**North Santiam Spring Chinook Pedigree Assignment Procedure**

1. Samples genotyped prior to implementing Progeny database have IDs created by Dave Jacobson, and those IDs should be retained for consistency across years. Note that different IDs were used in some steps of the pedigree analysis, but the first six characters and last four digits of the IDs remained the same. All other samples should retain their Progeny IDs.
2. Raw genotypes are filtered to exclude samples genotyped at <7 loci (Appendix A).
3. Duplicate samples are identified using “multilocus genotype matching” in GenAlEx (Appendix B).
4. For each progeny year (NOR returns), possible parents consist of adults that returned 3-5 years prior (e.g. parents of 2015 progeny returned in 2010, 2011, and 2012).
5. Parents are assigned to progeny in Colony (Appendix C) and Cervus (Appendix D).
6. The Colony output contains one set of assignments for each progeny, which is converted to a spreadsheet and sorted by progeny ID. Parent columns are labeled as “Colony Mother” and “Colony Father”. Columns are added for “Cervus Mother” and “Cervus Father”. (Note that Colony and Cervus report parents in different orders.)
7. The Cervus output contains a list of possible parents with positive LOD scores for each Progeny, along with the number of loci compared and the number of mismatching loci. Assignments are accepted if at least 7 loci match for each progeny-parent pair. Any assignment with >1 mismatch per progeny-parent pair or >2 mismatches per trio is excluded. Cervus assignments with significant LOD scores that meet matching criteria are highlighted in the Cervus output and parent IDs are copied to the Cervus columns on the spreadsheet.
8. Progeny-parent trios are checked to make sure the parents could have spawned together and disagreements are recorded in a “Comments” column:
   1. Both parents must have returned in the same year
   2. Both parents must have been released above Minto or above Detroit
   3. If one parent of a trio was collected as a carcass sample, recovery date must be after the release date of the other parent
9. For some progeny there are multiple assignments in Cervus and/or the Colony assignments do not match the Cervus assignments. The parent with the least number of mismatches is accepted. If multiple parents are assigned with the same number of mismatches and no other parents have fewer mismatches, then none of the assignments are accepted and the progeny is unassigned.
10. If no assignments pass criteria in Cervus, but there are near-assignments with >1 mismatched loci and a significant LOD, assignments are visually compared to determine if any mismatches are likely the result of allele dropout, sizing error, or other scoring errors.

**Appendix A**

**Filtering Samples Genotyped at < 7 Loci**

1. Open Excel file and create a new column to the right of the last marker, label header '#Loci'
2. Use '=COUNTIF' function in excel (e.g. =COUNTIF(D2:AA2,”>0”)/2)
3. Copy and paste this function for all individuals
4. Highlight entire sheet and select “Data” > “Filter”
5. Filter ‘#Loci’ column for values <7
6. Highlight samples with <7 loci
7. Copy tab in Excel and label “Filtered”
8. Delete samples with <7 loci
9. Clear all filtering on both sheets
10. Record number of individuals removed for each group

**Appendix B**

**Removing Duplicate Samples**

1. Copy tab containing filtered genotype table in Excel and format for GenAlEx
   1. Column A: sample ID
   2. Column B: population (year and/or location)
   3. Columns C-Z: genotypes
   4. Insert 2 rows at the top and insert the following information:

A1: Number of loci

B1: Number of individuals

C1: Number of populations

D1…: Individuals per population

D2…: Names of populations

A2: Name of the dataset (optional, will be in output file)

For example:

12 936 1 936

Title 2007

1. Load GenAlEx add-in and select the ‘Frequency-Based’ GenAlEx dropdown, select 'Multilocus’, then ‘Matches…' to match genotypes
2. Confirm numbers are right
3. Under the 'data format' select two columns/locus 'Codominant'
4. Select ‘List Pairs of Matches and Near Matches'
5. Specify N=11 loci to evaluate for near matches
6. Ignore missing data when finding matches
7. Click OK to run
8. Label created sheet “Matches”
9. Highlight a member of each pair to be removed
   1. If both samples were released above Minto, remove earlier sample
   2. If one sample was released above Minto and the other sample is a carcass sample, remove sample released above Minto
   3. If both samples were from the same jar (batch samples), remove first sample
   4. If samples cannot be results of duplicate sampling (i.e. at least one member of pair released above Detroit), remove both samples
10. Copy tab containing filtered genotype table in Excel and rename “Duplicates removed”
11. Copy highlighted samples and paste at the end of sample list on new tab
12. Highlight sample column and select ‘Conditional Formatting’ > ’Highlight Cell Rules’ > ’Duplicate Values’
13. Delete rows containing duplicate samples in red
14. Record number of duplicates removed for each group
15. Use “Duplicates removed” for the next step

**Appendix C**

**Assigning Parents in Colony**

#COLONY PARAMETERS USED BY ANDREW BLACK

CREATE NEW PROJECT, SELECTING EMPIRICAL DATASET.

#PARAMETER TAB

MATING SYSTEM: MALE/FEMALE POLYGAMY

MATING SYSTEM II: WITHOUT INBREEDING [default]

SPECIES: DIOECIOUS, DIPLOID [default]

LENGTH OF RUN: MEDIUM

ANALYSIS METHOD: FL-PLS COMBINED (FPLS)

LIKELIHOOD PRECISION: MEDIUM [default]

RUN SPECIFICATIONS: NO UPDATE ALLELE FREQUENCY, NO SIBSHIP SCALING, 1 RUN, 1234 RANDOM NUMBER SEED

SIBSHIP PRIOR: NO PRIOR

# MARKER TAB

NUMBER OF LOCI: 12

MARKER TYPES AND ERROR RATES: UPLOAD FILE FOUND IN DIRECTORY (C:\Users\blackand\Documents\PEDIGREE\_PROJECT\COLONY\North Santiam\2010\)

ALLELE FREQUENCY: UNKNOWN

CHECK DATA THEN PROCEED TO NEXT TAB

#OFFSPRING GENOTYPES (NOTE, YOU WILL NEED TO REMOVE HEADER PRIOR TO UPLOAD)

NUMBER OF OFFSPRING: NULL

CHECK DATA THEN PROCEED

#MALE GENOTYPES (NOTE, YOU WILL NEED TO REMOVE HEADER PRIOR TO UPLOAD)

NUMBER OF MALES: NULL

PROBABILITY OF DAD: NULL (DEFAULT TO 95% UNLESS YOU HAVE GOOD REASON TO DEVIATE)

CHECK DATA THEN PROCEED

#FEMALE GENOTYPES (NOTE, YOU WILL NEED TO REMOVE HEADER PRIOR TO UPLOAD)

NUMBER OF FEMALES: NULL

PROBABILITY OF MOM: NULL (DEFAULT TO 95% UNLESS YOU HAVE GOOD REASON TO DEVIATE)

CHECK DATA THEN PROCEED

#KNOWN PATERNAL SIBS TAB

Number of known paternal sibships: 0

#KNOWN MATERNAL SIBS TAB

Number of known maternal sibships: 0

#EXCLUDED PATERNITY

Number of offspring with excluded fathers: 0

#EXCLUDED MATERNITY

Number of offspring with excluded mothers: 0

#EXCLUDED PATERNITY SIBS

Number of Excluded Paternal Sibships: 0

#EXCLUDED MATERNITY SIBS

Number of Excluded Maternal Sibships: 0

#SAVE DATA

# UNDER 'RUN' SELECT 'RUN DATA', USING ONE THREAD

#WILL WANT TO USE THE 'BEST CLUSTER' RESULTS FILE

**Appendix D**

**Assigning Parents in Cervus**

## TO RUN ALLELE FREQUENCY ANALYSIS USING CERVUS

1. Create a genotype file containing ALL GENOTYPES (Offspring, Mothers, and Fathers). This data can be found in 'XX\_FINAL\_GENOTYPES\_2007-2014'

Change HMSC\_ID to ID and place in directory labeled for the corresponding 'offspring' year (e.g. 2015) at D:\Box Sync\N Santiam Chinook\analyses\replicating 2015\cervus

Save as a csv file and title all.genotypes\_NS15

2. 'Run Cervus' and select the allele frequency analysis

3. Designate 'input' file as 'All genotypes' file and save output in same directory

5. Tick 'read locus names and 'header row', designate ID as second column, first allele as third column, and 12 loci

###### TO SIMULATE PARENTAGE ANALYSIS

1. Designate input allele freq file (from previous step) [default]

2. List the number of offspring for this year

3. List the number of Mothers for the year range (e.g. 2007-2009)

4. Estimate the proportion of Mothers sampled 0.95

5. List the number of Fathers for the year range (e.g. 2007-2009)

6. Estimate the proportion of Fathers sampled 0.95

7. Prop. loci typed = 1

8. Prop. loci mistyped = 0.01 [default]

9. Specify '7’ typed loci

10. Specify 'LOD' confidence

11. Keep the confidence levels the same

######## TO RUN A PARENTAGE ANALYSIS

1. Create a csv file listing ONLY sample IDs for:

1. 20XX\_Offspring\_IDs

2. 20XX\_Fathers\_IDs

3. 20XX\_Mothers\_IDs

## CHECK TO MAKE SURE THE NUMBER OF ROWS IN THESE FILES MATCH WITH THE SIMULATION PARAMETERS (NUMBER OF INDIVIDUALS)

2. Designate these three files after selecting the 'parentage analysis with known sexes'

2. Follow the prompts and make sure to designate candidate ID's in column (NOT ROWS)

3. Designate the 'All genotypes' file

3. Select 'All parents with positive LOD scores'

4. Select 'Joint LOD'

######CONFIRMING/REJECTING ASSIGNED INDIVIDUALS OR PAIRS

1. Scroll through offspring row one sample at a time

2. If a mother or father pair are mismatched at more than 2 loci, this is a misassignment

3. If a mother and father (BUT NOT TRIO) match at the same number of loci, this is a misassignment for both

4. If an offspring was only typed at 7 loci, they must ALL match

5. For Trios, make sure the parents’ years MATCH