# Phylogeography, introgression, and population structure of the eastern North American birches *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta*

Ashley M. Thomson

A Thesis

in the Department

of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy in Biology at

Concordia University

Montreal, Quebec, Canada

July, 2013

**CONCORDIA UNIVERSITY**
**SCHOOL OF GRADUATE STUDIES**

This is to certify that the thesis prepared

By: Ashley Marie Thomson

Entitled: Phylogeography, introgression, and population structure of the eastern North American birches *Betula alleghaniensis, B. papyrifera, and B. lenta*

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biology

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Uri Shalev — Chair

Dr. Jonathan Davies — External Examiner

Dr. Paul Joyce — External to Program

Dr. Dylan Fraser — Examiner

Dr. Daniel McLaughlin — Examiner

Dr. Selvadurai Dayanandan — Thesis Supervisor

Approved by

Chair of Department or Graduate Program Director

July 28, 2013

Dean of Faculty

## ABSTRACT

**Phylogeography, introgression, and population structure of the eastern North American birches *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta***

**Ashley Thomson**, Ph.D.
Concordia University, 2013

This study examines phylogeography, population genetics, and introgression in birches (*Betula* spp.) of eastern North America with a specific focus on *B. papyrifera* Marsh., *B. alleghaniensis* Britt., and *B. lenta* L. *Betula papyrifera* (2n=56, 70, 84) occupies a wide variety of habitats across the transcontinental boreal zone and into northeastern temperate forests. *Betula alleghaniensis* (2n=84) is a mesophytic species occurring from southeastern Canada and New England to the southern Appalachians. *Betula lenta* (2n=28) is endemic to the Appalachian region of the United States, where it occurs primarily on moist, cool sites. Phylogeographic analyses based on chloroplast DNA (cpDNA) markers were used for inferring locations of glacial refugia and patterns of post-glacial migration. Nuclear microsatellite markers were used for investigating allelic richness, gene diversity, and genetic structure of populations of each species. The haplotype sharing among species was examined to elucidate patterns of cpDNA introgression, and nuclear microsatellite allele sharing was investigated to determine the incidence of nuclear genomic introgression. The cpDNA analyses revealed phylogeographic structuring of eastern and western populations of *B. alleghaniensis* and *B. papyrifera* suggesting postglacial colonization from separate refugia. The geographically structured patterns of regional cpDNA haplotype sharing between *B. alleghaniensis* and *B. papyrifera* suggest widespread introgression between species during the postglacial recolonization. In contrast, cpDNA markers were monomorphic in *B. lenta* and no haplotype sharing between *B. lenta* and other species were detected. Nuclear microsatellite data revealed low levels of population genetic substructure in each of the species with significant differentiation at larger spatial scales due to isolation by distance. The low level of population differentiation over relatively long distances probably reflects the long distance pollen and seed dispersal of these species. The analysis of nuclear microsatellite allele-sharing revealed that despite moderate to high proportions of shared alleles, each of the species was significantly genetically differentiated. No strong evidence of genetic introgression was found for *B. alleghaniensis* and *B. lenta*, or *B. papyrifera* and *B. lenta*. Admixture proportions were higher between *B. alleghaniensis* and *B. papyrifera* in the western Great Lakes, which is a previously-documented zone of hybridization based on the occurrence of morphological intermediates.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor, Dr. Dayanandan, for his support and guidance during my studies. Thanks to him for his knowledge and expertise and for helping me to develop a great project. Thanks also to my collaborator Dr. Christopher Dick, who co-supervised me during the year that I spent as a visiting scholar in his lab at University of Michigan. His knowledge and new perspectives have been influential to my research. Thanks very much to my committee members Dr. Dylan Fraser and Dr. Daniel McLaughlin for their insight and helpful comments.

I would like to thank the many personnel associated with state parks, national forests, and experimental forests of the eastern United States who provided assistance in locating sites, obtaining sample collections, and providing directions and lodging for the various field sites visited as part of this study. I would like to extend a sincere thank you to Fred Huber of the George-Jefferson National Forest for assisting with *B. uber* field collections. Thanks to Dale Simpson and Bernard Daigle of the CNTSC, as well as Dave Kolotelo of the BCTSC for providing seed collections. Also, thanks to Greg Kowawleski and Paul Bloese of Michigan State University for assistance in obtaining yellow birch provenance samples. A special thanks to Dr. Burton Barnes of the University of Michigan, whose expertise regarding eastern North American birches was instrumental in this study. I would like to thank the various organizations which provided funding including the National Research Council of Canada, the US National Science Foundation, Concordia University, the Quebec Centre for Biodiversity Research, and the Quebec Fonds de Recherche Nature et Technologies.

Thanks to all of my long-term and visiting lab-mates and fellow graduate students at Concordia and University of Michigan. I enjoyed their company during long hours in the lab.

Sincerest thanks to my undergraduate intern, Ana Pascoini. Her dedication to the project was impressive, and I know she will have a bright future in research.

A special thanks to my Master's supervisor Dr. William Parker of Lakehead University. His passion, dedication, and excellence in research have inspired me to

pursue a career in research. Thanks to him for his continual mentorship, support, and friendship.

I would like to extend a very special thank you to my fiancé Sebastian Belmar-Lucero. He is the most patient and kind man in the world, and I am very lucky to have him. Thanks to my parents, who have provided continuous support throughout my studies. A warm thanks to my future family in Chile who have provided support and encouragement despite the long-distance. A very special acknowledgement to my future nephew Lucas and nieces Agustina and Trinidad for sharing lots of fun during my trip to Chile. I really enjoyed our day at the zoo!

TABLE OF CONTENTS

LIST OF TABLES

**Chapter 1: Comparative phylogeography of eastern North American birches (*Betula* spp.) based on cpDNA markers**



**Chapter 2: Population genetic structure of *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta* as revealed by nuclear microsatellites**

**Chapter 3: Introgression and genetic relationships of eastern North American birches *Betula alleghaniensis*, *B. papyrifera* and *B. lenta* as revealed by nuclear microsatellites**

LIST OF FIGURES

**Chapter 1: Comparative phylogeography of eastern North American birches (*Betula* spp.) based on cpDNA markers**



**Chapter 2: Population genetic structure of *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta* as revealed by nuclear microsatellites**

**Chapter 3: Introgression and genetic relationships of eastern North American birches *Betula alleghaniensis*, *B. papyrifera* and *B. lenta* as revealed by nuclear microsatellites**

LIST OF APPENDICES

# GENERAL INTRODUCTION

Species are dynamic entities, and the divergence and structure of populations is shaped by both historic and contemporary population demographics and gene flow (Schaal et al., 1998; Widmer & Lexer, 2001). In temperate regions, historic climatic fluctuations have contributed strongly to shaping the population genetic structure and levels of genetic diversity (Hewitt, 1996, 2000). Repeated expansions and contractions of glacial ice sheets during the Pleistocene were associated with drastic reductions followed by rapid expansions of many species' ranges (Stewart & Lister, 2001). Expansion of glacial ice reduced available habitat for most temperate trees, and in Europe fossil and DNA evidence indicate that the majority of species were confined to isolated southern refugia in Italy, the Balkans, or the Iberian Peninsula at the time of the last glacial maxim (LGM) (Bennett et al., 1991; Comes & Kadereit, 1998; Taberlet et al., 1998). The locations of refugia for North American tree species are generally more poorly understood due to an inability to distinguish fossil pollen types below the generic level and due to a lack of comparative phylogeographic studies (but see Soltis et al., 1997, 2006; Jaramillo-Correa et al., 2009). As a result of long-term genetic isolation within glacial refugia, many temperate trees show high levels of population genetic substructure associated with isolation and divergence of different glacial lineages (Petit et al., 2003a). The finding of very low levels of DNA diversity in several northern hemisphere trees has been attributed to past population bottlenecks which occurred during the last glacial maximum or during post-glacial expansion; for example *Pinus resinosa* (Boys et al., 2005), *Pinus pinea* (Vendramin et al., 2008), and *Fagus sylvatica* (Demesure et al., 1996). It is generally hypothesized that after the retreat of the glaciers trees should have

expanded rapidly into newly-opened habitats via long-distance seed dispersal (LeCorre et al., 1997; Hewitt, 2000) and numerous species, for example *Quercus garryana* (Marsico et al., 2009), and *Thuja plicata* (O'Connell et al., 2008), demonstrate northward declines in allelic richness as a result of successive founding events along the postglacial recolonization route.

Phylogeographic studies provide insight into the location of glacial refugia, as well as likely post-glacial recolonization routes and provide information on how forest trees have responded to past changes in climate (McLachlan et al., 2005). Phylogeographic studies provide evidence that species do not necessarily respond similarly to climatic change, and responses may be largely individualistic and depend on species' adaptations and climatic tolerances (Stewart et al., 2010). Differences in dispersal ability also may play a role (Aguinagalde et al., 2005). In comparison to single-species phylogeographic studies, comparative phylogeographic studies of closely-related species may elucidate clearly the role of historic or ecological factors in influencing phylogeographic patterns by controlling for traits related to dispersal and colonization ability (Saeki et al., 2011). For example, the higher cpDNA diversity and stronger phylogeographic structure of red maple (*Acer rubrum)* in comparison to sugar maple (*Acer saccharinum*) is attributed to its broader geographic range and ecological amplitude (Saeki et al., 2011).

To date, the majority of phylogeographic studies of temperate angiosperms have been investigated using cpDNA markers. The cpDNA genome possesses a number of characteristics which make it especially useful for reconstructing plant phylogeographic histories (Avise, 2009). Chloroplasts are uniparentally inherited, and therefore do not

demonstrate genetic recombination. This clonal inheritance pattern infers that each haplotype has just a single ancestor in the previous generation, and allows for accurate reconstruction of phylogenetic lineages. Non-recombination also infers that in many plant species mutation serves as the only source of variation in the chloroplast (Wade et al., 1994). The effective population size of the chloroplast DNA (cpDNA) molecule is one half of that of nuclear DNA in monoecious plants, and one quarter in dioecious plants (Birky et al., 1983). Thus, chloroplast genes are subject to relatively rapid fixation by genetic drift and decreased gene diversity at mutation-drift equilibrium in comparison to nuclear markers. The higher rate of genetic drift results in distinct population structuring at chloroplast markers for many species (McCauley, 1995). The majority of angiosperms demonstrate maternal inheritance of chloroplasts (Corriveau & Coleman, 1988), while most gymnosperms (e.g. *Pineaceae*) demonstrate paternal chloroplast inheritance (Neale & Sederoff, 1989; Avise, 2009). In angiosperms with maternal chloroplast inheritance, cpDNA markers thus provide information about post-glacial colonization via seed dispersal (Oddou-Muratorio et al., 2001). Because gene dispersal by seeds is in many cases more limited than gene flow by pollen, this can contribute to stronger genetic structuring at cpDNA than nuclear markers (Petit et al., 2005). These properties imply that chloroplast genes are particularly sensitive to founder effects (LeCorre et al., 1997). Simulation studies have shown that approximately 10000 generations would be required to reach equilibrium between gene flow and drift after a colonization phase (LeCorre et al., 1997). However, for most long-lived forest trees only a few hundred generations may have passed since the time of post-glacial recolonization, and thus observed patterns of

cpDNA structure for many species may reflect recent range expansions associated with post-glacial recolonization.

In contrast to cpDNA, the nuclear genome readily recombines and is biparentally inherited (Hare, 2001). In plants, nuclear DNA has a higher mutation rate than cpDNA and appears to evolve at twice the rate as cpDNA (Wolfe et al., 1987). Due to their larger effective population sizes, nuclear genes should demonstrate reduced rates of drift and fixation in comparison to chloroplast DNA (Birky et al., 1983). The biparental inheritance of nuclear DNA implies that it reflects patterns of both seed and pollen dispersal (Oddou-Muratorio et al., 2001; Marsico et al., 2009). Since many forest trees demonstrate much more efficient dispersal of pollen than seeds, high rates of pollen flow may help explain the finding of relatively low levels of population substructure at nuclear markers in comparison to cpDNA studies (Petit et al., 2005). Indeed, the finding of relatively low differentiation among populations of *Juglans cinerea* (Ross-Davis et al., 2008), *Juglans nigra* (Victory et al., 2006), and *Quercus macrocarpa* (Craft & Ashley, 2007) in eastern North America has been attributed to efficient long-distance pollen dispersal.

Since many plants demonstrate incomplete reproductive isolation from closely-related congeners, gene flow via introgression also may have an impact on the genetic structure of plant populations (Schaal et al., 1998). Introgression may serve to increase genetic variation within species and generate novel gene combinations (Rieseberg et al., 2003). If levels of introgression vary between populations, this can serve to increase population differentiation (Oddou-Muratorio et al., 2001). For example, the finding of a unique haplotype group at the northern range of *Acer saccharinum* was attributed to

introgression of haplotypes from sympatric *A. rubrum* (Saeki et al., 2011). Chloroplast DNA-based phylogeographic studies have revealed strong evidence of introgression in the form of geographically-structured haplotype sharing in many plant species groups; for example *Acer* (Saeki et al., 2011), *Betula* (Palme et al., 2004; Maliouchenko et al., 2007), *Fraxinus* (Heuertz et al., 2006), and *Quercus* (Dumolin-Lapègue et al., 1997). The apparent introgression of a chloroplast genome from one species to another is commonly referred to as 'chloroplast capture' (Stegemann et al., 2012). Chloroplast capture through hybridization and introgression appears to be a widespread occurrence throughout a diverse array of plant taxa (Rieseberg & Soltis, 1991; Rieseberg, 1995). However, for many species, introgression at cpDNA markers is not necessarily associated with introgression at nuclear loci; for example, geographical structuring of chloroplast haplotypes suggestive of introgression has been found in *Alnus* (King & Ferris, 2000) and *Quercus* (Whittemore & Schaal, 1991) though nuclear genes did not appear to be exchanged freely. In many cases, it appears that cpDNA introgression may occur more readily than nuclear DNA introgression (Rieseberg & Soltis, 1991; Rieseberg et al., 1991). In other forest trees, studies of introgression have revealed allele-sharing at nuclear loci as a result of contemporary gene flow between species; for example *Fraxinus* (Fernandez-Manjarres et al., 2006; Gerard et al., 2006), *Juglans* (Hoban et al., 2009), *Populus* (Keim et al., 1989), and *Salix* (Hardig et al., 2000). Depending on the rate and direction of gene flow, species may remain morphologically differentiated or may eventually collapse into a species complex. For example, European white oaks (*Quercus spp.*) preferentially cross within species, but occasional hybrids may occur (Lepais & Gerber, 2011). Backcrosses of hybrids appear to be bidirectional, allowing for low rates

of gene flow between species. However, high rates of intraspecific mating ensure that parental types are recovered within a few generations, thus maintaining the general morphological distinction of species. In comparison, European birches *B. nana* (2x) and *B. pubescens* (4x) appear to hybridize frequently where they overlap, with about 10% of plants representing triploid hybrids (Thórsson et al., 2001; Anamthawat-Jónsson & Thórsson, 2003). Introgression is bidirectional so that morphological variation is largely continuous between *B. nana*, *B. pubescens* and the triploid hybrid. In a study of herbaceous plants *Senecio hercynicus* and *S. ovatus*, introgression was so extensive in one German population that virtually no pure individuals of the parental species could be observed, and nuclear markers confirmed continuous molecular variation suggesting that the species had essentially collapsed into a hybrid swarm at that location (Oberprieler et al., 2010). In contrast, studies of introgression between *Populus fremontii* and *P. angustifolia* revealed introgression only in the direction of *P. fremontii* to *P. angustifolia* (Keim et al., 1989).

*Thesis outline and chapter overview*

This study examines the phylogeography, population genetics, and incidence of introgression among birches (*Betula* spp.) of eastern North America. Specifically, the thesis will focus on three of the most common and wide-ranging species *Betula alleghaniensis* Marshall, *B. papyrifera* Britton, and *B. lenta* L. I have selected this as the topic of my PhD research for a number of reasons; first, current understanding of phylogeographic patterns of North American temperate trees is limited, as most studies

thus far have focused on a single-species or trees of unglaciated regions or widespread boreal conifers. Thus, there is a need for comparative phylogeographic studies to better understand biogeographic histories of eastern North American temperate trees. Birches are an ideal group of plants for comparative phylogeographic study, as the species share similar traits related to reproduction and dispersal but differ in climatic tolerance and habitat preference (Furlow, 1990). Thus, a comparative phylogeographic study of eastern North American birches will permit inference of shared aspects of biogeographic history among species while addressing how differences in phylogeographic structure among species may have been shaped by their differing ecological tolerances. Second, birches are widespread, ecologically and economically important species in temperate forests of northeastern North America, where they form a large component of the forest biomass in many areas. However, to the best of my knowledge, no studies examining genetic diversity or patterns of population structure at DNA markers currently exist for eastern North American birches, and there is a need to better understand patterns of genetic variation and population structure. This study will provide information which may inform future genetic management and conservation efforts, which will become increasingly important if climate change begins to threaten these species. Finally, no study of the population genetics of *Betula* could be complete without an understanding of how introgression among species may contribute to observed patterns of population structure and genetic diversity. Hybridization is frequent within the genus *Betula* (DeJong, 1993), and widespread chloroplast haplotype sharing and nuclear introgression have been documented for European birches (Thórsson et al., 2001; Anamthawat-Jónsson & Thórsson, 2003; Palme et al., 2004; Maliouchenko et al., 2007). Since introgression may

increase genetic diversity or population structure non-uniformly across a species' range, it is critical to understand the potential role of introgression in shaping patterns of genetic variation within eastern North America *Betula* species. Also, understanding the rate at which species share alleles due to introgression will allow for inference as to what extent the North American birches can be considered as genetically distinct, or whether they may be more appropriately considered as a species complex with continuous molecular variation.

**Chapter 1: Comparative phylogeography of eastern North American birches (*Betula* spp.) based on cpDNA markers.**

The first chapter will examine the phylogeography of six species of eastern North American *Betula* species using cpDNA markers with a focus on *B. alleghaniensis*, *B. papyrifera*, and *B. lenta*. As previously mentioned, current understanding of phylogeographic patterns of eastern North American temperate trees is currently limited. By comparing patterns of phylogeographic structure and haplotype sharing at cpDNA markers, this study will allow inference of shared aspects of the biogeographic history of these species which could potentially include shared glacial refugia and historic introgression. However, because these species demonstrate different habitat and ecological preferences, any differences in phylogeographic patterns between them could potentially reflect unique aspects of biogeographic history attributable to their different ecological tolerances. Finally, examination of the cpDNA structure and haplotype distribution will permit inference of the location of glacial refuges as well as individual post-glacial migration routes for each species.

**Chapter 2: Population genetic structure of *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta* as revealed by nuclear microsatellites.**

The second chapter examines patterns of population genetic structure and genetic diversity of eastern North American birches at nuclear microsatellite markers. In comparison to cpDNA markers, nuclear markers are biparentally inherited and recombinant, and thus patterns of genetic structure at nuclear markers may differ substantially from patterns at cpDNA markers, as revealed in Chapter 1. Also, patterns revealed by nuclear microsatellites may differ due to relative rates of seed and pollen flow, and differing rates of genetic drift between nuclear and chloroplast markers. Maps of allelic richness and gene diversity will be used to identify areas of relatively high genetic diversity for each species. Bayesian clustering methods will be used to identify patterns of population substructure, including the number and location of population groups within each species. The results will be interpreted with respect to how the observed patterns may have been shaped by both historic and contemporary population dynamics and gene flow. Finally, the observed patterns will be used to suggest populations of potential importance to future management and conservation efforts.

**Chapter 3: Introgression and genetic relationships of eastern North American birches *Betula alleghaniensis*, *B. papyrifera* and *B. lenta* as revealed by nuclear microsatellites.**

The third chapter examines the species genetic structure and incidence of introgression among *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* as assessed by allele-sharing at nuclear microsatellite loci. The species are largely morphologically distinct but occasional morphological intermediates have been reported and are thought to represent putative hybrids (Sharik & Barnes, 1971; Clausen, 1973, 1977; Barnes et al., 1974). However, the incidence of natural hybridization and introgression between eastern North American birches has never been investigated using nuclear DNA markers. Thus, it is

currently unclear to what extent species may share nuclear alleles as a result of genetic introgression, or whether they remain largely genetically distinct. This chapter will investigate the incidence of allele sharing and partitioning of nuclear microsatellite variation to assess to what extent *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* can be considered as genetically distinct. Further, principal components analysis and Bayesian analysis will be used to identify patterns of genetic clustering among individuals of different species. Bayesian analysis also will allow for the identification of any putatively admixed individuals. By analyzing patterns of isolation by distance and the geographic location of putatively admixed individuals, it should be possible to distinguish whether allele-sharing results from introgression or shared ancestry. The observed patterns will be interpreted in light of current understanding of the genetic relationships between the species, and documented incidences of putative natural hybridization.

## References


Aguinagalde I., Hampe A., Mohanty A., Martín J.P., Duminil J., & Petit R.J. (2005) Effects of life-history traits and species distribution on genetic structure at maternally inherited markers in European trees and shrubs. *Journal of Biogeography*, **32**, 329–339.

Anamthawat-Jónsson K. & Thórsson A.T. (2003) Natural hybridisation in birch: triploid hybrids between *Betula nana* and *B. pubescens*. *Plant Cell, Tissue, and Organ Culture*, **75**, 99–107.

Avise J.C. (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Barnes B.V., Dancik B.P., & Sharik T.L. (1974) Natural hybridization of yellow birch and paper birch. *Forest Science*, **20**, 215–221.

Bennett K.D., Tzedakis P.C., & Willis K.J. (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103–115.

Birky C.W., Maruyama T., & Fuerst P. (1983) An approach to population and evolutionary genetic theory for genes in mitochondrial and chloroplasts, and some results. *Genetics*, **103**, 513–527.

Boys J., Cherry M., & Dayanandan S. (2005) Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany*, **92**, 833–841.

Clausen K.E. (1973) *Genetics of Yellow Birch. USDA Forest Service Research Paper WO-18.* U.S. Department of Agriculture, Washington, DC.

Clausen K.E. (1977) The yellow x paper birch hybrid - a potential substitute for yellow birch on problem sites. *Proceedings of the Thirteenth Lake States Forest Tree Improvement Conference; Gen. Tech. Rep. NC-50.* pp. 166–171. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station., St. Paul, MN.

Comes P. & Kadereit J.W. (1998) The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**, 432–438.

Corriveau J.L. & Coleman A.W. (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443–1458.

Craft K.J. & Ashley M.V. (2007) Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. *Forest Ecology and Management*, **239**, 13–20.

DeJong P.C. (1993) An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. *Betula: Proceedings of the IDS Betula Symposium* (ed. by D. Hunt), pp. 7–18. Richmond, England.

Demesure B., Comps B., & Petit R.J. (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.

Dumolin-Lapègue S., Demesure B., Fineschi S., Le Corre V., & Petit R.J. (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.

Fernandez-Manjarres J.F., Gerard P.R., Dufour J., Raquin C., & Frascaria-Lacoste N. (2006) Differential patterns of morphological and molecular hybridization between *Fraxinus excelsior* L. and *Fraxinus angustifolia* Vahl (*Oleaceae*) in eastern and western France. *Molecular Ecology*, **15**, 3245–3257.

Furlow J.J. (1990) The genera of Betulaceae in the southeastern United States. *Journal of the Arnold Arboretum*, **71**, 1–67.

Gerard P.R., Fernandez-Manjarres J.F., & Frascaria-Lacoste N. (2006) Temporal cline in a hybrid zone population between *Fraxinus excelsior* L. and *Fraxinus angustifolia* Vahl. *Molecular Ecology*, **15**, 3655–3667.

Hardig T.M., Brunsfeld S.J., Fritz R.S., Morgan M., & Orians C.M. (2000) Morphological and molecular evidence for hybridization and introgression in a willow (*Salix*) hybrid zone. *Molecular Ecology*, **9**, 9–24.

Hare M.P. (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution*, **16**, 700–706.

Heuertz M., Carnevale S., Fineschi S., Sebastiani F., Hausman J.F., Paule L., & Vendramin G.G. (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (*Oleaceae*): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–2140.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt G.M. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hoban S.M., McCleary T.S., Schlarbaum S.E., & Romero-Severson J. (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. *Biology Letters*, **5**, 324–327.

Jaramillo-Correa J.P., Beaulieu J., Khasa D.P., & Bousquet J. (2009) Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research*, **39**, 286–307.

Keim P., Paige N., Whitham G., & Lark K.G. (1989) Genetic analysis of an interspecific hybrid swarm of *Populus*: occurrence of unidirectional introgression. *Genetics*, **123**, 557–565.

King R.A. & Ferris C. (2000) Chloroplast DNA and nuclear DNA variation in the sympatric alder species, *Alnus cordata* (Lois.) Duby and *A. glutinosa* (L.) Gaertn. *Biological Journal of the Linnean Society*, **70**, 147–160.

LeCorre V., Machon N., Petit R.J., & Kremer A. (1997) Colonization with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetical Research, Cambridge*, **69**, 117–125.

Lepais O. & Gerber S. (2011) Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. *Evolution*, **65**, 156–170.

Maliouchenko O., Palmé A.E., Buonamici A., Vendramin G.G., & Lascoux M. (2007) Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography*, **34**, 1601–1610.

Marsico T.D., Hellmann J.J., & Romero-Severson J. (2009) Patterns of seed dispersal and pollen flow in *Quercus garryana* (Fagaceae) following post-glacial climatic changes. *Journal of Biogeography*, **36**, 929–941.

McCauley D.E. (1995) The use of chloroplast DNA polymorphism in studies of gene flow in plants. *TREE*, **I0**, 198–202.

McLachlan J.S., Clark J.S., & Manos P.S. (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.

Neale D.B. & Sederoff R.R. (1989) Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics*, **77**, 212–216.

O'Connell L.M., Ritland K., & Thompson S.L. (2008) Patterns of post-glacial colonization by western redcedar (*Thuja plicata*, Cupressaceae) as revealed by microsatellite markers. *Botany*, **86**, 194–203.

Oberprieler C., Barth A., Schwarz S., & Heilmann J. (2010) Morphological and phytochemical variation, genetic structure and phenology in an introgressive hybrid

swarm of *Senecio hercynicus* and *S. ovatus* (Compositae, Senecioneae). *Plant Systematics and Evolution*, **286**, 153–166.

Oddou-Muratorio S., Petit R.J., Le Guerroue B., Guesnet D., & Demesure B. (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123–1135.

Palme A.E., Su Q., Palsson S., & Lascoux M. (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology*, **13**, 167–178.

Petit R.J., Aguinagalde I., De Beaulieu J.-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Muller-Starck G., Demesure-Musch B., Palme A., Martin J.P., Rendell S., & Vendramin G.G. (2003) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **1563**, 1563–1565.

Petit R.J., Duminil J., Fineschi S., Hampe A., Salvini D., & Vendramin G.G. (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Rieseberg L.H. (1995) The role of hybridization in evolution: old wine in new skins. *American Journal of Botany*, **82**, 944–953.

Rieseberg L.H., Choi H.C., & Ham D. (1991) Differential cytoplasmic versus nuclear introgression in *Helianthus*. *The Journal of Heredity*, **82**, 489–493.

Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., & Lexer C. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, **301**, 1211–1216.

Rieseberg L.H. & Soltis D.E. (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.

Ross-Davis A., Ostry M., & Woeste K.E. (2008) Genetic diversity of butternut (*Juglans cinerea*) and implications for conservation. *Canadian Journal of Forest Research*, **38**, 899–907.

Saeki I., Dick C.W., Barnes B.V., & Murakami N. (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): impacts of habitat specialization, hybridization and glacial history. *Journal of Biogeography*, **38**, 992–1005.

Schaal B.A., Hayworth D.A., Olsen K.M., Rauscher J.T., & Smith W.A. (1998) Phylogeographic studies in plants: problems and prospects. *Molecular Ecology*, **7**, 465–474.

Sharik T.L. & Barnes B.V. (1971) Hybridization in *Betula alleghaniensis* Britt. and *B. lenta* L.: A comparative analysis of controlled crosses. *Forest Science*, **17**, 415–424.

Soltis D.E., Gitzendanner M.A., Strenge D.D., & Soltis P.S. (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353–373.

Soltis D.E., Morris A.B., McLachlan J.S., Manos P.S., & Soltis P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stegemann S., Keuthe M., Greiner S., & Bock R. (2012) Horizontal transfer of chloroplast genomes between plant species. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 2434–2438.

Stewart J.R. & Lister A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stewart J.R., Lister A.M., Barnes I., & Dalén L. (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B*, **277**, 661–671.

Taberlet P., Fumgalli L., Wust-Saucy A.-G., & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Thórsson A.T., Salmela E., & Anamthawat-Jónsson K. (2001) Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. *The Journal of Heredity*, **92**, 404–408.

Vendramin G.G., Fady B., González-Martínez S.C., Hu F.S., Scotti I., Sebastiani F., Soto A., & Petit R.J. (2008) Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution*, **62**, 680–688.

Victory E.R., Glaubitz J.C., Rhodes O.E., & Woeste K.E. (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany*, **93**, 118–126.

Wade M.J., McKnight M.L., & Shaffer H.B. (1994) The effects of kin-structured colonization on nuclear and cytoplasmic genetic diversity. *Evolution*, **48**, 1114–1120.

Whittemore A.T. & Schaal B.A. (1991) Interspecific gene flow in sympatric oaks. *Evolution*, **88**, 2540–2544.

Widmer A. & Lexer C. (2001) Glacial refugia: sanctuaries for allelic richness, but not for gene diversity. *Trends in Ecology & Evolution*, **16**, 267–269.

Wolfe K.H., Li W.H., & Sharp P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054–9058.

**Chapter 1: Comparative phylogeography of eastern North American birches (*Betula* spp.) based on cpDNA markers.**

## Abstract

Comparative phylogeographic analysis was used to infer impacts of Quaternary climate change, species differences, and hybridization on patterns of phylogeographic structure of eastern North American birches with emphasis on *Betula papyrifera*, *B. alleghaniensis* and *B. lenta*. Chloroplast microsatellites (cpSSRs) and the *psb*A-*trn*H intergenic spacer were analysed for *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* sampled from 65, 80, and 12 populations, respectively. Co-occurring *B. uber*, *B. populifolia* and *B. cordifolia* were also sampled to examine haplotype relationships and account for potential introgression. Haplotype networks, spatial analysis of molecular variance (SAMOVA), and comparisons of $N_{ST}$ and $G_{ST}$ values were used to evaluate phylogeographic structure. Genetic diversity within and among species was compared using rarefaction analysis and haplotype richness maps. The most widespread species, *B. papyrifera*, demonstrated higher haplotype diversity and weaker population genetic subdivision than *B. alleghaniensis*, while the Appalachian endemic *B. lenta* was fixed for a single haplotype. The SAMOVA revealed two main phylogeographic groups for *B. papyrifera* and *B. alleghaniensis*, corresponding to populations of eastern and western origin. However, these two species showed extensive regional haplotype sharing indicating that introgression contributed to the shared phylogeographic pattern. The greatest haplotype diversity for the two species occurred in the Great Lakes Region, which is a postulated post-glacial biogeographic contact zone. Although similar phylogeographic patterns are often interpreted as evidence for common biogeographic histories, this study demonstrates that shared phylogeographic structure can arise through introgression, even among morphologically distinct species. This study provides further evidence that novel haplotype groups in post-glaciated areas can arise from introgression among closely related species, as an alternative to, or in conjunction with, cryptic northern refugia.

**Key Words** *Betula,* birches, comparative phylogeography, cpDNA, eastern North America, haplotype-sharing, introgression.

## Introduction

The genetic structure and diversity of temperate plant species were profoundly influenced by climatic fluctuations associated with Quaternary glacial-interglacial cycles (Hewitt, 2000). The prevalent view is that unfavourable climatic conditions and habitat loss due to advancing ice sheets, particularly during the last glacial maximum (LGM), truncated geographical ranges and caused many temperate tree species to retreat to refugia far south of the ice sheet margin (Stewart et al., 2010). This view is supported by studies in Europe, which indicate that temperate trees were mostly confined to southern regions corresponding to present day Italy, the Balkans, and the Iberian Peninsula (Bennett et al., 1991; Taberlet et al., 1998). As a result of long term genetic isolation, refugial populations are expected to be divergent from one another (Petit et al., 2003a) and display a northward decline in genetic diversity associated with founder events during recolonization (Hewitt, 2000). However, a meta-analysis of 22 European angiosperms found that haplotype diversity was often highest at intermediate latitudes, presumably owing to mixing of lineages in refugial contact zones (Petit et al., 2003a). Also, there is increasing evidence that cold-tolerant plants may have persisted at higher latitudes than previously expected (Willis et al., 2000; Stewart & Lister, 2001) as evidenced by endemic haplotypes close to the northern edge of species ranges (McLachlan et al., 2005).

In contrast to phylogeographic studies of angiosperms in Europe, current understanding of phylogeographic patterns of North American cold-temperate angiosperms is relatively poor. A majority of studies have focused on boreal conifers and trees of unglaciated regions (Soltis et al., 2006; Jaramillo-Correa et al., 2009), or have

included just a single species (but see Saeki et al. 2011). Thus, there is a need for comparative studies to better understand glacial history of cold-temperate angiosperms of eastern North-America. In particular, by examining closely related species, one can better distinguish shared biogeographic histories from idiosyncratic patterns attributable to ecological trait differences. A study of European ashes, for example, revealed higher levels of divergence among populations of *Fraxinus angustifolia* and *F. ornus* as compared to *F. excelsior* (Heuertz et al., 2006). This pattern was attributed to the relatively low cold tolerance of *F. angustifolia* and *F. ornus*, which would be associated with low LGM population sizes and high levels of genetic drift (Heuertz et al., 2006). Furthermore, studies of congeneric species often reveal introgression in the form of localized haplotype sharing, as has been documented in European birches (Palme et al., 2004; Maliouchenko et al., 2007), oaks (Petit et al., 1993; Dumolin-Lapègue et al., 1997; Belahbib et al., 2001), ashes (Heuertz et al., 2006) and North American maples (Saeki et al., 2011). Such introgression can potentially alter haplotype distributions and should be accounted for when drawing inferences of species' biogeographic histories; for example, the finding of a distinct phylogeographic group for *Acer rubrum* in the lower Mississippi River Valley is best explained by introgression of haplotypes from *A. saccharinum,* rather than by historical biogeographic contingencies (Saeki et al., 2011).

North American birches (*Betula*) are an ideal group of plants to study postglacial biogeography and to test for potential effects of introgression. *Betula* is composed of approximately 17 species in eastern North America, all of which share similar traits related to reproduction and dispersal but differ in climatic tolerance and habitat preference (Furlow, 1990). This study examines the cpDNA phylogeography and glacial

history of six eastern North American *Betula* species with a focus on three common species with contrasting distribution patterns: *Betula papyrifera* Marshall (paper birch), *B. alleghaniensis* Britton (yellow birch), and *B. lenta* L. (sweet birch). In addition, we sampled the codistributed species *B. populifolia* Marsh. (gray birch), *B. cordifolia* Regel (heartleaf birch), and *B. uber* (Ashe) Fernald (Virginia roundleaf birch) in order to obtain complete community-level sampling of *Betula*, which was needed to investigate the potential effects of introgression. Our main objectives were: (i) to infer locations of glacial refugia and postglacial migration routes based on phylogeographic structure (ii) to determine if differences in phylogeographic patterns could be attributed to habitat differences among these species, and (iii) to test for evidence of introgression as inferred by geographic structuring of shared haplotypes.

## Methods

### *Study Species*

All species of the genus *Betula* are monoecious with light, wind-dispersed seeds and pollen (Furlow, 1990). *Betula papyrifera* Marsh. is the most widely distributed and cold-tolerant of the North American birches, with a distribution that ranges across the boreal forest region of the United States and Canada to the northern limit of tree growth (Safford et al., 1990). In comparison, *Betula alleghaniensis* Britt. has a more southern distribution, occupying temperate forests of southeastern Canada through New England to the southern Appalachians (Erdmann, 1990). *Betula lenta* L. has the most restricted distribution of the three species and is endemic to the Appalachian region of the eastern United States; it occurs from southern Maine through to northern Alabama and Georgia

(Lamson, 1990). A comparative summary of select ecological and morphological characteristics of the six *Betula* species examined in this study is presented in Table 1. The study species fall into different subgenera; *Betulenta*, which contains species with generally dark, non-exfoliating bark (*B. alleghaniensis*, *B. lenta*, and *B. uber*) and *Betula*, comprised of white birches including *Betula papyrifera*, *B. populifolia*, and *B. cordifolia* (DeJong, 1993). This morphological classification is supported by molecular studies that separate the white and dark barked birches into different clades (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008). Two studies place *B. lenta* and *B. alleghaniensis* as sister-species (Jarvinen et al., 2004; Schenk et al., 2008) while another study places *B. uber* as sister to *B. lenta* (Li et al., 2005). Phylogenetic relationships of species within the subgenus *Betula* are poorly resolved (Jarvinen et al., 2004; Li et al., 2005), perhaps because hybridization is frequent within the genus *Betula.* Natural hybridization between *B. papyrifera* and B. *alleghaniensis* has been confirmed based on morphological traits of specimens from Minnesota, Wisconsin, and New Hampshire (Barnes et al., 1974), and their reproductive compatibility also has been shown by experimental crosses (Clausen, 1966; Alam & Grant, 1972). The natural hybrid of *B. cordifolia* and *B. populifolia* is *B. caerulea* (Grant & Thompson, 1975). Experimental crosses also have revealed compatibility of *B. papyrifera* with *B. populifolia, B. papyrifera* with *B. lenta,* and *B. lenta* with *B. alleghaniensis* (Clausen, 1966; Alam & Grant, 1972; Barnes et al., 1974).

Table 1. Comparison of important ecological and life history traits of six eastern North American birches examined in this study.

| | *B. papyrifera* | *B. alleghaniensis* | *B. lenta* | *B. populifolia* | *B. cordifolia* | *B. uber* | Reference |
|---|---|---|---|---|---|---|---|
| Habitat characteristics | | | | | | | |
| (i) Geographic distribution | Transcontinetal boreal zone of Canada and the northern United States | Southern Ontario to the Maritimes and New England through southern Appalachians | New England to the southern Appalachians | Eastern Ontario to New England and the Maritimes | Ontario to the Maritimes and New England, North Carolina, and Pennsylvania | Single known location in Smyth county, southwest Virginia | Furlow (1990, 1997) |
| (ii) Habitat | Virtually all site and soil types | Rich-mesic streambanks and forested slopes | Rich-mesic, cool forests and protected slopes | Dry to mesic open forests, abandoned fields, and waste places | Rich, open forests or moist, rocky slopes | Rich-mesic streambanks and floodplains | Furlow (1990,1997) |
| Morphology: | | | | | | | |
| (i) Bark | White, creamish, pinkish, or dark brown; loosely exfoliating | Yellowish to caramel; generally exfoliating but sometimes dark and close | Light brown and smooth; becoming dark and plated with age | Grayish-white; smooth, close | Reddish-white to reddish-tan or bronze; loosely exfoliating | Dark brown; smooth, close | Grant and Thompson (1975), Sharik and Ford (1984), Furlow (1990, 1997) |
| (ii) Leaves | Blade ovate; base rounded or cuneate; apex acute to acuminate | Blade ovate or oblong-ovate; base rounded to cuneate or cordate; apex acuminate | Blade ovate or oblong-ovate; base rounded to cordate; apex acuminate | Blade deltate or rhombic; base truncate to cuneate; apex long-acuminate | Blade ovate; base generally cordate; apex short-acuminate | Blade orbiculate; base rounded to cordate or truncate; apex rounded | Grant and Thompson (1975), Furlow (1990,1997) |
| Reproductive characteristics: | | | | | | | |
| (i) Chromosome number (x=14) | 5n=70 or 6n=84; rarely 4n=56 | 6n=84 | 2n=28 | 2n=28 | 2n=28; sometimes 3n=42 or 4n=56 | 2n=28 | Woodworth (1930), Grant and Thompson (1975), Sharik (1990) |

*Population Sampling*

Leaf samples for this study were obtained from throughout eastern Canada and the United States. Samples were obtained for 1-10 individuals per species when present in each of 36 natural populations. Leaf collections were made from individuals separated by at least 50 metres and dried in silica gel. We obtained additional samples representing various geographical regions from seed collections of the Canadian National Tree Seed Centre (CTNSC). In total, these collections contained seed from 57 localities with *B. papyrifera* from 24 sites, *B. alleghaniensis* from 17 sites, *B. populifolia* from 10 sites, and *B. cordifolia* from 10 sites. To increase the geographic coverage of the transcontinental-distributed *B. papyrifera*, we obtained seed collections from an additional 18 localities representing the western distributional range from the British Columbia Tree Seed Centre (BCTSC). Additional population samples of *B. alleghaniensis* were collected from a provenance trial located at the Kellogg Experimental Forest of Michigan State University near Kellogg, Michigan. During the winter of 2009, we collected vegetative buds from a single individual of each of 36 provenances at that trial. In total, the samples used for this study consist of 261 individuals of *B. papyrifera* at 65 sites, 314 individuals of *B. alleghaniensis* at 80 sites, 120 individuals of *B. lenta* at 12 sites, 54 individuals of *B. populifolia* at 16 sites, 16 individuals of *B. cordifolia* at 10 sites, and 10 individuals of *B. uber* at its single known locality Appendix 1.

*Laboratory Procedures*

DNA was extracted from mature leaf tissues using the modified CTAB protocol of Zeng et al. (2002). Qiagen DNeasy extraction kits (Qiagen Inc., Valencia, CA) were

used to extract DNA from vegetative buds and germinated seeds. Molecular variation was assessed using six chloroplast simple-sequence-repeat (cpSSR) markers. Three cpSSR markers were developed for *Betula* in this study with primers based on consensus sequences flanking variable microsatellite regions in *B. papyrifera*, *B. alleghaniensis*, and *B. lenta.* The remaining three cpSSR markers were obtained from the literature (Weising & Gardner, 1999). For all cpSSR markers, amplification was conducted using a tailed-primer procedure (Schuelke, 2000) in a total volume of 10 µl with 0.25 mM dNTPs, 2.5 mM $MgCl_2$, 0.1 mM tailed forward primer, 0.5 mM reverse primer, 0.1 mM of fluorescent-labeled primer, 1X buffer, 0.5 U Taq polymerase, and 2 µl of template DNA. PCR conditions for the cpSSR primers used in this study are provided in Table **2**. Amplified fragments were electrophoretically separated on the Licor 4200 DNA analyzer along with the manufacturer`s 50-350 bp size standard (LiCOR Biosciences, Lincoln, NE). Fragment lengths were scored using SAGA GT 2.1 software (Licor Biosciences) and verified manually.

Table 2. Primer information and thermocycling procedures for cpSSR markers.

| Primer | Target region | Repeat type in birch | Primer Sequence | Primer Source |
|---|---|---|---|---|
| ccmp4[1] | atpF intron | $(T)_{11}$ | 5'-AATGCTGAATCGAYGACCTA-3' | Weising and Gardner (1999) |
| | | | 5'-CCAAAATATTBGGAGGACTCT-3' | |
| ccmp5[1] | 3' to rps2 | $(T)_{14}$ | 5'-TGTTCCAATATCTTCTTGTCATTT-3' | Weising and Gardner (1999) |
| | | | 5'-AGGTTCCATCGGAACAATTAT-3' | |
| ccmp7[1] | atpB–rbcL intergenic | $(A)_7$ | 5'-CAACATATACCACTGTCAAG-3' | Weising and Gardner (1999) |
| | | | 5'-ACATCATTATTGTATACTCTTTC-3' | |
| BCMS1[2] | trnL3' exon–trnF | $(T)_{10}$ | 5'-GCTCTTTTCGTTAGCGGTTT-3' | This study |
| | | | 5'-ATTTGAAGCGGGGATACCTT-3' | |
| BCMS2[3] | trnL3' exon–trnF | $(T)_{11}$ | 5'-CCGCTTCAAATTTTAATGAT-3' | This study |
| | | | 5'-GATGACTTGGGTTTATGTCAA-3' | |
| BCMS3[2] | psbC-trnS intergenic | $(A)_8$ | 5'-CGGGCAAAACCAACAAAAT-3' | This study |
| | | | 5'-GGGTTCGAATCCCTCTCTCT-3' | |

[1] Thermocycling conditions consisted of initial denaturation at 94°C for 4:00 min, 35 cycles of 94°C for 1:00 min, 45°C for 1:00 min, 65°C for 1:00 min and final extension of 65°C for 10:00 min

[2] Initial denaturation of 94°C for 4:00 min, 35 cycles of 94°C for 45 s, 55°C for 45 s, 72°C for 45 s, and final extension of 72°C for 10:00 min

[3] Initial denaturation of 94°C for 4:00 min, 35 cycles of 94°C for 45 s, 52°C for 45 s, 72°C for 45 s, and final extension of 72°C for 10:00 min

We also obtained cpDNA sequence information from a subset of samples to compare phylogeographic patterns to those based on cpSSRs. The *psb*A-*trn*H intergenic spacer region was chosen based on its polymorphism and consistent PCR amplification. Amplification was conducted using primers of Sang et al. (1997). The PCR reaction mixture consisted of 0.25 mM dNTPs, 2 mM $MgCl_2$, 0.4 µM of each primer, 1X buffer, 1 U of Taq polymerase and 2 µL of template DNA in a total volume of 25 µL. Thermocycling consisted of 94°C for 4:00 min, 35 cycles of 94°C for 1:00 min, 55°C for 1:00 min, 72°C for 2:00 min, and a final elongation of 72°C for 10:00 min. PCR products were sequenced using an Applied Biosystems 3730xl DNA Analyzer. Sequence chromatograms were edited and aligned using GENEIOUS 4.8 software (Biomatters Ltd.,

Auckland, New Zealand). Indels were coded as single-characters. All forms of sequence polymorphisms including SNPs, indels, and cpSSRs were used in haplotype definitions (henceforth referred to as *psb*A-*trn*H or cpDNA haplotypes). We used a reduced within-population sampling intensity for sequencing compared to cpSSR genotyping since initial range-wide screening revealed that haplotypes were generally fixed across broad spatial scales. Individuals from all singleton populations were sequenced, while for multi-individual populations an average of three individuals per population was sequenced.

*Data analyses*

Haplotype frequencies were calculated using GENALEX 6.41 (Peakall & Smouse, 2006) and haplotype distribution maps were constructed using ARCGIS 10.0. The matrix of genetic distances among cpSSR haplotypes was based on the number of single-base-pair repeat differences summed across each of the loci (Heuertz et al. 2004). Calculation of the minimum spanning network for cpSSRs was performed using MINSPNET (Excoffier & Smouse, 1994) using the matrix of genetic distances. For *psb*A-*trn*H sequences, we constructed a statistical parsimony network using TCS 1.21 (Clement et al., 2000). Gaps were treated as $5^{th}$ state and the connection limit was set at five steps.

Analysis of molecular variance (AMOVA) was used for assessing the partitioning of molecular variance as implemented in ARLEQUIN 3.1 (Excoffier et al., 2005), treating each species individually as well as in species-groupings. To identify geographic patterns of genetic discontinuities, Spatial Analysis of Molecular Variance (SAMOVA) (Dupanloup et al., 2002) was performed with 10000 iterations and repeated with

increasing numbers of ($K$) groups until the $F_{CT}$ value reached a plateau. We also tested for the presence of a phylogeographic structure by comparing permutated values of $N_{ST}$ and $G_{ST}$ using the program PERMUTCPSSR 2.0 (Pons & Petit, 1996). Haplotype diversity among species was compared after standardization with rarefaction using the program CONTRIB 1.02 (Petit et al., 1998). The same program was used to estimate haplotype richness values for each population after rarefaction. Haplotype richness values for each population were mapped and interpolated using ARCGIS 10.0 to visualize geographic patterns of genetic diversity.

## Results

### *Haplotype distributions*

The six cpSSR markers revealed 21 haplotypes for the six eastern *Betula* species (Figure 1). In *B. papryrifera*, most of the eastern range from central Ontario through New England and the Maritimes was dominated by Hssr16 and related haplotypes Hssr13, Hssr17, and Hssr20 (Figure 1a). A second group composed of Hssr10 and Hssr11 was found along the Atlantic coast of Canada. *Betula papyrifera* populations in western Ontario and the Lakes States contained a mixture of endemic Hssr1 and Hssr7 with a low frequency of eastern haplotypes (Hssr16 and related). Haplotypes Hssr4, Hssr5, Hssr9 were common in southern British Columbia, while Hssr3 was common to northern British Columbia. The remaining haplotypes (Hssr2, Hssr19, and Hssr21) were unique to single individuals.

Of nine haplotypes recovered for *B. alleghaniensis*, seven were shared with *B. papyrifera* and showed a similar pattern of geographic distribution (Figure 1b). The

common haplotype Hssr16 and closely-related Hssr13 and Hssr17 formed a large eastern group. Hssr10 and Hssr11 occurred along the Atlantic coast. Populations west of the Great Lakes were composed predominantly of Hssr1 and Hssr7, with Hssr16 at a lower frequency. Haplotypes Hssr12 and Hssr18 were unique to single individuals. The haplotype distribution for *B. lenta* differed from that of *B. papyrifera* and *B. alleghaniensis*, as a single haplotype (Hssr6) was found throughout its geographic range (Figure 1c). CpSSR haplotype distributions for *B. populifolia* and *B. cordifolia* (Figures 1d & 1e) were similar to those of sympatric *B. papyrifera* and *B. alleghaniensis* with populations composed mostly of Hssr16 and closely-related haplotypes (Hssr13, Hssr14, Hssr15). *Betula uber* (Figure 1f) was fixed for the single haplotype Hssr6 that was found in *B. lenta*.

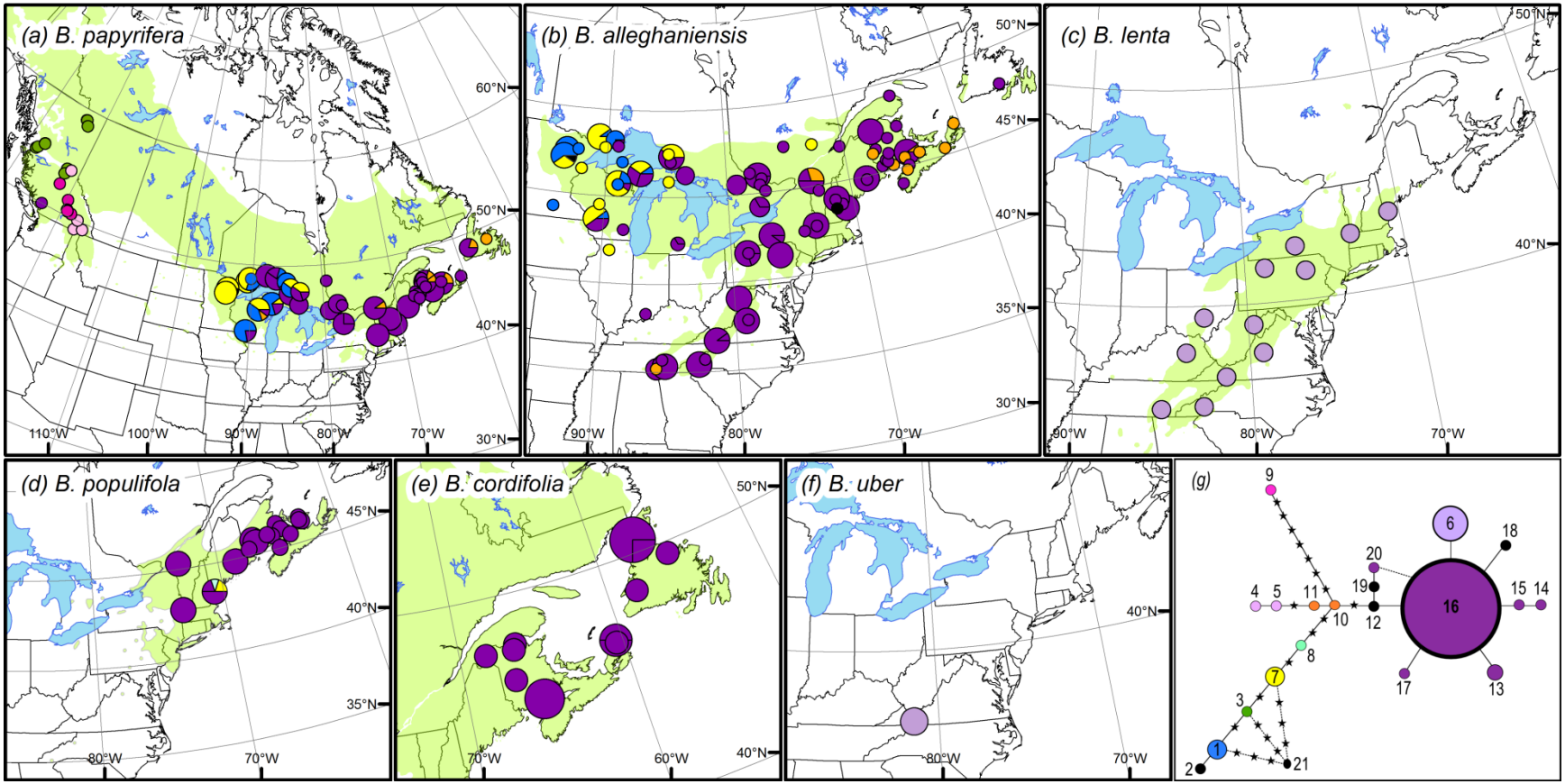



Figure 1. Geographic distribution of cpSSR haplotypes recovered for (a) *B. papyrifera*, (b) *B. alleghaniensis*, (c) *B. lenta*, (d) *B. populifolia*, (e) *B. cordifolia*, (f) *B. uber*, and (g) minimum spanning network of 21 cpSSR haplotypes.
Haplotypes were coloured-coded to represent major groups, while haplotypes unique to a single individual are coloured in black.

P*sb*A-*trn*H sequencing revealed 12 cpDNA haplotypes for the six eastern birches and haplotype distributions generally showed similar patterns as cpSSR data. The majority of the eastern range of *B. papyrifera* and *B. alleghaniensis* was composed of a group of closely-related haplotypes (H4, H5, H6, H12) which occurred from the northern Appalachians to New England and the Maritimes through Ontario (Figure 2a & b). Haplotype H8 was common along the Atlantic coast and was also found in several *B. papyrifera* populations from British Columbia. A separate haplotype group (H1, H9, H10, and H11) was predominant in the central and western Great Lakes region. In contrast to cpSSR data, the southern and central Appalachian range of *B. alleghaniensis* was composed of a mixture of two common haplotypes (H6 and H7). Three singleton haplotypes (H1, H3, and H5) were recovered for *B. papyrifera* and 1 singleton (H2) was recovered for *B. alleghaniensis*. *Betula lenta* was again fixed for a single haplotype (H7) (Figure 2c). *Psb*A-*trn*H haplotype distributions for *B. populifolia* and *B. cordifolia* (Figure 2d& e) were similar to those of *B. papyrifera* and *B. alleghaniensis* in sympatric areas, while *B. uber* (Figure 2f) was again fixed for a single haplotype (H7) which was shared with *B. lenta* and *B. alleghaniensis* at the same locality.

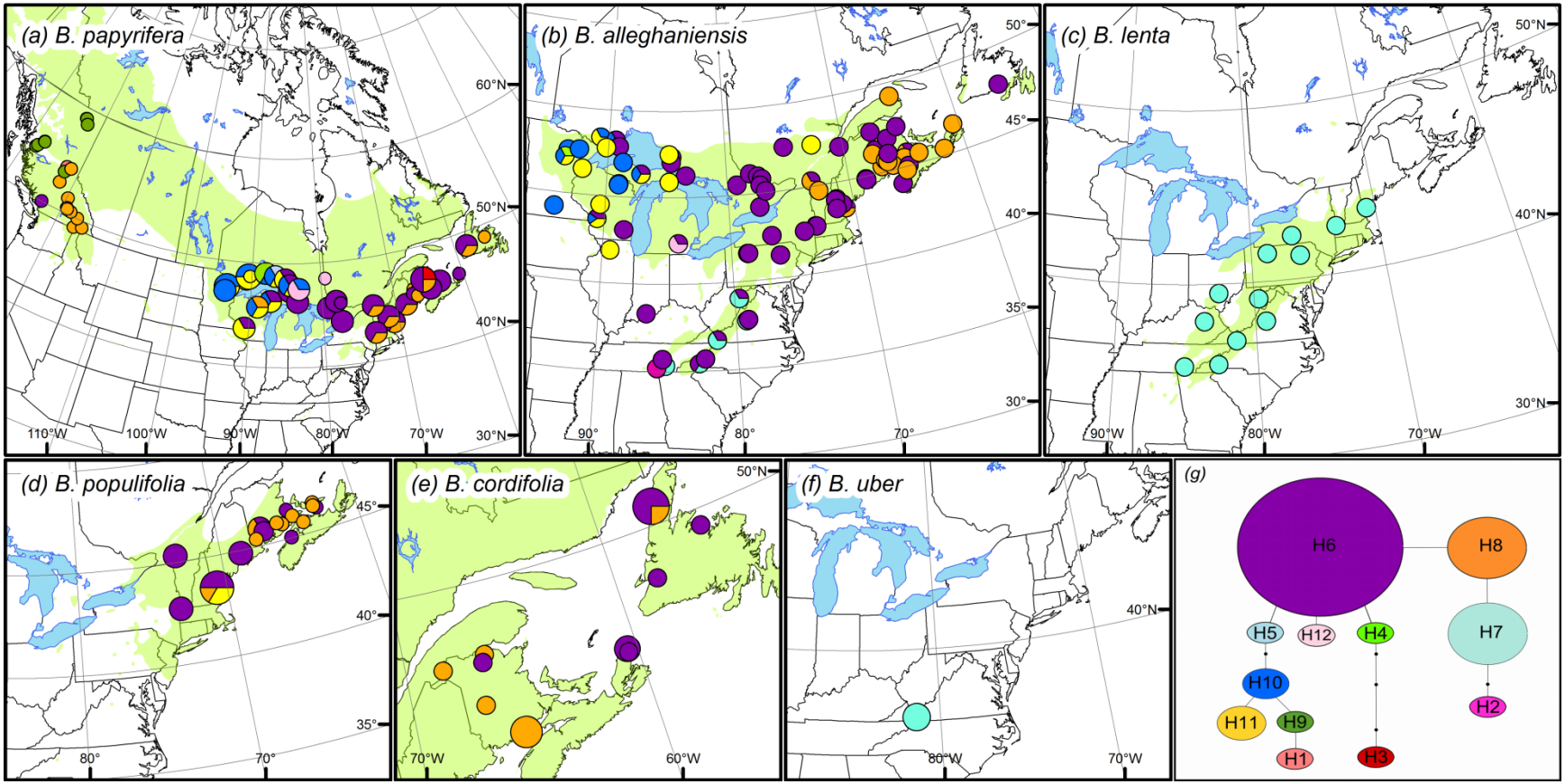



Figure 2. Geographic distribution and *psb*A-*trn*H haplotypes recovered for (a) *B. papyrifera*, (b) *B. alleghaniensis*, (c) *B. lenta* (d) *B. populifolia*, (e) *B. cordifolia*, (f) *B. uber*, and (g) statistical parsimony network of 12 *psb*A-*trn*H haplotypes.

*CpDNA Diversity*

Standardized estimates after rarefaction revealed that *B. papyrifera* had the highest cpSSR haplotype richness (11.2), followed by *B. alleghaniensis* (6.1), and *B. lenta* (0.0) (Table 3). Estimates of gene diversity were slightly higher for *B. papyrifera* than *B. alleghaniensis,* with values of 0.67 and 0.60, respectively. Interpolated maps of cpSSR haplotype richness for *B. papyrifera* and *B. alleghaniensis* revealed that genetic diversity was highest in the western Great Lakes region for both species (Figure 3). There were also areas of moderate diversity located in the Canadian Maritime range of *B. papyrifera* and in the Ontario/Quebec range of *B. alleghaniensis*.

Table 3. Standardized estimates of gene diversity and allelic richness after rarefaction (N=120) for *B. papyrifera*, *B. alleghaniensis*, and *B. lenta*.

| Species | N | No. Haplotypes | Gene diversity | Haplotype richness |
|---|---|---|---|---|
| *B. papyrifera* | 261 | 15 | 0.67 | 11.2 |
| *B. alleghaniensis* | 314 | 9 | 0.60 | 6.1 |
| *B. lenta* | 120 | 1 | 0.00 | 0.0 |

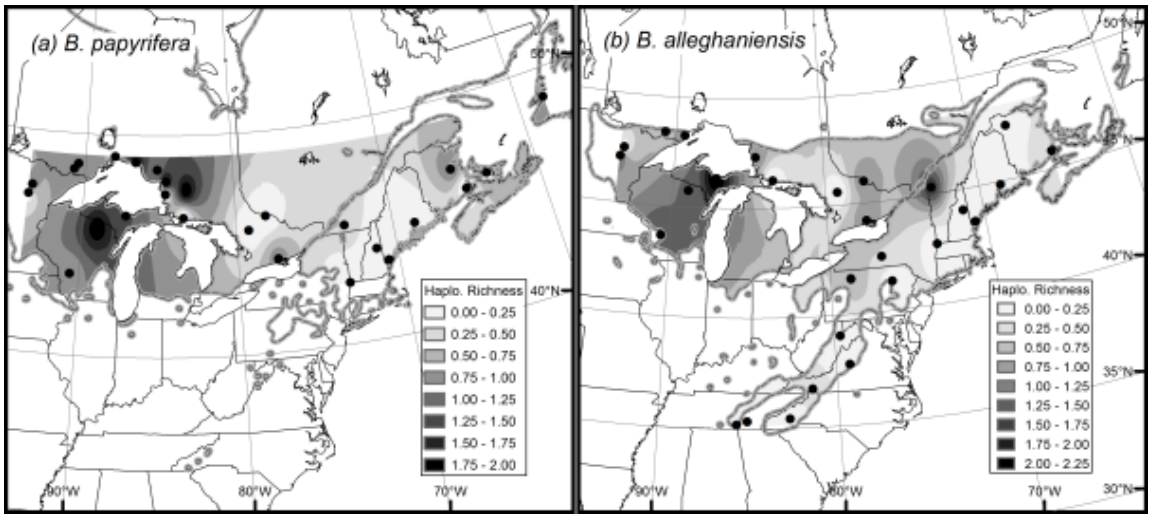



Figure 3. Mapped cpSSR haplotype richness for (a) *B. papyrifera*, and (b) *B. alleghaniensis*.

Values were determined by interpolation of haplotype richness values for sampled populations using ArcGIS. The interpolated grid is clipped to the extent of the sampled populations so as not to extrapolate outside of the study area.

*Partitioning of molecular variation and $N_{ST}$ vs. $G_{ST}$*

Analyses of molecular variance indicated substantial differentiation in haplotype frequencies among populations of *B. papyrifera* (56%) and *B. alleghaniensis* (61%) (Table 4). Population genetic subdivision ($G_{ST}$) was strong in both species, but was somewhat higher for *B. alleghaniensis* (0.601) than *B. papyrifera* (0.526). In contrast, *B. lenta* did not demonstrate variation within- or among-populations. $N_{ST}$ was significantly higher than $G_{ST}$ in *B. papyrifera* ($G_{ST} = 0.526$, $N_{ST} = 0.691$, $P < 0.05$) and *B. alleghaniensis* ($G_{ST} = 0.601$, $N_{ST} = 0.725$, $P < 0.05$) indicating the presence of a phylogeographic structure in both species.

Table 4. Analysis of molecular variance (AMOVA) and tests of phylogeographic structure ($N_{ST}$ vs. $G_{ST}$) for *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* based on cpSSR haplotype frequencies.

| Species | Source of variation | d.f. | Variation (%) | $G_{ST}$ | $N_{ST}$ | $N_{ST} > G_{ST}$ P-value |
|---|---|---|---|---|---|---|
| *B. papyrifera* | Among populations | 25 | 56.2 | 0.526 | 0.691 | 0.012* |
| | Within populations | 196 | 43.8 | - | - | - |
| *B. alleghaniensis* | Among populations | 28 | 61.3 | 0.601 | 0.725 | <0.0001* |
| | Within populations | 232 | 38.7 | - | - | - |
| *B. lenta* | Among populations | 11 | 0.0 | 0.000 | 0.000 | - |
| | Within populations | 108 | 0.0 | - | - | - |

When *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* were considered jointly, AMOVA revealed that the majority of cpSSR haplotype variation was partitioned among species (40.7%) and among populations within species (35.2%) (Table 5). AMOVA also revealed strong partitioning of molecular variance among ploidy levels (57.5%). Variation among polyploids *B. papyrifera* and *B. alleghaniensis* was low (0.3%) and non-significant.

Table 5. Analysis of molecular variance among species, ploidy levels, and polyploid species based on cpSSR haplotype frequencies. Only populations with 3 or more individuals were included in the analysis.

| Grouping | Source of variation | d.f. | Variation (%) | Fixation indices | P-value |
|---|---|---|---|---|---|
| (a) | Among species | 2 | 40.7 | $F_{CT}$=0.40710 | <0.001 |
| | Among populations within species | 64 | 35.2 | $F_{SC}$=0.59289 | <0.001 |
| | Within populations | 536 | 24.1 | $F_{ST}$=0.75862 | <0.001 |
| (b) | Among ploidy levels | 1 | 57.5 | $F_{CT}$=0.57539 | <0.001 |
| | Among populations within ploidies | 64 | 25.2 | $F_{SC}$=0.59452 | <0.001 |
| | Within populations | 536 | 17.2 | $F_{ST}$= 0.82783 | <0.001 |
| (c) | Among polyploid species | 1 | 0.3 | $F_{CT}$=0.00324 | 0.3287 |
| | Among populations within species | 53 | 58.7 | $F_{SC}$=0.58919 | <0.001 |
| | Within populations | 428 | 41.0 | $F_{ST}$=0.59052 | <0.001 |

(a) considering *B. papyrifera* , *B. alleghaniensis* , and *B. lenta* each as separate groups

(b) considering polyploids *B. papyrifera* and *B. alleghaniensis* as a single group and diploid *B. lenta* as a separate group

(c) considering polyploid species *B. papyrifera and B. alleghaniensis* as separate groups

**Spatial Analysis of Molecular Variation**

SAMOVA analyses based on cpSSRs revealed the highest $F_{CT}$ value for *B. papyrifera* when populations were grouped into 5 clusters ($F_{CT}$ = 0.78), consisting of three groups and two singleton populations (Figure 4a). The main group was composed of populations occupying the eastern range from New England and the Maritimes through Ontario. A second, smaller group consisted of several populations from the western Great Lakes while the third population group consisted of populations mostly from western Canada. For *B. alleghaniensis*, the highest $F_{CT}$ value was obtained when populations were grouped into 4 clusters ($F_{CT}$ = 0.93), consisting of 2 main groups and 2 singleton populations (Figure 4b). The largest group spanned most of the eastern distribution of *B. alleghaniensis* while a second, smaller group consisted mainly of western populations.

SAMOVA based on *psb*A-*trn*H data demonstrated similar geographic patterns (Figure 5), but revealed a greater number of population groups for both *B. papyrifera* ($K$=11, $F_{CT}$ = 0.77) and *B. alleghaniensis* ($K$=7, $F_{CT}$ = 0.75).

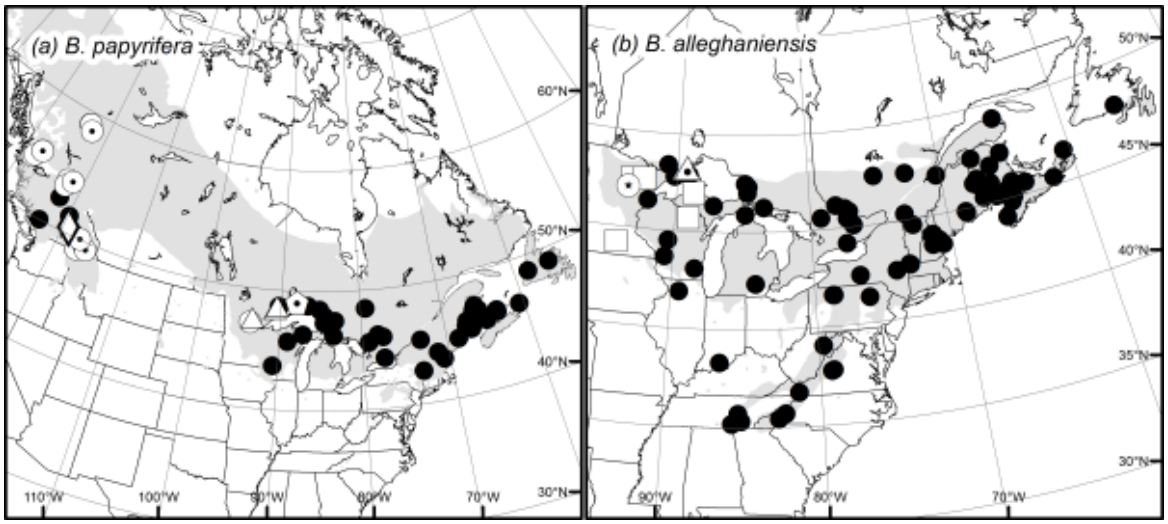


Figure 4. SAMOVA population groups based on cpSSR haplotypes for (a) *B. papyrifera* with 5 population groups and $F_{CT}$ = 0.78, and (b) *B. alleghaniensis* with 4 population groups and $F_{CT}$ = 0.93.

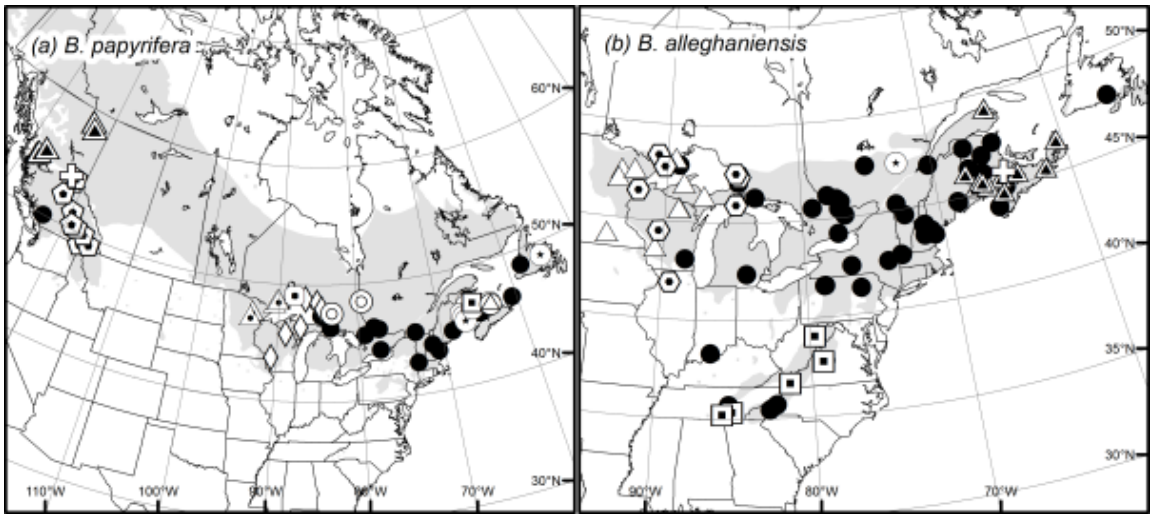


Figure 5. SAMOVA population groups based on *psb*A-*trn*H haplotyps for (a) *B. papyrifera* with 11 population groups and $F_{CT}$ = 0.77 and (b) *B. alleghaniensis* with 7 population groups and $F_{CT}$ = 0.75.

## Discussion

*Phylogeographic patterns, refugia, and post-glacial migration*

The present study revealed congruent spatial patterns of haplotype variation among species despite some differences in haplotype compositions (Figures 1 & 2). Haplotype distributions and SAMOVA analyses indicated two main population groups for *Betula* in eastern North America: (i) a main group occurring from the southern Appalachians to New England and the Maritimes, Quebec, and eastern Ontario; and (ii) a group of mixed haplotype composition occurring in the western Great Lakes region (Figure 4). These groups correspond well with LGM fossil pollen and macrofossil reconstructions for *Betula* in eastern North America which indicate the presence of low-density *Betula* populations in the central/southern Appalachians, the Midwest, and central Louisiana between 18000 and 15000 years ago (Delcourt & Delcourt, 1987; Jackson et al., 1997). The majority of the eastern portion of the study area, including the Appalachians, New England, the Maritimes, Quebec, and eastern Ontario, was probably colonized from Appalachian refugial populations, as indicated by the presence of a single predominant haplotype group across this area. Haplotypes originating in the Appalachian refugium (Hssr16 or closely-related haplotypes) would have colonized northwards and then westwards after the retreat of the glaciers. Western Ontario and the Great Lakes states appear to have been colonized by haplotypes originating from the Midwest region, but it is possible that some haplotypes may have originated from populations of far-western Canada and the United States. The lack of sampling in the Midwest (outside the natural range of study species) makes it difficult to know if the western Great Lakes haplotypes originated from the Midwest or elsewhere. Western Great Lakes populations

of jack pine (*Pinus banksiana*) and black spruce (*Picea mariana*) appear to have recolonized from the Midwest region (Jaramillo-Correa et al., 2004; Godbout et al., 2005), and this also may be the case for *Betula*. In addition to the two main population groups, a local haplotype group for *B. papyrifera* and *B. alleghaniensis* was recovered along the Atlantic coast of Canada (Hssr10 and 11). The presence of these endemic haplotypes suggests that this area may have been populated from low density populations or from a cryptic refuge located in this area, although the lack of fossil evidence suggests that such populations would have been present at low density. Local haplotype differentiation also has been found for jack pine and black spruce in the same area, suggesting that the Atlantic coast may have harboured low-density LGM populations of numerous boreal-temperate plants (Jaramillo-Correa et al., 2004, 2009; Godbout et al., 2005).

Haplotype diversity in *B. papyrifera* and *B. alleghaniensis* was highest in the northwestern Great Lakes region (Figure 3). This region appears to represent a zone of biogeographic contact between haplotype lineages originating from the Midwest or other western refugia (Hssr1 and Hssr7) along with eastern lineages derived from the Appalachian region (Hssr16 and related). The relatively high haplotype diversity observed in central Quebec and Maritime populations of *B. papyrifera* and *B. alleghaniensis* may represent another biogeographic contact zone; populations of *B. alleghaniensis* from central Quebec contained haplotypes which appear to come from the Appalachian, Atlantic coast, and Midwest refugial populations while *B. papyrifera* Maritime populations contained haplotypes from the proposed Appalachian and Atlantic lineages. The finding of areas of highest diversity at relatively northern latitudes for

eastern North American *Betula* species is in agreement with studies of other North American trees; high genetic diversity of Quebec and Maritime populations of jack pine (Godbout et al., 2005) and black spruce (Jaramillo-Correa et al., 2004) was also attributed to mixing of eastern and western haplotype lineages. The evidence for biogeographic contact for birches in the western Great Lakes region is unusual among the North American trees that have been studied (Jaramillo-Correa et al., 2004; Godbout et al., 2005; McLachlan et al., 2005; Saeki et al., 2011). The present study appears to be the first to report a zone of biogeographic contact in the western Great Lakes region. It is possible that the high haplotype diversity of *B. papyrifera* and *B. alleghaniensis* in the western Great Lakes also may be partially attributable to introgression between the two species, especially considering that this region represents a zone of natural hybridization between *B. papyrifera* and *B. alleghaniensis* (Barnes et al., 1974).

*Contrasts among species*

*Betula papyrifera* and *B. alleghaniensis* showed similar phylogeographic patterns and extensive haplotype sharing within geographic regions, although there were some notable differences. For example, *B. alleghaniensis* had higher among-population variation (Table 3), although haplotype richness after rarefaction was higher for *B. papyrifera* than for *B. alleghaniensis* (Table 2). The higher level of haplotype diversity and lower population genetic subdivision found in *B. papyrifera* may be related to its relatively high cold-tolerance and broad habitat distribution, with larger population sizes and greater population connectivity in the glacial and post-glacial landscapes. *B. papyrifera* is a habitat generalist, occurring across a broad range of site and soil types and

ranges nearly to the limit of tree growth in the arctic forest-tundra ecotone (Table 1). In comparison, *B. alleghaniensis* is more of a habitat specialist, and has a more southern distribution. Due to its greater ecological amplitude, *B. papyrifera* should have been more broadly distributed within shared glacial refuges and post-glacial habitats than *B. alleghaniensis*, and as a result would have maintained larger population sizes and experienced reduced genetic drift. This finding is in agreement with other studies that have demonstrated strong correlations between levels of genetic diversity and habitat distribution for boreal-temperate trees (Hamrick et al., 1992; Aguinagalde et al., 2005). Based on its transcontinental distribution, it is likely that *B. papyrifera* occupied refugia in far western North America, whereas *B. alleghaniensis* is thought to have persisted only in eastern and central refugia mentioned previously. These far-western refugia could have been located just south of the Cordilleran ice sheet, as has been documented for other western plant species (Soltis et al., 1997) or in Alaska, as has been documented for co-distributed white spruce (*Picea glauca*) (Anderson et al., 2006).

*Betula lenta* was the outlier in the present study, in that it was fixed for a single haplotype and therefore lacked phylogeographic structure. The low cpDNA diversity of *B. lenta* probably can be explained in terms of its low-cold tolerance and restricted geographic range. Based on the current geographical range of *B. lenta*, it should have been restricted to the Appalachian refugium during the LGM where low gene flow could have led to haplotype fixation through random genetic drift. We also can expect that the relatively low cold-tolerance of *B. lenta* would be associated with small population size within the refugial area, further exacerbating bottlenecks leading to rapid loss of haplotypes, as has been documented in a number of northern hemisphere trees (Demesure

et al., 1996; Echt et al., 1998; Vendramin et al., 2008). Still, the finding of just one cpSSR haplotype for *B. lenta* is surprising, given that levels of intraspecific polymorphism of cpSSR markers are generally high (Provan et al., 2001).

*Haplotype sharing and introgression*

The study uncovered extensive, geographically-structured haplotype sharing between *B. papyrifera* and *B. alleghaniensis*. These species also shared haplotypes with *B. populifolia* and *B. cordifolia* throughout their sympatric distribution. Haplotype H6 was shared between the four species throughout the eastern study area, and H8 was shared along the Atlantic coast of Canada (Figure 2). Haplotypes H10 and H11 were restricted to a relatively small area of the western Great Lakes region and were shared between *B. papyrifera* and *B. alleghaniensis*. In the Appalachian region, *B. lenta* and *B. uber* shared H7, for which they were both fixed. Possible explanations for interspecific haplotype-sharing include homoplasy, shared ancestral polymorphismss, or introgression (Palme et al., 2004). Homoplasy is an unlikely explanation since it would be expected to result in a random spatial distribution of shared haplotypes. Shared ancestral polymorphism, on the other hand, should result in widespread haplotype sharing and may be responsible for haplotype sharing between *B. lenta* and *B. uber*. Studies of *B. lenta* and *B. uber* based on morphology (Sharik & Ford, 1984) and nuclear DNA sequences (Li et al., 2005) indicate a close phylogenetic relationship between the two species and it also has been suggested that *B. uber* represents a variant of *B. lenta* (Sharik, 1990). Thus, it seems likely that the sharing of haplotype H7, for which both species were fixed, may reflect shared ancestral polymorphism. In contrast, haplotype sharing due to localized

introgression should result in geographic structuring of shared haplotypes, as was found for the shared haplotypes in *B. papyrifera*, *B. alleghaniensis*, *B. populifolia*, and *B. cordifolia* in our study. Similar patterns have been documented for other angiosperm tree groups including European birches (Palme et al., 2004; Maliouchenko et al., 2007), ashes (Heuertz et al., 2006), maples (Saeki et al., 2011), oaks (Petit et al., 1993, 1997; Belahbib et al., 2001) and eucalypts (McKinnon et al., 2001), suggesting that this is not an isolated phenomenon.

Despite extensive evidence of cpDNA introgression between birches in the present study, the species are morphologically distinct and do not display many morphological intermediates (Table 1). One explanation for cpDNA introgression amongst otherwise distinct species is related to interspecific pollen competition (Rieseberg et al., 1995). Formation of hybrids is likely to occur when the majority of the pollen load on a species that is present in low frequency originates from a reproductively compatible species present in high frequency (Rieseberg et al., 1995). If the resultant hybrids are receptive to pollen of the abundant species, then hybrids may serve as the seed parent in repeated backcrossings leading to the formation of hybrid individuals with the cytoplasmic genome of the minority species but an increasing nuclear genome content of the majority species. It has been suggested that such pollen competition may play a role in the formation of hybrids as a means of invasion of post-glacial habitats (Petit et al., 2003b). If related species are already present within post-glacial habitats, then colonization may be achieved through swamping of pollen of an invading species onto a local species. Knowing that *B. papyrifera* macrofossils were present across post-glacial habitats in eastern North America before those of *B. alleghaniensis* started to appear

(Jackson et al., 1997), this suggests that resident *B. papyrifera* may have contributed genes through seeds while *B. alleghaniensis* may have contributed genes mostly through hybridization via pollen.

The absence of shared cpSSR haplotypes between diploid *B. lenta* and the polyploid birches suggests that little hybridization and introgression occurs between them. However, haplotype distributions based on *psb*A-*trn*H sequences indicate that *B. lenta* and *B. alleghaniensis* share haplotype H7 in the southern Appalachian region. The sharing of H7 in this region could be due either to ancestral polymorphism or to introgression, but it seems likely that it represents ancestral polymorphism due to close phylogenetic relationship (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008) and the lower evolutionary rate in psbA-trnH in comparison to the cpSSRs. There are a number of possible explanations for the low incidence of introgression between *B. lenta* and the polyploid birches. The natural ranges of *B. papyrifera* and *B. lenta* overlap in only a small area, which limits their opportunity to interbreed. Also, controlled pollination experiments indicate that crosses between these two species produced seeds with low viability and germination rates (Clausen, 1966). The apparent isolation of *B. lenta* and *B. alleghaniensis* is somewhat more surprising, given that they are considered sister-species (Jarvinen et al., 2004; Schenk et al., 2008), overlap in phenology and geographic distribution, and lack prezygotic barriers to interbreeding (Sharik & Barnes, 1971). While controlled-crosses frequently produce viable seeds, germination rates are low and F1 progeny have reduced fitness, suggesting that post-zygotic barriers may play a significant role in maintaining genetic isolation between the two species (Clausen, 1966; Sharik & Barnes, 1971).

## Conclusion

The high level of chloroplast haplotype diversity observed in *B. papyrifera* and *B. alleghaniensis* in this study could be attributed to their broad geographic and habitat distributions translating into relatively large population sizes and greater population connectivity in the glacial and post-glacial landscape. In contrast, *B. lenta* was completely fixed for one haplotype suggesting that its low-cold tolerance and restricted distribution may have contributed to severe LGM population bottlenecks. High genetic diversity of *B. papyrifera* and *B. alleghaniensis* of the western Great Lakes region is suggestive of a zone of biogeographic contact between eastern and western glacial lineages, and has not been previously documented in other North American trees. *Betula papyrifera* and *B. alleghaniensis* demonstrated extensive sharing of localized haplotypes suggesting a high level of past introgression, whereas *B. lenta* demonstrated little evidence of introgression with the polyploid birches. Inclusion of *B. populifolia* and *B. cordifolia* revealed that these species also share localized haplotypes with *B. papyrifera* and *B. alleghaniensis* across their sympatric distribution while *B. uber* shared its single haplotype with *B. lenta*. Such high levels of haplotype sharing across multiple codistributed species suggest that hybridization and introgression within LGM habitats and during postglacial expansion has strongly contributed to the phylogeographic structure of the eastern North American birches.

**References**


Aguinagalde I., Hampe A., Mohanty A., Martín J.P., Duminil J., & Petit R.J. (2005) Effects of life-history traits and species distribution on genetic structure at maternally inherited markers in European trees and shrubs. *Journal of Biogeography*, **32**, 329–339.

Alam M.T. & Grant W.F. (1972) Interspecific hybridization in birch (*Betula*). *Le Naturaliste Canadien*, **99**, 33–40.

Anderson L.L., Hu F.S., Nelson D.M., Petit R.J., & Paige K.N. (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12447–12450.

Barnes B.V., Dancik B.P., & Sharik T.L. (1974) Natural hybridization of yellow birch and paper birch. *Forest Science*, **20**, 215–221.

Belahbib N., Pemonge M.H., Ouassou A., Sbay H., Kremer A., & Petit R.J. (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* Morocco. *Molecular Ecology*, **10**, 2003–2012.

Bennett K.D., Tzedakis P.C., & Willis K.J. (1991) Quaternary refugia of north European trees. *Journal of Biogeography*, **18**, 103–115.

Clausen K.E. (1966) Studies of compatibility in *Betula*. *Joint Proceedings of the Second Genetics Workshop of the Society of American Foresters and the Seventh Lake States Forest Tree Improvement Conference* pp. 48–52. U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Clement M., Posada D., & Crandall K.A. (2000) TCS : a computer program to estimate gene genealogies. *Molecular Ecology*, 1657–1659.

DeJong P.C. (1993) An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. *Betula: Proceedings of the IDS Betula Symposium* (ed. by D. Hunt), pp. 7–18. Richmond, England.

Delcourt P.A. & Delcourt H.R. (1987) *Long-term forest dynamics of the temperate zone :a case study of late-quaternary forests in eastern North America.* Springer-Verlag, New York.

Demesure B., Comps B., & Petit R.J. (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.

Dumolin-Lapègue S., Demesure B., Fineschi S., Le Corre V., & Petit R.J. (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.

Dupanloup I., Schneider S., & Exoffier L. (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**, 2571–2581.

Echt C.S., DeVerno L.L., Anzidei M., & Vendramin G.G. (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology*, **7**, 307–316.

Erdmann G.G. (1990) Yellow birch (*Betula alleghaniensis* Britton). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 133–147. U.S. Department of Agriculture Forest Service, Washington, DC.

Excoffier L., Laval G., & Schneider S. (2005) Arlequin V. 3.1. An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier L. & Smouse P.E. (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics*, **136**, 343–359.

Furlow J.J. (1990) The genera of Betulaceae in the southeastern United States. *Journal of the Arnold Arboretum*, **71**, 1–67.

Furlow J.J. (1997) Betulaceae. *Flora of North America North of Mexico. Volume 3* (ed. by Flora of North America Editorial Committee), pp. 507–538. Oxford University Press, New York.

Godbout J., Jaramillo-Correa J.P., Beaulieu J., & Bousquet J. (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology*, **14**, 3497–3512.

Grant W.F. & Thompson B.K. (1975) Observations on Canadian birches, *Betula cordifolia*, *B. neoalaskana*, *B. populifolia*, *B. papyrifera*, and *B. x caerulea*. *Canadian Journal of Botany*, **53**, 1478–1490.

Hamrick J.L., Godt M.J.W., & Sherman-Broyles S.L. (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Heuertz M., Carnevale S., Fineschi S., Sebastiani F., Hausman J.F., Paule L., & Vendramin G.G. (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (*Oleaceae*): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–2140.

Heuertz M., Fineschi S., Anzidei M., Pastorelli R., Salvini D., Paule L., Frascaria-Lacoste N., Hardy O.J., Vekemans X., & Vendramin G.G. (2004) Chloroplast DNA variation and postglacial recolonization of common ash (*Fraxinus excelsior* L.) in Europe. *Molecular Ecology*, **13**, 3437–3452.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Jackson S.T., Overpeck J.T., Webb-III T., Keattch S.E., & Anderson K.H. (1997) Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**, 1–70.

Jaramillo-Correa J.P., Beaulieu J., & Bousquet J. (2004) Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. *Molecular Ecology*, **13**, 2735–2747.

Jaramillo-Correa J.P., Beaulieu J., Khasa D.P., & Bousquet J. (2009) Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research*, **39**, 286–307.

Jarvinen P., Palme A., Morales L.O., Lannenpaa M., Keinanen M., Sopanen T., & Lascoux M. (2004) Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. *American Journal of Botany*, **91**, 1834–1845.

Lamson N.I. (1990) Sweet birch (*Betula lenta* L.). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 148–152. United States Department of Agriculture Forest Service., Washington, DC.

Li J., Shoup S., & Chen Z. (2005) Phylogenetics of *Betula* (Betulaceae) inferred from sequences of nuclear ribosomal DNA. *Rhodora*, **107**, 69–86.

Maliouchenko O., Palmé A.E., Buonamici A., Vendramin G.G., & Lascoux M. (2007) Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography*, **34**, 1601–1610.

McKinnon G.E., Vaillancourt R.E., Jackson H.D., & Potts B.M. (2001) Chloroplast sharing in the Tasmanian eucalypts. *Evolution*, **55**, 703–711.

McLachlan J.S., Clark J.S., & Manos P.S. (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.

Palme A.E., Su Q., Palsson S., & Lascoux M. (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology*, **13**, 167–178.

Peakall R. & Smouse P.E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Petit R.J., Aguinagalde I., De Beaulieu J.-L., Bittkau C., Brewer S., Cheddadi R., Ennos R., Fineschi S., Grivet D., Lascoux M., Mohanty A., Muller-Starck G., Demesure-Musch B., Palme A., Martin J.P., Rendell S., & Vendramin G.G. (2003a) Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, **1563**, 1563–1565.

Petit R.J., Bodenes C., Ducousso A., Roussel G., & Kremer A. (2003b) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Petit R.J., Kremer A., & Wagner D.B. (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, **87**, 122–128.

Petit R.J., El Mousadik A., & Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Petit R.J., Pineau E., Demesure B., Bacilieri R., Ducousso A., & Kremer A. (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*, **94**, 9996–10001.

Pons O. & Petit R.J. (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.

Provan J., Powell W., & Hollingsworth P.M. (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution*, **16**, 142–147.

Rieseberg L.H., Desrochers A.M., & Youn S.J. (1995) Interspecific pollen competition as a reproductive barrier between sympatric species of *Helianthus* (Asteraceae). *American Journal of Botany*, **82**, 515–519.

Saeki I., Dick C.W., Barnes B.V., & Murakami N. (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): impacts of habitat specialization, hybridization and glacial history. *Journal of Biogeography*, **38**, 992–1005.

Safford L.O., Bjorkbom J.C., & Zasada J.C. (1990) Paper birch (*Betula papyrifera* Marsh.). *Silvics of North America. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 158–171. United States Department of Agriculture, Washington, DC.

Sang T., Crawford D., & Stuessy T.F. (1997) Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American journal of botany*, **84**, 1120–1136.

Schenk M.F., Thienpont C.-L., Koopman W.J.M., Gilissen L.J.W.J., & Smulders M.J.M. (2008) Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics & Genomes*, **4**, 911–924.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 1–2.

Sharik T.L. (1990) *Virginia Round-Leaf Birch Recovery Plan.* U.S. Fish and Wildlife Service, Newton Corner, MA.

Sharik T.L. & Barnes B.V. (1971) Hybridization in *Betula alleghaniensis* Britt. and *B. lenta* L.: A comparative analysis of controlled crosses. *Forest Science*, **17**, 415–424.

Sharik T.L. & Ford R.H. (1984) Variation and Taxonomy of *Betula uber*, *B. lenta*, and *B. alleghaniensis*. *Brittonia*, **36**, 307–316.

Soltis D.E., Gitzendanner M.A., Strenge D.D., & Soltis P.S. (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**, 353–373.

Soltis D.E., Morris A.B., McLachlan J.S., Manos P.S., & Soltis P.S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, **15**, 4261–4293.

Stewart J.R. & Lister A.M. (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.

Stewart J.R., Lister A.M., Barnes I., & Dalén L. (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B*, **277**, 661–671.

Taberlet P., Fumgalli L., Wust-Saucy A.-G., & Cosson J.-F. (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Vendramin G.G., Fady B., González-Martínez S.C., Hu F.S., Scotti I., Sebastiani F., Soto A., & Petit R.J. (2008) Genetically depauperate but widespread: the case of an emblematic Mediterranean pine. *Evolution*, **62**, 680–688.

Weising K. & Gardner R.C. (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9–19.

Willis K., Rudner E., & Sümegi P. (2000) The full-glacial forests of central and southeastern Europe. *Quaternary Research*, **53**, 203–213.

Woodworth R.H. (1930) Cytological Studies on the *Betulaceae*. IV. *Betula*, *Carpinus*, *Ostrya*, *Ostryopsis*. *Botanical Gazette*, **90**, 108–115.

Zeng J., Zou Y.-P., Bai J.-Y., & Zheng H.-S. (2002) Preparation of total DNA from "recalcitrant plant taxa". *Acta Botanica Sinica*, **44**, 694–697.

## Chapter 2: Population genetic structure of *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta* as revealed by nuclear microsatellites


### Abstract

We sought to elucidate patterns of genetic diversity and population structure of eastern North American birches and infer how patterns have been shaped by glacial/postglacial population dynamics and contemporary gene flow. Polyploids *B. alleghaniensis* (6x), *B. papyrifera* (5x, 6x), and diploid *B. lenta* (2x) were genotyped at eight microsatellite loci. Maps of allelic richness and gene diversity were used to examine the spatial distribution of genetic diversity within each species and Bayesian analysis was used to identify patterns of population genetic subdivision. Populations of all species were weakly differentiated across the study area ($\Phi_{PT}$*alleghaniensis*=0.019, $\Phi_{PT}$*papyrifera*=0.028, $\Phi_{PT}$*lenta*=0.022). No clear pattern of genetic structure was found, and populations were differentiated only over large geographic scales due to isolation-by-distance. High levels of allelic richness were observed for *B. alleghaniensis* ($A_R$=12.5) and *B. papyrifera* ($A_R$=12.2), and *B. lenta* had low allelic richness ($A_R$=4.1). *Betula alleghaniensis* possessed its greatest allelic richness in the western Great Lakes region, which is a presumed region of secondary contact between eastern and western glacial lineages. The overall low levels of population genetic structure and lack of population genetic subdivision observed for birches in the current study is typical of what would be expected for species with high outcrossing rates and efficient dispersal of both pollen and seeds. High levels of nuclear microsatellite diversity observed for *B. alleghaniensis* and *B. papyrifera* probably reflect high current and historic effective population sizes with high contemporary gene flow, as well as their polyploid origins which contribute to fixed heterozygosity. For *B. lenta*, the finding of low microsatellite diversity can probably be explained in terms of bottlenecks before or during the time of the last glacial maximum.



Key Words: *Betula*, birches, eastern North America, population structure, genetic diversity, gene flow, postglacial recolonization

## Introduction

Patterns of genetic variation are shaped by complex factors, including both contemporary and historic population dynamics. In many northern hemisphere trees, population genetic studies have revealed evidence of historic population dynamics associated with glacial-interglacial cycles as well as contemporary patterns of gene flow which have developed since the last glacial maximum (LGM) (Heuertz et al., 2004b; Mimura & Aitken, 2007; O'Connell et al., 2008; Marsico et al., 2009; Potter et al., 2012). For example, the finding of distinct eastern and western population groups in *Fraxinus excelsior* is thought to be due to the separation of eastern and western populations within separate glacial refugia during the LGM (Heuertz et al., 2004b). Despite contemporary high levels of gene flow within regions, western and eastern European populations remain genetically differentiated. In western North American *Quercus garryana*, declining nuclear microsatellite diversity was observed towards the northward range periphery as a result of founding events during postglacial colonization (Marsico et al., 2009). However, the same study also revealed high levels of contemporary gene flow as a result of high pollen dispersal among populations. In contrast, a study of *Picea sitchensis* in western North America suggested that levels of gene flow were high among recently founded populations during postglacial recolonization but that contemporary gene flow is somewhat more limited as evidenced by the finding of a strong pattern of isolation-by-distance (Mimura & Aitken, 2007).

Most current information on patterns of genetic structure of North American trees is based on allozyme studies (e.g. Fins & Seeb 1986; Bousquet et al. 1988; Li & Adams 1989) or organelle DNA-based studies of phylogeographic history (e.g. Godbout et al.

2005; Lemieux et al. 2011; Saeki et al. 2011). For chloroplast DNA-based phylogeographic studies, the goal is generally to uncover patterns of variation and to relate the observed patterns to glacial history, including inference of the location of glacial refugia, and identify likely postglacial migration routes. While cpDNA-based phylogeographic studies have contributed greatly to understanding of patterns of glacial history of many northern hemisphere species, patterns of population structure revealed may not be reflective of patterns at nuclear markers due to the uniparental inheritance of the cpDNA genome. In angiosperms with maternal chloroplast inheritance, cpDNA markers provide information about colonization via seed dispersal, while biparentally-inherited nuclear markers reflect patterns of both seed and pollen dispersal (Oddou-Muratorio et al., 2001). Most nuclear-marker based studies of forest trees conducted thus far have been based on allozymes, and have revealed high levels of genetic diversity and low population differentiation which have been attributed to large geographic range sizes, high outcrossing rates, and high dispersal distances (Hamrick et al., 1992; Hamrick & Godt, 1996). However, these early studies suffer from low power due to the low polymorphism of allozymes and the lack of available models for inference of population structure (Heuertz et al., 2004b). In comparison to cpDNA and allozymes, nuclear microsatellite markers show higher levels of polymorphism, and when combined with recently developed software and models for the inference of population level processes, they can be highly effective in revealing subtle patterns of population genetic structure. For example, a recent study of eastern Hemlock (*Tsuga canadensis*) based on nuclear microsatellites revealed evidence of three or four distinct population groups which may correspond to different Pleistocene glacial refuges (Potter et al., 2012) whereas previous

studies based on cpDNA markers (Lemieux et al., 2011) did not reveal any distinct pattern of population structuring. Furthermore, the greater resolution of microsatellites has helped to reveal subtle or previously unexpected patterns of genetic diversity. For example, previous studies based on allozymes revealed very low levels of genetic diversity for red pine (*Pinus resinosa*) (Fowler & Morris, 1977; Mosseler et al., 1991) but a more recent study based on nuclear microsatellites revealed higher genetic diversity than previously inferred and also revealed structuring of eastern and western populations (Boys et al., 2005).

To date, few studies based on nuclear microsatellites have been conducted for eastern North American temperate trees. Most of the studies conducted thus far have focused on species with restricted geographic ranges or low population sizes which are believed to be at conservation risk (Victory et al., 2006; Craft & Ashley, 2007; Ross-Davis et al., 2008; Potter et al., 2012). Thus, there is a lack of information on patterns of population genetic structure for many widespread and ecologically important species. Furthermore, with declines in suitable habitat expected for many eastern North American tree species over the coming decades due to climate change (Iverson et al., 2004, 2007, 2008), there is an urgent need to generate information on patterns of genetic variation to be used for future development of genetic conservation and management strategies.

Birches (*Betula* spp.) are widespread, ecologically and economically important species in temperate forests of northeastern North America (Rustad et al., 2012). Three common and iconic species in this region include paper birch (*Betula papyrifera* Marshall), yellow birch (*Betula alleghaniensis* Britton), and sweet birch (*Betula lenta* L.). They are the most widely-utilized birches in the forest industry and are used for a variety

of products including lumber and veneer (Kaiser, 1993). From an ecological standpoint, these species are ubiquitous within temperate forests of the eastern United States and Canada and form a large component of the forest biomass in many areas. *Betula papyrifera* has been shown to increase soil and site productivity through deposition of calcium and other nutrients from its leaves and also reduces the acidity of rain falling through its crown (Safford et al., 1990). It is also important as habitat for cavity-nesting birds (Harestad & Keisker, 1989). *Betula alleghaniensis* is an important food species for browsing animals such as deer and moose, and its buds and seeds also form food for other animal species (Erdmann, 1990). It is probably the most economically-valued of the eastern birches, and its wood is used for products such as veneer, doors, cabinetry, and furniture (Kaiser, 2008). *Betula lenta* is valuable in providing nesting sites for northern hawks (Speiser & Bosakowski, 1987), and provides soil protection to slopes throughout its range in the Appalachian region (Lamson, 1990). Despite their importance, no studies of the population genetics of *B. alleghaniensis*, *B. papyrifera*, or *B. lenta* using nuclear DNA markers currently exist. A previous study of cpDNA diversity revealed high levels of population genetic substructuring between eastern and western populations of *B. papyrifera* and *B. alleghaniensis* due to the isolation of eastern and western lineages in separate glacial refugia during the LGM (Chapter 1). The same study revealed a lack of population genetic substructure and extremely low genetic diversity (haplotype fixation) in the Appalachian endemic *B. lenta*, presumably owing to LGM population genetic bottlenecks of this less cold-tolerant species. While these studies have been valuable in elucidating the post-glacial history of the eastern birches, they are probably not reflective of patterns of nuclear DNA diversity, as cpDNA studies generally show greater levels of

population substructure at cpDNA markers than at nuclear markers (Oddou-Muratorio et al., 2001; Marsico et al., 2009). Thus, there is a need to better-understand patterns of population genetic structure and genetic diversity of these important species using nuclear DNA markers. The primary objective of this study was to examine patterns of genetic diversity and population structure of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* in northeastern North America using nuclear microsatellite markers. Specifically we sought to (i) elucidate patterns of genetic diversity within and among populations across the study area, (ii) to investigate current patterns of population genetic substructure including inference of the number and location of population groups, and (iii) to infer how the observed patterns may have been shaped by both historic glacial-interglacial population dynamics and contemporary levels of gene flow.

## Methods

### *Population sampling*

Mature leaf samples were collected from natural populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* at 20 sites across the species` sympatric distribution in northeastern North America (Figure 1). At each site, sampled individuals were separated by a minimum distance of 50 m to avoid collection of multiple samples from the same clonal individual. Leaf material was dried in silica gel to preserve DNA quality. In total, samples were obtained for 483 individuals of *B. alleghaniensis* at 16 sites, 256 individuals of *B. papyrifera* at 9 sites, and 300 individuals of *B. lenta* at 12 sites (Table 1). *Betula alleghaniensis* and *B. papyrifera* co-occurred at eight sites, *B. alleghaniensis*

and *B. lenta* co-occurred at ten sites, *B. papyrifera* and *B. lenta* co-occurred at two sites, and all three species occurred together at two sites.

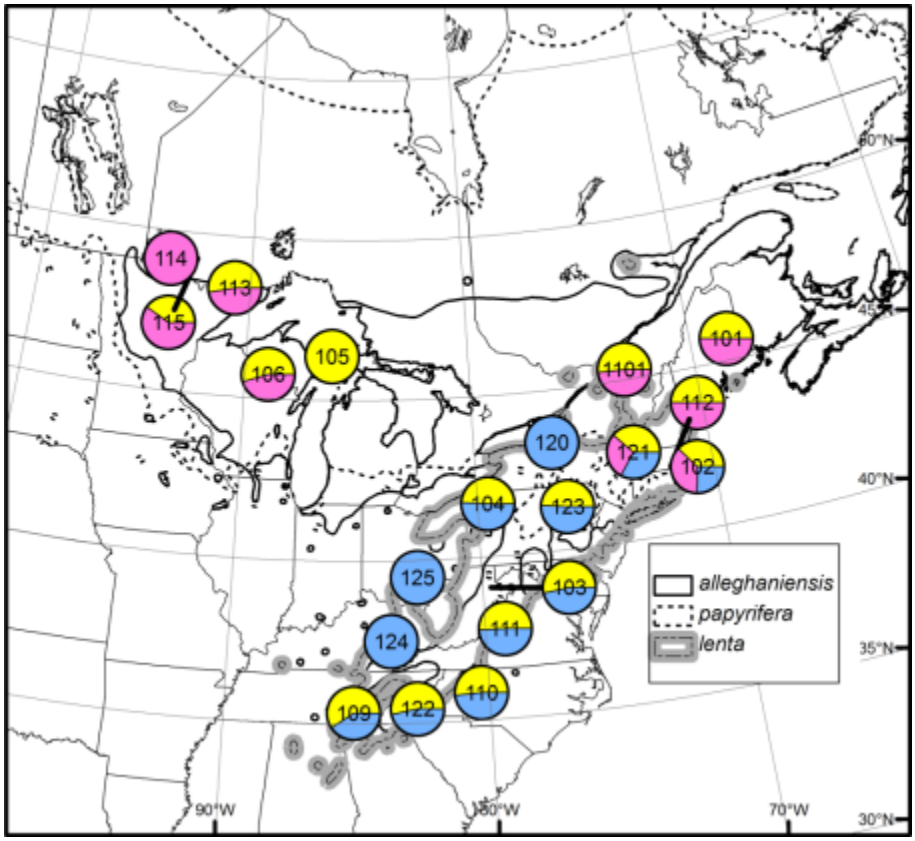



Figure 1. Distribution of *Betula* population sampling locations throughout northeastern North America.

Circles represent sampling locations and are shaded with different colours to represent the proportion of each species sampled at each location; *B. alleghaniensis* (yellow), *B. papyrifera* (pink), *B. lenta* (blue). Different line shadings are used to indicate the natural ranges of each species.

Table 1. Location and numbers of *Betula* population samples used to assess nuclear microsatellite diversity and population genetic structure.

| Population | Population Code | Province/ State | Lat. | Long. | (N) all | pap | len |
|---|---|---|---|---|---|---|---|
| Penobscot | 101 | ME | 44.86 | -68.63 | 30 | 30 | |
| Massabesic | 102 | ME | 43.44 | -70.67 | 30 | 29 | 20 |
| Fernow | 103 | WV | 39.05 | -79.67 | 30 | | 26 |
| Kane | 104 | PA | 41.60 | -78.77 | 29 | | 30 |
| Dukes | 105 | MI | 46.36 | -87.16 | 35 | | |
| Argonne | 106 | WI | 45.75 | -88.98 | 37 | 31 | |
| North River | 109 | TN | 35.33 | -85.09 | 29 | | 20 |
| Mt. Rogers | 110 | VA | 36.74 | -81.42 | 30 | | 28 |
| Glenwood | 111 | VA | 37.75 | -79.23 | 30 | | 29 |
| Bartlett | 112 | NH | 44.05 | -71.30 | 31 | 31 | |
| Greenwood | 113 | ON | 48.39 | -90.75 | 28 | 25 | |
| Marcell | 114 | MN | 47.53 | -93.47 | | 28 | |
| Blandin | 115 | MN | 47.12 | -93.68 | 20 | 30 | |
| Finger Lakes | 120 | NY | 42.49 | -76.77 | | | 19 |
| Hopkins | 121 | MA | 42.73 | -73.25 | 33 | 24 | 28 |
| Pisgah | 122 | NC | 35.43 | -82.73 | 31 | | 27 |
| Luzerne | 123 | PA | 41.32 | -76.29 | 30 | | 31 |
| Cumberland | 124 | KY | 37.88 | -83.66 | | | 27 |
| Hocking | 125 | OH | 39.47 | -82.58 | | | 15 |
| Gault | 1101 | QC | 45.32 | -73.09 | 30 | 28 | |

all, *alleghaniensis* ; pap, *papyrifera* ; len, *lenta*

*Laboratory procedures*

DNA was extracted from 20 milligrams of dried leaf tissue following the modified CTAB protocol of Zeng et al. (2002). Individuals were analyzed at eight polymorphic microsatellite markers developed in previous studies (Wu et al., 2002; Kulju et al., 2004; Truong et al., 2005; Tsuda et al., 2008). Information and sequences for primers used in this study are presented in Appendix 2. Forward primers were 5`-end labeled with infrared dye (IRD) and amplification was conducted in 10 µl total volume with multiplexed primers. Multiplex 1 contained 1 µl of 10x buffer, 0.2 mM dNTPs, 2.5

mM $MgCl_2$, 0.4 µM CD277302 forward and reverse primers, 0.2 µM L2.3 forward and reverse primers, 0.5 U Taq polymerase, and 2.0 µl of template DNA. Multiplex 2 contained 1 µl of 10x buffer, 0.2 mM dNTPs, 3.0 mM $MgCl_2$, 0.4 µM L1.10 forward and reverse primers, 0.1 µM L5.5 forward and reverse primers, 0.5 U Taq polymerase, and 2.0 µl of template DNA. For the remaining primers (L7.1a, L5.4, Bo.F330, Bp04) fluorescent labeling was accomplished through the use of a tailed-primer protocol (Schuelke, 2000). A 19-bp tail of sequence CACGACGTTGTAAAACGAC was added to the 5`-end of each forward primer and amplification was conducted in a volume of 10 µl containing 1 µl of 10X buffer, 0.2 mM dNTPs, 3.0 mM MgCl2, 0.1 µM forward primer, 0.1 µM IRD primer (CACGACGTTGTAAAACGAC), 0.5 µM reverse primer, 0.5 U taq polymerase, and 2.0 µl of template DNA. Thermocycling consisted of an initial denaturation at 94°C for 4:00 min, followed by 35 cycles of 94°C for 0:45 min, 50°C for 1:00 min, 72°C for 0:45 min and final elongation at 72°C for 10:00 min. Amplified fragments were electrophoretically separated on the Licor 4200 DNA analyzer along with the manufacturer`s 50-350 bp size standard (LiCOR Biosciences, Lincoln, NE). Fragment lengths were scored using SAGA GT 2.1 software (Licor Biosciences) and verified manually.

*Polyploid genotyping*

Due to partial heterozygosity of polyploid taxa, it is often not possible to determine the underlying allele copy number (Obbard et al., 2006). While a number of methods of estimating microsatellite allele frequencies of polyploids have been proposed (e.g. Esselink et al. 2004), these methods are generally not suited for analysis of taxa with

ploidies greater than 4x (Clark & Jasieniuk, 2011). Due to the difficulty in determining allele copy number in higher-level polyploids, many studies have used a dominant-marker approach to scoring of microsatellite alleles (e.g. Samadi et al. 1999; Rodzen et al. 2004; Andreakis et al. 2009; Helsen et al. 2009; Robertson et al. 2010). Specifically, the presence of an allele of a given fragment length is recorded as a 1 if that allele is present at a given locus, and as a 0 if it is absent. The presence/absence data for each allele and locus are then concatenated to produce multilocus allele phenotypes for each individual and the data can be analyzed in a similar manner to dominant marker data. While such an approach has in the past been the preferred method used in studies of high-order polyploids, it has some notable limitations including loss of information, since all allele phenotypes are treated as equidistant (Clark & Jasieniuk, 2011). However, the advent of recent software accommodating partial genotypic ambiguity of polyploid microsatellite data, including POLYSAT (Clark & Jasieniuk, 2011), STRUCTURE (Falush et al., 2007), and SPAGEDI (Hardy & Vekemans, 2002a) have the benefit of preserving information about genetic relationships between individuals, and have greatly improved the range of analyses available for polyploid taxa. Analyses using this approach have been successfully implemented in several recent studies of polyploid plant species including *Acacia senegal* (Assoumane et al., 2012), *Aster amellus* (Münzbergová et al., 2013), *Atriplex nummularia* (Sampson & Byrne, 2012), and *Rubus fruticosus* (Clark et al., 2012).

*Data analysis*

For genotyping of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta,* the different allele size-variants present within each individual were recorded. For the diploid *B. lenta*,

standard allele-frequency data were used. However, due to partial heterozygosity of the polyploids *B. alleghaniensis* and *B. papyrifera*, the underlying copy-number of each allele could not be determined. Thus, for *B. alleghaniensis* and *B. papyrifera* only the presence or absence of each fragment was considered and it was assumed, in all cases, that the underlying genotype of each individual was ambiguous.

Since *B. papyrifera* is a polyploid with varying chromosome numbers, the ploidy of each individual was determined based on the maximum number of alleles at a locus using the POLYSAT package for R software (Clark & Jasieniuk, 2011). Ploidies for all individuals of *B. alleghaniensis* and *B. lenta* were fixed as 6x and 2x, respectively. Genetic diversity measures including the number of alleles per locus ($A_N$), number of alleles standardized for unequal sample sizes by rarefaction ($A_R$), and gene diversity corrected for sample size ($H_E$) (Nei, 1978) were calculated using SPAGEDI software (Hardy & Vekemans, 2002a). The allelic richness after rarefaction was calculated based on the minimum sample size, which was equal to the smallest number of defined gene copies for a given population and locus. For the diploid *B. lenta*, observed heterozygosity was calculated using the program FSTAT (Goudet, 1995), and deviation from Hardy-Weinberg equilibrium was tested with 10000 randomizations to calculate significance levels. Deviation from HW expectations could not be calculated for the polyploids *B. alleghaniensis* and *B. papyrifera* as the underlying allele-frequencies were unknown. To test for the existence of recent population genetic bottlenecks in *B. lenta*, expected heterozygosity ($H_E$) based on the observed number of alleles for each population was tested against the expected heterozygosity under mutation drift equilibrium ($H_{eq}$) (Cornuet & Luikart, 1996). Since allelic richness is expected to decline more quickly than

heterozygosity under population bottlenecks, populations which have experienced a recent reduction in effective size are expected to demonstrate a significant excess of heterozygotes in comparison to the heterozygosity expected for a population at mutation-drift equilibrium. The significance of the deviation of $H_E$ from $H_{eq}$ was evaluated using the Wilcoxon signed-rank test calculated under both the IAM and the SMM using BOTTLENECK software (Piry et al., 1999). To investigate spatial patterns of genetic diversity, rarefacted allelic richness values were regressed against latitude using SIGMAPLOT software (Systat Software, San Jose, CA). In addition, values of rarefacted allelic richness and gene diversity for each population were mapped using ARCGIS (ESRI, Redlands, CA) software to visualize geographic patterns of genetic variation within each species. Values of genetic diversity parameters for individual populations were interpolated using the inverse distance-weighted (IDW) procedure to produce grids of allelic richness and heterozygosity for the sampled range of each species.

Genetic distances among individuals were calculated using two separate measures (i) Bruvo genetic distance (Bruvo et al., 2004) which is based on the stepwise mutation model and is scaled by ploidy level, and (ii) Lynch genetic distance (Lynch, 1990) which is based on the infinite allele model and calculates genetic distances between individuals based on the fraction of bands shared between them. Calculations of both distance measures were performed in POLYSAT software and the resultant matrices were used as input to principal components analysis (PCA) to visualize the genetic relationships among individuals of different populations. The partitioning of genetic variation among populations was examined using standard Analysis of Molecular Variance (AMOVA) (Excoffier et al., 1992) implemented in GENALEX software (Peakall & Smouse, 2006).

Two separate analyses were conducted for each species using Bruvo distances and Lynch distances, respectively. All standard AMOVA calculations were performed with 999 permutations of the data to calculate significance levels. Genetic differentiation among populations was estimated by the measure $\Phi_{PT}$, an $F_{ST}$-analog calculated as part of the AMOVA procedure in GENALEX. Both the Bruvo and Lynch distance measure assume that any two alleles have the potential to be identical or closely related by mutation (Clark & Jasieniuk, 2011). However, for an allopolyploid where alleles may be derived from two or more parental genomes, this assumption does not hold true. It is recommended that the Lynch distance measure only be used for allopolyploids if it is known that isoloci have no alleles in common. The Bruvo distance measure should only be used when alleles from different isoloci do not occur within five mutation steps of one another. The validity of this assumption is not clear within the context of the current study, as no information about the identity of alleles to different isoloci is currently available for *B. papyrifera* and *B. alleghaniensis*. Thus, we also calculated differentiation among populations based on a binary data matrix of the presence/absence of microsatellite allele-size variants using a Bayesian AMOVA procedure implemented in the computer program HICKORY (Holsinger et al., 2002). In contrast to regular AMOVA, the Bayesian-based method implemented in HICKORY does not require the assumption of Hardy-Weinberg equilibrium and is independent of sample size. The program randomly generates allele-phenotypes from a probability distribution based on the observed data to provide estimates of among-population differentiation based on the posterior probability point estimates. The parameter $\theta^{II}$ provided in the output is directly comparable to the $F_{ST}$-analog $\theta$ of Weir & Cockerham (1984). Estimates of the within-

population inbreeding coefficient *f*, a $F_{IS}$-analog, can also be obtained from the model. The program was run using the default parameters for each of the four models: (i) the full model, (ii) the $f = 0$ model, (iii) $\theta = 0$ model, and (iv) the *f*-free model. The deviance information criteria (DIC) and Dbar values provided in the output were used to assess the fit of each of the models to the data; Dbar is a measure of how well the model fits the data, and DIC is a measure of fit that accounts for the number of parameters. Models which produce a better fit are indicated by lower Dbar and DIC values. Genetic differentiation among populations was also assessed based on allele-identity under the infinite allele model (IAM) with the statistic $F_{ST}$ (Weir & Cockerham, 1984), and based on allele-size under the stepwise mutation model (SMM) with the statistic $R_{ST}$ (Slatkin, 1995) calculated in SPAGEDI with significance values obtained from 10000 random permutations of individuals. Estimates of the standard error of the within-population inbreeding coefficient ($F_{IS}$) were obtained by jackknifing over loci. Tests of phylogeographic structure were conducted by evaluating the significance of the observed $R_{ST}$ relative to the expected $R_{ST}$ obtained from random permutations of allele sizes among alleles following Hardy et al. (2003). If the permutated value of $R_{ST}$ is significantly lower than the observed value, this indicates the presence of a phylogeographic structure as alleles tend to be more closely-related within populations than between populations (Hardy & Vekemans, 2002b).

Patterns of spatial genetic structure were examined by testing for the presence of isolation-by-distance within each species using Mantel tests to examine the correlation between geographic and genetic distances between populations. Geographic distances were determined from the input geographic coordinates in GENALEX and genetic

differentiation values among pairs of populations ($\Phi_{PT}$) were estimated via AMOVA based on the input matrix of Bruvo genetic distances among individuals. Mantel tests were conducted in Genealx with 9999 random permutations to determine significance values. The matrix of genetic distances among population pairs ($\Phi_{PT}$) also was used to investigate the genetic relationships among populations through construction of UPGMA dendrograms using MEGA software (Tamura et al., 2011).

As an additional test of genetic structuring within each species, the program STRUCTURE (Pritchard et al., 2000) was used to examine the partitioning of individuals into genetically-differentiated populations; STRUCTURE uses a Bayesian clustering algorithm to assign individuals to *K* populations on the basis of their genotypes in such a way as to minimize Hardy-Weinberg and linkage disequilibrium. The latest version of the program accommodates genetic ambiguity due to partial heterozygosity at codominant loci in polyploids by incorporating a new algorithm which generates full genotypes for each individual based on their partial genotypes (Falush et al., 2007). The model was run with a burn-in of 100000 iterations and 100000 iterations of each MCMC chain each with five independent runs for *K*=1-12. The admixture model was used with no prior population information using the correlated allele frequencies option, since this method has been shown to be more effective at detecting subtle population substructuring (Falush et al., 2003). The solutions were evaluated to determine the true number of subpopulations (K) based on: (i) the estimated Ln probability of the data (ln Pr(X|K)) obtained from structure output, (ii) the rate of change in the posterior probability between successive runs ($\Delta K$) (Evanno et al., 2005) determined using STRUCTURE HARVESTER software (Earl & Bridgett, 2012), and (iii) the average maximum

correlation (avmaxcorr) of Q-matrices between independent runs at a given *K*-value determined in CORRSIEVE software (Campana et al., 2011).

## Results

*Genetic diversity*

The total number of alleles for the eight microsatellite loci ranged from 14 for Bp04 to 23 for Bo.F330, with a total of 126 alleles scored across all loci (Appendix 3). Multilocus averages of standardized allelic richness were similar for *B. alleghaniensis* and *B. papyrifera,* with values of 12.5 and 12.2 respectively (Table 2). In comparison, *B. lenta* demonstrated much lower allelic richness with a multilocus average of 4.1. Levels of gene diversity were also similar in *B. alleghaniensis* (0.78) and *B. papyrifera* (0.77), but were much lower in *B. lenta* (0.38).

Table 2. Total number of alleles ($A_N$), allelic richness after rarefaction ($A_R$), and gene diversity ($H_E$) for eight nuclear microsatellite loci scored for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta*.

| | Locus | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | CD277302 | L1.10 | L2.3 | L5.5 | L7.1a | L5.4 | Bo.F330 | Bp04 | average |
| *B. alleghaniensis* | | | | | | | | | |
| $A_N$ | 12 | 16 | 16 | 10 | 12 | 12 | 21 | 11 | 13.8 |
| $A_R$ | 11.2 | 14.9 | 14.7 | 9.9 | 11.1 | 9.3 | 18.5 | 10.1 | 12.5 |
| $H_E$ | 0.82 | 0.83 | 0.86 | 0.82 | 0.69 | 0.61 | 0.82 | 0.77 | 0.78 |
| *B. papyrifera* | | | | | | | | | |
| $A_N$ | 11 | 10 | 9 | 8 | 20 | 8 | 21 | 13 | 12.5 |
| $A_R$ | 11.0 | 9.8 | 8.5 | 7.8 | 18.9 | 7.8 | 20.7 | 13.0 | 12.2 |
| $H_E$ | 0.79 | 0.84 | 0.56 | 0.71 | 0.90 | 0.63 | 0.86 | 0.86 | 0.77 |
| *B. lenta* | | | | | | | | | |
| $A_N$ | 3 | 2 | 7 | 2 | 7 | 5 | 4 | 3 | 4.1 |
| $A_R$ | 3.0 | 2.0 | 7.0 | 2.0 | 7.0 | 5.0 | 4.0 | 3.0 | 4.1 |
| $H_E$ | 0.11 | 0.48 | 0.52 | 0.26 | 0.66 | 0.66 | 0.29 | 0.06 | 0.38 |

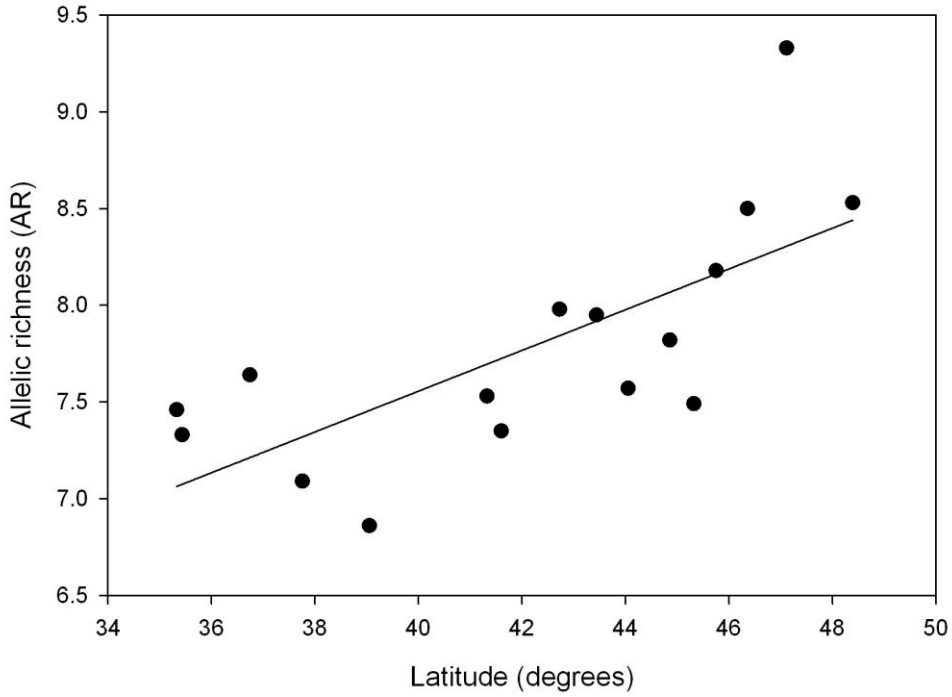
$A_N$, number of alleles; $A_R$, allelic richness after rarefaction (N=300); $H_E$, gene diversity

Markers CD277302, L1.10, L2.3, L5.5, L5.4, and Bp04 demonstrated standard dinucleotide repeat patterns in each of the species, whereas in *B. papyrifera* and *B. alleghaniensis* some alleles of Bo.F330 differed by a single base-pair and some alleles of L7.1a differed by a single base pair in *B. papyrifera*. Due to the single-base pair shifts, genotypes for polyploids were extremely difficult to score at these loci and so they were excluded from subsequent analyses. Thus, for *B. alleghaniensis* analyses were based on seven loci, analyses for *B. papyrifera* were based on six loci, and for *B. lenta* analyses were based on all eight loci. Four loci (CD277302, L5.4 Bo.F330, Bp04) showed significant departures from Hardy-Weinberg equilibrium in the diploid *B. lenta* (Appendix 4). Tests of population genetic bottlenecks for *B. lenta* indicated a significant excess of heterozygosity in comparison to mutation-drift expectations for a single central Appalachian population (103), while the remaining 11 populations did not present significant evidence of recent bottlenecks.

Standardized allelic richness values for individual *B. alleghaniensis* populations ranged from 6.9 to 9.3 and averaged 8.0 across all populations sampled (Appendix 5). Gene diversity values were high but similar across sampled populations, ranging from 0.75 to 0.82 and averaging 0.77 across populations. For populations of *B. papyrifera*, allelic richness values ranged from 6.6 to 7.5 and gene diversity ranged from 0.71 to 0.74 while for *B. lenta* allelic richness ranged from 2.5 to 3.0 and gene diversity ranged from 0.33 to 0.43. A significant heterozygote deficit was observed for four populations of *B. lenta* (103, 121, 122, 125) with distributions scattered across the range. Regressions of allelic richness against latitude indicated that for *B. alleghaniensis* allelic richness increased significantly moving northward ($R^2$=0.5193, P=0.0016), while allelic richness

declined slightly with increasing latitude for *B. papyrifera* ($R^2$=0.0585, P=0.5306) and *B. lenta* ($R^2$=0.2219, P=0.1221) though the relationship was not significant for either species (Figure 2).

(a)



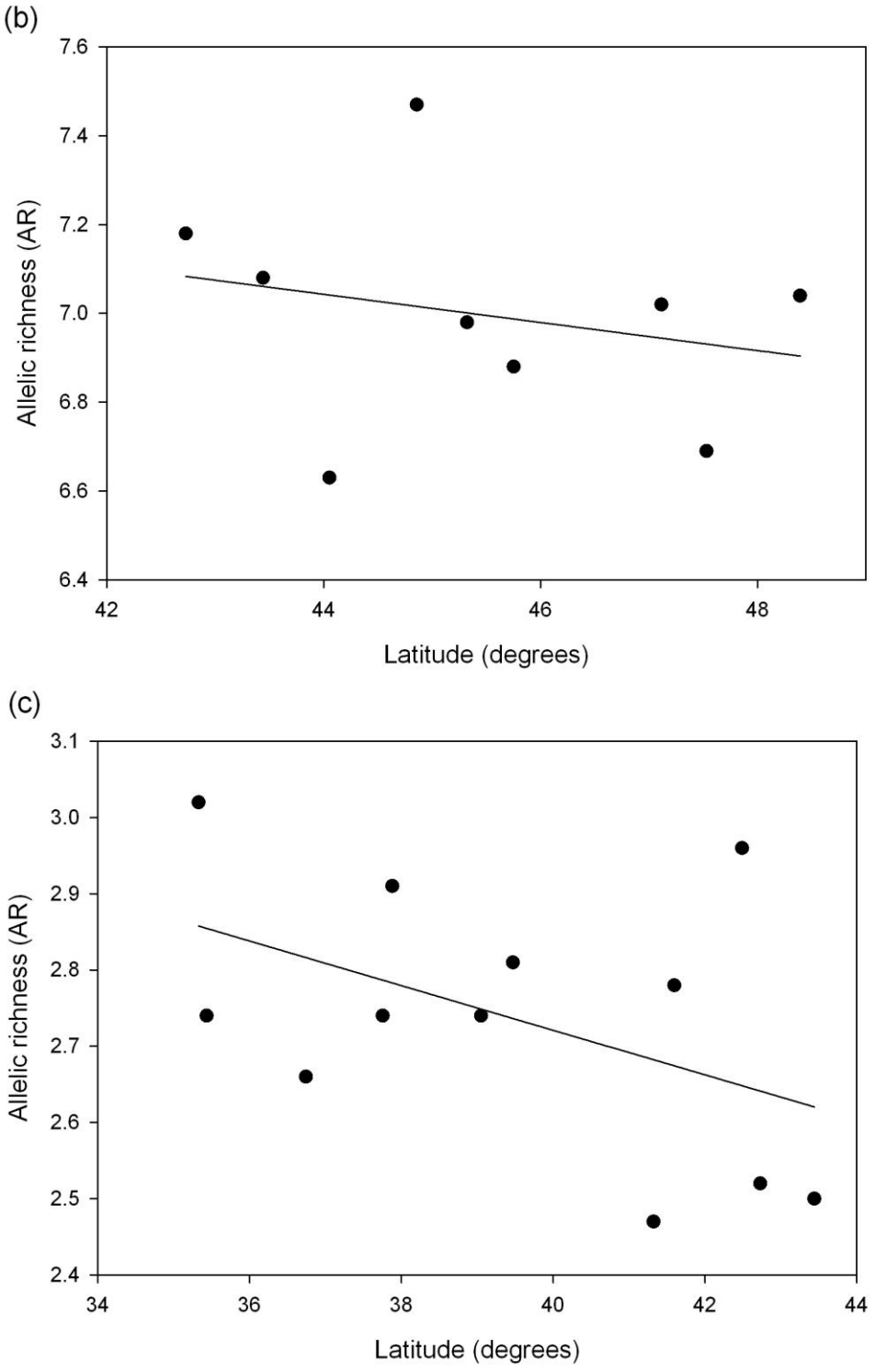



Figure 2. Correlations of standardized allelic richness with latitude for (a) *B. alleghaniensis* ($R^2$=0.5193, P=0.0016), (b) *B. papyrifera* ($R^2$=0.0585, P=0.5306), and (c) *B. lenta* ($R^2$=0.2219, P=0.1221).

Interpolated maps indicated that the highest allelic richness for *B. alleghaniensis* occurred for populations of the western Great Lakes region while southern and central Appalachian populations demonstrated lower values (Figure 3a). In contrast, *B. papyrifera* populations of the western Great Lakes region demonstrated relatively low allelic richness, and populations with the highest allelic richness were located in the eastern portion of the sampled range (Figure 3b). In *B. lenta*, southwestern Appalachian populations had the highest allelic richness and diversity was lowest for the three most northern Appalachian populations (Figure 3c).

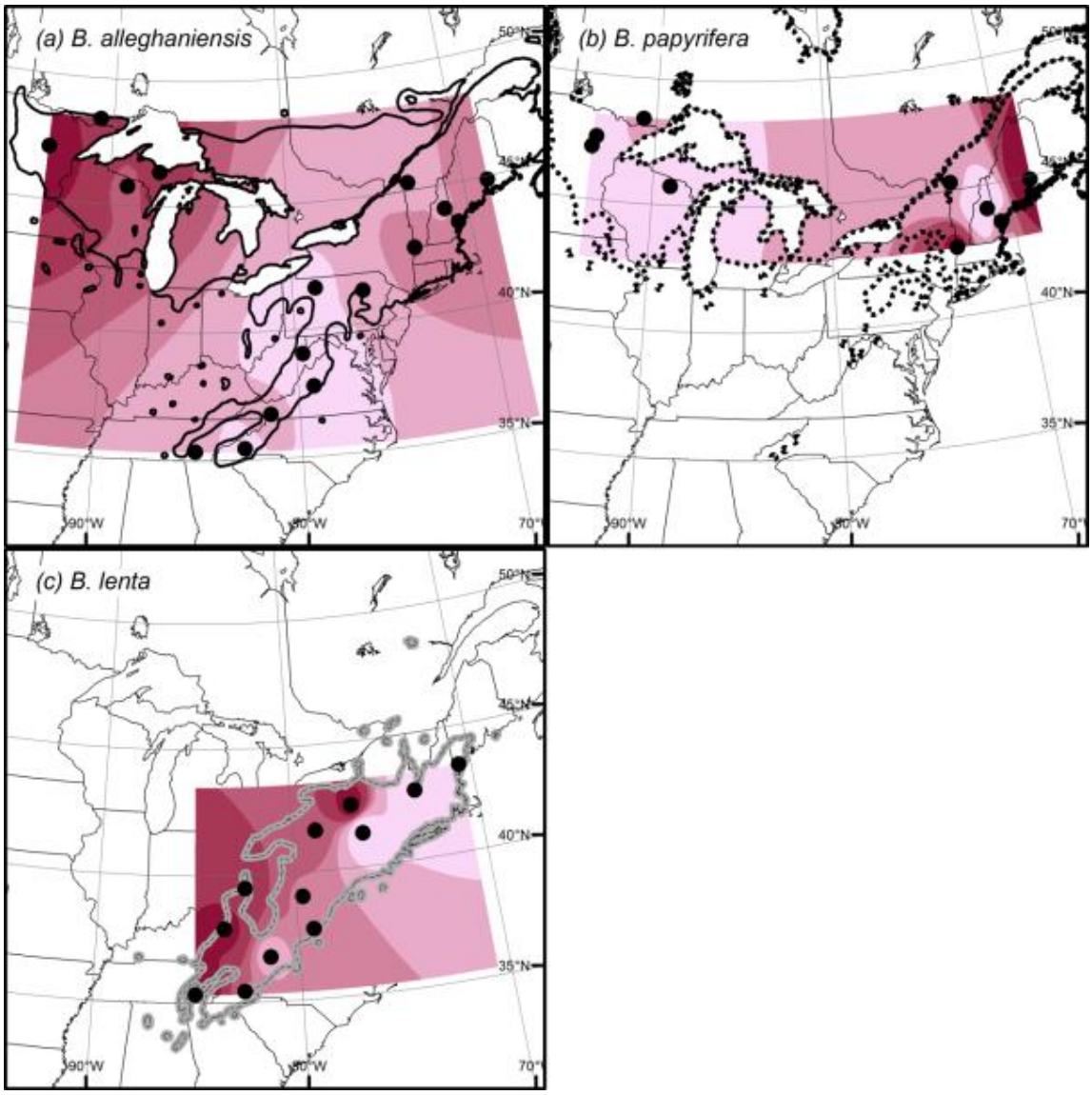


Figure 3. Mapped standardized allelic richness ($A_R$) for the sampled range of (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.
Areas of relatively low allelic richness are indicated by light shading while areas of higher allelic richness are indicated by darker shading.

Similar to allelic richness, gene diversity for *B. alleghaniensis* was highest in the western Great Lakes region (Figure 4a). For *B. papyrifera*, two areas of relatively high gene diversity were observed, with the first centered populations of the western Great Lakes region, and the second centered on a population in the northeastern portion of

thesampled range (Figure 4b). The distribution of gene diversity for *B. lenta* was similar to that of allelic richness, with highest gene diversity occurring in populations of the southwest Appalachians and declining gene diversity moving northward (Figure 4c).

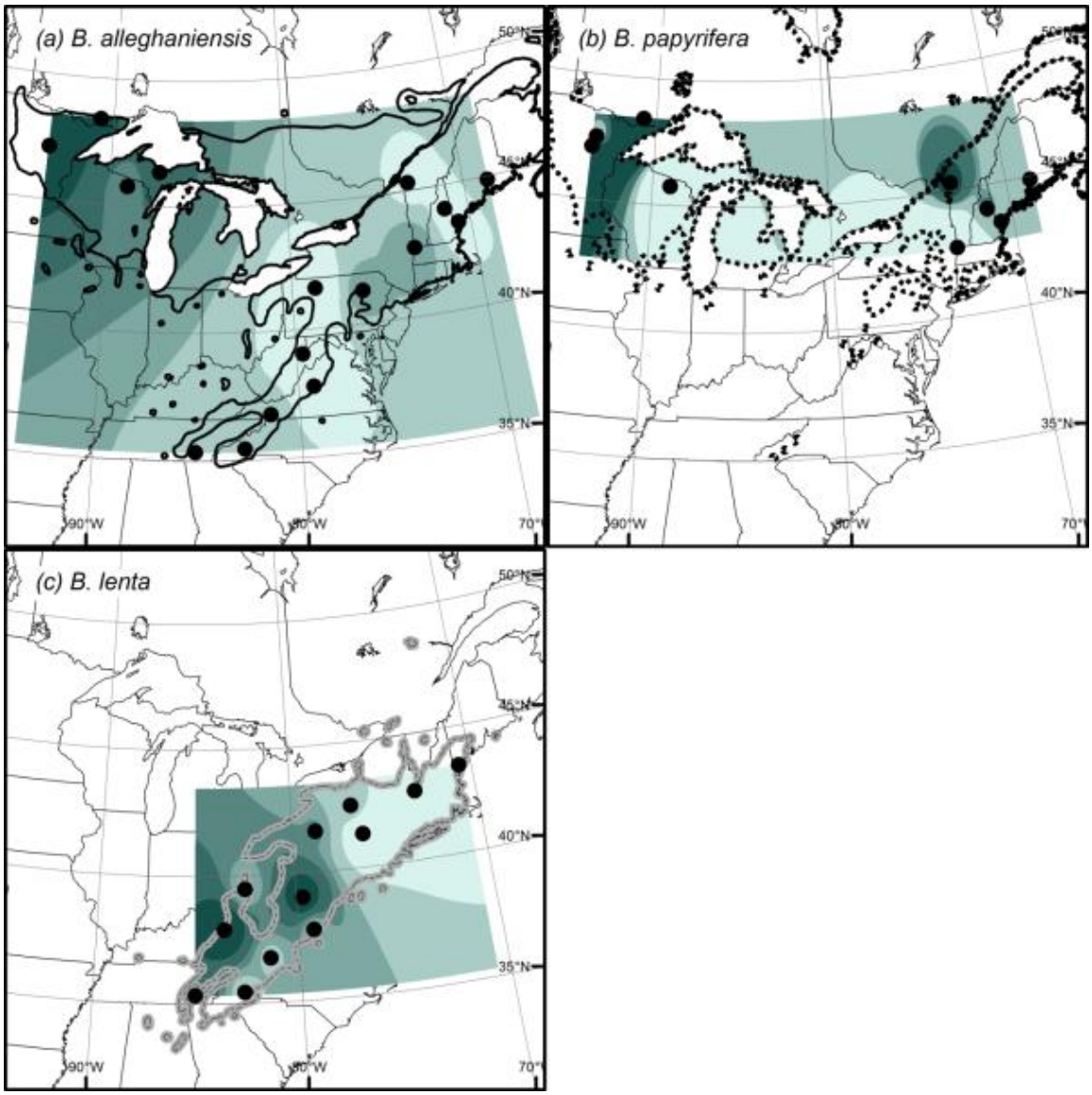
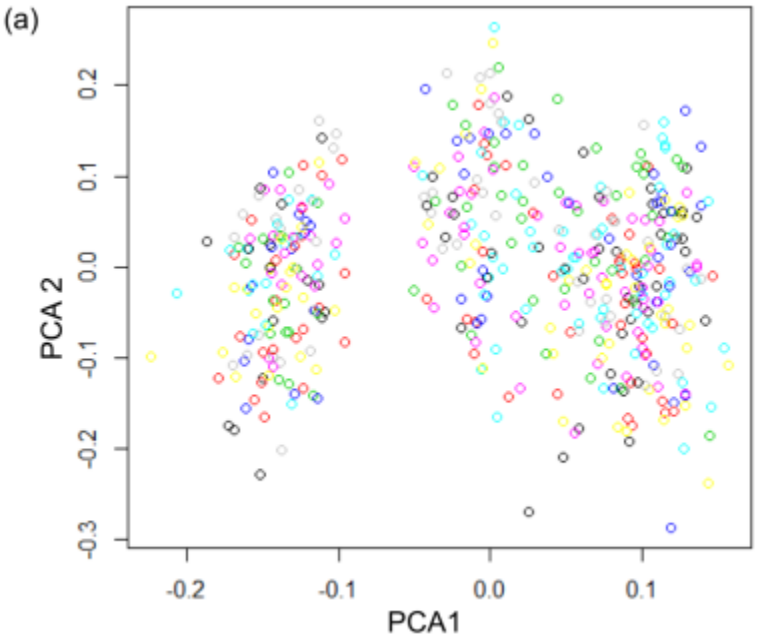



Figure 4. Mapped standardized gene diversity ($H_E$) for the sampled range of (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.
Areas of relatively low gene diversity are indicated by light shading while areas of higher gene diversity are indicated by darker shading.

*Population genetic structure*

Principal component analyses based on Bruvo genetic distances did not reveal strong patterns of genetic differentiation among populations of *B. alleghaniensis*, *B. papyrifera*, or *B. lenta* as there was no clear clustering of individuals from different populations (Figure 5). Similar results were obtained based on Lynch distances, which also did not indicate a clear separation of populations (Appendix 6).



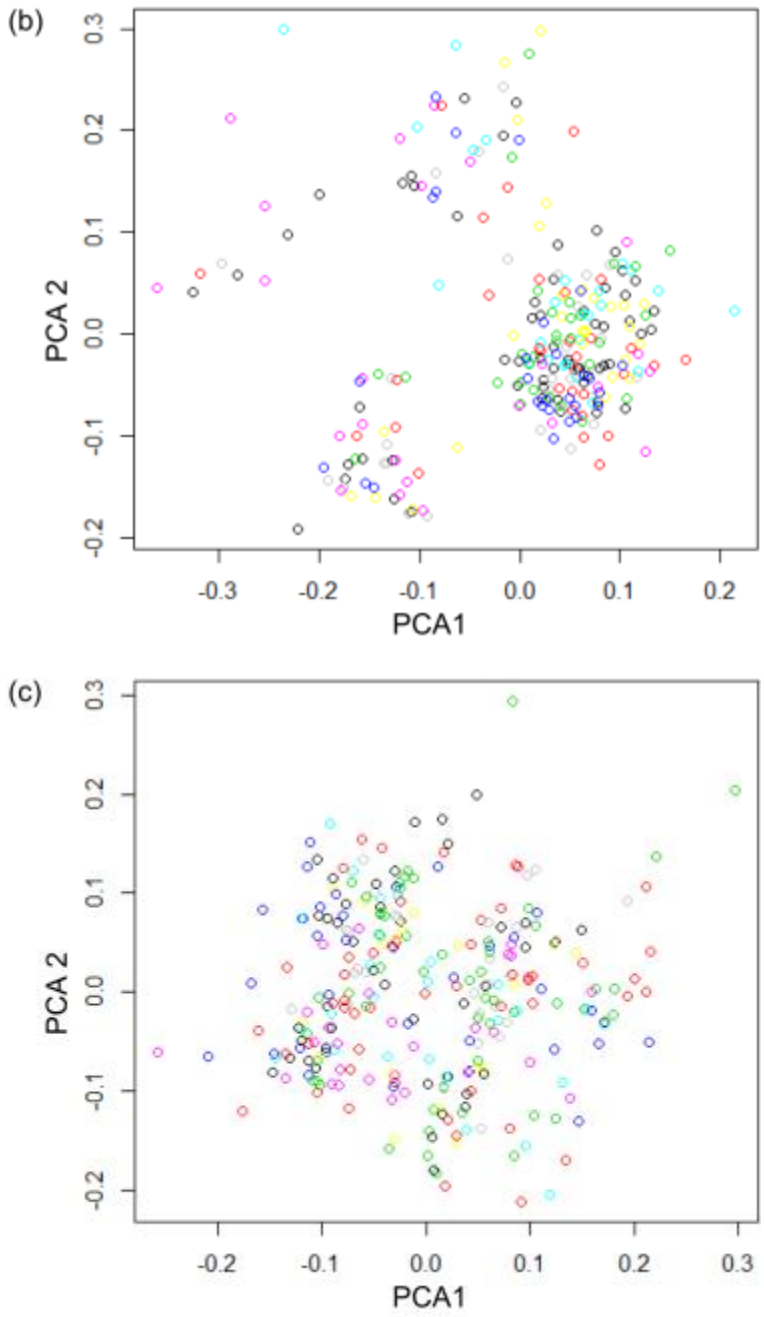



Figure 5. Principal components analysis of Bruvo distance matrices showing the pattern of genetic differentiation among individuals of different populations of (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.
Individuals of the same population are coded with the same colour.

Standard analyses of molecular variance indicated weak but significant partitioning of genetic variation among populations of *B. alleghaniensis* ($\Phi_{PT}$=0.019,

*P*=0.001) *B. papyrifera* ($\Phi_{PT}$=0.028, *P*=0.001) and *B. lenta* ($\Phi_{PT}$=0.022, *P*=0.001) based on Bruvo distances (Table 3). Similar results were obtained based on Lynch distances (Appendix 7).

Table 3. Standard analysis of molecular variance used to examine the partitioning of molecular variation among populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* based on Bruvo genetic distances across nuclear microsatellite loci.

| Species | Source of variation | df | Sum of squares | Variance components | Percent variation | $\Phi_{PT}$ | P |
|---|---|---|---|---|---|---|---|
| *B. alleghaniensis* | Among populations | 15 | 5.40 | 0.004 | 1.9 | 0.019 | 0.001 |
| | Within populations | 467 | 105.50 | 0.226 | 98.1 | | |
| *B. papyrifera* | Among populations | 8 | 2.92 | 0.006 | 2.8 | 0.028 | 0.001 |
| | Within populations | 247 | 49.96 | 0.202 | 97.2 | | |
| *B. lenta* | Among populations | 11 | 2.35 | 0.003 | 2.2 | 0.022 | 0.001 |
| | Within populations | 288 | 39.38 | 0.137 | 97.8 | | |

The Bayesian AMOVA conducted also indicated low levels of among-population differentiation for all species, with an estimated $\theta^{II}$ of 0.015 for *B. alleghaniensis*, 0.023 for *B. papyrifera* and 0.017 for *B. lenta* based on the full model (Table 4). In comparison to the full model, the θ=0 model indicated relatively higher Dbar and DIC values indicating a weaker fit to the data, which suggests that populations were significantly genetically differentiated. For all species, the full model and the *f*=0 model produced a similar fit to the data, suggesting that little inbreeding occurred within the studied populations. In contrast, high inbreeding coefficients were obtained from the full model, but these are probably not accurate due to the low reliability of Hickory in estimating *f*-values for dominant marker data (Holsinger & Lewis, 2007).

Table 4. Summary of Bayesian AMOVA describing the partition of molecular variation among populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* based on the treatment of the presence/absence of microsatellite alleles as a binary data matrix.

| Species | Model | $f$ | $\theta^{II}$ | Dbar | DIC |
|---|---|---|---|---|---|
| *B. alleghaniensis* | full | 0.618±0.284 | 0.015±0.003 | 4251.2 | 4638.3 |
| | $f$=0 | - | 0.010±0.001 | 4251.2 | 4654.3 |
| | θ=0 | 0.981±0.020 | - | 4896.3 | 4977.8 |
| | $f$-free | 0.502±0.287 | 0.031±0.007 | 4241.2 | 4967.8 |
| *B. papyrifera* | full | 0.729±0.269 | 0.023±0.005 | 1682.6 | 1865.1 |
| | $f$=0 | - | 0.015±0.003 | 1683.4 | 1875.7 |
| | θ=0 | 0.965±0.035 | - | 1981.9 | 2033.8 |
| | $f$-free | 0.506±0.289 | 0.041±0.010 | 1686.5 | 1990.2 |
| *B. lenta* | full | 0.413±0.294 | 0.017±0.005 | 1123.2 | 1227.8 |
| | $f$=0 | - | 0.014±0.003 | 1121.8 | 1238.8 |
| | θ=0 | 0.903±0.106 | - | 1299.7 | 1328.4 |
| | $f$-free | 0.505±0.286 | 0.039±0.009 | 1123.9 | 1319.8 |

$f$, within-population inbreeding coefficient; $\theta^{II}$, analog of Weir and Cockerham's θ; Dbar, a measure of how well the model fits the data; DIC, deviance information criteria

Global $F_{ST}$ values for *B. alleghaniensis* (0.008), *B. papyrifera* (0.015), and *B. lenta* (0.013) indicated significant but somewhat lower among-population differentiation than $R_{ST}$ values of 0.011, 0.019, and 0.032, respectively (Table 5). However, the comparison of $R_{ST}$ to $R_{ST}$ permuted yielded a significantly greater value of $R_{ST}$ only for *B. lenta*, indicating the presence of a phylogeographic structure within that species. The highest level of within population inbreeding was observed for *B. lenta* ($F_{IS}$=0.088), while *B. alleghaniensis* indicated only a low level of within-population inbreeding ($F_{IS}$=0.025) and *B. papyrifera* indicated an excess of heterozygotes ($F_{IS}$= -0.133).

Table 5. Global $F_{ST}$ and $R_{ST}$, and tests of $R_{ST}$ vs. $R_{ST}$ permuted based on nuclear microsatellite markers for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta*.

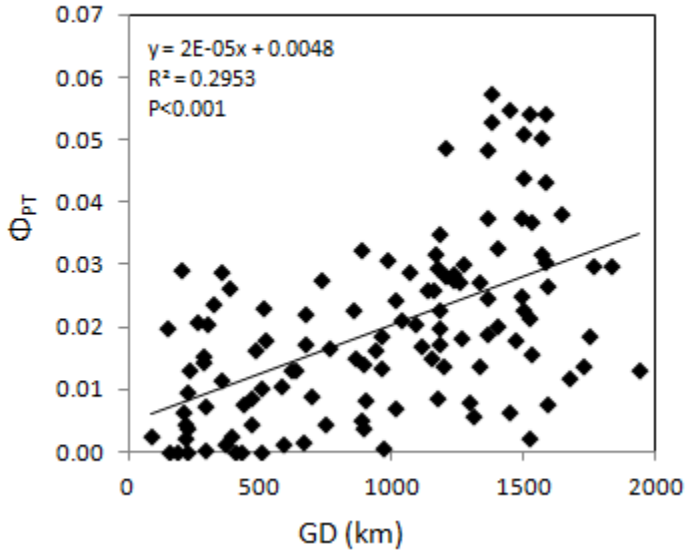
| Species | $F_{IS}$ | $F_{ST}$ | $R_{ST}$ | P($R_{ST}$>$R_{ST}$ permuted) |
|---|---|---|---|---|
| *B. alleghaniensis* | 0.025 (0.090) | 0.008* | 0.011* | 0.1799 |
| *B. papyrifera* | -0.133 (0.017) | 0.015* | 0.019* | 0.3137 |
| *B. lenta* | 0.088 (0.043) | 0.013* | 0.032* | 0.0198* |

Standard error of $F_{IS}$ in brackets, *significant at P<0.05

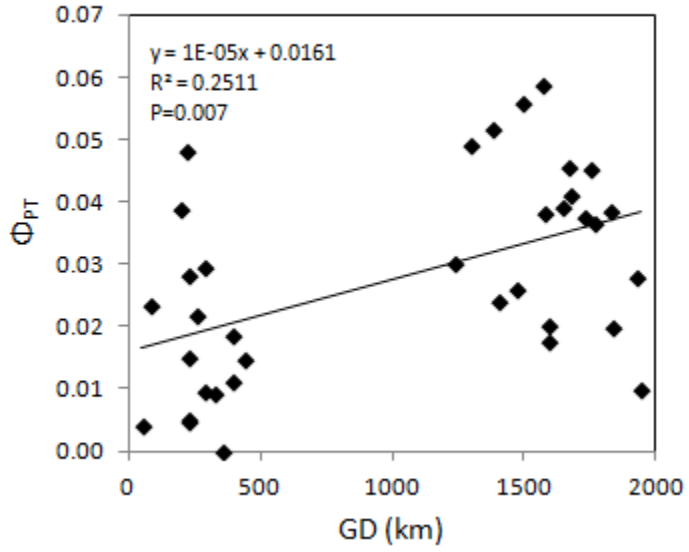
Pairwise population genetic differentiation values were generally low, with $\Phi_{PT}$ values ranging from -0.010 to 0.058 for *B. alleghaniensis*, 0.00 to 0.059 for *B. papyrifera*, and -0.022 to 0.073 for *B. lenta* (Appendix 8). For *B. alleghaniensis*, pairwise comparisons indicated that western populations 105, 106, and 115 were generally significantly differentiated from eastern populations, with the exception of two northeastern populations (101 and 121), which were not differentiated. Two populations, (113 and 1101) were relatively distinct from other populations, as indicated by numerous significantly different pairwise comparisons. Western populations of *B. papyrifera* (106, 113, 114, and 115) were, in general, significantly differentiated from eastern populations, and population 112 in the northeastern study area was divergent from most other populations. In *B. lenta*, there was no pronounced pattern of population differentiation with respect to geographic locality and significant differentiation generally only occurred between populations at extreme northern and southern sites. One northern population of *B. lenta* (121) differed significantly from all other populations with the exception of northern populations 102 and 120. Regressions of genetic distances against geographic distance revealed a significant positive relationship, indicating the presence of isolation-by-distance in each of the three species (Figure 6). The relationship was stronger for *B. alleghaniensis* ($R^2$=0.2953, P<0.001), in comparison to *B. papyrifera* ($R^2$=0.2511,

P=0.007) and *B. lenta* ($R^2$=0.2231, p=0.002). However, genetic distances increased more sharply with increasing geographic distance for *B. lenta* in comparison to the other two species as indicated by the greater slope of the regression line.

***(a) B. alleghaniensis***




***(b) B. papyrifera***



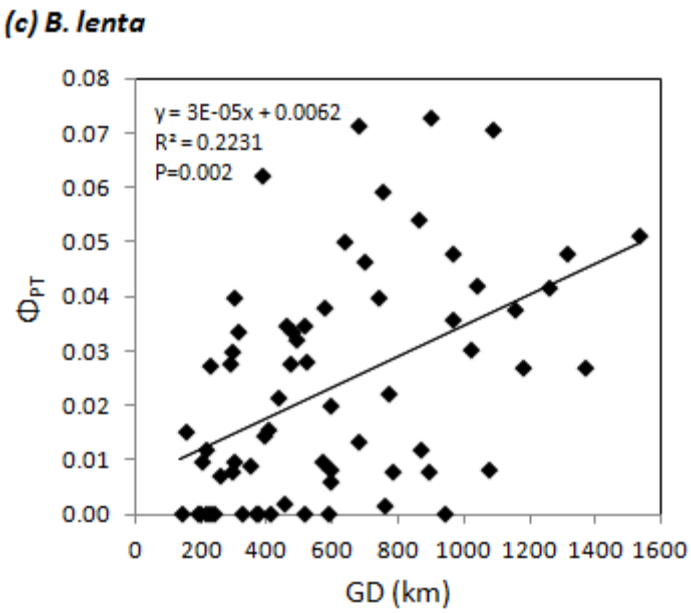
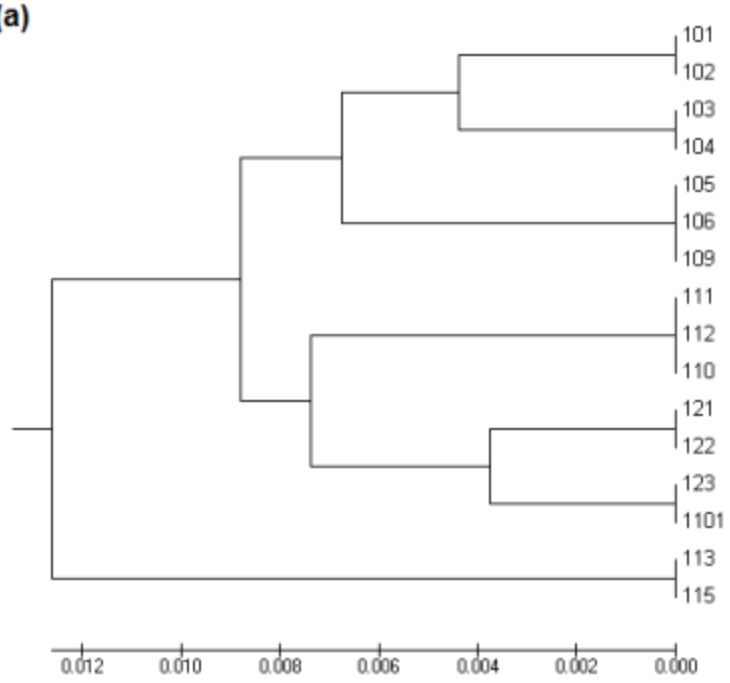


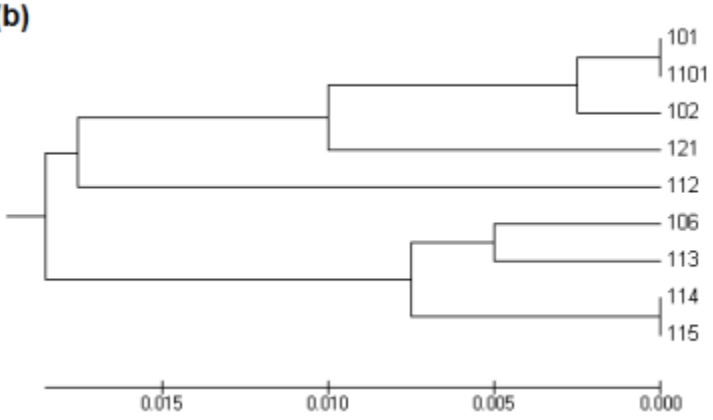

Figure 6. Plots of regressions of genetic distance ($\Phi_{PT}$) against geographic distance (km) between populations of (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.

The UPGMA tree for *B. alleghaniensis* indicated relative divergence of two western populations (113, 115) from the remaining populations (Figure 7). However, there was no clear geographic pattern to the relationships among the remaining populations. For *B. papyrifera*, the UPGMA tree indicated a clear separation differentiation of the western populations (106,113, 114, 115) from eastern populations (101, 1101, 102, 121, 112). Within both eastern and western groups, the grouping of populations was suggestive of a pattern of isolation by distance, with geographically-adjacent populations generally grouping together within the same clade. The UPGMA tree for *B. lenta* was also consistent with a pattern of isolation by distance, as clades were composed of populations that occurred in close geographic proximity to one another. However, the two most northern populations (102, 121) were relatively divergent and grouped separately from more southern populations.

(a)


101
102
103
104
105
106
109
111
112
110
121
122
123
1101
113
115
0.012
0.010
0.008
0.006
0.004
0.002
0.000

(b)


101
1101
102
121
112
106
113
114
115
0.015
0.010
0.005
0.000

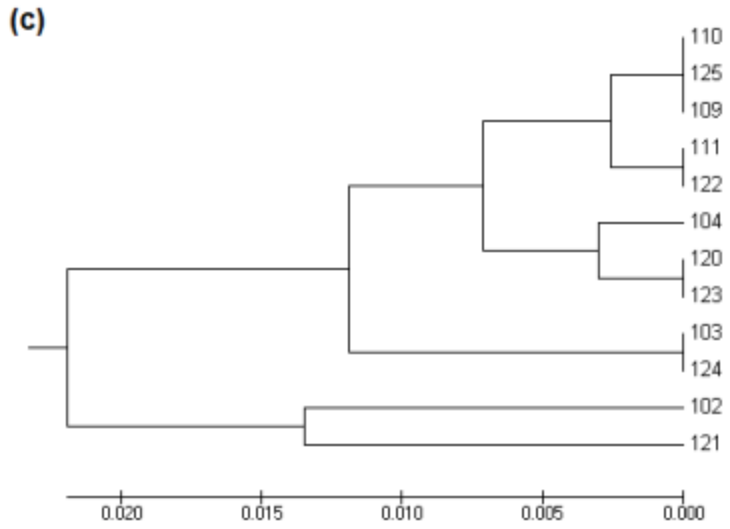



Figure 7. UPGMA trees constructed from pairwise population genetic distances ($\Phi_{PT}$) estimated via AMOVA of individual Bruvo distances for (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.

*Betula alleghaniensis* populations with the highest mean pairwise differentiation values were located in the far south of the range in the southern Appalachians and also in the far west of the range in the Great Lakes region (Figure 8). For *B. papyrifera*, the most divergent populations were relatively randomly distributed throughout the study area and did not demonstrate any clear geographic pattern while for *B. lenta* northern Appalachian populations demonstrated the highest mean differentiation.

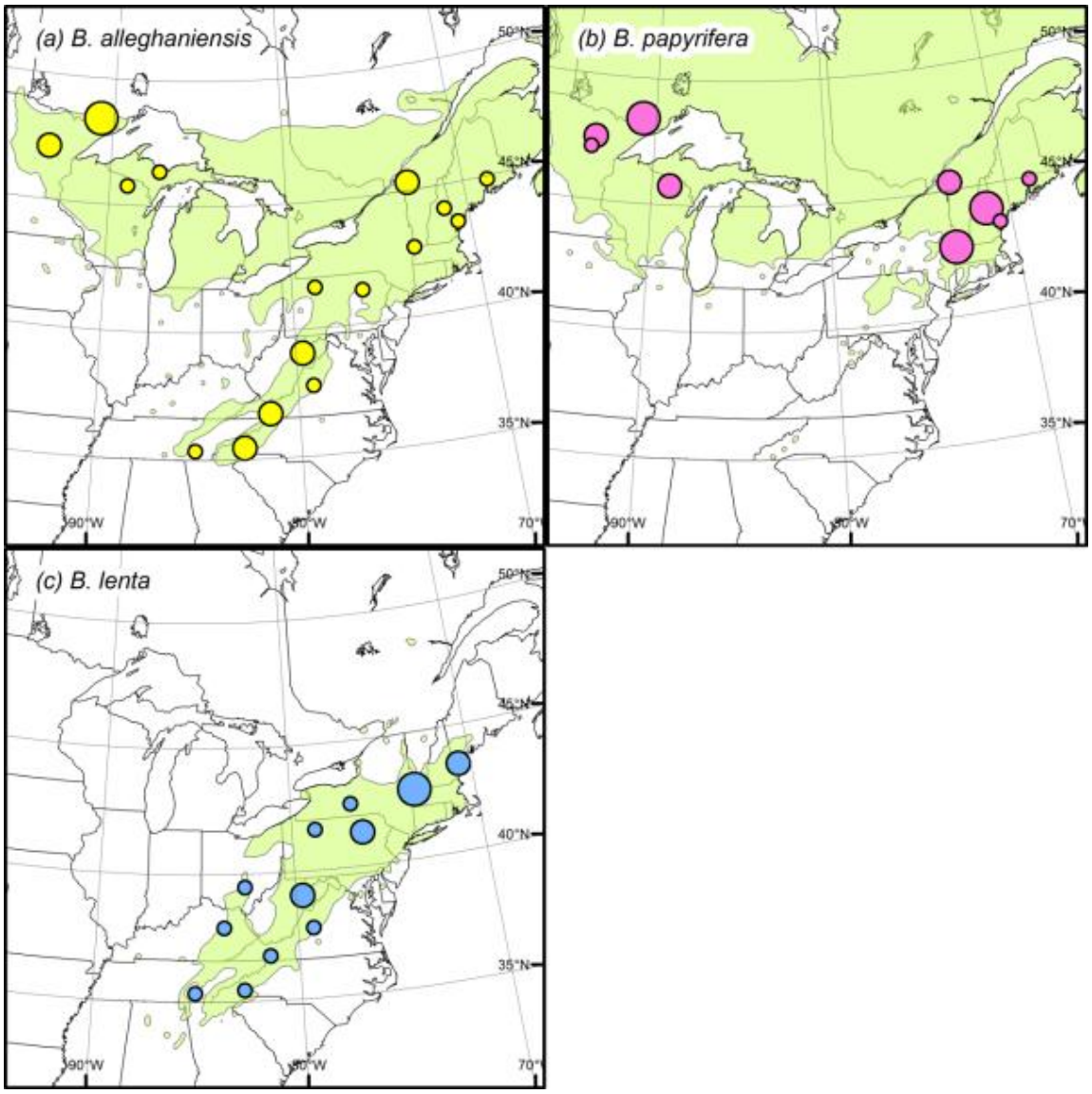



Figure 8. Mean pairwise differentiation for populations of (a) *B. alleghaniensis* ($\Phi_{PT}$ = -0.010 to 0.058), (b) *B. papyrifera* ($\Phi_{PT=}$ 0.00 to 0.059, and (c) *B. lenta* ($\Phi_{PT}$= -0.022 to 0.073)

Circles are scaled proportionally based on the population's mean $\Phi_{PT}$ value, with higher values indicated by larger circles, intermediate values indicated by medium circles, and low values indicated by the smallest circles.

STRUCTURE analysis for *B. alleghaniensis* indicated that the maximum ΔK value was obtained for K=2 (Table 6). However, the assignment of individuals to clusters was not significantly correlated between independent runs at K=2 (avmaxcorr=0.6000, P>0.05), suggesting that the solution was unstable (Campana et al., 2011). Also, at K=2,

individuals were assigned in roughly equal proportions to cluster 1 (0.502) and cluster 2 (0.498), which is an indication that there is no true population structure (Pritchard et al., 2007). Thus, the optimal assignment of individuals to populations occurred at K=1 for *B. alleghaniensis*.

Table 6. Summary of STRUCTURE output and parameters used to infer the number of population groups (K) for *B. alleghaniensis*.

| K | No. Runs | Mean LnPr(X\|K) | ΔK | avmaxcorr |
|---|---|---|---|---|
| 1 | 5 | -30354 | NA | 1.000 |
| 2 | 5 | -29199 | 9.2 | 0.600 |
| 3 | 5 | -28854 | 0.9 | 0.935 |
| 4 | 5 | -27837 | 5.7 | 0.982 |
| 5 | 5 | -27300 | 1.9 | 0.908 |
| 6 | 5 | -26880 | 0.7 | 0.933 |
| 7 | 5 | -26491 | 4.8 | 0.983 |
| 8 | 5 | -26231 | 0.1 | 0.995* |
| 9 | 5 | -25963 | 1.4 | 0.958 |
| 10 | 5 | -25765 | 0.3 | 0.953 |
| 11 | 5 | -25582 | 1.1 | 0.977 |
| 12 | 5 | -25450 | NA | 0.969 |

avmaxcorr, average maximum correlation; *, significant at r >=0.99, P<0.05)

STRUCTURE runs indicated increasing LnPr(X | K) values for *B. papyrifera* between K=1 and K=12, and the maximum ΔK value occurred for K=2 (Table 7). The proportion of individuals assigned to the two populations was roughly symmetric with 50.1 % of individuals assigned to cluster 1 and 49.9% assigned to cluster 2. This was an indication that the inferred structure at K=2 was not real and that populations of *B. papyrifera* were not genetically structured. Thus, K=2 was rejected in favour of K=1 for *B. papyrifera*.

Table 7. Summary of STRUCTURE output and parameters used to infer the number of population groups (K) for *B. papyrifera*.

| K | No. Runs | Mean LnPr(X\|K) | ΔK | avmaxcorr |
|---|---|---|---|---|
| 1 | 5 | -10528 | NA | 1.000 |
| 2 | 5 | -10146 | 193.7 | 0.999* |
| 3 | 5 | -9965 | 3.2 | 0.999* |
| 4 | 5 | -9829 | 8.1 | 0.999* |
| 5 | 5 | -9737 | 1.2 | 0.981 |
| 6 | 5 | -9687 | 3.9 | 0.999* |
| 7 | 5 | -9594 | 1.3 | 0.998* |
| 8 | 5 | -9534 | 0.3 | 0.991* |
| 9 | 5 | -9490 | 0.5 | 0.996* |
| 10 | 5 | -9455 | 0.4 | 0.997* |
| 11 | 5 | -9414 | 0.6 | 0.999* |
| 12 | 5 | -9388 | NA | 0.999* |

avmaxcorr, average maximum correlation; *, significant at r >=0.99, P<0.05)

For *B. lenta*, the maximum posterior probability of the data occurred for K=1 and the maximum ΔK occurred at K=2 (Table **8**). At K=2, individuals were assigned in roughly equal proportions to cluster 1 (0.499) and cluster 2 (0.501) and the assignment of individuals to clusters was not stable between independent runs at K=2 (avmaxcorr=0.527, P>0.05). Thus, K=2 was rejected in favour of K=1, suggesting that there existed no true population genetic structure for *B. lenta*.

Table 8. Summary of STRUCTURE output and parameters used to infer the number of population groups (*K*) for *B. lenta*.

| K | No. Runs | Mean LnPr(X\|K) | ΔK | avmaxcorr |
|---|---|---|---|---|
| 1 | 5 | -3331 | NA | 1.000 |
| 2 | 5 | -3526 | 1.7 | 0.527 |
| 3 | 5 | -3545 | 0.8 | 0.998* |
| 4 | 5 | -3584 | 0.6 | 0.999* |
| 5 | 5 | -3658 | 0.8 | 0.999* |
| 6 | 5 | -3656 | 1.5 | 0.999* |
| 7 | 5 | -3721 | 0.2 | 0.997* |
| 8 | 5 | -3809 | 1.2 | 0.995* |
| 9 | 5 | -3843 | 0.2 | 0.988 |
| 10 | 5 | -3859 | 0.4 | 0.980 |
| 11 | 5 | -3906 | 0.7 | 0.977 |
| 12 | 5 | -3884 | NA | 0.965 |

avmaxcorr, average maximum correlation; *, significant at r >=0.99, P<0.05)

**Discussion**

*Patterns of genetic variation*

The high levels of gene diversity of *B. alleghaniensis* (0.78) and *B. papyrifera* (0.77) are consistent with values reported for several other broad-scale studies of North American trees (Victory et al., 2006; O'Connell et al., 2008; Ross-Davis et al., 2008; Potter et al., 2012). Studies of the eastern temperate angiosperms black walnut (*Juglans nigra*) (Victory et al., 2006) and butternut (*Juglans cinerea*) (Ross-Davis et al., 2008) revealed gene diversity values of 0.79 and 0.76, respectively for populations collected throughout the species` natural ranges. Gene diversities of conifer species red cedar (*Thuja plicata*) (O'Connell et al., 2008) and eastern hemlock (*Tsuga canadensis*) (Potter et al., 2012) were also similar. This result is in keeping with the finding of generally high levels of heterozygosity for long-lived woody perennials with broad geographic distribution, wind-dispersed seeds, and predominantly outcrossed mating system

(Hamrick et al., 1992), such as exhibited by *B. alleghaniensis* and *B. papyrifera*. In contrast, the gene diversity of 0.38 found for *B. lenta* is surprisingly low, and should not be expected given that *B. lenta* possesses life history traits similar to *B. papyrifera* and *B. alleghaniensis*. Relatively higher expected heterozygosity (0.51) was reported even for the relatively depauperate red pine (*Pinus resinosa*), which is suggested to have suffered extreme population bottlenecks (Boys et al., 2005). A previous study based on cpDNA found evidence for only a single haplotype for *B. lenta* which was interpreted as evidence of population bottlenecks during the last glacial maximum (Chapter 1). However, based on the eight microsatellite markers used in this study, tests of recent population bottlenecks were significant only for a single central Appalachian population. It is possible that *B. lenta* has suffered recent population genetic bottlenecks, but not enough loci have been surveyed to have sufficient power of detection, given that the test suffers reduced power with a smaller number of loci and low heterozygosity (Cornuet & Luikart, 1996). In this study, tests of heterozygosity excess for *B. lenta* may have suffered from reduced power due to the near-fixation of alleles at four of the eight markers employed. Levels of observed heterozygosity were 0.097, 0.264, 0.163, and 0.048 for loci CD277302, L5.5, Bo.F330, and Bp04 respectively. While it is possible that inclusion of more highly polymorphic loci could have yielded greater power, initial screens of more than 30 microsatellite loci revealed generally low levels of polymorphism (results not shown). It has been shown that heterozygosity-excess tests are sometimes unable to detect significant population bottlenecks even when a large number of loci are surveyed and severe recent demographic bottlenecks are known to have occurred; for example BOTTLENECK software provided ambiguous results for tests of demographic

bottlenecks in the Scandinavian lynx (*Lynx lynx*) which was hunted to the brink of extinction between the early 19th and 20th century (Spong & Hellborg, 2002). Further, BOTTLENECK software did not clearly reveal demographic bottlenecks in the Antarctic seal (*Arctocephalus gazella*) which was severely overexploited during the 18th and 19th centuries (Hoffman et al., 2011). Thus, low levels of genetic variation in *B. lenta* could be explained by the occurrence of a strong historic population bottleneck if the heterozygosity-excess tests implemented in BOTTLENECK lacked sufficient power of detection. Alternatively, it is possible that population bottlenecks of *B. lenta* occurred sufficiently long ago that mutation-drift equilibrium has been re-established and bottlenecks are no longer detectable.

Significant heterozygote deficiencies were found at four of the eight loci scored for *B. lenta* (CD277302, L5.4 Bo.F330, Bp04), resulting in significant departures from HW equilibrium. Screening of individual loci using MICRO-CHECKER software (Van Oosterhout et al., 2004) also revealed heterozygote deficiencies for numerous populations, which could indicate the presence of null alleles at these loci. It is possible that the occurrence of null alleles could be partially responsible for the observed heterozygote deficiencies at one or more loci, given that the microsatellite primers used in this study were cross-amplified from congeners *B. pendula* (Kulju et al., 2004), *B. platyphylla* (Wu et al., 2002), *B. pubescens* (Truong et al., 2005), and *B. maximowicziana* (Tsuda et al., 2008), rather than specifically designed for *B. lenta*. If the primers contained significant mismatches with the target regions in *B. lenta*, then some alleles might not have amplified efficiently, leading to null alleles and an observed heterozygote deficiency. However, as mentioned previously, screening of over 30 primer pairs

generally revealed fixation or very low levels of heterozygosity at most loci for *B. lenta*. This, in combination with the low levels of cpDNA polymorphism found for *B. lenta* (Chapter 1), suggests that the low levels of heterozygosity observed in this study are biologically representative for *B. lenta*.

The observed high levels of gene diversity of *B. alleghaniensis* and *B. papyrifera* might partially be explained in terms of their allopolyploid origins, as polyploid species are expected to demonstrate higher heterozygosity than diploid congenerics (Soltis & Soltis, 2000). Since allopolyploids form from the combination of the genomes of two different parental species, they are expected to demonstrate fixed heterozygosity. Furthermore, most polyploids are not formed by a single polyploidization event, but rather from multiple recurrent formations. Recurrent formation serves to increase diversity of polyploids by capturing additional variation present within parental taxa. When assessed based on eight microsatellite loci in the current study, *B. papyrifera* and *B. alleghaniensis* demonstrated similar levels of allelic richness. However, higher levels of allelic richness were found for *B. papyrifera* than for *B. alleghaniensis* based on a previous study of cpDNA (Chapter 1). This inconsistency may be explained in that the previous study also included populations from the extreme western distribution range of *B. papyrifera* in British Columbia, whereas the present study includes only populations from the sympatric distribution with *B. alleghaniensis*. It is probable that higher levels of allelic richness would have been found for B. *papyrifera* if populations from across the full extent of the natural range had been included.

Northward declines in allelic richness are expected to occur for species which existed primarily within southern refugia during the last glacial maximum, as successive

founder events during postglacial migration along the northward migration path should be associated with a series of population bottlenecks (Hewitt, 2000). Based on cpDNA (Chapter 1), and fossil evidence (Jackson et al., 1997, 2000), *B. lenta* is suspected to have occupied a single southern refugia in the Appalachian mountains at the time of the last glacial maximum and is thought to have colonized sequentially northward through the Appalachian mountains to the present extent of its range in Maine, Vermont, and New Hampshire. The pattern of declining allelic richness in northern populations of *B. lenta* supports the theory of sequential population genetic bottlenecks during the northward migration of this species. Similar patterns of declining allelic richness with latitude have been demonstrated for a number of other temperate trees including *Quercus garryana* (Marsico et al., 2009) and *Thuja plicata* (O'Connell et al., 2008) which possess relatively narrow longitudinal but long latitudinal ranges and were explained by the presence of founding events during postglacial migration. In contrast, no significant declines in allelic richness with latitude were detected for either *B. alleghaniensis* or *B. papyrifera* in the current study. Rather, the highest allelic richness and heterozygosity for *B. alleghaniensis* was found in the western Great Lakes Region. For *B. papyrifera* the highest allelic richness occurred for two populations in the east of the study area (101, 102) while the highest gene diversity was found for populations 113 and 115 of the western Great Lakes area. These patterns coincide well with a previous study of *B. alleghaniensis* and *B. papyrifera* which found high cpDNA diversity for the same areas resulting from secondary contact between eastern, western, and Atlantic lineages (Chapter 1), and it seems likely that mixing of glacial lineages also is responsible for the high nuclear microsatellite diversity of these areas in the current study. Studies of the European shrub-

birch *Betula humilis* also revealed similar patterns of genetic diversity based on cpDNA (Jadwiszczak et al., 2012) and nuclear microsatellite markers (Jadwiszczak et al., 2011), with both chloroplast and nuclear markers revealing high diversity in northeastern Poland resulting from contact between different haplotype lineages.

*Patterns of population structure*

Weak patterns of among-population variation were recovered for *B. alleghaniensis* ($\Phi_{PT}$=0.019), *B. papyrifera* ($\Phi_{PT}$=0.028), and *B. lenta* ($\Phi_{PT}$=0.022). While low levels of among-population variation are expected based on other studies of long-lived woody perennials, the values reported in this study are lower than the mean value (0.086) for allozyme studies of long-lived perennials with wind-dispersed seeds (Hamrick & Godt, 1996). The among-population differentiation values found in this study also are lower than the values reported from several nuclear microsatellite studies of boreal and temperate trees, including: *Quercus garryana* ($F_{ST}$=0.049) (Marsico et al., 2009), *Fraxinus excelsior* ($F_{ST}$=0.076) (Heuertz et al., 2004b), *Pinus resinosa* ($F_{ST}$=0.28) (Boys et al., 2005), *Tsuga canadensis* ($F_{ST}$=0.077) (Potter et al., 2012), and *Thuja plicata* ($F_{ST}$=0.070) (O'Connell et al., 2008). Higher levels of differentiation were also found for Chinese populations of *Betula alnoides* ($\Phi_{ST}$=0.086) (Zeng et al., 2003) and *B. luminifera* (Xie et al., 2009) ($G_{ST}$=0.3486) based on RAPD markers. Few population genetics studies based on nuclear microsatellites currently exist for eastern North American angiosperms. Of the studies which are available, levels of population differentiation for *Juglans nigra* ($F_{ST}$=0.017) (Victory et al., 2006), *Juglans cinerea* ($F_{ST}$=0.025) (Ross-Davis et al., 2008), and *Quercus macrocarpa* ($F_{ST}$=0.027) (Craft & Ashley, 2007) were

similar to those determined for *Betula* species in the current study. For those species, the lack of distinct geographical population groupings was attributed to high rates of pollen flow over large spatial scales, which served to homogenize genetic variation among populations. For birches in this study, low levels of population genetic subdivision are probably also the result of relatively high rates of pollen flow across relatively large geographic distances as pollen studies in Europe have shown that birch pollen may be transported over distances of greater than 1000 km (Siljamo et al., 2008). *Betula papyrifera* and *B. alleghaniensis* both demonstrated strong population genetic substructure at cpDNA markers (*alleghaniensis* $G_{ST}$=0.601; *papyrifera* $G_{ST}$=0.526) (Chapter 1). Comparing these values to the $\Phi_{PT}$ values obtained based on nuclear microsatellites, the pollen-to-seed migration ratio calculated according to Ennos (1994) is approximately 76 for *B. alleghaniensis* and 37 for *B. papyrifera,* indicating relatively greater pollen dispersal than seed dispersal. While the pollen-to-seed migration ratio cannot be calculated for *B. lenta* due to its lack of structure at cpDNA markers ($G_{ST}$=0.00), its low population differentiation at nuclear markers is also suggestive of relatively high rates of pollen flow. In studies of *Juglans* (Victory et al., 2006; Ross-Davis et al., 2008) and *Quercus* (Craft & Ashley, 2007; Marsico et al., 2009), the relative importance of pollen flow is emphasized strongly due to the relatively limited dispersal ability of their large, heavy seeds which are dispersed primarily by animals. In contrast, seeds of *Betula* are light, small, and may be dispersed large distances by wind, especially over open snow-covered areas (Safford et al., 1990; Matlack, 1992). The relatively high pollen-to-seed migration ratios of 120 reported for *Quercus garryana* (Marsico et al., 2009), 500 for *Q. petraea* and 286 for *Q. robur* (ElMousadik & Petit, 1996) probably

reflect the limited dispersal of their heavier seeds. In comparison, lower pollen-to-seed migration ratios for *Betula* indicate that their relatively efficient seed dispersal may play an important role in maintaining gene flow among populations.

Significant patterns of isolation were found for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta,* though populations were generally significantly differentiated only over broad spatial scales, reflecting the relative efficiency of seed and pollen in maintaining gene flow among neighbouring populations. The Bayesian analysis of population genetic substructure indicated that the optimal assignment of individuals to populations occurred for K=1 for each of the species, since the assignments at K=2 were not biologically meaningful. This result is consistent with what should be expected under a pattern of isolation by distance whereby populations do not demonstrate genetic discontinuities but vary clinally across the species' range. While no clear pattern of genetic subdivision was observed, *B. papyrifera* and *B. alleghaniensis* populations from the western Great Lakes region were generally significantly differentiated from populations in the eastern portion of the range, and northern Appalachian populations of *B. lenta* were generally differentiated from central and southern populations. The differentiation of western populations of *B. alleghaniensis* and *B. papyrifera* probably reflects differences in allele frequencies contributed to by their relatively higher genetic diversity, including both high allelic richness and heterozygosity for *B. alleghaniensis*, and high heterozygosity for *B. papyrifera* as well as relative isolation from eastern populations due to the large geographic distances between them. Also, eastern and western populations appear to have originated from separate glacial refugia (Chapter 1), and at least some part of the variation at nuclear markers may be explained in terms of the historic long-term isolation

of these populations. Thus, the observed patterns are probably best explained by a combination of historic and contemporary population dynamics; for *B. alleghaniensis* and *B. papyrifera*, separation of eastern and western populations during the last glacial maximum would have lead to significant genetic divergence between them. Following the retreat of the glaciers, the populations from the eastern (Appalachian) refugia would have colonized first northwards then westwards where they came into contact with northward-expanding populations from the Great Lakes area resulting in the observed high levels of cpDNA structure (Chapter 1). Subsequent and extensive pollen flow among newly-established populations should then have helped to increase gene flow and roughly homogenize allele frequencies among populations, resulting in the weak pattern of genetic structure observed at nuclear markers. However, while the genetic similarity of populations implies that gene flow via pollen and seed is effective mostly in preventing population divergence, the observed pattern of isolation by distance reflects that populations are not panmictic and populations at large distances from one another may not exchange genes to a large extent. In *B. lenta*, the finding of just a single chloroplast haplotype throughout its distribution (Chapter 1) reflects that the observed genetic homogeneity among populations probably originated before the time of postglacial recolonization. However, subsequent reductions in effective population size associated with postglacial recolonization also seem to have occurred as evidenced by the lower genetic diversity of northern populations. In contrast, the relatively high levels of microsatellite variation observed across the range of *B. papyrifera* and *B. alleghaniensis* suggest that they maintained fairly large effective population sizes throughout the recolonization.

It has been hypothesized that peripheral populations should demonstrate relatively greater genetic divergence and lower genetic diversity than core populations as a result of lower gene flow and lower effective population size (Eckert et al., 2008). Since peripheral populations may often occupy sub-optimal habitat, populations at the range periphery are expected to occur in relatively low densities, resulting in relatively high rates of genetic drift in comparison to core populations. Furthermore, spatial isolation may reduce gene flow to peripheral populations. In the current study, we found evidence from *B. alleghaniensis* that peripheral populations were more highly differentiated on average than core populations, but these populations did not clearly demonstrate any reduced genetic diversity in comparison to core populations. In fact, diversity was highest in the western periphery of the *B. alleghaniensis* range, where genetic differentiation values were also high. This suggests that while peripheral populations of *B. alleghaniensis* might be relatively divergent from core populations due to spatial separation, peripheral populations still seem to maintain large enough effective population sizes that few alleles are lost to genetic drift. For *B. lenta*, northern populations clearly demonstrate greater genetic divergence and lower genetic diversity than core and southern populations. Initially, it is difficult to determine whether this pattern reflects contemporary low effective population sizes of northern populations, or whether it may have originated during postglacial recolonization as previously discussed. However, given the evidence for relatively high rates of pollen flow which are suggested by the lack of population genetic subdivision, and the lack of strong evidence for inbreeding in most northern populations, it seems likely that northern populations maintain large contemporary effective population sizes and reasonable rates of gene flow

with neighbouring populations. For *B. papyrifera* no clear trend with respect to core versus peripheral populations was observed, probably because the sampled populations did not cover a sufficiently large portion of the range for patterns to become clear.

*Conservation implications*

At present, none of the species examined in this study are at immediate conservation risk. The species occur across relatively broad geographic areas and are generally common across their ranges with large population sizes. However, studies of future habitat distribution under climate change indicate future declines in area and importance of these species across portions of their current range (Iverson & Prasad, 2002). Genetic diversity is the raw material for adaptation (Allendorf & Luikart, 2007). As such, considerable attention has been given recently to the importance of the preservation of genetic diversity in adaptation to climate change (Hampe & Petit, 2005; Pautasso, 2009; Loss et al., 2011; Keppel et al., 2012; Kremer et al., 2012). If neutral markers show correlations to levels of genome-wide diversity, they may be useful in identifying populations with high adaptive potential which can serve as source populations for *in situ* or *ex situ* conservation and reforestation efforts, or alternatively populations with low adaptive potential which may be at high risk of extinction (Petit et al., 1998). If levels of microsatellite variation are reflective of genome wide variation, then *B. lenta* might not demonstrate sufficient genetic variation to be able to adequately adapt to future changes in the climate. Given the very low genetic variation at cpDNA markers (Chapter 1) and low nuclear microsatellite variation in the current study, it seems prudent that *B. lenta* should be identified as a potential species of concern for

conservation under climate change. In contrast, the high levels of microsatellite variation indicated for *B. alleghaniensis* and *B. papyrifera* indicate that they potentially could be quite resilient to changing climate, as they could demonstrate high adaptive potential. Furthermore, the high range-wide variation for *B. alleghaniensis* and *B. papyrifera* indicates that they have probably maintained high effective population sizes during past postglacial warming periods, and may respond with similar resilience during future warming.

If *in situ* or *ex situ* conservation efforts were to be identified as a priority for these species, the population genetic patterns observed in this study could be informative. Based on the very low-among population differentiation values, 98.1%, 97.2%, and 97.8% of the total genetic variation could be represented for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* respectively by preservation of just a single population. This may indicate that relatively low sampling effort could be required for effective *ex situ* gene conservation programs for these species. For *B. alleghaniensis*, conservation efforts should potentially focus on western Great Lakes populations where allelic richness was at a maximum. In addition, populations from eastern Quebec and New England had somewhat elevated microsatellite diversity. This pattern also was reflected at cpDNA markers (Chapter 1) which indicated a mixing of Appalachian and Atlantic haplotype lineages within this area. Thus, sampling of populations from this area should obtain representation of high microsatellite diversity as well as cpDNA genes which are characteristic of both the Atlantic and Appalachian refugial populations. While the current study focused on only a limited portion of the total geographic distribution of *B. papyrifera*, it also provides valuable information for the conservation of this species. For

example, we found high microsatellite allelic richness in the Maritime region which represents an area of mixing haplotypes from the Appalachian and Atlantic lineages (Chapter 1). It is also recommended that Great Lakes *B. papyrifera* populations be afforded conservation priority, since this region contains high cpDNA haplotype diversity (Chapter 1), For *B. lenta*, southern populations should be given priority due to their high microsatellite allelic richness and also because these populations are the most likely to be at risk under climate change and may provide genotypes which are adapted to higher temperatures (Aitken et al., 2008). Northern populations should also be represented since these populations could provide the source material for northward-migrations under climate change.

It should be noted that the above conservation recommendations rely on the assumption that neutral genetic variation measured at microsatellite markers is positively correlated with levels of genome-wide or heritable adaptive genetic variation, and that this assumption is subject to much controversy (Bonin et al., 2007). While some studies have demonstrated positive correlations between neutral genetic diversity and genome-wide variation (Ryynänen et al., 2007; Väli et al., 2008), a recent meta-analysis suggests that neutral variation is generally a weak indicator of species' adaptive potential (Reed & Frankham, 2001). However, studies of molecular variation can be useful in setting conservation guidelines for species for which little information on quantitative trait variation exists. The information presented herein could be exploited by forest managers to set preliminary conservation guidelines given the current lack of information on quantitative genetic variation for these species. However, future studies using quantitative genetic markers will be required to confirm whether *Betula* populations identified for

conservation on the basis of neutral genetic variation in this study also possess high adaptive potential under climate change.

**Conclusions**

The overall low levels of population genetic structure and lack of population genetic subdivision observed for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* in the current study are typical of what would be expected for species with high outcrossing rates and efficient dispersal of both pollen and seeds. High levels of nuclear microsatellite diversity observed for both *B. alleghaniensis* and *B. papyrifera* probably reflect high current and historic effective population sizes with high contemporary gene flow, and their polyploid origins, which contribute to high heterozygosity. For *B. lenta*, the finding of low microsatellite diversity is in contrast to what would be expected for a widespread species with wind-dispersed seeds, but might be explained in terms of bottlenecks before or during the time of the last glacial maximum. Subsequent reductions in effective population size during postglacial recolonization are evident from the northward pattern of declining allelic richness. While not specifically designed for conservation purposes, the current study indicates that *B. lenta* should probably be designated as a species of high priority for conservation under climate change, while *B. alleghaniensis* and *B. papyrifera* may possess a high level of resilience.

## References


Aitken S.N., Yeaman S., Holliday J.A., Wang T., & Curtis-McLane S. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Allendorf F.W. & Luikart G.H. (2007) *Conservation and the Genetics of Populations.* Blackwell Publishing, Malden, MA.

Andreakis N., Kooistra W.H.C., & Procaccini G. (2009) High genetic diversity and connectivity in the polyploid invasive seaweed *Asparagopsis taxiformis* (Bonnemaisoniales) in the Mediterranean , explored with microsatellite alleles and multilocus genotypes. *Molecular Ecology*, 212–226.

Assoumane A., Zoubeirou A.M., Rodier-Goud M., Favreau B., Bezançon G., & Verhaegen D. (2012) Highlighting the occurrence of tetraploidy in *Acacia senegal* (L.) Willd. and genetic variation patterns in its natural range revealed by DNA microsatellite markers. *Tree Genetics & Genomes*, **9**, 93–106.

Bonin A., Nicole F., Pompanon F., Miaud C., & Taberlet P. (2007) Population adaptive index: a new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation biology : the journal of the Society for Conservation Biology*, **21**, 697–708.

Bousquet J., Cheliak W.M., & Lalonde M. (1988) Allozyme variation within and among mature populations of speckled alder (*Alnus rugosa*) and relationships with green alder (*A. crispa*). *Botanical Society of America*, **75**, 1678–1686.

Boys J., Cherry M., & Dayanandan S. (2005) Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany*, **92**, 833–841.

Bruvo R., Michiels N., D'Souza T., & Schulenburg H. (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, **13**, 2101–2106.

Campana M.G., Hunt H. V, Jones H., & White J. (2011) CorrSieve: software for summarizing and evaluating Structure output. *Molecular Ecology Resources*, 349–352.

Clark L. V & Jasieniuk M. (2011) POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources*, **11**, 562–566.

Clark L. V., Evans K.J., & Jasieniuk M. (2013) Origins and distribution of invasive *Rubus fruticosus* L. agg. (*Rosaceae*) clones in the Western United States. *Biological Invasions*, **6**, 1331–1342.

Cornuet J.M. & Luikart G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Craft K.J. & Ashley M.V. (2007) Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. *Forest Ecology and Management*, **239**, 13–20.

Earl D.A. & Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources*, **4**, 359–361.

Eckert C.G., Samis K.E., & Lougheed S.C. (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology*, **17**, 1170–1188.

ElMousadik A. & Petit R.J. (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547–555.

Ennos R.A. (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Erdmann G.G. (1990) Yellow birch (*Betula alleghaniensis* Britton). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 133–147. U.S. Department of Agriculture Forest Service, Washington, DC.

Esselink D.G., Nybom H., & Vosman B. (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting- peak ratios) method. *Theoretical and Applied Genetics*, **109**, 402–408.

Evanno G., Regnaut S., & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L., Smouse P.E., & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D., Stephens M., & Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Falush D., Stephens M., & Pritchard J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fins L. & Seeb L.W. (1986) Genetic variation in allozymes of western larch. *Canadian Journal of Forest Research*, **16**, 1013–1018.

Fowler D.P. & Morris R.W. (1977) Genetic diversity in red pine: evidence for low genic heterozygosity. *Canadian Journal of Forest Research*, **7**, 343–347.

Godbout J., Jaramillo-Correa J.P., Beaulieu J., & Bousquet J. (2005) A mitochondrial DNA minisatellite reveals the postglacial history of jack pine (*Pinus banksiana*), a broad-range North American conifer. *Molecular Ecology*, **14**, 3497–3512.

Goudet J. (1995) FSTAT (Version 1.2) A computer program to calculate F-statistics. *The Journal of Heredity*, **86**, 485–486.

Hampe A. & Petit R.J. (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

Hamrick J.L. & Godt M.J.W. (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transacations: Biological Sciences*, **351**, 1291–1298.

Hamrick J.L., Godt M.J.W., & Sherman-Broyles S.L. (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Hardy O.J., Charbonnel N., Fréville H., & Heuertz M. (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

Hardy O.J. & Vekemans X. (2002a) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Hardy O.J. & Vekemans X. (2002b) SPAGeDi 1.4: a program for spatial pattern analysis of genetic diversity. 2002.

Harestad A.S. & Keisker D.G. (1989) Nest tree use by primary cavity-nesting birds in south central British Columbia. *Canadian Journal of Zoology*, **67**, 1067–1073.

Helsen P., Verdyck P., Tye A., & VanDongen S. (2009) Low levels of genetic differentiation between *Opuntia echios* varieties on Santa Cruz (Galapagos). *Plant Systematics and Evolution*, **279**, 1–10.

Heuertz M., Hausman J.-F., Hardy O.J., Vendramin G.G., Frascaria-Lacoste N., & Vekemens X. (2004) Nuclear microsatellites reveal contrasting patterns of genetic

structure between western and southeastern European populations of the common ash (Fraxinus excelsior L.). *Evolution*, **58**, 976–988.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hoffman J.I., Grant S.M., Forcada J., & Phillips C.D. (2011) Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology*, **20**, 3989–4008.

Holsinger K.E. & Lewis P.O. (2007) *Hickory: A Package for Analysis of Population Genetic Data v1.1*.

Holsinger K.E., Lewis P.O., & Dey D.K. (2002) A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, **11**, 1157–1164.

Iverson L.R. & Prasad A.M. (2002) Potential redistribution of tree species habitat under five climate change scenarios in the eastern US. *Forest Ecology and Management*, **155**, 205–222.

Iverson L.R., Prasad A.M., & Matthews S. (2007) Modeling potential climate change impacts on the trees of the northeastern United States. *Mitigation and Adaptation Strategies for Global Change*, **13**, 487–516.

Iverson L.R., Prasad A.M., Matthews S.N., & Peters M. (2008) Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management*, **254**, 390–406.

Iverson L.R., Schwartz M.W., & Prasad A.M. (2004) How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology and Biogeography*, **13**, 209–219.

Jackson S.T., Overpeck J.T., Webb-III T., Keattch S.E., & Anderson K.H. (1997) Mapped plant-macrofossil and pollen records of late Quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**, 1–70.

Jackson S.T., Webb R.S., Anderson K.H., Overpeck J.T., Webb-III T., Williams J.W., & Hansen B.C.S. (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Science*, **19**, 489–508.

Jadwiszczak K. a., Banaszek A., Jabłońska E., & Sozinov O. V. (2011) Could *Betula humilis* have survived the last glaciation at a current margin of its distribution? Testing the hypothesis of a glacial refugium using nuclear microsatellites. *Plant Systematics and Evolution*, **297**, 147–156.

Jadwiszczak K. a., Banaszek A., Jabłońska E., & Sozinov O. V. (2012) Chloroplast DNA variation of *Betula humilis* Schrk. in Poland and Belarus. *Tree Genetics & Genomes*, **8**, 1017–1030.

Kaiser J. (1993) American birch: a furniture favourite. *Wood and Wood Products*, **98**, 29–29.

Kaiser J. (2008) A bundle of beauty: yellow birch yields an array of colors: red, white and two-toned veneers. *Wood and Wood Products*, **113**, 27–28.

Keppel G., Van Niel K.P., Wardell-Johnson G.W., Yates C.J., Byrne M., Mucina L., Schut A.G.T., Hopper S.D., & Franklin S.E. (2012) Refugia: identifying and understanding safe havens for biodiversity under climate change. *Global Ecology and Biogeography*, **21**, 393–404.

Kremer A., Ronce O., Robledo-Arnuncio J.J., Guillaume F., Bohrer G., Nathan R., Bridle J.R., Gomulkiewicz R., Klein E.K., Ritland K., Kuparinen A., Gerber S., & Schueler S. (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, **15**, 378–392.

Kulju K.K.M., Pekkinen M., & Varvio S. (2004) Twenty-three microsatellite primer pairs for *Betula pendula* (Betulaceae). *Molecular Ecology Notes*, **4**, 471–473.

Lamson N.I. (1990) Sweet birch (*Betula lenta* L.). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 148–152. United States Department of Agriculture Forest Service., Washington, DC.

Lemieux M.J., Beaulieu J., & Bousquet J. (2011) Chloroplast DNA polymorphisms in eastern hemlock: range-wide genogeographic analyses and implications for gene conservation. *Canadian Journal of Forest Research*, **41**, 1047–1059.

Li P. & Adams W.T. (1989) Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). *Canadian Journal of Forest Research*, **19**, 149–161.

Loss S.R., Terwilliger L.A., & Peterson A.C. (2011) Assisted colonization: Integrating conservation strategies in the face of climate change. *Biological Conservation*, **144**, 92–100.

Lynch M. (1990) The similarity index and DNA fingerprinting. *Molecular Biology and Evolution*, **7**, 478–484.

Marsico T.D., Hellmann J.J., & Romero-Severson J. (2009) Patterns of seed dispersal and pollen flow in *Quercus garryana* (Fagaceae) following post-glacial climatic changes. *Journal of Biogeography*, **36**, 929–941.

Matlack G.R. (1992) Influence of fruit size and weight on wind dispersal in *Betula lenta*, a gap-colonizing tree species. *American Midland Naturalist*, **128**, 30–39.

Mimura M. & Aitken S.N. (2007) Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity*, **99**, 224–32.

Mosseler A., Innes D.J., & Roberts B.A. (1991) Lack of allozyme variation in disjunct Newfoundland populations of red pine (*Pinus resinosa*). *Canadian Journal of Forest Research*, **21**, 525–528.

Münzbergová Z., Surinová M., & Castro S. (2013) Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity*, **110**, 123–130.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

O'Connell L.M., Ritland K., & Thompson S.L. (2008) Patterns of post-glacial colonization by western redcedar (*Thuja plicata*, Cupressaceae) as revealed by microsatellite markers. *Botany*, **86**, 194–203.

Obbard D.J., Harris S.A., & Pannell J.R. (2006) Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity*, 296–303.

Oddou-Muratorio S., Petit R.J., Le Guerroue B., Guesnet D., & Demesure B. (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123–1135.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M., & Shipley P. (2004) Micro-Checker: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data. *Molecular Ecology Notes*, **4**, 535–538.

Pautasso M. (2009) Geographical genetics and the conservation of forest trees. *Perspectives in Plant Ecology, Evolution and Systematics*, **11**, 157–189.

Peakall R. & Smouse P.E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Petit R.J., El Mousadik A., & Pons O. (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology*, **12**, 844–855.

Piry S., Luikart G., & Cornuet J.M. (1999) BOTTLENECK: A computer program for detecting recent reductions in effective population size using allele frequency data. *The Journal of Heredity*, **90**, 502–503.

Potter K.M., Jetton R.M., Dvorak W.S., Hipkins V.D., Rhea R., & Whittier W.A. (2012) Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer. *Conservation Genetics*, **13**, 475–498.

Pritchard J.K., Stephens M., & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pritchard J.K., Wen X., & Falush D. (2007) Documentation for structure software : Version 2.2. 36.

Reed D.H. & Frankham R. (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution; international journal of organic evolution*, **55**, 1095–103.

Robertson A., Rich T.C.G., Allen A.M., Houston L., Roberts C., Bridle J.R., Harris S.A., & Hiscock S.J. (2010) Hybridization and polyploidy as drivers of continuing evolution and speciation in *Sorbus*. *Molecular Ecology*, **19**, 1675–1690.

Rodzen J.A., Famula T.R., & May B. (2004) Estimation of parentage and relatedness in the polyploid white sturgeon *(Acipenser transmontanus*) using a dominant marker approach for duplicated microsatellite loci. *Aquaculture*, **232**, 165–182.

Ross-Davis A., Ostry M., & Woeste K.E. (2008) Genetic diversity of butternut (*Juglans cinerea*) and implications for conservation. *Canadian Journal of Forest Research*, **38**, 899–907.

Rustad L., Campbell J., Dukes J.S., Huntington T., Lambert K.F., Mohan J., & Rodenhouse N. (2012) *Changing Climate, Changing Forests: The Impacts of Climate Change on Forests of the Northeastern United States and Eastern Canada. Gen. Tech. Rep. NRS-99.* U.S. Department of Agriculture Forest Service, Northeast Research Station, Newton Square, PA.

Ryynänen H.J., Tonteri A., Vasemägi A., & Primmer C.R. (2007) A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (Salmo salar). *The Journal of heredity*, **98**, 692–704.

Saeki I., Dick C.W., Barnes B.V., & Murakami N. (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): impacts of habitat specialization, hybridization and glacial history. *Journal of Biogeography*, **38**, 992–1005.

Safford L.O., Bjorkbom J.C., & Zasada J.C. (1990) Paper birch (*Betula papyrifera* Marsh.). *Silvics of North America. Agriculture Handbook 654.* (ed. by R.M. Burns

and B.H. Honkola), pp. 158–171. United States Department of Agriculture, Washington, DC.

Samadi S., Mavarez J., Pointier J.P., Delay B., & Jarne P. (1999) Microsatellite and morphological analysis of population structure in the parthenogenetic freshwater snail *Melanoides tuberculata*: insights into the creation of clonal variability. *Molecular Ecology*, **8**, 1141–1153.

Sampson J.F. & Byrne M. (2012) Genetic diversity and multiple origins of polyploid *Atriplex nummularia* Lindl. (Chenopodiaceae). *Biological Journal of the Linnean Society*, **105**, 218–230.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 1–2.

Siljamo P., Sofiev M., Severova E., Ranta H., Kukkonen J., Polevova S., Kubin E., & Minin A. (2008) Sources, impact and exchange of early-spring birch pollen in the Moscow region and Finland. *Aerobiologia*, **24**, 211–230.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Soltis P.S. & Soltis D.E. (2000) The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7051–7057.

Speiser R. & Bosakowski T. (1987) Nest site selection by northern goshawks in northern New Jersey and southeastern New York. *The Condor*, **89**, 387–394.

Spong G. & Hellborg L. (2002) A Near-extinction Event in Lynx: Do microsatellite data tell the tale? *Conservation Ecology*, **6**, 1–8.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Truong C., Palme a. E., Felber F., & Naciri-Graven Y. (2005) Isolation and characterization of microsatellite markers in the tetraploid birch, *Betula pubescens* ssp. *tortuosa*. *Molecular Ecology Notes*, **5**, 96–98.

Tsuda Y., Ueno S., Ide Y., & Tsumura Y. (2008) Development of 14 EST-SSRs for *Betula maximowicziana* and their applicability to related species. *Conservation Genetics*, **10**, 661–664.

Väli U., Einarsson A., Waits L., & Ellegren H. (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular ecology*, **17**, 3808–17.

Victory E.R., Glaubitz J.C., Rhodes O.E., & Woeste K.E. (2006) Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *American Journal of Botany*, **93**, 118–126.

Weir B.S. & Cockerham C.C. (1984) Estimating F-Statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wu B., Lian C., & Hogetsu T. (2002) Development of microsatellite markers in white birch (*Betula platyphylla* var. *japonica*). *Molecular Ecology Notes*, **2**, 413–415.

Xie Y., Li Z., Huang R., Xiao X., & Huang Y. (2009) Genetic diversity of *Betula luminifera* populations at different elevations in Wuyi Mountain and its association with ecological factors. *Front. For. China*, **4**, 90–95.

Zeng J., Zou Y., Bai J., & Zheng H. (2003) RAPD analysis of genetic variation in natural populations of *Betula alnoides* from Guangxi, China. *Euphytica*, **134**, 33–41.

Zeng J., Zou Y.-P., Bai J.-Y., & Zheng H.-S. (2002) Preparation of total DNA from "recalcitrant plant taxa". *Acta Botanica Sinica*, **44**, 694–697.

## Chapter 3: Introgression and genetic relationships of eastern North American birches *Betula alleghaniensis*, *B. papyrifera* and *B. lenta* as revealed by nuclear microsatellites


### Abstract

Although natural hybridization among eastern North American birches *Betula alleghaniensis, B. papyrifera*, and *B. lenta* has long been suspected, no molecular genetic studies have assessed the incidence of hybridization and introgression among these species. We examined the patterns of allele-sharing at six nuclear microsatellite markers to assess introgression and the genetic distinctiveness of these species. We used Bayesian model-based clustering to identify putatively admixed individuals. The mapping of admixture proportions allowed us to examine if geographic patterns of allele-sharing were consistent with a scenario of introgression. Despite moderate to high levels of allele-sharing (*papyrifera/lenta*-22.2%, *alleghaniensis/lenta*-25.9%, *alleghaniensis*/*papyrifera*-68.1%), three examined species were significantly genetically differentiated ($\Phi_{PT}$=0.403, $F_{ST}$=0.2365, $R_{ST}$=0.3493). A number of individuals of *B. alleghaniensis* were identified as putatively admixed with *B. lenta*. However, the geographic pattern was more consistent with allele-sharing due to ancestral polymorphism than what would be expected under a scenario of introgression. Evidence for admixture between *B. alleghaniensis* and *B. papyrifera* was found in the western Great Lakes area, which based on previous morphological studies is a suspected zone of natural hybridization. The cause of the high levels of allele-sharing between *B. papyrifera* and *B. alleghaniensis* could not be definitively determined, but could be attributable to low levels of contemporary introgression, ancestral polymorphism, historic introgression, or a combination of these factors.



Key Words: *Betula*, birches, eastern North America, genetic relationships, hybridization, introgression, allele-sharing

## Introduction

The study of hybridization and genetic introgression has long-fascinated evolutionary researchers because of its relationship to plant diversification and speciation. Recent studies suggest that natural hybridization among closely-related plant species may occur much more frequently than previously thought, and recent estimates indicate that hybridization occurs in at least 25% of all plant species (Mallet, 2005). Natural hybridization followed by backcrossing to parental taxa can lead to genetic introgression, where a portion of the genome of one parental species may become incorporated into the other (Rieseberg & Ellstrand, 1993). Introgression may benefit one or more of the species involved by increasing genetic diversity and/or through introgression of fitness-related traits (Rieseberg, 1997). From an evolutionary standpoint, perhaps one of the most interesting outcomes of hybridization and introgression is its contribution to speciation. When hybridization is followed by a whole-genome duplication (WGD) event, it can lead to rapid speciation through reproductive isolation of the newly-formed polyploid from its progenitor species (Rieseberg & Willis, 2007). Alternatively, hybridization can, in some cases, result in the production of fertile hybrids which are not genetically isolated from the parents and repeated hybridization and backcrossing can lead to substantial introgression (Oberprieler et al., 2010). If introgression is substantial it can eventually lead to breakdown of species boundaries, a process which essentially results in 'speciation-reversal' (Seehausen et al., 2008). As such, hybridization and introgression with introduced species poses a significant extinction risk to many rare or endemic native plants as it may lead to relatively rapid species breakdown (Wolf et al., 2001). Alternatively, environmental selection (Dodd &

Afzal-Rafii, 2004) and preferential within-species mating (Lepais & Gerber, 2011) maintain morphological and molecular differentiation among parental species despite recurrent hybrid formation.

Numerous studies of temperate forest trees using cpDNA markers have revealed widespread evidence of genetic introgression among closely-related species; for example *Acer* (Saeki et al., 2011), *Fraxinus* (Heuertz et al., 2006), and *Quercus* (Dumolin-Lapègue et al., 1997). In fact, chloroplast capture through hybridization and introgression appears to be a widespread occurrence throughout a diverse array of plant taxa (Rieseberg & Soltis, 1991; Rieseberg, 1995). However, for many species introgression at cpDNA markers is not necessarily associated with introgression at nuclear loci; for example, geographical structuring of chloroplast haplotypes suggestive of introgression has been found in European *Alnus* (King & Ferris, 2000) and eastern North American *Quercus* (Whittemore & Schaal, 1991) though nuclear genes did not appear to be exchanged freely. In many cases, it appears that cpDNA introgression may occur more readily than nuclear DNA introgression (Rieseberg & Soltis, 1991; Rieseberg et al., 1991). Relatively high rates of cytoplasmic introgression in comparison to nuclear introgression are considered to be responsible for much of the incongruence between cpDNA and nuclear DNA-based phylogenies that is commonly observed (Rieseberg & Soltis, 1991); for example, as in hawkweeds (Fehrer et al. 2007). Studies of introgression based on cpDNA markers are inherently limited by the non-recombinant nature of the chloroplast, which makes it possible to recover only one genealogy (Muir and Schlotterer 2005). Also, the slow rate of mutation of the chloroplast means that the ancestral condition may persist for a relatively long period after divergence (Saeki et al., 2011) so

that ancestral polymorphism cannot be distinguished from introgression. In comparison, nuclear markers are biparentally-inherited and demonstrate higher mutation rates than chloroplast markers (Wolfe et al., 1987) and studies using codominant nuclear markers are valuable in detecting hybridization even when relatively small numbers of loci (4-5) are used (Boecklen & Howard, 1997). To date, most nuclear DNA based studies of introgression in temperate forest trees focus on a small number of relatively intensively-studied groups such as *Quercus* (Muir & Schlötterer, 2005; Valbuena-Carabaña et al., 2005; Lepais & Gerber, 2011), *Fraxinus* (Fernandez-Manjarres et al., 2006; Gerard et al., 2006; Thomasset et al., 2012), and *Populus* (Keim et al., 1989; Martinsen et al., 2001) (but see Hoban et al. 2009; Robertson et al. 2010).

The genus *Betula* provides a novel model for studies of introgression among temperate forest trees, as it is characterized by a high frequency of natural hybridization and associated polyploidy (DeJong, 1993). *Betula* is composed of approximately 35 species, which are predominantly long-lived perennial trees and shrubs of mostly cold-temperate and circumboreal distribution (Jarvinen et al., 2004). Studies of European *Betula* species have revealed widespread cpDNA haplotype sharing indicative of cpDNA introgression among sympatric *B. pendula*, *B. pubescens* and *B. nana* (Palme et al., 2004; Maliouchenko et al., 2007; Thórsson et al., 2010). The presence of nuclear introgression of European birches *B. nana* (2x) and *B. pubescens* (4x) also has been confirmed based on cytological and molecular studies (Thórsson et al., 2001; Anamthawat-Jónsson & Thórsson, 2003). It appears that gene flow between species is accomplished by way of a triploid-hybrid. Bi-directional gene flow between *B. nana* and *B. pubescens* is associated

with a continuous range of morphological variation between the diploid, triploid-hybrid, and tetraploid individuals.

In North America, *Betula* is composed of approximately 18 species with a wide range of morphological variation and natural hybrids (Furlow, 1990, 1997). Within eastern North America, three of the most common and wide-ranging species include paper birch (*Betula papyrifera* Marshall), yellow birch (*Betula alleghaniensis* Britton), and sweet birch (*Betula lenta* L.). Putative natural hybrids of *B. alleghaniensis* and *B. papyrifera* have been reported from areas of the northern United States and southern Canada (Barnes et al., 1974; Clausen, 1977). *Betula alleghaniensis* and *B. lenta* are closely-related and overlap sufficiently in distribution and flowering time that natural hybrids could be expected to occur (Sharik & Barnes, 1971). A previous study of these species based on chloroplast DNA demonstrated widespread haplotype sharing between *B. alleghaniensis* and *B. papyrifera* suggestive of genetic introgression between them, and there was evidence of limited haplotype sharing in the southern distribution of *B. lenta* and *B. alleghaniensis* (Chapter 1). Thus far, no studies have examined the genetic differentiation and incidence of introgression in *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* using nuclear DNA markers. As such, it is unclear to what extent these species might exchange nuclear alleles due to introgression, or whether the species might remain largely genetically isolated from one another. The objective of this study was to assess the genetic relationships and incidence of introgression among North American birches *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* using nuclear microsatellite markers. Specific objectives were (i) to examine the pattern of allele-sharing to assess the extent to which species can be considered genetically distinct, (ii) to examine whether these

species show evidence of genetic introgression at nuclear loci, and if so (iii) to identify the geographic pattern of introgression. It is expected that all analyses will reveal clear separation between the species, despite some level of allele-sharing since the species are relatively morphologically distinct and do not display many intermediates. Secondly, it is expected that there will be evidence of widespread introgression between *B. papyrifera* and *B. alleghaniensis*, as these species are suspected of producing natural hybrids and demonstrate strong patterns of cpDNA introgression (Chapter 1). In contrast, *B. alleghaniensis* and *B. lenta*, and *B. lenta* and *B. papyrifera* will not show introgressed genotypes as a previous study has suggested that these species are largely genetically isolated (Chapter 1).

**Methods**

*Study species*

*Betula papyrifera* Marsh. (2n=56, 70, 84) is the most widely distributed and cold-tolerant of the North American birches, with a distribution that ranges across the boreal forest region from the northern limit of tree growth in the Yukon and Alaska east to Newfoundland and Labrador (Grant & Thompson, 1975; Safford et al., 1990). The species occupies a wide variety of habitats, occurring on conditions ranging from xeric uplands to waterlogged muskegs but growth is best on moist, well-drained sandy loam sites (Safford et al., 1990). It is a predominantly shade-intolerant and characteristically pioneer species forming pure or mixed stands most often becoming established on recently-disturbed habitats. In the warm southern portions of its range, *B. papyrifera* is sympatric with the more southerly-distributed *B. alleghaniensis*. *Betula alleghaniensis* Britt (2n=84) occurs throughout temperate forests of northeastern North America from

southeastern Canada through New England to the southern Appalachians (Erdmann, 1990). *Betula alleghaniensis* is a mesophytic species, occupying streams and riverbanks throughout its range (Furlow, 1990). In the southern Appalachians, it occurs only at elevations greater than 1000 m. It is classified as intermediate in shade-tolerance, generally forming mixed stands in late-successional forests (Gilbert, 1960). *Betula lenta* L. (2n=24) is comparatively more restricted, and is endemic to the Appalachian region of the eastern United States, occurring from New England and southern Maine through the Appalachians to its southern limit in northern Alabama and Georgia (Lamson, 1990). *Betula lenta* is primarily a tree of moist, cool sites, occurring across a broad range of elevations from sea level in New England up to 1500 m in the southern Appalachians. It is sympatric with the closely-related *B. alleghaniensis* throughout the majority of its range. *B. lenta* is morphologically very similar to *B. alleghaniensis*, but may be distinguished by its dark, cherry-coloured, nonexfoliating bark and by the strong odour of wintergreen (methyl salicylate) in its twigs and bark (Sharik & Ford, 1984).

The three species fall into different subgenera based on their morphological characteristics. *Betula alleghaniensis* and *B. lenta* belong to subgenus *Betulenta,* which is composed mostly of species with dark, nonexfoliating bark which are believed to be closely-allied with the original birches of the Eocene (DeJong, 1993). In contrast, *B. papyrifera* falls within subgenus *Betula*, which is composed predominantly of white-barked, pioneer species. This morphological classification is supported by molecular studies that separate the white and dark barked birches into different clades (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008). Two studies place *B. lenta* and *B. alleghaniensis* as well-supported sister-species (Jarvinen et al., 2004; Schenk et al., 2008)

The phylogenetic placement of *B. papyrifera* is poorly resolved, perhaps because hybridization is frequent among the white-barked birches (DeJong, 1993; Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008). The relationships are further complicated by prevalent polyploidy within the genus; *B. papyrifera* appears to represent an allopolyploid, and there is evidence to suggest it may have formed from hybridization of a *pendula*-like diploid species and a tetraploid of subgenus *Betula* (Jarvinen et al., 2004). *Betula alleghaniensis* is thought to be an allopolyploid, and it has been suggested that it may have formed from hybridization of progenitor species of subgenera *Betulenta* and *Betula* (DeJong, 1993).

Putative natural hybrids of *B. papyrifera* and B. *alleghaniensis* have been identified based on morphological traits of specimens from Michigan, Minnesota, Wisconsin, Iowa, and New Hampshire (Barnes et al., 1974; Clausen, 1977), and their reproductive compatibility also has been shown by experimental crosses (Clausen, 1966; Barnes et al., 1974). In sympatric regions, natural hybridization could potentially be widespread as the flowering period of *B. alleghaniensis* overlaps sufficiently with pollen shed in *B. papyrifera* that cross-pollination could occur (Barnes et al., 1974). Experimental crosses have also revealed compatibility of *B. papyrifera* with *B. lenta* (Clausen, 1966), though natural hybrids have not been reported. Despite their taxonomic similarity and evolutionary relationship, natural hybrids of *B. lenta* and *B. alleghaniensis* have not been reported. However, controlled crosses have been successfully produced with *B. lenta* acting as the maternal parent (Clausen, 1966; Sharik & Barnes, 1971). Also, *B. alleghaniensis* and *B. lenta* overlap in range and flowering time sufficiently that natural hybrids might be expected to occur (Sharik & Barnes, 1971).

*Population sampling*

Mature leaf samples were collected from natural populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* at 20 sites across the species` sympatric distribution in northeastern North America (Figure 1). At each site, sampled individuals were separated by a minimum distance of 50 m to avoid collection of multiple samples from the same clonal individual. Leaf material was quick-dried in silica gel to preserve DNA quality. In total, samples from 483 individuals of *B. alleghaniensis* at 16 sites, 256 individuals of *B. papyrifera* at 9 sites, and 300 individuals of *B. lenta* at 12 sites were collected (Table 1). *Betula alleghaniensis* and *B. papyrifera* co-occurred at eight sites, *B. alleghaniensis* and *B. lenta* co-occurred at ten sites, *B. papyrifera* and *B. lenta* co-occurred at two sites, and all three species occurred together at two sites.

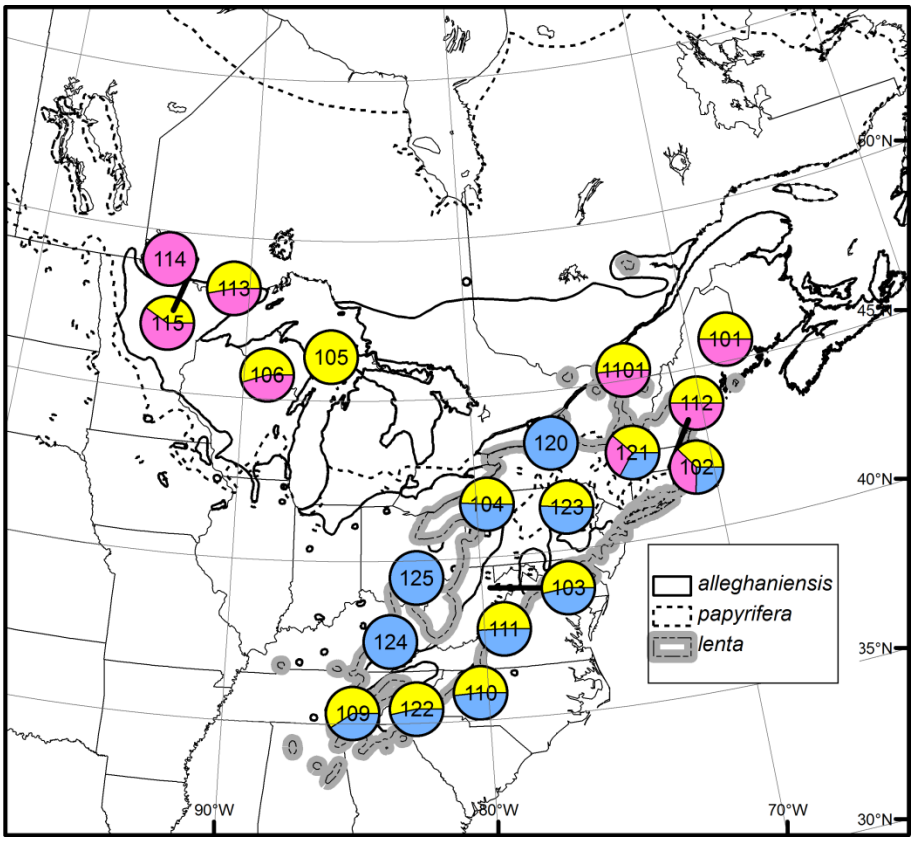



Figure 1. Distribution of *Betula* population sampling locations throughout northeastern North America.

Circles represent sampling locations and are shaded with different colours to represent the proportion of each species sampled at each location; *B. alleghaniensis* (yellow), *B. papyrifera* (pink), *B. lenta* (blue). Different line shadings are used to indicate the natural ranges of each species.

Table 1. Location and numbers of *Betula* population samples used in this study.

| Population | Population Code | Province/ State | Lat. | Long. | (N) all | pap | len |
|---|---|---|---|---|---|---|---|
| Penobscot | 101 | ME | 44.86 | -68.63 | 30 | 30 | |
| Massabesic | 102 | ME | 43.44 | -70.67 | 30 | 29 | 20 |
| Fernow | 103 | WV | 39.05 | -79.67 | 30 | | 26 |
| Kane | 104 | PA | 41.60 | -78.77 | 29 | | 30 |
| Dukes | 105 | MI | 46.36 | -87.16 | 35 | | |
| Argonne | 106 | WI | 45.75 | -88.98 | 37 | 31 | |
| North River | 109 | TN | 35.33 | -85.09 | 29 | | 20 |
| Mt. Rogers | 110 | VA | 36.74 | -81.42 | 30 | | 28 |
| Glenwood | 111 | VA | 37.75 | -79.23 | 30 | | 29 |
| Bartlett | 112 | NH | 44.05 | -71.30 | 31 | 31 | |
| Greenwood | 113 | ON | 48.39 | -90.75 | 28 | 25 | |
| Marcell | 114 | MN | 47.53 | -93.47 | | 28 | |
| Blandin | 115 | MN | 47.12 | -93.68 | 20 | 30 | |
| Finger Lakes | 120 | NY | 42.49 | -76.77 | | | 19 |
| Hopkins | 121 | MA | 42.73 | -73.25 | 33 | 24 | 28 |
| Pisgah | 122 | NC | 35.43 | -82.73 | 31 | | 27 |
| Luzerne | 123 | PA | 41.32 | -76.29 | 30 | | 31 |
| Cumberland | 124 | KY | 37.88 | -83.66 | | | 27 |
| Hocking | 125 | OH | 39.47 | -82.58 | | | 15 |
| Gault | 1101 | QC | 45.32 | -73.09 | 30 | 28 | |

all, *alleghaniensis* ; pap, *papyrifera* ; len, *lenta*

*Laboratory procedures*

DNA was extracted from 20 milligrams of dried leaf tissue following the modified CTAB protocol of Zeng et al. (2002). Individuals were analyzed at six polymorphic microsatellite markers developed in previous studies (Wu et al., 2002; Kulju et al., 2004; Truong et al., 2005; Tsuda et al., 2008). Information and sequences for primers used in this study are presented in Appendix 9. Forward primers were 5`-labeled with infrared dye (IRD) and amplification was conducted in 10 µl total volume with multiplexed primers. Multiplex 1 contained 1 µl of 10x buffer, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 0.4 µM CD277302 forward and reverse primers, 0.2 µM L2.3 forward and

reverse primers, 0.5 U Taq polymerase, and 2.0 µl of template DNA. Multiplex 2 contained 1 µl of 10x buffer, 0.2 mM dNTPs, 3.0 mM $MgCl_2$, 0.4 µM L1.10 forward and reverse primers, 0.1 µM L5.5 forward and reverse primers, 0.5 U Taq polymerase, and 2.0 µl of template DNA. For the remaining primers (L5.4, Bp04) fluorescent labeling was accomplished through the use of a tailed-primer protocol (Schuelke, 2000). A 19-bp tail of sequence CACGACGTTGTAAAACGAC was added to the 5`-end of each forward primer and amplification was conducted in a volume of 10 µl containing 1 µl of 10X buffer, 0.2 mM dNTPs, 3.0 mM $MgCl_2$, 0.1 µM forward primer, 0.1 µM IRD primer (CACGACGTTGTAAAACGAC), 0.5 µM reverse primer, 0.5 U Taq polymerase, and 2.0 µl of template DNA. Thermocycling consisted of an initial denaturation at 94°C for 4:00 min, followed by 35 cycles of 94°C for 45 s, 50°C for 1:00 min, 72°C for 45 s and final elongation at 72°C for 10:00 min. Amplified fragments were electrophoretically separated on the Licor 4200 DNA analyzer along with the manufacturer`s 50-350 bp size standard (LiCOR Biosciences, Lincoln, NE). Fragment lengths were scored using SAGA GT 2.1 software (Licor Biosciences) and verified manually.

*Data analysis*

For genotyping of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* the different allele size-variants present within each individual were recorded. For the diploid *B. lenta*, standard allele-frequency data were used. However, due to partial heterozygosity of polyploids *B. alleghaniensis* and *B. papyrifera*, the underlying copy-number of each allele could not be determined. Thus, for *B. alleghaniensis* and *B. papyrifera* only the presence or absence of each amplified DNA fragment was considered and it was assumed in all cases that the underlying genotype of each individual was ambiguous. Since *B.*

*papyrifera* is known to have varying chromosome numbers, the ploidy level of each individual was estimated based on the maximum number of DNA fragments or "alleles" at a locus using the POLYSAT package for R software (Clark & Jasieniuk, 2011). The ploidy of all individuals of *B. alleghaniensis* and *B. lenta* was fixed as 6x and 2x, respectively.

Simple diversity measures including the total number of alleles ($A_T$), effective number of alleles ($A_E$) (Nielsen et al., 2003), and gene diversity ($H_E$) (Nei, 1978) were calculated for each locus using SPAGEDI software (Hardy & Vekemans, 2002a). Due to the unknown allele-dosage of polyploids, calculated $H_E$ values are equal to the probability that two alleles drawn at random from one individual are the same (Münzbergová et al., 2013). To examine genetic relationships among *B. alleghaniensis*, *B. papyrifera*, and *B. lenta*, we calculated the number of unique alleles for each species ($A_U$), as well as the percentage alleles shared among different species-groupings. The correlation of allele-frequency distributions between different species-pairs was determined using Pearson's product-moment correlation calculated in Sigmaplot software (Systat Inc, Redlands, CA). Allele frequencies per locus and species were estimated using SPAGEDI under the assumption that each of the observed alleles was present in only a single copy. However, this simple allele-frequency calculation may tend to overestimate the frequency of rare alleles and underestimate the frequency of common alleles (Clark & Jasieniuk, 2011). Matrices of genetic distances among individuals of all species were calculated using two different measures suitable for polyploid microsatellite data with unknown allele-dosage; the first of these is a band-sharing similarity index which defines genetic similarity between individuals as the fraction of shared bands based on the

infinite allele model (IAM) (Lynch, 1990), while the second measure uses the stepwise mutation model (SMM) to determine genetic distances for polyploid microsatellite data (Bruvo et al., 2004). Distance matrix calculations were performed using the POLYSAT software package for R (Clark & Jasieniuk, 2011). POLYSAT is an R-software package designed for genetic analysis of polyploid microsatellite data. It is suitable for analysis of both autopolyploid and allopolyploid data of any ploidy, or mixed ploidy level. It accommodates partial heterozygosity of polyploids by assuming that allele copy-number is always ambiguous. To visualize the genetic relationships among species, principal components analysis (PCA) was performed on the matrices of genetic distances and the resultant PCA axes were plotted in POLYSAT. Genetic differentiation among species was examined through Analysis of Molecular Variance (AMOVA) (Excoffier et al., 1992) considering all species simultaneously, as well as different pairs of species-groupings. AMOVA was conducted in GENALX (Peakall & Smouse, 2006) based on matrices of genetic distances with 999 permutations of the data to calculate significance values. Genetic differentiation among species was estimated by the measure $\Phi_{PT}$, an $F_{ST}$-analog calculated as part of the AMOVA procedure in Genealx. Genetic differentiation among species also was assessed based on allele-identity under the infinite allele model (IAM) with the statistic $F_{ST}$ (Weir & Cockerham, 1984), and based on allele-size under the stepwise mutation model (SMM) with the statistic $R_{ST}$ (Slatkin, 1995) calculated in SPAGEDI with significance values obtained from 10000 random permutations of individuals. We also tested whether closely-related alleles tended to be located within-species by evaluating the significance of the observed $R_{ST}$ relative to the expected $R_{ST}$ obtained from random permutation of allele sizes among alleles following Hardy et al.

(2003). If the permutated value of $R_{ST}$ is significantly lower than the observed value, this indicates that closely-related alleles tend to be located within, rather than among species.

As another method of examining genetic structuring among the species, the program STRUCTURE (Pritchard et al., 2000) was used to examine the partitioning of individuals of each species into genetic clusters; STRUCTURE uses a Bayesian clustering algorithm to assign individuals to populations on the basis of their genotypes in such a way as to minimize Hardy-Weinberg and linkage disequilibrium. The latest version of the program accommodates genetic ambiguity due to partial heterozygosity at codominant loci in polyploids by incorporating a new algorithm which generates full genotypes for each individual based on their partial genotypes (Falush et al., 2007). The STRUCTURE analysis was run using the admixture model, which allows for the presence of individuals with ancestry from more than one population. Based on the user-specified *K*-value, the model clusters individuals into populations and estimates their proportion of membership within each population (Q). High Q-values indicate that the majority of an individual's alleles originated from the population to which it is assigned, while lower Q-values indicate that some portion of an individual's alleles may have originated from within another population. The model was run with a burn-in of 100000 iterations and 100000 iterations of each MCMC chain for K=1 to 6, each with five independent runs. The admixture model was run with the correlated allele frequencies option, since allele frequencies could potentially be correlated among the species due to hybridization or shared ancestral polymorphism. A threshold of $Q<0.8$ was used to identify admixed individuals which could potentially represent hybrids or introgressants.

To investigate spatial trends in the occurrence of putative introgressants (i.e. potential hybrid zones), the location of admixed individuals was mapped using ArcGIS 10.0.

As an additional means of investigating potential hybridization among species, we used tests of isolation by distance to investigate the genetic similarity between population pairs of different species at different spatial scales. If introgression was occurring, it would be expected that the genetic similarity between two hybridizing species at the same locality should be greater than the genetic similarity between populations of the same two species measured at different localities (Muir & Schlötterer, 2005). To test for the presence of isolation-by-distance, geographic distances among population-pairs of different species were compared to genetic distances between them. The pairwise geographic distances between populations were calculated using geographic coordinates and Genalx software. Genetic differentiation values among pairs of populations ($\Phi_{PT}$) were estimated via AMOVA using the input matrix of genetic distances among individuals in Genalx. The correlation between genetic distances and geographic distances was examined through simple linear regression implemented in Sigmaplot software.

**Results**

*Molecular diversity, allele-sharing, and correlations among species*

The number of alleles per locus ranged from 10 (L5.5) to 16 (L1.10, L2.3) with a total of 81 alleles across the six loci (Table 2). When allele frequencies were considered, the greatest number of effective alleles occurred for L1.10 (6.03) and the lowest effective number was for Bp04 (3.43). Gene diversity values were generally similar across loci, ranging from 0.7720 for locus L5.4 to 0.8341 for locus L1.10.

Table 2. Summary of molecular variation for six microsatellite loci used in this study.

| Locus | $A_T$ | $A_E$ | $H_E$ |
|---|---|---|---|
| CD277302 | 12 | 4.94 | 0.7977 |
| L1.10 | 16 | 6.03 | 0.8341 |
| L2.3 | 16 | 5.07 | 0.8027 |
| L5.5 | 10 | 4.63 | 0.7840 |
| L5.4 | 13 | 4.39 | 0.7720 |
| Bp04 | 14 | 3.43 | 0.7081 |
| multilocus | 81 | 4.75 | 0.7831 |

$A_T$, total number of alleles; $A_E$, effective number of alleles; $H_E$, gene diversity

In general, common alleles were shared between species, and most of the unique alleles were present in low frequencies (<5%). A greater number of unique alleles was found in *Betula alleghaniensis* ($A_U$=18) than *B. papyrifera* ($A_U$=3) and *B. lenta* ($A_U$=1). When unique alleles with less than 5% frequency were excluded, 2 unique alleles were found in *B. alleghaniensis*, no unique alleles were found in *B. papyrifera*, and a single unique allele was found in *B. lenta*. Eighteen alleles (22.2%) were shared among the three species. The highest level of allele-sharing (69.1%) was found between *B. papyrifera* and *B. alleghaniensis*, while levels of allele-sharing were much lower for *B. alleghaniensis* and *B. lenta* (25.9%) and *B. papyrifera* and *B. lenta* (22.2%). Allele-frequencies were significantly positively correlated between *B. alleghaniensis* and *B. papyrifera* ($r=0.285$, $P<0.01$), and *B. alleghaniensis* and *B. lenta* ($r=0.462$, $P<0.01$). In contrast, allele frequencies of *B. papyrifera* and *B. lenta* were not significantly correlated ($r=0.121$, $P=0.282$).

*Partitioning of molecular variation among species*

Analysis of molecular variance indicated substantial variation among the three species (40.3%) based on Bruvo distances (Table 3). In contrast, variation among populations within species was low (1.4%), and the majority of variation was harboured within populations (58.3%). When species-pairs were considered, the greatest variation was found between *B. papyrifera* and *B. lenta* (56.2%). Differentiation between *B. alleghaniensis* and *B. lenta* was somewhat lower (40.7%), and the lowest variation was found between *B. alleghaniensis* and *B. papyrifera* (27.5%). Percentages of among-species variation based on Lynch distance were similar to, but slightly higher than those based on Bruvo distances (Appendix 10).

When the three species were considered simultaneously, among-species variation based on the infinite allele model ($F_{ST}$=0.2365) was lower than the value calculated based on the stepwise mutation model ($R_{ST}$=0.3493), and both values were significant (P<0.01). However, the calculated $R_{ST}$ value was not significantly greater than the permuted value (P=0.0715), indicating that allele sizes did not significantly contribute to the partitioning of variation among species, as closely-related alleles tended to be distributed randomly between species instead of occurring mostly within species.

Table 3. Analyses of Molecular Variance (AMOVA) for the partitioning of molecular variation among species groupings based on Bruvo distances (SMM).

| Species[1] Grouping | Source of variation | df | Sum of squares | Variance components | Percent variation | $\Phi_{PT}$ | P-value |
|---|---|---|---|---|---|---|---|
| a) *all/pap/len* | Among species | 2 | 90.98 | 0.136 | 40.3 | 0.403 | 0.001 |
| | Among populations | 34 | 11.05 | 0.005 | 1.4 | | 0.001 |
| | Within populations | 1002 | 196.95 | 0.197 | 58.3 | | 0.001 |
| b) *all/pap* | Among species | 1 | 29.57 | 0.087 | 27.5 | 0.275 | 0.001 |
| | Among populations | 23 | 8.77 | 0.005 | 1.7 | | 0.001 |
| | Within populations | 714 | 160.40 | 0.225 | 70.8 | | 0.001 |
| c) *all/len* | Among species | 1 | 50.82 | 0.136 | 40.7 | 0.407 | 0.001 |
| | Among populations | 26 | 8.13 | 0.004 | 1.3 | | 0.001 |
| | Within populations | 755 | 146.99 | 0.195 | 58.0 | | 0.001 |
| d) *pap/len* | Among species | 1 | 59.15 | 0.213 | 56.2 | 0.562 | 0.001 |
| | Among populations | 19 | 5.19 | 0.004 | 1.1 | | 0.001 |
| | Within populations | 535 | 86.51 | 0.162 | 42.7 | | 0.001 |

[1] species abbreviations, all = *alleghaniensis;* pap = *papyrifera;* len = *lenta*

*Genetic clustering and admixture*

Principal components analyses based on both Bruvo and Lynch distances revealed a clear separation of *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* species' clusters, though a small number of intermediates occurred between *B. alleghaniensis* and *B. lenta*, and *B. alleghaniensis* and *B. papyrifera* (Figure 2). In contrast, no intermediates were observed between species' clusters of *B. papyrifera* and *B. lenta*.

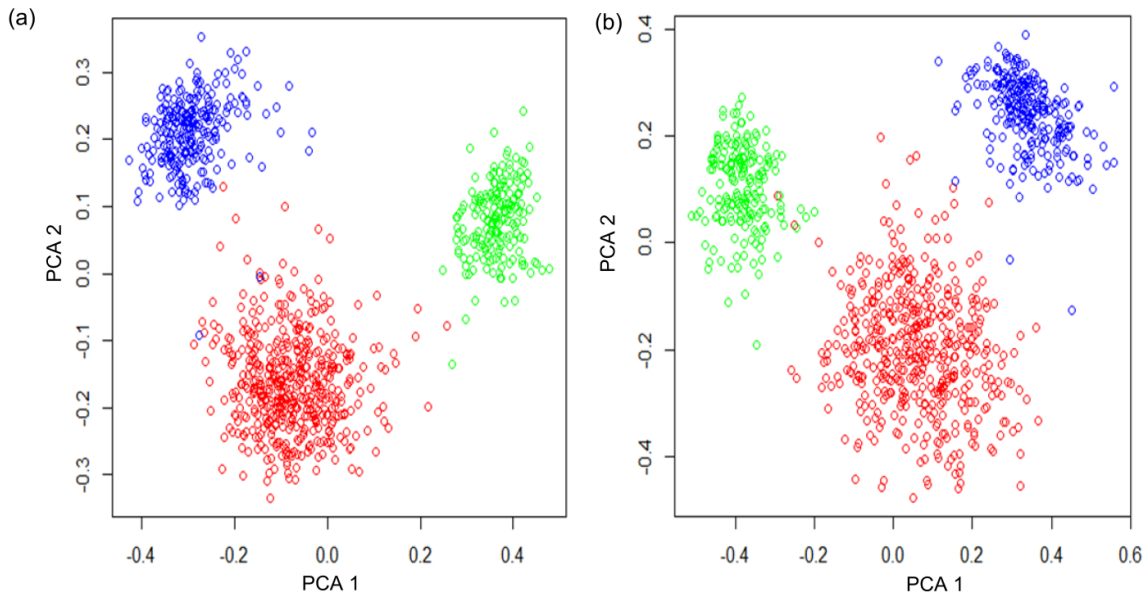



Figure 2. PCA plot showing species separation of *B. alleghaniensis* (red), *B. papyrifera* (blue), and *B. lenta* (green) calculated using a) Bruvo distance and b) Lynch distance.

When *B. papyrifera*, *B. alleghaniensis*, and *B. lenta* were considered jointly, STRUCTURE analyses results based on the correlated allele frequency model without prior population information indicated that the value of ln Pr($X|K$) continued increasing with $K$ values from 1 to 6. When the $K$-selection criteria proposed by Pritchard et al. (2000) was applied, it was found that the Ln Pr($X|K$) values increased substantially from $K$=1 to $K$=3 and then levelled off, suggesting that the number of optimal population groups was $K$=3 (Figure 3a). However, the rate of change in the posterior probability ($\Delta K$) was greatest between $K$=2 and $K$=3 (Figure 3b), indicating that the number of populations groups was $K$=2, based on the method of Evanno et al. (2005). For $K$=2, the first cluster was composed almost exclusively of *B. alleghaniensis* and *B. lenta*, and the second cluster was composed mostly of *B. papyrifera* and a small number of admixed *B. alleghaniensis*. When the cluster containing individuals of *B. alleghaniensis* and *B. lenta* was analyzed separately in STRUCTURE, it was found that *B. alleghaniensis* and *B.*

*lenta* individuals were assigned to separate species clusters with generally high membership proportions (Q>0.8). Thus, it appears that individuals putatively identified as *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* represent separate population groups and analyses presented herein are based on *K*=3.

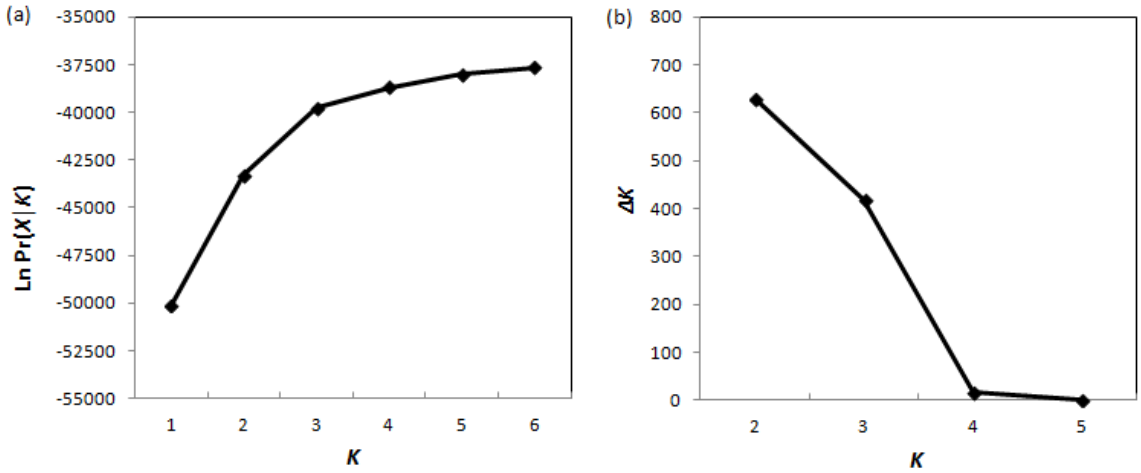



Figure 3. Plots of structure output parameters for *K*=1-6 used to identify the number of population groups (*K*) when *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* were considered jointly using the admixture model with correlated allele frequencies and no prior population information: a) mean of the posterior probabilities, ln Pr($X \mid K$), for each *K*-value, and b) change in the posterior probability ($\Delta K$) between successive *K*-values.

For *K*=3, the three clusters identified by STRUCTURE corresponded well to the putative species of origin of each individual (Figure 4). The majority of individuals had high membership coefficients (Q>=0.8) within their assigned clusters, with 84.9% of *B. alleghaniensis* assigned to cluster I, 98.4% of *B. papyrifera* assigned to cluster II, and 99.7% of *B. lenta* assigned to cluster III. *Betula alleghaniensis* possessed the greatest number of admixed individuals with 5.2% of individuals indicated as admixed with *B. papyrifera* and 10.1% admixed with *B. lenta* (Q>=0.2). In comparison, 0.8% of *B. papyrifera* and 0.3% of *B. lenta* were indicated as admixed with *B. alleghaniensis*

(Q>=0.2). No individuals of *B. papyrifera* were indicated as admixed with *B. lenta*, and a single individual of *B. lenta* was indicated as admixed with *B. papyrifera* (Q>=0.2)

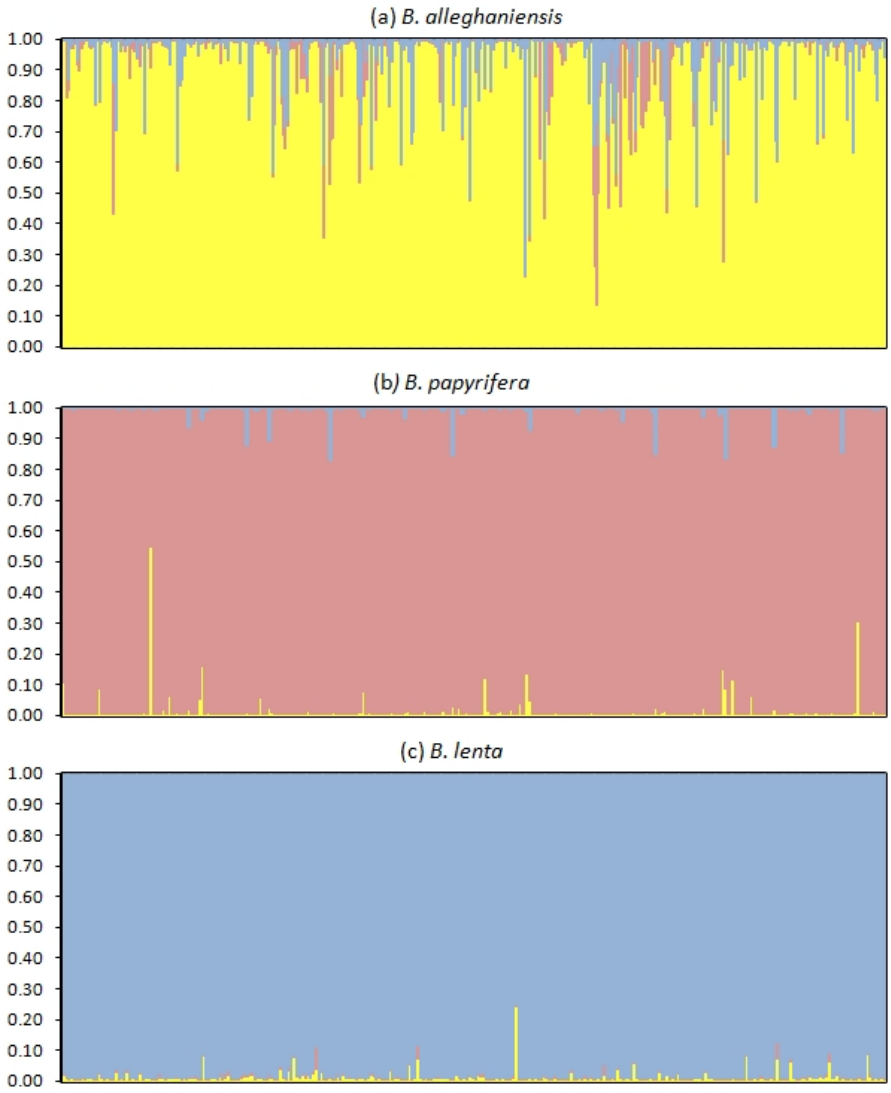
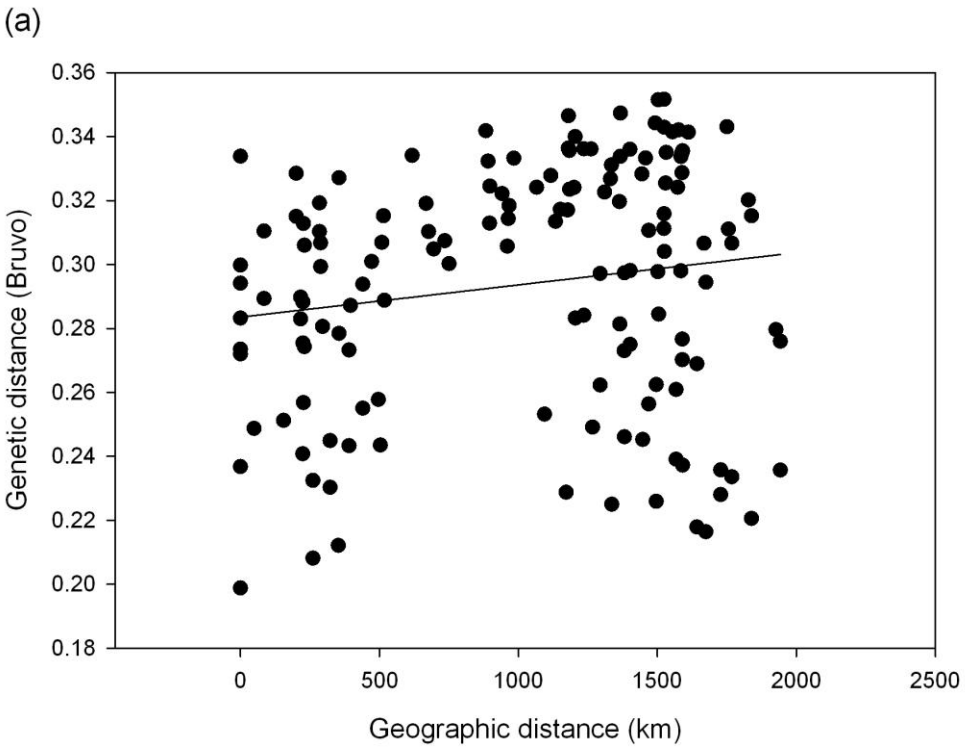



Figure 4. Assignments of individuals to population clusters identified by STRUCTURE at K=3 for (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.

*Geographic patterns of genetic similarity and admixture*

No significant pattern of isolation by distance was detected among population pairs of *B. alleghaniensis* and *B. papyrifera* ($R^2$=0.0246, P=0.0604), *B. alleghaniensis*

and *B. lenta* ($R^2$=0.0009, P=0.686), or *B. papyrifera* and *B. lenta* ($R^2$=0.0142, P=0.219) based on Bruvo genetic distances (Figure 5). Similarly, correlations based on Lynch distances did not indicate a significant relationship of geographic and genetic distances between *B. alleghaniensis* and *B. papyrifera* ($R^2$=0.0132, P=0.170), or *B. papyrifera* and *B. lenta* ($R^2$=0.0018, P=0.667). However, a slight but significant negative correlation was found for *B. papyrifera* and *B. lenta* ($R^2$=0.0292, P=0.0178).



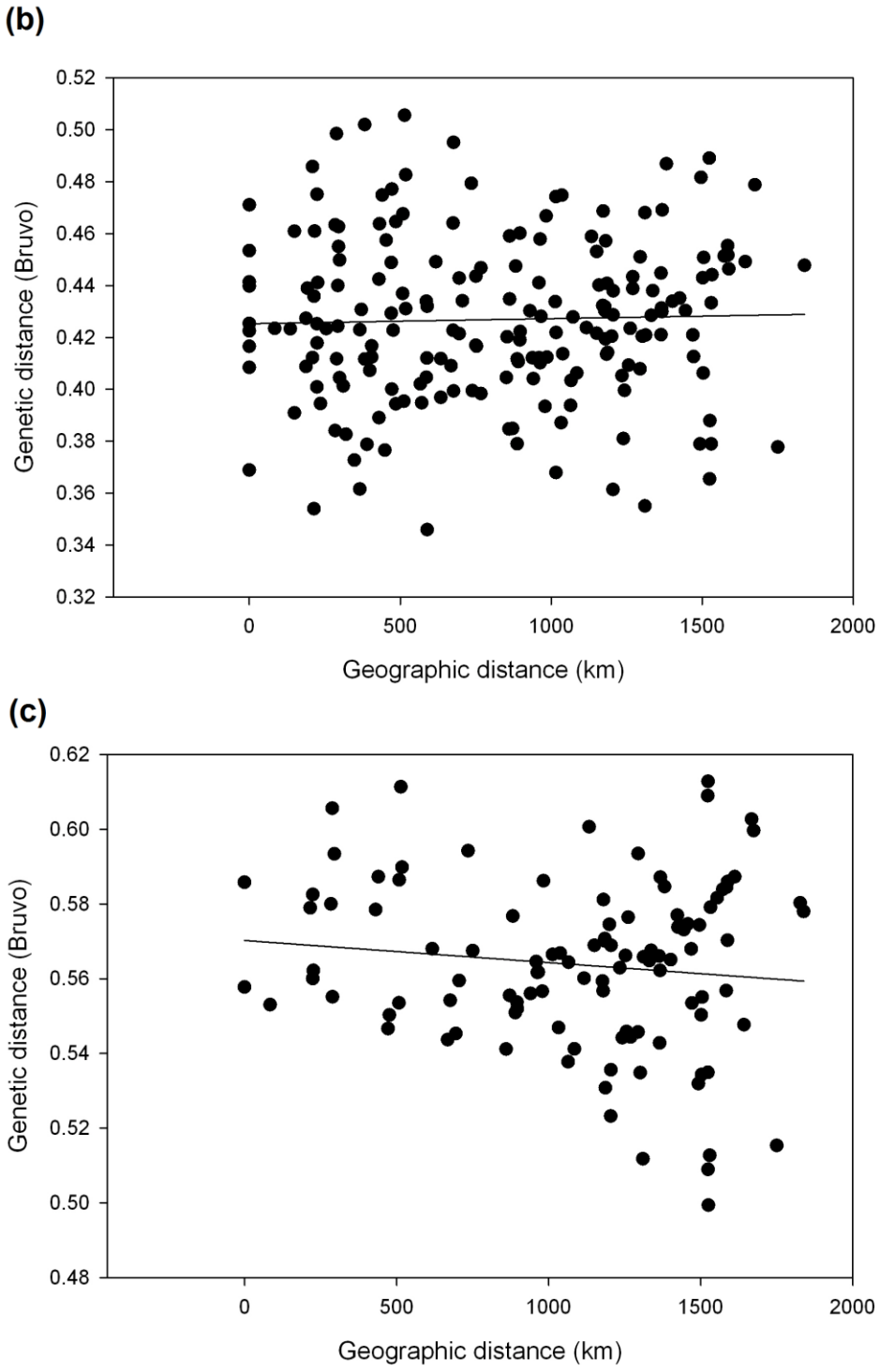
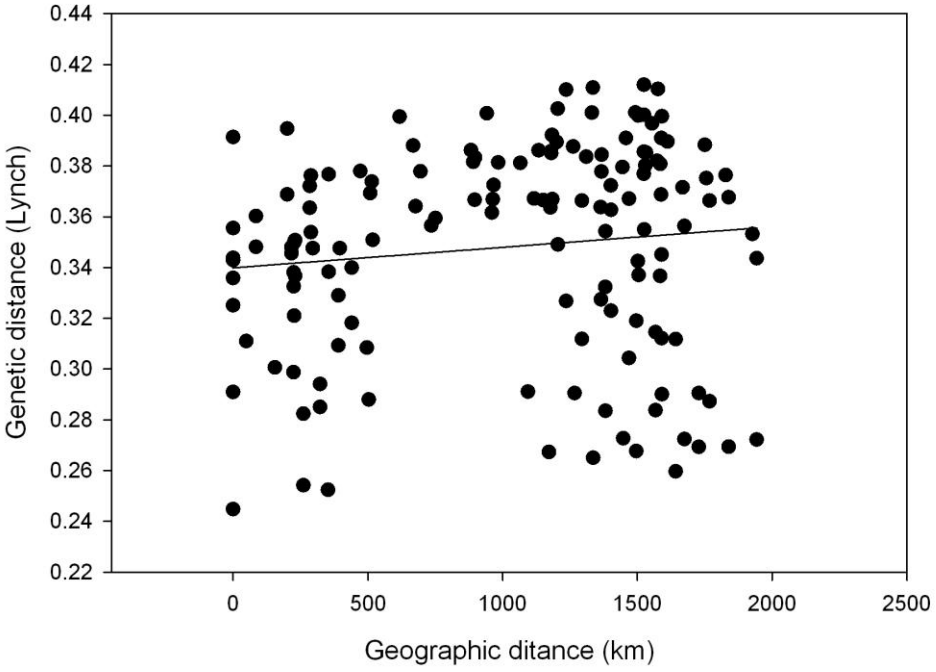
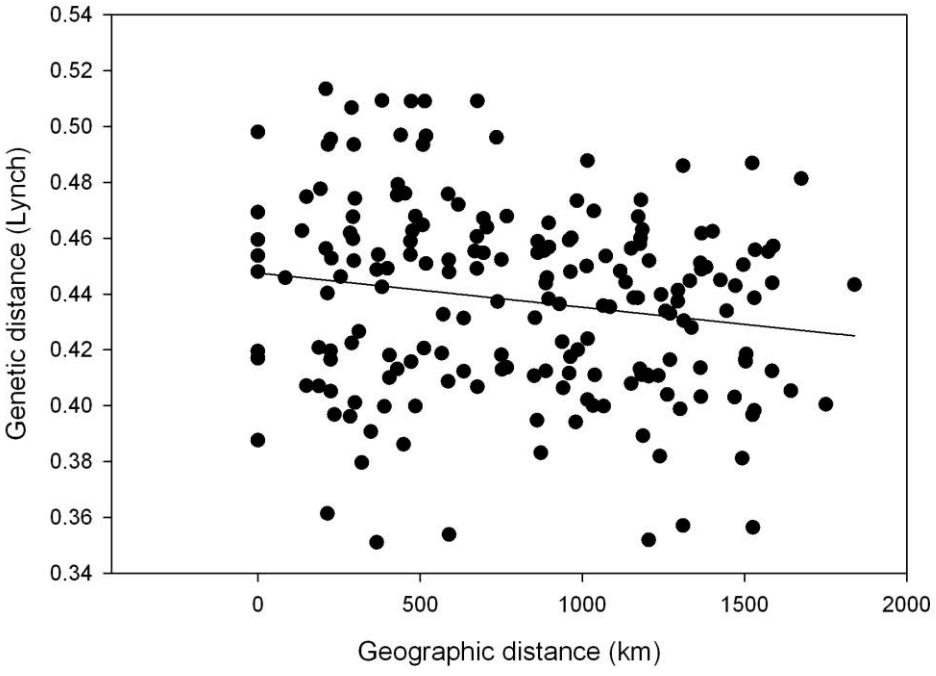



Figure 5. Regression analysis of geographic distances against Bruvo genetic distances between population pairs of (a) *alleghaniensis/papyrifera* (b) *alleghaniensis/lenta,* and (c) *papyrifera/lenta.*

(a)


0.44
0.42
0.40
0.38
0.36
0.34
0.32
0.30
0.28
0.26
0.24
0.22
Genetic distance (Lynch)
0
500
1000
1500
2000
2500
Geographic ditance (km)

(b)


0.54
0.52
0.50
0.48
0.46
0.44
0.42
0.40
0.38
0.36
0.34
Genetic distance (Lynch)
0
500
1000
1500
2000
Geographic distance (km)

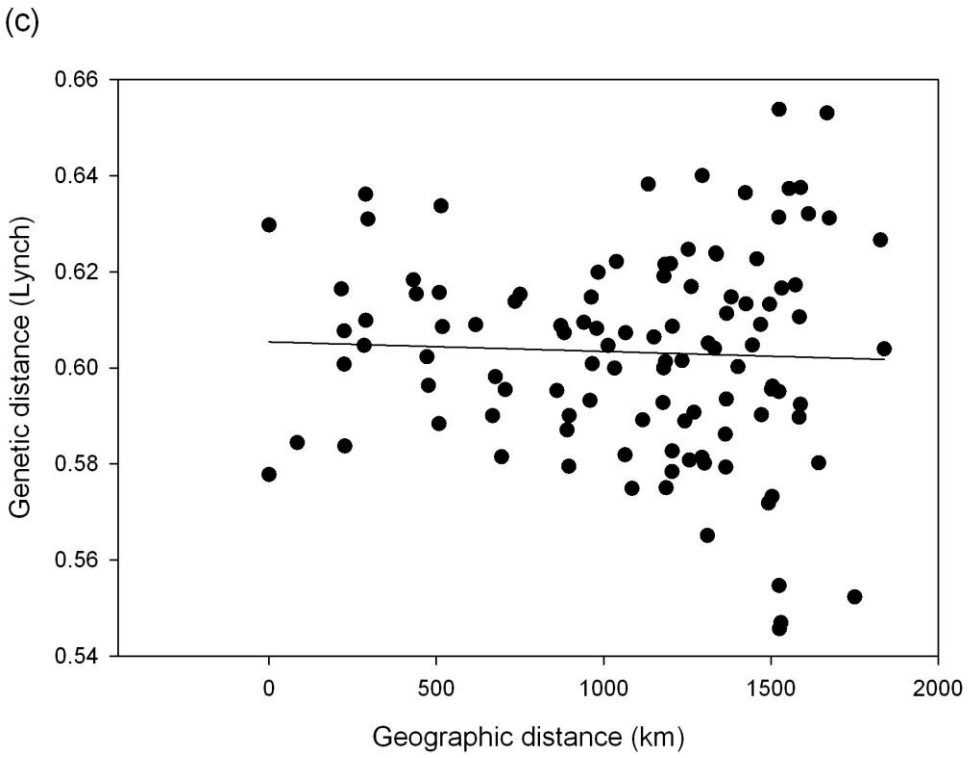
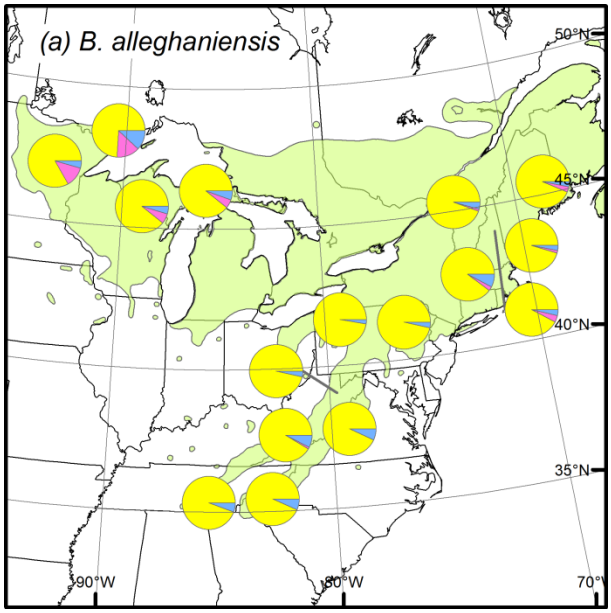
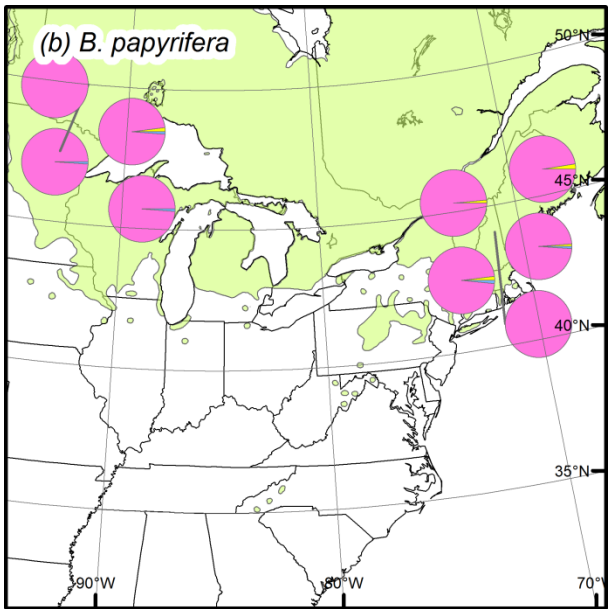


Figure 6. Regression analysis of geographic distances against Lynch genetic distances between populations pairs of (a) *alleghaniensis/papyrifera* (b) *alleghaniensis/lenta,* and (c) *papyrifera/lenta*.

Mapped admixture proportions for *B. alleghaniensis* indicated that the majority of alleles for populations throughout the study area were of *B. alleghaniensis* origin (Figure 7a). Some admixture with *B. lenta* was found as *B. alleghaniensis* populations throughout the study area possessed *B. lenta*-type alleles, though admixture proportions were generally small and similar across the geographic range. One site, S113, showed a slightly higher proportion of alleles originating from *B. lenta*. However, this site was outside the natural range of *B. lenta* and thus did not contain any individuals of that species. Eastern populations of *B. alleghaniensis* generally contained a low proportion of *B. papyrifera* alleles, while western Great Lakes populations demonstrated a somewhat higher proportion of *B. papyrifera* alleles. *Betula papyrifera* had very low proportions of alleles from *B. alleghaniensis* and *B. lenta* and there did not appear to be any clear

geographic pattern to the inferred admixture as the proportion of alleles from foreign species was roughly equal across the sampled populations (Figure 7b). *Betula lenta* populations across the study area were composed predominantly of *B. lenta*-type alleles, and the proportion of *B. alleghaniensis* alleles was low, indicating little admixture between the two species (Figure 7c). Also for *B. lenta*, little admixture with *B. papyrifera* was found throughout the range.



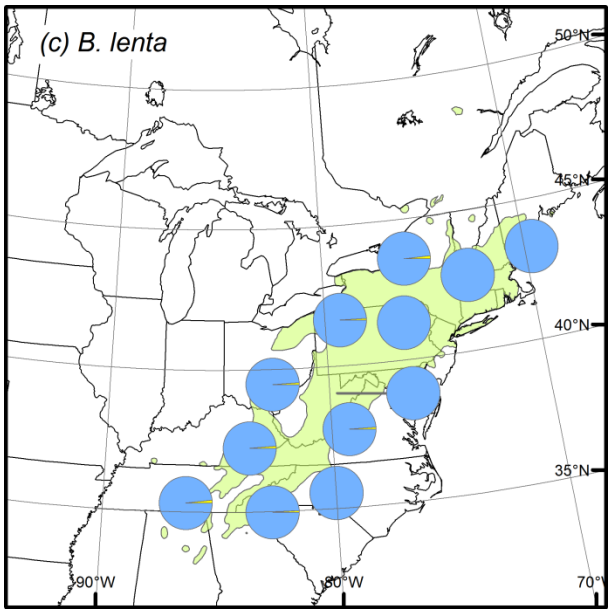


Figure 7. Proportion of alleles inferred to have originated from each species based on Structure analysis at K=3 for (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*. The proportion of alleles from each species is shown as pie chart with the proportion of *B. alleghaniensis* alleles indicated by yellow shading, the proportion of *B. papyrifera* alleles is in pink, and the proportion of *B. lenta* alleles is in blue.

**Discussion**

The analyses presented in this study suggest that *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* represent separate and distinct species with low incidence of genetic intermediates between them. Principal components analysis indicated a clear grouping of the three species into distinct genetic clusters and relatively few intermediates occurred between the species' groups. Despite moderate to high levels of allele-sharing among the species (all/pap/len=22.2%, all/pap=69.1%, all/len=25.9%, pap/len=22.2%), allele profiles for the species were distinct across the six loci and analysis of molecular variation indicated substantial and significant variation among the three species ($\Phi_{PT}$=0.403). Of the different species-pairs, *B. papyrifera* and *B. lenta* appeared to be the most highly-differentiated. Analysis of molecular variation indicated that these species

demonstrated greater differentiation ($\Phi_{PT}$=0.567) than *B. alleghaniensis* and *B. lenta* ($\Phi_{PT}$=0.407) or *B. alleghaniensis* and *B. papyrifera* ($\Phi_{PT}$=0.275). Furthermore, no genetic intermediates between *B. papyrifera* and *B. lenta* were revealed by PCA, and STRUCTURE analysis indicated only a single individual of *B. lenta* which was putatively admixed with *B. papyrifera*. The finding of a high level of genetic structuring between *B. papyrifera* and *B. lenta* is not surprising based on the current understanding of the phylogenetic relationships of these species. *Betula papyrifera* and *B. lenta* belong to different subgenera based on morphological classification; *B. papyrifera* belongs to subgenus *Betula*, while *B. lenta* belongs to subgenus *Betulenta* (DeJong, 1993). The separation of *B. papyrifera* and *B. lenta* also is supported by nuclear DNA-based phylogenetic studies of *Betula*, which show that *B. lenta* and *B. papyrifera* group within distinct and well-differentiated clades (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008). Furthermore, a phylogeographic study of eastern North American birches found little evidence for introgression between *B. papyrifera* and *B. lenta*, despite extensive evidence of introgression between *B. papyrifera* and most other species examined (Chapter 1). The predominant reproductive isolation of *B. papyrifera* and *B. lenta* could be the result of genomic incompatibilities between diploid *B. lenta* and polyploid *B. papyrifera* which might contribute to strong post-zygotic isolation. Furthermore, *B. papyrifera* and *B. lenta* overlap only within a small area of New England and the northern Appalachians, which should limit their opportunity to interbreed.

Despite being mostly genetically distinct in the current study, *B. alleghaniensis* included numerous individuals that were intermediate with *B. papyrifera* and *B. lenta*. Based on both PCA and STRUCTURE analysis, a number of individuals of *B.*

*alleghaniensis* were identified as being putatively admixed with *B. papyrifera* or *B. lenta*, though STRUCTURE analysis revealed that a relatively higher proportion of *B. alleghaniensis* individuals were admixed with *B. lenta* (10.1%) than *B. papyrifera* (5.2%). In contrast, analysis of molecular variation indicated a relatively greater differentiation of *B. alleghaniensis* and *B. lenta* ($\Phi_{PT}$=0.407) than of *B. alleghaniensis* and *B. papyrifera* ($\Phi_{PT}$=0.275). *Betula alleghaniensis* and *B. papyrifera* shared a higher proportion of alleles (69.1%) than *B. alleghaniensis* and *B. lenta* (25.9%). Sharing of alleles between species can be explained either by homoplasy, shared ancestral polymorphism, or hybridization and introgression (Palme et al., 2004). In the current study, shared ancestral polymorphism may be the most plausible explanation for the finding of shared alleles between *B. alleghaniensis* and *B. lenta*, as these two species are phylogenetically closely related (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008). The lack of strong geographic patterns of admixture between *B. alleghaniensis* and *B. lenta* provides relatively weak support for nuclear DNA introgression between these species. Under a scenario of localized introgression, one would expect to find that allele frequencies would be more highly correlated between populations occurring in close geographic proximity than between populations at more distant localities (Muir & Schlötterer, 2005). For example, a study of *Actinidia* species found that *A. chinensis* and *A deliciosa* tended to be more admixed at locations where the two species occurred in close geographic proximity (Liu et al., 2010). Also, generally higher rates of introgression might be expected in zones of sympatry compared to zones of allopatry as has been observed for numerous plant species groups; for example *Pinus* (Ye et al., 2002), *Alnus* (Bousquet et al., 1990), *Coffea* (Mahe et al., 2007), and *Silene* (Minder &

Widmer, 2008). However, in the current study, mapped values of population membership proportions for *B. alleghaniensis* indicated relatively low and roughly equal proportions of putative admixture with *B. lenta* throughout the *B. alleghaniensis* range. Several *B. alleghaniensis* populations containing putative *B. lenta* alleles originated from far outside the zone of sympatry with *B. lenta*, which suggests that allele sharing in these populations may be more likely to have arisen from shared ancestral polymorphism than introgression. Furthermore, the absence of a strong pattern of isolation by distance between *B. alleghaniensis* and *B. lenta* suggests either that these species do not interbreed or that gene flow between the species is so efficient that the species might be essentially panmictic across the geographic scale investigated. However, a previous study indicated strong patterns of isolation by distance for populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* at nuclear microsatellite markers (Chapter 2), suggesting that gene flow is limited between distant populations. Thus, the lack of a significant pattern of isolation by distance between *B. alleghaniensis* and *B. lenta* in the current study may be suggestive of a lack of nuclear introgression between them. However, it is possible that occasional hybrids of *B. alleghaniensis* and *B. lenta* might exist, as hybridization at low frequencies is not necessarily associated with permanent nuclear gene capture (Baack & Rieseberg, 2007).

Though no significant pattern of isolation by distance existed between *B. alleghaniensis* and *B. papyrifera*, the finding of slightly higher rates of admixture in sympatric areas of the western Great Lakes region might indicate that allele sharing between these species has arisen from hybridization and introgression. Putative morphological hybrids between *B. alleghaniensis* and *B. papyrifera* have been identified

from localities in the sympatric zone in Minnesota, Wisconsin, Michigan, Iowa, and New Hampshire (Clausen, 1973, 1977; Barnes et al., 1974) and it has been suggested that hybrids might occur somewhat frequently on disturbed sites where the species overlap (Barnes et al., 1974; Clausen, 1977). It appears that flowering times of some individuals of *B. alleghaniensis* overlap sufficiently with pollen shed of *B. papyrifera* that some hybrid individuals might be formed (Clausen, 1973; Barnes et al., 1974). Putative morphological hybrids of yellow x paper birch have been found to occur most frequently in provenances from the extreme western portion of the *B. alleghaniensis* range (Clausen, 1973). Interestingly, a relatively higher level of putative admixture between *B. alleghaniensis* and *B. papyrifera* also was found in the western range of *B. alleghaniensis* in the current study. The finding of presumed nuclear introgression between *B. alleghaniensis* and *B. papyrifera* in this study is supported by a previous study which found a strong pattern of cpDNA haplotype sharing arising from introgression during postglacial recolonization (Chapter 1). However, the relatively low frequency of admixed individuals in the current study indicates that hybridization between *B. alleghaniensis* and *B. papyrifera* might only occur AT low frequency.

In spite of a high percentage of allele-sharing between *B. papyrifera* and *B. alleghaniensis* (69.1%), the species were found to be clearly differentiated at nuclear microsatellite markers as indicated by AMOVA and PCA analyses. Muir & Schlötterer (2005) propose three explanations for the finding of a high level of allele-sharing among closely-related but distinct species; (i) gene flow occurs at a sufficiently low rate that species-specific alleles are not generated and drift results in different allele frequencies between the species, (ii) selection maintains species integrity despite gene flow between

species (iii) shared ancestral polymorphism results in allele-sharing among species assuming a recent speciation event and large effective population sizes. In the current study, it seems possible that low levels of gene flow might be responsible for the high levels of allele-sharing observed between *B. alleghaniensis* and *B. papyrifera*. While selection and shared ancestral polymorphism cannot be ruled out, it seems unlikely that selection/gene flow should be responsible for the observed pattern since species differentiation should be low at selectively neutral markers if gene flow was occurring. Shared ancestral polymorphism, on the other hand, could be a contributing factor. Both *B. alleghaniensis* and *B. papyrifera* are suspected allopolyploids and it has been suggested that at least one of the progenitors of both species might have been a member of Subgenus *Betula* (DeJong, 1993; Jarvinen et al., 2004), and so it is possible that *B. alleghaniensis* and *B. papyrifera* might share a common progenitor species. However, phylogenetic studies generally indicate a distant relationship of *B. alleghaniensis* and *B. papyrifera* (Jarvinen et al., 2004; Li et al., 2005; Schenk et al., 2008) so there is currently little support for the inference of a shared progenitor. Thus, it seems likely that a low level of introgression is the most suitable explanation for the high level of allele-sharing of *B. alleghaniensis* and *B. papyrifera*.

In comparison to a previous study based on cpDNA markers (Chapter 1), this study suggests a much lower level of introgression between *B. papyrifera* and *B. alleghaniensis* at nuclear DNA markers. The previous study found widespread and strongly geographically structured patterns of local haplotype sharing between *B. papyrifera* and *B alleghaniensis*. In comparison, this study found only low levels of admixture between the two species, and the potential zone of introgression was

concentrated in the western Great Lakes Region. The difference between these studies is probably best explained in terms of differences in historic and contemporary introgression. In the previous study, the strong pattern of local haplotype sharing was best explained in terms of historic introgression suggesting that relatively high levels of introgression may have occurred between *B. papyrifera* and *B alleghaniensis* within refugia or during postglacial recolonization. However, the current study indicates low levels of admixture suggesting that contemporary gene flow between species is low. Relatively low rates of nuclear introgression in spite of high rates of chloroplast DNA introgression have been documented in numerous plant species groups; for example oaks (Whittemore & Schaal, 1991; Muir & Schlötterer, 2005) and asters (Baack & Rieseberg, 2007), and has in some cases been explained in terms of interspecific pollen competition (Rieseberg et al., 1995) or pollen-swamping during postglacial migration (Petit et al., 2003b). Since it seems likely that postglacial pollen-swamping may have occurred between *B. papyrifera* and *B. alleghaniensis* (Chapter 1), it is probable that historic introgression also might be responsible for some of the observed sharing of nuclear microsatellite alleles.

Future research should be directed at gaining a better understanding of the origin of allele-sharing between *B. papyrifera* and *B. alleghaniensis*, since this study cannot definitively suggest whether allele sharing is due to ancestral polymorphism or introgression. Genomics-based approaches might be especially useful in this regard (Baack & Rieseberg, 2007). Alternatively, molecular and cytological studies could be used to confirm the hybridity of putatively admixed individuals (e.g. Thórsson et al. 2001). Also, greater sampling of populations in the potential hybrid zone of *B. papyrifera*

and *B. alleghaniensis* is required, as this might potentially reveal stronger patterns of localized introgression. Future studies might also include additional co-distributed *Betula* species as this could reveal patterns of introgression with other species and also might help to elucidate phylogenetic relationships among the eastern North American birches.

**Conclusion**

Despite moderate to high allele-sharing between the species, *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* were clearly genetically differentiated. The greatest differentiation was found between *B. papyrifera* and *B. lenta*, and this can be explained in terms of potential genomic incompatibilities as well as their generally non-overlapping geographic distributions. *B. alleghaniensis* and *B. lenta* demonstrated a moderate number of putatively admixed individuals. However, the geographic pattern was more consistent with allele-sharing from ancestral polymorphism than what would be expected under a scenario of introgression. *Betula alleghaniensis* and *B. papyrifera* demonstrated the highest level of allele-sharing. Some evidence of genetic admixture between *B. alleghaniensis* and *B. papyrifera* was found in the western Great Lakes area, which is a suspected zone of natural hybridization based on previous morphological studies. The high level of allele-sharing of *B. alleghaniensis* and *B. papyrifera* might be explained in terms of low levels of contemporary introgression, ancestral polymorphism, historic introgression, or a combination of these factors.

## References


Anamthawat-Jónsson K. & Thórsson A.T. (2003) Natural hybridisation in birch: triploid hybrids between *Betula nana* and *B. pubescens*. *Plant Cell, Tissue, and Organ Culture*, **75**, 99–107.

Baack E.J. & Rieseberg L.H. (2007) A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, **17**, 513–518.

Barnes B.V., Dancik B.P., & Sharik T.L. (1974) Natural hybridization of yellow birch and paper birch. *Forest Science*, **20**, 215–221.

Boecklen W.J. & Howard D.J. (1997) Genetic analysis of hybrid zones: Number of markers and power of resolution. *Ecology*, **78**, 2611–2616.

Bousquet J., Cheliak W.M., Wang J., & Lalonde M. (1990) Genetic divergence and introgressive hybridization between *Alnus sinuata* and *A. crispa* (*Betulaceae*). *Plant Systematics and Evolution*, **170**, 107–124.

Bruvo R., Michiels N., D'Souza T., & Schulenburg H. (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*, **13**, 2101–2106.

Clark L. V & Jasieniuk M. (2011) POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources*, **11**, 562–566.

Clausen K.E. (1966) Studies of compatibility in *Betula*. *Joint Proceedings of the Second Genetics Workshop of the Society of American Foresters and the Seventh Lake States Forest Tree Improvement Conference* pp. 48–52. U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station, St. Paul, MN.

Clausen K.E. (1973) *Genetics of Yellow Birch. USDA Forest Service Research Paper WO-18.* U.S. Department of Agriculture, Washington, DC.

Clausen K.E. (1977) The yellow x paper birch hybrid - a potential substitute for yellow birch on problem sites. *Proceedings of the Thirteenth Lake States Forest Tree Improvement Conference; Gen. Tech. Rep. NC-50.* pp. 166–171. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station., St. Paul, MN.

DeJong P.C. (1993) An introduction to *Betula*: its morphology, evolution, classification and distribution, with a survey of recent work. *Betula: Proceedings of the IDS Betula Symposium* (ed. by D. Hunt), pp. 7–18. Richmond, England.

Dodd R.S. & Afzal-Rafii Z. (2004) Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, **58**, 261–269.

Dumolin-Lapègue S., Demesure B., Fineschi S., Le Corre V., & Petit R.J. (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics*, **146**, 1475–1487.

Erdmann G.G. (1990) Yellow birch (*Betula alleghaniensis* Britton). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 133–147. U.S. Department of Agriculture Forest Service, Washington, DC.

Evanno G., Regnaut S., & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L., Smouse P.E., & Quattro J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D., Stephens M., & Pritchard J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Fehrer J., Gemeinholzer B., Chrtek J., & Bräutigam S. (2007) Incongruent plastid and nuclear DNA phylogenies reveal ancient intergeneric hybridization in *Pilosella* hawkweeds (*Hieracium*, Cichorieae, Asteraceae). *Molecular Phylogenetics and Evolution*, **42**, 347–361.

Fernandez-Manjarres J.F., Gerard P.R., Dufour J., Raquin C., & Frascaria-Lacoste N. (2006) Differential patterns of morphological and molecular hybridization between *Fraxinus excelsior* L. and *Fraxinus angustifolia* Vahl (*Oleaceae*) in eastern and western France. *Molecular Ecology*, **15**, 3245–3257.

Furlow J.J. (1990) The genera of Betulaceae in the southeastern United States. *Journal of the Arnold Arboretum*, **71**, 1–67.

Furlow J.J. (1997) Betulaceae. *Flora of North America North of Mexico. Volume 3* (ed. by E. Flora of North America Editorial Committee), pp. 507–538. Oxford University Press, New York.

Gerard P.R., Fernandez-Manjarres J.F., & Frascaria-Lacoste N. (2006) Temporal cline in a hybrid zone population between *Fraxinus excelsior* L. and *Fraxinus angustifolia* Vahl. *Molecular Ecology*, **15**, 3655–3667.

Gilbert A.M. (1960) Silvical characteristics of yellow birch (*Betula alleghaniensis*). Station Paper No. 134. U.S. Department of Agriculture Forest Service, North Central Forest Experiment Station, Upper Darby, PA.

Grant W.F. & Thompson B.K. (1975) Observations on Canadian birches, *Betula cordifolia*, *B. neoalaskana*, *B. populifolia*, *B. papyrifera*, and *B. x caerulea*. *Canadian Journal of Botany*, **53**, 1478–1490.

Hardy O.J., Charbonnel N., Fréville H., & Heuertz M. (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

Hardy O.J. & 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.

Heuertz M., Carnevale S., Fineschi S., Sebastiani F., Hausman J.F., Paule L., & Vendramin G.G. (2006) Chloroplast DNA phylogeography of European ashes, *Fraxinus* sp. (*Oleaceae*): roles of hybridization and life history traits. *Molecular Ecology*, **15**, 2131–2140.

Hoban S.M., McCleary T.S., Schlarbaum S.E., & Romero-Severson J. (2009) Geographically extensive hybridization between the forest trees American butternut and Japanese walnut. *Biology Letters*, **5**, 324–327.

Jarvinen P., Palme A., Morales L.O., Lannenpaa M., Keinanen M., Sopanen T., & Lascoux M. (2004) Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast matK sequences. *American Journal of Botany*, **91**, 1834–1845.

Keim P., Paige N., Whitham G., & Lark K.G. (1989) Genetic analysis of an interspecific hybrid swarm of *Populus*: occurrence of unidirectional introgression. *Genetics*, **123**, 557–565.

King R.A. & Ferris C. (2000) Chloroplast DNA and nuclear DNA variation in the sympatric alder species, *Alnus cordata* (Lois.) Duby and *A. glutinosa* (L.) Gaertn. *Biological Journal of the Linnean Society*, **70**, 147–160.

Kulju K.K.M., Pekkinen M., & Varvio S. (2004) Twenty-three microsatellite primer pairs for *Betula pendula* (Betulaceae). *Molecular Ecology Notes*, **4**, 471–473.

Lamson N.I. (1990) Sweet birch (*Betula lenta* L.). *Silvics of North America: Volume 2. Hardwoods. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 148–152. United States Department of Agriculture Forest Service., Washington, DC.

Lepais O. & Gerber S. (2011) Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. *Evolution*, **65**, 156–170.

Li J., Shoup S., & Chen Z. (2005) Phylogenetics of *Betula* (Betulaceae) inferred from sequences of nuclear ribosomal DNA. *Rhodora*, **107**, 69–86.

Liu Y., Liu Y., & Huang H. (2010) Genetic variation and natural hybridization among sympatric *Actinidia* species and the implications for introgression breeding of kiwifruit. *Tree Genetics & Genomes*, **6**, 801–813.

Lynch M. (1990) The similarity index and DNA fingerprinting. *Molecular Biology and Evolution*, **7**, 478–484.

Mahe L., LePierres D., Combes M.C., & Lashermes P. (2007) Introgressive hybridization between the allotetraploid *Coffea arabica* and one of its diploid ancestors, *Coffea canephora*, in an exceptional sympatric zone in New Caledonia. *Genome*, **50**, 316–324.

Maliouchenko O., Palmé A.E., Buonamici A., Vendramin G.G., & Lascoux M. (2007) Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography*, **34**, 1601–1610.

Mallet J. (2005) Hybridization as an invasion of the genome. *Trends in ecology & evolution*, **20**, 229–37.

Martinsen G.D., Whitham T.G., Turek R.J., & Keim P. (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

Minder A.M. & Widmer A. (2008) A population genomic analysis of species boundaries: neutral processes, adaptive divergence and introgression between two hybridizing plant species. *Molecular Ecology*, **17**, 1552–63.

Muir G. & Schlötterer C. (2005) Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology*, **14**, 549–561.

Münzbergová Z., Surinová M., & Castro S. (2013) Absence of gene flow between diploids and hexaploids of *Aster amellus* at multiple spatial scales. *Heredity*, **110**, 123–130.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nielsen R., Tarpy D.R., & Reeve H.K. (2003) Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology*, **12**, 3157–3164.

Oberprieler C., Barth A., Schwarz S., & Heilmann J. (2010) Morphological and phytochemical variation, genetic structure and phenology in an introgressive hybrid swarm of *Senecio hercynicus* and *S. ovatus* (Compositae, Senecioneae). *Plant Systematics and Evolution*, **286**, 153–166.

Palme A.E., Su Q., Palsson S., & Lascoux M. (2004) Extensive sharing of chloroplast haplotypes among European birches indicates hybridization among *Betula pendula*, *B. pubescens* and *B. nana*. *Molecular Ecology*, **13**, 167–178.

Peakall R. & Smouse P.E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Petit R.J., Bodenes C., Ducousso A., Roussel G., & Kremer A. (2003) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Pritchard J.K., Stephens M., & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rieseberg L.H. (1995) The role of hybridization in evolution: old wine in new skins. *American Journal of Botany*, **82**, 944–953.

Rieseberg L.H. (1997) Hybrid origins of plant species. *Annual Review of Ecology and Systematics*, **28**, 359–389.

Rieseberg L.H., Choi H.C., & Ham D. (1991) Differential cytoplasmic versus nuclear introgression in *Helianthus*. *The Journal of Heredity*, **82**, 489–493.

Rieseberg L.H., Desrochers A.M., & Youn S.J. (1995) Interspecific pollen competition as a reproductive barrier between sympatric species of *Helianthus* (Asteraceae). *American Journal of Botany*, **82**, 515–519.

Rieseberg L.H. & Ellstrand N.C. (1993) What can molecular and morphological markers tell us about plant hybridization? *Critical Reviews in Plant Sciences*, **12**, 213–241.

Rieseberg L.H. & Soltis D.E. (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, **5**, 65–84.

Rieseberg L.H. & Willis J.H. (2007) Plant speciation. *Science*, **317**, 910–914.

Robertson A., Rich T.C.G., Allen A.M., Houston L., Roberts C., Bridle J.R., Harris S.A., & Hiscock S.J. (2010) Hybridization and polyploidy as drivers of continuing evolution and speciation in *Sorbus*. *Molecular Ecology*, **19**, 1675–1690.

Saeki I., Dick C.W., Barnes B.V., & Murakami N. (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): impacts of

habitat specialization, hybridization and glacial history. *Journal of Biogeography*, **38**, 992–1005.

Safford L.O., Bjorkbom J.C., & Zasada J.C. (1990) Paper birch (*Betula papyrifera* Marsh.). *Silvics of North America. Agriculture Handbook 654.* (ed. by R.M. Burns and B.H. Honkola), pp. 158–171. United States Department of Agriculture, Washington, DC.

Schenk M.F., Thienpont C.-L., Koopman W.J.M., Gilissen L.J.W.J., & Smulders M.J.M. (2008) Phylogenetic relationships in *Betula* (Betulaceae) based on AFLP markers. *Tree Genetics & Genomes*, **4**, 911–924.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 1–2.

Seehausen O., Takimoto G., Roy D., & Jokela J. (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.

Sharik T.L. & Barnes B.V. (1971) Hybridization in *Betula alleghaniensis* Britt. and *B. lenta* L.: A comparative analysis of controlled crosses. *Forest Science*, **17**, 415–424.

Sharik T.L. & Ford R.H. (1984) Variation and Taxonomy of *Betula uber*, *B. lenta*, and *B. alleghaniensis*. *Brittonia*, **36**, 307–316.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Thomasset M., Fernández-Manjarrés J.F., Douglas G.C., Bertolino P., Frascaria-Lacoste N., & Hodkinson T.R. (2012) Assignment testing reveals multiple introduced source populations including potential ash hybrids (*Fraxinus excelsior* × *F. angustifolia*) in Ireland. *European Journal of Forest Research*, **132**, 195–209.

Thórsson A.T., Pálsson S., Lascoux M., & Anamthawat-Jónsson K. (2010) Introgression and phylogeography of Betula nana (diploid), B. pubescens (tetraploid) and their triploid hybrids in Iceland inferred from cpDNA haplotype variation. *Journal of Biogeography*, **37**, 2098–2110.

Thórsson A.T., Salmela E., & Anamthawat-Jónsson K. (2001) Morphological, cytogenetic, and molecular evidence for introgressive hybridization in birch. *The Journal of Heredity*, **92**, 404–408.

Truong C., Palme a. E., Felber F., & Naciri-Graven Y. (2005) Isolation and characterization of microsatellite markers in the tetraploid birch, *Betula pubescens* ssp. *tortuosa*. *Molecular Ecology Notes*, **5**, 96–98.

Tsuda Y., Ueno S., Ide Y., & Tsumura Y. (2008) Development of 14 EST-SSRs for *Betula maximowicziana* and their applicability to related species. *Conservation Genetics*, **10**, 661–664.

Valbuena-Carabaña M., González-Martínez S.C., Sork V.L., Collada C., Soto A., Goicoechea P.G., & Gil L. (2005) Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity*, **95**, 457–465.

Weir B.S. & Cockerham C.C. (1984) Estimating F-Statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whittemore A.T. & Schaal B.A. (1991) Interspecific gene flow in sympatric oaks. *Evolution*, **88**, 2540–2544.

Wolf D.E., Takebayashi N., & Rieseberg L.H. (2001) Predicting the risk of extinction through hybridization. *Conservation Biology*, **15**, 1039–1053.

Wolfe K.H., Li W.H., & Sharp P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 9054–9058.

Wu B., Lian C., & Hogetsu T. (2002) Development of microsatellite markers in white birch (*Betula platyphylla* var. *japonica*). *Molecular Ecology Notes*, **2**, 413–415.

Ye T.Z., Yang R., & Yeh F.C. (2002) Population structure of a lodgepole pine (*Pinus contorta*) and jack pine (*P . banksiana*) complex as revealed by random amplified polymorphic DNA. *Genome*, **540**, 530–540.

Zeng J., Zou Y.-P., Bai J.-Y., & Zheng H.-S. (2002) Preparation of total DNA from "recalcitrant plant taxa". *Acta Botanica Sinica*, **44**, 694–697.

# GENERAL CONCLUSION

The objective of this study was to examine the phylogeography, population genetics, introgression and genetic structure of birches of eastern North America, with a focus on the widespread and ecologically and economically important species *Betula alleghaniensis*, *B. papyrifera*, and *B. lenta*.

Chapter 1 examined the phylogeography of eastern North American birches to infer the glacial history and introgression of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* using chloroplast DNA markers. Eastern and western populations of *B. papyrifera* and *B. alleghaniensis* composed different phylogeographic groups, suggesting that they originated from separate glacial refugia located in the Appalachians and western United States, respectively. The greatest haplotype diversity for the two species occurred in the Great Lakes Region, which is a postulated post-glacial biogeographic contact zone. In contrast, the Appalachian-endemic *B. lenta* was fixed for a single haplotype, and this can probably be explained by its relatively restricted geographic distribution and low cold-tolerance, which could have contributed to severe LGM population bottlenecks. *Betula papyrifera* and *B. alleghaniensis* demonstrated extensive sharing of localized haplotypes suggesting a high level of past introgression, whereas *B. lenta* demonstrated little evidence of introgression with the polyploid birches.

Chapter 2 examined the population genetics of *B. alleghaniensis*, *B. papyrifera* and *B. lenta* at nuclear microsatellite markers. Populations did not appear to be genetically subdivided and were significantly differentiated only over very great distances due to isolation by distance. High genetic diversity of *B. alleghaniensis* and *B. papyrifera* in the western Great Lakes region was presumably the result of admixture

between eastern and western glacial lineages. The northward declining genetic diversity of *B. lenta* probably originated due to founding events during postglacial recolonization. The overall low levels of population genetic structure and lack of population genetic subdivision observed for *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* can be attributed to high rates of gene flow between populations facilitated by efficient wind-dispersal of their light seeds and pollen. While not specifically designed for conservation purposes, the current study indicates that *B. lenta* should probably be designated as a species of high priority for conservation under climate change, while *B. alleghaniensis* and *B. papyrifera* may possess a high level of resilience.

Chapter 3 examined the genetic structure and incidence of introgression among *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* at nuclear microsatellite loci. The species were strongly differentiated despite moderate to high levels of allele-sharing between them. Some evidence of genetic admixture was found between *B. alleghaniensis* and *B. lenta*, but the geographic pattern was more suggestive of allele-sharing due to shared ancestry rather than introgression. In contrast, *B. papyrifera* and *B. lenta* demonstrated little evidence of genetic admixture, which can probably be explained in terms of their relatively distant phylogenetic relationship and generally non-overlapping distributions which presents little opportunity for interbreeding. Some evidence of genetic admixture between *B. alleghaniensis* and *B. papyrifera* was found in the western Great Lakes area, which is a suspected zone of natural hybridization based on previous morphological studies. However, tests of isolation by distance were not significant which makes it unclear whether allele-sharing was due to introgression or ancestral polymorphism. Future studies should be directed at gaining a better understanding of the origin of allele-

sharing between *B. papyrifera* and *B. alleghaniensis*. Genomics-based approaches in combination with morphological and cytological studies could be used to confirm the hybridity of putatively admixed individuals. Also, greater sampling of putatively admixed populations in the potential western Great Lakes hybrid zone is required, as this might potentially reveal stronger patterns of localized introgression.

The results clearly show that patterns of population structure in eastern North American *Betula* are strongly shaped by glacial population isolation as well as by high contemporary gene flow and historic introgression. High levels of genetic diversity observed in *B. papyrifera* and *B. alleghaniensis* at chloroplast and nuclear DNA markers are explained in terms of their relatively broad glacial and contemporary geographic distributions, as well as their polyploid origins. In contrast, low chloroplast and nuclear DNA diversity observed for *B. lenta* is probably best explained by extreme past population bottlenecks, which would have been contributed to by its relatively narrow glacial distribution and low cold-tolerance. Past isolation within glacial refuges has contributed to the strong population differentiation observed at cpDNA markers for *B. papyrifera* and *B. alleghaniensis*. In contrast, all species showed low levels of population structuring at nuclear markers, and this can probably be explained in terms of high contemporary gene flow facilitated by the efficient dispersal of both pollen and seeds. Introgression has clearly played a strong role in shaping patterns of variation at cpDNA markers, as evidenced by the widespread haplotype sharing of *B. alleghaniensis* and *B. papyrifera*. In contrast, there was little evidence for introgression of *B. lenta* with *B. alleghaniensis* or *B. papyrifera*. In contrast to what would be expected based on current understanding of phylogenetic relationships among the species, *B. lenta* appeared to be

the most differentiated, suggesting that this species has remained largely genetically isolated from *B. alleghaniensis* and *B. papyrifera*. Despite their relatively distant phylogenetic relationships, *B. papyrifera* and *B. alleghaniensis* shared a large proportion of alleles, which may be explained in terms of past introgression as previously mentioned, but might also reflect limited contemporary gene flow within a restricted area of the western Great Lakes Region.

Overall, this study improves our understanding of the factors that shape patterns of genetic diversity and population structure of temperate trees. Clearly, historic glaciations have left a profound signature on the genetic structure of temperate trees, and genetic signatures of historic glaciations are in many cases still evident within contemporary populations despite high rates of gene flow. Historic introgression may strongly contribute to phylogeographic structure at cpDNA markers. Although similar phylogeographic patterns often are interpreted as evidence for common biogeographic histories, this study demonstrates that shared phylogeographic structure can arise through introgression, even among morphologically distinct species. This study provides further evidence that high rates of cytoplasmic introgression are not always associated with high rates of nuclear introgression, as haplotype sharing may reflect introgression during postglacial recolonization and may not be indicative of contemporary interspecific gene flow at nuclear loci.

# APPENDICES

Appendix 1. Location, information, and Genbank accession numbers for *Betula* population samples used in Chapter 1.

(a) *Betula cordifolia*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (pbsA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Broad Cove | NS | 46.73 | -60.42 | 1 | 1 | KC816071 |
| Budworm City | NB | 47.63 | -66.60 | 1 | 1 | KC816072 |
| First Lake | NB | 47.65 | -68.37 | 1 | 1 | KC816073 |
| Gander | NL | 49.50 | -55.50 | 1 | 1 | KC816074 |
| Hawkes Bay | NL | 50.55 | -57.29 | 4 | 4 | KC816075-KC816078 |
| Napadogon | NB | 46.38 | -66.90 | 1 | 1 | KC816079 |
| Paquette Lake | NS | 46.83 | -60.43 | 2 | 2 | KC816080-KC816081 |
| Ravine Big Gulch | NB | 47.85 | -66.45 | 1 | 1 | KC816082 |
| St. George's | NL | 48.51 | -58.17 | 1 | 1 | KC816083 |
| West Quaco | NB | 45.33 | -65.53 | 3 | 3 | KC816084-KC816086 |

(b) *Betula populifolia*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (psbA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Penobscot | ME | 44.86 | -68.63 | 10 | 3 | KC816087-KC816089 |
| Massabesic | ME | 43.44 | -70.67 | 10 | 6 | KC816090-KC816095 |
| Gault | QC | 45.32 | -73.09 | 7 | 3 | KC816096-KC816098 |
| Hopkins | MA | 42.73 | -73.25 | 5 | 3 | KC816099-KC816101 |
| Afton Road | PE | 46.38 | -62.93 | 1 | 1 | KC816102 |
| Bishop Mountain | NS | 45.03 | -64.98 | 1 | 1 | KC816103 |
| Canoose | NB | 46.37 | -67.37 | 1 | 1 | KC816104 |
| Coates Mills | NB | 46.38 | -64.85 | 1 | 1 | KC816105 |
| Coles Island | NB | 45.90 | -65.72 | 1 | 1 | KC816106 |
| Memramcook | NB | 46.03 | -64.55 | 1 | 1 | KC816107 |
| Mount Albion | PE | 46.23 | -62.95 | 1 | 1 | KC816108 |
| Newmarket | NB | 45.81 | -66.96 | 10 | 3 | KC816109-KC816111 |
| Penobsquis | NB | 45.77 | -65.37 | 1 | 1 | KC816112 |
| Tracy | NB | 45.65 | -66.73 | 2 | 2 | KC816114 |
| Valleyfield | PE | 46.12 | -62.75 | 1 | 1 | KC816115 |
| Williamsdale | NS | 45.60 | -63.90 | 1 | 1 | KC816116 |

(c) *Betula lenta*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (psbA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Massabesic | ME | 43.44 | -70.67 | 10 | 3 | KC816117-KC816119 |
| Fernow | WV | 39.05 | -79.67 | 10 | 3 | KC816120-KC816122 |
| Kane | PA | 41.60 | -78.77 | 10 | 3 | KC816123-KC816125 |
| North River | TN | 35.33 | -85.09 | 10 | 3 | KC816126-KC816128 |
| Mt.Rogers | VA | 36.74 | -81.42 | 10 | 3 | KC816129-KC816131 |
| Glenwood | VA | 37.75 | -79.23 | 10 | 3 | KC816132-KC816134 |
| Finger Lakes | NY | 42.49 | -76.77 | 10 | 3 | KC816135-KC816137 |
| Hopkins | MA | 42.73 | -73.25 | 10 | 3 | KC816138-KC816140 |
| Pisgah | NC | 35.43 | -82.73 | 10 | 3 | KC816141-KC816143 |
| Luzerne | PA | 41.32 | -76.29 | 10 | 3 | KC816144-KC816146 |
| Cumberland | KY | 37.88 | -83.66 | 10 | 3 | KC816147-KC816149 |
| Hocking | OH | 39.47 | -82.58 | 10 | 3 | KC816150-KC816152 |

(d) *Betula uber*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (psbA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Mt. Rogers | VA | 36.74 | -81.42 | 10 | 10 | KC816153-KC816162 |

(e) *Betula papyrifera*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (psbA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Indian Brook | NS | 46.35 | -60.55 | 1 | 1 | KC816163 |
| Riverdale | PE | 46.23 | -63.37 | 10 | 3 | KC816164-KC816166 |
| Prosser Brook | NB | 45.80 | -64.88 | 10 | 3 | KC816167-KC816169 |
| Semiwagon | NB | 46.80 | -65.59 | 10 | 4 | KC816170-KC816173 |
| Duparquet | QC | 48.46 | -79.39 | 1 | 1 | KC816174 |
| Gallants | NL | 48.68 | -58.20 | 5 | 3 | KC816175-KC816177 |
| Adams Lake | BC | 51.43 | -119.83 | 1 | 1 | KC816178 |
| Grand Falls House | NL | 48.83 | -55.33 | 1 | 1 | KC816179 |
| Petawawa CNTSC | ON | 45.97 | -77.47 | 1 | 1 | KC816180 |
| Cheakumus River | BC | 50.00 | -123.17 | 1 | 1 | KC816181 |
| Beaver Lake | BC | 59.02 | -123.18 | 1 | 1 | KC816182 |
| McGraw Brook | NB | 46.87 | -66.20 | 1 | 1 | KC816183 |
| Dungarvon River | NB | 46.65 | -66.33 | 1 | 1 | KC816184 |
| Valleyfield | NB | 46.12 | -67.25 | 1 | 1 | KC816185 |
| Wayerton | NB | 47.22 | -65.93 | 1 | 1 | KC816186 |
| Tracy | NB | 45.65 | -66.73 | 1 | 1 | KC816187 |
| Pineville | NB | 46.80 | -65.92 | 1 | 1 | KC816188 |
| Oromocto Lake | NB | 45.70 | -66.65 | 1 | 1 | KC816189 |
| Lincoln | NB | 45.88 | -66.60 | 1 | 1 | KC816190 |
| Jewetts Creek | NB | 45.83 | -66.98 | 1 | 1 | KC816191 |
| Millvale | PE | 46.40 | -63.40 | 1 | 1 | KC816192 |
| Sixth Lake | NB | 45.73 | -67.47 | 1 | 1 | KC816193 |
| Mount Pleasant | NB | 45.42 | -66.83 | 1 | 1 | KC816194 |
| Salmon River | NB | 46.33 | -65.67 | 1 | 1 | KC816195 |
| Larch Hills Creek | BC | 50.41 | -119.08 | 1 | 1 | KC816196 |
| Aleza Lake | BC | 54.06 | -122.03 | 1 | 1 | KC816197 |
| Hooch Cr | BC | 49.13 | -117.20 | 1 | 1 | KC816198 |
| Beaver Lake | BC | 59.01 | -123.11 | 1 | 1 | KC816199 |
| Sardine Creek | BC | 52.47 | -122.14 | 1 | 1 | KC816200 |
| Lee Creek | BC | 50.56 | -119.32 | 1 | 1 | KC816201 |
| Skeena River | BC | 54.30 | -128.34 | 1 | 1 | KC816202 |
| Little Oliver Creek | BC | 54.48 | -128.16 | 1 | 1 | KC816203 |
| Burdick Creek | BC | 55.11 | -127.47 | 1 | 1 | KC816204 |
| Juniper Creek | BC | 55.08 | -127.43 | 1 | 1 | KC816205 |
| St. Mary River | BC | 49.38 | -116.03 | 1 | 1 | KC816206 |
| Wilson Creek | BC | 50.04 | -117.23 | 1 | 1 | KC816207 |
| Eaglet Lake | BC | 54.06 | -122.21 | 1 | 1 | KC816208 |
| Tabor Lake | BC | 53.55 | -122.22 | 1 | 1 | KC816209 |
| Amanita Lake | BC | 54.08 | -121.47 | 1 | 1 | KC816210 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Barnes Creek | BC | 50.34 | -118.50 | 1 | 1 | KC816211 |
| Mt. Tappen | BC | 50.45 | -119.20 | 1 | 1 | KC816212 |
| McConachie Cr | BC | 58.54 | -122.51 | 1 | 1 | KC816213 |
| Penobscot | ME | 44.86 | -68.63 | 10 | 3 | KC816214-KC816216 |
| Massabesic | ME | 43.44 | -70.67 | 10 | 3 | KC816217-KC816219 |
| Dukes | MI | 46.36 | -87.16 | 10 | 3 | KC816220-KC816222 |
| Argonne | WI | 45.75 | -88.98 | 10 | 3 | KC816223-KC816225 |
| Gault | QC | 45.32 | -73.09 | 10 | 3 | KC816226-KC816228 |
| Superior | ON | 47.34 | -84.57 | 10 | 3 | KC816229-KC816231 |
| Thessalon | ON | 46.27 | -83.42 | 5 | 3 | KC816232-KC816234 |
| Petawawa | ON | 46.18 | -78.07 | 5 | 3 | KC816235-KC816237 |
| Burchell | ON | 48.64 | -90.48 | 10 | 4 | KC816238-KC816241 |
| Athelstane | ON | 48.67 | -90.25 | 1 | 1 | KC816242 |
| Nipigon | ON | 49.04 | -88.00 | 10 | 3 | KC816243-KC816245 |
| Neys | ON | 48.78 | -86.59 | 10 | 3 | KC816246-KC816248 |
| White River | ON | 48.46 | -85.12 | 5 | 3 | KC816249-KC816251 |
| Wawa | ON | 47.91 | -84.54 | 5 | 3 | KC816252-KC816254 |
| Chapleau | ON | 47.52 | -83.22 | 5 | 3 | KC816255-KC816257 |
| Bartlett | NH | 44.05 | -71.30 | 10 | 3 | KC816258-KC816260 |
| Greenwood | ON | 48.39 | -90.75 | 10 | 3 | KC816261-KC816263 |
| Marcell | MN | 47.53 | -93.47 | 10 | 4 | KC816264-KC816267 |
| Blandin | MN | 47.12 | -93.68 | 10 | 3 | KC816268-KC816270 |
| Kickapoo | WI | 43.66 | -90.57 | 9 | 3 | KC81627-1KC816273 |
| Algonquin | ON | 45.58 | -79.21 | 5 | 3 | KC816274-KC816276 |
| Farm | ON | 44.20 | -77.49 | 8 | 3 | KC816277-KC816279 |
| Hopkins | MA | 42.73 | -73.25 | 10 | 3 | KC816280-KC816282 |

(f) *Betula alleghaniensis*

| Locality | Prov. /State | Lat. | Long. | N (cpSSR) | N (psbA-trnH) | Genbank Accession |
|---|---|---|---|---|---|---|
| Grundy State Park | TN | 35.20 | -85.70 | 1 | 1 | KC816283 |
| Falls Creek Falls | TN | 35.70 | -85.30 | 1 | 1 | KC816284 |
| Parsonfield | ME | 43.70 | -70.90 | 1 | 1 | KC816285 |
| Isle Royale | MI | 47.90 | -89.10 | 1 | 1 | KC816286 |
| English | IN | 38.30 | -86.50 | 1 | 1 | KC816287 |
| Blue Ridge Parkway | NC | 35.70 | -82.30 | 1 | 1 | KC816288 |
| Moran | MI | 45.90 | -84.80 | 1 | 1 | KC816289 |
| Delafield | WI | 43.10 | -88.40 | 1 | 1 | KC816290 |
| Minneopa State Park | MN | 44.20 | -94.10 | 1 | 1 | KC816291 |
| Virginia | MN | 47.60 | -92.50 | 1 | 1 | KC816292 |
| Grand Marais | MN | 47.80 | -90.20 | 1 | 1 | KC816293 |
| Reichuster Lake | WI | 46.50 | -92.10 | 1 | 1 | KC816294 |
| Williamstown | MA | 42.70 | -73.20 | 1 | 1 | KC816295 |
| Penobscot Exp. Forest | ME | 44.80 | -68.60 | 1 | 1 | KC816296 |
| South Richford | VT | 44.70 | -72.60 | 1 | 1 | KC816297 |
| Grand Detour | IL | 41.90 | -89.40 | 1 | 1 | KC816298 |
| Bartlett Exp. Forest | NH | 44.00 | -71.40 | 1 | 1 | KC816299 |
| Laconia | NH | 43.50 | -71.40 | 1 | 1 | KC816300 |
| Houghton | MI | 47.00 | -88.70 | 1 | 1 | KC816301 |
| Livingstonville | NY | 42.50 | -74.20 | 1 | 1 | KC816302 |
| L'Anse-a-Gilles | QC | 47.00 | -70.30 | 1 | 1 | KC816303 |
| La Tuque | QC | 47.40 | -72.60 | 1 | 1 | KC816304 |
| Lake Mitchinamecus | QC | 47.50 | -75.00 | 1 | 1 | KC816305 |
| Grande-Vallee | QC | 49.20 | -65.10 | 1 | 1 | KC816306 |
| Massanoga | ON | 45.10 | -76.90 | 1 | 1 | KC816307 |
| East Kemptville | NS | 44.10 | -65.80 | 1 | 1 | KC816308 |
| Guysborough | NS | 45.40 | -61.80 | 1 | 1 | KC816309 |
| Tabusintac River | NB | 47.40 | -65.20 | 1 | 1 | KC816310 |
| Fredericton | NB | 46.00 | -66.40 | 1 | 1 | KC816311 |
| Victoria | NS | 46.60 | -60.50 | 1 | 1 | KC816312 |
| Bridgetown | NS | 44.80 | -65.20 | 1 | 1 | KC816313 |
| Granton | WI | 44.50 | -90.40 | 1 | 1 | KC816314 |
| Argonne Exp. Forest | WI | 45.70 | -89.00 | 1 | 1 | KC816315 |
| Blue Ridge Mtns. | VA | 37.80 | -79.10 | 1 | 1 | KC816316 |
| Lake Provincial Park | ON | 47.50 | -84.80 | 1 | 1 | KC816317 |
| Kane Exp. Forest | PA | 41.60 | -78.70 | 1 | 1 | KC816318 |
| Penobscot | ME | 44.86 | -68.63 | 10 | 3 | KC816319-KC816321 |
| Massabesic | ME | 43.44 | -70.67 | 10 | 3 | KC816322-KC816324 |
| Fernow | WV | 39.05 | -79.67 | 10 | 3 | KC816325-KC816327 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Kane | PA | 41.60 | -78.77 | 10 | 3 | KC816328-KC816330 |
| Dukes | MI | 46.36 | -87.16 | 10 | 3 | KC816331-KC816333 |
| Argonne | WI | 45.75 | -88.98 | 10 | 3 | KC816334-KC816336 |
| Foster Falls | TN | 35.18 | -85.68 | 7 | 3 | KC816337-KC816339 |
| North River | TN | 35.33 | -85.09 | 10 | 3 | KC816340-KC816342 |
| Gault | QC | 45.32 | -73.09 | 10 | 3 | KC816343-KC816345 |
| Squaretop | ON | 48.27 | -89.40 | 5 | 3 | KC816346-KC816348 |
| Superior | ON | 47.34 | -84.57 | 10 | 3 | KC816349-KC816351 |
| Thessalon | ON | 46.27 | -83.42 | 5 | 3 | KC816352-KC816354 |
| Petawawa | ON | 46.18 | -78.07 | 1 | 1 | KC816355 |
| Mt.Rogers | VA | 36.74 | -81.42 | 10 | 3 | KC816356-KC816358 |
| Glenwood | VA | 37.75 | -79.23 | 10 | 3 | KC816359 -KC816361 |
| Bartlett | NH | 44.05 | -71.30 | 10 | 3 | KC816362-KC816364 |
| Greenwood | ON | 48.39 | -90.75 | 10 | 3 | KC816365-KC816367 |
| Marcell | MN | 47.53 | -93.47 | 9 | 3 | KC816368-KC816370 |
| Blandin | MN | 47.12 | -93.68 | 10 | 3 | KC816371-KC816373 |
| Kickapoo | WI | 43.66 | -90.57 | 10 | 5 | KC816374-KC816378 |
| Waterloo | MI | 42.33 | -84.10 | 3 | 3 | KC816379-KC816381 |
| Algonquin | ON | 45.58 | -79.21 | 6 | 3 | KC816382-KC816384 |
| Farm | ON | 44.20 | -77.49 | 6 | 3 | KC816385-KC816387 |
| Finger Lakes | NY | 42.49 | -76.77 | 10 | 3 | KC816388-KC816390 |
| Hopkins | MA | 42.73 | -73.25 | 10 | 3 | KC816391-KC816393 |
| Pisgah | NC | 35.43 | -82.73 | 10 | 3 | KC816394-KC816396 |
| Luzerne | PA | 41.32 | -76.29 | 10 | 3 | KC816397-KC816399 |
| Bay Despoir | NL | 47.85 | -55.70 | 1 | 1 | KC816400 |
| Bishop Mountain | NS | 45.03 | -64.98 | 1 | 1 | KC816401 |
| Black Brook | NB | 47.45 | -67.45 | 10 | 3 | KC816402-KC816404 |
| Canoose | NB | 46.37 | -67.37 | 1 | 1 | KC816405 |
| Cormac | ON | 45.47 | -77.30 | 1 | 1 | KC816406 |
| Dumbarton | NB | 45.35 | -67.13 | 1 | 1 | KC816407 |
| Grand Bay | NB | 45.27 | -66.22 | 1 | 1 | KC816408 |
| Martin Head | NB | 45.53 | -65.18 | 1 | 1 | KC816409 |
| McGraw Brook | NB | 46.87 | -66.20 | 1 | 1 | KC816410 |
| Oromocto Lake | NB | 45.70 | -66.65 | 1 | 1 | KC816411 |
| Pancake River | ON | 47.03 | -84.63 | 2 | 1 | KC816412 |
| Pembroke | ON | 45.82 | -77.12 | 2 | 1 | KC816413 |
| Petawawa NTSC | ON | 45.97 | -77.47 | 10 | 3 | KC816414-KC816416 |
| Prosser Brook | NB | 45.80 | -64.88 | 10 | 3 | KC816417-KC816419 |
| Tracy | NB | 45.65 | -66.73 | 1 | 1 | KC816420 |
| Valley | NB | 46.17 | -62.67 | 1 | 1 | KC816421 |
| Williamsdale | NS | 45.60 | -63.90 | 1 | 1 | KC816422 |

Appendix 2. Information and sequences of eight microsatellite primers used in Chapter 2.

| Primer Name | Working Code | Published Repeat | Forward Primer 5'-3' | Reverse Primer 5'-3' | Reference |
|---|---|---|---|---|---|
| CD277302 | BNP1 | $(TC)_{10}$ | ACCAAGTGAACTTACAAAGGCGA | GTTTCCTGCTATAAGGCCTGCACCT | Tsuda et al. 2008 |
| L1.10 | BNP2 | $(GA)_4AA\ (GA)_{10}$ | ACGCTTTCTTGATGTCAGCC | TCACCAAGTTCCTGGTGGAT | Kulju et al. 2004 |
| L2.3 | BNP3 | $(AG)_{16}$ | CAGTGTTTGGACGGTGAGAA | CGGGTGAAGTAGACGGAACT | Kulju et al. 2004 |
| L5.5 | BNP8 | $C_{12}CTCC(CT)_7$ | GAGGAAGTCTCAGCTGACGTG | TCCTTTTCAGTTTTCTGATTTCTG | Kulju et al. 2004 |
| L7.1a | BNP9 | $(CT)_{12}CCTT$ | GTTTTGGGTTTCCACTTCCA | ACTGGTAATACCTTTACCAAGCC | Kulju et al. 2004 |
| L5.4 | BNP18 | $(TC)_{26}$ | GAAAGCATGAGACCCGTCTT | AACCTAAACAGCCTGCCAAA | Truong et al. 2005 |
| Bo.F330 | BNP20 | $(TC)_{14}$ | TGGCAGCACGAAAGT | TGGGAATGAGAGAACAAG | Truong et al. 2005 |
| Bp04 | BNP29 | $(GT)_{12}$ … $(GA)_5$ | GGCAACCAGCAGCAATCTGAC | ATGCCCAAGGACGACTAGACC | Wu et al. 2002 |

… intervening sequence

Appendix 3. Summary of allele-size variants at eight microsatellite loci for *B. alleghaniensis*, B. *papyrifera* and *B. lenta.*

| Species | CD277302 | L1.10 | L2.3 | L5.5 | L7.1a | L5.4 | Bo.F330 | Bp04 |
|---|---|---|---|---|---|---|---|---|
| *all* | 226, 228, 230, 232, 234, 236, 238, 240, 242, 246, 248 | 159, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 203 | 188, 190, 192, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220 | 110, 112, 114, 116, 118, 120, 124, 130, 136, 142 | 158, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182 | 156, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180 | 192, 194, 196, 198, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216 | 160, 166, 168, 170, 172, 174, 176, 178, 182, 186, 188 |
| *pap* | 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248 | 171, 173, 175, 177, 179, 181, 183, 185, 187, 189 | 198, 200, 202, 206, 210, 212, 216, 218, 220 | 112, 114, 116, 118, 124, 130, 136, 142 | 158, 162, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178. 179, 180, 181 | 160, 162, 164, 166, 168, 174, 176, 180 | 192, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 209, 210, 211, 212, 213, 214, 215, 216 | 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188 |
| *len* | 230, 232, 234 | 175, 179 | 188, 198, 200, 202, 204, 210, 212 | 114, 116 | 162, 164, 166, 178, 180, 182, 194 | 158, 164, 166, 168, 170 | 192, 196, 198, 206 | 166, 168, 170 |

all, *alleghaniensis* ; pap, *papyrifera* ; len, *lenta*

Appendix 4. Total number of genotyped individuals (N), total number of alleles ($A_T$), allelic richness after rarefaction ($A_R$), observed heterozygosity ($H_O$), gene diversity ($H_E$), global F- and R-statistics, and tests of HW equilibrium for *B. lenta* at each locus.

| Locus | N | $A_T$ | $A_R$ | $H_O$ | $H_E$ | $F_{IS}$ | $F_{ST}$ | $R_{ST}$ | HW |
|---|---|---|---|---|---|---|---|---|---|
| CD277302 | 299 | 3 | 2.0 | 0.097 | 0.112 | 0.121 | 0.013 | 0.015 | * |
| L1.10 | 298 | 2 | 2.0 | 0.443 | 0.482 | 0.057 | 0.012 | 0.012 | |
| L2.3 | 298 | 7 | 4.3 | 0.531 | 0.521 | -0.011 | 0.006 | 0.010 | |
| L5.5 | 300 | 2 | 2.0 | 0.264 | 0.262 | -0.027 | 0.025 | 0.025 | |
| L7.1a | 296 | 7 | 3.7 | 0.620 | 0.661 | 0.037 | 0.021 | 0.036 | |
| L5.4 | 300 | 5 | 4.7 | 0.572 | 0.658 | 0.112 | 0.014 | 0.057 | * |
| Bo.F330 | 300 | 4 | 2.1 | 0.163 | 0.291 | 0.427 | 0.001 | -0.003 | * |
| Bp04 | 300 | 3 | 1.6 | 0.048 | 0.062 | 0.291 | 0.007 | 0.006 | * |

* significant departure from HW expectations at $P<0.05$

Appendix 5. Summary of genetic diversity parameters for individual populations of (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.

(a) *B. alleghaniensis*

| Population | N | $A_N$ | $A_R$ | $H_E$ |
|---|---|---|---|---|
| 101 | 30 | 8.6 | 7.8 | 0.77 |
| 102 | 30 | 8.6 | 8.0 | 0.75 |
| 103 | 30 | 7.3 | 6.9 | 0.75 |
| 104 | 29 | 8.0 | 7.4 | 0.75 |
| 105 | 35 | 9.4 | 8.5 | 0.80 |
| 106 | 37 | 9.3 | 8.2 | 0.78 |
| 109 | 29 | 7.9 | 7.5 | 0.76 |
| 110 | 30 | 8.3 | 7.6 | 0.76 |
| 111 | 30 | 7.6 | 7.1 | 0.75 |
| 112 | 31 | 8.4 | 7.6 | 0.76 |
| 113 | 28 | 9.1 | 8.5 | 0.80 |
| 115 | 20 | 9.6 | 9.3 | 0.82 |
| 121 | 33 | 8.9 | 8.0 | 0.78 |
| 122 | 31 | 7.9 | 7.3 | 0.77 |
| 123 | 30 | 8.0 | 7.5 | 0.77 |
| 1101 | 30 | 8.1 | 7.5 | 0.75 |

N, number sampled; $A_N$, average number of alleles; $A_R$, allelic richness after rarefaction (N=61); $H_E$, gene diversity

(b) *B. papyrifera*

| Population | N | $A_N$ | $A_R$ | $H_E$ |
|---|---|---|---|---|
| 101 | 30 | 8.2 | 7.5 | 0.73 |
| 102 | 29 | 7.3 | 7.1 | 0.72 |
| 106 | 31 | 7.3 | 6.9 | 0.71 |
| 112 | 31 | 7.0 | 6.6 | 0.73 |
| 113 | 25 | 7.2 | 7.0 | 0.74 |
| 114 | 28 | 7.0 | 6.7 | 0.72 |
| 115 | 30 | 7.3 | 7.0 | 0.74 |
| 121 | 24 | 7.3 | 7.2 | 0.71 |
| 1101 | 28 | 7.5 | 7.0 | 0.73 |

N, number sampled; $A_N$, average number of alleles; $A_R$, allelic richness after rarefaction (N=61); $H_E$, gene diversity

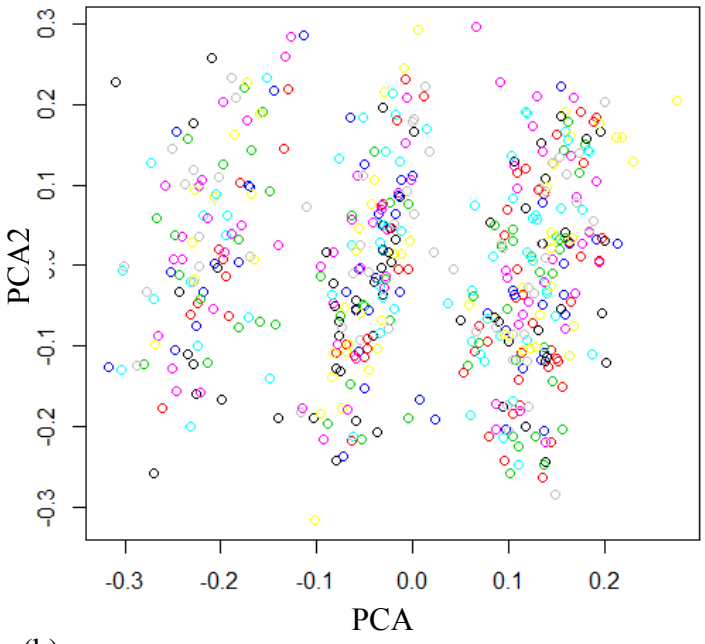
(c) *B. lenta*

| Population | N | $A_N$ | $A_R$ | $H_E$ | $F_{IS}$ |
|---|---|---|---|---|---|
| 102 | 20 | 2.6 | 2.5 | 0.35 | 0.116 |
| 103 | 26 | 3.0 | 2.7 | 0.42 | 0.125* |
| 104 | 30 | 3.1 | 2.8 | 0.38 | 0.040 |
| 109 | 20 | 3.1 | 3.0 | 0.40 | 0.091 |
| 110 | 28 | 2.9 | 2.7 | 0.37 | 0.072 |
| 111 | 29 | 3.1 | 2.7 | 0.38 | 0.083 |
| 120 | 19 | 3.1 | 3.0 | 0.38 | 0.073 |
| 121 | 28 | 2.9 | 2.5 | 0.34 | 0.117* |
| 122 | 27 | 3.0 | 2.7 | 0.37 | 0.132* |
| 123 | 31 | 2.8 | 2.5 | 0.33 | 0.040 |
| 124 | 27 | 3.3 | 2.9 | 0.43 | 0.053 |
| 125 | 15 | 2.9 | 2.8 | 0.37 | 0.173* |

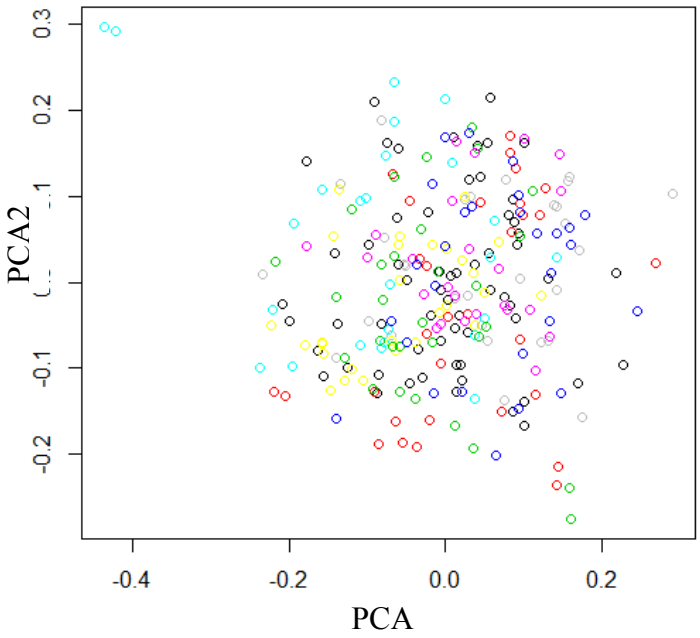
N, number sampled; $A_N$, average number of alleles; $A_R$, allelic richness after rarefaction; $H_E$, gene diversity; * significant deviation from HWE $P<0.05$

Appendix 6. Principal components analysis showing grouping of individuals from different populations based on Lynch genetic distance for (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*.

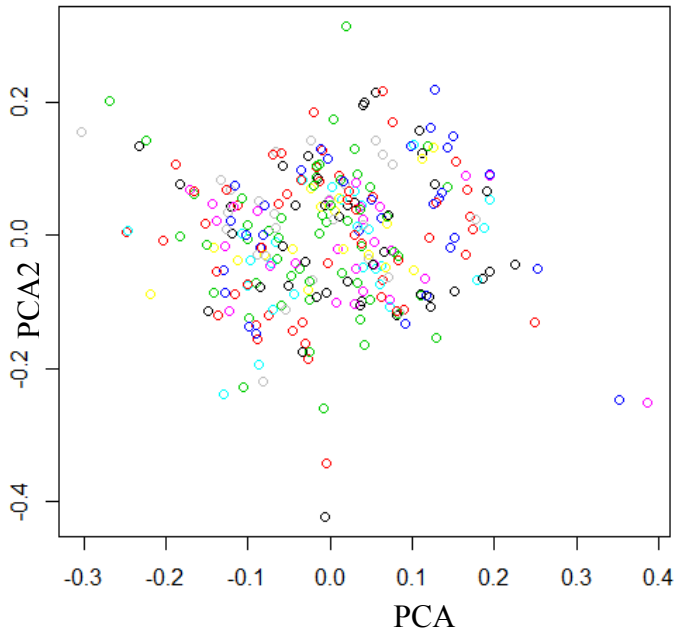
(a)




(b)



(c)



Appendix 7. Standard AMOVA summarizing the partitioning of molecular variation among and within populations of *B. alleghaniensis*, *B. papyrifera*, and *B. lenta* based on Lynch genetic distances.

| Species | Source of variation | df | Sum of squares | Variance components | Percent variation | $\Phi_{PT}$ | P |
|---|---|---|---|---|---|---|---|
| *B. alleghaniensis* | Among populations | 15 | 5.9 | 0.005 | 2.0 | 0.020 | 0.001 |
| | Within populations | 467 | 113.4 | 0.243 | 98.0 | | |
| *B. papyrifera* | Among populations | 8 | 3.3 | 0.007 | 3.6 | 0.036 | 0.001 |
| | Within populations | 247 | 49.4 | 0.200 | 96.4 | | |
| *B. lenta* | Among populations | 11 | 2.4 | 0.003 | 2.5 | 0.025 | 0.001 |
| | Within populations | 288 | 38.6 | 0.134 | 97.5 | | |

Appendix 8. Pairwise population differentiation values ($\Phi_{PT}$) for (a) *B. alleghaniensis*, (b) *B. papyrifera*, and (c) *B. lenta*. Pairwise $\Phi_{PT}$ values are shown below the diagonal and significant differences (P<0.05) are indicated above the diagonal.

(a) *B. alleghaniensis*

| Population | 101 | 102 | 103 | 104 | 105 | 106 | 109 | 110 | 111 | 112 | 113 | 115 | 121 | 122 | 123 | 1101 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 101 | 0.00 | | * | | | | * | * | * | * | * | | | * | * | * |
| 102 | 0.00 | 0.00 | * | | * | * | * | * | * | | * | * | | * | * | * |
| 103 | 0.02 | 0.01 | 0.00 | | * | * | * | * | * | | * | * | * | * | * | * |
| 104 | 0.01 | 0.01 | 0.01 | 0.00 | * | * | | | | | * | * | | * | | * |
| 105 | 0.01 | 0.01 | 0.02 | 0.02 | 0.00 | | * | * | * | * | * | | | * | * | * |
| 106 | 0.01 | 0.02 | 0.03 | 0.02 | 0.00 | 0.00 | * | * | * | * | * | | | * | * | * |
| 109 | 0.02 | 0.02 | 0.01 | 0.01 | 0.03 | 0.03 | 0.00 | | | | * | * | | | | * |
| 110 | 0.03 | 0.03 | 0.02 | 0.01 | 0.03 | 0.02 | 0.00 | 0.00 | | * | * | * | * | | * | * |
| 111 | 0.02 | 0.01 | 0.02 | 0.00 | 0.03 | 0.01 | 0.00 | 0.00 | 0.00 | | * | * | | | | * |
| 112 | 0.01 | 0.00 | 0.00 | 0.00 | 0.02 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | * | * | | * | | * |
| 113 | 0.01 | 0.04 | 0.05 | 0.05 | 0.01 | 0.02 | 0.05 | 0.04 | 0.05 | 0.05 | 0.00 | * | * | * | * | * |
| 115 | 0.01 | 0.03 | 0.05 | 0.03 | 0.00 | 0.00 | 0.04 | 0.04 | 0.03 | 0.03 | 0.02 | 0.00 | | * | * | * |
| 121 | 0.01 | 0.01 | 0.02 | 0.01 | 0.01 | 0.01 | 0.01 | 0.02 | 0.00 | 0.00 | 0.03 | 0.01 | 0.00 | * | | * |
| 122 | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 | 0.03 | 0.00 | 0.00 | 0.00 | 0.02 | 0.05 | 0.04 | 0.01 | 0.00 | * | * |
| 123 | 0.03 | 0.02 | 0.03 | 0.01 | 0.02 | 0.03 | 0.01 | 0.02 | 0.00 | 0.01 | 0.06 | 0.02 | 0.00 | 0.01 | 0.00 | * |
| 1101 | 0.03 | 0.01 | 0.03 | 0.01 | 0.02 | 0.03 | 0.02 | 0.04 | 0.03 | 0.03 | 0.05 | 0.03 | 0.02 | 0.04 | 0.02 | 0.00 |

ΦPT of -0.005 between pops by dukes and by argonne converted to zero

ΦPT of 0.000 between pops by kane and by glenwood converted to zero

ΦPT of -0.010 between pops by mt rogers and by glenwood converted to zero

ΦPT of -0.003 between pops by dukes and by blandin converted to zero

ΦPT of -0.002 between pops by mt rogers and by pisgah converted to zero

ΦPT of -0.003 between pops by glenwood and by pisgah converted to zero

(b) *B. papyrifera*

| Population | 101 | 102 | 106 | 112 | 113 | 114 | 115 | 121 | 1101 |
|---|---|---|---|---|---|---|---|---|---|
| 101 | 0.00 | | * | * | * | * | | | |
| 102 | 0.00 | 0.00 | * | * | * | * | * | | |
| 106 | 0.02 | 0.03 | 0.00 | * | | | * | * | * |
| 112 | 0.03 | 0.02 | 0.02 | 0.00 | * | * | * | * | * |
| 113 | 0.04 | 0.04 | 0.01 | 0.06 | 0.00 | | * | * | * |
| 114 | 0.03 | 0.04 | 0.01 | 0.05 | 0.01 | 0.00 | | * | * |
| 115 | 0.01 | 0.02 | 0.02 | 0.04 | 0.02 | 0.00 | 0.00 | * | * |
| 121 | 0.01 | 0.02 | 0.05 | 0.05 | 0.06 | 0.05 | 0.04 | 0.00 | * |
| 1101 | 0.00 | 0.01 | 0.03 | 0.04 | 0.05 | 0.04 | 0.02 | 0.03 | 0.00 |

(c) *B. lenta*

| Population | 102 | 103 | 104 | 109 | 110 | 111 | 120 | 121 | 122 | 123 | 124 | 125 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 102 | 0.000 | * | * | * | | * | | | | * | * | * |
| 103 | 0.073 | 0.000 | * | * | | | * | * | * | * | | |
| 104 | 0.046 | 0.030 | 0.000 | | | * | | * | * | | | |
| 109 | 0.051 | 0.050 | 0.008 | 0.000 | | | | * | | * | * | |
| 110 | 0.027 | 0.010 | 0.000 | 0.000 | 0.000 | | | * | | | | |
| 111 | 0.048 | 0.015 | 0.021 | 0.020 | 0.000 | 0.000 | | * | | * | | |
| 120 | 0.035 | 0.034 | 0.000 | 0.008 | 0.002 | 0.010 | 0.000 | | | | | |
| 121 | 0.027 | 0.071 | 0.034 | 0.048 | 0.036 | 0.059 | 0.008 | 0.000 | * | * | * | * |
| 122 | 0.027 | 0.032 | 0.022 | 0.000 | 0.000 | 0.000 | 0.000 | 0.037 | 0.000 | | * | |
| 123 | 0.028 | 0.062 | 0.012 | 0.030 | 0.013 | 0.028 | 0.000 | 0.040 | 0.012 | 0.000 | * | * |
| 124 | 0.042 | 0.000 | 0.008 | 0.033 | 0.000 | 0.014 | 0.008 | 0.042 | 0.028 | 0.040 | 0.000 | |
| 125 | 0.071 | 0.007 | 0.015 | 0.000 | 0.000 | 0.009 | 0.006 | 0.054 | 0.002 | 0.038 | 0.010 | 0.000 |

ΦPT of -0.002 between pops bl kane and bl mt rogers converted to zero

ΦPT of -0.002 between pops bl north river and bl mt rogers converted to zero

ΦPT of -0.016 between pops bl mt rogers and bl glenwood converted to zero

ΦPT of -0.022 between pops bl kane and bl finger lakes converted to zero

ΦPT of -0.009 between pops bl north river and bl pisgah converted to zero

ΦPT of -0.010 between pops bl mt rogers and bl pisgah converted to zero

ΦPT of -0.006 between pops bl glenwood and bl pisgah converted to zero

ΦPT of -0.004 between pops bl finger lakes and bl luzerne converted to zero

ΦPT of -0.016 between pops bl fernow and bl cumberland converted to zero

ΦPT of -0.003 between pops bl mt rogers and bl cumberland converted to zero

ΦPT of -0.002 between pops bl north river and bl hocking converted to zero

ΦPT of -0.011 between pops bl mt rogers and bl hocking converted to zero

Appendix 9. Information and sequences for six nuclear microsatellite primers used in Chapter 3.

| Primer Name | Published Repeat | Forward Primer 5'-3' | Reverse Primer 5'-3' | Reference |
|---|---|---|---|---|
| CD277302 | $(TC)_{10}$ | ACCAAGTGAACTTACAAAGGCGA | GTTTCCTGCTATAAGGCCTGCACCT | Tsuda et al. 2008 |
| L1.10 | $(GA)_4AA\ (GA)_{10}$ | ACGCTTTCTTGATGTCAGCC | TCACCAAGTTCCTGGTGGAT | Kulju et al. 2004 |
| L2.3 | $(AG)_{16}$ | CAGTGTTTGGACGGTGAGAA | CGGGTGAAGTAGACGGAACT | Kulju et al. 2004 |
| L5.5 | $C_{12}CTCC(CT)_7$ | GAGGAAGTCTCAGCTGACGTG | TCCTTTTCAGTTTTCTGATTTCTG | Kulju et al. 2004 |
| L5.4 | $(TC)_{26}$ | GAAAGCATGAGACCCGTCTT | AACCTAAACAGCCTGCCAAA | Truong et al. 2005 |
| Bp04 | $(GT)_{12}$ … $(GA)_5$ | GGCAACCAGCAGCAATCTGAC | ATGCCCAAGGACGACTAGACC | Wu et al. 2002 |

… intervening sequence

Appendix 10. Analyses of Molecular Variance (AMOVA) for the partitioning of molecular variation among species groupings based on Lynch distances (IAM).

| Species[1] Grouping | Source of variation | df | Sum of squares | Variance components | Percent variation | P-value |
|---|---|---|---|---|---|---|
| a) *all/pap/len* | Among species | 2 | 106.25 | 0.159 | 43.2 | <0.001 |
| | Among populations | 34 | 11.91 | 0.005 | 1.4 | <0.001 |
| | Within populations | 1002 | 203.99 | 0.204 | 55.4 | <0.001 |
| b) *all/pap* | Among species | 1 | 39.69 | 0.117 | 32.5 | <0.001 |
| | Among populations | 23 | 9.71 | 0.006 | 1.7 | <0.001 |
| | Within populations | 714 | 169.27 | 0.237 | 65.7 | <0.001 |
| c) *all/len* | Among species | 1 | 55.44 | 0.149 | 41.6 | <0.001 |
| | Among populations | 26 | 8.63 | 0.005 | 1.3 | <0.001 |
| | Within populations | 755 | 154.60 | 0.205 | 57.2 | <0.001 |
| d) *pap/len* | Among species | 1 | 67.61 | 0.244 | 60.0 | <0.001 |
| | Among populations | 19 | 5.48 | 0.005 | 1.2 | <0.001 |
| | Within populations | 535 | 84.11 | 0.157 | 38.7 | <0.001 |

1/ species abbreviations: all = *alleghaniensis* , pap = *papyrifera* , len = *lenta*