HW3\_bkretzle

Bailey Kretzler

11/15/2021

### Background

Trees in the genus *Populus* (poplars) are important for the bioenergy sector. Within this genus, two closely related species (*Populus trichocarpa* and *Populus balsamifera*) are of interest due to their wide geographic range and ability to hybridize and adapt to local environments (references). In the pacific northwest, there is overlap in these species geographic ranges which leads to mix ancestry of trees along this gradient. Such mixed ancestry may contribute to trees abilities to adapt to their local environment and as such may regulate phenological response to environmental cues. To explore this relationship, stem cuttings were collected from 6 transcripts spanning the hybridization gradient. These cuttings were rooted and leaf tissue was collected for DNA extraction and paired end sequencing via the Illumina NovaSeq platform. Cuttings were transplanted to a common garden in Vermont where they were monitored for phenological changes. During this monitoring period, climate data was also collected to assess if phenological processes varied by environmental cues. I hypothesize that their will be overlap in the genomic regions associated with phenological traits such as bud flush and response to climate characteristics. Further the genes shared among these traits are likely to be associated with breaking of dormancy at the cellular level.

### Bioinformatics pipeline

In order to assign local ancestry, admixture was determined for each individual with a K of 5. The admixture file, along with the metadata was used to assign broad ancestry for each individual (balsamifera, trichocarpa, or admixed) based on the proportion of each genome attributed to each K group. vcftools was used to make vcf files containing the reference individuals for *P.trichocarpa* and *P.balsamifera* as well as the SNP positions. A while read loop was used to make the vcf file for the admixed reference individuals by reading in files one line at a time. These vcf files were then matched to a sample ID and passed to loter which inferred local ancestry along chromosome 18. The text file output was ran through datamash, which converted haploid format (0/1 for each homolog) to diploid (0/1/2 for both homologs combined) at each SNP site. The diploid file was converted to a format Plink would understand by adding SNP positions and mock alleles at each site. A .fam file containing information about each sample was also generated. These files were converted to Plink format and used to determine local ancestry at each SNP site. Finally, Plink was used for admixture mapping by modeling the generalized linear relationship between local ancestry and phenotype and climate conditions. All Plink files were brought into R to analyze and visualize the relationship between climate or phenotype and proportion of *P.balsamifera* ancestry in chromosome 18. Admixture mapping results were used to determine outliers for each trait and if genomic regions containing these outliers overlapped between traits using the GenomicRanges package in R. Genes of interest were determined by aligning regions with reference genomes then matched to gene ontology (GO) terms using PopGenie.

### Results

From the analysis, it is clear that ancestry has a relationship with chilling degree days (CDD), growing degree days (GDD), final freezing event (FFE), and bud flush. All traits showed a significant linear relationship with the proportion of *P.Balsamifera* ancestry - which is displayed in figure 1. CDD, GDD, and FFE all show significant positive relationships with ancestry where lower levels of *P.balsamifera* ancestry correlate to lower values for each trait. Bud flush, however, shows a signficant negative relationship with ancestry where lower levels of *P.balsamifera* ancestry correlate to higher values for budflush. Additionally, certain genomic regions show strong association with each trait (figure 2). In general the regions associated with these traits are clustered together and the genome regions seem to be close but different for each trait. Further, when we look at inferred local ancestry along the chromosome the regions seem to be near where changes in frequency of *P.trichocarpa* ancestry occur. Only GDD associated regions showed overlap with the bud flush associated regions, indicating that it may be the only climatic variable influencing bud flush. Unfortunately, when GO analysis was done on the overlapping regions there was no evidence of functional enrichment or at least the functions of the genes have yet to be classified. CDD, GDD, and FFE had 4, 6, and 4 candidate gene regions with 54, 49, and 27 genes respectively. GDD and FFE genes showed little enrichment for previously classified functions, where as CDD showed considerable function enrichment in GO categories “Biological Process” and “Molecular function”. Most of the enriched functions for CDD were associated with transported activity.

### Conclusion

It is apparent that ancestry has an impact on both phenology and response to climatic conditions. Candidate gene regions for all traits were concentrated toward the end of the chromosome. Little overlap was seen for climatic traits as the phenological trait bud flush, however GDD did show some overlap. Additionally, there is enrichment among CDD genes for functions associated with transport which may be linked to the transport of biochemicals for winter storage or breaking of dormancy. It is worth noting that the lack of functional enrichment seen for GDD and FFE and the overlap of GDD with bud flush may be due to unknown functions for candidate genes. Therefore it would be naive to assume that these candidate gene regions have no important function. In future studies it would be helpful to pin down the roles of these genes in overall plant growth and phenology. Synthesizing the results, it is clear that admixed individuals may be more cold adapted as they are associated with a higher number of chilling degree day and a later final frost event. In future studies it would be useful to determine where specific gene regions were inherited from, I had hoped to make a chromosome map to visualize this but wasn’t sure how. Lastly, it would be important expand the species used in the analysis to all North American poplars to understand how ancient introgression may impact modern phenotypes.