The Allopolyploid Origin and Population Genetics of the Rare *Solidago Houghtonii* (Asteraceae)

Pamela J. Laureto

*Western Michigan University*

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

**Recommended Citation**


https://scholarworks.wmich.edu/dissertations/605

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE ALLOPOLYPLOID ORIGIN AND POPULATION GENETICS OF THE RARE *SOLIDAGO HOUGHTONII* (ASTERACEAE)

by

Pamela J. Laureto

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Biological Sciences
Advisor: Todd J. Barkman, Ph.D.

Western Michigan University
Kalamazoo, Michigan
May 2010
Solidago houghtonii is a federally threatened hexaploid plant species of likely hybrid origin, limited geographic distribution, and high habitat specificity. An understanding of the evolutionary history and population genetic structure of rare species is critical to ensuring their long-term survival. To study the hybrid origin of *S. houghtonii* I sequenced four noncoding chloroplast (cp) DNA loci, and the nuclear ribosomal DNA ITS and 3' ETS regions from four individuals of *S. houghtonii*, which span its geographic range, and 25 other species of *Solidago* including all sympatric species. Polymorphisms within the nrDNA sequences indicated the presence of multiple homeologue types which were separated by molecular cloning. Phylogenetic analyses of cloned sequence data indicated that *S. riddellii*, *S. ptarmicoides*, and *S. ohioensis* are parents of *S. houghtonii*, however cpDNA sequence data indicated a fourth species, *S. gigantea*, as the maternal genome donor. These data reveal both a single origin and a complex pattern of reticulation that is consistent not only with the hypothesized allohexaploid nature of this species, but also with chloroplast capture of cpDNA from an unexpected source through introgression.

The amount of genetic diversity and the degree of population subdivision was analyzed in 23 *S. houghtonii* populations spanning the geographical range and habitat
specificities of the species. 452 individual plants were characterized for cpDNA haplotype by screening amplicons of three noncoding cpDNA regions for length polymorphism. A total of 14 haplotypes were detected. Gene diversity ($H_E$) across populations ranged from 0.00 – 0.59 and a high degree of population structure was found ($F_{ST} = 0.76$). An AMOVA partitioned most of the variation among populations (76%), and the remainder among individuals within populations (24%). Similar structure was found when the populations were grouped according to either geographic region or substrate type. A Mantel test found significant isolation-by-distance between all populations except the most disjunct population ($P = 0.003$). A statistical parsimony network analysis of the 14 $S. houghtonii$ haplotypes indicated a northern Michigan origin for the species. The genetic structure of $S. houghtonii$ populations is consistent with postglacial range expansion and contraction during the hypsithermal period $\sim 6000 \text{YBP}$.
ACKNOWLEDGMENTS

This work would not have been completed were it not for the assistance of numerous people. My sincere gratitude goes to my major advisor, Dr. Todd Barkman who welcomed me as a non-traditional student into his lab. You were an excellent Ph.D. supervisor who provided encouragement and guidance along the way. Thanks for your willingness to discuss my work, even when I interrupted yours. I would like to thank the other members of my graduate committee, Drs. David Cowan, Stephen Malcolm, Aaton Reznicek (University of Michigan), and Maarten Vonhof for helpful conversations and continued support throughout this project. Special thanks go to Dave Cowan for materials when supplies ran short, they kept the work going.

I also thank the staff, students, and volunteers from both the Barkman and Cowan labs who have contributed to this work. I could not have completed it without the assistance and encouragement of Cathy Foune, Talline Martin, Seok-Hong Lim, Nabeeh Hasan, Aida Shafreena Ahmad Puad, and numerous undergraduates who spent hours helping with the DNA extractions.

Mike Penskar of the Michigan Natural Features Inventory, Steve Young of the New York Natural Features Inventory, and Mike Oldham of the Canadian Ministry of Natural Resources are thanked for providing locality information, assistance in securing collection permits, and helpful discussions regarding Solidago houghtonii.

The Bergen Swamp Preservation Society, Mr. Clifford Fielding, the Whitefish River First Nation, the Nature Conservancy of Canada, the Ontario Ministry of
Natural Resources, the US Fish and Wildlife Service and the Michigan Department of Natural Resources are thanked for permission to collect on their properties, and Michael Schuldman and Craig Thomsen are thanked for field collections of *Chrysothamnus* and *Solidago velutina*, respectively.

Dr. Edward G. Voss (University of Michigan) first introduced me to *Solidago houghtonii* and its fascinating shoreline habitat in the summer of 1992. Thanks for letting me into your summer course those many years ago and for planting the seed which grew into this research project.

My gratitude goes to Dr. James S. Pringle (Royal Botanical Garden, Hamilton, Ontario, Canada) for sharing details of his extensive morphological and cytological work on *Solidago houghtonii* including his hypotheses regarding parentage.

Special thanks go to two very special friends, first, James D. Doyle, PhD, who roused my interest in botany and fostered my love of plants. You’ve been a great mentor and friend. Thanks for helping with the New York field work; I couldn’t have done it without you. Second, to Peg Haviland who traversed dune and stream, spent a week in a tent, and overcame bad tunafish sandwiches to help with fieldwork. Thanks for sticking it out with me.

Most importantly, I’d like to thank my family. To my husband and best friend, Tom Laureto, you withstood my bad research days, long hours away from home, and endless conversation about PCR, DNA sequencing, and fragment analysis. Your love and support has been invaluable and your company on collecting trips and
Acknowledgments—continued

in the lab has kept me going. And to my sons, daughter-in-laws, and grandchildren, thanks for hanging in there with me, I am finally graduating.

Field and lab work upon which this dissertation is based has been supported in part by citizens who contributed to the Nongame Wildlife Fund on their Michigan income tax form or by a direct donation to the Fund, by the Clarence and Florence Hanes Foundation, and by the Michigan Botanical Foundation.

Pamela J. Laureto
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ ii
LIST OF TABLES .................................................................................................................. ix
LIST OF FIGURES .............................................................................................................. x

CHAPTER

I. LITERATURE REVIEW: TAXONOMY AND NATURAL HISTORY OF SOLIDAGO HOUTONII ............................ 1

Classification ..................................................................................................................... 1

The Asteraceae Family ..................................................................................................... 1

The Genus Solidago ......................................................................................................... 3

The Species Solidago houghtonii ...................................................................................... 6

Type Locality ................................................................................................................... 6

Species Description ......................................................................................................... 7

Geographic Distribution and Habitat ............................................................................. 9

Reproductive Biology ...................................................................................................... 14

Hybridization and Hypotheses of Allopolyploid Origin ........................................... 18

Population Structure ...................................................................................................... 22

Aims of Dissertation ....................................................................................................... 24

Phylogeny ........................................................................................................................ 25

Population Genetics ......................................................................................................... 26

II. NUCLEAR EVIDENCE FOR THE HYBRID ORIGIN OF SOLIDAGO HOUTONII ................................................... 27

Introduction ................................................................................................................... 27
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials and Methods</td>
<td>31</td>
</tr>
<tr>
<td>Taxon Sampling</td>
<td>31</td>
</tr>
<tr>
<td>DNA Extraction</td>
<td>32</td>
</tr>
<tr>
<td>PCR Amplification and Sequencing</td>
<td>34</td>
</tr>
<tr>
<td>Confirmation of ITS Homology</td>
<td>35</td>
</tr>
<tr>
<td>Cloning of the 3'ETS-ITS nrDNA Region</td>
<td>36</td>
</tr>
<tr>
<td>Phylogenetic Analyses</td>
<td>37</td>
</tr>
<tr>
<td>Phylogenetic Test of Hybridization</td>
<td>39</td>
</tr>
<tr>
<td>Results</td>
<td>39</td>
</tr>
<tr>
<td>3'ETS and ITS Sequence Variation</td>
<td>39</td>
</tr>
<tr>
<td>Homeologous Sequence Types</td>
<td>41</td>
</tr>
<tr>
<td>Test of Hybridization</td>
<td>51</td>
</tr>
<tr>
<td>Discussion</td>
<td>52</td>
</tr>
<tr>
<td>Nuclear Evidence of Parentage</td>
<td>52</td>
</tr>
<tr>
<td>Parentage by Solidago riddellii</td>
<td>54</td>
</tr>
<tr>
<td>Parentage by Solidago ptarmicoides</td>
<td>56</td>
</tr>
<tr>
<td>Parentage by Solidago ohioensis</td>
<td>57</td>
</tr>
<tr>
<td>Alternative Explanations for Molecular Results</td>
<td>59</td>
</tr>
<tr>
<td>Morphological Support of Hybridization</td>
<td>60</td>
</tr>
<tr>
<td>Single vs. Polytopic Origin</td>
<td>61</td>
</tr>
<tr>
<td>III. CHLOROPLAST EVIDENCE FOR THE HYBRID ORIGIN OF Solidago Houghtonii</td>
<td>64</td>
</tr>
</tbody>
</table>
# Table of Contents—continued

## CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype Variation</td>
<td>95</td>
</tr>
<tr>
<td>Genetic Diversity</td>
<td>96</td>
</tr>
<tr>
<td>Population Structure of <em>Solidago houghtonii</em></td>
<td>99</td>
</tr>
<tr>
<td>Haplotype Distribution</td>
<td>101</td>
</tr>
<tr>
<td>Test for Isolation by Distance</td>
<td>103</td>
</tr>
<tr>
<td>Haplotype Relationships</td>
<td>105</td>
</tr>
<tr>
<td>Discussion</td>
<td>107</td>
</tr>
<tr>
<td>Genetic Diversity and Population Differentiation</td>
<td>107</td>
</tr>
<tr>
<td>Geographic Structure among Chloroplast Haplotypes</td>
<td>109</td>
</tr>
<tr>
<td>Isolation by Distance</td>
<td>114</td>
</tr>
<tr>
<td>Phylogeography of <em>Solidago houghtonii</em></td>
<td>115</td>
</tr>
<tr>
<td>Historical Consideration Regarding Population Structure</td>
<td>117</td>
</tr>
<tr>
<td>Implications for Conservation of <em>Solidago houghtonii</em></td>
<td>119</td>
</tr>
</tbody>
</table>

## APPENDIX

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voucher Information for Species Used in This Study</td>
<td>124</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>127</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

1. Nuclear Ribosomal DNA 3'ETS and ITS Nucleotide Site Comparison for *Solidago houghtonii* Homeologues and Progenitor Taxa.......................... 44

2. Collection Details for 23 *Solidago houghtonii* Populations ....................... 87

3. Haplotype Matrix .................................................................................................. 92

4. Distribution of cpDNA Haplotypes among 23 *Solidago houghtonii* Populations ........................................................................................................... 98

5. Chloroplast Polymorphism within Five Sampled Geographic Regions ...... 98

6. Chloroplast Polymorphism within Six Regions Defined by Habitat ............ 99

7. Hierarchical Analysis of Molecular Variance (AMOVA) of *Solidago houghtonii* Samples based on Haplotype Frequencies...................................... 101
**LIST OF FIGURES**

1. *Solidago houghtonii* at Waugoshance Point, Emmet County, Michigan .................................................. 8
2. Geographic Distribution of *Solidago houghtonii* ......................................................................................... 10
3. Nuclear Ribosomal DNA Cistron Showing Locations of the ITS and ETS Regions and Primers .......................... 28
4. Nuclear Ribosomal DNA Strict Consensus Tree ............................................................................................. 46
5. Statistical Parsimony Network ....................................................................................................................... 50
6. Informative Chloroplast DNA Indels ............................................................................................................. 70
7. Chloroplast DNA Strict Consensus Tree ......................................................................................................... 73
8. Relative Frequency of 14 cpDNA Haplotypes ................................................................................................. 97
9. Collection Localities and the Distribution of 14 *Solidago houghtonii* cpDNA Haplotypes ................................ 102
10. Relationship between Genetic Distance ($F_{ST}$) and Geographical Distance of All Sampling Pairs ............. 104
11. Relationship between Genetic Distance ($F_{ST}$) and Geographical Distance of All Sampling Pairs minus the New York Disjunct Population .............................................................................. 105
12. Statistical Parsimony Network of 14 *Solidago houghtonii* cpDNA Haplotypes ........................................... 106
CHAPTER I

LITERATURE REVIEW: TAXONOMY AND NATURAL HISTORY OF
SOLIDAGO HOUGHTONII

Classification

The Asteraceae Family

The Asteraceae (Aster, Daisy or Sunflower Family) is the largest family of vascular plants, consisting of approximately 1,620 genera and 23,600 species which accounts for nearly 8% of all flowering plants (Panero and Funk 2008). The family has a worldwide distribution, except for Antarctica, and is especially diverse in the tropical and subtropical regions of North America, the Andes, eastern Brazil, southern Africa, the Mediterranean, central Asia, and southwestern China (Stevens 2001 onwards). The family contains several species of anthropogenic importance as sources of cooking oils, sweetening agents, tea infusions, and important garden ornamentals zinnias, marigolds, dahlias, and chrysanthemums (Panero and Crozier 2008).

Taxonomic work within the Asteraceae has a long history, yet the family is still without a well-resolved phylogenetic estimate of the relationships between genera and higher taxa (Panero and Funk 2008). Henri Cassini, a French botanist regarded by many as the founder of Asteraceae classification, established the tribal classification within the Asteraceae in the nineteenth century (Bremer 1994). It was not until the 1970s that the relationship between the tribes began to be understood when both Carlquist (1976) and Wagenitz (1976) circumscribed two subfamilies, Cichorioideae and Asteroideae, based
primarily on differences in corolla, anther, and style morphology. According to Bremer (1994), the Asteroideae are characterized (with some exceptions) as having true ray florets, disc corollas with short lobes, caveate pollen, stigmatic surfaces of style branches separated into two marginal lines sometimes confluent at apices, and a distinctive secondary chemistry. This combination of morphological characteristics is seldom seen in the Cichorioideae. With the advent of molecular phylogenetic studies, the classification of the Asteraceae changed relatively quickly (Panero and Funk 2008). Following the chloroplast DNA (cpDNA) restriction fragment length polymorphism (RFLP) work of Jansen and Palmer (1987), a third subfamily, the Barnadesioideae, was erected (Bremer and Jansen 1992). In his book on the cladistics and classification within the Asteraceae, Bremer (1994) recognized three subfamilies (Asteroideae, Barnadesioideae, and Chichorioideae) and 17 tribes. In the family-wide analysis of cpDNA \textit{ndhF} gene sequences done by Kim and Jansen (1995) the Barnadesioideae was resolved as sister to all other members of the Asteraceae, the subfamily Asteroideae formed a monophyletic clade, but the subfamily Cichorioideae was shown to be paraphyletic. This led Bremer (1996) to formally recognize five branches within the Asteraceae; subfamilies Asteroideae, Barnadesioideae, Carduoideae, Cichorioideae, and the paraphyletic (Kim and Jansen 1995) tribe Mutisieae. Panero and Funk (2008) used cpDNA sequence data obtained from 10 markers, along with extensive sampling of 108 taxa, including several taxa of uncertain tribal position, to construct a phylogenetic hypothesis of the Asteraceae at the subfamily and tribal levels. Their analysis recognized 12 monophyletic subfamilies and resolved many of the problematic relationships within the family. However, systematic work within the Asteraceae is not complete as the phylogenetic relationships among the three main lineages of the Cichorioideae and the
position of subfamily Stifftioideae remain unresolved. In addition, the monophyly of only five tribes has been tested (Panero and Funk 2008).

The Asteroideae, is the largest subfamily of the Asteraceae. It is comprised of 1210 genera and approximately 17,000 species, accounting for 72% of the diversity within the Asteraceae (Panero and Funk 2008). Most of the eleven subfamilies basal to Asteroideae are species poor, with the exception of Chichoroideae, with 3,600 species, and Carduoideae, with 2,800 species (Stevens 2001). The Asteroideae is comprised of 8 tribes for which the tribe Astereae is one of the largest with approximately 170 genera and 3000 recognized species (Bremer 1994). The Astereae tribe is comprised of 3 subtribes, the Asterinae which contains the majority of the genera (117) and species (2100) and which has a world-wide distribution, the Grangeinea which is a small group of paleotropical taxa, and the Solidagininae, a predominately North American subtribe comprised of 48 genera and approximately 700 species (Bremer 1994). *Solidago houghtonii* is a member of the Asteraceae, subfamily Asteroideae, tribe Astereae, and subtribe Solidagininae.

The Genus *Solidago*

The name *Solidago* L. originated from the Latin *solidus*, meaning whole, and – *ago*, resembling or becoming (Semple and Cook 2006), owing to the belief, still held by some today, that the plants have medicinal value. The genus includes about 100 North American species, with the greatest number of these occurring in eastern North America (Semple and Cook 2006) where they are an important component of the autumn flora. Additionally, there are 6 – 10 species that are native to Eurasia, eight in Mexico, and four in South America. Some species, such as *S. canadensis* and *S. nemoralis*, are ubiquitous
and often occur in such dense populations that they are commonly described as “weeds”. However, many other species are sparsely distributed throughout their native range, and four species, *S. albopilosa*, *S. houghtonii*, *S. shortii*, and *S. spithamaea*, have such restricted distributions that they are federally listed as either “Threatened” or “Endangered” (USFWS 1988a; USFWS 1998b; USFWS 1985b; USFWS 1985a respectively).

*Solidago* is comprised of herbaceous perennial plants which arise from a short, stout caudex that produces clumps of stems or a more slender and elongate rhizome that produces scattered stems. The stems can be decumbent to ascending or erect and range in size from 0.05 – 1.0 (–2.0) m and are glabrous to strigose to hispid. The leaves are alternate and range from basally disposed to chiefly cauline, the lower ones are petiolate to sessile, elliptic to lanceolate or oblanceolate to obvate, are primarily toothed although some species have entire margins. Plants bloom in late summer or fall and are gynomonoecious meaning they bear both pistillate (female) and bisexual flowers. Flowers are yellow, very rarely white, and are borne in small capitula which are arranged into an inflorescence. In a few species the inflorescence is axillary, but most species have a terminal inflorescence that can broadly be described as either paniculiform, racemiform, or corymbiform. An inflorescence is composed of few to many capitula, and each capitulum is composed of several miniature flowers each called a floret (Mani and Seravanan 1999). Within each capitulum, several fertile ray (pistillate) florets, each with a single petal, surround a small cluster of fertile disc (bisexual) florets bearing five fused petals. The ray florets open before the disc florets and disc florets are protandrous, so that pollen presentation precedes stigma receptivity (Bertin and Gwisc 2002 and references within). Fruits are compressed obconic achenes subtended by a pappus of 25 – 45 long,
short-barbed, bristles. The chromosomal base number for *Solidago* is $x=9$ (Semple and Cook 2006).

Although there are no distinct morphological apomorphies that define *Solidago* (Schilling et al. 2008), most species are easily recognized as “goldenrods” by their numerous, small capitula of yellow flowers. However, morphological complexity within the genus often makes identification to species difficult. Clear delineation of species is further complicated, not only by the occurrence of interspecific hybrids, but also because many species are polytypic, and several species have multiple cytotypes including both diploids and polyploids (see Semple and Cook 2006 for polytypic species and ploidy numbers). Based primarily on morphology, the genus has been divided into differing numbers of sections and subsections. Nesom (1993) presented a taxonomic overview in which he recognized two sections, section *Solidago* with four subsections and section *Unilaterales* with seven subsections. Nesom also segregated two taxa that had previously been treated within *Solidago*: *Oligoneuron* Small and *Oreochrysum* Rydberg.

According to this view, the genus *Oligoneuron* includes six species that are distinguished morphologically from the other members of *Solidago* by their corymboid (flat-topped) inflorescence, punctate leaves, broader phyllaries, and larger achenes. Additionally, the author cites a “paucity of natural hybridization” between taxa of *Oligoneuron* and those of *Solidago* as further support for the segregation of *Oligoneuron*. Most recently, Semple and Cook (2006) have presented a taxonomic treatment of *Solidago* in which they recognized section *Solidago* with 11 subsections and section *Ptarmicoidei*. They placed the six corymboid taxa, recognized by Nesom (1993) as *Oligoneuron*, within *Solidago* section *Ptarmicoidei*. This is significant because my species of interest, *Solidago houghtonii* Torr. & Gray ex Gray, is one of the corymboid taxa cited in Nesom’s revision
as *Oligoneuron houghtonii* (Torr. & Gray ex Gray) Nesom. Because the species is listed in the Federal Register (USFWS 1988b) as *S. houghtonii* that name will be used throughout. To date, there has not been a comprehensive molecular phylogenetic study of *Solidago*; however, recent molecular studies involving members of the subtribe Solidagininae (Zhang 1996; Urbatsch et al. 2003; Beck et al. 2004; Schilling et al. 2008) found *Oligoneuron* to be most closely related to *Solidago*, which may support the inclusion of *Oligoneuron* within *Solidago*. Circumscription of *Solidago* awaits a definitive molecular-based phylogeny (Beck et al. 2004).

The Species *Solidago houghtonii*

**Type Locality**

*Solidago houghtonii* A. Gray, commonly called Houghton’s goldenrod, was discovered on the north shore of Lake Michigan on August 15, 1839 by a team from the Michigan Geological Survey (McVaugh 1970). The survey team was led by Dr. Douglass Houghton, Michigan’s first appointed State geologist. Between August 2 and August 26, 1839 the survey team traveled by rowboat from Mackinac Island, Mackinac Country Michigan USA to Greenbay, Wisconsin USA (McVaugh 1970). On the day of August 15th, Houghton’s notes indicate that the survey team traveled from a campsite east of Epoufette to their next campsite, west of Naubinway. Therefore the type locality is between approximately 10.4 km southeast of Naubinway and 10.4 km west of Epoufette in western Mackinac County, Michigan USA (Guire and Voss 1963). Even though Douglass Houghton was an experienced botanical collector (Voss 1978), the type specimens of *S. houghtonii* were likely collected by Houghton’s assistant George Bull.
The specimens were sent to Asa Gray, a well known American botanist, who named the new discovery in honor of Douglass Houghton in his first edition of Gray’s Manual (Gray 1848).

Species Description

_Solidago houghtonii_ is listed as a federally threatened species and has been awarded protection through the Endangered Species Act of 1973 (USFWS 1988b). Additionally, it is listed as “state threatened” in Michigan (MDNR, Michigan Special Plants List 1991), as “state endangered” in New York (Young 1996), and as “rare” in Canada (COSEWIC 2005).

_Solidago houghtonii_ is described by Semple and Cook (2006) as producing a short, branching caudex from which a basal rosette of leaves and flowering stems arise (Fig. 1). In addition, the caudex forms numerous short rhizomes which give rise to cespitose clumps of clonal plants. The basal leaves of _S. houghtonii_ are ovate with an acute apex and taper to a long winged petiole that is somewhat conduplicate. The basal leaves range to 22 mm wide, are glabrous, and have entire (sometimes irregularly and finely toothed) margins. The lower cauline leaves are linear-oblanceolate, acute to obtuse, and taper to a winged petiole-like base that partially clasps the stem. These are also glabrous, entire, range from 158.0 – 177.0 × 7.0 – 22.0 mm, and display a unique 3-nerved venation pattern in which the 2 lateral nerves arise proximal to and run alongside the mid-nerve for some distance before diverging. The upper cauline leaves are similar to the lower but are sessile and reduced upward ranging from 47.0 – 100.0 × 4.0 – 10.0 mm. The slender stems of _S. houghtonii_ are glabrous, sometimes reddish, can reach heights of
solidago houghtonii at Waugoshance Point, Emmet County, Michigan

30.0 - 60.0+ cm, and terminate in a corymbiform inflorescence composed of (2-) 5 - 50 (100+) capitula borne on finely but distinctly pubescent pedicles. Each capitulum consists of an involucre that is 5.0 - 8.0 mm in length and composed of an unequal series of phyllaries that are blunt, rough-margined, and weakly striated. In addition, each
capitulum consists of 6–12 large (7.4 – 7.9 × 0.5 – 0.6 mm), bright yellow ray florets that surround 8–15 disc florets. The achenes of S. houghtonii are obconic, 1.4 – 1.8 mm long, glabrous with ribs that are sometimes dark. The pappus bristles attached to the apex of the achene are numerous, 4.0–5.5 mm in length with clavate tips. The plants flower from early August through early September with some plants blooming well into October. Solidago houghtonii is a hexaploid species with 2n = 6x = 54 chromosomes (Semple and Cook 2006).

Solidago houghtonii is most easily confused with two morphologically similar species; the widespread Euthamia graminifolia (grass-leaved goldenrod) and S. ohioensis (Ohio goldenrod). Both these species are commonly found in association with S. houghtonii. Euthamia graminifolia can be distinguished by its leafier stem which lacks basal leaves at anthesis, its narrower 3–5 nerved leaves, and its inflorescence which is composed of distinctly smaller capitula with short ray florets and pubescent achenes. Solidago ohioensis can be distinguished as a more robust species, with relatively broad, flat, ovate-lanceolate leaves and a dense inflorescence consisting of numerous small capitula borne on smooth pedicles. The ray florets of S. ohioensis are distinctly smaller than those of S. houghtonii and the glabrous achenes lack ribs.

Geographic Distribution and Habitat

Solidago houghtonii is an Upper Great Lakes endemic, occurring mainly on the northern shores of Lakes Michigan and Huron in the State of Michigan, USA and Ontario, Canada where approximately 71 isolated populations are known to occur; 58 in Michigan and 13 in Ontario (Fig. 2). The majority, 51, of these populations occur in northern Michigan around the Straits of Mackinac which separates Lake Michigan from
Figure 2. Geographic Distribution of *Solidago houghtonii*.

Lake Huron (USFWS 1997; COSEWIC 2005). The distribution of *S. houghtonii* is associated with a geological feature known as the Niagara Escarpment. This bedrock formation begins east of Rochester New York USA, runs westward along the southern shore of Lake Ontario and then turns and runs northwest where outcrops form the Bruce Peninsula and islands of the Manitoulin region in northern Lake Huron. From the Manitoulin region, the escarpment turns westward where it underlies the southern edge of Michigan’s Upper Peninsula. Finally, it heads south along the western coast of Lake Michigan following the Door Peninsula in the State of Wisconsin USA and ending near the Wisconsin-Illinois boarder.

Note: Red dots indicate approximate location of known populations based on USFWS (1997) and COSEWIC (2005). Boxes indicate four hypothesized entities for *Solidago houghtonii* – see text (USFWS 1997). Collection localities for individuals included in this study are indicated by arrows and referenced by their accession number. Figure modified from Guire and Voss (1963).
The escarpment, as well as the whole of the Great Lakes region, was covered by a vast ice sheet during the Wisconsin glaciation which peaked about 20,000 years ago (Morton and Venn 2000). As the glaciers melted, they left behind vast deposits of silt, sand, gravel and boulders, and their melt water filled depressions giving rise to the Great Lakes. As the glaciers receded, water levels in the early lakes fell as they drained southwards so that by 9000 years ago only a river separated Michigan’s Upper Peninsula from the land now known as the Manitoulin Region (Morton and Venn 2000). The discovery of 9900 year old tree stumps on the floor of the Main Channel between Manitoulin Island and the Bruce Peninsula provide strong evidence that during this time the islands of the Manitoulin region were connected to the Bruce Peninsula by dry land (Tovell 1978). This period of low lake levels was also marked by climatic warming and the uninterrupted migration of both plants and animals into these previously glaciated regions (Morton and Venn 2000). However, the period of low lake levels did not last and by about 8,700 years ago lake levels again began to rise. The period of warm climate and high lake levels ended about 4,000 years ago and lake level again fell, giving rise to the Great Lakes in their present form (Morton and Venn 2000). As the waters of the lakes receded, weathering processes formed the various alkaline substrates, with which *S. houghtonii*’s habitat is associated.

Throughout much of its Michigan distribution, *Solidago houghtonii* populations are restricted to sand beach flats, cobblestone shores, the lee side of foredunes, and fen-like interdunal wetlands that parallel the shorelines of Lakes Michigan and Huron. The sands associated with the Niagara Escarpment are composed of about 1%-5% ground mollusk shells and are therefore circumneutral (pH 7.0) to alkaline (pH 8.0) and may have a thin organic covering (USFWS 1997). Jolls (1994) found a higher organic content in
soils of cobble beaches vs. those of sand flats. The Canadian populations of *S. houghtonii* occur primarily in the Manitoulin region, including the La Cloche Peninsula, Cockburn Island, Great La Cloche Island, Little La Cloche Island, Manitoulin Island, Strawberry Island, and also on the eastern tip of the Bruce Peninsula. Here, *Solidago houghtonii* is primarily associated with smooth limestone pavements, or alvars, of the Niagara Escarpment.

Alvar habitats were created as erosion of glacial deposits exposed the Niagara Escarpment. These pavements are composed of either limestone or the harder dolomite limestone. Alvar formed on dolomite limestone is characterized by large flat expanses of bare rock which occasionally fractures and characteristically weathers into small pot-holes of roughly 2.0 – 4.0 cm in diameter. These fill with water and organic debris to provide habitat for calcareous loving plants (Morton and Venn 2000; personal observation). Alvars formed on softer limestone are typically covered by a thin layer of mineral soil which supports a grass- and sedge- dominated plant community along with other calcareous loving forbes. Both alvar types are characterized by annual cycles of flooding followed by extreme drought for extended periods of the growing season (Stephenson 1986). In Ontario Canada, over 86% of known *S. houghtonii* populations occur on alvar (COSEWIC 2005 and reference therein) as compared to Michigan populations which more commonly occur in interdunal fen-like habitats (USFWS 1997; personal observation). The restriction of *S. houghtonii* to calcareous sand and dolomitic limestone suggests that the species may require substrates with relatively high amounts of calcium and magnesium (USFWS 1997).

*Solidago houghtonii* most often occurs where there is a relatively low density of competing vegetation (personal observation). In Michigan, this is primarily along the
edges of fen-like interdunal wetlands, although the plants also grow on beach flats, the leeward side of foredunes, and on cobble beaches. These shoreline habitats are dynamic; wind and wave action leads to the continual movement of sand which can destabilize populations and lead to cycles of plant extinction and recolonization as suitable habitat develops. Additionally, changes in the water table associated with the fluctuating water levels in the Great Lakes can greatly affect these shoreline habitats (USFWS 1997). It has been suggested, that the establishment of *S. houghtonii* in some of its well-drained habitat (i.e. backside of foredune and alvar) may indicate the species has some resistance to desiccation (USFWS 1997). Buchelé et al. (1989) found that *S. shortii*, a federally endangered goldenrod, persists in xeric conditions by extending its roots 50+ cm through cracks in rocky soil. *Solidago houghtonii* may similarly be adapted to xeric conditions.

In addition to its Lakes Michigan and Huron shoreline distribution, there are four disjunct inland populations of *Solidago houghtonii*. One of these disjunct locations is in Genesee County, New York, USA at a site known as the Bergen Swamp. This population is approximately 500 km SE of the nearest Ontario population. Here *S. houghtonii* grows in a marl fen on the slightly higher ground that surrounds open pools and scattered throughout the grass and sedge dominated community. Marl is a mud formed under freshwater conditions from clay and calcium carbonate and therefore provides the alkaline conditions which *S. houghtonii* prefers. Interestingly, a marl fen, which is floristically similar to the marl fen in New York, occurs at a disjunct inland site in Michigan’s Upper Peninsula known as Summerby Swamp. Here, *S. houghtonii* also occurs along the edges of open pools and scattered among the graminoids.

A disjunct location occurs inland in Michigan’s Lower Peninsula and is approximately 96 km SE of the nearest known Lake Michigan shoreline population. This
is the only population not associated with the Niagara Escarpment. The habitat is described as a wet prairie, a community type with a strong affinity to both northern and southern fens (USFWS 1997). The area over which *S. houghtonii* occurs is just to the north of Lake Margarethe, a large inland lake, and may have formed as a remnant of a larger postglacial lake that once covered the area (M. Penskar, per. comm.). The soil is alkaline but with a much deeper organic layer than is found in other *S. houghtonii* habitats. This population is unique because it is the only population not associated with the Niagara Escarpment and also because the plants are reported to be octoploid (2n = 8x = 72) (J. Pringle per. comm.).

Reproductive Biology

*Solidago houghtonii* is an herbaceous perennial plant that persists for “at least” ten years (Jolls 1994). Plants form a short, branching, caudex from which stems arise. In addition, the caudex also forms numerous short rhizomes so that a cespitose clump of ramets is produced. This indicates that, as in other goldenrod species, vegetative propagation is a common form of reproduction for *S. houghtonii*. Plants have between 2 – 12 ramets that rapidly disarticulate and become established as independent genets (USFWS 1997). Ramets die back to the underground caudex each year. In the next growing season a new node is produced on the caudex and a new ramet appears (Jolls 1994). In general, ramets persist for only 1 year. Jolls (1994) found that flowering was correlated both with plant size and whether the plant had flowered in the previous year. Most plants do not re-flower in the following growing season; rather, they produce only vegetative growth. Field observations of *S. houghtonii* indicate that only 6% – 31% of
individuals flower in any given year possibly due to the metabolic cost of sexual reproduction (Jolls 1994).

Solidago houghtonii produces a corymboid inflorescence composed of (2-) 5 – 50 (100+) capitula. The structurally complex capitulum is gynomonoecious, having both pistillate and bisexual florets; however, it behaves ecologically and functionally as if it were a single blossom (Mani and Seravanan 1999). Within the capitulum of S. houghtonii florets develop as one whorl per day; first the outer ray florets emerge and become receptive, followed by the two to three inner whorls of disc florets (Jolls 1994). Bertin and Gwisc (2002) examined the gynomonoecious sexual system in six Solidago species in order to determine if the ratio of pistillate to bisexual florets varies in Solidago as a response to changes in environmental conditions. This type of sexual system occurs in approximately 2.8% of angiosperms and is especially common in the Asteraceae (Bertin and Kerwin 1998 and references therein). Gynomonoecy is believed to have arisen as an adaptive mechanism that would allow plants flexibility in allocating resources to either male or female reproductive functions in response to varied environmental conditions (Bertin and Kerwin 1998; Bertin and Gwisc 2002). Under such a scenario the female function is expected to be emphasized when resources are readily available as female fitness is tied to the availability of resources for making fruits and seeds. Alternatively, the male function is expected to be emphasized under harsh environmental conditions or when resources are scarce as male fitness is limited by access to ovules (Bertin and Gwisc 2002). Bertin and Gwisc (2002) varied light, nutrient, and water regimes. The effects on both the proportion of ray florets in a capitulum and the number of florets per head were then evaluated. Their results did not support the hypothesis that the gynomonoecious sexual system in Solidago is advantageous in
permitting flexibility in the allocation of resources to male and female reproductive functions. However, they point out that the lack of environmental influences on floret ratios in *Solidago* does not rule out the idea that plants respond to changing environmental conditions by changing sex allocation. Allocation of resources to either ovule production, or to the number of pollen grains produced per anther, is a possible mechanism for regulating sex expression in response to environment. Because the environment fluctuates continuously in both the shoreline and alvar habitats of *S. houghtonii*, the allocation of resources to femaleness or maleness should be investigated for this species.

While the study of Bertin and Gwisc (2002) showed little evidence of environmental influence on pistillate/bisexual floret ratios, it did show considerable variation among individuals of the same species. Andersson (1993) found similar among-plant variation when examining seed/ovule ratios, a trait directly related to plant fitness. Andersson (1993) suggested that among-plant variation could reflect a genetic load generated by sexual reproduction in outcrossing species particularly when coupled with the maintenance of less reproductively fit genotypes through clonal reproduction. Compatibility experiments in four species of *Solidago* concluded that they are largely self-incompatible and reliant on pollinators for successful seed-set (Gross and Werner 1983). As many species of *Solidago*, including *S. houghtonii*, reproduce clonally, similar pressures could underlie the variation among individuals observed in *Solidago* (Bertin and Gwisc 2002).

Jolls (1994) study of seed and seedling ecology found *Solidago houghtonii* to be self-incompatible. Achene production was shown to be quite low, averaging 12% - 17% fertile seed per outcrossed ramet. The average germination rate of *S. houghtonii* seed was
shown to be relatively low, 56%, and germination rates declined with seed age. Therefore, seeds appeared to remain viable for only 1 year, which suggests that there is not a persistent seed bank for *S. houghtonii* and that asexual reproduction is essential for persistence. Additionally, winter stratification and light were found to be necessary for germination, and germination was not correlated to moisture or substrate type. Similar findings have been reported for the Federal endangered *Solidago shortii* (sect. *Triplinerviae*). Buchelé et al. (1991a) found that *S. shortii* and several of its weedy and/or geographically widespread congener species were self-incompatible, and that fertile seed production by outcrossed individuals averaged 22% – 25.2% (Buchelé et al. 1992). Furthermore, Buchelé et al. (1991b) indicated that *S. shortii* did not produce a viable seed bank. Gross and Werner (1983) also found a low overall average fertile seed-set in two widespread weedy *Solidago* species; 12.4% fertile seed in *S. juncea* and 33.2% in *S. canadensis*. In addition, seed production by *S. canadensis* varied from 13.2% to 33.2% depending on year. These results are similar to those suggested for *S. houghtonii* and may be typical for the genus.

The similarity in size, color, and general floral morphology between the different species of *Solidago* suggests that they are able to attract and use the same insect pollinators (Gross and Werner 1983). *Solidago* florets provide both nectar and pollen for foraging insects. While each individual floret provides only a small quantity of nectar, a large clone will produce thousands of florets and ample rewards for insect pollinators. The major insects carrying *Solidago* pollen are bees (Hymenoptera), butterflies (Lepidoptera), flies (Diptera), moths (Lepidoptera), and wasps (Hymenoptera) (Gross and Werner 1983; Jolls 1994). Because *Solidago* pollen is characteristically heavy and sticky, it is well suited for invertebrate pollen vectors (Semple et al. 1999). Note, these traits
differ from wind-borne pollen, mitigating the common belief that Goldenrods (*Solidago* spp.) cause seasonal allergies.

**Hybridization and Hypotheses of Allopolyploid Origin**

Interspecific hybridization and polyploidization have contributed greatly to the diversification and speciation of plants (Grant 1981; Arnold 1997; Rieseberg 1997; Soltis and Soltis 2000). Several lines of evidence have been used to elucidate the origins and evolutionary history of polyploid plant taxa. Early studies focused on morphological, cytogenetic, and allozyme data. More recently, DNA sequence data obtained from both the uniparentally inherited chloroplast genome and the biparentally inherited nuclear genome have been used to clarify the phylogenetic affinities of several polyploid species (reviewed in Arnold 1997; Rieseberg 1997). However, the phylogenetic reconstruction of hybrid speciation is often difficult due to the reticulate nature of hybrid origin which violates the cladistic assumption that speciation occurs in a bifurcated manner. The use of chloroplast and nuclear markers can provide complementary evidence regarding hybridization events and their combined use allows different questions to be investigated (Sang and Zhong 2000). Since in most angiosperms the chloroplast genome is maternally inherited (Corriveau and Coleman 1988), phylogenetic analysis of chloroplast DNA (cpDNA) sequence data can be used to investigate the maternal source of an allopolyploid species where the hybrid would be expected to appear in a clade with its maternal parent.

In contrast, the biparentally inherited nuclear genome is potentially useful for studying the origins of polyploid species as all F₁ hybrids are expected to have the full genomic complement of both parents. Nuclear markers can be used to determine the number of independently derived genomes that make up an allopolyploid, and also to
identify the extant diploid species whose genomes share the same evolutionary history and that may have acted as the progenitors of the hybrid.

Molecular phylogenetic analyses of allopolyplloid plant taxa have shown that many have formed multiple times from independent hybridization events (Soltis and Soltis 1993; 1999), that complex relationships which involve multiple unexpected genome donors occur (Cronn et al. 2003; Mason-Gamer 2004), and that cpDNA introgression into polyploid lineages is common (Wendel 1989; Dorado et al. 1992; Liston and Kadereit 1995; Cronn et al. 2003).

Kaplan and Fehrer (2007) point out that the majority of literature on plant hybridization is based on binary hybrids and that most of those are focused on the origin of allotetraploids. In contrast, the literature contains very few examples of studies using molecular tools to reveal the origin of recently formed natural hybrids between three (or more) species. According to these authors, most of the literature on triple hybrids 1) is based on morphology and presented in the context of local floras, 2) reports on the experimental production of triple hybrids, or 3) is focused on ancient hybridization events involving three or more species and that the majority of these studies are based on grasses and grain crops.

Recent naturally occurring hybrids between three of more species have been reported for Aesculus (dePamphilis and Wyatt 1990) and Quercus (Dodd and Afzal-Rafii 2004). Using allozyme and AFLP markers, respectively, each study found that, in populations devoid of morphologically identifiable hybrids, genes from at least two species had been introgressed into a third. Arnold et al. (1990) investigated the origin of Iris nelsonii, a stabilized hybrid from Southern Louisiana which displayed morphological and chromosomal characteristics of three sympatric Iris species. Allozyme data
confirmed that *Iris nelsonii* possessed a combination of genetic markers and indicated the involvement of all three sympatric taxa as parental donors. However, in none of these cases were the naturally occurring triple hybrids reported to be hexaploid. Kaplan and Fehrer (2007) used nrDNA ITS and cpDNA sequence data to reveal the presence of three genomes in a persistent, but sterile, naturally occurring hexaploid *Potamogeton* clone. *Solidago houghtonii* is a fertile hexaploid species (2n = 6x = 54) and likely derived through natural hybridization (Morton 1979; Semple et al. 1999).

Ellstrand et al. (1996) determined that the distribution of naturally occurring hybrid taxa is not uniform across higher plant families but rather is concentrated among certain “hotspot” families. Among the families included in their study, the Asteraceae accounted for the largest number of hybrid taxa. Hybridization in several genera of this family have been reported in the literature including *Tragopogon* (Soltis et al. 1995; Soltis and Soltis 1999), *Senecio cambrensis* (Lowe and Abbott 1996), *Microseris acuminata* and *M. campestris* (Roelofs et al. 1997), *Helianthus anomalus* (Schwarzbach and Rieseberg 2002), *Eupatorium godfreyanum* (Siripun and Schilling 2006), *Doellingeria* (Saito et al. 2007), and Solidaster cv. Lemore (Schilling et al. 2008) to name a few. The genus *Solidago* is characterized as having a perennial habit, outcrossing breeding system, and clonal reproduction. These are all characteristics identified by Ellstrand et al. (1996) as advantageous to the formation and stabilization of hybrid taxa. Gleason and Cronquist (1991) consider interspecific hybridization to be widespread within *Solidago*.

There have been several hypotheses regarding the hybrid origin of *Solidago houghtonii*; none of which have been formally tested. Morton (1979) proposed that *S. houghtonii* might have arisen by amphidiploidy (chromosome doubling) of naturally
occurring hybrids between *S. ohioensis* and *S. ptarmicoides*. He, as well as Mitchell and Sheviak (1981), further proposed that the New York plants were of an independent origin from the Great Lakes plants. Semple et al. (1999) proposed that *S. riddellii* was involved in the hybrid with *S. ptarmicoides*, rather than *S. ohioensis*, citing morphological similarities in leaf structure between *S. riddellii* and *S. houghtonii*. Pringle (USFWS 1997) proposed that *S. houghtonii* is actually comprised of 4 separate entities, each resulting from separate hybridization events involving different putative parents and corresponding to distinct geographic locations. Pringle’s hypotheses are as follows:

“entity 1” - true *S. houghtonii*, represented by the nomenclatural type and located around the Straits of Mackinac in Michigan, is the result of hybridization between *S. ptarmicoides* (*2n = 18*) and *S. riddellii* (*2n = 18*); “entity 2” - the Canadian plants that occur around the Manitoulin Island region and on the east side of the Bruce Peninsula, resulted from hybridization between *S. ptarmicoides* (*2n = 18*) and *S. ohioensis* (*2n = 18*); “entity 3” - the octoploid population disjunct in Michigan, resulted from hybridization between *S. ptarmicoides* (*2n = 18*) and *S. uliginosa* (*2n = 36*); and “entity 4” - the New York plants, derived from hybridization between *S. ptarmicoides* (*2n = 18*) and *S. uliginosa* (*2n = 18*) (Fig. 2). Pringle cites variation in chromosome number and morphology as evidence for these hypotheses. Edwards-Wilson (1999) attempted to test these hypotheses using electron microscopy to evaluate 22 morphological characters of *S. houghtonii* and all putative parental taxa. The multivariate analysis of variation for each of the morphological characters suggested that *S. houghtonii* individuals from across its geographic range were one species. However, the statistical results were inconclusive with regard to parentage, suggesting that *S. ptarmicoides* and either *S. riddellii* or *S. uliginosa* were involved.
Population Structure

Most natural populations of species show some level of genetic structure (Haig 1998; Balloux and Lugon-Moulin 2002). Several authors (Gray 1996; Haig 1998; Balloux and Lugon-Moulin 2002) identify both the “intrinsic biological properties” of the species and “extrinsic dynamic processes” as factors which affect the genetic diversity in plant populations. The intrinsic biological property which has the largest influence on genetic diversity of plant populations is the type of breeding system (Gray 1996). On average, self-incompatible species, such as *Solidago houghtonii*, have a significantly higher percentage of individuals that are heterozygous at polymorphic loci than do selfing species (Gray 1996). Additionally, outcrossing individuals, such as *S. houghtonii*, have significantly higher levels of overall genetic diversity (Gray 1996). Hamrick and Godt’s (1989) analysis of allozyme diversity in more than 400 plant species showed that selfing species had more than 50% of their total diversity among populations as compared with wind-pollinated, outcrossing species that had approximately 10%. Hamrick and Godt (1989) also showed that annual species had nearly four times the diversity among their populations than long-lived woody perennials indicating that life form is another important intrinsic biological property. The species ploidy level, mode of pollination, and method of seed dispersal are other intrinsic factors acting to shape the genetic diversity within plant populations (Gray 1996). Fluctuations in population size caused by bottlenecks, founding events, plant community succession, and a range of historical and contemporary events including glaciation, environmental uncertainty due to predators or pathogens, abiotic variation such as fire, drought, lake level, and anthropogenic habitat fragmentation are examples of extrinsic dynamic processes which could affect the genetic structure of populations (Hamrick and Godt 1989; Gray 1996; Bonnin et al. 2002).
In addition to showing some level of genetic structure, most plant species live in naturally fragmented habitats (Bonnin et al. 2002). Because habitat is not continuous, species occur as discrete populations (subpopulations) that are more or less connected by migration, a concept explained by metapopulation theory (Opdam, 1991; Baguette 2004; Hanski 2004). As ecological conditions change in response to various extrinsic processes a given habitat site may become unsuitable while at the same time new habitat sites become available. Therefore, many species are characterized by a balance between local population extinctions and the establishment of new populations. *Solidago houghtonii* populations are more or less contiguous patches of plants separated from other patches by gaps of at least 1.6km but many populations are separated by much greater distances. Thus populations of *S. houghtonii* are best thought of as subpopulations of a metapopulation.

Migration among existing subpopulations is a factor that can affect the genetic structure within a given subpopulation. Such migration may act as a buffer against the consequences of extrinsic processes and thereby increase the probability that subpopulations persist. However, migration between subpopulations can also prevent them from adapting to their local environmental conditions (Bonnin et al. 2002). Within the metapopulation structure, a balance between the opposing forces of migration and selection, as well as, extinction and colonization act to maintain viability (Bonnin et al. 2002). Human activities such as development on the Great Lakes shoreline and the increased use of shoreline habitat for recreational purposes act to disrupt such balances by increasing habitat fragmentation (USFWS 1997; Bonnin et al. 2002). As the landscape becomes more fragmented the amount of suitable habitat is reduced which leads to smaller population sizes and increased rates of extinction. Additionally, as the amount of
suitable habitat is reduced, the distances between habitat sites increases. This leads to a decrease in the rate of colonization as the abundance and effectiveness of pollinators and seed dispersers is altered (England et al. 2002). In highly fragmented populations, the geographic distance between subpopulations often exceeds an individual’s dispersal capacity and subpopulations become genetically differentiated through isolation by distance; populations in close proximity are genetically more similar than more distant populations (Strand et al. 1996; Balloux et al. 2002).

Numerous studies of both plant and animal species have employed the use of molecular markers to delineate population genetic structure. Information from these types of studies is central to developing an effective conservation plan for the species of interest. An understanding of the genetic differentiation within and among populations can help focus conservation efforts on specific populations in need of recovery (Haig 1998). Additionally, if founder relationships, the rates of effective dispersal among subpopulations, and the effective population size are understood then more accurate population management strategies can be implemented (Haig 1998). Currently, there is no information on population structure and genetic diversity within and among populations of *Solidago houghtonii*.

**Aims of the Dissertation**

*Solidago houghtonii* is a federally threatened plant species which has been awarded protection under the Endangered and Threatened Species Act of 1973. At the very heart of conserving biological diversity is the need for an understanding of the biology of the species for which preservation is sought (Falk and Holsinger 1991).
Conservation programs look to plant biologists and ecologists to provide the genetic and population biology data necessary for the development of biologically sound strategies by which to manage populations of rare species. Therefore, it is of not surprise that the United States Fish and Wildlife Service’s Recovery Plan for Solidago houghtonii calls for the scientific community to provide both biosystematic and population level studies (USFWS 1997). The plan states that the taxonomic relationships of S. houghtonii populations in Michigan, New York, and Canada, need to be clarified, and that the genetic variability within and among populations throughout the global range of the species should to be assessed. The study described herein aims to provide information on both the taxonomic relationships of S. houghtonii and the population genetic structure of this rare species. I hope that this information will aid in the development of a successful long-term management plan leading to the recovery of Solidago houghtonii.

**Phylogeny**

This study will use molecular data in order to assess the origin and evolutionary relationships of Solidago houghtonii. Specifically, the nrDNA ITS1 and ITS2 spacers and intervening 5.8S gene, the 3' end of the nrDNA ETS spacer, and four cpDNA intergenic spacers accD-psal, psbA-trnH, trnL (UAA)-trnF (GAA), and rps16-trnQ to investigate the putative hybrid origin of Solidago houghtonii. Here the following questions are asked: 1. Is Solidago houghtonii of hybrid origin? 2. If S. houghtonii is of hybrid origin, then who are its most likely progenitors? 3. Was S. houghtonii originated in a single hybridization event or is S. houghtonii of polytopic origin?
Population Genetics

This study will use cpDNA haplotype analysis to investigate the population genetic structure of *S. houghtonii* with the following objectives: 1. To assess how cpDNA diversity is structured, both within and among populations, 2. To estimate the relationship between geographical distance and cpDNA gene flow, 3. To determine if habitat fragmentation has significant effects on the distribution of cpDNA genetic diversity within and among populations, 4. To determine if there is a correlation between cpDNA genetic structure and substrate type, and 5. To investigate the cpDNA phylogeographic origin of *S. houghtonii*. 
CHAPTER II

NUCLEAR EVIDENCE FOR THE HYBRID ORIGIN OF
SOLIDAGO HOUGHTONII

Introduction

The biparentally inherited internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (nrDNA) cistron has proved useful for the inference of phylogeny in a large and diverse number of plant groups and for the study of hybridization, introgression, and the origin of polyploids (Nieto Feliner and Rosselló 2007). The nrDNA cistron is composed of the 18S, 5.8S, and 26S rRNA genes which are separated by two internal transcribed spacers (ITS1 and ITS2) and flanked by an external transcribed spacer (ETS). The 18S-5.8S-26S genes and the spacers are transcribed together and appear to function in the ribosomal maturation process, where the spacers are spliced out of the RNA molecule (Baldwin et al. 1995). Together, the three genes, the ITS, and the ETS regions form a transcriptional unit that is tandemly repeated hundreds to thousands of times at a locus (Fig. 3). In addition, plant genomes typically contain several ribosomal loci and these frequently occur on different chromosomes. The nuclear ribosomal region is known to undergo concerted evolution, a process which involves genomic mechanisms of turnover such as gene conversion and unequal crossing-over that act to homogenize the various repeats within an individual so that sequence similarity of the repeated gene family is greater within a species than among species (Baldwin et al. 1995). Therefore, ITS and ETS sequences obtained through direct sequencing of PCR products are the
consensus of many loci that share the same priming site (Nieto Feliner and Rosselló 2007). In part, it is this sequence homogeneity that has led to the phylogenetic utility of the nrDNA ITS region in many species (Andreasen and Baldwin 2003). Recently, the ETS region of nrDNA, a marker whose mode of molecular evolution is similar to that of the ITS but whose nucleotide substitution rate may be as much as 1.4 times faster, has increased in popularity for phylogeny reconstruction, often in conjunction with ITS sequence data, as it may provide a larger number of phylogenetically informative characters (i.e. Markos and Baldwin 2001; Andreasen and Baldwin 2003; Urbatsch et al. 2003). While the nrDNA ITS region has been the marker of choice for well over a decade, several authors have cautioned against its use, citing several molecular and organism-level processes that may affect its utility for phylogeny reconstruction (reviewed in Álvarez and Wendel 2003; Small et al. 2004; Soltis et al. 2008). In particular, the process of ITS homogenization following polyploidy formation can limit its utility for the study of polyploidy origins. Additionally, nrDNA sequence data may display intra-individual polymorphisms that are not associated with the formation of a hybrid individual but rather due to the evolution of pseudogenes (nonfunctional repeats)
or to the evolution of functional, but divergent, paralogous gene copies. Despite these
cautions, the ITS region has successfully been used to reveal the hybrid origin of several
allopolyploid taxa. Rauscher et al. (2002; 2004) used nrDNA homeologous loci to detect
and characterize hybridity in the *Glycine tomentella* allopolyploid complex, Guiggisberg
et al. (2006) found complete additivity at several diagnostic polymorphic ITS sites which
supported an allopolyploid origin for *Centaurium bianoris* and, in conjunction with
cpDNA, revealed the paternal source of the hybrid, and Kaplan and Fehrer (2007) found
that direct sequencing of the nrDNA ITS region revealed the presence of three ITS
variants which corresponded to the three parental taxa of the hexaploid *Potamogeton ×
torssanderi*.

In allopolyploid species, nuclear genes are duplicated as homeologues so that
tetraploid species are expected to have two divergent sequence types, hexaploid species
three, and octoploid species four divergent sequence types (Wendel 2000). The presence
of homeologues is frequently evidenced as polymorphic nucleotide positions through
direct sequencing of consensus nrDNA PCR products (e.g. Baldwin et al. 1995; Sang et
al. 1995; Barkman and Simpson 2002; Andreasen and Baldwin 2003). The expectation is
that the genome of the allopolyploid will be additive with respect to its parental species
(Wendel 2000). Molecular cloning can then separate the divergent sequence types for
phylogenetic analysis with the parental sequence types (e.g. Barkman and Simpson 2002;
Andreasen and Baldwin 2003; Siripun and Schilling 2006). This assumes the hybrid has
maintained both parental genome complements and that the sequences will be
phylogenetically sister to their orthologues from the respective parental species (Wendel
2000).
The interpretation of nrDNA sequence data for the study of polyploid origins may seem straightforward but, when different nrDNA sequences are united within a single genome through allopolyploid hybridization or introgression, the speed and direction of homogenization cannot be predicted and may not be consistent across different descendant lineages (Álvarez and Wendel 2003). Concerted evolution can act to completely or partially eliminate one of the parental homeologues from hybrid individuals thus obscuring the inference of parental genome donors from the sequence data (e.g. Wendel et al. 1995; Fuertes Aguilar et al. 1999a). However, sequence polymorphisms that are additive between parental sequence types persists in many plant groups (e.g. Sang et al. 1995; Campbell et al. 1997; Ge et al. 1999; Barkman and Simpson 2002; Andreasen and Baldwin 2003; Saito et al. 2007; Schilling et al. 2008) as concerted evolution can be slowed by processes such as high mutation rates, the loss of sexual recombination, the presence of nrDNA loci on nonhomologous chromosomes, and also the specific chromosomal location of the repeat types (Wendel 2000). However, when hybrids maintain the parental sequence types, those sequences may undergo various degrees of genomic recombination and produce chimeric sequence types as demonstrated by Barkman and Simpson (2002). These chimeric sequences will be evidenced by varying degrees of additivity and will resolve basally to parental taxa in phylogenetic analysis (Sang et al. 1995; Álvarez and Wendel 2003; Soltis et al. 2008). Additionally, it can be difficult to infer the origin of relatively old allopolyploids because the extant parental genomes may have diverged significantly from those of the original genotype donors or one or more of the parental taxa may have become extinct. It is generally believed that recently formed allopolyploids will maintain the parental repeat types but Fuertes Aguilar
et al. (1999a) demonstrated that concerted evolution can act to eliminate a parental sequence type as early as the F$_2$ generation.

This chapter describes the use of the nrDNA ITS1 and ITS2 spacers and intervening 5.8S gene, and the 3' end of the nrDNA ETS spacer to investigate the putative hybrid origin of *Solidago houghtonii*.

Materials and Methods

**Taxon Sampling**

Samples for DNA analysis were primarily obtained from field collections of natural populations and in a few cases from purchased seed or plants. For each individual sampled, a single leaf was obtained from the stem, just below the inflorescence, and silica gel-dried following the protocol of Chase and Hills (1991). In total, 26 *Solidago* taxa were sampled. *Euthamia graminifolia*, *Ericameria nauseosus*, and *Heterotheca villosa*, were chosen as outgroups because they belong to the same tribe and subtribe (Astereae: Solidagininae) as *Solidago* and because recent phylogenetic studies have shown them to be closely related to *Solidago* (Bremer 1994; Goertzen et al. 2003; Urbatsch et al. 2003; Beck et al. 2004). Voucher data and collection localities are presented in Appendix 1. Michigan plant material of *Solidago houghtonii* was collected under Michigan Department of Natural Resources endangered species permits #00-1133 and #00-1257 and Federal Fish and Wildlife permit TE074805-0. Canadian material was collected with permission from the Ontario Ministry of Natural Resources and private landowners. The New York collections were made with permission of the Bergen Swamp Preservation
Society. Vouchered specimens for each species examined in this study are deposited in the Western Michigan University (WMU) Hanes herbarium unless otherwise noted.

To investigate the parentage of *S. houghtonii*, taxon sampling was designed to include all species of *Solidago* that are sympatric with my species of interest. In addition, at least one representative from each section and subsection of the genus *Solidago* sensu Semple and Cook (2006) was included to represent the phylogenetic breadth of the genus. All members of *Solidago* section *Ptarmicoidei* (House) Semple and Gandhi (2004), of which *S. houghtonii* is a member and likely derived, were also included in the sampling. Multiple samples of *S. houghtonii*, of all proposed parental taxa (*S. ptarmicoides, S. ohioensis, S. riddellii, and S. uliginosa*), as well as *S. gigantea*, were collected. For the majority of species included in this study, multiple individuals were sequenced for some markers; however, only those included in the phylogenetic analyses are reported as no intraspecific phylogenetically informative variation was observed. Data for additional ITS and ETS sequences were obtained from GenBank (Appendix). To investigate the alternate hypotheses of a single or polytopic origin of *S. houghtonii*, samples of *S. houghtonii* were collected from across the geographic range of the species corresponding to each of Pringle’s four entities (Fig. 2).

**DNA Extraction**

Leaves were stored at room temperature in individual plastic bags containing silica gel until DNA extraction. Total DNA was extracted from approximately 1 cm² of dried leaf tissue which was added to a 1.5 μl UV-sterilized Eppendorf tube. Multiple samples (up to 24) were typically extracted at one time. The 2xCTAB (2% hexadecyltrimethylammonium bromide) extraction method of Doyle and Doyle (1987),
modified with the addition of 1% PEG8000 to assist in the precipitation of polysaccharide-like components from the sample (Li et al. 1994), was followed by aliquoting 500 μl of room temperature extraction buffer to each tube. A 2.5 μl aliquot of 2-mercaptoethanol was added and the leaf tissue was further ground until a dark green color was visible. When multiple samples were prepared at one time, the leaf/buffer mixture was held on ice until all preparations were complete. Samples were mixed well by vortexing for 30 seconds and then incubated in a water bath at 71°C for 30 minutes.

After incubation the samples were placed on ice for a minimum of 1 minute and then 500 μl of -20°C SEVAG (24:1 chloroform:isoamyle alcohol) was added. Samples were again vortexed for about 10 seconds and then centrifuged at high speed for 3 minutes. The aqueous (top) layer was removed and added to a new UV-sterilized 1.5ml Eppendorf tube. Pellets of plant tissue, precipitated proteins and carbohydrates were discarded. A 500 μl aliquot of -20°C SEVAG was added to each tube, samples were vortexed for 10 seconds and again centrifuged at high speed for 3 minutes. The aqueous layer was again removed and transferred to a UV-sterilized 1.5 μl O-ring collection tube. Pelleted waste products were discarded. A 500 μl aliquot of -20°C isopropanol was added to each tube; tubes were capped and gently mixed by inverting 3 – 4 times. The DNA-isopropanol mixture was then centrifuged at high speed for 10 minutes.

Following final centrifugation the supernatant was carefully decanted and the DNA pellet was washed with 400 μl of -20°C 95% ethanol for 1 minute. Following the wash the ethanol was decanted from the pellet. The uncapped tubes with DNA pellets were laid on their sides, covered with a clean paper towel, and allowed to air-dry overnight. Dried DNA was resuspended in 100 μl of sterile deionized water and stored at -20°C.
PCR Amplification and Sequencing

Total genomic DNA was used in all amplification reactions. For each sampled individual the entire nrDNA ITS region, an 837 bp fragment which includes ITS1, ITS2, and the intervening 5.8S gene, was amplified by Polymerase Chain Reaction (PCR) (Mullis and Faloona 1987) using the primers 17SE and 26SE (Sun et al. 1994) (Fig. 3). PCR amplifications were performed in 50 µl volumes containing 35.4 µl sterile water, 5.0 µl of the provided 10X enzyme buffer, 2.5 µl magnesium chloride (50 mmol), 2.0 µl total dNTPs (10 mmol) (USB Corp.), 1.0 µl (10 µM) of each primer, 0.1µl of Platinum® Taq polymerase (Invitrogen Corp.), and 1µl of template DNA diluted 1:10. The reaction mixtures were cycled on an Eppendorf Mastercycler Programmable Thermal Cycler (Eppendorf Scientific Inc.). Temperature and cycling conditions consisted of an initial 3 min denaturation at 94°C, followed by 34 cycles of 1 min denaturation (94°C), 1 min annealing (55°C), and 1 min 20 sec elongation (72°C), followed by a 5 min final extension at 72°C. A 559 bp fragment from the 3' end of the ETS (hereafter referred to as 3'ETS) was amplified for each sampled individual using primers 18S-R1 (Urbatsch et al. 2003) and Ast-8 (Markos and Baldwin 2001). Temperature and cycling conditions consisted of an initial 3 min denaturation at 94°C, followed by 40 cycles of 1 min denaturation (94°C), 1 min annealing (55°C), and 1 min 20 sec elongation (72°C), followed by a 7 min final extension at 72°C. PCR products were verified on a 1.5% agarose TAE (Tris, Acetic Acid, and EDTA) gel using a 100 bp size standard (Invitrogen Corp.). All PCR products were purified using either the QIAquick PCR Purification kit (QIAGEN, Inc.) or AMPure magnetic bead PCR purification system (Agencourt Bioscience) following the manufacturer’s protocols. Both strands of the purified PCR products were sequenced using a Beckman-Coulter CEQ2000XL Genetic Analyzer and
the manufacturer’s kits and protocols. Sequences for both the ITS and 3'ETS regions were obtained using their respective PCR primers. In order to ensure complete forward and reverse reads through the entire ITS region, sequences were also obtained using the internal primers ITS 2 and ITS 3 (White et al. 1990) (Fig. 3).

**Confirmation of ITS Homology**

Numerous authors have cautioned against the use of paralogous sequences in phylogenetic analysis, therefore, the guidelines provided by Nieto Feliner and Roselló (2007) for obtaining reliable ITS sequence data were followed. Using one of the ITS sequences obtained from *Solidago houghtonii*, a BLAST search was performed in GenBank which indicated significant alignment with high scores to other ITS sequences from the *Asteraceae*. All additional ITS sequences were aligned to the one used in the initial BLAST search and no differences in length were observed. Electropherograms of the direct sequences were inspected for polymorphic sites and/or “dirty” sequence. As expected, direct sequences of *S. houghtonii* indicated the presence of several polymorphic positions. Also, forward and reverse sequence reads from one of the hypothesized parental species, *S. ptarmicoides*, were observed to be clean up to a point and then become unreadable, indicating the presence of ITS types of slightly different length. Sequencing of two additional *S. ptarmicoides* individuals produced the same results (data not shown). Therefore, four *S. houghtonii* individuals (accessions HOU002, HOU008, HOU015, and HOU060), and one of *S. ptarmicoides* (accession PTR011), were cloned (see below). *S. gigantea* (GIG004) had a polymorphism at position 585 in ITS2. The remaining species of *Solidago* showed no within individual ITS polymorphisms and had “clean” sequence.
Pairwise comparisons of base pair differences and the presence of indels across sequences of the normally highly conserved 5.8S gene have been used to determine the presence of nonfunctional pseudogenes within ITS sequence data (Buckler et al. 1997; Hershkovitz et al. 1999; Hughes et al. 2002). Therefore, the conserved 5.8S region was investigated and found to be 164 bp in length for all Solidago individuals and clones, well within the expected 160 – 170 bp. Additionally, there was very little sequence divergence in the 5.8S region. The Solidago 5.8S sequences differed at only one nucleotide position from the published consensus sequence (Goertzen et al. 2003) for the Astereae. Therefore, the lack of 5.8S variation is taken as an indicator of functionality for all directly amplified and cloned nrDNA sequences.

Cloning of the 3'ETS-ITS nrDNA Region

Due to the presence of polymorphic nucleotide positions in direct sequences of purified PCR products from all samples Solidago houghtonii individuals, and also S. ptarmicoides, the entire 3'ETS – 18S gene – ITS1 – 5.8S gene – ITS2 region, a total of 3055bp, was cloned using a TOPO TA cloning kit (Invitrogen) following the manufacturer’s protocol. By cloning this entire region it was possible to positively link the 3'ETS sequence with the ITS sequence from the same clone. Approximately 70 colonies from each cloning reaction were selected and further grown on LB agar streak plates containing 100µg/mL ampicillin. Colonies were then screened for inserts of the appropriate length by direct PCR using primers Ast-8 and 18S-R1 (Markos and Baldwin 2001 and Urbatsch et al. 2003 respectively) for the 3'ETS region and primers 17SE and 26SE (Sun et al. 1994) for the ITS region. Temperature and cycling conditions consisted of lysing the cells and initial denaturation for 10 min at 94°C, followed by 35 cycles of 1
min denaturation at 94°C, 1 min 20 sec annealing at 55°C, and 3 min elongation at 72°C, with a 30 min final extension at 72°C. For each cloning reaction, the PCR products from 20 to 30 positive colonies were cleaned and sequenced as described above using the respective 3'ETS and ITS primers. Similar clones were grouped, and their consensus sequence was used in phylogenetic analysis (see results).

Phylogenetic Analyses

For each marker, contigs were assembled in Sequencher ver. 4.01 (Gene Codes Corporation, Inc.). Sequence alignments were made using CLUSTAL_X v1.8 (Thompson et al. 1997) and further adjusted by eye. Complete sequences have been deposited in GenBank (Appendix). Sequence alignments were straightforward for much of their lengths, but in the regions where insertions and deletions (indels) occurred, manual adjustments were necessary and gaps were inserted. The nrDNA 3'ETS and ITS data sets were combined and analyzed together as they are both part of the same transcriptional unit (Baldwin and Markos 1998; Markos and Baldwin 2001).

Several authors (e.g. Golenberg et al. 1993; Vijverberg et al. 1999; Freudenstein and Chase 2001; Ingvarsson et al. 2003) have shown that indels can be extremely useful for inferring phylogenetic relationships both at the intra- and interspecific level. Therefore, indels in the aligned nrDNA data set were coded and added to the data matrix as separate binary characters. Indels with identical 5' and 3' termini were treated as homologous and those that differed in their 5' and/or 3' termini as having been derived independently of one another. Following the simple indel-coding method of Simmons and Ochoterena (2000), the character state 1 was assigned when the indel was present and
the character state 0 if the indel was absent. 3.46% of the nrDNA data matrix cells were scored as missing data.

All phylogenetic analyses were performed using PAUP* 4.0b4a (Swofford 1998). Maximum parsimony (MP) analysis of the nrDNA data set was performed using the heuristic search algorithm with 100 random taxon addition sequences, TBR (tree bisection reconnection) branch swapping, and the MulTrees option selected. Tree statistics and measures of homoplasy were calculated using PAUP* and the multiple most parsimonious trees were combined to construct a strict consensus tree. The robustness of clades was assessed using Jackknife analyses (Farris et al. 1996). As in Freudenstein et al. (2004), PAUP* was set to emulate Jac resampling, 2000 pseudoreplicates, 37% character deletion, TBR branch swapping with MulTrees on, and 1 random addition search replicate per jackknife replicate with 2 trees saved in each search replicate.

A maximum likelihood (ML) analysis (Felsenstein 1981) was also performed on the nrDNA data set using the best-fit model GTR+G+I (Rodríguez et al. 1990), as chosen from likelihood ratio tests in Modeltest 3.06 (Posada & Crandell 1998). A heuristic search was performed with 1 random addition sequence and TBR branch swapping with the MulTrees option selected.

An unrooted haplotype network was constructed for the nrDNA data set using the parsimony method of Templeton et al. (1992) as implemented in the TCS 1.21 program (Clement et al. 2000). *Solidago rigida* and *S. ohioensis* sequences obtained from GenBank had IUPAC coded nucleotide positions indicating the presence of any base (N). These caused problems with the TCS analysis. Following Rauscher et al. (2004) the ambiguous positions were scored as identical to the individual that was most similar at all other sites and included in the analysis. Gaps were treated as missing data, but an
additional three characters reflecting indel structure were added to the data matrix. The 95% probability limit of parsimonious connections was applied to produce the network.

**Phylogenetic Test of Hybridization**

The hypothesis that *Solidago houghtonii* is of hybrid origin was phylogenetically tested by comparing alternative topologies as in Barkman and Simpson (2002). Using MacClade 4 (Maddison and Maddison 2001), a constraint topology that forced the cloned nrDNA homeologue types of *S. houghtonii* to be monophyletic (nonhybrid origin) was designed. This constraint was applied to ML searches using the combined nrDNA data set and maximum likelihood criteria. The ML score from an unconstrained analysis, where relationships of the cloned homeologue types were freely estimated, was compared to the score from the constrained analysis using the conservative Shimodaira-Hasegawa (S-H) test (Shimodaira and Hasegawa 1999).

**Results**

**3'ETS and ITS Sequence Variation**

In general, the nrDNA 3'ETS and ITS sequence data displayed a lack of divergence among the 26 *Solidago* species sampled in this study. A similar lack of sequence variation was observed for *Solidago* by Schilling et al. (2008). Careful observation of sequence electropherograms revealed that most species lacked evidence of single site polymorphisms. However, *S. ptarmicoides* had detectable polymorphisms at positions 55 (C/G) and 484 (C/T) in the 3'ETS sequence, at position 698 (G/T) in the ITS2 sequence, and became unreadable in ITS1, indicating the possibility of multiple
ITS1 types of differing length. Cloning of *S. ptarmicoides* revealed that it was heterozygous for divergent ITS types (Table 1). *Solidago ptarmicoides*-Type 1 has a three bp deletion in ITS1 (positions 140 – 142 of the ITS alignment), a thymine (T) at position 698 in ITS2 and a cytosine (C) at positions 55 and 484 in the 3'ETS. *Solidago ptarmicoides*-Type 2 has a five base pair deletion in ITS1 (positions 138 – 142 of the ITS alignment), a guanine (G) at positions 698 in ITS2 and 55 in the 3'ETS, and a thymine (T) at position 484 in the 3'ETS.

Sequence divergence between four of the morphologically distinct species of section *Ptarmicoidei* (*S. nitida*, *S. ohioensis*, *S. ptarmicoides*, and *S. rigida*) is notably lacking; pairwise differences ranged from 0% - 0.37%. One bp difference was found between sequences of *S. rigida* (RIG002) and *S. ohioensis* (OHI005) and zero differences were found between *S. rigida* and the GenBank accession of *S. ohioensis*. There were two bp differences between *S. nitida* (NIT001) and *S. ohioensis* (OHI005) and three bp differences between *S. nitida* (NIT001) and *S. rigida* (RIG002). Only ITS1 length variation was found between *S. ptarmicoides*-Type 1 and *S. ohioensis* (OHI005). *S. ptarmicoides*-Type 2 was the most divergent with ITS 1 length variation and 3 bp differences when compared to *S. ohioensis* (OHI005). These results are consistent with Beck et al. (2004) who showed that sequence divergence between *S. shortii* and *S. discoidea*, two other morphologically distinct *Solidago* species, was also low. In fact, low levels of sequence divergence between closely related species has frequently been observed to be less than 1% for the nrDNA ITS and ETS markers (Hershkovitz et al. 1999). In contrast, there were 7 – 9 bp differences between the nrDNA sequence of *S. riddellii* (section *Ptarmicoidei*) and all other section *Ptarmicoidei* members with pairwise differences ranging from 0.52% - 0.75%.
Following direct sequencing of PCR products from four individuals (accessions HOU002, HOU008, HOU015, HOU060) of the polyploid *Solidago houghtonii*, several single site polymorphisms were observed in the electropherograms of aligned 3'ETS and ITS sequence data. In total, twelve sites were polymorphic among the four *S. houghtonii* individuals, but only three of these positions were found in all four individuals (Table 1).

**Homeologous Sequence Types**

Molecular cloning of the four *S. houghtonii* individuals, representing Pringle’s four hypothesized entities (USFWS 1997; Fig. 2), was performed. A total of 111 clones were sequenced for the 3'ETS and ITS regions; 30 clones from individual HOU002 representing Northern Michigan, 22 from individual HOU008 representing Canada, 30 from individual HOU015 representing the New York disjunct population, and 29 clones from individual HOU060 representing the Michigan disjunct population. In the process of independently aligning the cloned sequences from each individual, it was observed that relatively few were completely identical. It appeared that many of the variable sequences were the result of “Taq errors”, “recombination” (PCR mediated or meiotic), or a combination of these processes. When PCR products from heterogeneous DNA templates (e.g. allopolyploids) are cloned, they frequently produce PCR recombinants (Bradley and Hillis 1997; Cronn et al. 2002), and the multicopy rDNA ETS and ITS markers seem especially prone to PCR recombination (Kovarik et al. 2005).

PCR recombination involves the in vitro formation of erroneous sequences that combine features of the different homeologues in the template DNA into a single sequence. The process occurs when DNA polymerases, such as the *Taq* polymerase (from *Thermus aquaticus*) used in this study, dissociate from the template, producing
incomplete PCR products which then act as primers in subsequent replication cycles. If the partially amplified sequence re-anneals to a different template, the resulting PCR product will be a recombinant of the two homoeologous sequences (Bradley and Hillis 1997; Cronn et al. 2002). The amplification of large PCR products for use in cloning reactions, such as the approximately 3055 bp product used in this study, seems to be especially prone to the formation of recombinants. Bradley and Hillis (1997) and Cronn et al. (2002) showed that PCR recombination can account for a relatively large percentage of the resulting clones, 43% and up to 89%, respectively. Theoretically, the proportion of PCR recombinants may be expected to increase with ploidy level as the greater number of template strands provides a greater chance of re-annealing to an alternative template. It is generally understood that distinguishing between PCR recombinants and genomic recombinants can be difficult. However, in some cases, PCR recombinants can be detected because *Taq* polymerase has a tendency to insert an adenosine when it reaches the end of a template strand. Therefore, the incomplete PCR product, which acts as a primer during replication, will have an additional adenosine added to the fragment and an autapomorphic A/T nucleotide will be incorporated at the point of recombination (Pääbo et al. 1990). Although true genomic recombination is a viable explanation, most of the recombinants are likely artifacts of cloning large PCR products. However, a few recombinants were amplified from multiple clones and those may represent true genomic recombination (Barkman and Simpson 2002; Kovarik et al. 2005). Because we aligned the cloned sequences with their likely extant diploid progenitors, as examples of the non-recombinant or “ancestral” condition, we were able to confidently identify recombinant sequences. Fifteen (15) recombinants (50%) were identified from individual HOU002. An additional 15 recombinants (68%) were identified from individual HOU008, 10 (32%)
from individual HOU015 and individual HOU060 had 1 recombinant sequence (3%). Most putative recombinants were excluded from the analysis as they occurred almost exclusively in single clones and were believed to be the result of PCR recombination.

For the individual from New York (HOU015), four identical recombinant clones were recovered which showed evidence of two of the hypothesized parental taxa and may be the result of genomic recombination. Therefore, this clone type was included in the analysis. Likewise, the single recombinant clone recovered from individual HOU060 (Michigan disjunct) was included in the analysis because it also showed evidence of two of the hypothesized parental taxa and is also likely the result of genomic recombination.

In addition to PCR recombinants, several cloned sequences contained singleton changes that were attributed to "Taq error". Taq error is associated with DNA polymerases that lack proofreading abilities and are therefore prone to occasionally substituting incorrect nucleotides (Päabo et al. 1990). Essentially Taq errors were ignored by using a consensus sequence from clones of similar type in the analysis. Four consensus homeologue types from clones of *S. houghtonii* individual HOU002, three from individual HOU008, five from individual HOU015, and six from individual HOU060 were identified (Table 1).

*Solidago houghtonii* individual HOU002, representing Pringles entity 1 (Northern Michigan) shows perfect additivity of polymorphic sites between *S. riddellii*, *S. ohioensis*, and *S. ptarmicoides*. Individual HOU008 (entity 2 - Canada) shows perfect additivity of polymorphic sites between *S. riddellii* and *S. ohioensis*. Individual HOU060 (entity 3 – Michigan disjunct) and individual HOU015 (entity 4 – New York disjunct) shows perfect additivity between *S. ohioensis* and *S. ptarmicoides* and partial additivity of *S. riddellii*.

In addition to *Solidago ptarmicoides* and the polyploid *S. houghtonii*, both of which were cloned, the *S. gigantea* individual (accession GIG004) had single site
Table 1. Nuclear Ribosomal DNA 3'ETS and ITS Nucleotide Site Comparison for *Solidago houghtonii* Homeologues and Progenitor Taxa

<table>
<thead>
<tr>
<th>Nucleotide Sites</th>
<th>3'ETS</th>
<th>ITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entity Taxa and Clone Types</td>
<td>E T S</td>
<td>I T S</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td><em>S. ptarmicoides</em> - Type 1</td>
<td>1 G A G A C AG A G A G A G A G A G T T G T</td>
<td>(1)</td>
</tr>
<tr>
<td><em>S. ptarmicoides</em> - Type 2</td>
<td>1 G A G A C AG A G A G A G A G A G T T G T</td>
<td>(1)</td>
</tr>
</tbody>
</table>

Note: Consensus clone types from PCR products of *S. houghtonii* individuals representing each of the four geographic entities are shown followed by the number of clones recovered per consensus type. *S. riddellii* informative bases are shaded and *S. ptarmicoides* informative bases are in bold face type. * = deletion, R = A&G, Y = C&T, M = A&C, S = C&G, K = T&G.

polymorphisms; one in the ITS and two in the 3'ETS markers. No ITS polymorphisms were observed in the remaining 23 sampled species of *Solidago*, but 5 of the 23 species were found to possess a single polymorphism in their 3'ETS sequence. For these individuals we coded the polymorphisms using IUPAC coding. Including coded polymorphisms in the MP analyses could increase the number of most parsimonious trees.
and cause lack of resolution (Soltis et al. 2008), but since the majority of the polymorphisms occurred at uninformative sites, and since molecular cloning of these taxa was not expected to increase an understanding of the origin of *S. houghtonii*, those taxa were not cloned and their sequences were included with coded polymorphisms.

In total, sixty sequences including two homeologues recovered from clones of *Solidago ptarmicoides*, eighteen consensus homeologue types recovered from clones of four *S. houghtonii* individuals, ten sequences obtained from GenBank, and three outgroup taxa were aligned to generate a combined 3'ETS and ITS1-5.8S-ITS2 nrDNA data matrix which contained 1455 characters including gaps; 520 characters from 3'ETS, 123 from the 3' end of the 18S gene, 252 from ITS1, 164 from the 5.8S gene, 212 from ITS2, and 69 from the 5' end of the 26S gene. Alignment required the insertion of seventeen gaps, mostly due to indels in the outgroup taxa. Gaps were coded for presence/absence (0/1) (Simmons and Ochoterena 2000). Of the 1455 characters, 178 were variable and 92 (6.3%) were parsimony-informative. MP analysis resulted in 33,522 equally parsimonious trees (335 steps, CI = 0.842 and RI = 0.789). The strict consensus tree is presented (Fig. 4). The ML analysis resulted in 2 trees (-lnL = 3757.7823) that were congruent with the strict consensus from the MP analysis with respect to the placement of species belonging to both section *Solidago* subsection *Triplinerviae* and section *Ptarmicoidei*, and also with respect to the placement of *S. houghtonii* clone types (data not shown).

A notable aspect of the combined *Solidago* nrDNA phylogeny is that section *Ptarmicoidei* is paraphyletic; *S. riddellii* is not resolved with other members of section *Ptarmicoidei* (Fig. 4). This is consistent with Zhang’s (1996) cpDNA RFLP analysis of
Figure 4. Nuclear Ribosomal DNA Strict Consensus Tree

Note: Numbers above the branches represent jackknife support.
Solidago where S. riddellii also did not resolve with other members of section Ptarmicoidei (Zhang’s Corymbosea).

While the nrDNA analysis (Fig. 4) was unable to resolve most relationships within the genus Solidago, it did resolve the genus as monophyletic (Jackknife = 100). Also resolved was a well-supported clade (Jackknife = 92) that contains all members of section Solidago subsection Triplinerviae: S. canadensis, S. altissima, and all sampled individuals of S. gigantea. There are three additional species included in this clade that are not members of the subsection Triplinerviae: S. juncea and S. missouriensis, which are members of subsection Junceae and which form a moderately supported subclade (Jackknife = 72), and S. multiradiata from subsection Multiradiatae. Notably, this clade does not include any clones of Solidago houghtonii. Within this “Triplinerviae Clade”, S. altissima, considered in some classifications (e.g. Gleason and Cronquist 1991) as S. canadensis var. scabra, forms a weakly supported (Jackknife = 60) clade with S. canadensis. Two S. gigantea individuals, both from within the State of Michigan, USA, located approximately 240 km apart, were also weakly supported (Jackknife = 58) as sisters.

Solidago riddellii (section Ptarmicoidei) forms a strongly supported clade (Jackknife = 99) with S. riddellii sequences obtained from GenBank and two S. houghtonii cloned homeologue types: HOU002-#1 representing S. houghtonii entity 1 (Northern Michigan) and HOU008-#1 representing S. houghtonii entity 2 (Ontario, Canada) per Pringle (USFWS 1997, Fig. 2). Solidago houghtonii chimeric homeologue type HOU060-#1 (Pringle’s entity 3 – Michigan disjunct), is weakly supported (Jackknife = 60) as sister to S. riddellii and S. houghtonii homeologue types HOU002-#1 and HOU008-#1. These are identified as the “Riddellii Clade” (Fig. 4). Chimeric
homeologue type HOU015-#1 (Pringle’s entity 4 – New York disjunct), occupies a position within the unresolved polytomy of *Solidago* species.

The remainder of the species within *Solidago* section *Ptarmicoidei* (*S. nitida*, *S. ohioensis*, *S. ptarmicoides*, and *S. rigida*), along with the remainder of homeologues generated from cloning of *S. houghtonii*, form a monophyletic clade (Jackknife < 50) identified here as the “Ptarmicoidei Clade” (Fig. 4). The paucity of sequence variation between species and cloned homeologue types within the “Ptarmicoidei Clade” resulted in a polytomy within which several subclades were resolved. *Solidago nitida* is sister to the GenBank sequence of *S. nitida* (Jackknife = 76). The subclade identified as “A” (Jackknife = 87) is composed of three homeologues, HOU002-#4, HOU060-#4, and HOU015-#4, which are sister to the *S. ptarmicoides*-Type 1 homologue. The “B” subclade (Jackknife = 85) is composed of homeologue HOU060-#6 which is sister to the *S. ptarmicoides*-Type 2 homologue. *Solidago houghtonii* homeologues HOU002-#3, HOU008-#3, HOU015-#3, and HOU060-#3 form a monophyletic clade (Jackknife = 61) identified as the “C” subclade. These homeologues share a unique cytosine (C) at bp 618 in the ITS2. The “D” subclade (Jackknife = 61) is an unresolved polytomy indicating a sister relationship between *S. rigida*, *S. ohioensis* (GB), and two *S. houghtonii* homeologues, HOU015-#5 and HOU060-#5. These homeologues have a guanine (G) at position 352 in the 3'ETS which is shared with *S. rigida* and *S. ohioensis* (GB), but none of the other species of *Solidago* included in the analysis. However, visual inspection of electropherograms of 3'ETS sequence data from 14 additional *S. ohioensis* individuals (data not shown) revealed that five individuals had single site polymorphisms at 1 – 3 nucleotide positions with accession OHI011 having an A/G polymorphism at bp 352, suggesting that *S. ohioensis* is variable for its ETS sequence type. All other nucleotides
within *S. houghtonii* homeologues HOU015-#5 and HOU060-#5 are identical to those of *S. ohioensis* (OHI005) except for an additional autapomorphy in HOU060-#5.

To gain better insight into the relationships between the nrDNA sequences of potential diploid progenitor species and the homeologous repeat types recovered from cloning of *Solidago houghtonii*, we estimated a parsimony haplotype network (Fig. 5). TCS (Clement et al. 2000) calculated a 95% parsimony connection limit of 16 steps and resulted in a haplotype network of the 57 nrDNA sequences obtained from *Solidago* species and clones (outgroup taxa were removed from the data matrix prior to the analysis). Fifty-six haplotypes inferred by TCS were not found in the analyzed individuals and are identified as missing intermediates in the network. Six closed loops, which could not be unambiguously resolved, also occurred in the network. These loops were likely caused by homoplastic alignment positions that prevented TCS from distinguishing between alternative connections. The network is broadly divided into three groups which correspond to the “Ptarmicoidei Clade”, the “Triplinerviae Clade” and the remainder of the *Solidago* species included in the MP analysis (Fig. 4). The “Riddellii Clade” is nested within the group of *Solidago* species.

TCS grouped the 18 nrDNA homeologue repeat types from the four *Solidago houghtonii* individuals (accessions HOU002, HOU008, HOU015, and HOU060) into ten haplotypes because the program collapses identical sequences into single haplotypes (Fig. 5). Each accession of *S. houghtonii* contained a homeologue type that formed a single haplotype with the putative diploid *S. ohioensis* accession OHI005. *Solidago houghtonii* homeologues, HOU002-#3, HOU008-#3, and HOU015-#3 formed a second haplotype which differed from *S. ohioensis* (OHI005) by a single synapomorphy, a unique cytosine...
Figure 5. Statistical Parsimony Network

Note: Lines represent single mutation steps and small circles are hypothetical intermediate repeat types inferred by TCS. Gray lines separate sequence types identified in MP analysis as belonging to the “Triplinerviae Clade”, the “Ptarmicoidei Clade” and the “Riddelli Clade”.
Solidago ptarmicoides possessed two nrDNA repeat types; *S. ptarmicoides*-Type 1 formed a single haplotype with *S. houghtonii* homeologues HOU015-#4 and HOU060-#4 in the network analysis, and homeologue HOU002-#4 differed from this by one mutation step, again attributed to a single autapomorphy. The *S. houghtonii* homeologue HOU060-#6 differed from *S. ptarmicoides*-Type 2 by one autapomorphy. *Solidago ohioensis* (GB) and *S. rigida* (RIG002) formed a single haplotype with *S. houghtonii* homeologues HOU015-#5 and HOU060-#5. This haplotype formed a closed loop with the *S. ohioensis* (OHI005) haplotype and an unsampled intermediate leading to *S. ptarmicoides*—Type 1 suggesting that homoplasy exists.

*Solidago houghtonii* homeologue types HOU002-#1 and HOU008-#1 formed a single haplotype with the putative diploid *S. riddellii* while homeologue types HOU060-#1 and HOU015-#1, which are likely chimeric sequences between *S. riddellii* and *S. ohioensis*, are 11 and 15 mutation steps away from the *S. riddellii* haplotype respectively. Each of these chimeric homeologues is only 6 mutation steps away from the haplotype composed of *S. ohioensis* (GB), *S. rigida*, and two *S. houghtonii* homeologue types.

Test of Hybridization

The phylogenetic test for hybridization was designed to evaluate a hybrid vs. non-hybrid status for *Solidago houghtonii*. The unconstrained analysis resulted in a $-\log$ likelihood (-lnL) of 3767.8 for the hybrid hypothesis, while a $-\ln L$ of 3843.1 was obtained for the monophyletic (non-hybrid) hypothesis. The Shimodaira-Hasegawa test
significantly rejected monophyly of *S. houghtonii* homeologues \((P = 0.002)\), consistent with a hybrid origin for the intra-individual nrDNA homeologue types in *S. houghtonii*.

### Discussion

**Nuclear Evidence of Parentage**

The interpretation of results from the nrDNA sequence data is somewhat complex. This is expected as nuclear DNA can reflect the contributions of multiple donors to the allopolyploid genome. *Solidago houghtonii* is hexaploid so three parental genome donors are expected; although these do not necessarily have to come from three separate species. This study has not only investigated hybridization in a taxon of potential triple hybrid origin, but also one hypothesized to have had four separate origins involving varied combinations of different parental taxa. In effect, a combined analysis of four studies in which each of Pringle’s (USFWS 1997) hypothesized entities is presented. Many *Solidago* species, including potential progenitors, have multiple cytotypes \((2n = 18, 36, \text{ or } 54)\) which must be taken into account when evaluating their potential contribution to *S. houghtonii*. Additionally, *S. houghtonii* exists in fragmented populations across much of its geographic range (entities 1 and 2; Fig. 2) and also in two disjunct populations (entities 3 and 4; Fig. 2). Gene flow is therefore expected to be greatly restricted among populations. This may potentially result in different patterns of genome evolution across the range of the species (Schall et al. 1998) so that nrDNA sequence variation may display geographic structuring (Fuertes Aguilar et al. 1999a; Nieto Feliner et al. 2004). Chapter 4 of this study addresses the population genetic structure of *S. houghtonii*. Lastly, the process of polyploidization is associated with extensive genome restructuring.
and additional gene duplications beyond those expected from the direct combination to parental genomes in the polyploid (Paterson et al. 2000). Therefore, additional sequence types beyond those indicated by ploidy level alone may result (e.g. Mason-Gamer 2004).

With nuclear DNA sequence data, such as the ITS and ETS spacers, all parental taxa contribute to the hybrid offspring. An important feature of these markers, which must be taken into account when contemplating their use for phylogenetic analysis of allopolyploid taxa, is that the homeologous loci inherited from the parental taxa consists of multiple tandemly repeated copies that are frequently homogenized by concerted evolution (Wendel et al. 1995). Even in allopolyploids of recent origin, the divergent repeat types can be rapidly homogenized so that only one parental type is evident (Wendel et al. 1995; Nieto Feliner et al. 2004). However, homogenization is believed to proceed more efficiently within rather than among chromosomes so that divergent rDNA types are expected to persist longer in allopolyploids than in diploid taxa (Lihová et al. 2004). Therefore, polymorphic nucleotide positions can provide evidence for divergent parental rDNA types within an allopolyploid genome (Sang et al. 1995; Whittall et al. 2000). The usefulness of the ITS and ETS markers for my analysis of hybrid origin of S. houghtonii is indicated by the presence of sequence polymorphisms that appeared to be additive between hypothesized parental species.

Another aspect to consider when using nrDNA for the inference of allopolyploid parentage involves its behavior during the molecular cloning process used to identify the parental types. PCR amplification of the allopolyploid results in a homeologous mixture of parental rDNA types from which high frequencies of artificial recombinants can result (Bradley and Hillis 1997; Cronn et al. 2002). Additionally, cloning of higher ploidy level taxa, such as S. houghtonii, further increases the likelihood of recombinants. Due to the
strong hypotheses with regards to extant parental species, cloning was deemed suitable for the analysis of parentage in *S. houghtonii*. Again, the usefulness of the ITS and ETS markers for the assessment of hybrid origin was indicated by the phylogenetic analyses of cloned homeologue types.

An additional consideration is that phylogenetic analyses using traditional methods, such as maximum parsimony and maximum likelihood, assume a hierarchical structure between genes sampled from different species. Phylogenetic trees describe this hierarchical pattern of ancestry in a bifurcating diagram. Hybridization, however, results in histories that may not be adequately modeled by a bifurcating tree since the genes sampled from allopolyplloid species represent reticulate rather than hierarchical relationships (Posada and Crandall 2001). Network analysis takes into account processes that act at the species level, such as the coexistence of ancestors with their derivatives, multifurcations, and reticulations. Therefore, more of the phylogenetic information contained in a data set is incorporated by a network analysis (Posada and Crandall 2001). Because MP analysis may not completely reveal the phylogenetic information contained in the nrDNA data set, the sequences were additionally analyzed with a statistical parsimony network approach. In statistical parsimony, genealogical relationships are reflected as like sequences that are collapsed into haplotypes and single mutation steps separate adjacent haplotypes in the network (Posada and Crandall 2001).

**Parentage by *Solidago riddellii***

Maximum parsimony analysis of the nrDNA sequences indicates *Solidago* section *Ptarmicoidei* is paraphyletic. Species within *Solidago* section *Ptarmicoidei* resolved into two clades, the “Ptarmicoidei Clade” and the “Riddellii Clade” (Fig. 4) with seven
informative nucleotide positions separating them (Table 1). Molecular cloning of the entire 3'ETS – ITS regions from four individuals of *S. houghtonii* resulted in consensus homeologue types that nested within both clades. MP analysis indicted that one class of cloned homeologue types, represented by HOU002-#1 and HOU008-#1, was closely related to *S. riddellii*. Statistical parsimony analysis supported the MP analysis (Fig. 4) by collapsing the sequences of two homeologues and two *S. riddellii* individuals into a single haplotype. This provides evidence that *S. riddellii* is one progenitor of the allohexaploid *S. houghtonii*.

It is not surprising that *Solidago riddellii* is a contributor to the genome of *S. houghtonii* since it was proposed by Semple et al. (1999) and by Pringle (USFWS 1997), albeit only for entity 1 (Northern Michigan), to be a progenitor species. However, a relative paucity of *S. riddellii* homeologues were recovered from cloning of *S. houghtonii* individuals representing each geographic entity: two from accession HOU002 (entity 1 – Northern Michigan) and one from HOU008 (entity 2 - Canada); although several *S. riddellii* recombinants were recovered from each. Accession HOU060 (entity 3 – Michigan disjunct) and HOU015 (entity 4 – New York disjunct) each yielded chimeric *S. riddellii* repeats. The paucity of *S. riddellii* homeologues may be due to unpredictable events during PCR amplification such as PCR selection or drift. Alternatively, the detection of a nrDNA homeologue within a polyploid genome is dependent on its copy number (Rauscher et al. 2002) and the number of genomic rDNA copies for any of the species included in the study is unknown. Additionally, differential partial homogenization (Soltis et al. 2008) of nrDNA homeologues through the evolutionary process of concerted evolution may be acting to eliminate the *S. riddellii* homeologue from the polyploid genome of *S. houghtonii*.
Parentage by *Solidago ptarmicoides*

A second class of homeologue type cloned from *Solidago houghtonii* was found to be closely related to *S. ptarmicoides*. Two divergent rDNA sequence types were identified from *S. ptarmicoides* accession PTR011. Homeologues were recovered from the genomes of three *S. houghtonii* individuals HOU002, HOU060, and HOU015 (entities 1, 3, and 4 respectively) that resolved with the *S. ptarmicoides*-Type 1 sequence in the MP analysis. Network analysis indicated *S. ptarmicoides*-Type 1 sequence and homeologues HOU015-#4 and HOU060-#4 share a common haplotype; while homeologue HOU002-#4 was distinguished from *S. ptarmicoides*-Type 1 by a single base pair autapomorphy. An additional homeologue was recovered from HOU060 (entity 3) which resolved with the *S. ptarmicoides*-Type 2 sequence in the MP analysis. Network analysis indicated that the HOU060-#6 homeologue is one mutation step away from the *S. ptarmicoides*-Type 2 homeologue which we attribute to a single autapomorphy. Therefore, as hypothesized by Morton (1979), Pringle (USFWS 1997 for each entity), and Semple (Semple et al. 1999) and as indicated by both MP and network analysis, *S. ptarmicoides* is a second parental contributor to the genome of the allohexaploid *S. houghtonii*.

*Solidago ptarmicoides* homeologues were not recovered from *S. houghtonii* individual HOU008 (entity 2 – Canada). Rauscher et al. (2004) showed that in tetraploids where universal primers yielded clean sequence (no observable polymorphisms) for a single ITS homeologue, the second ITS homeologue could be amplified with repeat-specific primers. Assuming that *S. ptarmicoides* would be present in this individual, as it was in all other individuals, we attempted to design repeat specific and exclusion primers; however, likely due to the lack of sequence divergence between *S. ohioensis* and *S.*
The procedure failed to amplify the missing homeologue and instead amplified the *S. ohioensis* homeologue. In addition to low copy number, there are several other possible explanations for why the *S. ptarmicoides* homeologue was not recovered from *S. houghtonii* individual HOU008 representing entity 2 (Ontario, Canada). These range from PCR associated problems (See Raucher et al. 2002) to incomplete sampling of clones or evolutionary processes such as concerted evolution. With regard to incomplete sampling, we screened a seemingly large number of clones, 22, for HOU008. But Small et al. (2004) indicated that screening a minimum of 40 clones is required to recover all homeologues from higher ploidy level individuals and *S. houghtonii* is hexaploid. Biased concerted evolution may have eliminated *S. ptarmicoides* or perhaps the *S. ptarmicoides* homeologous locus has been silenced or lost in individuals from this geographic region. Alternatively, *S. ptarmicoides* may not have been involved in the parentage of individuals from this geographic region although morphology suggests otherwise.

**Parentage by Solidago ohioensis**

A third class of cloned homeologue type was recovered that corresponded to a C/T polymorphism at bp 618 in ITS2 of *Solidago houghtonii*. We could not account for the origin of the cytosine (C) as it was not present in any of the sampled *Solidago* species or outgroup taxa. The C at bp 618 was the only character to separate this homeologue from the rDNA sequence of *S. ohioensis* (OHI005) for *S. houghtonii* individuals HOU002, HOU008, and HOU015. Whereas two and three character differences separated it from *S. rigida* and *S. nitida* respectively (See nrDNA parsimony network analysis Fig. 5). This unique “C” homeologue may represent a parental taxon which is missing in my dataset or is now extinct. Because the genus was broadly sampled
including all sympatric taxa, the probability that the parental species corresponding to this rDNA type is extant, but not sampled, is rather low; although, an extinct taxon remains a possibility. Parsimony network analysis (Fig. 5) suggests this homeologue and S. 

\textit{ohioensis} share recent common ancestry as it is only 1 mutation step (the synapomorphic “C” at bp 618) away from the \textit{S. ohioensis} (OHI005) haplotype. The fact that this unique “C” homeologue is found in all \textit{S. houghtonii} individuals from across the geographic range of the species suggests either it was present in the diploid ancestral population of \textit{S. ohioensis} prior to the hybridization event which gave rise to \textit{S. houghtonii} or that it arose in an early polyploid ancestor that gave rise to the modern \textit{S. houghtonii}. Regardless of the origin of this haplotype, its presence in all individuals is suggestive of a single origin of \textit{S. houghtonii}.

A fourth class of cloned homeologue type was recovered from \textit{S. houghtonii} individuals HOU015 and HOU060 that was not recovered from individuals HOU002 or HOU008. This homeologue type, HOU015-#5 and HOU060-#5, forms a common haplotype with \textit{S. rigida} (RIG002), and \textit{S. ohioensis} (GB) in the network analysis (Fig. 5). Because these cloned homeologue types are identical at all nucleotide positions to both \textit{S. rigida} (RIG002) and \textit{S. ohioensis} (GB), we can not infer its parental donor. If one considers the potentially high level of 3'ETS variability within \textit{S. ohioensis}, that bp 352 is the only nucleotide position separating my \textit{S. ohioensis} (accession OHI005) from the GenBank sequence, and that \textit{S. rigida} is morphologically quite distinct from other members of section \textit{Ptarmicoidei}, then \textit{S. rigida} would be ruled out and \textit{S. ohioensis} accepted as the source for this homeologue type. Parsimony network analysis clearly indicates that \textit{S. ohioensis} is the third parental donor of the allohexaploid \textit{S. houghtonii}. 

58
The additional cloned homeologue types of *S. houghtonii*, HOU002-#2, HOU008-#2, HOU015-#2, and HOU060-#2, were not resolved with any extant species or with any other clone types in the MP analysis; instead, they formed a polytomy (multifurcation) due to the lack of parsimony informative characters (Fig. 4). Thus, although these homeologues clearly represent Ptarmicoidei-type sequences, MP analysis was unable to directly indicate parentage for *S. houghtonii* from these homeologues. Network approaches, however, are particularly well suited for handling this type of data (Posada and Crandall 2001). The network analysis indicated that homeologues HOU002-#2, HOU008-#2, HOU015-#2, and HOU060-#2 from each of the four geographical entities are completely identical to the homeologue of *S. ohioensis* (OHI005), forming a single haplotype in the network.

**Alternative Explanations for Molecular Results**

The Shimodaira-Hasegawa test indicated that within-individual cloned nrDNA homeologues were not monophyletic; instead they arose from various members of section *Ptarmicoidei*. The polyphyletic origin of *S. houghtonii* nrDNA homeologue variants is consistent with the hypothesis of a hybrid origin for this species. However, processes other than hybridization could account for within-individual sequence variation. For example, recent gene duplication could produce within-individual sequence variation. In this case, one would expect that the divergent sequences would be closely related in phylogenetic analysis. This was not the case for sequences recovered from *S. houghtonii*. Additionally, one would not expect the varied sequence types to be monophyletic with the proposed progenitor species if they were the result of recent gene duplication. If the gene duplication was more ancient, then the variation could be due to divergent paralog
sequences or pseudogenes. Careful assessment of homology following the guidelines set forth by Nieto Feliner and Roselló (2007) indicated that the variable sequences recovered from *S. houghtonii* were likely not pseudogenes or divergent paralogs but rather duplicated homeologues resulting from hybridization.

**Morphological Support of Hybridization**

A common assumption regarding hybrid species is that they are morphologically intermediate and display a blend of parental characters rather than novel ones. Rieseberg and Ellstrand (1993) showed that F$_1$ hybrids are a mosaic of both parental species and display intermediate morphological characters. They also showed that 89% of later generation hybrids display extreme or novel characters. The morphology of *S. houghtonii* clearly “favors” that of *Solidago* section *Ptarmicoidei* whose members have a corymbiform inflorescence and a basal rosette of more or less linear leaves. More specifically, *S. houghtonii* strongly resembles the naturally occurring hybrid of *S. ohioensis* × *S. ptarmicoides* as they both have a strigillose pubescence on the peduncles and large capitula. These characters are used to distinguish *S. houghtonii* from narrow-leaved forms of the morphologically similar *S. ohioensis*, with which it often occurs, as *S. ohioensis* has glabrous peduncles and small capitula. *Solidago houghtonii* also resembles *S. riddellii* as its lower stem and basal leaves are slightly conduplicate. *Solidago riddellii* has conduplicate leaves and also pubescent peduncles. Lastly, the large ray florets of *S. houghtonii* are similar in size to the ray florets of *S. ptarmicoides* and the inflorescence of *S. houghtonii* appears to be intermediate in density and flower-head number (2−) 5−50 (−100+) to those of either *S. riddellii* or *S. ohioensis*. The flowering heads of *S. ptarmicoides*, usually 1 − 25 (−50) per inflorescence, are widely separated giving an open
appearance to the inflorescence. In marked contrast, the inflorescences of both *S. riddellii* and *S. ohioensis* are very dense with *S. riddellii* having 30 – 450 capitula and *S. ohioensis* having 10 – 500+ capitula per inflorescence (See Semple and Cook 2006 for morphological descriptions). Edwards-Wilson’s (1999) microscopic analysis of *S. houghtonii* and its putative parents found the pappus of *S. houghtonii* to be most like that of *S. ptarmicoides*, robust with a clavate tip, and the transverse sectioned achenes of *S. houghtonii* to be similar to those of *S. riddellii*. Additionally, the marginal leaf cilia of *S. houghtonii* displayed characteristics of two of the putative parents: a long sharp point similar to that of *S. ptarmicoides* and the cilia were a constant length from leaf tip to base as observed in *S. riddellii*. Therefore, morphology supports *S. ptarmicoides*, *S. riddellii*, and *S. ohioensis* as paternal sources of *S. houghtonii*.

**Single vs. Polytopic Origin**

Mayr et al. (1953) employed the term polytopic to describe a collective subspecies consisting of “several unrelated and widely separated populations” that have independently acquired the same diagnostic characters. The phylogenetic and network analyses of nrDNA provide support for *Solidago ohioensis*, *S. ptarmicoides*, and *S. riddellii* being involved in the parentage of *S. houghtonii*. Each of the four individuals, representing Pringle’s four hypothesized entities and spanning the geographic range of the species, share an identical or nearly identical homeologue identified in Figure 4 as subclade “C”. Network analysis indicated this homeologue is likely diverged from *S. ohioensis* and was either present at the time of hybridization or arose in an early ancestor of modern *S. houghtonii*. 

61
Solidago ohioensis is further suggested as a parental taxon for S. houghtonii individuals from each geographic region as the parsimony network analysis indicated that all individuals share a common S. ohioensis haplotype. It should be noted that MP analysis was unable to resolve these homeologue types thus placing them in a polytomy labeled as the “Ptarmicoidei Clade” in Figure 4. Statistical parsimony network analysis is designed to deal with the unique characteristics of intraspecific genetic variation and proved a useful tool for elucidating the involvement of S. ohioensis in the parentage of S. houghtonii individuals from each geographic region.

Solidago riddellii is also implicated in the parentage of S. houghtonii. Homeologue types from individuals representing entities 1 and 2 were identical to sequences of S. riddellii, forming a single haplotype in network analysis and a monophyletic clade in MP analysis. Solidago houghtonii individuals representing the two disjunct populations, entities 3 and 4, each possessed chimeric homeologues which showed evidence of S. riddellii which suggests that S. riddellii was involved in the parentage of these entities as well.

Solidago ptarmicoides was found to possess two nrDNA homeologues that differed in length and at 3 nucleotide positions. S. houghtonii individuals representing entities 1, 3, and 4 (Northern Michigan, Michigan disjunct, and New York disjunct respectively) each possessed a homeologue that was identical or nearly identical to the S. ptarmicoides-Type 1 homeologue. In addition, S. houghtonii representing entity 3 possessed a homeologue that was nearly identical to the S. ptarmicoides-Type 2 homeologue. The S. houghtonii individual from Ontario, Canada, entity 2, was the only entity in which a S. ptarmicoides homeologue type was not recovered. Since S. houghtonii is a sexually reproducing hexaploid species, the existence of three genomes is
expected. Essentially two genomes, *S. ohioensis* and *S. riddellii*, were recovered from this individual as the homeologue HOU008-#3 is likely a divergent *S. ohioensis* sequence. The third parental genome of individuals from entity 2 remains unclear; however, cloning additional individuals from this geographic region may yet reveal the presence of the *S. ptarmicoides* genome as it is believed to be the most likely contributor of the large ray flowers of *S. houghtonii*. It is possible that this entity is of different parentage as the third genome may have been contributed by a species other than *S. ptarmicoides*.

Several species of *Solidago* are known to possess multiple cytotypes (See Semple and Cook 2006 for ploidy numbers). *Solidago houghtonii* is known to be hexaploid throughout its range; however, the Michigan disjunct population, entity 3, is reported to be octoploid. Both MP analysis and network analysis indicate that the putative octoploid population of *S. houghtonii* is closely related to the hexaploid populations as they share nrDNA homeologue types related to *S. ohioensis*, *S. ptarmicoides*, and *S. riddellii*. If this population is octoploid it would be expected to possess four genomes and would therefore be of separate origin. Perhaps a backcross to one of the paternal species is the cause for the increased ploidy level of this population. If the population is confirmed to be octoploid, then further work is needed to reveal the origin of the forth genome. These results indicate that this population is at least derived from the hexaploid *S. houghtonii* with which it shares the same parental lineages.

While the shared nrDNA homeologue types are suggestive of a single origin for *Solidago houghtonii*, multiple origins involving the same three parental taxa remain a viable alternative. Additional studies to examine the variation in the parental taxa are necessary in order to address this question.
CHAPTER III

CHLOROPLAST EVIDENCE FOR THE HYBRID ORIGIN OF
SOLIDAGO HOUGHTONII

Introduction

For nearly two decades, the chloroplast genome (cpDNA) of plants has been extensively used for inferring plant phylogenies at different taxonomic levels (Shaw et al. 2005). The genome is well suited for use in phylogenetic studies as it is a relatively abundant component of total plant DNA and varies little in size, structure, and gene content among angiosperms (Clegg et al. 1994). Additionally, there is a wealth of comparative genetic data as representatives of all major lineages of land plants have completely sequenced chloroplast genomes. The rate of nucleotide substitution within coding regions of the chloroplast genome is relatively slow, making the genome useful for studies of plant phylogeny at deep taxonomic levels. Conversely, the noncoding regions of the chloroplast genome tend to evolve more rapidly, primarily by the accumulation of insertions and deletions (indels), so that noncoding cpDNA is useful for inferring phylogenies at lower taxonomic levels (Shaw et al. 2005). The utility of noncoding cpDNA for studying phylogenetic relationships between closely related species is evidenced by the large number of studies employing its use in recent years (Shaw et al. 2005). Another important feature of the chloroplast genome is its predominantly maternal mode of inheritance in angiosperms (Corriveau and Coleman 1988; Zhang et al. 2003). However, biparental chloroplast inheritance has been shown to
occasionally occur within some species (Mason et al. 1994; Ellis et al. 2008). Corriveau and Coleman (1988) found that chloroplast inheritance in *Solidago speciosa* is maternal which perhaps supports the assumption that cpDNA is maternally inherited within the genus *Solidago*. In addition to its use in phylogeny reconstruction, cpDNA has been used in studies of hybridization where incongruence among phylogenies produced from independent molecular markers with different modes of inheritance and sequence evolution provide a further line of evidence for the hybrid origin of polyploids. Several studies have used both the biparentally inherited nrDNA and uniparentally inherited cpDNA to infer the hybrid origins of several polyploid species including *Gossypium* (Wendel et al. 1995), *Paeonia* (Sang et al. 1997), *Oryza* (Ge et al. 1999), *Silene aegaea* (Popp and Oxelman 2001), *Dendrochilum acuiferum* (Barkman and Simpson 2002), *Cardamine* (Lihová et al. 2006), and *Potamogeton × torssanderi* (Kaplan and Fehrer 2007). In all of these cases, the hybrid species was found to have conflicting placement on the nuclear and chloroplast phylogenies.

The maternally inherited chloroplast genome is also useful for revealing single versus polytopic origins. Soltis and Soltis (1989) used cpDNA restriction site data to show two independent origins for the allopolyploid *Tragapogon miscellus* (*T. dubius × T. pratensis*) where different polyploid populations had incorporated the chloroplast genome of each of the two different parents. Chloroplast DNA has also been used to infer multiple origins for several other polyploid species including, *Senecio cambrensis* (Ashton and Abbott 1992), *Stebbinsoseris heterocarpa* and *S. decipiens* (Wallace and Jansen 1995), *Helianthus anomalus* (Schwarzbach and Rieseberg 2002), *Aegilops triuncialis* (Vanichanon et al. 2003) and *Platanthera huronensis* (Wallace 2003) where in each case the hybrid species was shown to have multiple chloroplast haplotypes inherited.
from different maternal sources. In contrast, Guggisberg et al. (2006) found a single chloroplast haplotype for the polyploid *Centaurium bianoris* which formed a clade with one of its hypothesized parental donors suggesting a single origin for the species. Likewise, Slotte et al. (2006) found that cpDNA sequences from 20 accessions of the tetraploid *Capsella bursa-pastoris* lacked intraspecific variation, suggesting a single and recent origin for this species.

This chapter describes the use of four cpDNA intergenic spacers, *accD-psal*, *psbA-trnH, trnL (UAA)-trnF* (GAA), and *rps16-trnQ*, to investigate the origin and maternal source of the putative hybrid *Solidago houghtonii*.

**Materials and Methods**

**DNA Samples, PCR Amplification, and Sequencing**

DNA samples used in the nuclear analysis of hybrid origin of *Solidago houghtonii*, Chapter 2, were also used in the chloroplast analysis. For each individual sampled, four noncoding cpDNA intergenic spacers (IGS), *accD-psal, psbA-trnH, trnL-trnF, rps16-trnQ* were amplified using polymerase chain reaction (PCR) (Mullis and Faloon 1987). PCR amplifications were performed in 50 µl volumes containing 35.4 µl sterile water, 5.0 µl of the provided 10X enzyme buffer, 2.5 µl magnesium chloride (50 mmol), 2.0 µl total dNTPs (10 mmol) (USB Corp.), 1.0 µl (10 µM) of each primer, 0.1µl of Platinum® *Taq* polymerase (Invitrogen Corp.), and 1µl of template DNA diluted 1:10. The reaction mixtures were cycled on an Eppendorf Mastercycler Programmable Thermal Cycler (Eppendorf Scientific Inc.). The PCR primers ACCD-769F and PSAI-75R (Barkman and Simpson 2002) were used to amplify the *accD-psal* IGS region. The
**accD-psal** PCR protocol consisted of an initial 3 min denaturation at 94°C, followed by 34 cycles of 1 min denaturation (94°C), 1 min annealing (55°C), and 1 min 20 sec elongation (72°C), followed by a 7 min final extension at 72°C. The *psbA-trnH* intergenic spacer was amplified using primers TRNH(GUG) and PSBA (Saltonstall 2001) and the following PCR protocol: an initial 3 min denaturation at 94°C, followed by 35 cycles of 1 min denaturation (94°C), 1 min annealing (50°C), and 45 sec elongation (72°C), followed by a 3 min final extension at 72°C. Primers E and F (Taberlet et al. 1991) and the *accD-psal* PCR protocol described above were used to amplify the *trnL-trnF* intergenic spacer. Lastly, the *rps16-trnQ* intergenic spacer was amplified with primers RPS16F and TRNQ (Saltonstall 2001) and the above *accD-psal* PCR protocol with the annealing temperature modified to 50°C. PCR products were purified and sequenced as described in Chapter 2. The *trnL-trnF*, *psbA-trnH*, and *rps16-trnQ* spacers were sequenced using their respective PCR primers. The *accD-psal* spacer was sequenced using *Solidago* specific primers (SOL-446F 5'-'AAATCGGAAATCTGACCC and newPSAI-R 5'-TCATAGAATGGGTACCTCGA) that were designed for this study.

**Phylogenetic Analyses**

For each marker, contigs were assembled and sequences aligned as described for the nuclear markers (Chapter 2). Complete sequences have been deposited in GenBank (Appendix). The cpDNA IGS data sets (*accD-psal*, *psbA-trnH*, *trnL-trnF*, *rps16-trnQ*) were combined and analyzed together since they are inherited as a single unit.

As in the nrDNA data set (Chapter 2), indels in the aligned cpDNA data sets were coded for presence/absence by assigning the character state 1 if the indel was present and
the character state 0 if the indel was absent (Simmons and Ochoterena 2000). 1.36% of the cpDNA data matrix cells were scored as missing data.

All phylogenetic analyses were performed using PAUP* 4.0b4a (Swofford 1998). Maximum parsimony (MP) analyses using the heuristic search algorithm were conducted on the combined cpDNA data set as described for the nrDNA data set (Chapter 2). Due to computational constraints, the cpDNA data set was analyzed using 10 random addition sequences. Tree statistics and measures of homoplasy were calculated using PAUP* and, the multiple most parsimonious trees were combined to construct a strict consensus tree. The robustness of clades was assessed using Jackknife analyses (Farris et al. 1996) as described for the nrDNA data set (Chapter 2).

Results

Chloroplast DNA Sequence Variation

Overall, there was a low level of nucleotide variability for each of the four cpDNA IGS sequences. Of the 3349 characters, including gaps, in the cpDNA data matrix, 121 (3.61%) are variable in at least one of the sampled individuals, and of these, 44 (1.31%) are parsimony-informative. In contrast to the relatively low rate of nucleotide substitution, there was high variability in sequence length across the taxa. The minimum / maximum sequence lengths obtained for each of the four cpDNA regions were 763 / 939 bp (accD-psal IGS), 212 / 290 (psbA-trnH IGS), 407 / 514 bp (trnL-trnF IGS), and 869 / 999 bp (rps16-trnQ IGS). Length variation is attributed primarily to insertions or deletions of short (1 – 30 bp) repeats. Sequence alignment can be difficult when there is a large number of indels flanked by regions with numerous nucleotide substitutions;
however, because my sequences had very few nucleotide substitutions, alignment was relatively straightforward. This allowed us to confidently code the indels for inclusion in the data matrix (Fig. 6). A total of 203 length mutations in the cpDNA alignment were coded for presence/absence (1/0). Of these, 95 (47%) occurred between the outgroup and ingroup taxa. Following the addition of the coded indels, the combined cpDNA data matrix contained 3552 characters, 263 (7.4%) are variable, and of these 87 (2.45%) are parsimony-informative.

The accD-psal sequence alignment included a total of 132 indels, the vast majority of which are unique to a single species or shared by species thought to be in the same taxonomic section and/or subsection. Sequence alignment of this data set revealed a unique 29 bp insertion that is shared by all sampled S. houghtonii individuals regardless of their geographic location or ploidy level. Surprisingly, this insertion was also shared, not with one of the proposed parents as we had expected, but rather with Solidago gigantea and no other sampled species. All sampled S. gigantea individuals from diverse localities in Michigan, USA and Ontario, Canada share this insertion, but the individual (accession GIG005) from New York, USA does not (Fig. 6a).

The trnL-trnF data set contained 16 indels. Sequence length variation is primarily attributed to two factors: 1) variation in the number of repeats in an “AT” microsatellite, and 2) an 11 bp deletion. These length variations are shared with all sampled individuals of S. gigantea, except the individual from New York (accession GIG005) and also with S. canadensis, S. nemoralis, S. nitida, S. riddellii, and S. velutina (Fig. 6b). This data set also includes a 95 bp insertion which was shared by S. rigida and S. simplex. In the same region, S. uliginosa has a 101 bp insertion which is identical to that of S. rigida and S. simplex but with 6 additional nucleotides. These insertions were not present in other
**Figure 6. Informative Chloroplast DNA Indels**

<table>
<thead>
<tr>
<th>A.</th>
<th>B.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. multiradiata</td>
<td>S. multiradiata</td>
</tr>
<tr>
<td>S. simplex</td>
<td>S. simplex</td>
</tr>
<tr>
<td>S. petiolaris</td>
<td>S. petiolaris</td>
</tr>
<tr>
<td>S. bicolor</td>
<td>S. bicolor</td>
</tr>
<tr>
<td>S. hispida</td>
<td>S. hispida</td>
</tr>
<tr>
<td>S. speciosa</td>
<td>S. speciosa</td>
</tr>
<tr>
<td>S. caespitosa</td>
<td>S. caespitosa</td>
</tr>
<tr>
<td>S. flexicaulis</td>
<td>S. flexicaulis</td>
</tr>
<tr>
<td>S. patula</td>
<td>S. patula</td>
</tr>
<tr>
<td>S. sempervirens</td>
<td>S. sempervirens</td>
</tr>
<tr>
<td>S. uliginosa</td>
<td>S. uliginosa</td>
</tr>
<tr>
<td>S. juncea</td>
<td>S. juncea</td>
</tr>
<tr>
<td>S. missouriensis</td>
<td>S. missouriensis</td>
</tr>
<tr>
<td>S. rugosa</td>
<td>S. rugosa</td>
</tr>
<tr>
<td>S. ulmifolia</td>
<td>S. ulmifolia</td>
</tr>
<tr>
<td>S. altissima</td>
<td>S. altissima</td>
</tr>
<tr>
<td>S. canadensis</td>
<td>S. canadensis</td>
</tr>
<tr>
<td>S. gigantea-GIG02</td>
<td>S. gigantea-GIG02</td>
</tr>
<tr>
<td>S. gigantea-GIG03</td>
<td>S. gigantea-GIG03</td>
</tr>
<tr>
<td>S. gigantea-GIG04</td>
<td>S. gigantea-GIG04</td>
</tr>
<tr>
<td>S. gigantea-GIG05</td>
<td>S. gigantea-GIG05</td>
</tr>
<tr>
<td>S. gigantea-GIG07</td>
<td>S. gigantea-GIG07</td>
</tr>
<tr>
<td>S. gigantea-GIG08</td>
<td>S. gigantea-GIG08</td>
</tr>
<tr>
<td>S. mosquitalis-HOU002</td>
<td>S. mosquitalis-HOU002</td>
</tr>
<tr>
<td>S. houghtonii-HOU008</td>
<td>S. houghtonii-HOU008</td>
</tr>
<tr>
<td>S. houghtonii-HOU015</td>
<td>S. houghtonii-HOU015</td>
</tr>
<tr>
<td>S. houghtonii-HOU060</td>
<td>S. houghtonii-HOU060</td>
</tr>
<tr>
<td>S. ohoeosa</td>
<td>S. ohoeosa</td>
</tr>
<tr>
<td>S. pharnicoides</td>
<td>S. pharnicoides</td>
</tr>
<tr>
<td>S. petiolaris</td>
<td>S. petiolaris</td>
</tr>
<tr>
<td>Heterotheca villosa</td>
<td>Heterotheca villosa</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Heterotheca villosa</td>
</tr>
<tr>
<td>Euthamia graminifolia</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Euthamia graminifolia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.</th>
<th>D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. multiradiata</td>
<td>S. multiradiata</td>
</tr>
<tr>
<td>S. simplex</td>
<td>S. simplex</td>
</tr>
<tr>
<td>S. petiolaris</td>
<td>S. petiolaris</td>
</tr>
<tr>
<td>S. bicolor</td>
<td>S. bicolor</td>
</tr>
<tr>
<td>S. hispida</td>
<td>S. hispida</td>
</tr>
<tr>
<td>S. speciosa</td>
<td>S. speciosa</td>
</tr>
<tr>
<td>S. caespitosa</td>
<td>S. caespitosa</td>
</tr>
<tr>
<td>S. flexicaulis</td>
<td>S. flexicaulis</td>
</tr>
<tr>
<td>S. patula</td>
<td>S. patula</td>
</tr>
<tr>
<td>S. sempervirens</td>
<td>S. sempervirens</td>
</tr>
<tr>
<td>S. uliginosa</td>
<td>S. uliginosa</td>
</tr>
<tr>
<td>S. juncea</td>
<td>S. juncea</td>
</tr>
<tr>
<td>S. missouriensis</td>
<td>S. missouriensis</td>
</tr>
<tr>
<td>S. rugosa</td>
<td>S. rugosa</td>
</tr>
<tr>
<td>S. ulmifolia</td>
<td>S. ulmifolia</td>
</tr>
<tr>
<td>S. altissima</td>
<td>S. altissima</td>
</tr>
<tr>
<td>S. canadensis</td>
<td>S. canadensis</td>
</tr>
<tr>
<td>S. gigantea-GIG02</td>
<td>S. gigantea-GIG02</td>
</tr>
<tr>
<td>S. gigantea-GIG03</td>
<td>S. gigantea-GIG03</td>
</tr>
<tr>
<td>S. gigantea-GIG04</td>
<td>S. gigantea-GIG04</td>
</tr>
<tr>
<td>S. gigantea-GIG05</td>
<td>S. gigantea-GIG05</td>
</tr>
<tr>
<td>S. mosquitalis-HOU002</td>
<td>S. mosquitalis-HOU002</td>
</tr>
<tr>
<td>S. houghtonii-HOU008</td>
<td>S. houghtonii-HOU008</td>
</tr>
<tr>
<td>S. houghtonii-HOU015</td>
<td>S. houghtonii-HOU015</td>
</tr>
<tr>
<td>S. houghtonii-HOU060</td>
<td>S. houghtonii-HOU060</td>
</tr>
<tr>
<td>S. ohoeosa</td>
<td>S. ohoeosa</td>
</tr>
<tr>
<td>S. pharnicoides</td>
<td>S. pharnicoides</td>
</tr>
<tr>
<td>Heterotheca villosa</td>
<td>Heterotheca villosa</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Heterotheca villosa</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Euthamia graminifolia</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Eriogonum nauseosus</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td>Euthamia graminifolia</td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Eriogonum nauseosus</td>
</tr>
<tr>
<td></td>
<td>Euthamia graminifolia</td>
</tr>
</tbody>
</table>

Note: A. A 29 bp insertion from the accD-psaI IGS. B. Microsatellite length variation and a 5 bp deletion from the trnl-trnF IGS. C. A 15 bp deletion from the psbA-trnH IGS. D. Microsatellite length variation from the rps16-trnQ IGS.
samples of these species (data not shown). For ease of data presentation these indels have been removed from Fig. 6b.

In the psbA-trnH sequence alignment we found 29 indels, including a 15 bp deletion that was present in all sampled S. houghtonii individuals and in all sampled individuals of S. gigantea, except for the individual from New York (GIG005) (Fig. 6c). This 15 bp indel was not found in any other sampled species.

Lastly, the rps16-trnQ sequence alignment included a total of 26 indels, of which there was a single informative indel at nucleotide position 632 (Fig. 6d). This indel is the result of variation in the number of repeats in an “AT” microsatellite. Here S. houghtonii, S. gigantea (except for the New York sample), S. juncea, S. nemoralis, S. ohioensis, S. patula, S. ptarmicoides, and S. speciosa have the same number of “AT” repeats.

Analysis of Pairwise Differences

The analysis of pairwise differences indicates that the mean cpDNA divergence between all sampled individuals of S. houghtonii and individuals of S. gigantea from Michigan, USA and Ontario, Canada is quite low (0.044%) and that there are zero character differences between two S. gigantea individuals (GIG003 northernmost Michigan and GIG004 Ontario, Canada) and three individuals of S. houghtonii (HOU002 from the northern tip of Michigan’s lower peninsula, HOU008 from the Bruce Peninsula, Ontario, Canada, and HOU015 from New York). Solidago gigantea (GIG005) from New York differs considerably from other samples of S. gigantea in its combined cpDNA sequence data, having 6 unique indels and 2 nucleotide substitutions. However, the nrDNA 3’ETS and ITS sequences from the New York sample of S. gigantea (GIG005) does not differ from other S. gigantea individuals (see nuclear data – Chapter 2).
Chloroplast DNA Phylogeny

The MP analysis of the four combined cpDNA IGS data sets (Fig. 7) resulted in 143,316 equally parsimonious trees (400 steps, CI = 0.885, and RI = .720). Although the overall species relationships are not well resolved, the analysis strongly supports (Jackknife = 100) the monophyly of the genus *Solidago*. *Solidago bicolor* and *S. speciosa*, both members of section *Solidago* subsection *Squarrosae* form a well supported clade (Jackknife = 87). *Solidago missouriensis* (subsection *Junceae*) and *S. multiradiata* (subsection *Multiradiatae*) form a moderately supported clade (Jackknife = 70) and *S. ohioensis* and *S. ptarmicoides*, both members of section *Ptarmicoidei*, form a strongly supported clade (Jackknife = 98). The analysis also strongly supports (Jackknife = 90) a clade comprised of *S. houghtonii* (section *Ptarmicoidei*) individuals from each geographic region and three *S. gigantea* (subsection *Triplinerviae*) individuals. All other relationships are weakly supported (Jackknife < 70).

Discussion

Incongruent Phylogenies as Evidence of Hybridization

Hybridization resulting in allopolyploids has been shown to be a widespread and frequent mode of plant evolution and speciation (Soltis and Soltis 1993; Wendel and Doyle 1998; Soltis and Soltis 2000). These events can often be detected from the incongruence of species relationships reconstructed from molecular markers with differing modes of inheritance. The expectation, when using cpDNA, is that sequences generated from the maternally inherited chloroplast genome of the allopolyploid will form a clade with those generated from the maternal parent. In contrast, the expectation
Figure 7. Chloroplast DNA Strict Consensus Tree

Note: Numbers above the branches represent jackknife support.
is that sequences generated from the biparentally inherited nuclear genome of the hybrid will be additive with those of the diploid progenitors and that in phylogenetic analysis they will be sister to the orthologues from their progenitor taxa (Wendel 2000). Therefore, nuclear and chloroplast DNA sequences frequently produce different phylogenetic patterns which are useful for inferring interspecific hybridization events. However, recent studies have shown that genome evolution in polyploids can be much more complex than suggested. For example, polytopic origins (Soltis and Soltis 1993, 1999), nonadditivity due to non-Mendelian changes following polyploidization (reviewed in Soltis and Soltis 1999; Wendel 2000), gene amplification above the number expected from direct genome duplication during polyploidization (Rabinowicz et al. 1999), and unexpected introgression from divergent sources (Cronn et al. 2003; Mason-Gamer 2004) can complicate the inference of parentage.

Visual examination of the MP trees depicting species relationships reconstructed from the combined cpDNA sequence data (Fig. 7) and the combined nrDNA sequence data (Chapter 2 – Fig. 4) shows evidence of an incongruent topological placement for *S. houghtonii* as is expected for a species of hybrid origin. Phylogenetic analysis of cpDNA strongly supports (Jackknife = 90) the placement of all *S. houghtonii* individuals in a clade with *S. gigantea*. In contrast, phylogenetic analysis of the nrDNA supports the nesting of *S. houghtonii* homeologous 3'ETS and ITS loci with various members of section *Ptarmicoidei* and supports the placement of *S. gigantea* with other members of section *Solidago* subsection *Triplinerviae* (Jackknife = 92%). This is consistent with the taxonomical placement of *S. houghtonii* with other members of section *Ptarmicoidei* and also with morphology.
Chloroplast Evidence of Parentage

In the Asteraceae, the chloroplast genome is usually maternally inherited (Corriveau and Coleman 1988; Zhang et al. 2003) and therefore serves as a suitable marker for investigating the maternal source of the allopolyploid Solidago houghtonii. The alignment of combined cpDNA sequence data from S. houghtonii and 25 other Solidago species revealed a unique 29 bp insertion in the psaI-accD IGS and a unique 15 bp deletion in the psbA-trnH IGS. Solidago houghtonii shared these unique indels with S. gigantea and with no other Solidago species. This suggests that S. gigantea is the source of S. houghtonii cpDNA. The trnL-trnF and rps16-trnQ markers revealed a deletion and microsatellite length variation which was shared by S. houghtonii and S. gigantea, as well as other Solidago species. While the relationship between S. houghtonii and S. gigantea cannot be inferred from the trnL-trnF and rps16-trnQ markers alone, the data from these markers do not refute the data obtained from the psaI-accD and psbA-trnH markers. Furthermore, there is zero chloroplast sequence divergence between S. gigantea (accessions GIG003 and GIG004) and S. houghtonii (accessions HOU002, HOU008, and HOU015). All sampled S. houghtonii and S. gigantea individuals form a strongly supported clade in MP analysis (Fig. 7). These data can be explained by two alternative scenarios: 1) maternal parentage of S. houghtonii by S. gigantea or, 2) capture of S. gigantea chloroplast DNA by one of the parental taxa of S. houghtonii.

The presence of three separate genomic contributions is expected for organisms of allohexaploid origin. The nrDNA data (Chapter 2) supports the presence of at least three parental genomes of S. houghtonii involving S. ptarmicoides, S. riddelli, and S. ohioensis, but does not support S. gigantea. This conflicts with the expectation that the maternal source of the allopolyploid would be represented in the nuclear genome. The
existence *S. gigantea* sequences within the nuclear genome of *S. houghtonii* appears improbable because *S. gigantea*-specific ITS primers were unable to amplify *S. gigantea* sequences from *S. houghtonii*. In addition to sequencing a large number of clones, 22 – 30 per individual for the entire 3'ETS – ITS region of nrDNA, we initially cloned only the ITS region and sequenced 10 – 15 clones per individual. We failed to isolate any clones resembling *S. gigantea* from any individual of *S. houghtonii*. The absence of any evidence of *S. gigantea*-like nrDNA within the genome of *S. houghtonii* is consistent with an absence of shared morphological characters between these species.

The alternative scenario, that *Solidago gigantea* cpDNA in *S. houghtonii* arose as a result of chloroplast capture by a parent of *S. houghtonii*, may be more probable. Rieseberg and Soltis (1991) examined the extent to which chloroplast capture through hybridization and introgression occurs in plants and found the literature to contain a high number of conclusive examples that, like the presence of *S. gigantea* cpDNA in *S. houghtonii*, were largely unexpected by the investigators. Furthermore, they indicated that cytoplasmic gene flow occurred more often than nuclear gene flow and that cytoplasmic gene flow frequently occurs without nuclear DNA introgression. Soltis et al. (1991) analyzed cpDNA variation across five sections and the majority of subsections of the genus *Heuchera* (Saxifragaceae) and found that one section, *Rhodoheuchera*, was highly divergent from the remaining four sections. Their analysis indicated that three species, *H. hallii*, *H. nivalis*, and *H. parvifolia* resolved with members of section *Rhodoheuchera* although morphological and allozyme data had indicated that their closest relatives were not from within this section. These authors proposed that the distinctive chloroplast genotype of section *Rhodoheuchera* was captured by populations of *H. hallii*, *H. nivalis*, and *H. parvifolia* via hybridization and introgression, indicating
that, in *Heuchera*, intersectional hybridization is common. My results provide another example of intersectional hybridization as *S. gigantea* belongs to section *Solidago* subsection *Triplinerviae* and the nuclear genome donors of *S. houghtonii* belong to section *Ptarmicoidei*.

Since *S. gigantea*-like sequences were not recovered from the nuclear genome of *S. houghtonii*, a scenario involving introgressive hybridization leading to chloroplast capture is implicated. This would involve *S. gigantea* cpDNA introgression into one of the putative, section *Ptarmicoidei*, parents of *S. houghtonii*. Or perhaps an extinct intermediate tetraploid ancestor of *S. houghtonii* captured the *S. gigantea* cpDNA. Regardless of which ancestor of *S. houghtonii* captured the cpDNA of *S. gigantea*, that parent would have had to serve as the maternal parent for any hybridization event leading to the formation of both the hexaploid and octoploid cytotypes of *S. houghtonii*. My data do not indicate which maternal species captured *S. gigantea* chloroplasts. Likewise, it can not be determined in which of the sequential hybridization events the introgression took place or, alternatively, if the introgression was into an already formed triple hybrid. In some species, sequence elimination seems to be related to the ploidy level of the parents where sequence elimination is more common from the lower ploidy level parent (Wendel 2000). If *Solidago* follows the common pattern and assuming that *S. gigantea* possessed the Midwestern cytotype of \(2n = 36\) (Semple et al. 1984; Semple et al. 1999), then *S. gigantea* would only be the lower ploidy level parent if the introgression was into an already formed triple hybrid. It is not clear how *S. gigantea* cpDNA passed into *S. houghtonii*, but its presence in all examined populations suggests that it was introgressed/captured early in the history of *S. houghtonii*. 
Alternative Explanations for Molecular Results

Although it appears that *S. gigantea* is the maternal genome donor of *S. houghtonii*, finding that the cpDNA of *S. gigantea* from New York did not match the cpDNA from *S. gigantea* individuals from the Great Lakes region (Fig. 6a – d) is confounding. Sequence divergence between *S. houghtonii* and the chloroplast haplotype of *S. gigantea* from New York was almost an order of magnitude higher than when *S. houghtonii* was compared to individuals of *S. gigantea* from the Great Lakes region. It is possible that the specimen was incorrectly identified but, its placement in the nrDNA phylogeny (Chapter 2 – Fig. 4) supports the correct identification. Another explanation may be that *S. gigantea* individuals from the Great Lakes region were introgressed from *S. houghtonii*. To resolve this question the *psal-accD* IGS for individuals of *S. gigantea* from Illinois (GIG007), Colorado (GIG008), and Texas (GIG009) (Fig. 6a) were sequenced. These additional sampled individuals also contained the unique 29 bp insertion which supports the rejection that local (i.e. Great Lakes region only) introgression from *S. houghtonii* into *S. gigantea* occurred. Alternatively, the *S. gigantea* individual from New York may be an introgressed individual from yet another species. A final possibility involves differences associated with cytotype. Eastern populations of *S. gigantea* are diploid (2n = 18) while Midwest populations are tetraploid (2n = 36) (Semple et al. 1984; Semple et al. 1999). A more extensive sampling of eastern populations of *S. gigantea* would be necessary in order to explain the source of chloroplast variation within *S. gigantea*.

Topological discrepancies between trees generated from chloroplast DNA and those generated from nuclear DNA can potentially complicate phylogenetic interpretations as they may result not from hybridization but rather from lineage sorting of
ancestral polymorphic chloroplast haplotypes (Wendel and Doyle 1998; Sang and Zhong 2000). If we hypothesize lineage sorting to account for _S. houghtonii_ and _S. gigantea_ sharing the same chloroplast haplotype then the chloroplast polymorphism would need to be maintained through the various _Solidago_ speciation events separating _S. houghtonii_ from _S. gigantea_. This would then have to be followed by extinction of one allele in all species except _S. houghtonii_ and _S. gigantea_ and the subsequent extinction of the other allele in these two species. Because of the complicated nature of the lineage sorting scenario, the scenario of introgressive hybridization leading to chloroplast capture is accepted as the more parsimonious explanation for the observed topologies.

**Single vs. Polytopic Origin**

As reviewed in Chapter 1, there have been several hypotheses regarding the origin of _Solidago houghtonii_ involving both single and polytopic origins. The analysis of cpDNA haplotype diversity in _S. houghtonii_ resulted in _S. houghtonii_ cpDNA haplotypes from across the range of the species forming a well supported monophyletic clade, suggesting a single origin. Taken together, the nuclear and chloroplast DNA phylogenetic and network analyses implicate a single origin for _Solidago houghtonii_ through chloroplast capture and reticulate hybridization. Alternatively, but perhaps unlikely, a polytopic origin for _S. houghtonii_ involving the independent formation of the same hexaploid taxon in four widely separated geographic areas and each undergoing introgressive hybridization with _S. gigantea_ can be postulated. Or, perhaps a polytopic origin involving a widespread allotetraploid with _S. gigantea_ cpDNA that obtained additional paternal genomes through separate hybridization events can be postulated. In either case, the most parsimonious explanation for these results involves a single origin.
CHAPTER IV

POPULATION GENETICS OF SOLIDAGO HOUGHTONII

Introduction

Rare species are of particular interest to evolutionary biologists seeking to understand how observed patterns of genetic variation have been influenced by historical evolutionary processes. This knowledge allows for predictions of how populations may respond to future events of both natural and anthropogenic origin (Huenneke 1991; Wallace 2002). Therefore, an understanding of both the patterns and causes of genetic diversity are of great importance when developing recovery and management strategies for rare, threatened, or endangered species (Templeton 1991).

Rarity in plants is generally defined in terms of geographic distribution, abundance, and habitat specificity (Gaston 1997). Generally, rare species will have a narrow geographic range, high habitat specificity, and small population size (Rabinowitz 1981). Species with restricted ranges and high habitat specificity, such as Solidago houghtonii, the subject of this study, are often identified as endemic species and many of these are threatened or endangered (Kothera et al. 2007).

The genetic consequences of rarity generally follow the predictions made for small populations. Populations of rare plant species are expected to have reduced genetic variation within populations and increased differentiation among populations due to the affects of genetic drift, founder events, and genetic bottlenecks (Ellstrand and Elam 1993; Hannan and Orick 2000). While limited genetic variation has been reported for many
rare plant species (reviewed in Hamrick and Godt 1989), this view may be an overgeneralization as there are now several reports of highly variable rare species (reviewed in Gitzendanner and Soltis 2000). In any case, genetic variation is believed to be an important predictor of the long-term success of a species or population as it should determine a population’s ability to adapt to changing environments (reviewed in Booy et al. 2000).

The loss of genetic variation can potentially be overcome by gene flow or the movement of alleles between populations. Wright (1931) suggested that the adverse effects of genetic drift can be reduced with the immigration of one or more individuals per generation into a population. To a large extent, the amount of genetic drift and gene flow a population experiences depends on its size and its degree of isolation. Wright (1943) proposed in his Isolation by Distance model, that gene flow will decrease with increased geographic distance so that geographically close populations are expected to be genetically more similar than more distant populations.

Most natural populations of species show some level of genetic structure (Balloux and Lugon-Moulin 2002; Haig 1998). Fluctuations in population size caused by historical processes such as founder events, bottlenecks, and climatic changes can affect the genetic structure of populations through reduced gene flow and genetic drift. In addition, contemporary events such as environmental uncertainty due to invasive species, abiotic factors (i.e. fire, drought, flood, or cyclical fluctuations in lake levels), or anthropogenic habitat fragmentation can affect the genetic structure of populations (Hamrick and Godt 1989; Bonnin et al. 2002).

Most plant species live in naturally fragmented habitats and therefore exist in a number of discrete populations (subpopulations) (Bonnin et al. 2002). As ecological
conditions change in response to stochastic environmental conditions, a given habitat site may become unsuitable while at the same time new sites become available. Therefore, many species are characterized by a balance between local population extinctions and the establishment of new populations which is a concept explained by metapopulation theory (Opdam 1991; Baguette 2004; Hanski 2004).

Migration to new suitable sites is the first step in establishing a new population. As such, migration may act as a buffer against the consequences of environmental stochasticities and thereby increase the probability that subpopulations persist. However, migration between subpopulations can also prevent them from adapting to local environmental conditions (Bonnin et al. 2002). Within the metapopulation structure, a balance between the opposing forces of migration and selection, as well as, extinction and colonization act to maintain species viability (Bonnin et al. 2002). Human activities often disrupt this balance by increasing habitat fragmentation (USFWS 1997; Bonnin et al. 2002). As the landscape becomes more fragmented, the amount of suitable habitat is reduced which leads to smaller population sizes, loss of genetic variation, and increased rates of extinction. Consequently, the amount of suitable habitat is reduced and the distances between habitat sites increase. This leads to a decrease in the rate of colonization as the abundance and effectiveness of pollinators and seed dispersers is altered (England et al. 2002). In highly fragmented populations, the geographic distance between subpopulations often exceeds an individual’s dispersal capacity and subpopulations become genetically differentiated through isolation by distance (Bonnin et al. 2002).

Solidago houghtonii is a rare plant species endemic to the northern shores of Lakes Michigan and Huron. As with other endemic plant species, it has a very limited
global distribution and high habitat specificity. *Solidago houghtonii* is a specialist on alkaline substrates, especially those associated with the Niagara escarpment. Therefore, appropriate habitat for this species is naturally fragmented and occurs over a narrow geographic range (see Chapter 1 for a complete description of the species and its habitat).

According to the standardized methodology of NatureServe (2009), *S. houghtonii* has a global status of G3 thus indicating that the species is vulnerable across its entire range. In Canada the species has a national rank of N2 (imperiled) and a subnational rank of S2 (imperiled). In the United States, the species is designated nationally as vulnerable (N3), vulnerable (S3) in the State of Michigan and critically imperiled (S1) in the state of New York. In the United States the species has been awarded protected status through the Endangered Species Act of 1973 (USFWS 1988b) and approximately half of the known populations occur on protected lands (USFWS 1997). In contrast, the species does not have protected status in Ontario, Canada and only two populations occur on protected habitat. In areas where the species is not protected, it faces threats of increased habitat fragmentation due to heavy recreational use, residential development, ATV traffic, limestone quarrying and other activities. *Solidago houghtonii* populations are also vulnerable to natural disturbances such as drought, flooding, fluctuating lake levels which alter the water table, and the invasion of alien species (USFWS 1997; COSEWIC 2005).

In recent decades, chloroplast DNA (cpDNA) has been employed as a genetic marker for studies focusing on estimates of genetic diversity, population structure and phylogeographic distribution (McCauley 1995, Hamilton et al. 2003). Since chloroplast DNA is maternally inherited in angiosperms, it is dispersed in seeds but not through pollen. The importance of seed dispersal in the establishment of new plant populations is well established (Loveless and Hamrick 1984; Cain et al. 1998; Willson and Traveset
Therefore, an understanding of the distance by which genes can move between populations via seed is critical to understanding the genetic structure of populations. Seed dispersal is the one stage in the life-cycle of most plants that allows populations to expand (Willson and Traveset 2000). Within the metapopulation framework, long-distance dispersal events are critically important for the colonization of new sites (Cain et al. 2000).

Several features of the chloroplast genome make this marker particularly well suited to for estimating population structure and phylogeographic distribution. First, the lack of cpDNA recombination and its maternal inheritance allows the spatial pattern of cpDNA haplotypes to be interpreted as an estimate of past seed dispersal (McCauley 1995). Second, because the chloroplast genome is haploid, it has a smaller effective population size than the diploid nuclear genome. This acts to accelerate the process of genetic drift so that the levels of differentiation between populations will be displayed sooner for cpDNA than for nuclear markers (Hamilton et al. 2003; Butaud et al. 2005). Thirdly, cpDNA intergenic regions have been shown to harbor substantial intraspecific variation due to the presence of insertion/deletion (indel) polymorphisms within and among plant populations that are useful for estimating population genetic structure and gene flow (e.g., McCauley 1994; Hamilton et al. 2003; McCauley et al. 2003).

The use of cpDNA indel polymorphisms in studies of population genetic structure is not without its challenges. Indels do not occur at random throughout the chloroplast genome but rather occur at “hot-spots” associated with specific features of the DNA sequences such as microsatellite repeats and stem-loop secondary structures (reviewed in Kelchner 2000). This leads to the possibility that haplotypes defined by indels will have a greater possibility of being homoplastic (identical by state rather than being identical by
Indel homoplasies can reduce the inferred population subdivision as haplotypes are shared between populations due to both gene flow and recurrent mutation (Hamilton et al. 2003). In addition, indel homoplasies may make it difficult to distinguish between the homogenization of populations due to the evolutionary processes of gene flow and selection, and the convergence of populations to a common haplotype (Hamilton et al. 2003). The use of multiple cpDNA markers can overcome these difficulties as it is unlikely that multiple markers would have identical homoplasies.

To date, there have been thousands of studies performed for both plant and animal species that employ the use of molecular markers to delineate population genetic structure. Information from these types of studies is central to developing effective conservation plans for rare, threatened and endangered species. There is currently no information on population structure and genetic diversity within and among populations of *Solidago houghtonii*. This study will use cpDNA hyplotypes to investigate the population genetic structure of *S. houghtonii* with the following objectives: 1) to assess how cpDNA diversity is structured both within and among populations; 2) to estimate the relationship between geographical distance and cpDNA genetic distance; 3) to determine if habitat fragmentation has had significant effects on the distribution of cpDNA genetic diversity within and among populations; 4) to determine if there is a correlation between cpDNA genetic structure and geographical region; 5) to determine if there is a correlation between cpDNA genetic structure and substrate type; and 6) to investigate the cpDNA phylogeography of *S. houghtonii*. 

85
Material and Methods

Study Species

Solidago houghtonii is a perennial herb endemic to the shores of Lakes Michigan and Huron in the State of Michigan, USA, and Ontario Canada. There are two disjunct populations; one is in Genesee Co., New York and the other is at an inland location in Michigan’s Lower Peninsula (Fig. 2; Chapter 1). Individual plants reproduce vegetatively by short rhizomes. Mature plants are gynomonoecious, insect pollinated, and flower in alternate years (reviewed in Chapter 1). Phylogenetic analysis based on chloroplast and nuclear genes (Chapters 2 and 3) suggest a complex origin involving reticulate evolution and introgression for S. houghtonii. Solidago houghtonii is considered to be a hexaploid species (2n = 54); however, the Michigan disjunct population is reported to be octoploid (2n = 72) (Jim Pringle - unpublished).

Population Sampling

In the summer of 2002, 23 populations of Solidago houghtonii from throughout its geographic range were sampled, including two disjunct populations (Table 2). Sampling was designed to include populations from all substrate types; alvar, calcareous beach flats, cobble beaches, interdunal wetlands, marl fen, and wet prairie fen.

Local populations of Solidago houghtonii are defined as distinct groups of plants separated from other groups by at least 1km (USFWS 1997). The 23 populations sampled in this study were separated by an average of 185 km; the within region distances were less. Within Canada populations were separated by distances of 1 – 98 km, within Michigan’s Upper Peninsula by 7 – 108 km and within Michigan’s Lower Peninsula by
Table 2. Collection Details for 23 *Solidago houghtonii* Populations

<table>
<thead>
<tr>
<th>Region/Population</th>
<th>Population Code</th>
<th>Location</th>
<th>Haplotypes</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York – Disjunct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genesee Co.</td>
<td>BRG</td>
<td>33°06'08&quot; N 78°00'58&quot; W</td>
<td>3, 9</td>
<td>Marl Fen</td>
</tr>
<tr>
<td>Ontario, Canada</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bruce Peninsula</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cabot’s Head</td>
<td>CAB</td>
<td>17 E4762 N 50098</td>
<td>7</td>
<td>Alvar</td>
</tr>
<tr>
<td>Little LaCloche Island</td>
<td>LLC</td>
<td>17 E4428 N 50917</td>
<td>11</td>
<td>Alvar</td>
</tr>
<tr>
<td>Manitoulin Island</td>
<td>SBY</td>
<td>17 E4339 N 50582</td>
<td>4, 7</td>
<td>Beach Flat</td>
</tr>
<tr>
<td>Strawberry Island</td>
<td>STR</td>
<td>17 E4339 N 50905</td>
<td>10, 11</td>
<td>Alvar</td>
</tr>
<tr>
<td>LaCloche Peninsula</td>
<td>WFS</td>
<td>17 E4411 N 50965</td>
<td>7, 10, 11</td>
<td>Alvar</td>
</tr>
<tr>
<td>Great LaCloche Island</td>
<td>WGL</td>
<td>17 E4388 N 50941</td>
<td>10, 11</td>
<td>Alvar</td>
</tr>
<tr>
<td>Manitoulin Island</td>
<td>WHT</td>
<td>17 E4314 N 50901</td>
<td>11</td>
<td>Beach Flat</td>
</tr>
<tr>
<td>Michigan, USA – Upper Peninsula</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chippewa Co.</td>
<td>ALB</td>
<td>T41N R03E</td>
<td>8</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Mackinac Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albany Creek Mouth</td>
<td>BKB</td>
<td>T42N R10W</td>
<td>1, 6, 7, 12</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Castle Rock</td>
<td>CAS</td>
<td>T41N R04W</td>
<td>6, 7</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Hog Island State Park *</td>
<td>HOG</td>
<td>T43N R08W</td>
<td>6</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Point La Barbe</td>
<td>PLB</td>
<td>T40N R04W</td>
<td>6</td>
<td>Cobble Beach</td>
</tr>
<tr>
<td>St. Martin’s Point</td>
<td>SMP</td>
<td>T41N R02W</td>
<td>7, 9</td>
<td>Cobble Beach</td>
</tr>
<tr>
<td>Summerby Fen</td>
<td>SUM</td>
<td>T41N R04W</td>
<td>9, 11, 13, 14</td>
<td>Marl Fen</td>
</tr>
<tr>
<td>Schoolcraft Co.</td>
<td>SCP</td>
<td>T41N R13W</td>
<td>1, 2, 6, 12</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Michigan, USA – Lower Peninsula</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlevoix Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beaver Island</td>
<td>BVR</td>
<td>T39N R10W</td>
<td>6</td>
<td>Beach Flat</td>
</tr>
<tr>
<td>Fisherman’s Island State Park</td>
<td>FSH</td>
<td>T33N R09W</td>
<td>6</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Cheboygan Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duncan Bay</td>
<td>DUN</td>
<td>T38N R01W</td>
<td>7, 10</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Emmet Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail’s End Bay/Cecil Bay</td>
<td>MAC</td>
<td>T39N R04W</td>
<td>7</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Waugoshance Point</td>
<td>WAG</td>
<td>T38N R05W</td>
<td>6</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Presque Isle Co.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hammond Bay</td>
<td>HAM</td>
<td>T36N R03E</td>
<td>7</td>
<td>Interdunal Wetland</td>
</tr>
<tr>
<td>Michigan, USA – Disjunct</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crawford Co.</td>
<td>HOW</td>
<td>T27N R04W</td>
<td>4, 5, 7, 9</td>
<td>Wet Prairie Fen</td>
</tr>
</tbody>
</table>

* Approximate Site of Type Specimen
16 – 122 km. The disjunct population in New York was separated from all other populations by distances of 352 – 701 km and 78 – 577 km separated the Michigan disjunct population from all others. Local population sizes ranged from 14 to more than 1000 individuals; most sampled populations were estimated to have 400 – 500 individuals. Although most populations were spatially discrete and geographically isolated, a few of the shoreline populations were less isolated and lacked discrete limits. Therefore, a standard sampling protocol was used at each population: 60 plants were randomly sampled along transects that were placed well within the boundaries of the population. Because *S. houghtonii* reproduces vegetatively by short rhizomes which disarticulate to produce independent plants, I sampled plants that were at least a minimum of 4 meters apart. Shoreline populations generally occur in a narrow band behind the foredune that parallels the lakeshore; consequently, transects also paralleled the shoreline. Inland populations tended to spread out over much broader areas; as a result, 2+ transects were sampled in order to cover the length and breadth of the population. Populations BVR and WHT contained <60 plants; consequently, all individuals were sampled. A nondestructive sampling approach was taken whereby a single healthy cauline leaf was collected from just below the inflorescence of each plant and dried in silica gel. A random subsample of 20 individuals per population (14 from BVR and 18 from WHT) was used for genetic analysis.

The entire set of sampled *Solidago houghtonii* populations were divided into regions which were defined in two ways. First, we divided the samples into regions (subpopulations) based on geographic location and defined the Canadian populations as one region, the Michigan Upper Peninsula populations as a second, the Michigan Lower Peninsula populations as a third and each of the two disjunct populations as separate
regions for a total of five regions. Regions were also defined by substrate type: alvar, beach-flat, cobble beach, interdunal wetland, and wet-prairie fen (Table 2). With the exception of the disjunct populations, regions are known to contain additional populations to those actually sampled. We sampled approximately 40% of the known Canadian populations and approximately 22% of the known Michigan populations.

DNA Extraction, PCR Amplification, and Fragment Analysis

Total genomic DNA was extracted from 1 cm$^2$ silica-gel dried leaf tissue following the CTAB method of Doyle and Doyle (1987), modified by adding PEG 8000 (Li et al. 1994) to precipitate the polysaccharide-like components from the sample. (A more complete description is given in Chapter 2).

The study of chloroplast DNA variation in Solidago (Chapter 3) indicated the presence of several indels in each of the four non-coding intergenic spacers investigated. From these, three regions were chosen for use in this study as each displayed a number of potentially informative indels. These three markers were amplified for each sampled individual using polymerase chain reaction (PCR). The amplified regions of cpDNA were an intergenic spacer separating the trnL(UAA) and trnF genes using primers e and f of Taberlet et al. (1991), an intergenic spacer separating the psbA and trnH genes using primers TRNH(GUG) and PSBA of Saltonstall (2001), and the 5'end of the accD-psal IGS. The study of cpDNA variation in Solidago (Chapter 3) found the accD-psal IGS to be largely invariant in Solidago for the 1st 800 bp of the nearly 1300 bp region (data not shown). Therefore, primers PSAI Frag 5'-TAAGTAGTAATAATTCCAATTTAG -3' and PSAInewR 5'- TCATAGAATGGGTACCTCGA - 3' were designed and used to amplify just the variable region at the 3' end of the marker.
Due to the large number of individuals studied and because information from three separate cpDNA regions was required to determine the haplotype of each individual, fluorescent dye fragment length analysis data rather than DNA sequence data was used. Horning and Cronn (2006) showed that this is a simple yet powerful method for identifying haplotype variation for the purpose of population-level studies. Fragment analysis required that one primer of each primer pair be 5'-labeled with a WellRED (Sigma-Proligo) fluorescent dye. These dyes excite in the infrared range, 650nm – 750nm, where there is little interference from biological materials present in the sample so that results are highly reproducible.

All Polymerase Chain Reactions were performed in 12.5 μl reaction volumes with 0.5 μl of template DNA diluted 1:10, 0.25 μl total dNTPs (10 mmol), and 0.05 μl of Platinum® Taq polymerase (Invitrogen, Corp.). The remainder of the PCR cocktail varied by marker; for the accD-psal marker 0.375 μl of the supplied 10X reaction buffer (Invitrogen), 1.25 μl MgCl₂, and 0.5 μl (10 μM) of each primer was used. Multiplexing was performed for the psbA-trnH and trnL-trnF markers using 1.25 μl of 10X reaction buffer, 0.375 μl MgCl₂, 0.5 μl (10 μM) each of the PSBA and TRNH primers, and 0.25μl (10 μM) each of the TRNL-e and TRNL-f primers. All reaction mixtures were cycled on an Eppendorf Mastercycler Programmable Thermal Cycler (Eppendorf Scientific Inc.) with the following PCR conditions: 94°C for 3 min followed by 39 cycles of 94°C for 1 min, 1 min at 45°C for the accD-psal marker (50°C for the multiplexed markers) and 72°C for 45 s, and a final 5 min extension at 72°C. Following PCR, 1 μl of product was directly loaded along with 38 μl of Sample Loading Solution (SLS) and 0.5 μl of DNA size standard-600 (Beckman-Coulter) for analysis. Fragment analysis was run on a CEQ 2000XL Genetic Analyzer (Beckman-Coulter) and fragments were sized using CEQ™
FRAGMENT ANALYSIS software and the 600 bp size standard for reference. About 10% of all samples for each marker were re-amplified and scored a second time in separate runs to verify repeatability.

Two to twelve individuals per size fragment of each marker were selected for sequencing to reveal sequence variation among Solidago houghtonii individuals. PCR-amplification was performed using identical but unlabeled primers and following the protocol outlined above. PCR products were purified using AMPure magnetic bead purification system (Agencourt Bioscience) following the manufacturers protocol. Both strands of all purified PCR products were sequenced on a Beckman-Coulter CEQXL2000 Genetic Analyzer using the manufacturer’s kits and protocols and the marker specific primers (more detailed procedures may be found in Chapter 2).

Statistical Analysis of Haplotype Variation

Following the fragment analysis, the individuals were assigned a haplotype designation that represented the fragment size profile of each of the three regions of cpDNA that were amplified (Table 3). Genetic variation within Solidago houghtonii was described by treating the chloroplast genome as a single locus and the haplotypes as alleles at that locus (McCauley 1994).

Variation within local populations, within regions, and within the entire sample was first estimated by dividing the number of haplotypes present in a population by the number of individuals analyzed \( (D_{hap}, \text{ haplotype diversity}) \). Then gene diversity based on haplotype composition within each population, within regions, and over all populations was estimated as \( H_E = 1 - \sum p_i^2 \) where \( p_i \) is the frequency of the \( i \)th allele (Nei 1987). Since gene diversity is calculated from the sum of squares of allele frequencies it is an
Table 3. Haplotype Matrix

| Marker Fragment Length |  
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                        | psaI-accD              | trnL-trnF              | psbA-trnH              |
| 3 3 4 4 4 4 4 4 4 3 2 2 3 3 3 3 3 |
| 9 9 0 1 2 3 3 4 5 9 9 0 0 1 1 1 |
| 6 7 8 1 5 0 9 8 4 7 9 2 1 7 2 6 9 9 |

Haplotypes

| Haplotype |  
|-----------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| H1        | 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 |
| H2        | 1 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 |
| H3        | 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 |
| H4        | 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 |
| H5        | 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 |
| H6        | 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 |
| H7        | 0 0 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 |
| H8        | 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 |
| H9        | 0 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 |
| H10       | 0 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 |
| H11       | 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 |
| H12       | 0 0 0 0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 |
| H13       | 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 |
| H14       | 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 |

The amount of variation within the entire sample, within regions, and within local populations was calculated by the hierarchical analysis of molecular variance (AMOVA) framework (Excoffier et al. 1992) carried out using ARLEQUIN (Excoffier et al. 2005). The haplotype matrix (Table 3) of presence or absence of each fragment length per marker was used as input. Significance was tested using 16000 nonparametric permutations; this number is reported to yield less than 1% difference with the exact
probability in 99% of the cases (Guo and Thomson 1992). Haplotype frequencies within local populations were used to calculate the haploid version of Wright’s hierarchical $F$-statistics ($F_{ST}$) (Wright 1978; Weir and Cockerham 1984), providing an estimate of genetic variation among populations within regions, and among regions. In addition, $F_{ST}$ values were calculated by grouping populations into five geographic regions (New York, Ontario Canada, Upper Peninsula of Michigan, Lower Peninsula of Michigan, and Michigan Disjunct) and also by grouping populations into six regions based on substrate type (Alvar, Beach Flat, Cobble Beach, Interdunal Wetland, Marl Fen, and Wet Prairie Fen) (Table 2).

Gene flow among populations was approximated as $Nm$, the number of reproductively successful migrants per generation between populations. Because the cpDNA used in this study is haploid and maternally inherited, $Nm$ can be estimated using the expression $F_{ST} = 1/(1 + 2Nm)$, where $N$ is the number of individuals in each population and $m$ is the fraction of migrants in each population in each generation (Slatkin 1993).

The relationship between geographic distance and genetic distance was examined using a Mantel matrix correlation analysis. This test evaluates whether haplotype frequencies vary continuously over space which is a distribution consistent with an isolation-by-distance (IBD) model (Slatkin 1993). Geographic distances between sampling sites were measured as linear distances and represented by a distance matrix calculated between pairwise sites. Genetic divergence between all pairwise combinations of local populations ($F_{ST} > 0$) was estimated and regressed against the geographical distance matrix using a Mantel test as implemented in ARLEQUIN (Excoffier et al. 2005). The significance of the regression coefficient of matrix correspondence ($r$) was assessed.
by 10,000 randomized permutations. In order to determine if the Mantel test was biased by the extremely disjunct New York population, the analysis was repeated with the New York population removed from the data set.

Assuming that it may be more difficult for seed to travel among the geographic regions than within a geographic region, isolation-by-distance was evaluated separately for each geographic region (excluding the two disjunct regions each with only one population). Because population CAB is disjunct from the other Canadian populations by being separated by a large expanse of water, the Canadian region was evaluated after removing CAB from the data set.

Isolation-by-distance was also evaluated by asking whether genetic distance is greater between populations in different regions than between populations in the same region over and above what is expected due to geographical distance alone? For this analysis, a partial correlation between the matrix of pairwise $F_{ST}$ values, the matrix of geographic distances, and a third matrix designating geographic regions was calculated using a multiple factor Mantel permutation test (Smouse et al. 1986) as implemented in ARLEQUIN (Excoffier et al. 2005). The region matrix was composed of zeros and ones following Sacks et al. (2004) where 0 indicated two sampling locations in the same region and 1 indicated two sampling locations in different geographical regions. Since the partial correlation coefficient between genetic distance and geographic region acts as a control for the effects of geographical distance, its significance would support the hypothesis that geographical regions account for genetic distance over and above that explained by geographical distance alone (Sacks et al. 2004). In order to evaluate whether habitat type influences genetic structure, a similar multiple factor Mantel test was
performed as described above. For this test a substrate matrix (0 = same substrate type, 1 = different substrate type) was substituted for the geographical region matrix.

Lastly, the relationships between all chloroplast haplotypes were estimated and a haplotype network was inferred by the statistical parsimony method of Templeton et al. (1992) as implemented in the TCS 1.21 program (Clement et al. 2000). The network was produced using sequence data generated from individuals representing each haplotype and a 95% probability limit was applied. Indels in the aligned haplotype sequences were coded for presence/absence and added to the matrix prior to analysis.

Results

Haplotype Variation

In total, 452 individuals were sampled from 23 populations (see Table 4 for # per population). Nine fragment lengths were recovered in PCR products amplified from the *accD-psaI* IGS, five in the *psbA-trnH* IGS product, and three in the *trnL-trnF* IGS PCR product. All fragment length variation was attributed to insertion/deletion length mutations. When considering the size fragments from all three markers simultaneously, there were 13 unique combinations each designated as a unique haplotype.

DNA sequencing of several individuals for each fragment length confirmed the presence of shared indels and also revealed the presence of three nucleotide polymorphisms. Individuals from populations SUM and BRG who share the *psbA-trnH* 319 bp size fragment also share a guanine (G) at nucleotide position 241 where all other sequenced individuals for all other *psbA-trnH* size fragments have a thyamine (T). At bp 57, located within a *psbA-trnH* indel shared by individuals from populations HOW, SBY,
and SUM there is a guanine (G) and individuals from population BRG have an adenine (A). This resulted in an additional haplotype identified as haplotype 14. In addition, sequencing revealed that individuals from populations SUM and BRG share a thyamine (T) as bp 212 in the trnL-trnF IGS where all other sequenced individuals for all other size fragments have a cytosine (C). The fourteen haplotypes compiled from the data are characterized in Table 3. Of the fourteen cpDNA haplotypes, 2 were present at a frequency >0.30 and 7 were present at a frequency between 0.16 and 0.01 (Fig. 8). The remaining 5 haplotypes were present at a frequency < 0.005; two occurred in only two of the sampled individuals and three were each found in just single individuals. Haplotypes 6 and 7 were the most common occurring in 139 of 452 sampled individuals.

Genetic Diversity

Of the 23 sampled S. houghtonii populations, 12 were polymorphic (Tables 4 and 5). The haplotype diversities ($D_{hap}$) for individual populations ranged from 0.05 to 0.2. Because this measure is directly affected by the number of individuals sampled per population, the overall haplotype diversity for the entire sample set was markedly lower ($D_{hap} = 0.03$). Populations BKB, SCP, SUM (Mich. – UP region) and HOW (Mich. – disjunct region) each contained four haplotypes and relatively higher haplotype diversities of 0.20 as compared to the 11 populations fixed for a single haplotype ($D_{hap} = 0.05 – 0.07$) (Tables 2 and 4).

The estimates of gene diversity ($H_{E}$) are generally insensitive to sample size and ranged from 0.00 for monomorphic populations (no diversity) to 0.59 in population SCP. A gene diversity value of 1.0 would indicate the haplotypes were evenly distributed throughout the sampled population. The overall gene diversity across the 23 populations
of *S. houghtonii* was relatively high, $H_E = 0.76$, however, the mean was considerably lower, $H_E = 0.19$, reflecting the large number of populations (11 of 23) that were monomorphic (Table 4).

The number of haplotypes found within geographical regions ranged from two in the New York disjunct region to ten in the Michigan Upper Peninsula region (Table 5). Although four haplotypes were present in both the Canadian and Michigan–disjunct regions, more individuals were sampled in the former and thus its haplotype diversity was lower ($D_{hap} = 0.029$) compared to that of the Michigan–disjunct region ($D_{hap} = 0.200$) (Table 4). A relatively high haplotype diversity ($D_{hap} = 0.200$) but lower gene diversity ($H_E = 0.41$) was observed in the Michigan–disjunct region. Gene diversity was also relatively low ($H_E = 0.42$) in the NY–disjunct region while gene diversity was highest ($H_E = 0.67$) in the Michigan–Upper Peninsula region. Each of the disjunct populations, located in their own geographical region, had a single private haplotype. The remaining 4 private haplotypes were all located in the Michigan – Upper Peninsula region (Table 5).
Table 4. Distribution of cpDNA Haplotypes among 23 *Solidago houghtonii* Populations

<table>
<thead>
<tr>
<th>Region</th>
<th>Population Code</th>
<th>Haplotypes</th>
<th>Total</th>
<th>$D_{hap}$</th>
<th>$H_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY-Disjunct</td>
<td>BRG 6</td>
<td>14</td>
<td>20</td>
<td>0.10</td>
<td>0.42</td>
</tr>
<tr>
<td>Ont. CA</td>
<td>CAB 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>LLC 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>SBY 7 13</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>STR 2 18</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>WFS 10 8 2</td>
<td></td>
<td>20</td>
<td>0.15</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>WGL 7 13</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>WHT 18</td>
<td></td>
<td>18</td>
<td>0.06</td>
<td>0.00</td>
</tr>
<tr>
<td>Mich.-UP</td>
<td>ALB 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>BKB 16 2 1</td>
<td></td>
<td>20</td>
<td>0.20</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>CAS 19 1</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>HOG 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>PLB 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>SMP 19 1</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>SCP 8 1</td>
<td></td>
<td>20</td>
<td>0.20</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>SUM 15 2 1</td>
<td></td>
<td>20</td>
<td>0.20</td>
<td>0.42</td>
</tr>
<tr>
<td>Mich.-LP</td>
<td>BVR 14</td>
<td></td>
<td>14</td>
<td>0.07</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>DUN 19 1</td>
<td></td>
<td>20</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>FSH 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>HAM 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>MAC 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>WAG 20</td>
<td></td>
<td>20</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>Mich.-Disjunct</td>
<td>HOW 3 1 15 1</td>
<td></td>
<td>20</td>
<td>0.20</td>
<td>0.41</td>
</tr>
<tr>
<td>Total</td>
<td>23 9 1 6 10 139 139 20 31 18 73 2 1</td>
<td>452</td>
<td>0.03 0.76</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Underlined values are private haplotypes. $D_{hap}$ = hyplotype diversity; $H_E$ = haploid gene diversity

Table 5. Chloroplast Polymorphism within Five Sampled Geographic Regions

<table>
<thead>
<tr>
<th>Geographical Region</th>
<th>No. of populations</th>
<th>No. of individuals</th>
<th>No. of haplotypes</th>
<th>$D_{hap}$</th>
<th>$H_E$</th>
<th>No. of polymorphic populations</th>
<th>No. of private haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NY - Disjunct</td>
<td>1</td>
<td>20</td>
<td>2</td>
<td>0.100</td>
<td>0.42</td>
<td>1 (100%)</td>
<td>1 (50%)</td>
</tr>
<tr>
<td>Ont. CA</td>
<td>7</td>
<td>138</td>
<td>4</td>
<td>0.029</td>
<td>0.62</td>
<td>4 (57%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Mich. - UP</td>
<td>8</td>
<td>160</td>
<td>10</td>
<td>0.063</td>
<td>0.67</td>
<td>5 (50%)</td>
<td>4 (40%)</td>
</tr>
<tr>
<td>Mich. - LP</td>
<td>6</td>
<td>114</td>
<td>3</td>
<td>0.026</td>
<td>0.51</td>
<td>1 (17%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Mich. - Disjunct</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>0.200</td>
<td>0.41</td>
<td>1 (100%)</td>
<td>1 (25%)</td>
</tr>
</tbody>
</table>
The number of haplotypes found within regions grouped by habitat type ranged from three for both the alvar and cobble beach habitats to seven for the interdunal wetland habitat (Table 6). The wet prairie fen habitat is composed of only one population, HOW, from which 4 haplotypes were recovered; one of which is a private haplotype. Two populations (BRG and SUM) occur in habitats identified as marl fens. Both these populations are polymorphic. They share haplotype 9 and also have three private alleles recovered between them. The beach flat habitat had a relatively low haplotype diversity ($D_{hap} = 0.077$) but the highest gene diversity ($H_E = 0.73$) (Table 6).

Table 6. Chloroplast Polymorphism within Six Regions Defined by Habitat

<table>
<thead>
<tr>
<th>Substrate Type</th>
<th>No. of populations</th>
<th>No. of individuals</th>
<th>No. of haplotypes</th>
<th>$D_{hap}$</th>
<th>$H_E$</th>
<th>No. of polymorphic populations</th>
<th>No. of private haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvar</td>
<td>5</td>
<td>100</td>
<td>3</td>
<td>0.030</td>
<td>0.60</td>
<td>3 (60%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Beach Flat</td>
<td>3</td>
<td>52</td>
<td>4</td>
<td>0.077</td>
<td>0.73</td>
<td>1 (33%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Cobble Beach</td>
<td>2</td>
<td>40</td>
<td>3</td>
<td>0.075</td>
<td>0.52</td>
<td>1 (50%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Interdunal Wetland</td>
<td>10</td>
<td>200</td>
<td>7</td>
<td>0.035</td>
<td>0.62</td>
<td>4 (40%)</td>
<td>2 (29%)</td>
</tr>
<tr>
<td>Marl Fen</td>
<td>2</td>
<td>40</td>
<td>5</td>
<td>0.125</td>
<td>0.46</td>
<td>2 (100%)</td>
<td>3 (60%)</td>
</tr>
<tr>
<td>Wet Prairie Fen</td>
<td>1</td>
<td>20</td>
<td>4</td>
<td>0.200</td>
<td>0.41</td>
<td>1 (100%)</td>
<td>1 (25%)</td>
</tr>
</tbody>
</table>

Population Structure of *Solidago houghtonii*

The hierarchical analysis of molecular variance (AMOVA) among populations indicated that 76.1% of variation was partitioned among populations with 23.9% of variation within populations. The $F_{ST}$ value was 0.760 ($P = 0.00$) for the entire study area indicating significant departure from the null hypothesis of $F_{ST} = 0$ (no variation between subpopulations) (Table 7). Gene flow within the population was estimated to be $Nm =$
0.16 indicating that only limited gene flow occurs between populations. When the populations were grouped according to geographical regions, 15.51% of the variation was partitioned among regions, 61.44% among populations within regions, and 23.05% of the variation was within populations. Grouping the populations according to substrate type produced the same trend; 10.42% of the variation was partitioned among regions, 66.19% among populations within regions, and 23.39% of the variation was within populations (Table 7). When the populations were divided into regions by either geographical location or habitat, the differentiation among populations for geographical regions ($F_{ST} = 0.770$) or for habitat ($F_{ST} = 0.766$) was not significantly different than $F_{ST}$ for the entire study area. In all cases, most of the variation occurred among populations (Table 7). A separate AMOVA was performed for each geographic region (except for the two disjunct regions with 1 population each) and for each substrate type (data not shown). In all geographic regions and for all substrate types, except for the marl substrate, variation occurred primarily among populations; $F_{ST}$ values ranged from 0.602 – 0.971 (all partitions were significant at $P = 0.00$). Among population variation was only 7.48% for the region defined by a marl substrate (populations BGR and SUM) with 92.52% of the variation occurring within subpopulations. The $F_{ST}$ value was 0.075 ($P = 0.07$) which indicates that $F_{ST}$ was not significantly different from the null hypothesis of no differentiation between subpopulations. This result is surprising since populations BGR and SUM are geographically separated by 626 km. Seed movement between them would seem unlikely at such a great distance.
Table 7. Hierarchical Analysis of Molecular Variance (AMOVA) of Solidago houghtonii Samples based on Haplotype Frequencies

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation indices ($F_{ST}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The entire study area</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among populations</td>
<td>22</td>
<td>133.943</td>
<td>0.30496</td>
<td>76.1</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>429</td>
<td>41.300</td>
<td>0.09627</td>
<td>23.99</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>451</td>
<td>175.243</td>
<td>0.40123</td>
<td></td>
<td>0.760*</td>
</tr>
<tr>
<td>Divided into five geographical regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>4</td>
<td>41.631</td>
<td>0.06478</td>
<td>15.51</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>18</td>
<td>92.313</td>
<td>0.25660</td>
<td>61.44</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>429</td>
<td>41.300</td>
<td>0.09627</td>
<td>23.05</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>451</td>
<td>175.243</td>
<td>0.41765</td>
<td></td>
<td>0.770*</td>
</tr>
<tr>
<td>Divided into six regions by substrate type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Among regions</td>
<td>5</td>
<td>412.20</td>
<td>0.04290</td>
<td>10.42</td>
<td></td>
</tr>
<tr>
<td>Among populations within regions</td>
<td>17</td>
<td>92.724</td>
<td>0.27247</td>
<td>66.19</td>
<td></td>
</tr>
<tr>
<td>Within populations</td>
<td>429</td>
<td>41.300</td>
<td>0.09627</td>
<td>23.39</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>451</td>
<td>175.243</td>
<td>0.41765</td>
<td></td>
<td>0.766*</td>
</tr>
</tbody>
</table>

*P = 0.00

Haplotype Distribution

The distribution of 14 haplotypes recovered from 23 populations of Solidago houghtonii was not random but showed strong geographic patterns (Fig. 9; Table 4). Several haplotypes occurred in multiple populations and were distributed over multiple regions. Six private haplotypes were recovered from five populations: population SCP (haplotype 2; Michigan-UP region / interdunal wetland region), population BGR (haplotype 3; NY-disjunct / marl fen region), population HOW (haplotype 5; MI-disjunct / prairie fen region), population ALB (haplotype 8; Michigan-UP / interdunal wetland region), and population SUM (haplotype 13 and haplotype 14; Michigan-UP / marl fen region) (Tables 5 and 6). Haplotype 3 and haplotype 14 from populations BGR and SUM differ at only 1 nucleotide position. They share 5 unique indels and two unique nucleotide substitutions which were not recovered in any other haplotypes. In addition to these two
Figure 9. Collection Localities and the Distribution of 14 Solidago houghtonii cpDNA Haplotypes
closely related haplotypes, populations BGR and SUM also share haplotype 9 which was recovered in only 2 other sampled individuals (one from HOW and one from SMP). The most widespread haplotypes were haplotypes 6 and 7 (Fig. 9; Table 4). Haplotype 6 was recovered in eight Michigan populations and was spread across both the Upper and Lower Peninsula regions. Haplotype 7 was geographically more widespread occurring in ten populations and all geographic regions except New York. Haplotype 11 occurred in five of the seven Canadian populations and was also recovered from 2 individuals in population SUM located 218 km from the nearest Canadian population. Three of the seven Canadian populations were monomorphic for either haplotype 7 or haplotype 11. Of the eight Michigan Upper Peninsula populations, three were monomorphic for either haplotype 6 or haplotype 8, and five of the six Michigan Lower Peninsula populations were monomorphic for either haplotype 6 or haplotype 7. The single polymorphic Michigan Lower Peninsula population, DUN, was nearly monomorphic as 19 sampled individuals possessed haplotype 7 and only 1 possessed haplotype 10 (Table 4).

Test for Isolation by Distance

Population pairwise genetic distance ($F_{ST}$) values were plotted against the geographic distance separating all pairs of populations in order to evaluate whether populations are isolated by distance (Fig. 10). The slight positive slope of the regression line was not significantly different from zero indicating that there is no evidence for isolation by distance (IBD) although genetic differentiation and geographic distance appeared to be moderately correlated ($r = 0.13, P = 0.08$). After removing the extremely disjunct New York sample site (identified in Fig. 10 with a box), the Mantel test revealed a highly significant $r$-value of 0.27. The permutation analysis found a larger value for $r$
in only 31 of 10,000 comparisons ($P = 0.003$) indicating that all populations except the disjunct New York population are isolated by distance (Fig. 11). This finding is counter intuitive because removal of the furthest sampling sites should lead to the remaining populations being less isolated by distance.

The multiple factor Mantel test indicated that the partial correlation between genetic distance and geographical region ($r_{y2-1} = 0.11$, $P = 0.06$) was slightly but not significantly greater than the partial correlation between genetic distance and geographical distance ($r_{y1-2} = 0.05$). Genetic distance between populations did not differ significantly according to whether pairs consisted of the same or different substrate types ($r_{y2-1} = 0.02$, $P = 0.39$). These results indicate that neither geographical region nor substrate type account for genetic distance above that which would be explained by geographical distance alone.
Figure 11. Relationship between Genetic Distance ($F_{ST}$) and Geographical Distance of all Sampling Pairs minus the New York Disjunct Population

Haplotype Relationships

The rooted haplotype network depicts the relationships between the 14 *Solidago houghtonii* haplotypes (Fig. 12). The network shows that haplotype 6 and haplotype 7, each represented by $n = 139$ sampled individuals, are closely related and occupy a central position in the network. Two unresolved loops resulting from homoplasies connect these haplotypes with four others. The predominantly Canadian haplotypes, haplotype 10 and haplotype 11 appear to be derived from Haplotype 7. Haplotype 6 which occurs throughout Michigan’s Upper and Lower Peninsulas, but not in Canada, appears to have given rise to haplotypes 2 and 12 which also occur only in Michigan. Haplotype 4 is closest to the root, *S. ceasia*. Haplotypes 3, 4, 5, and 14 are all found in disjunct populations, they occur along a long branch, and are clearly in derived positions. Haplotypes 3 and 14, both private haplotypes, are separated by one autapomorphy with haplotype 3 being derived from haplotype 14. The network suggests a Michigan origin.
for *Solidago houghtonii* since haplotypes 6 and 7 predominate throughout the region and are centrally located in the network.

Figure 12. Statistical Parsimony Network of 14 *Solidago houghtonii* cpDNA Haplotypes

Note: Each line in the network represents a single base-pair mutational change. Small circles between lines represent intermediate sequence haplotypes that were not detected in the study. \(n\) = the numbers of individuals displaying this sequence haplotype.
Discussion

Genetic Diversity and Population Differentiation

In this study, genetic differentiation, geographical patterns of variation, and phylogeographical origin of populations of Solidago houghtonii from across the range of the species were demonstrated based on variability in three noncoding regions of cpDNA: accD-psal, psbA-trnH, and trnL-trnF intergenic spacer regions. These cpDNA intergenic regions each contain a number of indels that varied within and among S. houghtonii populations revealing not only genetic variation within the population but also providing information on the contribution of seed movement to the overall population structure.

The results indicate a high level of genetic variation across the 23 populations of Solidago houghtonii from New York, Canada, and Michigan with 14 different cpDNA haplotypes detected. This is high when compared to other endangered and geographically restricted species. For example, Penstemon haydenii (Scrophulariaceae) displayed a monomorphic cpDNA RFLP banding pattern (Caha et al. 1998) and Walker and Metcalf (2008), using cpDNA sequence data, found that Astragalus jaegerianus (Fabaceae) was monomorphic for the trnL-F intergenic spacer. However, a limited number of individuals were sampled in each case which may account for the lack of detected variation.

When evaluating the viability of rare, threatened, or endangered species, it is important to understand the amount of genetic diversity within the species and how that variation is partitioned. Rare species are expected to have reduced genetic variation within populations and increased differentiation among populations (Ellstrand and Elam 1993), although this is not always the case (reviewed in Gitzendanner and Soltis 2000). Solidago houghtonii is one of three rare plant species endemic to the shore-lines of the
Great Lakes (Guire and Voss 1963). The other two, *Iris lacustris* (Iridaceae) and *Cirsium pitcheri* (Asteraceae), have been shown to be genetically depauperate. Hannan and Orick (2000) evaluated genetic diversity in *I. lacustris* and found that it lacked detectable polymorphisms in 18 isozyme loci. Loveless and Hamrick (1988) examined genetic diversity in *C. pitcheri* and its closely related progenitor species *C. canescens*. Using isozyme electrophoresis they found that only four of 14 loci were polymorphic in *C. pitcheri*. This resulted in an expected heterozygosity of $H_E = 0.024$ which was significantly lower than the more widespread *C. canescens* ($H_E = 0.174$). Because *S. houghtonii* likely shares a similar post-glacial history with these two endemics, it is surprising that the 23 *S. houghtonii* populations examined in this study contain moderately high levels of genetic variation. Although a substantial amount of genetic variation was also observed in *S. shortii* which is a rare endemic in northern Kentucky (USA) (Calie et al. 2001).

When considering the amount of genetic diversity within a species, Nei’s (1987) measure of gene diversity ($H_E = 1 - \sum p_i^2$) may be the most useful measure of diversity because it takes into account both the number and frequency of haplotypes within the population (Apsit and Dixon 2001). Total gene diversity for *Solidago houghtonii* was moderately high ($H_E = 0.76$) which indicates a moderately high probability that two plants chosen at random from throughout the species range will have different cpDNA haplotypes. However, the distribution of cpDNA haplotypes in *S. houghtonii* was not uniform across the species range as is indicated by the average gene diversity of $H_E = 0.19$. Ten haplotypes, including 4 private haplotypes, were recovered from eight populations in Michigan’s Upper Peninsula. In contrast, Michigan’s Lower Peninsula had three haplotypes represented across six populations; but one of the haplotypes was
recovered from only one of the 114 sampled individuals. Haplotypes 6 and 7 were the most common haplotypes throughout these regions. While haplotype 7 was widely distributed across all regions except the New York disjunct, haplotype 6 was only present in the Michigan populations. The Ontario Canada populations had four haplotypes distributed across seven populations. Haplotypes 10 and 11 predominated in the northeastern alvar populations and were sparsely represented in other regions. Haplotype 10 was recovered from 1 individual in Michigan’s Lower Peninsula and haplotype 11 was recovered from 2 individuals in Michigan’s Upper Peninsula. The two marl fen populations, BRG disjunct in New York and SUM in Michigan’s Upper Peninsula, each had private haplotypes as did the Michigan disjunct population (HOW) which occurs in a wet prairie fen habitat. The private haplotype from BRG, haplotype 3, was closely related to private haplotype 14 from SUM. Haplotype 9 was recovered from a majority of sampled individuals in each of these two populations but was only recovered from 2 other sampled individuals across the species range. Despite the overall high gene diversity for the species, the unequal distribution of genetic variation among the 23 Solidago houghtonii populations suggests that all populations are important to maintaining genetic diversity in this species.

Geographic Structure among Chloroplast Haplotypes

The differentiation in chloroplast haplotypes among Solidago houghtonii populations is high, with > 76% of haplotype variation \( F_{ST} = 0.760 \) being distributed among populations. High \( F_{ST} \) values indicate there are few migrants between subpopulations and imply that gene flow has been restricted in this species for some time. This starkly contrasts to the two other Great Lakes endemics. *Iris lacustris* was
monomorphic and therefore lacked any identifiable subpopulation structure. *Cirsium pitcheri* showed only a moderate amount of subpopulation differentiation ($G_{ST} = 0.139$). The lack of population structure for both these species may be a consequence of low variability since both studies were based on isozyme data which frequently does not exhibit high levels of intraspecific variability, especially in rare species (reviewed in Hamrick & Godt 1989; Gitzendanner and Soltis 2000). Mention should be made here regarding the expected differences in $F_{ST}$ values based on the use of nuclear or chloroplast DNA markers. As cpDNA is maternally inherited, it disperses in seeds but not pollen. This leads to higher $F_{ST}$ values for cpDNA when compared to those of nuclear DNA because of differences in the effective population size for each type of marker. $F_{ST}$ values for cpDNA are typically higher due to two demographic effects associated with its maternal inheritance (McCauley 1994). First, maternal inheritance acts to reduce the effective population size of the organelle genome to half that of the nuclear genome, assuming a 1:1 sex ratio. Second, because cpDNA is considered to be haploid the effective number of copies is again reduced by half when compared to the diploid nuclear genome. These two demographic effects will cause the chloroplast genome to experience higher rates of genetic drift which will increase the level of population structure and result in a higher $F_{ST}$ value (Hamilton and Miller 2002).

In plants, gene flow occurs through the movement of both pollen and seeds. Seeds carry two copies of the nuclear genome and pollen grains carry only a single copy. Therefore the movement of nuclear genes in seeds can have a greater influence on local population size and overall population subdivision than gene flow via pollen as it contributes twice as much (Hamilton and Miller 2002). As reviewed in Hamilton and Miller (2002), the vast majority of studies examining gene flow in plants focus only on
pollen movement. Therefore the estimate of overall gene flow is often downwardly biased. Because the maternally inherited chloroplast genome is only dispersed in seeds, it is useful for estimating the levels of gene flow via seed (McCauley 1994, 1995). In a highly fragmented population such as that of *S. houghtonii*, knowledge of the rate at which seeds are colonizing new habitats would be useful. Gene movement within a plant population, whether by seed or pollen, is often limited to the point that local populations consist of a group of plants that are more or less related individuals (Willson and Traveset 2000). In addition, local population size is often small due to selection for localized conditions and/or due to very restricted gene flow (Linhart and Grant 1996). Therefore, the dispersal pattern of seeds contributes to the genetic structure of populations, and to the potential for local populations to experience genetic drift or respond to natural selection.

How plants disperse their seeds is one factor that affects its ability to successfully colonize a new area since suitable habitat may occur some distance from the seed source. Wind- and vertebrate-dispersed seeds are better adapted for long-distance dispersal than for example, ant-dispersed seeds (Willson and Traveset 2000). According to Willson and Traveset (2000), even wind-dispersed seeds do not routinely colonize habitats located at extreme distances. Long distance dispersal events typically rely on birds to ferry seed in their guts, in feathers, or in mud attached to their feet. Indeed, many plants species in the Manitoulin island region of the Great Lakes rely on birds to disperse their seed although many seeds also travel by water (Morton and Hogg 1989).

The achenes of *S. houghtonii* are topped by a hairy pappus which would seem to make them well adapted for wind dispersal. Linhart and Grant (1996) indicate that when drawing inferences about the mechanism of seed dispersal, it is best to make direct
comparisons with species that employ the same mode of seed dispersal because the same scale of spatial differentiation can be expected. Walck et al. (2001) reviewed the biology of two closely related Solidago species; S. altissima and S. shortii. Solidago altissima is widely distributed throughout North America (> 5,000,000 km$^2$) and S. shortii is a rare endemic restricted to an area of 12.2 km$^2$ in northern Kentucky (USA). Wind was shown to be the primary dispersal agent for seeds in both these species even though the seeds of S. altissima can also disperse via water (Walck et al. 2001). Seed dispersal is a function of its size and the wind velocity it faces. The seeds of S. altissima measure 1.8mm long $\times$ 0.4 mm wide, have a pappus length of 2.5–3.5 mm, and weigh ~70µg. The seeds of S. shortii measure 2.3 mm long $\times$ 0.6 mm wide, have a pappus length of 2.0–3.0 mm and weigh ~370µg. The seeds of S. altissima are reported to travel up to 14.9 m in 8–10 km h$^{-1}$ wind while the somewhat larger and heavier seed of S. shortii is believed to travel at least 15 m in 10–15 km h$^{-1}$ wind. Platt and Weis (1977) showed that S. rigida, a species closely related to S. houghtonii, has seeds that disperse up to 15 m in 10–15 km h$^{-1}$ wind. The seeds of S. rigida measure 0.8 to 1.7 mm in length and its pappus measures 3.0–4.0 mm long. The seeds of S. houghtonii are somewhat larger and its pappus is quite a bit longer than those of S. rigida. The seed measures 1.4–1.8 mm in length and the pappus measures 4–5.5 mm. Being within the range of both S. altissima and S. rigida, S. houghtonii may also be capable of dispersing seed a distance of 15 m and the longer pappus may allow it to disperse seed at even greater distances. Morton and Hogg (1989) indicate that the pappus of most members of the Asteraceae, including Solidago, is too short to facilitate wind dispersal over the distances that are typical separations between the islands of the Manitoulin region of Canada or across the straits that separate Michigan’s peninsulas.
Suitable habitat for *S. houghtonii* is fragmented with at least 1 km separating populations; most populations are separated by even greater distances. Therefore, it seems unlikely that seed would be regularly dispersed by wind between these populations. Since *S. houghtonii* habitat runs parallel to the shoreline, it may be possible that, like the seed of *S. altissima*, some seed is dispersed by water. Species of *Solidago* release the majority of their seed in late fall but some seed is retained in the capitulum and released throughout the winter and early spring (Morton and Hogg 1989; Walck et al. 2001). The inflorescences of *Solidago*, including *S. houghtonii*, are tall enough and rigid enough to remain above snow level permitting seeds to be released during the winter months when they can be swept across the frozen land and water (Morton and Hogg 1989). Morton and Hogg (1989) collected seed from several *Solidago* species in windblown debris that collected on the lee side of obstacles on the frozen landscapes of the Bruce Peninsula and Manitoulin Island, Ontario Canada. Perhaps some seed of *S. houghtonii* moves between populations in this manner. Perhaps another dispersal mechanism is that of shoreline visitors such as birds. Jolls (1994) found that only 6% - 31% of *S. houghtonii* individuals flower in a given year, that on average only 12% - 17% fertile seed is set, and that seeds remain viable for only one year. Taken together, these ecological and demographic traits are consistent with the finding of limited gene flow between populations. The estimate of gene flow between *S. houghtonii* populations was $N_m = 0.16$. This value far less than 1 and indicates that ongoing gene flow is severely hindered between populations.

The effect of population history has resulted in high $F_{ST}$ values for several rare or endemic species. For example, Wallace (2002) examined population structure in *Platanthera leucophaea* which is a threatened orchid species that colonized the prairie peninsula following the Wisconsin glacial period. Populations were once abundant east
of the Mississippi River but have been reduced to 70% of their original range by anthropogenic habitat fragmentation and invasion of exotic species. Wallace found that population structure was highly subdivided ($F_{ST} = .754$). She attributed this level of structure to the fixation of different alleles in populations. Eleven of the 23 Solidago houghtonii populations examined in this study were also fixed for different haplotypes which may account for the high degree of observed differentiation in this species.

Pleasants and Wendel (1989) found that population subdivision was greater ($G_{ST} = 0.33$) in Erythronium propullans, a narrow endemic which arose following the Wisconsin glacial period, than it was for its more widespread congener S. albidum ($G_{ST} = 0.01$). In the Great Lakes region, Geum triflorum exists in isolated alvar populations disjunct from its primary distribution in the prairies of the Midwest. Hamilton and Eckert (2007) attributed greater genetic differentiation among these alvar populations, as compared to the prairie populations, to a post-glacial range expansion of G. triflorum which was followed by range contraction thus leaving populations isolated on Great Lakes alvar.

Since S. houghtonii also has a post-glacial distribution, population history may have played a critical role in its genetic structuring as well. The significant heterogeneity of substrate specific $F_{ST}$ values ranging from 0.075 for marl substrate, to 0.602 for alvar, to 0.905 for interdunal wetland, suggests that substrate type may also be acting to shape local population structure.

Isolation by Distance

The initial Mantel test for isolation by distance seemed to indicate that populations were isolated but not significantly. This was perplexing as seemingly suitable habitat sites (personal observation) remain uncolonized which suggests that seed
is not migrating even over relatively short distances. Examination of individual pairwise $F_{ST}$ values between the marl fen population in Michigan’s Upper Peninsula (SUM) and the marl fen population disjunct in New York (BRG) revealed a close genetic relationship suggestive of a long-distance dispersal event. A high percentage of sampled individuals from both populations have haplotype 9 which is much underrepresented in the remainder of the overall population. In addition, haplotype 3 found in the New York population appears to be derived from haplotype 14. Neither of these haplotypes is represented in the overall population. It is well documented that when plant species disperse their pollen or seeds at distances greater than 1km, the long distance dispersal events act to homogenize their population structure at the landscape scale (Linhart and Grant 1996). Therefore, it appears that the close genetic relationship between these two populations was having a homogenizing affect and skewed the overall IBD results. When the disjunct New York population was removed from the analysis, there was significant evidence for isolation by distance over all remaining subpopulations. The results of the multiple factor Mantel test indicated that neither the geographical region nor the substrate type accounted for genetic structure above that explained by geographical distance. Therefore, IBD suggests that overall gene flow is strongly limited at the geographical scale.

Phylogeography of *Solidago houghtonii*

Phylogeography is a useful tool for detecting historic gene flow events and can provide useful insight into the geographic structuring of genealogical lineages within plant species. Phylogeographic analysis relies on interpreting patterns of congruence, or lack thereof, between the geographical distribution of haplotypes and their genealogical relationships (Schall et al. 1998). A pattern of congruence is observed if clades of closely
related haplotypes are geographically restricted and occur in close proximity. This pattern results when mutations remain localized within the geographic context of their origin due to highly restricted gene flow between populations (Butaud et al. 2005). In general, the network of Solidago houghtonii haplotypes is congruent with genealogy. Haplotype 7 is both centrally located in the network and geographically widespread which suggests that it is the ancestral haplotype from which the remaining haplotypes were derived. Haplotype 6 is closely related to haplotype 7 and is widespread throughout Michigan. These results are suggestive of a Northern Michigan origin for S. houghtonii.

Derived haplotypes should occur at the tips of the network and should be localized geographically (Schall et al. 1998). Haplotype 10 occurs at a tip in the network and appears to be derived from haplotype 11. These two haplotypes are geographically localized, occurring predominantly in the more northerly region of the Canadian distribution although they do occur infrequently in Michigan. Haplotype 10 was recovered in 1 individual in population DUN in Michigan’s Lower Peninsula. Haplotype 11 was recovered in 2 individuals in population SUM in Michigan’s Upper Peninsula. This suggests that ancestral polymorphisms may exist. The Michigan disjunct population, HOW, appears to be closely related to other Michigan populations as 15 of 20 sampled individuals have the common haplotype 7. In addition, HOW has three derived haplotypes; 4, 5, and 9. Population SBY from the southern end of Manitoulin Island in Ontario, Canada was the only other population to have haplotype 4. This suggests that either a long distance dispersal event has occurred between HOW and SBY or there is persistence of an ancestral polymorphism. Haplotypes 3 and 14 occur at the end of a long branch and are clearly in derived positions. Both of these haplotypes are restricted to marl fen habitats located inland from the typical shoreline habitats of S. houghtonii.
While not geographically localized, Haplotype 14 occurs in SUM in Michigan’s Upper Peninsula and Haplotype 3 occurs in BRG in New York, they are clearly closely related as there is only 1 nucleotide bp that separates them. These two populations also share haplotype 9. The genealogical relationship between haplotypes 3 and 14 is clearly incongruent with geography and is suggestive of a long-distant dispersal event from SUM in northern Michigan to New York.

Historical Considerations Regarding Population Structure

For species that occur in regions which have undergone some form of climatic change such as glaciation, the geographic variation in their genetic structure may be a legacy of historical fluctuations in population size and gene flow (Hamilton and Eckert 2007). With the end of the last glacial maximum, and the climatic oscillations that subsequently followed, the distribution of plant and animal species across North America has dramatically changed (Pielou 1991). During the time of glaciation, many temperate species were isolated in southern glacial refugia. As ice sheets retreated populations migrated northward leaving isolated disjunct populations behind. One alternative for the origin of *S. houghtonii* in the Great Lakes region is that it migrated to its current distribution from a more southerly region as the glaciers receded. However, all the historically known populations are in areas that were covered by ice during the Wisconsin glacial period. Since *S. houghtonii* is restricted to open alkaline habitats primarily associated with the shorelines of the Great Lakes, it seems unlikely that it survived glaciation in the forested habitats that existed south or east of the Great Lakes (Loveless and Hamrick 1988).
Another explanation is that *Solidago houghtonii* arose in Northern Michigan sometime after the retreat of the last glacial advance and spread east throughout the Manitoulin and Bruce Peninsula regions of Canada. The area currently occupied by *S. houghtonii* has remained ice-free for only \( \sim 11,000 \text{ YBP} \) (Karrow 1987; Karrow et al. 2007). Beginning about 10,000 \text{ YBP} the water levels in Lakes Michigan and Huron were much below their present level so that about half of the present-day lake bottoms were exposed as dry land (Karrow 1987; Karrow et al. 2007). During this time the Bruce Peninsula and Manitoulin Islands were connected and only a small river separated the easternmost edge of Michigan from the Manitoulin region. Waters in the Great Lakes fluctuated greatly during this time and crested about 6m above their present level some \( 5,000 \text{ YBP} \), finally settling to their present level about \( 3,000 \text{ YBP} \) (Karrow 1987; Karrow et al. 2007). Therefore, current shoreline populations of *S. houghtonii* cannot be older than 11,000 \text{ YR} and they may be as young as 3,000 \text{ YR}.

The data from this study suggests that the origin of *S. houghtonii* may have been in northern Michigan as this region contains a high percentage of widespread haplotypes in addition to several private hyplotypes. The current fragmented distribution of *S. houghtonii* populations is suggestive of an episode of range expansion during the warmer Holocene hypsithermal period which lasted from about 9000 to 6000 \text{ YBP}. During this time *S. houghtonii* may have migrated from northern Michigan into Canada following the continuous land mass of the Niagara escarpment. As climate cooled *S. houghtonii* populations may have contracted leaving isolated populations scattered along the shorelines of Lakes Michigan and Huron in Northern Michigan and Canada. As spatial isolation increases between populations, genetic drift is expected to become more influential than gene flow and populations are expected to display greater genetic
differentiation and may drift to fixation. Populations of *S. houghtonii* are clearly
differentiated with 11 of 23 populations being monomorphic for their cpDNA haplotype.
Haplotype 6 predominates throughout Michigan but is not found in Canada. Haplotypes
10 and 11 predominate through the northeastern Canadian populations and are poorly
represented in Michigan (Fig. 9). Present population size may be sufficiently large to
maintain current levels of genetic variation. However, if *S. houghtonii* is of recent origin
there may not yet have been enough time for genetic drift to bring all populations to
fixation.

The distribution of *Solidago houghtonii* haplotypes throughout Michigan and
Canada suggests that this species may have undergone a range expansion and a
subsequent contraction since all Canadian haplotypes are represented in the Michigan
populations. In addition, the New York Bergen Swamp population (BRG) appears to
have been founded by a long-distance dispersal event from Summerby Swamp (SUM) in
Michigan’s Upper Peninsula; although it is not clear by what means the dispersal would
have occurred.

**Implications for Conservation of *Solidago houghtonii***

According to the recovery plan for *Solidago houghtonii* (USFWS 1997), this
species is particularly vulnerable to extirpation because it is restricted to narrow shoreline
habitats of the northern Great Lakes. Anthropogenic habitat destruction is considered to
be the greatest threat to this shoreline species. While the species may have experienced
fluctuations in lake level throughout its history, the current fluctuations appear to be more
dramatic and unpredictable (USFWS 1997). Lake level changes that occur rapidly over
short time scales have the potential to limit available habitat and accentuate the
fragmentation among existing populations. Consider the mining of limestone from alvar. Or, consider the construction of beach retaining walls, marinas, roads and homes. These events as well as recreational foot and off-road vehicle traffic act to destabilize dunes and beach flats, level dune ridges and prevent dune formation. All have acted to further fragment the habitat by decreasing the size and increasing the isolation of populations.

The affects of anthropogenic habitat fragmentation are perhaps most evident in Michigan’s Lower Peninsula. Interesting, in this highly fragmented region all shoreline populations were monomorphic, or nearly monomorphic, perhaps as a result of restricted gene flow.

My results indicate that populations of *Solidago houghtonii* generally fit the expectations for rare or endemic species (Ellstrand and Elam 1993). They are highly differentiated and isolated by distance. However, they unexpectedly contain relatively high levels of genetic diversity. High levels of genetic diversity were also been observed in *S. shortii* which is an endangered species endemic to a 2.2 km² area in northern Kentucky (Calie et al. 2001). It seems that both *S. shortii* and *S. houghtonii* have several life history traits in common and may be maintaining high levels of diversity in similar ways. For example, both are perennial outcrossing clonal species that are self incompatible. Both inhabit a variety of habitats (*S. shortii* – cedar glades, rock outcrops and fallow pastures, and *S. houghtonii* – interdunal wetland, cobble beaches, marl fens, and alvar). Both prefer circumneutral to alkaline substrates. (See Buchele et al. 1989, 1991a for *S. shortii* life history traits.) Beck et al. (2001) determined through morphometric analysis that *S. shortii* displays a wide range of morphological variation. While a formal study of morphological variation has not been performed for *S. houghtonii*, the wide variation within the species has been noted and has led several
botanists (Mitchell and Sheviak 1981; Morton 1979; Pringle pers. comm.) to propose the species was actually comprised of several entities that were all called $S. \text{houghtonii}$; but see Chapters 2 and 3.

A commonly held belief is that rare species exhibit less phenotypic variation than wide-spread species due in part to their limited ecological tolerances or genetic homozygosity (Beck et al. 2001). For both $S. \text{houghtonii}$ and $S. \text{shortii}$, morphological distinction between populations could be the result of restricted gene flow due to geographic distance between the populations. Phenotype variability also results from epigenetic phenotypic plasticity. Through phenotypic plasticity plants can modify the expression of their phenotype in varying environments without a genetic change. This phenomenon is very widespread in plants and often is most pronounced under stressful circumstances (Booy et al. 2000). In her review of the evolutionary implications for phenotypic plasticity, Sultan (1987) emphasized the importance of environmental factors with respect to phenotypic variation in natural plant communities. In her view, the differences in growth between individual plants are a reflection of environmental factors rather than genetic factors. While Buchelé et al. (1989) found no striking differences among the habitats in which $S. \text{shortii}$ grows a similar study has not been performed for $S. \text{houghtonii}$. Both species appear to grow in stressful habitats. *Solidago shortii* occupies rocky, xeric habitat (Buchelé et al. 1989). *Solidago houghtonii* occupies the shorelines of Lakes Michigan and Huron where the availability of water fluctuates along with the lake level, on alvar that is seasonally flooded and followed by xeric conditions, and in poorly drained marl. These conditions could account for the morphological variation observed in $S. \text{houghtonii}$. Further study may reveal the underlying causes of phenotypic variation
across the range of *S. houghtonii* and would be invaluable to designing appropriate conservation management strategies.

High genetic variation may be maintained in these species as they are long-lived perennials that reproduce both clonally and by the establishment of seed. According to Pleasants and Wendel (1989), long-lived species are able to maintain higher genetic diversity as they have more opportunities within their lifetime to pass on their alleles. Clonality can increase genetic diversity by increasing generation time in species that initially have high levels of genetic diversity; the longevity of a genet can greatly exceed the longevity of a given ramet. However, for some species such as the Great Lakes endemic *Iris lacustris* which diverged from *Iris cristata* with only a subset of its genetic diversity, clonality may increase genetic drift by reducing its effective population size. Thus, I would expect that the high level of genetic diversity observed in *S. houghtonii* would be maintained by its clonality. While the survival of adapted clones and the ability to rapidly colonize favorable habitats is most likely ensured by vegetative reproduction, the establishment of seedlings is important to the maintenance of genetic diversity within a population. Through sexual reproduction, new combinations of alleles are introduced into populations which may allow them to adapt as environments change. While it is clear that there is limited seed dispersal between populations, $Nm = 0.16$, even one migrant per generation can offset the adverse effect of genetic drift (Wright 1931).

The results of this study are relevant to the management of the threatened species *Solidago houghtonii*. The lack of new populations coupled with the loss of at least 20 percent of the historically known populations over the last 35 years (NaturServe 2009) can be explained by both its limited seed migration and its increased habitat fragmentation. While individual populations of *Solidago houghtonii* may be genetically
differentiated and geographically isolated, they collectively all contribute to the high genetic diversity of the species. Many of the established populations appear to have drifted to fixation which may be associated with a postglacial expansion and contraction of the species range. Populations in Michigan’s Lower Peninsula have a within region \( F_{ST} \) value of 0.971. This indicates that this region has experienced the greatest loss of genetic variation which is perhaps due to the more extensive habitat fragmentation that has occurred in the Lower Peninsula. Protecting the habitat from further fragmentation will preserve the present level of genetic variation within this species. This seems especially important for Michigan’s Upper Peninsula which has the highest gene diversity and the largest percentage of polymorphic populations. Further studies focused on the levels of nuclear variation would reveal the extent to which pollen moves between populations and its contribution to the overall genetic differentiation between populations. And lastly, detailed demographic studies should be conducted in order to more effectively assess the conservation status of \( S. \ houghtonii \).
APPENDIX

Voucher Information for Species Used in This Study
Species are listed by section and subsection following Semple and Cook (2006). Taxon name is followed by voucher number, DNA accession number, and collection locality. All vouchers are deposited at WMU unless otherwise noted. * = GenBank accession; † = purchased seed or plant (collection locality not available).

gigantea Aiton, TEX408224, GIG009, Texas: Hunt Co.; Solidago gigantea Aiton*; Solidago subsec.

Nemorales. Solidago nemoralis Aiton subsp. nemoralis, PJL1045, NEM001, Michigan: Kent Co.;

Solidago velutina D.C. subsp. californica (Nutt.) Semple, PJL1185, CAL001, California: Yolo Co.;


143