Population Genomics Tutorial 1: Intro to our dataset on red spruce genomics

SRK

Fall 2025

1. Red spruce, *Picea rubens*

Red spruce is a coniferous tree that plays a prominent role in montane communities throughout the Appalachians. It thrives in the cool, moist climates of the high elevation mountains of the Appalachians and northward along the coastal areas of Atlantic Canada. In the low-latitude trailing edge of the range, populations are highly fragmented and isolated on mountaintops. These "island" populations are remnants of spruce forests that covered the southern U.S. glaciers extended as far south as Long Island, NY. As the climate warmed at the end of the Pleistocene (~20K years ago), red spruce retreated upward in elevation to these mountaintop refugia, where they are now highly isolated from other such stands and from the core of the range further north.

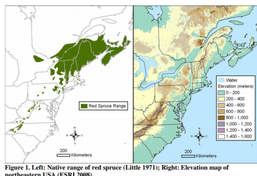

Figure 1. Left: Native range of red spruce (Little 1971); Right: Elevation map of northeastern USA (ESRI 2008).

Because of its preference for cool, moist climates, red spruce shows climate sensitivities that may make it especially vulnerable to climate change. This makes assessing the amount and distribution of genetic diversity across the landscape of red spruce an important conservation issue! Ultimately, we want to use genomic insights to help inform conservation biologists working to restore red spruce and evaluate the potential for assisted migration (a form of human-mediated dispersal) to offset the loss of adaptation under climate change.

A close partner in this effort is the Nature Conservancy and the Central Appalachian Spruce Restoration Initiative (CASRI) – a multi-partner group dedicated to restoring and enhancing red spruce populations to promote their resilience under climate change.

CASRI | CENTRAL APPALACHIAN SPRUCE RESTORATION INITIATIVE

Some videos of the collaboration between UVM and CASRI on red spruce restoration:

- Building Resilience
- Seeds of Hope

Since 2015, the Keller Lab has been studying the genetic basis of climate adaptation across the entire distribution of *P. rubens*. Our main goal is to use genomics to aid conservation of red spruce under climate change through a better understanding of **(1) how genetic diversity diversity is distributed across the range and how that reflects the demographic history of population expansions, bottlenecks, gene flow, and divergence** and **(2) identify regions of the genome that show evidence of adaptation in response to abiotic climate gradients.** We hope to use this information to inform areas of the range most likely to experience climate mal-adaptation, and to help guide mitigation strategies such as sourcing seed for restoration and assisted migration.

2. Summary of prior population genomics research on red spruce

Our recent work funded by NSF focused on early-life fitness of seedlings in response to population genomic variation and climate adaptation. That work sampled seeds and needle tissue from 340 mother trees at 65 populations spread throughout the range and generated population genomic data through exome capture sequencing. Seedlings from each mother were grown in multiple common gardens and measured for fitness traits. Based on these data, some of the insights we gleaned were:

- Red spruce has very low genetic diversity and has an Ne that has been declining for thousands of years Capblancq et al. 2020
- Populations are differentiated into 3 geographically separated clusters of genetic ancestry in the north (core) mid-latitude (margin) and southern (edge) regions of its range Capblancq et al. 2020)
- Local populations level of genetic diversity and frequency of deleterious mutations (aka, "genetic load") were related to how well seedlings survived and grew under greenhouse conditions Capblancq et al. 2021
- Seedling growth traits in common garden experiments showed heritable genetic variation and genetically-based trait divergence among the 3 ancestry groups Prakash et al. 2022
- Certain genomic regions showed strong allele frequency clines along climatic gradients indicative of selection, with gene functions often related to abiotic stress (heat and drought) Capblancq et al. 2023
- There was a hint in some of the results that hybridization with black spruce might be playing a role in some of the above results, but we lacked genomic data from black spruce to nail that down.
- Other studies have also found evidence for hybridization between red and black spruce (e.g., de Lafontaine et al. 2015, de Lafontaine and Bousquet 2017, Bashalkhanov et al. 2023), but the evolutionary outcomes for what this means for adaptation to local environments are still not well understood.

Let's brainstorm about some issues these data couldn't address, or where the inference of evolutionary history or adaptation were constrained by aspects of the experimental design or sampling?

What questions would we want to ask next?

3. A new dataset for analysis

The data we'll be analyzing here consist of exome capture data for adult red spruce growing in a provenance trial in northern New Hampshire near the town of Colebrook. Here are the details:

- 95 individuals sampled from 12 populations across the range (N=190 red spruce)
- Individuals were grown from seed and planted out into the provenance trial as 2 year old seedlings in 1960
- Multiple studies have assessed survival, growth (height DBH), and cold tolerance of these individuals at multiple time points over the last 60 years
- Needle tissue was sampled from surviving red spruce in the trial in May 2020 for genomic DNA
- We also sampled 18 black spruce individuals from natural stands in 2 locations distant from red spruce's range (MN and MI).
    - These will be useful for detecting black spruce ancestry in the sampled red spruce populations, if it exists!
- We used the same exome-capture probe set as Capblancq et al. (2020). *Why exome capture instead of alternatives (WGS, or RAD/GBS)?*
- Exome-capture was designed based on transcriptomes from multiple tissues and developmental stages in the related species, white spruce (*P. glauca*).
- Bait design used 2 transcriptomes previously assembled by Rigault et al. (2011) and Yeaman et al. (2014).
- A total of 80,000 120 bp probes were designed, including 75,732 probes within or overlapping exomic regions, and an additional 4,268 probes in intergenic regions.
- Each probe was required to represent a single blast hit to the *P. glauca* reference genome of at least 90bp long and 85% identity, covering **38,570 unigenes**.
- Libraries were made by random mechanical shearing of DNA (250 ng -1ug) to an average size of 400 bp followed by ligation of barcoded adapters, and PCR-amplification of the library. SureSelect probes (Agilent Technologies; Santa Clara, CA) were used for targeted enrichment following the SureSelect Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library protocol.
- Libraries were sequenced on an Illumina HiSeq to generate paired-end 150-bp reads.

Here's the table of sample population codes used in file naming and their source localities

PopCode	PopName	State/Province	Country	Latitude	Longitude	Elevation
2019	Indian_Gap	North_Carolina	USA	-83.45000	35.60000	1620
2020	Glade_Run	West_Virginia	USA	-79.83333	38.63333	1160
2021	Bear_Meadows	Pennsylvania	USA	-77.75000	40.72910	550
2022	October_Mtn_State_Forest	Massachusetts	USA	-73.25000	42.36667	550
2024	Upper_Jay	New-York	USA	-73.66667	44.41667	610
2027	Pillsbury_State_Forest	New-Hampshire	USA	-72.08333	43.20000	500
2030	Amherst	Maine	USA	-68.38333	44.90000	150
2030	Amherst	Maine	USA	-68.38333	44.90000	150
2032	Valcartier_Forest_Experimental_Station	Quebec	Canada	-71.55000	46.91667	270
2100	Sheet_Harbour_Waters	Nova-Scotia	Canada	-62.73333	45.20000	60
2101	Corberrie	Nova-Scotia	Canada	-65.90000	44.16667	60
2103	Centra_Acadia_Forest_Experiment_Station	New-Brunswick	Canada	-66.33333	46.03333	60
2505	Eastern_Acadia_Forest_Experiment_Station	New-Brunswick	Canada	-66.20000	46.03333	60

And here's a map of the sample sites within the red spruce range: