2bRAD Data

Using a 2bRAD reduced-representation sequencing approach, we sequenced 279 individuals and 18 technical replicates for a total of 297 samples across 4 lanes of Illumina HiSeq. The technical replicates are necessary for quality assessment and recalibration of downstream analyses. 245 of these samples had sufficient sequencing depth for subsequent genotyping and analysis (greater than 300,000 sequencing reads). We mapped reads from 19-20 individuals and 2 or 3 technical replicates from each of the three populations to a draft O. lurida genome of X # of scaffolds over 10kb in length. These individuals included 13 of the MBD-BS-Seq samples. For each sample in this subset, approximately 30% of reads mapped to the draft genome using Bowtie2. After all filtering and genotyping steps using GATK, VCFTools, and custom scripts by Mikhail Matz, we identified 3,510 single nucleotide polymorphisms (SNPs) confidently genotyped in at least 50% of these samples. With the full genome, we anticipate this number increasing substantially.

To get an estimate of population differentiation among our three populations, we thinned our dataset to one biallelic SNP per 2bRAD tag and selected 882 SNPs genotyped in at least 70% of the 59 individuals. Using Weir and Cockerham's 1984 FST estimator for each SNP, we found a mean FST of 0.0528 for all 59 individuals as three subpopulations. Eight SNPs elevated values of FST over 0.4. Average pairwise FST between the Oyster Bay and Hood Canal populations is 0.0668, with 24 SNPs having an FST over 0.4. The program Bayescan will be used on the full dataset of 245 individuals to identify other potential outlier loci that may be under selection. Once the genome has been annotated, we can assign putative function to these loci.

To further investigate population structure, we used the maximum likelihood-based program Admixture which assumes a set number of ancestor populations and then assigns proportions of each individual's genome to these ancestor populations. To meet the program's criteria for linkage equilibirum among loci, we filtered our dataset to 825 biallelic SNPs located at least 5kb apart. Admixture uses a cross-validation scheme in order to help the user estimate K – the number of ancestor populations. This cross-validation scheme indicates a K of 1 for our samples. Plots of the genetic admixture proportions for K = 2,3 and 4 are below.

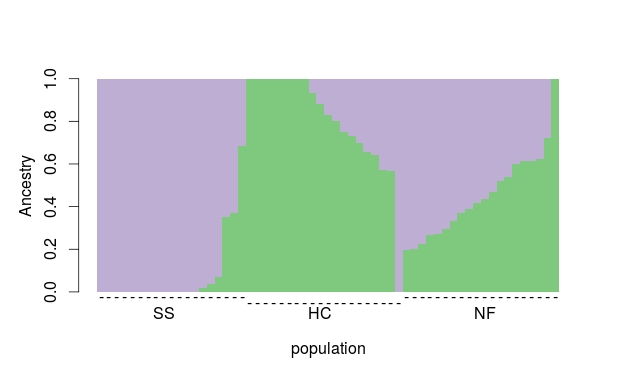
Illustration 1: K=2

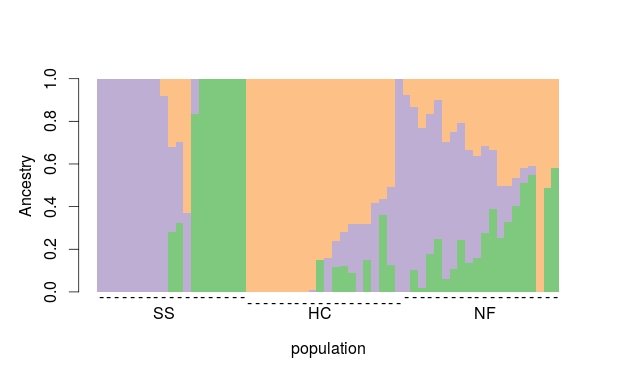
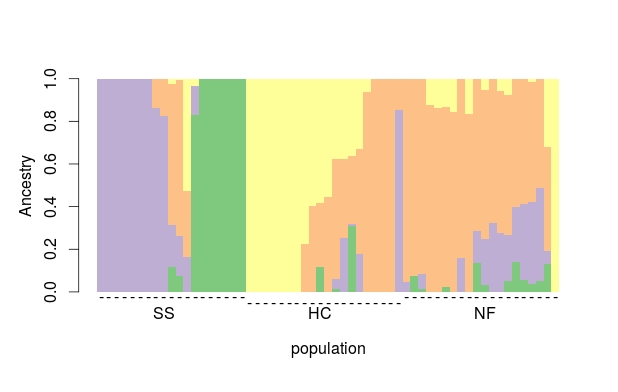
Illustration 2: K = 3

Illustration 3: K = 4

These results indicate that these three populations likely did not arise from distinct ancestor populations and may have ongoing gene flow.