# Linking spatial distances to genetic distances in *Betula Alleghaniensis* (Yellow Birch)

BIOS 35502: Practicum in Environmental Field Biology

Student:

Paloma A. Nieves Rodríguez

Advisor:

Candice Lumibao

## 1. Abstract

Genetic ditances between individuals may be determined by their spatial distances. This correlation may be particularly present within range populations. I tested out this hypothesis by analizing DNA from individuals of the *Betula alleghaniensis* (yellow birch) species and relating it to the coordinates taken from each site. The population studied was located in the Upper Peninsula of Michigan/Wisconsin region, where yellow birch reaches its range limit. The results of this study showed no correlation between genetic and spatial distances. I discussed the possible influence of long distance dispersal on the lack of relationship between these variables, as well as the effects of random mating and inbreeding. Additionaly, I discussed the limitations encountered during this research, such as the time and spatial scale in which it was conducted. I concluded that strong correlations may still exist between genetic and spatial distances, and that these could be evidenced by more extensive studies.

## 2. Introduction

Along with most shrub species, forest trees commonly present low extinction rates. Because of their longevity and high rates of pollen flow, they are generally resistant to loss of genetic diversity (Hamrick 2004). Genetic diversity is an important factor because it determines the responses of organisms to environmental stress, natural selection and susceptibility to different diseases (Poczai et al. 2012). Higher genetic diversity allows populations to adapt in response to environmental changes (Fairbanks and Andersen 1995). Therefore, it is expected that many forest tree communities will remain unaffected against most disturbances (Lavergne et al 2010).

However, due to their slow reproduction rates, forest trees may still be vulnerable to rapid environmental changes. This is particularly important given the pace of present day changes,

mostly caused by an increase in human activities (Bell and Collins 2007). As the rate of human activities continues to intensify, the need to integrate genetic disciplines into environmental approaches becomes more urgent (Escudero et al. 2003). Ensuring that organisms possess sufficient genetic diversity so as to be able to adapt to new environments is the next step in conservation (Kimatu et al. 2012).

The colonization and establishment of a tree species into a new environment is determined by seed and pollen dispersal processes (Houle 1998). These processes however are influenced by numerous factors such variable wind patterns and canopy structures, as well as certain geographic barriers between source populations to the colonizing populations (Houle 1998). These are particularly more pronounced at range margins, where populations are small and disjunct. Subsequent extensive pollen and seed flow among these newly established individuals can lead to a lack of strong spatial structure in the distribution of genetic diversity as well as in individual haplotypes. On the other hand, lack of migration or gene flow will result in strong spatial genetic structure and with some haplotypes becoming fixed in a given locality and population (e.g. Petit et al. 1997). The proccess of dispersal can work in various ways. Due to all the factors that influence dispersal, it doesn't necessarily guarantee a favorable distribution. (Pease et al. 1989). Individuals may find themselves within a high population density in a poor environment, while favorable environments may have a low population density (Pease et al. 1989). Dispersal can also prevent range populations from diverging from central populations (Lavergne et al. 2010). This could lead to the development of strong spatial genetic structures (SGS) within marginal individuals.

The presence of SGS implies that there is a relationship between both spatial and genetic distances between individuals. Individuals found in one same area can, essentially share similar

genetic information (Escudero et al. 2003). Here, I examined the distribution of chloroplast DNA haplotype in *Betula allegheniensis* as a function of spatial distance. I proposed the hypothesis that, at closer distances, tree individuals would be genetically similar than those farther apart. This would especially be true within range populations. As a null hypothesis, spatial distances would have no relation to genetic distances.

## 3. Methods

### Study system and Site

I conducted research at the University of Notre Dame Environmental Research Center's East property, near the Great Lakes region. I studied individuals of the *Betula alleghaniensis* (yellow birch) species. *B. alleghaniensis* is an ecologically, and economically, important tree species, being a significant source of hardwood lumber (Houle 1998). It has characteristics of both late-succesional and mature forests trees. Although it has a high wind-dispersal capacity, its rates of establishment are quite low (Houle and Payette 1990).

*B. alleghaniensis* also reaches its range limit in the Wisconsin and Upper Peninsula of Michigan region. Because of dispersal limitations, range populations are expected to contain lower genetic diversity, and be more vulnerable to environmental changes than populations in center areas (Eckert et al. 2008).

To corroborate this, I made use of chloroplast markers, a reliable method of measuring genetic diversity in populations (Fairbanks and Andersen 1995). Another reliable method to measure genetic diversity is a spatial autocorrelation analysis, also used in this research, which looks for direct associations between the individual's traits by studying its Spatial Genetic Structure (SGS) (Barbujani 1987).

**Sample Collection**

I randomly collected leaf samples from 16 individuals of *B. alleghaniensis* across the property, and the geographic coordinates of each individual were taken. Samples were then stored at a low temperature for DNA analysis.

**Molecular analysis**

I extracted the DNA from the preserved leaf samples using the *E.Z.N.A. DNA Isolation System Kit* (Omega Biotek Company). I processed a total of 16 samples and prepared them for amplification.

DNA extracts were amplified initially at two chloroplast DNA regions: spacer region between *trnL-trnF* and *atpB-rbcL*. Amplifications were carried out as follows: A master mix was created for all 16 samples, containing 10x PCR buffer, 25 mM MgC12, 10.2 µm of dNTP mix, 0.6 µm of each primer and 3.4 µm of Taq The master mix was cycled with the following parameters: at 95C for 5 minutes, at 95C for 2 minutes, at 56C for 1 minute and 30 seconds, at 72C for 1 minute and at 72 C for 5 minutes for a final elongation.

PCR products were purified with Exo-Sap, using a master mix for all samples. The master mix was prepared with a concentration of 10.2 µm of Premix Dye, 42.5 µm of ABI buffer, 34 µm of aTPB primer and 110.5 µm of H2O over 13.4 µm of each PCR product. The Exo-Sap product was sent to the University of Notre Dame to get sequenced, using an ABI 3730 sequencing machine.

**Sequence Analysis**

Sequences were first aligned and manually edited using *Sequencher 5.2.4* (Gene Codes Corporation, Inc. 1995). I identified program errors by corroborating the chromatograms of each

sequence and proceeded to distinguish all Single Nucleotide Polymorphisms (SNPs). Because of sequencing errors that cut of some sequences short, the sample size was further reduced to 13 individuals.

cpDNA haplotypes were inferred based on SNPS as well as indels (insertions/deletions). Haplotype distances based on molecular distances were first calculated using *Arlequin 3.5* (Excoffier et al. 2005).

**Genetic and statistical analysis**

In order to examine whether genetic relatedness among individuals was correlated with spatial distance, genetic distance (or kinship coefficient) was calculated among all pairs of individuals using *SpaGedi v1.4* (Hardy and Vekemans 2002). A simple linear regression was performed using *R 3.1.1.*, utilizing genetic distances as a response variable against the corresponding pair-wise geographic distances, the explanatory variable. To provide a more robust analysis, I performed a Mantel Test using *IBDWS* (Jensen et al. 2005). I visualized the spatial distribution of haplotypes across my study site through *ArcGis* (ESRI, Redlands, CA, USA).

## 4. Results

A total of 729 bases from the *trnl-tnF* and 729 bases from the *atpB-rbcL* region were sequenced. However, I only considered the *atpB-rbcL* region since only a single nucletiotide variation (SNP) was detected from the *trnL-trnF* region. Additonally, the *trnL-trnF* region did not amplify or sequence well in a majority of the samples.

I identified a total of 5 different haplotypes within the *atpB-rbcL* region of the samples. The results of the linear regression demonstrated that the association between the genetic distance and spatial distance of these haplotypes was not significant ($P>0.005$), suggesting that individuals did

not demonstrate closer genetic characteristics to those closer to them than to those farther apart. The results of the Mantel Test (Figure 1) further disprove my hypothesis of a spatial autocorrelation between the individuals. Likewise, the haplotype map (Figure 2), showed no visible correlation.

## 5. Discussion

The spatial distributions of genetic diversity among population or haplotypes among individuals are influenced by numerous factors such as gene flow and geographic distances among individuals or populations. My results suggest that genetic relatedness among birch individuals was not defined by the spatial distance among them. The lack of relationship between spatial and genetic distances may be most likely due to the high wind-dispersal capacity of *Betula alleghaniensis* as it is both a wind-dispersed and a wind-pollinated species. At such fine-scale analysis, extensive gene flow both from pollen and seed dispersal might have occurred, resulting in the random distribution of cpDNA haplotypes throughout my study site. This also indicates potential random mating among individuals, along with a probability of inbreeding. Inbreeding prevents outcrossing and reduces dispersal capacities, providing more genetic diversity (Charlesworth and Barton 2003).

Although yellow birch reaches its northern range limit in northeastern Wisconsin, where I conducted my research, my findings indicate that the genetic processes commonly expected to play out under those conditions (i.e. high spatial genetic structure among populations due to limited gene flow at peripheral populations) were not present. Other studies, however, suggest otherwise. For instance, Petit et al. (1997) found strong fine-scale spatial genetic structures in two species of oaks (*Quercus robur* and *Quercus petraea*) in western France. This strong spatial cpDNA haplotype distribution were attributed to both rare long-distance dispersal events during the post-

glacial recolonization of oaks and subsequent lack of gene flow among these individuals and populations. Lack of gene flow and potential genetic drift in these populations have lead to some haplotypes becoming fixed in a single location.

The results of this research, however, may not be sufficient to discard a correlation between spatial and genetic distances among individuals. The reduced number of samples studied, due to limitations in time and resources, may not have given an accurate representation of the genetic diversity of the *Betula alleghaniensis* population inside the property. Likewise, the use of more chloroplast markers or other types of molecular markers could have allowed the discovery of more haplotypes, facilitating a more accurate representation of the genetic diversity between the individuals. A broader spatial scale than the one used in this study, in which the largest distance between individuals was around 4 km, could have provided a better representation, as well.

## 6. Acknowledgements

I would like to give a special acknowledgement to Candice Lumibao, my mentor during this research, who provided her full support and guidance during the whole process. I would also like to thank Elyssa Schwendy and Jose Carlos Wharton Soto for providing additional help. Special thanks goes to Jennifer and Gary Belovsky, from the University of Notre Dame's Environmental Research Center, for making all their undergraduate researchers feel welcome. I'd also like to give special thanks to the University of Notre Dame itself, for allowing undergraduate students from other universities to conduct research inside their properties, and to the Bernard J. Hank Family Endowment, for providing all the funds neccesary for this research.

## 7. Literature Cited

Barbujani, Guido. (1987) "Autocorrelation of gene frequencies under isolation by distance." Genetics 117.4: 777-782.

Bell, Graham, and Sinead Collins. 2008. "Adaptation, extinction and global change." *Evolutionary Applications* 1.1: 3-16.

Charlesworth, Brian, Deborah Charlesworth, and Nicholas H. Barton. 2003. "The effects of genetic and geographic structure on neutral variation." Annual Review of Ecology, Evolution, and Systematics: 99-125.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. "Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond." *Molecular ecology* 17.5: 1170-1188.

Escudero, Adrian, Jose M. Iriondo, and M. Elena Torres. 2003. "Spatial analysis of genetic diversity as a tool for plant conservation." *Biological Conservation* 113.3: 351-365.

ESRI. 2011. "ArcGIS Desktop: Release 10". Redlands, CA: Environmental Systems Research Institute.

Excoffier, L. and H.E. L. Lischer. 2010. "Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows". *Molecular Ecology Resources*. 10: 564-567.

Fairbanks, D. J.; Andersen, W. R.,. 1995. "Molecular analysis of genetic diversity: advantages and limitations". General Technical Report Intermountain Research Station, USDA Forest Service ( INT-GTR-315): 267-270

Gene Codes Corporation, Inc. 1995. “Sequencher 5.2.4: The complete software solution for sequencing DNA.”

Hamrick, J. L. 2004. "Response of forest trees to global environmental changes." *Forest ecology and management* 197.1: 323-335.

Hardy OJ, Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618-620.

Houle, Gilles. 1998 "Seed dispersal and seedling recruitment of Betula alleghaniensis: spatial inconsistency in time." Ecology 79.3: 807-818.

Jensen, J.L., Bohonak, A.J., and Kelley, S.T. 2005. Isolation by distance, web service. BMC Genetics 6: 13. v.3.23 http://ibdws.sdsu.edu/

Kimatu, Josphert N., Geoffrey M. Muluvi, and Bao Liu. 2012 "Genetic Diversity, Habitat Fragmentation and Epigenetic Variations."

Lavergne, Sébastien, et al. 2010 "Biodiversity and climate change: integrating evolutionary and ecological responses of species and communities." Annual Review of Ecology, Evolution, and Systematics 41: 321-350.

Pease, Craig M., Russell Lande, and J. J. Bull. 1989 "A model of population growth, dispersal and evolution in a changing environment." Ecology: 1657-1664.

Petit, Rémy J., et al. 1997 "Chloroplast DNA footprints of postglacial recolonization by oaks." Proceedings of the National Academy of Sciences 94.18: 9996-10001.

Poczai, Péter, et al. 2012 "Genomics Meets Biodiversity: Advances in Molecular Marker Development and Their Applications in Plant Genetic Diversity Assessment." The Molecular Basis of Plant Genetic Diversity.

## 8. Figures

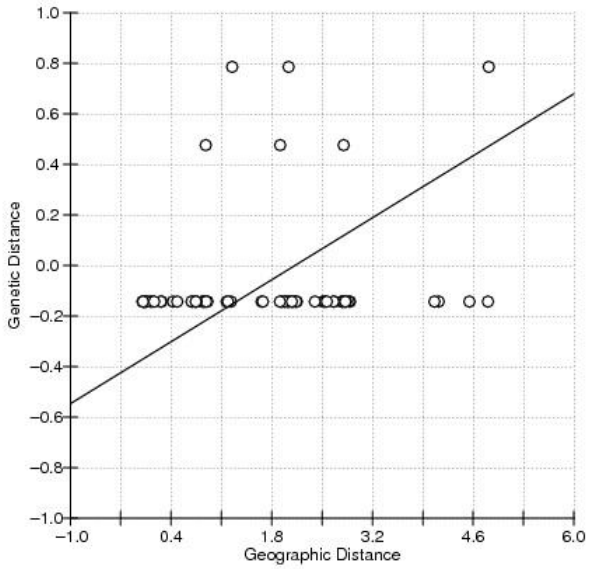


FIGURE 1. Results of a Mantel Test for matrix correlation between genetic distance and geographic distance. ran by *Isolation by Distance, Web Service (Jensen et al 2005)*

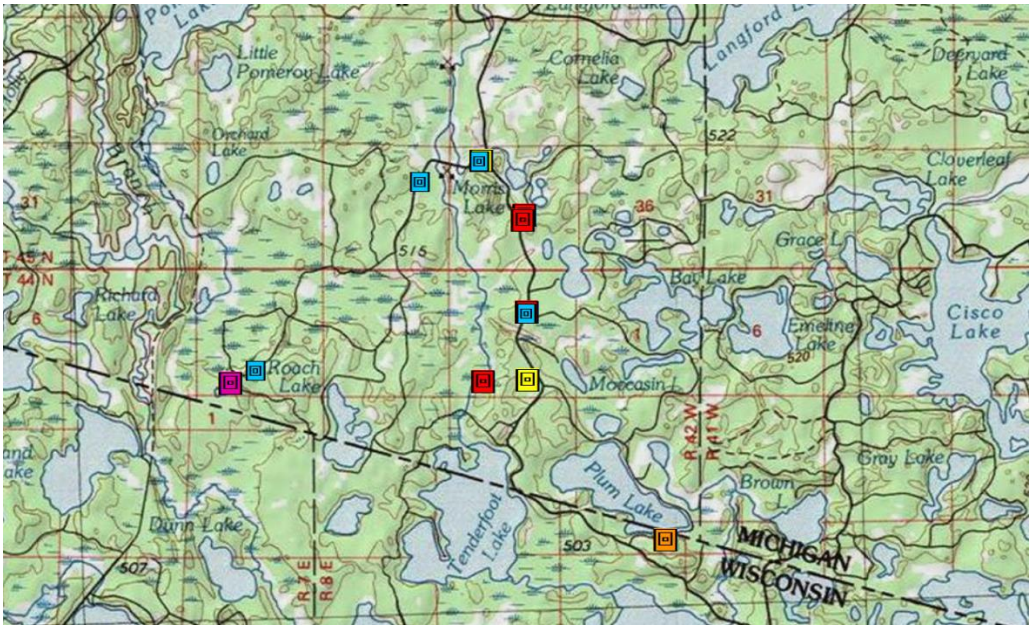


FIGURE 2. Mapping of haplotypes obtained through *ArcGis*.