don’t have significance values for these measures with this program).   
   
3. Estimate the number of migrants for each sex and subpopulation using the formula:   
Nm = ((1 / FST) – 1)   
 4   
a. What do these results indicate about dispersal between males and females   
across the breeding regions? How may this affect conservation efforts?   
   
References Cited:   
Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output   
and implementing the Evanno method. Conservation genetics resources, 4(2), 359-361.   
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus   
genotype data. Genetics, 155(2), 945-959.

Prugnolle, F., & De Meeûs, T. (2002). Inferring sex-biased dispersal from population genetic tools: a   
review. Heredity, 88(3), 161-165.   
Rosenberg, N. A., Burke, T., Elo, K., Feldman, M. W., Freidlin, P. J., Groenen, M. A., . . . Vignal, A. (2001).   
Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken   
breeds. Genetics, 159(2), 699-713.