

Analysis of Variance and Minimum Spanning Tree of Chlorotypes Within Populations of *Larix laricina*, a Boreal Conifer

Carlyn Albright, Evan Bishop, Madison Beckstedt, Sierra Blair
Molecular Ecology and Evolution (FNR 505000) Data Analysis Project

Introduction

- Tamarack (*Larix laricina*) is an early-successional transcontinental boreal conifer. The authors of the original paper (Warren et al 2016) aimed to infer the species biogeographical history and to assess the putative genetic imprint left by interspecific competition during post-glacial migration, using cytoplasmic DNA and fossil data of the tamarack trees spanning the North American boreal forest.
- The authors surveyed 621 trees for chloroplast DNA which they ran through BAPS 6.0 using a Bayesian analysis of population structure to identify groups of genetically homogeneous populations based on cpDNA polymorphisms. They found two major groups and one distinct minor group. Their results are found in Fig. 3.
- We used Analysis of Molecular Variance (AMOVA) to test if these three groups are differentiated. AMOVA produces estimates of variance components and F-statistic analogs reflecting the correlation of haplotypic diversity at different levels of hierarchical subdivision. This test will show the percent variation among groups, among populations within groups, and within populations.
- Running an AMOVA on the three groups, using chlorotypes will further confirm that these three groups are genetically different from each other.
- A minimum spanning tree of the chlorotypes was made using the AMOVA output from Arlequin. We wanted to make a new tree that clearly shows which chlorotypes are most common within all three groups and their relationship to each other.

Methods

- Chlorotype data was taken from the supplemental materials found in Warren et al 2016.
- AMOVA was used using Arlequin to test if the three groups identified by Bayesian analysis are differentiated among groups, among populations within groups and within populations.
- A minimum spanning tree of the chlorotypes was made using the output from Arlequin and running it in Hapstar.

Results

Source of Variation	Degrees of Freedom	Sum of Squares	Percentage Variation
Among groups	2	40.418	9.46%
Among populations within groups	42	124.78	10.29%
Within populations	539	601.871	80.25%
Total	583	767.068	100.00%

Table 1: AMOVA results between the three populations including degrees of freedom, sum of squares, and percentage of variation as specified in the original paper

Source of Variation	Fixation Indices	P-value
Among groups within populations	FSC	0.11366
Among groups	FST	0.19755
Within populations	FCT	0.09464

Table 2: AMOVA results of fixation indices along with P-values among different sources of variation among groups within populations, among groups, and within populations.

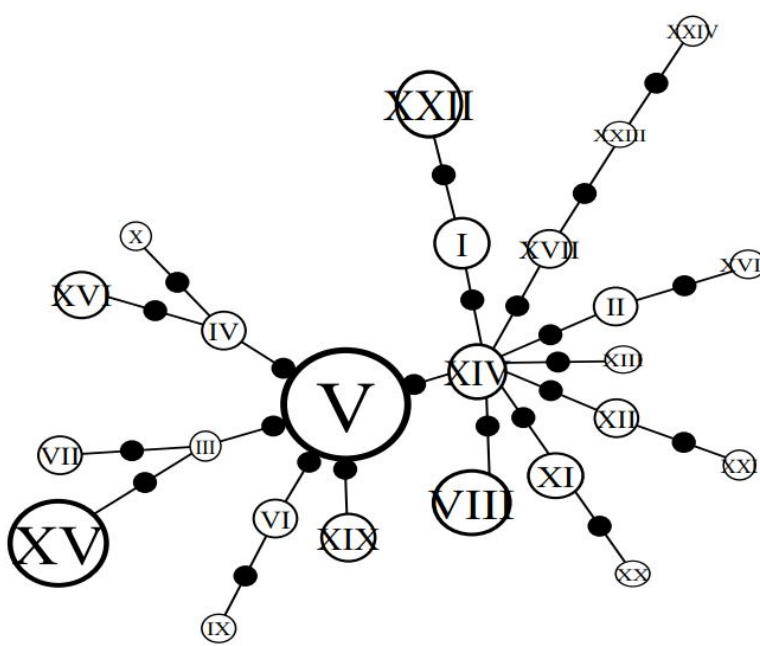


Figure 1: Minimum spanning tree of chlorotypes with no alternative links. Each roman numeral represents a different chlorotype. Most common chlorotypes are enlarged. V=260 individuals, XV=83 individuals, XXII=37 individuals.

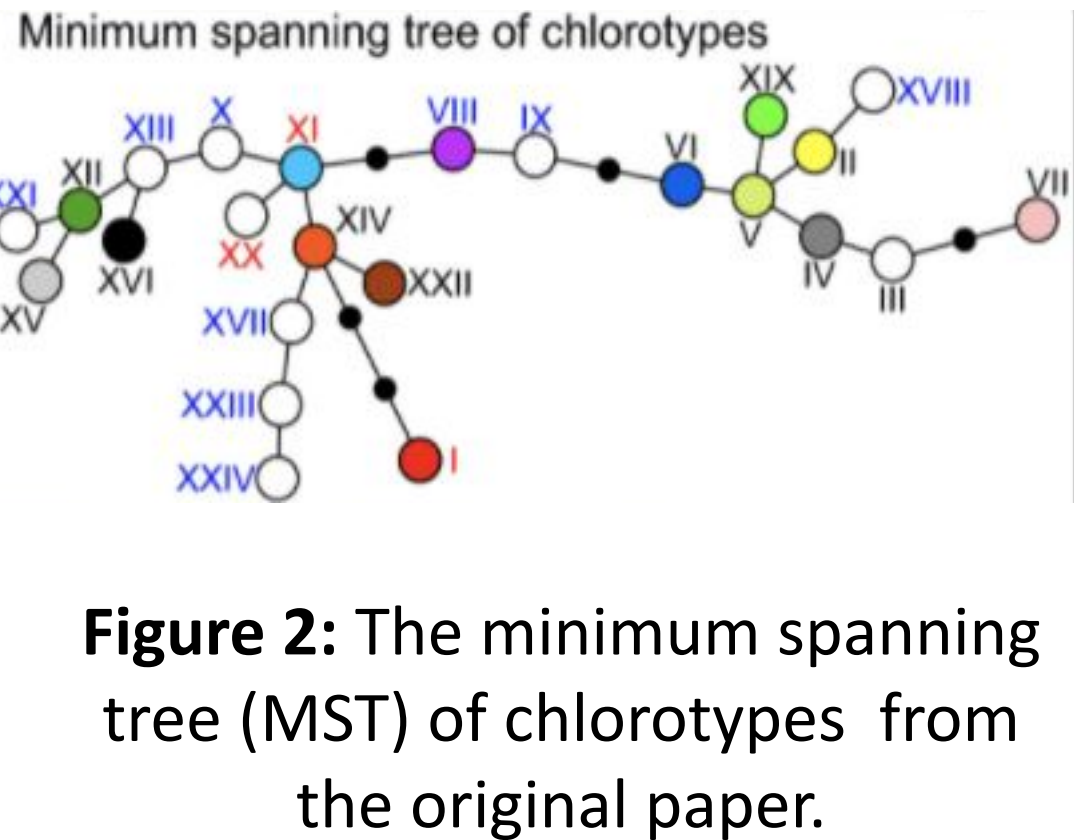


Figure 2: The minimum spanning tree (MST) of chlorotypes from the original paper.

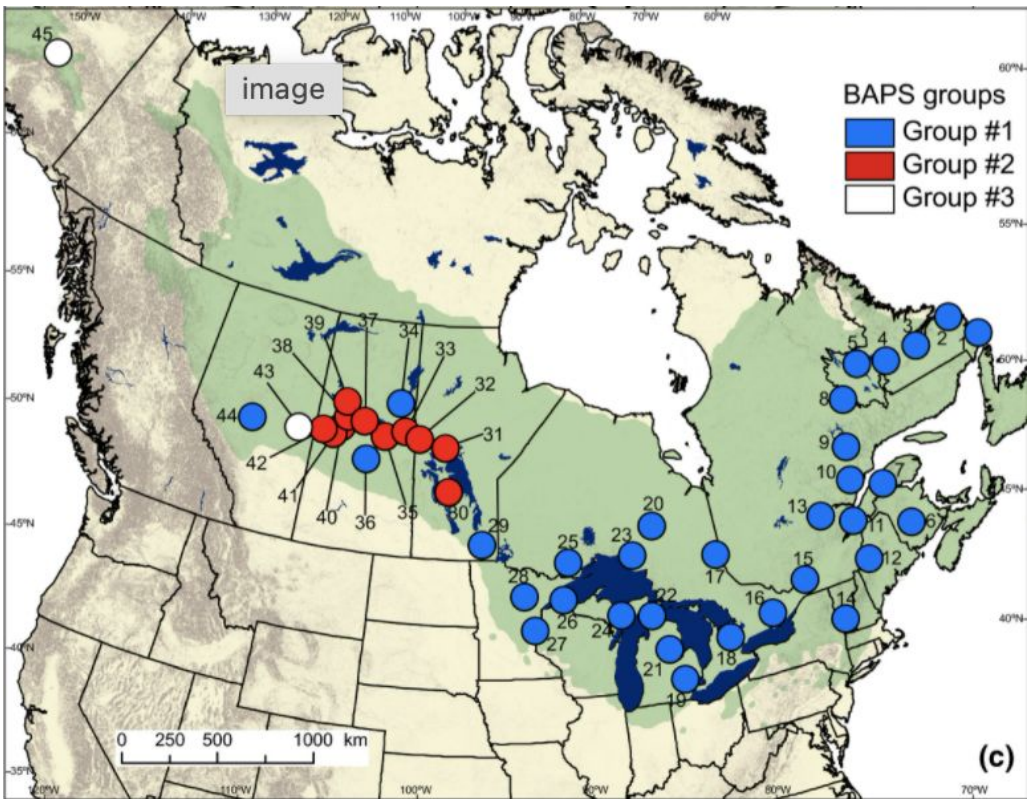


Figure 3: Bayesian clustering of three different groups from the original paper.

Discussion

- The AMOVA analysis made it clear that the three groups identified by Warren et al 2016 were indeed genetically different from each other. The AMOVA results showed the percent variation among groups, among populations between groups and within groups. Each source of variation had low p-values of <0.05. Even though among groups had only 9.46% variation, there was a significant difference between the three groups. This analysis added to the original paper by confirming that there is a clear, genetic differentiation between three groups of tamarack trees in the North American boreal forest.
- The p-values of the fixation indices in Table 2 indicate that there is some isolation amongst the three different groups. Since the FST p-value among groups is 0.19755 this indicates that there is some isolation to some degree among groups and that the group is most likely not breeding with one another, but are not high enough to be considered fixed. The p-values of FSC and FCT are not as high and therefore are not as isolated.
- The minimum spanning tree was used to show the relationships between chlorotypes from populations within the groups. Our version of the MST shows the most common chlorotypes as being enlarged. These enlarged chlorotypes were more helpful compared to the original MST, shown in Figure 2, since the authors did not visually show which chlorotypes were most common. Without the raw data it was not clear which chlorotypes were most common.
- The limitations of this analysis were that the data was not easy to put into different programs which resulted in less than optimal figures. Creating a better minimum spanning tree by identifying the geographic area the chlorotypes are in is recommended. A SAMOVA analysis would aid in understanding the role that geographic distance plays in genetic differentiation.

Literature Cited

Warren, E., de Lafontaine, G., G rardi, S., Senneville, S., Beaulieu, J., Perron, M., Jaramillo-Correa, J.P. and Bousquet, J. (2016), Joint inferences from cytoplasmic DNA and fossil data provide evidence for glacial vicariance and contrasted post-glacial dynamics in tamarack, a transcontinental conifer. J. Biogeogr., 43: 1227-1241. <https://doi.org/10.1111/jbi.12675>