Populus Genetic Variation Networks

Matthew K. Lau

February 2, 2011

Contents

1 Project Concept

In order to understand the evolution of species it is important to understand the structure of genetic variation of the population. Primarily, populations have been treated as well mixed spatially homogenous units with the intent of reducing the complexity of evolutionary models. Recent developments in the field of landscape genetics have shown that the spatial context of populations for many species can drastically improve our understanding of the evolutionary dynamics of these populations. The integration of evolutionary biology and network theory has yielded useful tools for approaching questions related to 1) the structure of spatial variation in populations and 2) the scale at which management efforts should focus.

Here we apply the network theoretic methods to:

1. Elucidate the structure of genetic variation in populations of *Populus spp.*
2. Search for evolutionarily relevant units
3. Model the potential evolutionary dynamics of the species

2 Relict Data

2.1 Metadata

Email from Scott Woolbright Here are the relict SSR marker data. First row is marker name. Columns are sample ID and pop ID. The first 102 are parental samples from across each range (Fremont, narrowleaf, and trichocarpa). The rest are allele sizes. Missing data is coded as -1. Some populations have a lot of missing data (NPV and NGM in particular). Not sure how you want to handle that. Also, I've been looking at Eastern ranges vs. Central ranges. Eastern ranges are all of those starting at row 103 and ending on row. 228-230 are from a southern range. The rest are from the central part of the state. Let me know if you need anything else.

Original data file = RelictsforGenAlEx.xls

2.2 Analysis Notes

2.2.1 Outline of Network Construction Process

Calculate multi-locus genetic distances

Calculate genetic covariance matrix

Generate independence matrix

3 Research Notes and Avenues for Further Study

3.1 2 Feb 2011

1. WRITE-UP METHODS AND RESULTS FROM ANALYSES THUS FAR
2. To deal with missing values, assign the most abundant allele at that locus with probability determined by the abundances of that allele in the population.
3. Try doing an individual based network and then detect modules.