Abstract

To do

Introduction

- Puget Sound chum salmon populations
  - Population structure
  - life history variation
  - ESA listing - summer chum ESU
  - Effective population size
- Genotyping duplicates
  - Legacy of the salmonid WGD
  - uncharacterized regions of the genome
  - first approach in salmon using next-gen seq data.
- Genome scan
  - map-assisted
  - paired population design
  - draw on synteny / orthology to interpret results
- Map
  - consensus map w/centromeres
  - synteny
  - annotation

Methods

Population Genetics

Sequence analysis and genotyping

- reference-based Stacks analysis. Reference constructed from Waples 2015
- alignment with BWA-mem, remove low-quality alignments
- identify varainats and assign genotypes with pstacks
- filter genotypes and individuals to produce final data set.
- Does allelic bias filtration have a place here?

Does allelic bias filtration have a place here?

Individual-based analyses

- Genotyping duplicate loci using the dominance coding suggested by Patterson (2006). Individual-based analyses
- PCAs - How do these two methods compare? - maybe measure info loss? Use a Procrustes analysis to find an optimal transformation. This allows the superimposition of one result onto another set of axes, through rotation and stretching. produced a procrustes similarity. Other option is a CCA (canonical correlation analysis)
- Formal tests for population structure - tracey-widom stats. Breakdown of population structure and lower PC axes

Population-based analyses

- MAF, Heterozygosity
- phylogenetic tree
- Genome scan - gloabal or paired populations -
- Fst across the genome. Use LOESS (local regression) to reveal regions of elevated differentation. Benefits of this approach vs a bootstrapping methods (eg. Hohenlohe 2010)
- Effective population size -
  Effective population size was estimated for each population using the LD method implemented in the LDNe software package (Waples and Do). The LD methods estimates average correlation of alleles at pairs of loci (r2). The mean pairwise r2 value across unlinked loci provides

an estimate of contemporary effective population size (WAPLES). The linkage map was used to ensure that only pairs of loci not co-located on a chromosome were used to calculate mean r2.

Linkage map

Sequence analysis and genotyping

Identification of paralogs follows Waples (2015)

Map construction follows McKinney (2015), Waples (2015)

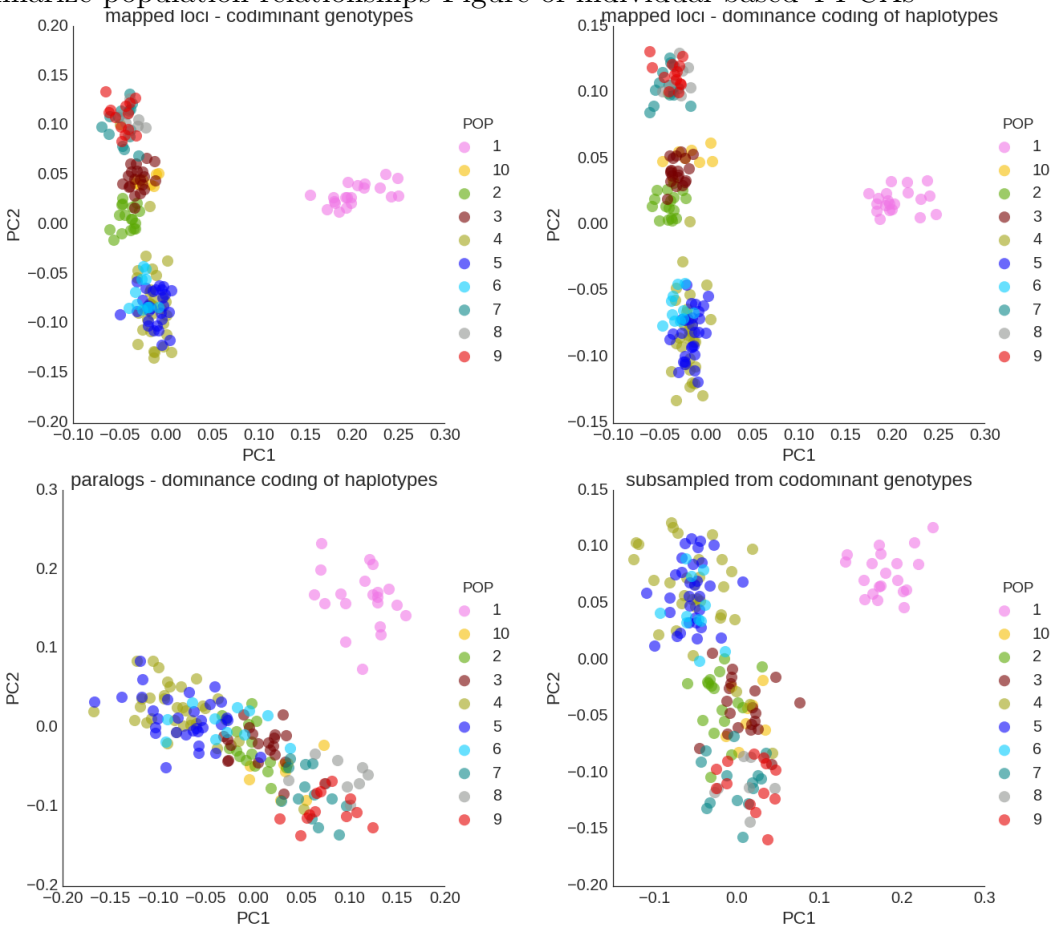
Synteny - relation to genetic resources

- Chinook salmon
- Atlantic salmon

Results

Population Genetics

Summarize population relationships Figure of individual-based 4 PCAs



paralogs have similar neutral patterns of population structure. measure information loss from dominance coding

discuss population vs individual based results

can we demonstrate contained within paralogs by bootstrapping Genome scans - LG regions highlighted.

Effective population size

Ascertainment Bias

Demonstrate ascertainment effect when using only loci on linkage map - effect on allele frequencies.

Linkage map

Identification of paralogs

congruence of identified homeologs across families (supplemental table)

Consensus linkage map

Table (Figure?) placement of centromeres

table (supplemental) paralogs note the distribution of paralogs matches the pattern found in other salmonids syntenic relationships - per LG Table (supplemental?)

Discussion

To do

Coalescent!

Does this turn into two papers?

- linkage map and individual-based analyses - including duplicated loci

- inference of adaptation -associated life history variation - Fst across the genome.

Supplemental

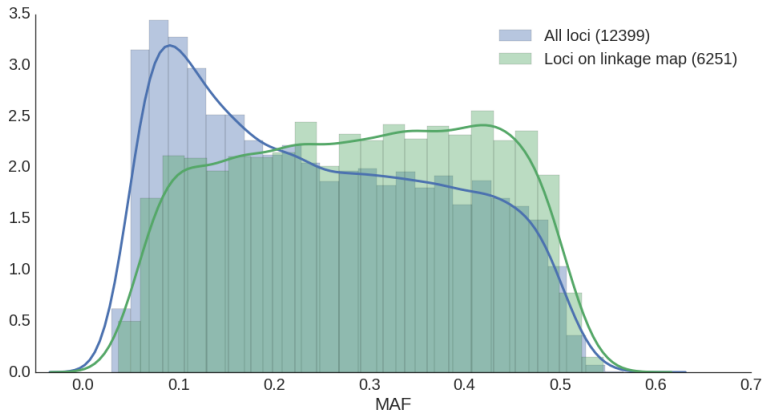


Figure 1: Minor allele frequency (MAF) distribution for all genotyped loci (blue) and just the loci placed on the linkage map (green). The rightward shift in the MAF distributino shows the effect of ascertainment bias within the mapped loci.

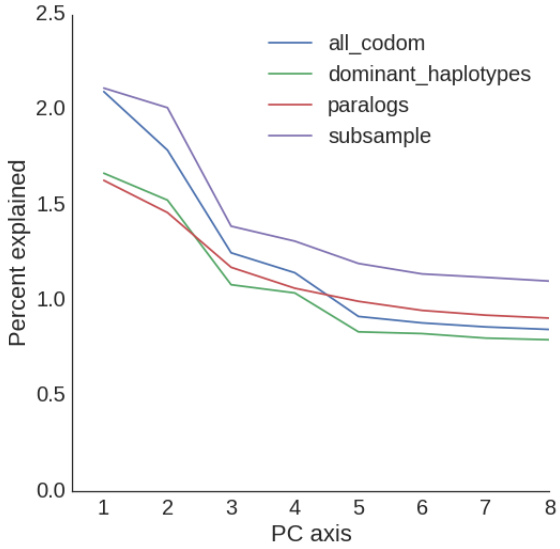


Figure 2: Percent variance explained (eigenvalue) for the first eight PC axes of each locus set. Notice the similarity between the two bi-allelic sets and the two haplotypic sets.

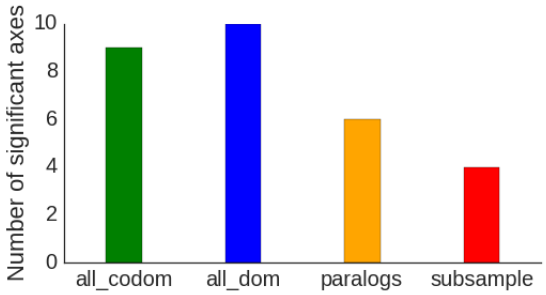


Figure 3: Number of significant PC axes as determined by the Tracey-Widom test.

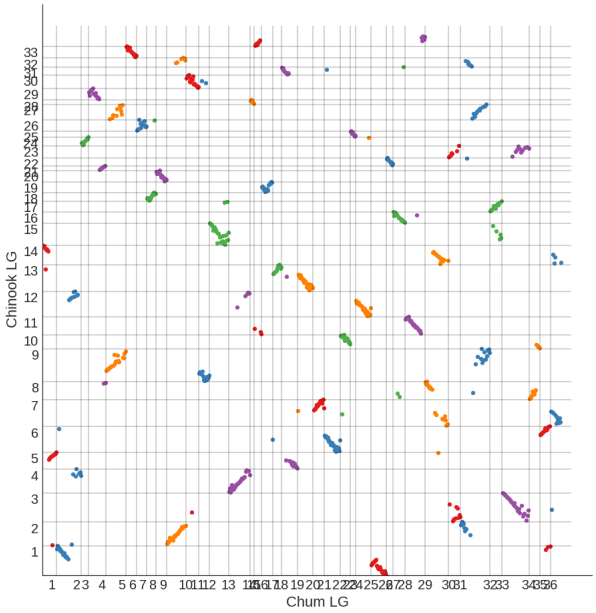


Figure 4: Oxford grid - Chum and Chinook linkage groups. Loci are positioned according to the order within each genome.