Population genomics of *Picea rubens*

Thomas O'Leary

February 25, 2020

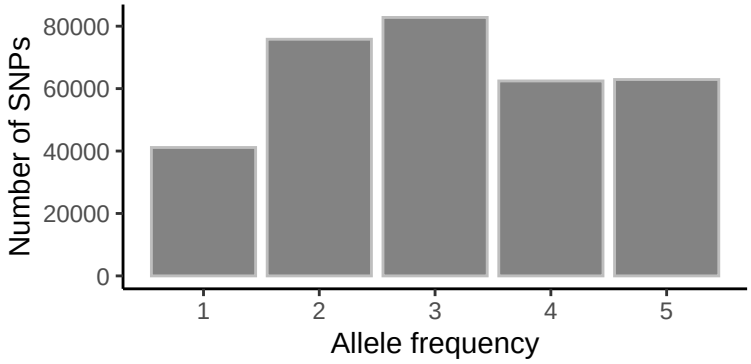
Background

Picea rubens Ecology

In the context of a rapidly changing climate many organisms will be forced to either adapt, migrate, or go extinct. In the case of *Picea rubens*, a red spruce that thrives in cool and moist climates, this may lead to futher range contraction or eventual extinction if the species is unable to adapt in time. As glaciers began to melt approximately 20,000 years ago, *Picea rubens* was forced to retreat to isolated populations along the mountain tops of the Appalacian range, isolated from the northen core of red spruce. As the climate continues to warm, these isolated populations may pop out of existence due to the increased environmental stress. However, because evolution can only act on standing genetic variation these isolated populations may represent an important genetic resource for the species at large, and may help inform conservation efforts. The ultimate aims of this study are to (i) describe the population structure and genetic diversiy of *Picea rubens* along its current range, (ii) identify loci that show signs of positive selection and (iii) map the genetic basis of these adaptive phenotypes. In this write up, I will only begin to address the first aim using a few genetic diversity metrics including nucleotide diversity (π), Wattserson's estimator (θ), and Tajima's D (Nei and Li 1979; Watterson 1975; Tajima 1989).

Sample collection, library preparation, and sequencing

Whole genomic DNA was extracted from needle tissue that was collected from a total of 340 mother trees in 65 populations, of which 110 trees were from 23 edge populations. 80,000 120 bp probes were designed for exome capture based on the multiple developmental and tissue type transcriptomes from the related white spruce *Picea glauca* (Rigault et al. 2011; Yeaman et al. 2014). Approximately 95% of the probes were designed within exomic regions with the remaining 5% in intergenic regions, covering a total of 38,570 unigenes. The probes were blasted against the *P. glauca* reference genome to ensure at least 90bp of 85% identity. 250ng to 1 μg of genomic DNA was mechanically sheared to an average length of 400 bp, exome fragments were enriched using designed probes and following barcode adaptation the libraries were pooled and paired-end 150 bp sequenced on an Illumina HiSeq X.



Bioinformatics Pipeline

The quality of the raw reads were accessed using FastQC VERSION (Fastqc citation). To remove low quality sequence data the raw fastq files were trimmed using Trimmomatic VERSION (Bolger, Lohse, and Usadel 2014) and the cleaned reads were vizualized again with FastQC. The cleaned reads were mapped to a reduced verion of the Norway spruce

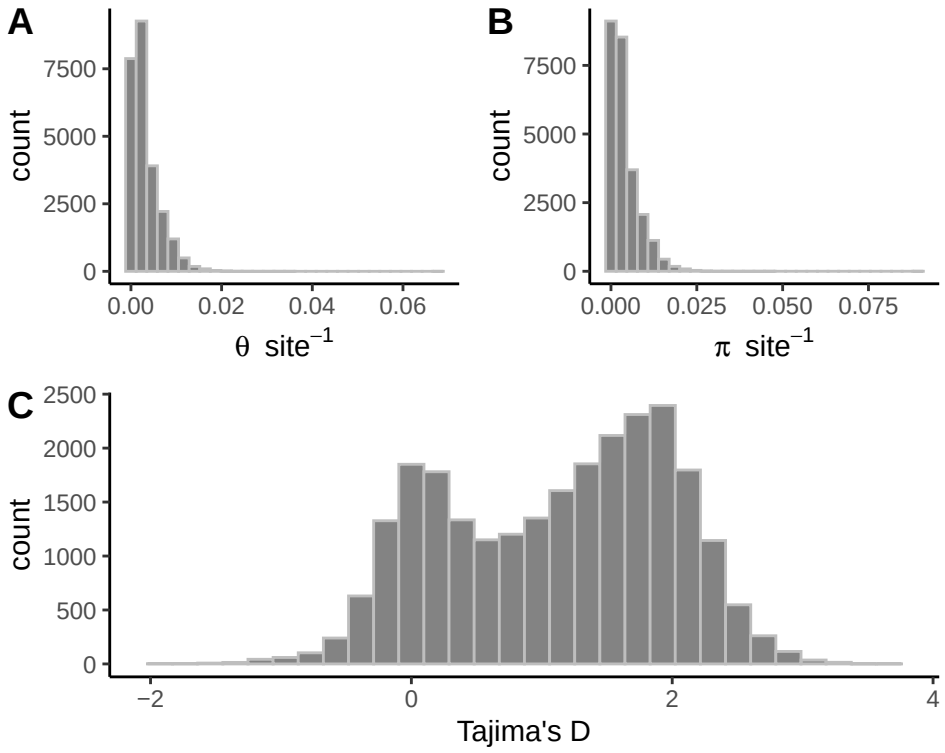
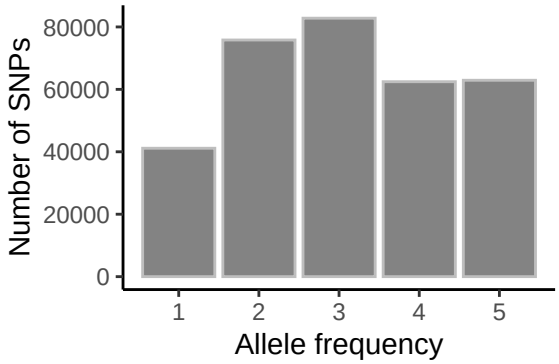


Figure 1: Figure caption goes here!!!!!

Picea abies reference genome (Nystedt et al. 2013) using bwa VERSION (Li and Durbin 2009). A reduced *Picea abies* reference genome was used because there is no available *Picea rubens* genome and the exome capture technique meant that only a small fraction of the genome near the designed probes would be sequenced.

Results

The



$\theta = 4N_e\mu$

Given that θ is the .... and the effective population size (N_e). The estimated per base per year mutation rate of *Picea* is $2.2\text{x}10^{-9}$ (Nystedt et al. 2013).

$D = \frac{\pi}{S}$

Conclusion

References

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A flexible trimmer for Illumina sequence data." *Bioinformatics* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Li, Heng, and Richard Durbin. 2009. "Fast and accurate short read alignment with Burrows-Wheeler transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Nei, M., and W. H. Li. 1979. "Mathematical model for studying genetic variation in terms of restriction endonucleases." *Proceedings of the National Academy of Sciences of the United States of America* 76 (10): 5269–73. https://doi.org/10.1073/pnas.76.10.5269.

Nystedt, Björn, Nathaniel R. Street, Anna Wetterbom, Andrea Zuccolo, Yao Cheng Lin, Douglas G. Scofield, Francesco Vezzi, et al. 2013. "The Norway spruce genome sequence and conifer genome evolution." *Nature* 497 (7451). Nature Publishing Group: 579–84. https://doi.org/10.1038/nature12211.

Rigault, Philippe, Brian Boyle, Pierre Lepage, Janice E.K. Cooke, Jean Bousquet, and John J. MacKay. 2011. "A white spruce gene catalog for conifer genome analyses." *Plant Physiology* 157 (1): 14–28. https://doi.org/10.1104/pp.111.179663.

Tajima, Fumio. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." *Genetics* 123: 585–95.

Watterson, G. A. 1975. "On the Number of Segregating Sites in Genetical Models without Recombination." *Theoretical Population Biology* 7: 256–76. https://doi.org/10.1039/b316709g.

Yeaman, Sam, Kathryn A. Hodgins, Haktan Suren, Kristin A. Nurkowski, Loren H. Rieseberg, Jason A. Holliday, and Sally N. Aitken. 2014. "Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (Pinus contorta) and interior spruce (Picea glauca × Picea engelmannii)." *New Phytologist* 203 (2): 578–91. https://doi.org/10.1111/nph.12819.