

4-4-2010

DNA Barcoding Analyses of White Spruce (*Picea glauca* var. *glauca*) and Black Hills Spruce (*Picea glauca* var. *densata*)

Janessa Ann James

Follow this and additional works at: http://knowledge.e.southern.edu/senior_research



Part of the [Biology Commons](#)

Recommended Citation

James, Janessa Ann, "DNA Barcoding Analyses of White Spruce (*Picea glauca* var. *glauca*) and Black Hills Spruce (*Picea glauca* var. *densata*)" (2010). *Senior Research Projects*. Paper 4.
http://knowledge.e.southern.edu/senior_research/4

This Thesis is brought to you for free and open access by the Southern Scholars at KnowledgeExchange@Southern. It has been accepted for inclusion in Senior Research Projects by an authorized administrator of KnowledgeExchange@Southern. For more information, please contact dbravo@southern.edu.

Southern Adventist University

Collegedale, TN

**DNA Barcoding Analyses of White Spruce (*Picea glauca* var. *glauca*) and
Black Hills Spruce (*Picea glauca* var. *densata*)**

A thesis submitted in partial satisfaction

of the requirements for the degree of

Bachelor of Science

in

Biology

by

Janessa Ann James

April 4, 2010

Thesis Committee:

Lee A. Spencer, Ph.D., Chairperson

Joyce L. Azevedo, Ph.D.

Introduction

The white spruce [*Picea glauca* (Moench) Voss] is an ecologically and economically important tree with a wide range throughout Alaska, Canada, and parts of the northern continental United States (O'Connell et al. 2006, Taylor 1993). *P. glauca* is the state tree of South Dakota (as the Black Hills spruce) as well as the provincial tree of Manitoba (Taylor 1993). Four taxonomic varieties of *P. glauca* are recognized: *P. glauca* var. *glauca*, *P. glauca* var. *albertiana*, *P. glauca* var. *densata*, and *P. glauca* var. *porsildii* (Strong and Hills 2006).

All varieties of white spruce except var. *glauca* are commonly thought to be hybrids between white spruce and Engelmann spruce (*P. engelmannii*) (Taylor 1993). Strong and Hills (2006) analyzed seed-cone scales in *P. glauca* and *P. engelmannii* specimens from 676 locations to determine the extent of hybridization between these species. They proposed an emended species (*P. albertiana*) to account for intermediates between *P. glauca* and *P. engelmannii*, including all varieties of white spruce except *P. glauca* var. *glauca*.

In this experiment, DNA barcoding will be used to determine the genetic relatedness of the white spruce *P. glauca* var. *glauca* and *P. glauca* var. *densata*, commonly referred to as the Black Hills spruce. *P. glauca* var. *densata* is an isolated population of white spruce found in the Black Hills, South Dakota, and Wyoming (Strong and Hill 2006). Because of the genetic isolation of these varieties, we hypothesize that the white spruce (var. *glauca*) and the Black Hills spruce (var. *densata*) are in fact distinct species. This research uniquely adds to the *P. glauca* classification debate through the introduction of DNA barcoding as a tool of investigation.

DNA Barcoding

Hebert et al. (2003) proposed DNA barcoding as means of rapid species identification through the use of a standardized gene region. This short gene region, referred to as a DNA barcode, must be found in all specimens under study and must contain sufficient variation to allow for differentiation between species. DNA barcodes can be used to identify a species by amplifying the barcode from a small piece of DNA from an unknown organism and comparing it to the sequences of known species. This procedure is becoming increasingly important for the classification and identification of species in modern taxonomy (Kress 2007).

Barcoding has been used extensively in the animal kingdom, facilitated in large part by development of a universal DNA barcode. This animal barcode is a mitochondrial gene known as cytochrome c oxidase I (COI or *cox1*) (Herbert 2003). Mitochondrial DNA in plants, however, contains too low of variation between species to be used as a reliable barcode in these organisms (Fazekas et al. 2008). A number of gene regions have been proposed as potential plant barcodes, but a universal barcode remains controversial. Most of the proposed barcodes come from the plastid region of the plant cells (Kress et al. 2005; Newmaster et al. 2006; Lahaye et al. 2008).

This research endeavor examined the gene sequences of five different potential barcoding genes from *Picea glauca* var. *glauca* and *Picea glauca* var. *densata* specimens. The five gene regions investigated were *rpoC1*, *rpoB*, *matK*, *trnH-psbA*, and *rbcL*. Chase et al. (2007) proposed the first four of these regions as useful barcodes. The four regions were suggested for use in two different combinations (1) *rpoC1*, *rpoB*, and *matK* or (2) *rpoC1*, *matK*, and *trnH-psbA*.

The *trnH-psbA* region is an intergenic spacer region found in plastid DNA. This region is non-coding and was originally suggested as one of the two leading candidates for barcoding

regions in angiosperms (Kress et al. 2005). Later research alternatively suggested the use of *rbcL* as a core barcoding gene (Newmaster et al. 2006). Researchers felt that the difficulty in aligning sequences in *trnH-psbA* above the genera rank (as reported in Kress et al., 2005) made the *trnH-psbA* gene region a poor choice for a single universal barcode. They instead suggested a tiered approach using a primary barcode to differentiate between highly-divergent specimens and a more specific secondary barcode to differentiate at the species level. The *rbcL* gene was tested as a potential primary core gene and found to be an acceptable gene region for this task. The barcoding gene has several advantages such as its ease of amplification and alignment. This gene is also the most characterized plastid coding region in GenBank, a widely-used gene sequence database (Newmaster et al. 2006), allowing researchers easy access to a large number of sequences from a variety of plant species.

Kress and Erickson (2007) proposed the use of *rbcL* with *trnH-psbA* as a global plant barcode. This combination was chosen after testing nine potential barcode loci across 48 genera. With this combination, the researchers reported nearly 88 percent success in species differentiation in their test specimens. More recently the *matK* gene was proposed as a universal DNA barcode for flowering plants (Lahaye et al. 2008). Fazekas et al. (2008) reported in their study that though *matK* provided the highest species resolution of any of the single regions they tested, its success was complicated by its relatively low amplification rates. The gene region *rbcL* was the only region amplifiable in all test subjects and generally sequenced the easiest. The *rbcL* region, however, showed significantly less sequence variability (along with the other coding regions) than *matK* or the non-coding regions. The researchers thus proposed that multiple gene regions should be used in a barcoding protocol. They suggested a combination of

genes from the coding genes *rbcL*, *rpoB*, and *matK* as well as the non-coding regions *trnH-psbA* and *atpF-atpH*.

The Consortium for the Barcode of Life (CBOL) was formed in 2004 to develop DNA barcoding as a standard for species identification (Ratnasingham and Hebert 2007). CBOL now includes over 120 organizations in 45 countries. In August 2009 the CBOL's Plant Working Group published *rbcL* and *matK* as their chosen 2-locus universal plant barcode. The research data, however, revealed greater success in obtaining *rpoB*, *rpoCl*, and *trnH-psbA* from gymnosperms than *matK* (CBOL's Plant Working Group 2009). For this reason, *rpoB*, *rpoCl*, and *trnH-psbA* were included in our research along with the prescribed 2-locus universal plant barcode of *rbcL* and *matK*.

Materials and Methods

Plant specimens were obtained from their native environments. A white spruce (var. *glauca*) and two black spruce (*Picea mariana*) specimens were collected by Dr. Lee Spencer from Inuvik, Northwest Territories, Canada. Five Black Hills spruce (var. *densata*) specimens were collected by Dr. Spencer from the Black Hills in South Dakota. The plants were processed in a research laboratory in Hickman Science Center at Southern Adventist University. After reaching the lab, specimens were stored at -80 degrees Celsius to minimize DNA degradation. DNA was extracted from needle and/or cone samples for each specimen using the PowerPlant™ DNA Isolation Kit (MO BIO Laboratories, Inc.) and the ZR Plant/Seed DNA Kit™ (Zymo Research). DNA extracts were analyzed with standard gel electrophoresis using 1.5% agarose in TAE (Sambrook et al. 1989). Gels were stained using ethidium bromide at a concentration of 2.5µg/ml to visualize DNA (Sambrook et al. 1989). Stained gels were analyzed with a UVP EC3

documentation system equipped with VisionWorks Software located at Southern Adventist University.

Following successful isolation of DNA, *matK*, *rpoB*, *rpoC1*, *rbcL*, and *trnH-psbA* gene regions were amplified for each extract using polymerase chain reaction (PCR) protocols as outlined by the Royal Botanical Gardens, Kew (<http://www.kew.org/barcoding/protocols.html>). PCR uses specific primers to amplify desired segments of DNA from the volume of DNA extracted from each plant. Through cycles of heating and cooling, the primers anneal to the DNA, and a heat-resistant DNA polymerase replicates the specific genes. After many cycles, sufficient copies of the gene sequences are made to be analyzed further.

After completion of PCR, the amplified chloroplast gene segments were separated using standard gel electrophoresis and analyzed in the same manner as the original DNA extracts. PCR products which were successfully amplified were outsourced to a commercial laboratory (Macrogen, Inc.) for DNA sequencing. Sequence alignments were completed using *Geneious Pro* 4.8.5 software. The sequence database program GenBank, accessed via *Geneious*, was used to obtain and compare sequences from other researchers (where available) to increase the validity of experimental data.

Results

The *matK* gene

The *matK* region proved the most difficult to amplify of all five gene regions analyzed. Of the eight spruce specimens examined, only one (*P. glauca* 1) yielded a quality *matK* sequence. Three additional white spruce (*P. glauca*) and two additional black spruce (*P.*

mariana) sequences were obtained from GenBank for comparison. Alignments with these six total *matK* sequences revealed a distinct pattern of four nucleotide differences between black spruce and white spruce specimens (Fig. 1). Note that the very ends (~75-175 nucleotides) of

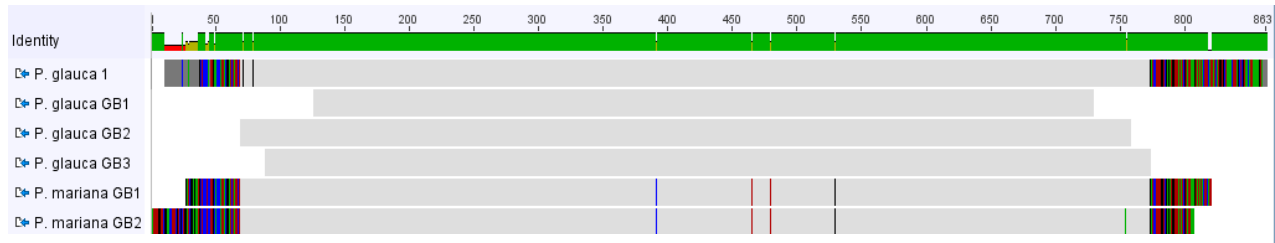


Figure 1. Alignment of *matK* gene sequences

Alignment of four white spruce (*P. glauca*) and two black spruce (*P. mariana*) sequences revealed a pattern of four nucleotide differences (~ nucleotides 390-530 above) which distinguished between these species. Sequences labeled GB were retrieved from GenBank. *P. glauca* GB1 through GB3 correspond to GenBank accession numbers EU749471 through EU749473. *P. mariana* GB1 and GB2 correspond to EU74974 and EU74975, respectively.

the *matK* sequences showed significant variability (partly due to sequence quality at distant ends) and thus were not used during analysis. No Black Hills spruce sequence data for *matK* was available for comparison.

The *rpoB* gene

Sequences for the *rpoB* gene region were obtained from three of the eight spruce specimens. A total of eight sequences (four using forward primers and four using reverse primers) were obtained from these three specimens. Four white spruce and two black spruce *rpoB* sequences were additionally retrieved from GenBank. Alignments with the experimental and GenBank specimens revealed nearly identical nucleotide sequences for white spruce, black spruce, and Black Hill spruce specimens (Fig. 2).

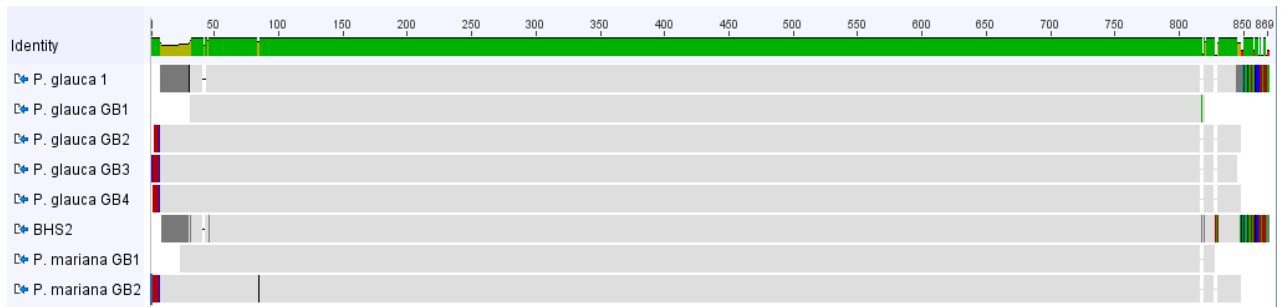


Figure 2. Alignment of *rpoB* gene sequences

Alignment of five white spruce (*P. glauca*), two black spruce (*P. mariana*), and one Black Hills spruce (BHS2) specimens revealed nearly identical sequences for the *rpoB* gene sequence. Sequences labeled GB were retrieved from GenBank. *P. glauca* GB1 through GB4 correspond to GenBank accession numbers EU749242 through EU749245. *P. mariana* GB1 and GB2 correspond to EU749246 and EU749247, respectively.

The *rpoC1* gene

Sequences for the *rpoC1* gene region were obtained from three of the eight specimens examined. A total of seven sequences were determined for these three specimens. GenBank provided an additional three white spruce and one black spruce *rpoC1* sequences. As with *rpoB*, alignments with *rpoC1* revealed nearly identical sequences for the white spruce and black spruce specimens examined (Fig. 3). No Black Hills spruce data was available for this gene.

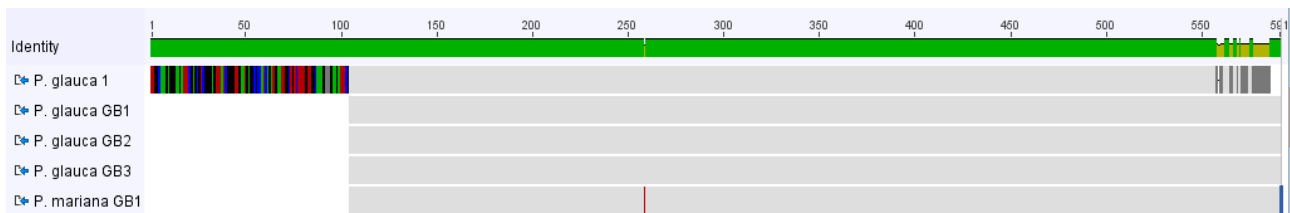


Figure 3. Alignment of *rpoC1* gene sequences

Alignment of four white spruce (*P. glauca*) and one black spruce (*P. mariana*) specimens revealed nearly identical sequences for the *rpoC1* gene sequence. Sequences labeled GB were retrieved from GenBank. *P. glauca* GB1 through GB3 correspond to GenBank accession numbers EU750383 through EU750385. *P. mariana* GB1 corresponds to EU750386.

The *rbcL* gene

The *rbcL* gene region, the most successfully sequenced locus, was obtained for seven of the eight specimens studied. A total of eleven sequences were obtained from these specimens in addition to four white spruce and two black spruce sequences retrieved from GenBank. Initial

comparisons between white spruce and black spruce from GenBank showed 99.5% pairwise identity which at first raised concerns about the effectiveness of this barcode for species differentiation. Alignments for all *rbcL* sequences obtained, however, revealed a distinct three nucleotide difference between black and white spruces (Fig. 4). The Black Hill spruce *rbcL* sequences (BHS2 through BHS4) showed the identical pattern of white spruce sequences at these three nucleotide locations.

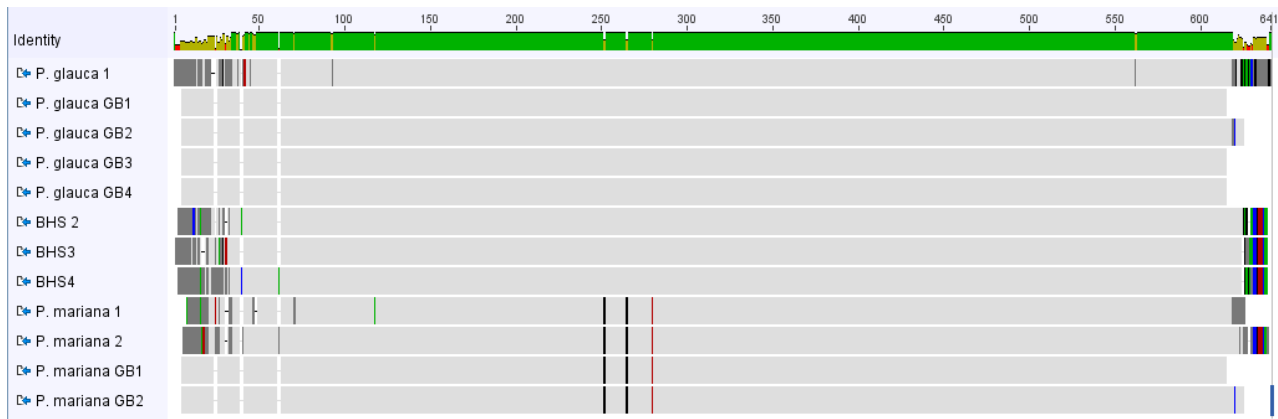


Figure 4. Alignment of *rbcL* gene sequences

Alignment of five white spruce (*P. glauca*), four black spruce (*P. mariana*), and three Black Hills spruce specimens revealed a pattern of three different nucleotides (~ nucleotides 250-280 above) between white spruce and black spruce. Black Hills spruce specimens were indistinguishable from white spruce based on this pattern. Sequences labeled GB were retrieved from GenBank. *P. glauca* GB1 through GB4 correspond to GenBank accession numbers EU677080 through EU677083. *P. mariana* GB1 and GB2 correspond to EU677084 and EU677084, respectively.

The *trnH-psbA* intergenic spacer

The *trnH-psbA* intergenic spacer region was obtained for five of the eight spruce specimens studied. Nine sequences were obtained from these five specimens which included a white spruce, a black spruce, and three Black Hills spruce. Four white spruce and two black spruce were additionally retrieved via GenBank. Alignment of *trnH-psbA* sequences revealed two distinct differences between black spruce and white spruce specimens. First, the black spruces lacked a distinct five nucleotide segment found in white spruce sequences (Fig. 5). Second, there was a single nucleotide difference present between the black and white spruces

found in each sequence evaluated (Fig. 5). Black Hills spruce sequences (BHS2-BHS4) showed identical patterns as white spruce sequences at these two locations. A couple other regions showed slight variability, but the differences were not consistent within species and thus could not be used for species differentiation.

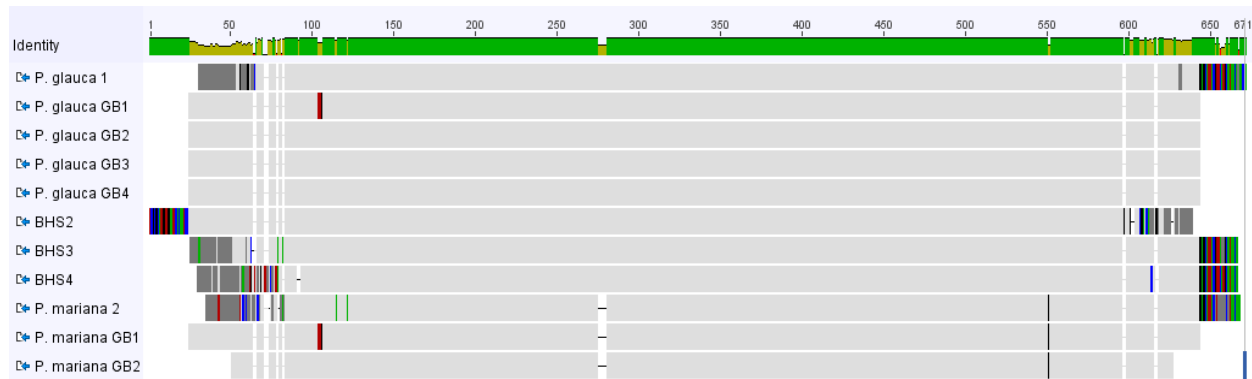


Figure 5. Alignment of *trnH-psbA* intergenic spacer sequences. Alignment of five white spruce (*P. glauca*), three black spruce (*P. mariana*), and three Black Hills spruce specimens revealed a five nucleotide segment (~ nucleotide 275 above) missing in black spruce samples but present in white spruce and Black Hills spruce sequences. There was also a distinct single nucleotide difference present (~ nucleotide 550 above) between these two groups. Sequences labeled GB were retrieved from GenBank. *P. glauca* GB1 through GB4 correspond to GenBank accession numbers EU750621 through EU750624. *P. mariana* GB1 and GB2 correspond to EU750625 and EU750626, respectively.

Discussion

Of the five potential barcoding regions analyzed, *matK*, *rbcL*, and *trnH-psbA* showed potential as effective barcodes within the *Picea* genus. In each case, these three barcoding regions revealed distinct differences between black spruce (*P. mariana*) and white spruce (*P. glauca*). These differences of three to six nucleotides initially seemed fairly insignificant when compared to the 400-800 nucleotide sequences, but detecting these small but consistent variations was critical to using barcoding sequences effectively.

The remaining regions, *rpoB* and *rpoC1*, were not variable enough to differentiate between the separate species of white and black spruce. This inability indicated that *rpoB* and *rpoC1* may not be effective as barcoding regions, especially within genus *Picea*. Because of their

inability to distinguish between two known distinct *Picea* species, these barcoding regions were unable to be used to provide meaningful analysis of the relationship between the white spruce (*Picea glauca* var. *glauca*) and Black Hills spruce (*Picea glauca* var. *densata*).

Though *matK* showed potential as an effective barcode, difficulty in obtaining *matK* sequences prevented comparison of Black Hills spruce specimens. These results are consistent with earlier studies which showed the lowest amplification success of *matK* in gymnosperms when compared to *rpoB*, *rpoC1*, *rbcL*, and *trnH-psbA* (CBOL's Plant Working Group 2009). The challenges associated with obtaining *matK* sequences may inhibit its potential effectiveness as a barcode. Further perfection of extraction/amplification procedures appear to be necessary to increase the effectiveness of this barcoding region. As *matK* is one of the two barcoding regions proposed as the 2-locus universal plant barcode (CBOL's Plant Working Group 2009), increasing amplification success in this gene region is crucial to improving barcoding efforts as a whole.

Analysis of the relationship between white spruce and Black Hills spruce was obtained through alignments with *rbcL* and *trnH-psbA*. Both of these barcoding regions showed indistinguishable sequences for white spruce and Black Hills spruce specimens. These results indicate that Black Hills spruce (var. *densata*) is not in fact a genetically distinct species from white spruce (var. *glauca*), contrary to the initial hypothesis. These results support the current classification of Black Hills spruce as a variety of white spruce.

DNA barcoding is becoming an increasingly important tool for the identification of plant species for taxonomic and other practical purposes. As barcoding protocols are developed, it will continue to be important to obtain sequences for putative barcoding regions from a large variety of plant specimens. In this experiment, the use of five major potential barcoding regions allowed

for comparison between specimens at several loci. As the barcoding controversy continues to unfold, the results of this experiment provide additional evidence to determine which loci provide the most accurate means of identification and differentiation among *Picea* specimens.

Obtaining the sequences for *matK*, *rpoB*, *rpoC1*, *rbcL*, and *trnH-psbA* in this experiment also created a bank of known sequences for further research. This base of known sequences could allow for more rapid identification of *Picea* specimens in the future as well as potential identification of new *Picea* species. Identification and sequencing of these barcoding genes is essential to the continual improvement of DNA barcoding accuracy and efficiency.

References

- CBOL Plant Working Group. 2009. A DNA barcode for land plants. PNAS 106(31): 12794-12797.
- Chase MW, et al. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56(2): 295-299.
- Fazekas AJ, et al. 2008. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE 3(7): e2802.
- Hebert, PDN, Cywinksa A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270: 313-321.
- Kress WJ, Erickson, DL. 2007. A Two-Locus Global DNA Barcode for Land Plants: The Coding *rbcL* Gene Complements the Non-Coding *trnH-psbA* Spacer Region. PLoS ONE 2(6): e508.
- Kress WJ, Wurdack KJ, Zimmer EA, Weight LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102: 8369-8374.
- Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA 105: 2923-2928.
- Newmaster SG, Fazekas AJ, Ragupathy S. 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. Can J Bot 84: 335-341.
- O'Connell L, Mosseler A, Rajora O. 2006. Impacts of forest fragmentation on the reproductive success of white spruce (*Picea glauca*). Can J Bot 84(6): 956-965.
- Ratnasingham S, Hebert PDN. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes. 7: 355-364.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual, 2nd ed.*

Cold Spring Harbour, NY: Cold Spring Harbor Laboratory Press.

Strong WL, Hills LV. 2006. Taxonomy and origin of present-day morphometric variation in

Picea glauca (xengelmannii) seed-cone scales in North America. *Can J Bot* 84: 1129-

1141.

Taylor RJ. 1993. *Picea*. In: Flora of North America Editorial Committee, editor. *Flora of North*

America: North of Mexico. Vol. 2. Pteridophytes and Gymnosperms. New York: Oxford

University Press. pp 369-373.