Molecular Identification of Boreal Forest Roots: An Expansion of Techniques and Investigation of Limitations and Biases

by

Paul Metzler

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Forest Biology and Management

Department of Renewable Resources
University of Alberta

Abstract

Plant identification is a fundamental ecological tool. While identifying flowers and leaves is relatively straightforward, identifying roots can be difficult. Here, I expand the use of fluorescent amplified fragment length polymorphisms (FAFLPs) as a tool to identify roots. Using this molecular tool, I examine possible limitations of identifying a large set of boreal plant species and compare the utility of FAFLPs to DNA barcoding. In addition, I address some challenges specific to belowground detection of roots, namely, the influence of species and root size class. To identify roots, fragment lengths of three non-coding cpDNA regions, the *trn*T-*trn*L intergenic spacer, *trn*L intron, and *trn*L-*trn*F intergenic spacer, were resolved using capillary electrophoresis for 194 plant species common to the Canadian boreal forest. To determine whether DNA sequencing increases successful identification of closely related species, Sanger sequencing of the *trn*L intron of a subset of 24 species across nine genera was compared to FAFLPs. FAFLPs produced unique size profiles for 74% of species using all three cpDNA regions. In contrast, only 27 species (14%) could be identified using the relatively conserved *trn*L intron alone. DNA sequencing did not increase detection success: eight (33%) species could be distinguished by sequences of the *trn*L region, nine (38%) by fragment lengths of the same region. Fifteen (63%) congeneric species could be distinguished by FAFLPs of all three regions. Fine roots yielded higher DNA concentrations as well as higher DNA purity than larger root classes. Fine roots of the grass species *Poa pratensis*, produced the highest yield and quality of DNA. This suggests that false-positives in belowground assays of roots may most likely to occur from fine roots of specific species. Overall, I found that molecular tools can be effective in identifying roots, but FAFLPs and DNA sequencing have strengths and limitations, and more assumptions of the methods presented here need to be tested before accurate multiplexing of roots from large species pools can occur.

Acknowledgments

First, I would like to thank my advisor, Justine Karst, who truly cares about the work that she does and the people with whom she works. Justine has been tremendously supportive and still has allowed me to work independently and follow my interests. I have learned so much from her and from all of my coworkers in the lab group. Additionally, I would like to thank JC Cahill for serving on my committee.

I would like to thank Alison Wilson whose company and plant identification skills made field work a breeze and Pak Chow who introduced me to molecular techniques and is a wealth of knowledge and advice. I would like to thank Angie Rice and Chloe Christenson who endured huge amounts of tedious lab work in support of this project. This work was funded by an Alberta Conservation Association biodiversity grant and an NSERC Industrial Research Chair awarded to Justine Karst in collaboration with Canada's Oil Sands Innovation Alliance.

My partner, Noelle Petrowski, has been with me for almost a decade, moved with me to Canada in support of my studies, and managed to thrive here. Finally, love and thanks to my grandmother, Phyllis Brower, who found me my first job, supported me always, and passed before the completion of this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 The challenges and gains of studying roots

Roots provide the anchorage and access to water that has allowed plants to gain global success. Roots drive global processes such as carbon cycling, nutrient cycling, and soil pedogenesis (Bardgett et al. 2014). Additionally, 50-90% of primary production in temperate forests and grasslands occurs in roots (Ruess et al. 2003, Stanton 1988, Steinaker & Wilson 2008), and of the 20 elements essential to most higher plants, 18 are attained primarily through the roots (Mengel et al. 2001).

Despite the importance of roots, much of our understanding of vascular plants comes from measurements of aboveground tissues. However, making predictions based on aboveground tissues (i.e., stems and leaves) does not always reflect belowground processes. For example, rates of fine root decomposition do not mirror those of leaves (Hobbie et al. 2010, McCormack et al. 2012), root and leaf growth are often asynchronous, (Steinaker et al. 2010, Blume-Werry et al. 2015), and carbon storage can be grossly asymmetrical between above and belowground. Specifically, forest soils below 1 m depth can contain more carbon than aboveground biomass (Nepstad et al. 1994) due to sequestration by deep roots, and generally, most soil carbon is from roots (Rasse et al. 2005). By improving the accuracy of root biomass estimates, estimates of global terrestrial vegetation carbon stocks have increased by 12% (Mokany et al. 2006).

Roots are difficult to study, in part, because they are belowground (under more than 50 m of soil in extreme cases, (Stone & Kalisz 1991)), but also because they are difficult to identify to species. Unlike aboveground, where individual plants are straightforward to identify, the identity of belowground plant tissues are often inferred from a suite of tools and techniques

used to see into the “black box” of soil (Tiedje et al. 1999, Pickles & Pither 2014). Prior to the use of DNA-based tools (Bobowski et al. 1999, Jackson et al. 1999), identification of roots was limited. Tracing roots back to their stem is effective, but labor intensive, and not practical for fine, deep, or excised roots (Maeght et al. 2013, Rewald et al. 2012). Using keys based on morphological features, biochemical markers, and/or staining, can be effective for distinguishing among trees and shrubs (Cutler 1987). However, these methods are time consuming, depend on environmental conditions, and often fail to identify fine roots of woody perennials or roots of herbaceous species ((Rewald et al. 2012), but see Endara et al. (2018) for a recent successful use of defense chemicals as a marker to distinguish between closely related species).

1.2 History and current state of molecular methods for root identification

Successful molecular identification of roots was first achieved by Bobowski et al. (1999) using PCR-based restriction fragment length polymorphism (PCR-RFLP) on the chloroplast *rbc*L gene. This method was subsequently used with the *trn*L intron (Brunner et al. 2001, Ridgway et al. 2003) of the chloroplast DNA (cpDNA) and the internal transcribed spacer (ITS) of the nuclear ribosomal DNA (Moore & Field 2005). The latter of which had success identifying species from a mixed pool of roots belonging to four grassland species.

Direct sequencing of the ITS region to identify deep roots from caves was first achieved by Jackson et al. (1999). Since then, DNA barcoding, which uses the nucleotide sequence of one or more regions to classify an organism, has been the standard in identifying roots to species (for example, Linder et al. 2000, Taberlet et al. 2007, Kesanakurti et al. 2011). Although there is general consensus on which regions to target for barcoding bacteria (Woese et al. 1990),

animals (Hebert et al. 2003), and fungi (Schoch et al. 2012), there is continued debate for land plants, and this is partially due to the relatively slow evolutionary rate of plant plastids (Palmer 1992), and the low resolving power of current plant barcodes (Fazekas et al. 2009). Ideal barcodes are ~700 base pairs (bp) in length, amplified by universal primers, and their sequences vary more between species than within (Cowan et al. 2006). Often, the concatenation of two or three barcodes has been proposed for use in plant identification (Chase et al. 2007, Hollingsworth et al. 2011, Burgess et al. 2011, Hosein et al. 2017).

One major drawback of 'first-generation' sequencing, however, is it is restricted to identifying a single species; i.e., sequencing mixed-DNA template is not possible with these techniques. Being able to identify multiple species from one sample (multiplexing) has immense applications including: analyzing environmental DNA or deposits of ancient DNA (Sønstebø et al. 2010, Willerslev et al. 2014), detecting cryptic species (Hebert et al. 2004), identifying plant communities in pollen or diet analysis studies (Pompanon et al. 2012, Valentini et al. 2009), and verifying medicinal herbs and commercial products (Ganie et al. 2015).

One method to identify multiple species within an environmental sample, relies on species-specific primers and fragment lengths of PCR products. This method is similar to PCR-RFLPs but avoids the use of restriction enzymes downstream of PCR (Anderson & Cairney 2004). For instance, McNickle et al. (2008) used species-specific primers based on public sequence information to amplify fragments of distinct sizes characteristic of 10 co-occurring grassland species, and Bockstette (2017) used these methods to distinguish among tree roots growing at depth in reclaimed soils. This method has been adapted to quantify relative amounts of DNA through quantitative real-time PCR (Mommer et al. 2008, Haling et al. 2011). However, creating

species-specific primers may be impractical for plant communities rich in species because sequence information is not readily available for many species, primers take time to test and optimize, and there is a limitation to how many unique species-specific size profiles that can be created for a region that is only a few hundred base pairs long.

Fluorescent amplified fragment length polymorphisms (FAFLPs, also called FFLPs) use long-established universal plant primers to detect polymorphisms in lengths of DNA fragments. This method is less expensive than sequencing and the use of universal, rather than species-specific primers means that it is applicable to large species pools. Polymorphisms in the *trn*L region, or part of the *trn*L region, have been used successfully to identify roots (Ridgway et al. 2003, Frank et al. 2010). In addition, using FAFLPs of multiple regions can increase the likelihood of identifying unique polymorphisms. Specifically, the *trn*T-*trn*L intergenic spacer, the *trn*L intron, and *trn*L-*trn*F intergenic spacer, have been used to identify roots to much success (Taggart et al. 2011, Randall et al. 2014). Expanding these methods to boreal forest species is the focus of Chapter 2.

Next-generation sequencing (NGS) platforms, such as Illumina and Roche 454 enabled sequencing of multiple species in one sample. Roche 454 pyrosequencing produces long reads (~800 bp, Shokralla et al. 2012), and has been successfully used on the *trn*L intron to assess belowground diversity of plants, (Hiiesalu et al. 2012), and relationships between root and fungal diversity (Hiiesalu et al. 2014). Pyrosequencing is now off the market in response to a move toward Illumina sequencing, which targets smaller regions (~150 bp) at a much lower cost and produces higher quality reads (Glenn 2011). In plants, often the P6 loop of the *trn*L intron is targeted, which has a stable structure and can be amplified with universal primers (Taberlet et

al. 2007). This region has proven useful at characterizing highly degraded samples such as in diet analysis (Popanon et al. 2012, Valentini et al. 2009), and in ancient DNA (Sønstebø et al. 2010, Willerslev et al. 2014), as well as in assays of belowground plant diversity in temperate grasslands and arctic tundra (Lamb et al. 2016).

1.3 Sources of false-positives/negatives when identifying roots to species

Molecular methods for belowground plant identification have consistently detected species that are not seen aboveground (i.e., 'false-positives') (Hiiesalu et al. 2012, Kensanakurti et al. 2011). Researchers have speculated that this result could be due to the presence of plants with roots that lay dormant for years (Pärtel et al. 2012), or from roots persisting in soil after aboveground tissue has died or been removed (Kensanakurti et al. 2011, Mommer et al. 2011). In intact ecosystems, detecting dormant or recently dead roots could be one benefit of using molecular methods in diversity surveys (Hiiesalu et al. 2012). However, in disturbed ecosystems where vegetation has been removed, the presence of dead roots could bias belowground diversity surveys through false-positives. In addition to false-positives, false-negatives may also emerge with DNA amplification. DNA quality and PCR inhibitors released from roots can vary by species (Mommer et al. 2011, Karst et al. 2015), root size, and time since death. Poor DNA quality and/or an increase in inhibitors can mask species known to occur. Taken together, both biological and methodological issues can give arise to false detections.

These issues highlight the importance of the objective of my third chapter which is to measure the quality and quantity of DNA across root size classes and species in preparation for burial in a five-year experiment. The goal of this long-term experiment is to determine how long dead roots can be detected by molecular methods but is not formally included in this thesis.

1.4 Study region

My research is located in the boreal forest, the largest terrestrial ecoregion in Canada. In northeastern Alberta, *Picea mariana* (Miller) Britton, Sterns & Poggenburgh (black spruce), *Picea glauca* (Moench) Voss (white spruce) *Pinus banksiana* Lambert (Jack pine), and *Populus tremuloides* Michaux (aspen) dominate a landscape mosaic driven by topography, distance to water tables, and frequent fires (Larson 1980). The relatively low vascular plant species richness in this area (45 species per 0.25 ha plots, Zhang et al. 2014) makes this an ideal place to test the utility of FAFLPs as genetic markers. Additionally, unique FAFLP markers have already been established for a majority of plants present in the aspen parkland (Taggart et al. 2011), a region immediately south of the boreal forest.

1.5 Study objectives

The first objective of my research is to create a fragment size key of FAFLPs for common plant species in the boreal forest of Alberta. This size key will expand those existing for the region (Taggart et al. 2011, Randall et al. 2014), and be a valuable resource to identify roots to species using inexpensive molecular techniques. Towards building this database, I will also address the possible shortcomings of FAFLPs in distinguishing closely related species by comparing the detection of species using FAFLPs and DNA barcoding. The final objective is to address bias caused by false-positives from roots as a function of root size and species. Specifically, I assessed how DNA quantity and quality varies in roots differing in size and species.

Chapter 2: Molecular identification of roots from a boreal forest ecosystem

2.1 Introduction

Plant identification is a fundamental ecological tool. While identifying flowers and leaves is relatively straightforward, identifying roots can be difficult. First, roots are difficult to sample because they are buried. The most common method of sampling roots *in situ* is by soil coring, which excises roots from aboveground stems and as a result, cores typically contain roots from multiple species. Notably, within a core we expect lower species richness than that occurring across cores (e.g. tens versus hundreds of species). Core volumes range in size depending on research question and ecosystem, but typically are small enough to use manually, and large enough to sample heterogeneously distributed roots. Tracing roots back to their stem can also be an effective method to identify roots to species, but it is labor intensive, and impractical for fine, deep, or excised roots (Maeght et al. 2013, Rewald et al. 2012).

The second issue making excised roots difficult to identify to species is that roots are morphologically similar. Using keys based on morphological features, biochemical markers, and/or staining, can be effective for distinguishing among trees and shrubs (Cutler 1987; Endara et al. 2018), but these methods are time consuming, depend on environmental conditions, and often fail to identify fine roots of woody perennials or those of herbaceous species (Rewald et al. 2012). To date, polymerase chain reaction (PCR)-based molecular tools are the most effective and reliable methods of identifying roots. When identifying roots from soil cores, users require molecular tools that are high throughput, affordable, and distinguish species within a mixed sample of relatively low species richness.

A variety of PCR-based tools exist to distinguish among species. Species-specific primers that amplify DNA of target species has been used to distinguish among forest (Bockstette 2017) and grassland (McNickle et al. 2008) species and has been used in concert with quantitative real-time PCR to quantify relative amounts of DNA (Mommer et al. 2008, Haling et al. 2011). However, creating species-specific primers may be impractical because sequence information is not readily available for many plant species, primers take time to test and optimize, and there is a limitation to how many unique species-specific size profiles that can be created for a region that is only a few hundred base pairs long. 'Next generation sequencing' can generate thousands to millions of DNA sequences from multiple species present in a sample, however this level of sequencing may be superfluous when identifying roots excised within a core, where we expect species richness to be relatively low.

Other candidate tools include 'first generation sequencing', which generates a DNA sequence from an individual organism, and fluorescently amplified fragment length polymorphisms (FAFLPs), molecular markers which differentiate species using size differences in fluorescently labelled PCR amplicons (a fragment of DNA produced by PCR). Using FAFLPs, size profiles derived from unknown roots are compared to those developed from known species. Fluorescently amplified fragment length polymorphisms of multiple regions can be used to increase the likelihood of identifying unique polymorphisms. Fragment lengths have correctly identified species in mixed samples (Ridgway et al. 2003) of up to sixteen species (Taggart et al. 2011). In particular, FAFLP size keys have been previously developed for plants of two common ecosystems in western Canada, aspen parkland (Taggart et al. 2011) and the boreal forest (Randall et al. 2014).

One known issue with FAFLPs, however, is the inability to distinguish among some closely related species (Ridgway et al. 2003, Taggart et al. 2011, Randall et al. 2014). Sanger sequencing generates data of higher resolution than that derived from fragment lengths of a given amplified region, and as such, sequences may be more effective to differentiate congenerics than FAFLPs. While DNA can be extracted from bulk roots for FAFLPs, it must be separately extracted from each root fragment for Sanger sequencing, adding a considerable cost to the latter method (single versus multiple extractions). Additional costs for Sanger sequencing arise in the actual sequencing step, which is otherwise unnecessary in FAFLPs because it is a size-based technique. Thus, DNA sequences potentially provide higher resolution to species identification, but do so at a higher cost.

Here, I expand and test two candidate molecular tools for use in identifying species in the western Canadian boreal forest. To accomplish this, 209 boreal forest plant species were collected and analyzed for FAFLPs using the *trn*T-*trn*L intergenic spacer, the *trn*L intron, and the *trn*L-*trn*F intergenic spacer to generate a size key for identifying roots to species. As a first objective, I double the number of species serving as references for future studies. As a second objective, a subset of the congenerics present in the current study are sequenced and the species identification success from sequence data is compared to that of FAFLPs of the *trn*L region alone, and in combination with the *trn*T-*trn*L and *trn*L-*trn*F intergenic spacers. The *trn*L intron was selected for comparison because it is amplified with established primer sets (Taberlet et al. 1991), it contains a short and stable secondary structure, i.e., the P6 loop, useful for identifying highly degraded samples (Taberlet et al. 2007), and of the three regions targeted for the FAFLP analysis, it is the most easily and consistently resolved.

2.2 Methods

2.2.1 Field collection of reference plant tissue: leaves

I collected leaf tissue from one to six individuals from 209 species common to the boreal forest in northeastern Alberta, Canada (Figure 1). Sixty 30 x 30 m plots were chosen to represent a range of natural and disturbed habitats from across the region. These plots were intensively searched by walking 15 transects, each the length of the plot, and checking for new species within one meter of the transect. Sites included *Pinus banksiana* Lambert (Jack pine), *Picea glauca* (Moench) Voss (white spruce), *Picea mariana* (Miller) Britton, Sterns & Poggenburgh (black spruce), *Populus tremuloides* Michaux (aspen), and mixed-wood dominated upland sites, forested wetlands, and disturbed sites, such as abandoned well pads and roadsides. My sampling covered an approximately 30,000 km^2 region from 56°0'21.49"N to 54°32'27.73"N latitude (NAD 83). Replicates of the same species were taken from different plots separated by at least 25 m to capture intraspecific genetic variation.

Approximately 20 g of disease-free leaves, showing no signs of herbivory or infection, were collected for each sample in paper bags and kept on ice until frozen (-20°C) at the end of the day. For smaller herbs, stems were collected as well. For each species, a voucher specimen was collected, mounted, and deposited at the University of Alberta Herbarium (ALTA).

Frozen samples were thawed and washed with deionized water and left to air-dry until excess moisture was removed. Aluminum packets were folded around the plant samples and lyophilized using a benchtop freeze dryer (Labconco FreeZone 2.5, Kansas City, MO, USA) for three to four days. Using sterilized forceps, approximately 40 mg of plant material was placed in a 2 mL tube along with three sterilized 3 mm tungsten carbide beads. Samples were tissue-

lysed on a TissueLyser II (Qiagen Inc, Hilden, Germany) for two minutes at 30 rotations per second, repeated if necessary until pulverized.

2.2.3 Determining fragment size profiles for species

In total, FAFLP analysis was run on 2040 samples (680 individuals x 3 cpDNA regions). Total genomic DNA of leaves was extracted based on modified 2% hexadecyltrimethylammonium bromide (CTAB) protocol (Roe et al. 2010; Appendix I, Using this extraction method, only 44% of samples produced fragment lengths. Specifically, success rates for the *trn*L intron, the *trn*T-*trn*L intergenic spacer, and *trn*L-*trn*F intergenic spacers were 30%, 55%, and 47%, respectively. Owing to the low success, I re-extracted DNA of common species for which one or more regions were unresolved using a second method, 5% CTAB and a polyethylene glycol (PEG) precipitation (Griffiths et al. 2001; Appendix II).

In total, FAFLP analysis was run on 422 samples re-extracted with this new method, of which 61% produced fragment lengths. Success rates for the *trn*L intron, the *trn*T-*trn*L intergenic spacer, and *trn*L-*trn*F intergenic spacers were higher: 54%, 67%, and 68%, respectively, even though these samples were non-randomly chosen from a group more likely to fail (i.e., from samples that were unsuccessful using the 2% CTAB method). Of the 15 species for which no fragments were recovered, six, *Cinna latifolia, Equisetum fluviatile, Gymnocarpium dryopteris, Senecio eremophilus, Symphyotrichum puniceum* var. *puniceum,* and *Vaccinium caespitosum,* were not recovered by either extraction method, and nine, *Campanula rapunculoides*, *Geranium bicknellii, Geum rivale, Juncus bufonius, Juniperus horizontalis, Lathyrus venosus, Lonicera villosa, Maianthemum trifolium,* and *Malaxis monophyllos*, were not

tried with the second extraction method because they are less common in the boreal forest of northeastern Alberta.

Three regions were targeted with the universal primer sets established by Taberlet et al. (1991): the *trn*T-*trn*L intergenic spacer, the *trn*L intron, and the *trn*L-*trn*F intergenic spacer with a modified *trn*T-*trn*L forward primer (Cronn et al. 2002) (Table 1). Forward primers in each primer pair were fluorescently labelled (A2: FAM; C: VIC; E: NED; Integrated DNA Technologies, Coralville, Iowa, USA). Polymerase chain reactions (PCR) were carried out in 25 µL volumes: 12.5 µL of EconoTaq PLUS 2X Master Mix (Lucigen Corp., Middleton, WI, USA), 2.5 µL of each forward and reverse primer at 10 µM, 5.5 µL autoclaved deionized water, and 2 µL of 5-10 ng μL^{-1} DNA template. Reactions were performed using a Model 6321 Eppendorf Mastercycler Pro S gradient thermal cycler (Eppendorf Canada, Mississauga, ON, Canada). Reaction conditions were the same for all three regions, slightly modified from Taggart et al. (2011): 94°C for five minutes, followed by 35 cycles of 94°C for 60 seconds, 60°C for 60 seconds, 72°C for 80 seconds and a final extension of 72°C for 30 minutes.

Amplified product from each region was diluted 200x then 2 µL was added to 8 µL of Hi-Di formamide and 0.15 µL of GeneScan 1200 LIZ Size Standard (Applied Biosystems, Foster City, CA, USA). Future studies could increase the throughput of FAFLPs by fluorescently labeling the three primer sets and running PCR on a mixed-sample, co-amplifying all three regions. For the current study, I chose to separate species and region to reduce potential error in the creation of the fragment size key. Fragment lengths were resolved using capillary electrophoresis (ABI 3730 DNA Analyzer, Applied Biosystems, Foster City, CA, USA) and sized to the nearest base pair

using GeneMapper 4.0 software (Applied Biosystems, Foster City, CA, USA). The fragment length is determined by a peak in relative fluorescent units (RFUs).

Visualization with GeneMapper 4.0 showed that many samples contained multiple peaks and peak height varied depending on region amplified, PCR run, and species. When visualization showed multiple peaks, this could be due to multiple binding sites, "primer dimers", contamination, or noise. Because of the large variation in peak height, this could not be determined by simple RFU cutoffs and incidences of multiple peaks had to be determined within the context of the amplified region, PCR run, and species. Specifically, a peak was considered a "primer dimer", if the length of the fragment was less than 150 bp and was consistently present in multiple samples in the same PCR run. A peak was considered noise if there were one or more peaks within the same sample that were at least 10x higher than the peak in question. Contamination within a plate was identified when multiple species across distantly related taxa on the same PCR run were resolved as the same size fragment length; these peaks were removed. In some species, the selected primer sets had multiple binding sites (see Results), which lead to two or more tall peaks that were consistent across individuals within a species. In this case, up to four peak heights were recorded.

I considered a species to have a unique identifier if at least one region differed in length from that of another species by at least two base pairs. When fragment lengths were unresolved for one or more regions, I used an additional set of criteria. In the case where one region matched that of another species, but the two other regions were unresolved in one species and resolved in the other species, I considered these to be unique fragment length profiles (see Table 2). Just one missing region could be due to laboratory conditions, but two

missing regions is more likely to be characteristic, especially if the species with missing regions is compared to a distantly related species that has all three regions. Fragment lengths or a range of fragment lengths associated with each species were recorded and categorized differently. In the cases where fragment lengths varied within a species by more than 15 bp, I categorized this as a "v", denoting a highly variable species and that a consistent and useful identifier for that region could not be found. In the cases where only one sample was resolved for a particular region, and thus the length could not be verified by a replicate, a stricter standard was used. A sample with a single replicate was recorded (and denoted with a "*") only if the species had a clear, high peak height (above 2000 RFU), and good ladder size quality (SQ > 0.4, this score reflects how well the data from the GeneScan 1200 LIZ Size Standard (Applied Biosystems, Foster City, CA, USA) matches expected values), or, if the samples peak height was lower, but could be confirmed by a closely related species in the current or other published studies.

2.2.4 Sanger sequencing of congeneric DNA

The *trn*L intron that was used to establish fragment size profiles was also sequenced for individuals of species within the genera *Alnus*, *Betula*, *Carex*, *Cornus*, *Fragaria*, *Picea, Populus, Ribes*, and *Rosa* (species listed in Table 4). DNA was extracted and amplified using the methods described above. Amplified DNA was cleaned using ExoSAP (Exonuclease 1 10 units μL^{-1} (New England BioLabs M0293S) and Shrimp Alkaline Phophatase 1 unit μL^{-1} (New England BioLabs M0371S)) following the manufacturers protocol. Big Dye sequencing reactions and bi-directional sequencing was performed on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) carried out by the University of Alberta Molecular Biology Facility.

Sequence data was manually edited in Geneious v11.0.5 (Kearse et al., 2012) by replacing bases denoted as "N" that were clearly either G, C, A, or T, based on a distinct single peak. Poor quality 5' and 3' ends were trimmed (error probability limit = 0.01) and heterozygotes were detected using the heterozygote plugin (peak similarity = 50%). Bidirectional reverse sequences were aligned using the Geneious de novo assembly alignment tool. Bidirectional sequences were manually searched for inconsistencies, edited if needed (all heterozygotes were edited or replaced by the International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes), and the consensus sequence was extracted. All sequencing data was verified in GenBank databases using nucleotide BLAST.

Species were successfully identified when sequences formed monophyletic clades that include one member of the genus and not the others. For this analysis, I built a tree using neighbor-joining based on the Tamura-Nei genetic distance model (global alignment with free end gaps, cost matrix: 65% similarity) implemented through Geneious v11.0.5 (Kearse et al., 2012).

2.3 Results

2.3.2 Size key of Fluorescent Amplified Fragment Length Polymorphisms

Final success rates for the *trn*L intron, the *trn*T-*trn*L intergenic spacer, and *trn*L-*trn*F intergenic spacers were 58%, 90%, and 78% respectively. Fifteen species would not amplify, and 51 of the remaining 194 species did not produce unique size profiles, i.e., fragment lengths were identical across multiple species. Twenty-eight of these species were closely related (i.e., congenerics), and 23 were distantly related species. In other words, unique size profiles were created for 74% of species using FAFLPs of all three regions and 88% of genera in this study

(Table 3). In contrast, only 27 species (14%) could be identified to species using the relatively conserved *trn*L intron alone, leaving the remaining 167 species ambiguous.

The amount of intraspecific variation as well as nonspecific primer annealing varied by region. For the *trn*T-*trn*L, *trn*L, and *trn*L- *trn*F regions, 30, 7, and 15 species had multiple binding sites that caused multiple peak heights. Furthermore, for the *trn*T-*trn*L, *trn*L, and *trn*L- *trn*F regions, 29, 37, and 28 species showed intraspecific variation (< 15bp). In total, 75 species had some amount of intraspecific variation, 47 species had multiple binding sites, and 17 species had a large range in fragment length (> 15 bp, and up to hundreds of base pairs) that was found across individuals in the same species with relatively high confidence (peak height > 5,000 RFU).

2.3.3 Comparing identification success between FAFLPs and sequencing

Almost all GenBank searches on sequences produced a closely related species in the expected genera with high query cover (mean = 99%) and pairwise identity (mean = 99%). The exceptions were two *Fragaria vesca* sequences, which produced a blast search result in the *Festuca* genus, and one *Picea glauca* sequence that produced a *Pinus* BLAST result. These samples were considered contamination, representing 4% of the sequenced samples, and not considered further.

Eight of the nine genera that were sequenced were placed in monophyletic clades containing only that genus (Figure 2), with one exception: one individual *Alnus alnobetula* subsp. *crispa* was placed just outside of a monophyletic *Alnus* group. No species within *Alnus, Betula, Rosa, or Fragaria* could be distinguished based on sequences of the *trn*L region. Both members of the *Cornus* genus could be distinguished with sequencing, but only some members of the remaining species formed truly monophyletic clades. Most *Carex* species could be

distinguished with the exception of *Carex aurea,* for which one sample was placed slightly outside of a monophyletic group. *Ribes triste* and *Ribes glandulosum* could be distinguished, but *Ribes hirtellum* and *Ribes lacustre*, although in separate clades from each other, were both grouped with *Ribes oxycanthoides*. *Ribes oxycanthoides* showed a much smaller fragment length for the *trn*L region than other members of the *Ribes* genus (Table 4), so it is possible that this misplacement is due to *Ribes oxycanthoides* missing an important characteristic section of DNA. Finally, *Populus tremuloides* and *Picea glauca* formed monophyletic clades, but their corresponding congenerics did not produce true monophyly. The lack of discrimination within *Betula* and *Rosa* is not surprising considering all members of the species studied within these genera hybridize with at least one other species within that genera (Brouillet et al. 2010).

Of the 24 congenerics tested, eight (33%) could be distinguished by sequencing of the *trn*L region, and nine (38%) could be distinguished by FAFLPs of the *trn*L region alone. When all three regions were included fifteen congenerics (63%) could be distinguished by FAFLPs (Table 4).

2.4 Discussion

2.4.1 Amplicon size profiles: successes and areas of improvement

I achieved moderate success in identifying unique size profiles for a large pool of plant species occurring in the western boreal forest. To date, this is the largest plant FAFLP size key compiled (194 versus 95 species (Taggart et al. 2011)), and the current study doubles the number of reference species available for future studies. Fragment size differences from the *trn*L intron, and the *trn*T-*trn*L and *trn*L-*trn*F intergenic spacers differentiated 74% of the 194 species and 88% of genera in the study.

A common finding across studies using FAFLPs is the presence of intraspecific variation in fragment lengths. In the current study, 75 of 194 species showed some amount (<15 bp) of intraspecific variation. Randall et al. (2014) found intraspecific variation in all seven tree species studied, and Ridgway et al. (2003) found intraspecific variation in nine of 10 grassland species. Taggart et al. (2011) recorded only one incidence of intraspecific variation in the 49 species that were replicated (i.e., 2-10 replicates), but this could be a result of limited geographical sampling. In the current study, and in the context of the studies mentioned above, intraspecific variation in fragment sizes may be more common than previously assumed (Randall et al. 2014, Ridgway et al. 2003, Taggart et al. 2011), especially in the boreal forest. In the current study, intraspecific variation caused a high level of ambiguous fragment lengths among closely related species, but also among unrelated species. In addition, 17 species had lengths that were highly variable within species (>15 bp). These data were treated as unresolved lengths because it is unclear if they represent biological variation or if they are lab errors. Treating these values as correct would have decreased the number of ambiguous species by two. Many of the ambiguities observed in FAFLP size profiles due to intraspecific variation can be limited by restricting the number of species in an identification key to the number of species identified aboveground through field surveys. By restricting the species pool, species with the same size polymorphism are less likely to be encountered in analysis, but dormant or seasonal species may be missed. Another way to resolve ambiguities in size profiles is to test for a phylogenetic signal in intraspecific variation to predict in which taxa this variation will occur.

Multiple peaks are another consistent finding associated with FAFLPs (Ridgeway et al. 2003, Taggart et al. 2011, Randall et al. 2014). At the molecular level, multiple peaks may be a

result of nonspecific binding or binding to repeated sequences in the genome; both events can generate fragments of different sizes. While it is usually clear which peak is signal and which is noise in a single-species sample, in samples containing multiple species, it may be difficult to identify an individual species in the presence of multiple peaks, since peak height may vary depending on species and PCR inhibitors present. To improve the throughput of FAFLPs in the present study, I applied a 1 to 200 dilution to the entire plate of PCR product regardless of the band brightness of gel electrophoresis of the PCR product. In the future, it may be necessary to adjust the dilution based on band brightness to standardize the expected peak heights (RFUs). Next steps to improve FAFLP analysis in multiplexing must include developing standard methods for parsing out signal peaks from noise.

Resolving unique FAFLP size profiles is a task that needs development for application to mixed-species communities. Ideally, one value (i.e., fragment length) is associated with each region for a given species to create a unique size profile. As I show here, fragments may be absent entirely (amplification failed), or range in size owing to intraspecific variation and multiple binding sites. Many of the size profiles in this study were ambiguous because of variability that causes a range of values or missing peaks and reduces the potential for unique values. Taggart et al. (2011) offers four analysis methods to identify unknown species using these fragment size profiles. Which of the four analysis methods used depends on whether one fragment length from one region is enough (liberal) or if all known fragment lengths must be detected (conservative), and whether or not the user limits the species pool to what was detected aboveground (constrained or unconstrained). However, none of these analysis methods consider multiple peaks or intraspecific variation. Analysis of fragment size profiles in

the future should use analytical methods that take into consideration that: 1) not all known fragment lengths may be resolved for each species size profile, 2) species detected aboveground are the most likely to be found belowground (except dead, dormant, or seasonally present species), 3) some size profiles include multiple peaks for one region and, importantly, 4) intraspecific variation occurs. Moving forward, it may be possible to determine the extent of intraspecific variation and frequency of multiple peaks across species by using *in silico* methods, e.g., analysis of publicly available sequences from databases such as GenBank (National Centre for Biotechnology Information; NCBI). Bioinformatics have also been developed to probabilistically assign taxonomy to include uncertainty owing to incomplete reference databases, mislabeled reference sequences, intraspecific variation, and errors in DNA sequences (Somervuo et al. 2016, 2017, Abarenkov et al. 2018). These analytical tools could be adapted to probabilistically assign identities to species in mixed-species samples of roots.

Because FAFLPs are PCR-based methods, they are adaptable to a variety of other well-established techniques. For example, steps to resolve ambiguities in fragment lengths could be performed on a case by case basis by further processing PCR product with restriction enzymes (Ridgway et al. 2003). To reduce error and determine relative abundance of species in a mixed-sample, quantitative PCR methods along with use of internal standards could be applied to FAFLP techniques (Mommer et al. 2008, Haling et al. 2011). Overall, there have been few studies that use FAFLPs on plant roots and there is ample opportunity for expansion and improvement of these methods.

2.4.2 Identification success using FAFLPs compared to sequencing

My second objective in this study was to compare identification success between Sanger sequencing and FAFLPs of congeners, i.e., species within a genus. I determined fragment size profiles and sequenced the *trn*L intron for individuals of species within the genera *Alnus*, *Betula*, *Carex*, *Cornus*, *Fragaria*, *Picea, Populus, Ribes*, and *Rosa*. FAFLPs of the *trn*L intron were unexpectedly more highly resolved than sequencing of the *trn*L intron. Furthermore, FAFLPs of all three cpDNA regions were more resolved than relying on the size of the *trn*L intron alone. One implication of these findings is that FAFLPs can be as effective as sequencing in identifying species, and FAFLPs require less time and costs than sequencing. This finding is unexpected because sequencing gives more detailed information, i.e., the sequence of hundreds of base pairs, whereas FAFLPs provides only the region length.

Comparing the resolution of these two methods is difficult because they produce fundamentally different data. FAFLPs, in this study, produce one to three lengths of DNA that can be used as a distinguishing character to uniquely identify a species. Alternatively, sequence data relies on the percent identity, i.e., sequences are placed into monophyletic clades based on how similar they are to other sequences (see Methods). One probable reason for the lower resolving power of sequencing is the different analytical methods and standards that define species identification success. I used a commonly used neighbor-joining tree building method based on the Tamura-Nei genetic distance model (Hebert et al. 2003). However, the determination of a successful species identification may be influenced by the particular analytical tool used. For example, the choice of distance-based, coalescent-based, and character-based barcode analysis methods can have drastic effects on resolution of

phylogenetic relationships (Yu et al. 2015). In fact, character-based approaches (DeSalle et al. 2005) to sequence analysis may be most appropriately compared to FAFLPs as fragment lengths are just one characteristic of the target region.

Using FAFLPs of all three regions was much more effective than using the sequence of just one region. One implication of this finding is that it may not be as important which molecular techniques are used, rather the number of regions targeted may be the key step to identify species using DNA-based methods. With barcoding strategies, it has been suggested that the number of regions is more important than the identity of those regions for correct species identification (Fazekas et al. 2008). In this same line, the Barcode of Life Data System (BOLD, Ratnasingham & Hebert 2007) recommends the combination of *rbc*L and *mat*K regions to identify plant species. Seberg and Petersen (2009) suggest that it is unlikely that a single barcode will allow us to identify more than 70-75% of known species, although a concatenation of four barcodes allowed them to identify 92% of species within the *Crocus* L. (Iridaceae) genus. Regardless of target region, some researchers have suggested there is an "upper limit" on detecting species using barcodes (Fazekas et al. 2009). Using multiple loci to create a barcode seems to be necessary for in-depth taxonomy, but this may be impractical for applying barcodes to species identification in mixed-species samples.

Similar methods were used to extract DNA for FAFLP and Sanger sequencing in the current study, and these methods require development to increase DNA yields. The recovery rates of size lengths for each region in the current study (*trn*T-*trn*L: 58%, *trn*L: 90%, and *trn*L-*trn*F: 78%) are similar to those found in Taggart et al. (2011) (58%, 100%, and 98%). The higher recovery rates found for the *trn*L intron and the *trn*L-*trn*F by Taggart et al. (2011) could be

attributed to the use of a different DNA extraction method (DNeasy PowerPlant Pro Kit, Qiagen Inc, Hilden, Germany) or the different species present in grasslands versus the boreal forest. Additionally, I used a different forward primer for the *trn*L-*trn*F region, which was more successful in terms of number of samples amplified (unpublished data), which could explain why my recovery rate was the same for the *trn*T-*trn*L region, but lower for the other regions. These results in light of those by Taggart et al. (2011) suggest that the amplification of DNA, especially for plants, may depend on DNA extraction method and PCR inhibitors specific to species (Mommer et al. 2011). Therefore, optimization of extraction and PCR condition based on species may be a prerequisite to using PCR-based methods in species-rich systems.

2.5 Conclusion

I found that FAFLPs are an effective tool for identifying plants and may be as effective as DNA barcoding, especially when using multiple regions. However, like all DNA-based identification methods, they are not without their limitations. In future studies, I suggest using large datasets or publicly available sequences to understand how intraspecific variation and incidences of multiple peaks function in relation to species, and how we can incorporate these features into analysis of FAFLPs. In addition, future work with FAFLPs should focus on improving and standardizing lab techniques so that FAFLPs can be used in mixed-community analysis with limited noise.

2.6 Figures

Figure 1. Map of sites from which leaves of 209 plant species were collected to develop DNA size profiles used to identify roots. Sixty collection sites are denoted by pink dots.

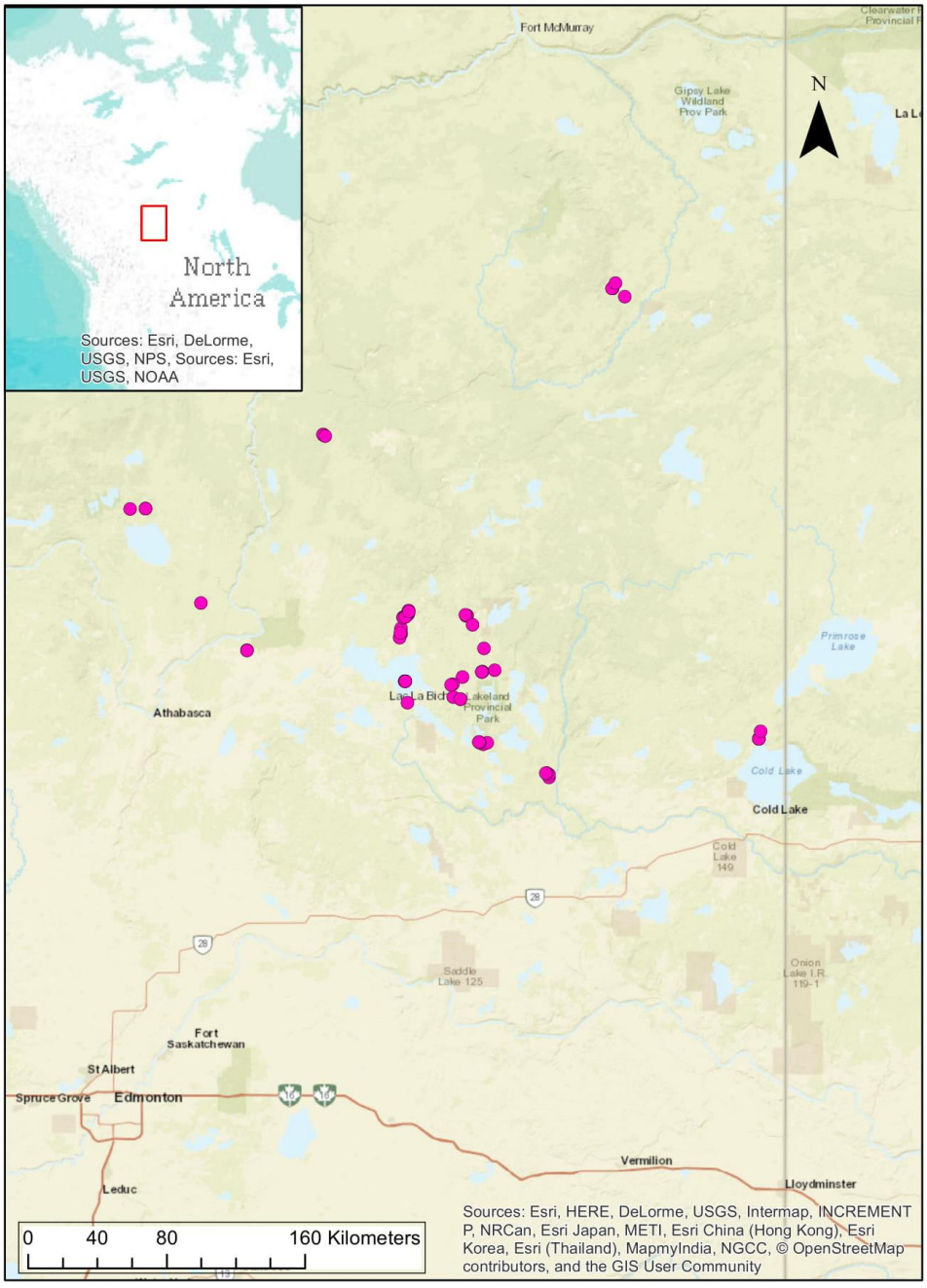
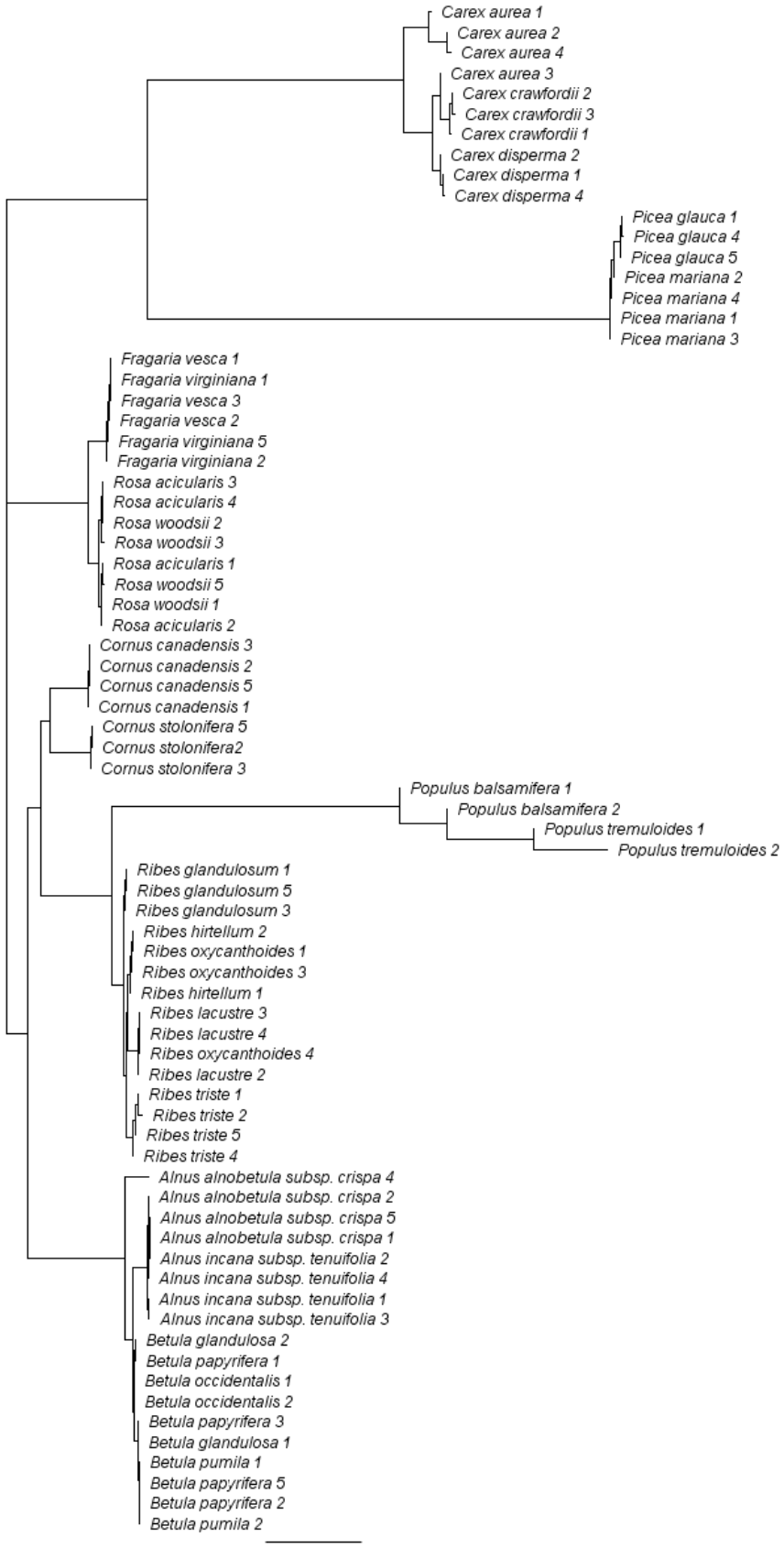


Figure 2. DNA barcoding for 24 species and nine genera included in this study. Phylogenetic tree built using neighbor-joining based on the Tamura-Nei genetic distance model on the trnL intron of each individual. Numbers represent the individual replicate.


Carex aurea 1
Carex aurea 2
Carex aurea 4
Carex aurea 3
Carex crawfordii 2
Carex crawfordii 3
Carex crawfordii 1
Carex disperma 2
Carex disperma 1
Carex disperma 4
Picea glauca 1
Picea glauca 4
Picea glauca 5
Picea mariana 2
Picea mariana 4
Picea mariana 1
Picea mariana 3
Fragaria vesca 1
Fragaria virginiana 1
Fragaria vesca 3
Fragaria vesca 2
Fragaria virginiana 5
Fragaria virginiana 2
Rosa acicularis 3
Rosa acicularis 4
Rosa woodsii 2
Rosa woodsii 3
Rosa acicularis 1
Rosa woodsii 5
Rosa woodsii 1
Rosa acicularis 2
Cornus canadensis 3
Cornus canadensis 2
Cornus canadensis 5
Cornus canadensis 1
Cornus stolonifera 5
Cornus stolonifera2
Cornus stolonifera 3
Populus balsamifera 1
Populus balsamifera 2
Populus tremuloides 1
Populus tremuloides 2
Ribes glandulosum 1
Ribes glandulosum 5
Ribes glandulosum 3
Ribes hirtellum 2
Ribes oxycanthoides 1
Ribes oxycanthoides 3
Ribes hirtellum 1
Ribes lacustre 3
Ribes lacustre 4
Ribes oxycanthoides 4
Ribes lacustre 2
Ribes triste 1
Ribes triste 2
Ribes triste 5
Ribes triste 4
Alnus alnobetula subsp. crispa 4
Alnus alnobetula subsp. crispa 2
Alnus alnobetula subsp. crispa 5
Alnus alnobetula subsp. crispa 1
Alnus incana subsp. tenuifolia 2
Alnus incana subsp. tenuifolia 4
Alnus incana subsp. tenuifolia 1
Alnus incana subsp. tenuifolia 3
Betula glandulosa 2
Betula papyrifera 1
Betula occidentalis 1
Betula occidentalis 2
Betula papyrifera 3
Betula glandulosa 1
Betula pumila 1
Betula papyrifera 5
Betula papyrifera 2
Betula pumila 2
0.04

2.7 Tables

Table 1. Primers used to isolate the three regions in this study, the *trn*T-*trn*L intergenic spacer, the *trn*L intron, and the *trn*L-*trn*F intergenic spacer. For each region, forward primer is listed first, followed by the reverse primer. These universal primer sets were established by Taberlet et al. (1991) with a modified trnT-trnL forward primer (Cronn et al. 2002).

Region	Name	Sequence 5'-3'
*trn*T-*trn*L	A2	CAAATGCGATGCTCTAACCT
	B	TCTACCGATTTCGCCATATC
*trn*L	C	CGAAATCGGTAGACGCTACG
	D	GGGGATAGAGGGACTTGAAC
*trn*L-*trn*F	E	GGTTCAAGTCCCTCTATCCC
	F	ATTTGAACTGGTGACACGAG

Table 2. Conceptual example of which combinations of fragment lengths are considered unique identifiers for that species. Note that one missing value is not enough to differentiate species one and two. However, no replicates of species four could produce fragment lengths for the *trn*T-*trn*L or *trn*L-*trn*F introns. In this study, this is considered a unique feature or this species.

Species	*trn*T-*trn*L	*trn*L	*trn*L-*trn*F	Unique?
Species 1	300	450	600	No
Species 2	301	450	X	No
Species 3	400	300	600	Yes
Species 4	X	300	X	Yes

Table 3. Size key of species included in this study and resolved fragment lengths (bp) for three regions, the *trn*T-*trn*L intergenic spacer, the *trn*L intron, and the *trn*L-*trn*F intergenic spacer. Species were collected from the boreal forest of north eastern Alberta, Canada. The *trn*L region was sequenced for the species in bold. "x" denotes a region where amplification failed for a species. "v" denotes a highly variable species (>15 bp) and indicates that a consistent and useful identifier for that region could not be found. "/" separates lengths from multiple binding sites and "-" represents variability found for a specific fragment length. "*" represents a fragment length that was found in only one replicate and could not be confirmed by a closely related species in the current or other published studies. Numbers in parentheses are replicates.

Family		Species	*trn*T-*trn*L	*trn*L	*trn*L-*trn*F
Amaranthaceae	(2)	*Blitum capitatum subsp. capitatum* Linnaeus	792-793	x	x
	(1)	*Chenopodium album* Linnaeus	813*/823*	x	x
Apiaceae	(1)	*Cicuta maculata* Linnaeus	846*	559*	x
	(2)	*Heracleum maximum* W. Bartram	840/430	571	447
	(2)	*Hieracium umbellatum* Linnaeus	x	509	442
	(1)	*Osmorhiza depauperata* Philippi	839	574	430
	(2)	*Sanicula marilandica* Linnaeus	x	571-572	x
Apocynaceae	(3)	*Apocynum androsaemifolium* Linnaeus	815	418	397
Araliaceae	(5)	*Aralia nudicaulis* Linnaeus	852-853	575	440
Asparagaceae	(3)	*Maianthemum canadense* Desfontaines	x	601-602	432
	(3)	*Smilacina stellata* (Linnaeus) Desfontaines	707*	602	417*
Asteraceae	(5)	*Achillea millefolium* Linnaeus	562	491	425-426
	(4)	*Achillea sibirica* Ledebour	768*	491	426-427
	(2)	*Artemisia campestris* Linnaeus	771*	495	440
	(1)	*Bidens cernua* Linnaeus	563*	601/503	396*
	(3)	*Cirsium arvense* (Linnaeus) Scopoli	873*	508	v
	(3)	*Erigeron philadelphicus* Linnaeus	x	453*	324
	(3)	*Eurybia conspicua* (Lindley) G.L. Nesom	x	495	432*
	(2)	*Matricaria discoidea* de Candolle	547*	492-493	440
	(2)	*Petasites frigidus var. palmatus* (Aiton) Cronquist	664	511	x
	(2)	*Petasites frigidus var. sagittatus* (Pursh) Cherniawsky	876-881	489	420

	(4)	*Solidago canadensis* Linnaeus	x	500	v
	(3)	*Solidago spathulata* de Candolle	x	500	452-458
	(2)	*Sonchus arvensis subsp. uliginosus* (M. Bieberstein) Nyman	642	507-508	417
	(1)	*Symphyotrichum boreale* (Torrey & A. Gray) A. Love & D. Love	x	x	432
	(5)	*Symphyotrichum ciliolatum* (Lindley) A. Love & D. Love	x	504	432
	(1)	*Symphyotrichum laeve var. laeve* (Linnaeus) A. Love & D. Love	896*	504	432
	(1)	*Symphyotrichum lanceolatum subsp. hesperium* (A. Gray) G.L. Nesom	x	x	432
	(2)	*Tanacetum vulgare* Linnaeus	642	503	440
	(4)	*Taraxacum officinale* F.H. Wiggers	621-622	522	402
	(2)	*Tripleurospermum inodorum* (Linnaeus) Schultz-Bipontinus	x	494*	518*
Betulaceae	(3)	***Alnus alnobetula subsp. crispa* (Aiton) Raus**	x	602-603	464
	(4)	***Alnus incana subsp. tenuifolia* (Nuttall) Breitung**	x	603-605	464
	(2)	***Betula glandulosa* Michaux**	v	440	477
	(2)	***Betula occidentalis* Hooker**	1042-1043/1033	440	446
	(4)	***Betula papyrifera* Marshall**	1043	440	475-476
	(2)	***Betula pumila* Linnaeus**	x	440	476*
	(4)	*Corylus cornuta* Marshall	854	602-603	470
Boraginaceae	(2)	*Mertensia paniculata* (Aiton) G. Don	780-781	553	453
Brassicaceae	(1)	*Arabidopsis lyrate* (Linnaeus) O'Kane & Al-Shehbaz	x	576	v
	(1)	*Lepidium densiflorum* Schrader	x	590*	x
	(1)	*Thlaspi arvense* Linnaeus	x	401*	x
Campanulaceae	(3)	*Campanula rotundifolia* Linnaeus	831	588-589	x
Caprifoliaceae	(4)	*Linnaea borealis* Linnaeus	804	578-579	447-448
	(2)	*Lonicera dioica* Linnaeus	813/175/210/365	583	443*
	(3)	*Lonicera involucrata* (Richardson) Banks ex Sprengel	x	587	442-448
	(3)	*Symphoricarpos albus* (Linnaeus) S.F. Blake	815*	587	397*
Caryophyllaceae	(1)	*Cerastium nutans* Rafinesque	537*	668*	448*
	(1)	*Moehringia lateriflora* (Linnaeus) Fenzl	707*	629*	417*
	(4)	*Stellaria longifolia* Muhlenberg ex Willdenow	642/633	637-638	433
Celastraceae	(1)	*Parnassia palustris* Linnaeus	x	686*	382*
Colchicaceae	(4)	*Disporum trachycarpum* (S. Watson) Bentham & Hooker f.	1008	582	483
Cornaceae	(3)	***Cornus canadensis* Linnaeus**	857	582-584	434
	(2)	***Cornus stolonifera* Michaux**	x	590	423*
Cyperaceae	(4)	*Carex aenea* Fernald	626	689-694	456/444
	(2)	***Carex aurea* Nuttall**	v	686	437

	(2)	*Carex bebbii* (L.H. Bailey) Olney ex Fernald	624-625	x	x
	(1)	*Carex brunnescens* (Persoon) Poiret	626*	x	x
	(1)	*Carex concinna* R. Brown	v	334-337/616*	x
	(3)	***Carex crawfordii* Fernald**	623-625	679	x
	(4)	***Carex disperma* Dewey**	627-628/262/277/618-619	689	443
	(1)	*Carex magellanica subsp. irrigua* (Wahlenberg) Hiitonen	426/417	x	x
	(2)	*Carex utriculata* Boott	426/417	x	x
	(3)	*Scirpus microcarpus* J. Presl & C. Presl	x	690	x
Dryopteridaceae	(2)	*Dryopteris carthusiana* (Villars) H.P. Fuchs	x	x	375
Elaeagnaceae	(4)	*Shepherdia canadensis* (Linnaeus) Nuttall	887	550	476
Equisetaceae	(3)	*Equisetum arvense* Linnaeus	x	334	458/431
	(2)	*Equisetum hyemale* Linnaeus	785-799	333	280-281
	(1)	*Equisetum palustre* Linnaeus	x	606*	281
	(2)	*Equisetum pratense* Ehrhart	x	334	345
	(3)	*Equisetum scirpoides* Michaux	x	325-333	366
	(4)	*Equisetum sylvaticum* Linnaeus	x	306	345
Ericaceae	(5)	*Arctostaphylos uva-ursi* (Linnaeus) Sprengel	960/951	575-576	262-263
	(1)	*Empetrum nigrum* Linnaeus	178*	483*	345*
	(2)	*Moneses uniflora* (Linnaeus) A. Gray	x	575	310
	(4)	*Orthilia secunda* (Linnaeus) House	v	593	315
	(4)	*Pyrola asarifolia* Michaux	920	623*	321
	(4)	*Pyrola chlorantha* Swartz	917-918	580-581	x
	(2)	*Rhododendron groenlandicum* (Oeder) Kron & Judd	x	581	452*
	(4)	*Vaccinium microcarpum* (Turczaninow ex Ruprecht) Schmalhausen	x	563	472-773
	(2)	*Vaccinium myrtilloides* Michaux	x	561-562	x
	(5)	*Vaccinium vitis-idaea* Linnaeus	x	567	461
Fabaceae	(1)	*Astragalus cicer* Linnaeus	659*	623	x
	(3)	*Lathyrus ochroleucus* Hooker	x	510	176
	(2)	*Medicago sativa* Linnaeus	547	x	221*
	(4)	*Melilotus albus* Medikus	1147-1148/1138	310	205
	(3)	*Melilotus officinalis* (Linnaeus) Lamarck	1149	319	216
	(3)	*Trifolium hybridum* Linnaeus	x	615-617	203-209
	(3)	*Trifolium pratense* Linnaeus	x	585	x
	(4)	*Trifolium repens* Linnaeus	x	617/305	203
	(4)	*Vicia americana* Muhlenberg ex Willdenow	x	522	179

Grossulariaceae	(3)	*Ribes americanum* Miller	1122-1123/1112-1113	x	403*
	(3)	***Ribes glandulosum* Grauer**	1103/1193-1194	586	x
	(2)	***Ribes hirtellum* Michaux**	1105/1096	586	x
	(1)	*Ribes hudsonianum* Richardson	1121*/1112*	586	x
	(5)	***Ribes lacustre* (Persoon) Poiret**	1128/1119	585	411
	(4)	***Ribes oxyacanthoides* Linnaeus**	1117-1119/1108-1109/1127	319-320	x
	(5)	***Ribes triste* Pallas**	1109-1110	580	411
Iridaceae	(4)	*Sisyrinchium montanum* Greene	740-741	551	308
Juncaceae	(2)	*Juncus balticus* Willdenow	811/625	679*	v
Lamiaceae	(3)	*Agastache foeniculum* (Pursh) Kuntze	602*	v	369-370
	(4)	*Galeopsis tetrahit* Linnaeus	x	v	342
	(1)	*Mentha arvensis* Linnaeus	x	565*	x
	(3)	*Scutellaria galericulata* Linnaeus	740	553-554	386
Lilaceae	(2)	*Lilium philadelphicum* Linnaeus	x	608	255
	(2)	*Streptopus amplexifolius* (Linnaeus) de Candolle	x	v	454-458
Lycopodiaceae	(4)	*Diphasiastrum complanatum* (Linnaeus) Holub	420/223/411	590	457
	(4)	*Lycopodium annotinum* Linnaeus	439/457	589-590	438
	(3)	*Lycopodium obscurum* Linnaeus	423	598	961
Myricaceae	(1)	*Myrica gale* Linnaeus	x	589*	x
Onagraceae	(4)	*Chamaenerion angustifolium* (Linnaeus) Scopoli	x	603-604	504/497
Ophioglossaceae	(2)	*Botrypus virginianus* (Linnaeus) Michaux	x	x	452
Orchidaceae	(3)	*Corallorhiza maculata* (Rafinesque) Rafinesque	x	772/898	471/430*
	(2)	*Corallorhiza trifida* Chftelain	x	310/542/609	x
	(1)	*Galearis rotundifolia* (Banks ex Pursh) R.M. Bateman	x	680*	350*
	(1)	*Goodyera repens* (Linnaeus) R. Brown	916*	663	480
	(1)	*Platanthera hyperborea* (Linnaeus) Lindley	882*	610-620	394*/433*
	(1)	*Platanthera obtusata* (Banks ex Pursh) Lindley	906*	619	x
	(1)	*Platanthera orbiculata* (Pursh) Lindley	x	600*	492*
Orobanchaceae	(4)	*Castilleja miniata* Douglas ex Hooker	x	548-553	432-433
	(4)	*Melampyrum lineare* Desrousseaux	803	544	394
	(3)	*Rhinanthus minor subsp. groenlandicus* (Chabert) Neuman	791-792/782-783	x	x
Papaveraceae	(1)	*Corydalis aurea* Linnaeus	x	553*	x
Pinaceae	(3)	*Abies balsamea* (Linnaeus) Miller	470	554-555	465
	(3)	*Larix laricina* (Du Roi) K. Koch	472/463	548*	x
	(6)	***Picea glauca* (Moench) Voss**	470/461	559-560	460-465

	(5)	***Picea mariana* (Miller) Britton, Sterns & Poggenburgh**	469/460	559-560	460
	(3)	*Pinus banksiana* Lambert	501/492	557	464
Plantaginaceae	(5)	*Plantago major* Linnaeus	764	578	426
	(1)	*Veronica americana* (Rafinesque) Schweinitz ex Bentham	x	553	405
Poaceae	(4)	*Beckmannia syzigachne* (Steudel) Fernald	890	609	421
	(3)	*Bromus ciliatus* Linnaeus	v	647	v
	(2)	*Bromus inermis* Leysser	v	649	443-444/394
	(4)	*Calamagrostis canadensis* (Michaux) Palisot de Beauvois	876	490	420
	(3)	*Elymus trachycaulus subsp. trachycaulus* (Link) Gould ex Shinners	668	641-645/423-428	430-432/394
	(2)	*Festuca saximontana* Rydberg	867-874	571*	v
	(3)	*Hordeum jubatum* Linnaeus	661/652	634	x
	(2)	*Koeleria macrantha* (Ledebour) Schultes	842	406	x
	(4)	*Leymus innovatus subsp. innovatus* (Beal) Pilger	x	557	497
	(4)	*Oryzopsis asperifolia* Michaux	861-884	600	421-424
	(1)	*Phalaris arundinacea* Linnaeus	880*	v	420-425/349
	(5)	*Phleum pratense* Linnaeus	881-882	608-609	425
	(4)	*Piptatheropsis pungens* (Torrey ex Sprengel) Romaschenko	358-359/349-350	597-600	424-426/394
	(1)	*Poa compressa* Linnaeus	x	611	432
	(3)	*Poa palustris* Linnaeus	882	597	425/394/444
	(3)	*Poa pratensis* Linnaeus	v	620	464*
	(4)	*Schizachne purpurascens* (Torrey) Swallen	816/740/806	605	417/394
Polemoniaceae	(3)	*Collomia linearis* Nuttall	x	592-593	440-441
Polygonaceae	(1)	*Rumex occidentalis* S. Watson	692*	623*	x
Primulaceae	(5)	*Lysimachia borealis* (Rafinesque) U. Manns & Anderberg	x	557	336-337
Ranunculaceae	(5)	*Actaea rubra* (Aiton) Willdenow	746-747	542	459
	(2)	*Anemone canadensis* Linnaeus	x	608	448
	(4)	*Anemone patens* Linnaeus	x	571	507
	(1)	*Anemone virginiana* Linnaeus	x	561*	476*
	(5)	*Caltha palustris* Linnaeus	681	564	457
	(1)	*Coptidium lapponicum* (Linnaeus) Gandoger ex Rydberg	x	544*	492*
	(1)	*Delphinium glaucum* S. Watson	760*	v	417-429
	(2)	*Ranunculus acris* Linnaeus	761	647*	444*/349*
	(1)	*Ranunculus sceleratus* Linnaeus	x	565*	x
	(4)	*Thalictrum venulosum* Trelease	746-748	609-615	469
Rosaceae	(1)	*Agrimonia striata* Michaux	x	539*	428*

	(5)	*Amelanchier alnifolia* (Nuttall) Nuttall ex M. Roemer	x	586	484
	(2)	***Fragaria vesca* Linnaeus**	998	490	497
	(4)	***Fragaria virginiana* Miller**	x	490	428-430/394
	(3)	*Geum macrophyllum* Willdenow	x	615	476-477
	(3)	*Potentilla norvegica* Linnaeus	x	599-601	432-492
	(2)	*Potentilla palustris* (Linnaeus) Scopoli	918	580	321
	(1)	*Prunus pensylvanica* Linnaeus f.	760*	560*	417*/488*
	(4)	*Prunus virginiana* Linnaeus	920*	592	210/433
	(5)	***Rosa acicularis* Lindley**	x	616-618	482
	(3)	***Rosa woodsii* Lindley**	525-526	617	482
	(2)	*Rubus arcticus* Linnaeus	x	569	493
	(3)	*Rubus chamaemorus* Linnaeus	x	556	483
	(5)	*Rubus idaeus* Linnaeus	501*	556	476
	(5)	*Rubus pubescens* Rafinesque	501*	569	492
	(1)	*Sibbaldia tridentata* (Aiton) Paule & Soj k	x	499	x
	(2)	*Sorbus scopulina* Greene	x	586	484*
Rubiaceae	(2)	*Galium boreale* Linnaeus	846	607	483
	(2)	*Galium trifidum* Linnaeus	x	592	442*
	(2)	*Galium triflorum* Michaux	x	585	470*
Salicaceae	(5)	***Populus balsamifera* Linnaeus**	x	653	399-403
	(5)	***Populus tremuloides* Michaux**	525-526	693-695	391-392
	(1)	*Salix myrtillifolia* Andersson	546	652	432
	(4)	*Salix spp.* Linnaeus	547	653-654	422-432
Santalaceae	(3)	*Comandra umbellata* (Linnaeus) Nuttall	x	572-573	182
	(2)	*Geocaulon lividum* (Richardson) Fernald	697	578	375
Saxifragaceae	(4)	*Mitella nuda* Linnaeus	377-378	537/580	438*
Typhaceae	(3)	*Typha latifolia* Linnaeus	x	x	389
Urticaceae	(3)	*Urtica dioica* Linnaeus	772	475	443
Violaceae	(5)	*Viola adunca* Smith	406-407/397-398	583	443
	(5)	*Viola canadensis* Linnaeus	406/397	583	315
	(1)	*Viola palustris* Linnaeus	405	547	437
	(3)	*Viola renifolia* A. Gray	377-378	538	438

Table 4. Comparison of identification success with FAFLPs of trnL alone, FAFLPs of trnL with trnT-trnL and trnL-trnF, and with sequencing data. Whether sequencing was successful in identifying genus and species is marked with a "y" for yes and "n" for no. FAFLP profiles of all three regions are presented next as well as whether they create a unique size profile and whether the trnL is unique on its own. "x" denotes a region where amplification failed for a species. "v" denotes a highly variable species (>15 bp) and indicates that a consistent and useful identifier for that region could not be found.

			Barcoding		FAFLPs				
Genus	Species	n	Genus	Species	*trn* T-*trn* L	*trn* L	*trn* L-*trn* F	Unique?	*trn* L Alone
Alnus	*alnobetula subsp. crispa*	4	n	n	x	602-603	464	n	n
Alnus	*incana subsp. tenuifolia*	4	n	n	x	603-605	464	n	n
Betula	*glandulosa*	2	y	n	v	440	477	n	n
Betula	*occidentalis*	2	y	n	1042-1043/1033	440	446	y	n
Betula	*papyrifera*	4	y	n	1043	440	475-476	n	n
Betula	*pumila*	2	y	n	x	440	476*	n	n
Carex	*aurea*	4	y	n	v	686	437	y	y
Carex	*crawfordii*	3	y	y	623-625	679	x	y	y
Carex	*disperma*	3	y	y	627-628/262/277/618-619	689	443	y	y
Cornus	*canadensis*	4	y	y	857	582-584	434	y	y
Cornus	*stolonifera*	3	y	y	x	590	423*	y	y
Fragaria	*vesca*	3	y	n	998	490	497	y	n
Fragaria	*virginiana*	3	y	n	x	490	428-430/394	y	n
Picea	*glauca*	4	y	y	470/461	559-560	460-465	n	n
Picea	*mariana*	4	y	n	469/460	559-560	460	n	n
Populus	*balsamifera*	2	y	n	x	653	399-403	y	y
Populus	*tremuloides*	2	y	y	525-526	693-695	391-392	y	y
Ribes	*glandulosum*	3	y	y	1103/1193-1194	586	x	y	n
Ribes	*hirtellum*	2	y	n	1105/1096	586	x	y	n
Ribes	*lacustre*	3	y	n	1128/1119	585	411	y	n
Ribes	*oxyacanthoides*	3	y	n	1117-1119/1108-1109/1127	319-320	x	y	y
Ribes	*triste*	4	y	y	1109-1110	580	411	y	y
Rosa	*acicularis*	4	y	n	x	616-618	482	n	n
Rosa	*woodsii*	4	y	n	525-526	617	482	n	n

Chapter 3: Baseline DNA quality and quantity of roots across species and size classes

3.1 Introduction

Molecular techniques have helped us capture belowground diversity that is missed by aboveground surveys (Hiiesalu et al. 2012), understand the function of deep roots (Maeght et al. 2013), and describe communities of ancient plants (Sønstebø et al. 2010), however, these techniques are not without issues. False positives and negatives can occur due to low resolution of molecular markers, polymerase chain reaction (PCR) inhibitors, as well as dead or dormant roots persisting in the soil, all of which can cause biases by misrepresenting a belowground community.

Dead roots have long been recognized as a potential bias in assessing living belowground diversity with molecular methods (Hiiesalu et al. 2012). Kesanakurti et al. (2011) found short root fragments (< 2 cm in length) amplified poorly and speculated that this was due to some short fragments being dead. This problem of 'false positives' is exacerbated if the environment of interest contains salvaged soil placed in a new location, which is a common practice used in restoration and reclamation (Macdonald et al. 2015, Wubs et al. 2016), because salvaged soils may be laden with dead roots from the source plant community. If research questions focus on the role living roots play in communities or ecosystems, we must be able to distinguish the alive roots from the dead, or, determine when dead roots no longer contribute DNA to sampled soils.

3.1.1 Root decay

How DNA decays *in situ* is relatively unknown, but how roots decay has been well explored (Zhang & Wang 2015). In particular, root chemistry, environmental factors, and root

size are drivers of root decay rates. In some systems, roots can retain most of their mass after two years (Xiong et al., 2013) and after four years, depending on species and size class, most temperate tree roots retain half of their mass (Sun et al. 2013). Milchunas (2009) divided roots into short-lived and long-lived categories and calculated a turnover rate of five and ten years, respectively. Root order or size class may affect whether a root is long- or short-lived. In general, root orders are better measures of root function than diameter cutoffs (e.g., fine roots: < 2 mm, medium roots: 2-5 mm, large roots: > 5 mm), and measurements made on entire root systems can capture critical functions missed by root orders (Freschet et al. 2017). However, size classes based on diameter cutoffs are commonly used and easy to implement and thus are used in this study. When determining size class, greater than 2 mm and less than 2 mm are most relevant categories (Hobbie et al. 2010).

Root chemistry strongly influences root decay rates and is intricately linked to root size class. Some studies have suggested that larger diameter classes decompose more slowly (Silver & Miya 2001), while other studies suggest that fine roots decay more slowly (Xiong et al. 2013) because of complex C molecules in very fine roots (Sun et al. 2013). Goebel et al. (2011) found that fine roots decay more slowly, but root N is retained for longer in higher order roots. Hobbie et al. (2010) found that in fine roots, N immobilization and decomposition rates did not mirror those in leaves due to different tissue traits and N dynamics that influence decomposition. Silver and Miya (2001) found that root chemistry appears to be the primary controller of root decomposition, but that environmental factors and soil texture, can strongly influence the rate of decay, for example, decay rates are higher in sandy loam or clay loam soils. Important to the boreal forest, colonization by ectomycorrhizal fungi may slow root

decomposition rates (Langely et al. 2006), or ectomycorrhizal fungi may compete with saprotrophic fungi inhibiting their breakdown of roots (Brzostek et al. 2015). Taken together, I expect that fine roots may lose mass and complex C molecules more slowly than coarse roots, but it is unknown if DNA degradation will follow similar patterns to root decay.

3.1.2 DNA degradation

Environmental damage to DNA that causes failure of amplification using PCR-based methods can come from two sources. The DNA molecule can degrade by physically breaking down, or the sample can become inundated with PCR inhibitors. Understanding the timeline of these processes is key to addressing false positives associated with dead roots.

In general, PCR inhibitors are usually concentration dependent, interfere more with smaller amplicons, and can be sequence-dependent. They can interfere with the polymerase enzyme, bind DNA, or both. One inhibitor of concern is humic acid from soil which binds to DNA effectively reducing the concentration of the DNA template (McCord et al. 2011). This inhibitor may have sequence-specific effects and can affect both large and small amplicons in human DNA (McCord et al. 2011). Other inhibitors, such as the variety of secondary metabolites created by the roots themselves, may inhibit amplification in ways that are highly species-specific. For example, polyphenols such as flavonoids and tannins can affect enzyme activity (Gegengeimer 1990) or bind to DNA giving it a brown color (Katterman & Shattuck 1983). Other examples of inhibitors likely to be found in roots are acid polysaccharides, which likely inhibit the polymerase enzyme (Pandey et al. 1996). Importantly, the presence and composition of root secondary metabolites vary across species (Rasmann & Agrawal 2008, van Dam 2009). Overall, a variety of secondary compounds are likely to be found in roots across lineages and

PCR inhibition is likely to depend on phylogenetic relationships (Senior et al. 2016) as well as root size and function (van Dam 2009, Poirier et al. 2018).

In the absence of repair processes maintained by a living cell, the DNA molecule degrades due to a variety of lesions which interfere with DNA recovery. Lesions can be caused by enzymatic degradation, such as from lysosomal nucleases in the postmortem cell, chemical oxidation, radiative crosslinking of purines, hydrolytic lesions which cause a change in the molecular structure of nucleic acids, or strand breaks caused by microbial decomposition or other damage (Pääbo et al. 2004). Ultraviolet radiation commonly produces cyclobutane pyrimidine dimers as well as a variety of other DNA lesions such as single and double strand breaks (Jiang et al. 2007). Kumar et al. (2004) found that longer fragments of DNA would not be amplified after UV-degradation owing to a higher probability of experiencing a degradation event. McCord et al. (2011) found that chemical oxidation does not seem to be a major factor in environmental damage to forensic DNA samples, but little else is known about the relative importance of these factors to samples degrading in the environment. Overall, DNA degradation is less likely to depend on species and root size, and more likely do depend on time and environmental conditions.

Despite these many factors acting to degrade DNA, well-preserved DNA can potentially persist in the environment for up to 1.5 million years (Willerslev et al. 2004). In plants, the P6 loop of the *trn*L intron has been used to characterize ancient plant communities (Sønstebø et al. 2010) because of its highly stable secondary structure, its short size, and its ability to be amplified by universal primers (Taberlet et al. 2007). Even poorly preserved DNA can persist in the environment for a long time, for example, extracellular DNA from fungi and other microbes

can persist in soil for weeks to years (Carini et al. 2016). This environmental DNA (eDNA) is commonly used for biodiversity monitoring and conservation (Thomsen & Willerslev 2015). Overall, there has been much progress made in amplifying highly degraded DNA from dead material, but little progress in recognizing and excluding degraded DNA from belowground surveys.

Degradation of DNA and most forms of PCR inhibition are likely to have a more pronounced effect on amplification of a sample when the target region is longer, when environmental conditions are less favorable, and when the sample has been degrading for a longer period of time. In addition, certain PCR inhibitors may be highly dependent on species and size class of roots. The effects of damage to DNA on amplification can be visualized and quantified by gel electrophoresis (among other methods) and are characterized by allele dropout of larger regions and lower intensities of fluorescent bands (Sutherland et al. 2003). These factors could potentially be used to determine if a sample is from a dead root, but not without knowing how the baseline quality and quantity of DNA changes by species and size class.

In this chapter, I describe the installation of a five-year field experiment to measure DNA degradation with time, and predict the origin of false positives, which may interfere with the methodology for species identification described in Chapter 2. However, the current chapter will focus on only the baseline measurements of DNA quality and quantity across different root size classes and species.

3.2 Methods

3.2.1 Setup of field experiment

I set up a field experiment to test the roles of species, size class, and time on DNA degradation in roots. The experiment is located in the southeast corner of the University of Alberta Botanic Gardens. The climate of this region is characterized by 2.2 ± 1.1°C mean annual temperature and 451 ± 78 (SD) mm annual precipitation. The forest in the Botanic Garden is characterized by sandy soil with a rolling topography covered by a mixed woodland of *Picea glauca* (Moench) Voss (white spruce), *Pinus banksiana* Lambert (Jack pine), and *Populus tremuloides* Michaux (aspen). On June 21, 2017, I installed a 2.85 m by 2.85 m plot (marked by PVC at the northwest corner; 53°23'59.45"N and 113°45'19.40"W, NAD 83) on flat grassy terrain. Other plant species within the plot include: *Maianthemum canadense* Desfontaines, *Linnea borealis* Linnaeus, *Rosa acicularis* Lindley, *Galium boreale* Linnaeus, *Aralia nudicaulis* Linnaeus, *Arctostapholos uva-urs* (Linnaeus) Sprengeli, *Vaccinium mytilloides* Michaux, *Vaccinium vitus-idaea* Linnaeus, *Corylus cornuta* Marshall, *Lathyrus venosus* Hooker, *Achillea millefolium* Linnaeus, *Vicia Americana* Muhlenberg ex Willdenow, *Bromus inermis* Leysser, and *Poa pratensis* Linnaeus. I removed large (> 3 cm) aboveground woody vegetation and debris from the plot.

Using twine, I marked a grid consisting of 361 15 x 15 cm squares comprising the plot. In the southern portion of each gridded square, I cut a 15 cm slit into the soil just deep enough that a mesh bag containing roots (see below) could be inserted into the slit flush with the soil surface. Mesh bags were positioned randomly (see Appendix III).

3.2.2 Collection and preparation of roots

In June 2017, I collected roots from 25 individuals of eight species: *Cornus canadensis, Rosa acicularis, Populus tremuloides, Linnaea borealis, Viburnum edule, Picea glauca, Maianthemum canadense,* and *Poa pratensis*. Sampling locations included crown land west of Lakeland Provincial Park, Alberta (54°38'45.40"N, 111°33'13.88"W, NAD 83) and from the University of Alberta, Woodbend park, Alberta (53°23'41.49"N, 113°45'26.75"W, NAD 83). Individuals of a single species were located > 5 m apart. Roots were traced back to aboveground plant parts to insure proper identification, and voucher specimens were submitted to the University of Alberta Herbarium (ALTA). Excess soil was removed from roots in the field, and roots were stored in dry paper bags within plastic bags on ice for no more than three days. Roots were then rinsed with deionized water and categorized into three size classes: diameters < 2 mm (fine), between 2 and 5 mm (medium), and > 5 mm (large). Roots were air dried for 3-5 days, and for each individual, approximately 0.5 g of root tissue was sampled for each size class. Not all species had roots in all size classes. Root samples were put into labelled 5 x 5 cm fiberglass mesh bags (standard window screen, 2 mm mesh) and fastened with office staples. Mesh size of bags is important to consider for the large root size class (Scheu & Schauerman 1994). Size was chosen to be large enough for most soil fauna to enter, with the exception of some larger soil fauna such as earthworms (Bocock & Gilbert 1957). As this study was created to assess DNA degradation over 5 years, enough samples were installed to allow for yearly harvests. Four hundred and fifty samples were prepared from 25 individuals (5 replicates x 5 sequential harvests) x 8 species, where size class of each individual was

sampled, if present. For the baseline measurements, an additional 5 individuals of each species were collected and separated into relevant size classes for a total of 75 samples.

3.2.3 Baseline Measurements of root DNA quality and quantity

Total genomic DNA was extracted from 20 mg of lyophilized and finely tissue-lysed root material from 75 single species samples using the DNeasy PowerPlant Pro Kit (Qiagen Inc, Hilden, Germany). DNA template was amplified, and subsequently measured for quality (260/280 nm and 260/230 nm absorbance ratios), and extraction efficiency (quantity (ng μL^{-1}) divided by dry weight prior to extraction (mg)) using a Nanodrop 2000 (Thermo Fisher Scientific, Wilmington, DE, USA). Light absorbed at 260/280 and 260/230 nm indicates presence of protein or phenol, and carbohydrates or phenol, respectively. Methods for sample lyophilization, pulverization and amplification are described in Chapter 2.

3.2.4 Data analysis

There were eight species with roots in the fine root class (*Cornus canadensis, Rosa acicularis, Populus tremuloides, Linnaea borealis, Viburnum edule, Picea glauca, Maianthemum canadense,* and *Poa pratensis*) and four species with roots across all three size classes (*Rosa acicularis, Populus tremuloides, Viburnum edule, and Picea glauca*). To test if species differed in two measures of DNA quality and extraction efficiency of fine roots, separate one-way analyses of variance (ANOVA) were used with a type III sum of squares, to account for unequal sample size. To test if the same three measures of DNA differed by species, root size class, or their interaction, a two-way ANOVA was conducted on the subset of four species. Post-hoc comparisons were done with a Tukeys HSD test and an alpha of $p < 0.05$ was used in all tests.

3.3 Results

Since fine roots could be collected for all species in the experiment, this size class was used to compare DNA across species. The two measures of DNA quality differed among species. Quality differed by species for both absorbance ratios (260/230 nm: $F(7,32) = 10.39$, $p < 0.001$, and 260/280 nm: $F(7,32) = 7.67$, $p < 0.001$) (Table 1). Post hoc Tukey HSD test showed that *Poa pratensis* had the highest 260/230 nm absorbance ratio, and *Cornus canadensis* had significantly lower absorbance ratios than the rest of the species, with the exception of *Linnaea borealis* and *Viburnum edule* (Figure 3). Trends were similar for DNA quality measured by 260/280 nm absorbance ratios. Though *Poa pratensis* also had the highest 260/280 nm absorbance ratio (mean: 1.72 ± 0.06 SE), it was not significantly different than the other species (data not shown). *Cornus canadensis* had low absorbance levels at 260/280 nm (mean = 1.06± 0.03 SE), and was similar to *Rosa acicularis* (data not shown). Species did not differ by extraction efficiency ($F(7,32) = 0.93$, $p = 0.49$).

Of the subset of species with roots in all size classes, the main effects of species and size class, but not their interaction, had a significant effect on 260/230 and 260/280 nm ratios (Table 2). Fine roots contained DNA of higher quality than other root classes for each of the two measures of quality and fine roots had a higher extraction efficiency. *Viburnum edule* had the lowest 260/230 nm absorbance ratio of the four species. *Picea glauca* had significantly higher 260/280 nm ratios than *Viburnum edule*, which was the lowest value. Extraction efficiency differed by root size class ($F(2,43) = 4.53$, $p = 0.02$). Fine roots had the highest extraction efficiency (3.02 ± 0.38 SE) followed by large and medium roots size classes.

3.4 Discussion

3.4.1 DNA quality

The two measurements of DNA quality presented above are measurements of secondary metabolites such as phenols, carbohydrates, or proteins, that are coextracted with nucleic acids, and may cause inhibition in PCR-based molecular methods. As expected, these potential inhibitors varied by species and root size class. The species with the highest DNA quality was a grass, *Poa pratensis*. Although more studies would be needed to confirm that this trend applies generally to grass species, grasses may have the highest potential to bias molecular studies aimed to detect living species belowground. Other species, such as *Cornus canadensis* and *Viburnum edule*, have very low DNA quality. All else being equal, we would expect DNA from Poa pratensis to be amplified with greater success than *Cornus canadensis* and *Viburnum edule.* Owing to differences in DNA quality, these latter species may be more likely to produce false negatives in belowground assays of living plant biodiversity. However, after death, they may be less likely to produce false positives in circumstances where soils have been salvaged and placed in a new environment.

Fine roots generally have higher quality DNA than other root size classes so DNA extracts from fine roots are less likely to be coextracted with PCR inhibitors. This combined with the knowledge that fine roots tend to last longer in soils (Sun et al. 2013, Milchunas et al. 2009) suggests that fine roots may have the highest potential to bias molecular studies aimed to detect living species belowground. In the boreal, many conifer fine roots are associated with ectomycorrhizal fungi which may lengthen their decomposition rates (Langely et al. 2006). This further emphasizes the point that dead fine roots may be a large potential bias (i.e., false

positive) in estimates of living below ground plant species, especially in the context of restoration and reclamation in the boreal forest.

3.4.2 DNA extraction efficiency

Extraction efficiency, or the concentration of DNA extracted standardized by the weight of the plant sample, is effectively a measure of DNA quantity. If DNA from a sample has a high extraction efficiency, it is more likely to be successfully amplified simply because it may have a proportionally higher amount of template. Extraction efficiency did not differ by species but did differ by size class. More DNA was extracted from fine roots than other root size classes probably because fine roots are younger (Goebel et al. 2011). The implication of fine roots having a higher extraction efficiency corroborates the idea presented above that fine roots may be the most likely to cause false positives. Since extraction efficiency does not vary by species, we can expect that species-specific effects on PCR come from the quality of DNA extracts and not from the quantity of DNA recovered.

3.4.3 Future directions

One important distinction to make is that neither extraction efficiency nor either measure of DNA quality directly measures DNA degradation, even though both will have implications for recovering degraded DNA. In fact, highly fragmented and altered DNA can show up as high quality, based on absorption ratios, as long as the extraction successfully removed other compounds. Future research, and future researchers associated with this long-term study, will have to use PCR-based or other methods to directly detect DNA degradation (Sutherland et al. 2003, Kumar et al. 2004, Deagle et al. 2006). Although we can detect that DNA has degraded, we do not yet know the dominant mechanism of degradation of DNA in soil.

Knowing the primary mechanism of DNA degradation in soil may help us to quantify degradation in dead roots. This will be complicated, however, by the presence of PCR inhibitors that vary widely by species. Cataloguing the different PCR inhibitors likely to be found for common species could help us predict problems associated with PCR inhibitors. However, this may be impractical for a large number of species. A more elegant solution would be to improve methods of removing these inhibitors from DNA template across a wide variety of species. Cleaner DNA would allow us to get past the species-specific effects of amplification success and we could more easily study DNA degradation. Still, future studies will be able to detect when degradation and PCR inhibition begin to interfere with molecular methodologies to assess belowground diversity. Using gradients of lengths of target regions, we can determine the time of PCR failure as long alleles drop out when there is significant inhibition or degradation (Sutherland et al. 2003, Kumar et al. 2004). Additionally, real-time quantitative PCR has been used as a tool to identify the mechanism of inhibition of several known PCR inhibitors (McCord et al. 2011) and the frequency of polymerase blocking DNA damage (Deagle et al. 2006). Being able to quantify this degradation and to see how PCR inhibition changes through time may lead future researchers to being able to distinguish between alive and dead roots belowground.

3.5 Figures

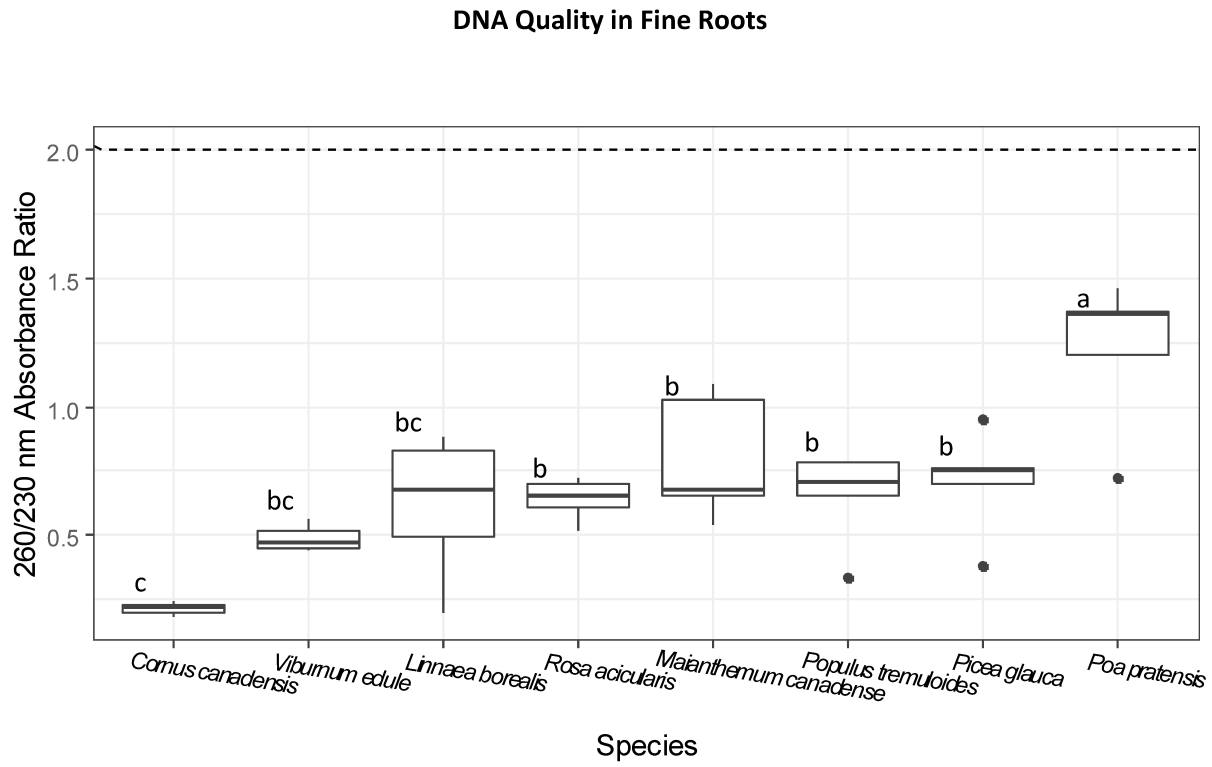


Figure 3. DNA quality across species as measured by the 260/230 nm absorbance ratio. Dotted line (2.0) indicates "pure" DNA. Boxplots that do not share a letter have means that significantly differ from one another.

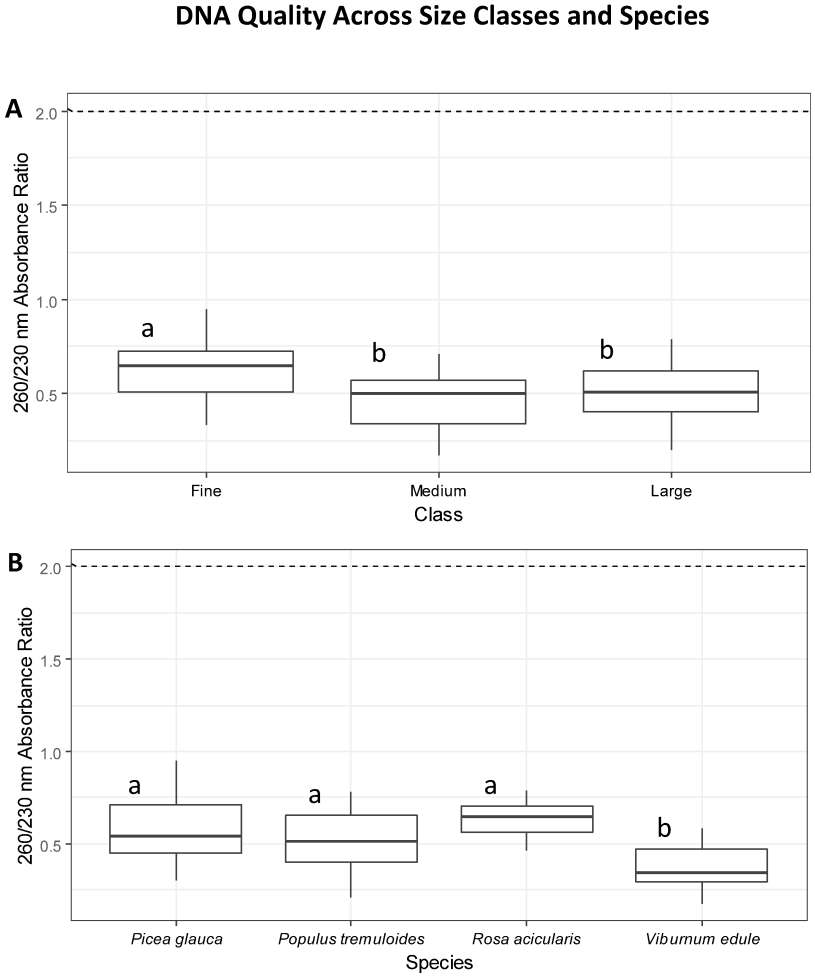


Figure 4. DNA quality across size class (A) and species (B) as measured by the 260/230 nm absorbance ratio. Dotted line (2.0) indicates "pure" DNA. Boxplots that do not share a letter have means that significantly differ from one another.

3.6 List of Tables

Table 5. ANOVA tables for the effect of plant species on measures of DNA quality, 260/280 absorbance ratios (a), and 260/230 absorbance ratios (b), and extraction efficiency (c).

ANOVA Tables for 3 Measures of DNA Quality

a.

	Sum of Squares	df	F value	Sig.
Species	1.53	7	7.67	0.00
Error	0.91	32		
Total	2.44	39		

b.

	Sum of Squares	df	F value	Sig.
Species	2.84	7	10.39	0.00
Error	1.25	32		
Total	4.09	39		

c.

	Sum of Squares	df	F value	Sig.
Species	40.26	7	0.93	0.49
Error	197.11	32		
Total	237.37	39		

Table 6. Two-way ANOVA tables for measures of DNA quality, 260/280 absorbance ratios (a), and 260/230 absorbance ratios (b), and extraction efficiency (c), across plant species and root size classes.

Two-Way ANOVA Tables for 3 Measures of DNA Quality

a.

	Sum of Squares	df	F value	Sig.
Species	0.49	3	10.71	0.00
Size Class	0.35	2	11.43	0.00
Interaction	0.17	6	1.86	0.11
Error	0.66	43		
Total	1.68	54		

b.

	Sum of Squares	df	F value	Sig.
Species	0.35	3	4.06	0.01
Size Class	0.27	2	4.69	0.01
Interaction	0.27	6	1.53	0.19
Error	1.25	43		
Total	2.14	54		

c.

	Sum of Squares	df	F value	Sig.
Species	6.23	3	1.64	0.19
Size Class	11.47	2	4.53	0.02
Interaction	8.77	6	1.16	0.35
Error	54.41	43		
Total	80.88	54		

Chapter 4: General Discussion

4.1 Research summary

The objective of this thesis is to build a size key of FAFLPs to identify roots in the boreal forest and to address some predicted limitations and biases. I expected that FAFLPs would be limited in their ability to identify closely related species and that belowground detection of roots would be influenced by species and root size class. I found that sequencing was not superior to FAFLPs in distinguishing closely related species. In addition, baseline DNA extract from roots varied in quality by species and size class, which is likely to bias belowground surveys of roots. I also encountered some unexpected limitations to FAFLPs, specifically, high levels of intraspecific variation of fragment length and lower resolving power of a dataset of this size. Still, unique FAFLPs were found for 74% of the 194 boreal species in this study which expands a dataset in the adjacent aspen parkland ecoregion (Taggart et al. 2011). Overall, I found that molecular tools can be effective, but each method has its strengths and limitations, and each could benefit from improved and standardized techniques as well as more data from similar studies to help us understand overarching mechanisms that apply to analyzing DNA from roots in soil.

I identified three important future directions that should be explored to improve our confidence in root identification methods: (1) optimization of DNA extractions to reduce species-specific effects from PCR inhibitors, (2) integration of intraspecific variation into FAFLP analysis, and (3) testing how environmental conditions affect DNA degradation in dead roots. These research topics should be explored before more advanced multiplexing of samples occurs.

4.2 Controlling species variation through better DNA extraction

DNA can fail to produce FAFLPs or sequences due to problems during DNA extraction, PCR amplification, capillary electrophoresis, and downstream data processing. For large species pools or multi-species samples, it is impractical to tailor primer sets, PCR conditions, or DNA extraction techniques to individual species. For this reason, I advocate for improvements in DNA extraction techniques widely applicable to roots across distantly related taxa to reduce species-specific affects from secondary metabolites that interfere with PCR.

Commercial DNA extraction kits such as DNeasy PowerPlant or PowerSoil kits (Qiagen Inc, Hilden, Germany) are designed to remove impurities such as polysaccharides in leaves or humic acids in soil, but neither kit is optimized for roots. Using both kits (the first optimized for plants and the second optimized for soil) would be expensive, and more importantly, would greatly reduce the concentration of DNA extracted. There are a variety of 'homebrew' extraction methods designed for roots (Linder et al. 2000, Brunner et al. 2001, Khan et al. 2007), but they are often optimized for a limited number of species. In my study, the allele recovery of the second extraction method (Griffiths et al. 2001) was much higher than the first (Roe et al. 2010), even though these samples were non-randomly chosen from a group that were unsuccessful using the 2% CTAB method. This finding suggests that more troubleshooting and optimization may produce significant advances in DNA extraction purity. A robust extraction method optimized for roots across all lineages would help improve the resolution of FAFLP data but also reduce the species-specific variation in DNA quality from degraded roots.

4.3 Accounting for intraspecific variation

Even if we can extract reasonable quantities of very clean DNA and remove inhibitor effects on a DNA sample, there are still other sources of variation that are likely to affect our confidence in DNA data. For example, we cannot control intraspecific variation, but future research should be able to account for and incorporate it in analysis. Fluorescent amplified fragment length polymorphisms in small datasets are often used to identify if a specimen is either present or absent. In large datasets, some FAFLP profiles overlap, especially when there is intraspecific variation. This makes data analysis complicated and we cannot always apply a single value to one region for a species because we do not yet know how intraspecific variation operates. Based on the FAFLP size key presented in Chapter 1, intraspecific variation will likely depend on species and there could be a phylogenetic signal in the fragment length sizes. In addition, variation likely increases with geographic distance, so variation may be dependent on whether a global or regionally specific database is implemented (Lamb et al. 2016). Since there are few studies related to FAFLPs on roots, these effects could be tested using electronic PCR (*in silico*) of publicly available chloroplast plastid sequences.

4.4 Environmental effects on DNA degradation

After the long-term experiment is completed we will have a better understanding of how root size class and species affect DNA degradation. The next most important factor is likely the environment (Silver & Maya 2001), which was held constant in the current study. Since there is a scarcity of data on root DNA degradation it will be hard to determine how environmental factors affect degradation without labor and time intensive field work. It is possible that some environmental factors could be controlled and accounted for in a

greenhouse. Alternatively, if DNA degradation is correlated with chemical or other measures of root degradation, we could use the many root degradation studies to inform how DNA may degrade *in situ*.

4.5 Challenges and limits with multiplexing

Originally, we hoped that FAFLPs would be a quick and affordable alternative to NGS for analyzing mixed species samples. Though FAFLPs remain a strong candidate, more optimization of FAFLPs needs to occur so that a single species can be identified correctly and confidently. It is not yet clear if we can distinguish dead roots from live roots, especially in a mixed-species pool, but this may be possible after the completion of the long-term root decay experiment. Another approach to quantify relative species abundance in mixed samples is using species-specific primers and quantitative real-time PCR (Mommer et al., 2008), though this approach as well as NGS have similar limitation to FAFLPs.

4.6 Applications

Applications for molecular methods for identification or quantification of degradation are wide and varied. With this tool, we can reexamine well-explored questions in belowground ecology related to primary production, biomass, and phenology, but answer these questions at the species level. Furthermore, we can further explore root competition, how roots interact with novel soil profiles, and how belowground diversity relates to the aboveground.

Other applications range from identifying the root that grew into your water pipe to judging the age of a forensic sample in a shallow grave. In ecology, these methods could be applicable to seed and bud bank studies. For example, we could predict what kind of weeds will establish in a reclamation site based on DNA identified seeds removed from soil, or, we could

predict if these seeds and buds are likely to be viable based on DNA degradation. Beyond plants, these methods, using different primer sets, could be applied to soil fauna, fungi, and bacteria to further illuminate the cryptic world belowground.

5 Literature

Abarenkov K., Somervuo P., Nilsson R.H., Kirk P.M., Huotari T., Abrego N. and Ovaskainen O. (2018). PROTAX-fungi: a web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences *New Phytologist* 220: 517–525doi: 10.1111/nph.15301

Anderson, I. C., & Cairney, J. W. G. (2004). Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Environmental Microbiology*, *6*(8), 769–779.

Algeo, T. J., Berner, R. A., Maynard, J. B., & Scheckler, S. E. (1995). Late Devonian oceanic anoxic events and biotic crises: "rooted" in the evolution of vascular land plants. *GSA Today*, *5*(3), 45–66.

Bardgett, Richard D., Liesje Mommer, and Franciska T. De Vries (2014) Going Underground: Root Traits as Drivers of Ecosystem Processes. Trends in Ecology & Evolution 29(12): 692–699.

Blume-Werry, G., Wilson, S. D., Kreyling, J., & Milbau, A. (2016). The hidden season: growing season is 50% longer below than above ground along an arctic elevation gradient. *New Phytologist*, *209*(3), 978–986.

Bobowski, B. R., Hole, D., Wolf, P. G., & Bryant, L. (1999). Identification of roots of woody species using polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) analysis. *Molecular Ecology*, *8*(3), 485–491.

Bockstette, S. W., Pinno, B. D., Dyck, M. F., & Landhäusser, S. M. (2017). Root competition, not

soil compaction, restricts access to soil resources for aspen on a reclaimed mine soil. *Botany*, *95*(7), 685–695.

Bocock, K. L., & Gilbert, O. J. W. (1957). The disappearance of leaf litter under different woodland conditions. *Plant and Soil*, *9*(2), 179–185.

Brouillet, L., F. Coursol, S.J. Meades, M. Favreau, M. Anions, P. Bélisle & P. Desmet. 2010+. VASCAN, the Database of Vascular Plants of Canada.

Brunner, I., Brodbeck, S., BÜchler, U., & Sperisen, C. (2001). Molecular identification of fine roots of trees from the Alps: reliable and fast DNA extraction and PCR–RFLP analyses of plastid DNA. *Molecular Ecology*, *10*(8), 2079–2087.

Brzostek, E. R., Dragoni, D., Brown, Z. A., & Phillips, R. P. (2015). Mycorrhizal type determines the magnitude and direction of root-induced changes in decomposition in a temperate forest. *New Phytologist*, *206*(4), 1274–1282.

Burgess, K. S., Fazekas, A. J., Kesanakurti, P. R., Graham, S. W., Husband, B. C., Newmaster, S. G., Barrett, S. C. H. (2011). Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. *Methods in Ecology and Evolution*, *2*(4), 333–340.

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S., & Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology*, *2*, 16242.

Chase, M. W., Cowan, R. S., Hollingsworth, P. M., van den Berg, C., Madriñán, S., Petersen, G., … Wilkinson, M. (2007). A Proposal for a Standardised Protocol to Barcode All Land Plants. *Taxon*, *56*(2), 295–299.

Cowan, R. S., Chase, M. W., Kress, W. J., & Savolainen, V. (2006). 300,000 Species to Identify:

Problems, Progress, and Prospects in DNA Barcoding of Land Plants. *Taxon*, *55*(3), 611–616.

Cronn, R. C., Small, R. L., Haselkorn, T., & Wendel, J. F. (2002). Rapid diversification of the cotton genus (Gossypium: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany*, *89*(4), 707–725.

Cutler, D. F. (1987). *Root identification manual of trees and shrubs: a guide to the anatomy of roots of trees and shrubs hardy in Britain and Northern Europe*. London: Chapman & Hall.

Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology*, *3*, 11.

DeSalle, R., Egan, M. G., & Siddall, M. (2005). The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1905–1916.

Endara María-José, Coley Phyllis D., Wiggins Natasha L., Forrister Dale L., Younkin Gordon C., Nicholls James A., … Kursar Thomas A. (2018). Chemocoding as an identification tool where morphological- and DNA-based methods fall short: Inga as a case study. *New Phytologist*, *218*(2), 847–858.

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Husband, B. C., Barrett, S. C. H. (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLOS ONE*, *3*(7), e2802.

Fazekas, A. J., Kesanakurti, P. R., Burgess, K. S., Percy, D. M., Graham, S. W., Barrett, S. C. H., …

Husband, B. C. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Molecular Ecology Resources*, *9*, 130–139.

Frank, D. A., Pontes, A. W., Maine, E. M., Caruana, J., Raina, R., Raina, S., & Fridley, J. D. (2010). Grassland root communities: species distributions and how they are linked to aboveground abundance. *Ecology*, *91*(11), 3201–3209.

Freschet Grégoire T., Roumet Catherine, & Treseder Kathleen. (2017). Sampling roots to capture plant and soil functions. *Functional Ecology*, *31*(8), 1506–1518.

Ganie, S. H., Upadhyay, P., Das, S., & Prasad Sharma, M. (2015). Authentication of medicinal plants by DNA markers. *Plant Gene*, *4*, 83–99.

Gegengeimer, P. (1990). Preparation of Extracts from Plants. In M. P. Deutscher (Ed.), *Methods in Enzymology* (Vol. 182, pp. 174–193). Academic Press.

Gielly, L., & Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular Biology and Evolution*, *11*(5), 769–777.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, *11*(5), 759–769.

Goebel, M., Hobbie, S. E., Bulaj, B., Zadworny, M., Archibald, D. D., Oleksyn, J., … Eissenstat, D. M. (2011). Decomposition of the finest root branching orders: linking belowground dynamics to fine-root function and structure. *Ecological Monographs*, *81*(1), 89–102.

Griffiths, R., Whiteley, A., O'Donnell, A., & Bailey, M. (2001). Rapid Method for Coextraction of DNA and RNA from Natural Environments for Analysis of Ribosomal DNA- and rRNA-Based Microbial Community Composition. *Applied and Environmental Microbiology*, *66*, 5488–5491.

Haling, R. E., Simpson, R. J., McKay, A. C., Hartley, D., Lambers, H., Ophel-Keller, K., … Richardson, A. E. (2011). Direct measurement of roots in soil for single and mixed species using a quantitative DNA-based method. *Plant and Soil*, *348*(1–2), 123–137.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1512), 313–321.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14812–14817.

Hiiesalu, Inga, Maarja Öpik, Madis Metsis, et al. (2012) Plant Species Richness Belowground: Higher Richness and New Patterns Revealed by next-Generation Sequencing. *Molecular Ecology* 21(8): 2004–2016.

Hiiesalu, Inga, Meelis Pärtel, John Davison, et al. (2014) Species Richness of Arbuscular Mycorrhizal Fungi: Associations with Grassland Plant Richness and Biomass. *New Phytologist* 203(1): 233–244.

Hobbie, Sarah E., Jacek Oleksyn, David M. Eissenstat, and Peter B. Reich. 2010. “Fine Root Decomposition Rates Do Not Mirror Those of Leaf Litter among Temperate Tree Species.” *Oecologia* 162 (2): 505–13.

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and Using a Plant DNA Barcode. *PLOS ONE*, *6*(5), e19254.

Hosein, F. N., Austin, N., Maharaj, S., Johnson, W., Rostant, L., Ramdass, A. C., & Rampersad, S.

N. (2017). Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. *Ecology and Evolution*.

Jackson, R. B., Moore, L. A., Hoffmann, W. A., Pockman, W. T., & Linder, C. R. (1999). Ecosystem rooting depth determined with caves and DNA. *Proceedings of the National Academy of Sciences*, *96*(20), 11387–11392.

Jiang, Y., Ke, C., Mieczkowski, P. A., & Marszalek, P. E. (2007). Detecting Ultraviolet Damage in Single DNA Molecules by Atomic Force Microscopy. *Biophysical Journal*, *93*(5), 1758–1767.

Katterman, F. R. H., & Shattuck, V. I. (1983). An Effective Method of DNA Isolation from the Mature Leaves of Gossypium Species That Contain Large Amounts of Phenolic Terpenoids and Tannins. *Preparative Biochemistry*, *13*(4), 347–359.

Karst, J., Pak Chow, & Landhäusser, S. M. (2015). Biases underlying species detection using fluorescent amplified-fragment length polymorphisms yielded from roots. *Plant Methods*, *11*(1), 1–12.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Kesanakurti, P. R., Fazekas, A. J., Burgess, K. S., Percy, D. M., Newmaster, S. G., Graham, S. W., … Husband, B. C. (2011). Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Molecular Ecology*, *20*(6), 1289–1302.

Khan, S., Qureshi, M. I., Alam, T., & Abdin, M. Z. (2007). Protocol for isolation of genomic DNA from dry and fresh roots of medicinal plants suitable for RAPD and restriction digestion. *African Journal of Biotechnology*, *6*(3), 175.

Kumar, A., Tyagi, M. B., & Jha, P. N. (2004). Evidences showing ultraviolet-B radiation-induced damage of DNA in cyanobacteria and its detection by PCR assay. *Biochemical and Biophysical Research Communications*, *318*(4), 1025–1030.

Lamb, E. G., Winsley, T., Piper, C. L., Freidrich, S. A., & Siciliano, S. D. (2016). A high-throughput belowground plant diversity assay using next-generation sequencing of the trnL intron. *Plant and Soil*, *404*(1–2), 361–372.

Langley, J., Chapman, S. K., & Hungate, B. A. (2006). Ectomycorrhizal colonization slows root decomposition: the post-mortem fungal legacy. *Ecology Letters*, *9*(8), 955–959.

Larson, J. A. 1980. *The Boreal Ecosystem*. New York: Academic Press.

Linder, C. R., Moore, L. A., & Jackson, R. B. (2000). A universal molecular method for identifying underground plant parts to species. *Molecular Ecology*, *9*(10), 1549–1559.

Macdonald, S. E., Landhäusser, S. M., Skousen, J., Franklin, J., Frouz, J., Hall, S., … Quideau, S. (2015). Forest restoration following surface mining disturbance: challenges and solutions. *New Forests*, *46*(5–6), 703–732.

Maeght, J., Rewald, B., & Pierret, A. (2013). How to study deep roots - and why it matters. *Frontiers in Plant Science*, *4*.

McCord, B., Opel, K., Funes, M., Zoppis, S., & Jantz, L. M. (2011). An Investigation of the Effect of DNA Degradation and Inhibition on PCR Amplification of Single Source and Mixed Forensic Samples.

McCormack, M. L., Adams, T. S., Smithwick, E. A. H., & Eissenstat, D. M. (2012). Predicting fine root lifespan from plant functional traits in temperate trees. *New Phytologist*, *195*(4), 823–831.

McNickle, G. G., Cahill, J. F., & Deyholos, M. K. (2008). A PCR-based method for the identification of the roots of 10 co-occurring grassland species in mesocosm experiments. *Botany*, *86*(5), 485–490.

Mengel, K., Kirkby, E. A., Kosegarten, H., & Appel, T. (2001). Plant Nutrients. In Principles of Plant Nutrition (pp. 1–13). Springer, Dordrecht.

Milchunas, D. G. (2009). Estimating Root Production: Comparison of 11 Methods in Shortgrass Steppe and Review of Biases. *Ecosystems*, *12*(8), 1381–1402.

Mokany, K., Raison, R. J., & Prokushkin, A. S. (2006). Critical analysis of root : shoot ratios in terrestrial biomes. *Global Change Biology*, *12*(1), 84–96.

Mommer, L., Wagemaker, C. a. M., De Kroon, H., & Ouborg, N. J. (2008). Unravelling below-ground plant distributions: a real-time polymerase chain reaction method for quantifying species proportions in mixed root samples. *Molecular Ecology Resources*, *8*(5), 947–953.

Mommer, L., Dumbrell, A. J., Wagemaker, C. (Niels) A. M., & Ouborg, N. J. (2011). Belowground DNA-based techniques: untangling the network of plant root interactions. *Plant and Soil*, *348*(1–2), 115.

Moore L.A., & Field C.B. (2005). A technique for identifying the roots of different species in mixed samples using nuclear ribosomal DNA. Journal of Vegetation Science, 16(1), 131–134.

Nepstad, D. C., Carvalho, C. R. de, Davidson, E. A., Jipp, P. H., Lefebvre, P. A., Negreiros, G. H., … Vieira, S. (1994). The role of deep roots in the hydrological and carbon cycles of Amazonian forests and pastures. *Nature*, *372*(6507), 666–669.

Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., … Hofreiter, M. (2004). Genetic Analyses from Ancient DNA. *Annual Review of Genetics*, *38*(1), 645–679.

Palmer, J. D. (1992). Mitochondrial DNA in Plant Systematics: Applications and Limitations. In *Molecular Systematics of Plants* (pp. 36–49). Springer, Boston, MA.

Pandey, R. N., Adams, R. P., & Flournoy, L. E. (1996). Inhibition of random amplified polymorphic DNAs (RAPDs) by plant polysaccharides. *Plant Molecular Biology Reporter*, *14*(1), 17–22.

Pärtel, M., Hiiesalu, I., Öpik, M., Wilson, S. D., & Fox, C. (2012). Below-ground plant species richness: new insights from DNA-based methods. *Functional Ecology*, *26*(4), 775–782.

Pickles, B. J., & Pither, J. (2014). Still scratching the surface: how much of the “black box” of soil ectomycorrhizal communities remains in the dark? *The New Phytologist*, *201*(4), 1101–1105.

Poirier, V., Roumet, C., & Munson, A. D. (2018). The root of the matter: Linking root traits and soil organic matter stabilization processes. *Soil Biology and Biochemistry*, *120*, 246–259.

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, *21*(8), 1931–1950.

Randall, M. J., Karst, J., Pec, G. J., Davis, C. S., Hall, J. C., & Cahill, J. F. (2014). A Molecular

Identification Protocol for Roots of Boreal Forest Tree Species. *Applications in Plant Sciences*, *2*(11), 1400069.

Rasmann, S., & Agrawal, A. A. (2008). In Defense of Roots: A Research Agenda for Studying Plant Resistance to Belowground Herbivory. *Plant Physiology*, *146*(3), 875–880.

Rasse, D. P., Rumpel, C., & Dignac, M.-F. (2005). Is soil carbon mostly root carbon? Mechanisms for a specific stabilisation. *Plant and Soil*, *269*(1–2), 341–356.

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355–364.

Rewald, B., Meinen, C., Trockenbrodt, M., Ephrath, J. E., & Rachmilevitch, S. (2012). Root taxa identification in plant mixtures – current techniques and future challenges. *Plant and Soil*, *359*(1–2), 165–182.

Ridgway, K. P., Duck, J. M., & Young, J. P. W. (2003). Identification of roots from grass swards using PCR-RFLP and FFLP of the plastid trn L (UAA) intron. *BMC Ecology*, *3*, 8.

Roe, A. D., Rice, A. V., Bromilow, S. E., Cooke, J. E. K., & Sperling, F. a. H. (2010). Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the mountain pine beetle. *Molecular Ecology Resources*, *10*(6), 946–959.

Ruess, R. W., Hendrick, R. L., Burton, A. J., Pregitzer, K. S., Sveinbjornssön, B., Allen, M. F., & Maurer, G. E. (2003). Coupling Fine Root Dynamics with Ecosystem Carbon Cycling in Black Spruce Forests of Interior Alaska. *Ecological Monographs*, *73*(4), 643–662.

Scheu, S., & Schauermann, J. (1994). Decomposition of roots and twigs: Effects of wood type (beech and ash), diameter, site of exposure and macrofauna exclusion. *Plant and Soil*, *163*(1), 13–24.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., … Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, *109*(16), 6241–6246.

Seberg, O., & Petersen, G. (2009). How Many Loci Does it Take to DNA Barcode a Crocus? *PLOS ONE*, *4*(2), e4598.

Senior, J. K., Potts, B. M., Davies, N. W., Wooliver, R. C., Schweitzer, J. A., Bailey, J. K., & O'Reilly-Wapstra, J. M. (2016). Phylogeny Explains Variation in The Root Chemistry of *Eucalyptus* Species. *Journal of Chemical Ecology*, *42*(10), 1086–1097.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., … Small, R. L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, *92*(1), 142–166.

Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, *21*(8), 1794–1805.

Silver, Whendee L., and Ryan K. Miya. 2001. "Global Patterns in Root Decomposition: Comparisons of Climate and Litter Quality Effects." *Oecologia* 129 (3): 407–19.

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., … Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, *10*(6), 1009–1018.

Stanton, N. L. (1988a). The Underground in Grasslands. *Annual Review of Ecology and Systematics*, *19*, 573–589.

Steinaker, D. F., & Wilson, S. D. (2008). Phenology of fine roots and leaves in forest and

grassland. *Journal of Ecology*, *96*(6), 1222–1229.

Steinaker, D. F., Wilson, S. D., & Peltzer, D. A. (2010). Asynchronicity in root and shoot phenology in grasses and woody plants. *Global Change Biology*, *16*(8), 2241–2251.

Stone, E. L., & Kalisz, P. J. (1991). On the maximum extent of tree roots. *Forest Ecology and Management*, *46*(1), 59–102.

Somervuo P., Koskela S., Pennanen J., Nilsson R.H., Ovaskainen O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics* 32: 2920–2927.

Somervuo P., Yu D., Xu C., Ji Y., Hultman J., Wirta H., Ovaskainen O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution* 8: 398–407.

Sun, T., Mao, Z., & Han, Y. (2013). Slow decomposition of very fine roots and some factors controlling the process: a 4-year experiment in four temperate tree species. *Plant and Soil*, *372*(1–2), 445–458.

Sutherland, B. M., Georgakilas, A. G., Bennett, P. V., Laval, J., & Sutherland, J. C. (2003). Quantifying clustered DNA damage induction and repair by gel electrophoresis, electronic imaging and number average length analysis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, *531*(1), 93–107.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14.

Taberlet, P., Gielly, L., Pautou, G., & Bouvet, J. (1991). Universal primers for amplification of

three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, *17*(5), 1105–1109.

Taggart, J. M., Cahill Jr, J. F., McNickle, G. G., & Hall, J. C. (2011). Molecular identification of roots from a grassland community using size differences in fluorescently labelled PCR amplicons of three cpDNA regions. *Molecular Ecology Resources*, *11*(1), 185–195.

Terrestrial Subgroup Cumulative Environmental Management Association. 2010. Guidelines for reclamation to forest vegetation in the Athabasca oil sands region. CEMA: Alberta Environment, Province of Alberta

Tiedje, J. M., Asuming-Brempong, S., Nüsslein, K., Marsh, T. L., & Flynn, S. J. (1999). Opening the Black Box of Soil Microbial Diversity. *ResearchGate*, *13*(2), 109–122.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18.

Valentini, A., Miquel, C., Nawaz, M. A., Bellemain, E., Coissac, E., Pompanon, F., … Taberlet, P. (2009). New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources*, *9*(1), 51–60.

van Dam, N. M. (2009). Belowground Herbivory and Plant Defenses. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 373–391.

Willerslev, E., Hansen, A. J., Rønn, R., Brand, T. B., Barnes, I., Wiuf, C., … Cooper, A. (2004). Long-term persistence of bacterial DNA. *Current Biology*, *14*(1), R9–R10.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486), 47–51.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, *87*(12), 4576–4579.

Wubs, E. R. J., Putten, W. H. van der, Bosch, M., & Bezemer, T. M. (2016). Soil inoculation steers restoration of terrestrial ecosystems. *Nature Plants*, *2*(8), 16107.

Xiong, Y., Fan, P., Fu, S., Zeng, H., & Guo, D. (2013). Slow decomposition and limited nitrogen release by lower order roots in eight Chinese temperate and subtropical trees. Plant and Soil, 363(1–2), 19–31.

Yu, Z., Li, Q., Kong, L., & Yu, H. (2015). Utility of DNA Barcoding for Tellinoidea: A Comparison of Distance, Coalescent and Character-based Methods on Multiple Genes. *Marine Biotechnology*, *17*(1), 55–65.

Zhang, J., Nielsen, S. E., Grainger, T. N., Kohler, M., Chipchar, T., & Farr, D. R. (2014). Sampling Plant Diversity and Rarity at Landscape Scales: Importance of Sampling Time in Species Detectability. *PLoS ONE*, *9*(4), 1–11.

Zhang, X., & Wang, W. (2015). The decomposition of fine and coarse roots: their global patterns and controlling factors. *Scientific Reports*, *5*(1).

6 Appendix

I. DNA extraction by 2% CTAB

Procedures follow the steps from Cahill Lab (modified from Roe et al. 2010)

1. **Equipment and supplies**

- CTAB (cetyltrimethyl ammonium bromide) = Sigma 52365
- NaCl = Sigma S3014 (cheaper alternative = Fisher S271)
- EDTA = Sigma E9884 (cheaper alternative = Sigma EDS)
- Tris = Sigma T4661
- NaOH = Sigma S5881
- HCl = Sigma 320331
- PVP 40,000 = Sigma PVP40
- Spermidine = Sigma 85558 (cheaper alternative = Sigma S2626)
- B-mercaptoethanol = Sigma M3148
- Proteinase K = Sigma P2308 (better to use Qiagen 19131)
- Chloroform : isoamyl alcohol 24:1 = Sigma C0549
- Isopropanol = Fisher BP26181
- Ethanol 98% (v/v) = Sigma 24194
- mL safe-lock microcentrifuge tube, PCR clean = Fisher 05-402-95

2. **Chemicals and solutions**

Check the MSDS before using any chemicals. Protective clothing (gloves, glasses and lab coat) should be worn throughout the analytical process and WHMIS Material Safety Data Sheets for chemicals should be consulted prior to their use. All waste must be disposed properly according to chemical disposal guidelines.

2.1 **2% CTAB (w/v) buffer (good for months at room temperature)**

Dissolve 20 g of CTAB (cetyltrimethyl ammonium bromide) in 500 mL of deionized water (dH20) first. It forms clumps, needs heat up to ~50 °C on the stir plate to dissolve. The solution is foamy. The solution may crystalize after cooling. To re-dissolve, heat it up to 50 °C and stir. Then add:

280 mL 5 M NaCl

40 mL 0.5 M EDTA (pH 8.0)

100 mL 1 M Tris-HCl (pH 8.0)

Bring total volume to 1 L with dH20

2.2 **5 M NaCl**

292.2 g of NaCl (sprinkle to dH20 slowly on a stir plate)

700 mL dH20

Takes time to dissolve (don't add NaCl all at once, it will never go into solution), bring to 1 L

2.3 **7.5M NaOH**

Dissolve 45.0 g of NaOH pellets in 100 mL of dH20, on a stir plate

After cooled down to room temperature, bring total volume to 150 mL with dH20

2.4 **0.5 M EDTA in NaOH (pH 8.0)**

To a 250 mL Erlenmeyer flask containing 110 mL of dH20, add 40 mL of 7.5 M NaOH, set on a stir plate and stir

Add 29.22 g of EDTA very slowly by sprinkling into flask (via. a funnel) while stirring

While the EDTA is dissolving, adjust to pH 8.0 by adding slowly 7.5 M NaOH (~5 mL). EDTA won't dissolve until the pH is near 8.0. Add the NaOH in drops when approaching the end, because the pH end point is sharp.

Top up to 200 mL with dH20

2.5 **1 M Tris-HCl pH 8.0**

To a 250 mL Erlenmeyer flask containing 130 mL of dH20, add and dissolve 24.22 g of Tris by sprinkle slowly into flask, on a stir plate

Bring pH down to 8.0 by adding concentrated HCl (~17-18 mL)

Bring total volume to 200 mL with dH20

2.6 **1 M spermidine solution**

Spermidine is in liquid form at room temperature, but turns into solid in cold.

Warm up spermidine to 25°C. Mix 0.235 mL of spermidine with 1.265 mL of dH20 in a 2 mL vial. Invert several times to mix.

Keep the solution in the freezer for long term storage, or in the fridge for short term storage (weeks).

2.7 **TE buffer**

10 mL 1 M Tris-HCl pH 8.0

2 mL 0.5 M EDTA

Bring total volume to 1L with dH20

3. **Procedure**

1. Prepare CTAB buffer and all other solutions (see the reagents). Keep isopropanol and 70% ethanol in freezer.
2. Turn block heater to 65°C. Dispense the required amount of CTAB buffer solution in a flask which contains the required amount of polyvinylpyrrolidone (PVP) as shown in the table below, warm up to 65°C. After the PVP is completely dissolved, add the required amount of spermidine under the fumehood. The solution is good for 1-2 days. Prior to starting extraction, add β-mercaptoethanol under the fumehood and mix well. This is the CTAB extraction solution.

CTAB buffer	PVP 40,000 (final concentration 2 % (w/v))	1M spermidine solution (final concentration 0.5 g/L)	β-mercaptoethanol (final concentration 4 % (v/v))
5 mL	0.1 g	17.2 µL	200 µL
25 mL	0.5 g	86.0 µL	1000 µL

3. Under the fumehood, add 700 µL of CTAB extraction solution to the sample (20 mg of roots or foliage material, 50 mg of woody material, or 1 g of soil material, freeze-dried and ball milled to powder (1:1 ratio with glass powder for soil samples) contained in a 2.0 mL safe-lock microcentrifuge tube. Vortex until all solids are broken up and suspended.
4. Under the fumehood, add 10 µL Proteinase K (Qiagen, 600 mAU/mL) to each sample tube.
5. Incubate tube(s) at 65°C for 1 hr. Vortex every 15 min. to suspend the solids.
6. Cool samples to room temperature and under the fumehood add 600 µL of 24:1 chloroform:isoamyl alcohol to each tube. Vortex for 30 seconds. Centrifuge for 10 min. at 17,000 g at room temperature. Four layers are formed.

7. Set a pipet at 500 µL, carefully transferred the top brown layer (absolutely no solid materials) to a new 2.0 mL safe-lock microcentrifuge tube. Add 600 µL of ice-cold isopropanol. Mix briefly by repeated inversion (10 times) and place in a -20 °C freezer for 2 hr.
8. Centrifuge at 17,000 g at room temperature for 15 min.
9. Supernatant(s) are decanted to waste. Immerse pellet(s) in 500 µL of 70% ice cold ethanol. Vortex briefly and re-centrifuge at 17,000 g at room temperature for 5 min.
10. Supernatant(s) are carefully decanted to waste. Excess ethanol is pulled off using pipette. The pellet(s) are dried (~10 min) upside down on Kimwipe. If needed, continue to dry upright covered by a Kimwipe (~20-30 min). Alternatively, the pellet(s) can be dried in a Speedvac for 20 min.
11. Re-dissolve in either 50 µL nuclease-free water (for short-term stability (months in the fridge)) or 1x TE buffer (for long-term stability). Incubate at room temperature for 10 min. and mix with gentle agitation.
12. Optional: Fully re-dissolved DNA can be placed on ice, characterized spectrophotometrically and frozen (-20°C) for future use. (Typically, yields of 500+ ng/µL can be achieved with A260/280 ratios varying widely depending on species).

Warning - Chemical safety

- Read the MSDS before use
- CTAB: very toxic to aquatic life and human, skin irritation, severe eye damage, damage to organs if swallowed or inhaled.
- Spermidine: very toxic, severe damage to skin, eye & respiratory system, severe damage to organs if swallowed or inhaled.

- Mercaptoethanol: very toxic to human, severe damage to organs if swallowed, fatal, avoid inhalation.
- Chloroform:isoamyl alcohol: very toxic to human, severe damage to organs if swallowed, carcinogenic, avoid inhalation.
- Proteinase: skin and eye irritation.
- Fumehood must be used for this procedure.
- All waste must be contained for special disposal
- PPE (personal protective equipment) must be worn at all times, including gloves, safety glasses and lab coat.

G. Pec 2012 Notes on CTAB

CTAB – detergent which disrupts cell wall and biological membranes and at the same time denaturing or inhibiting proteins

5 M NaCl – provides positive ions to neutralize charges in nucleic acids bringing nucleic acid molecules together

0.5 M EDTA – chelates divalent metal ions and prevents magnesium-mediated aggregation of nucleic acids with proteins

1 M Tris-HCl pH 8.0 – provides proper pH for nucleic acid extraction. Helps cell membranes become more permeable

2% PVP – precipitates polyphenolics and removes them from extract

0.5 g/L spermidine – binds and precipitates nucleic acids

4 % β-mercaptoethanol – reducing agent that breaks down intramolecular protein disulfide bonds improving denaturing of proteins

Proteinase K – used to digest protein and remove contamination that might otherwise degrade DNA or RNA during purification

II. DNA extraction by 5% CTAB

Slightly modified from Griffiths et al. (2000).

1. **Equipment and supplies:**

- Fume hood
- Pipettes, 1-10 µL, 10-100 µL and 100-1000 µL
- Vortex mixer, with speed control and adapter for 2 mL microcentrifuge tubes
- Refrigerated microcentrifuge, with rotor for 2 mL tubes, speed up to 18,000 g
- 4°C fridge and -20°C freezer
- pH meter
- Stir plate and stir bars
- Speedvac
- Nanodrop
- 2 mL microcentrifuge tubes, with safe lock (e.g. Eppendorf 022363344). For alternatives, make sure the caps can make a tight seal, otherwise, it will leak during extraction.

2. **Chemicals and solutions**

Check the MSDS before using any chemicals. Protective clothing (gloves, glasses and lab coat) should be worn throughout the analytical process and WHMIS Material Safety Data Sheets for chemicals should be consulted prior to their use. All waste must be disposed properly according to chemical disposal guidelines.

2.1 **Deionized water (dH2O)**

- Deionized water, grade Type 2, conductivity at 25°C < 1 µS/cm
- Filter through 0.22 µm to remove microorganisms
- Nuclease-free and Rnase-free

2.2 **5 M NaCl (sodium chloride) solution**

- Dissolve 146.1 g of NaCl (e.g. Fisher S271) in 350 mL of dH2O using a 1L Erlenmeyer flask. Do it by sprinkling the NaCl to dH2O slowly on a stir plate. It takes time to dissolve, so don't add NaCl all at once, otherwise it will never go into solution. To accelerate, more dH2O can be added while stirring, but keep the total volume below 450 mL. After all dissolved, transfer the solution to a measuring cylinder and bring the volume to 500 mL with dH2O.

Storage: In a bottle at room temperature for years.

2.3 **Solution A – 5% CTAB buffer**

A.1 **CTAB Solution (10%)**

- In a 250 mL Erlenmeyer flask with a stir bar, add 100 mL of dH2O and 28 mL of 5M NaCl solution. Add 20.0 g of CTAB (cetyltrimethylammonium bromide = hexadecyltrimethylammonium bromide, e.g. Sigma 52365) slowly, in little amount at a time while stirring. To accelerate, more dH2O can be added while stirring, but keep the total volume below 180 mL. The solution is very thick. After all dissolved, transfer the solution to a measuring cylinder and bring the volume to 200 mL with dH2O. This is Solution A.1.

Storage: In a bottle at room temperature for years. Keep a stir bar inside the bottle. The solution may recrystallize after long storage. Stir the solution on a stir plate will re-dissolve the crystals.

A.2 **240 mM potassium phosphate buffer pH 8**

- 1 M K2HPO4 solution: Dissolve 34.8 g of K2HPO4 (potassium phosphate dibasic, e.g. Fisher P288) in 120 mL of dH2O. Bring the volume to 200 mL with dH2O.
- 1 M KH2PO4 solution: Dissolve 27.2 g of KH2PO4 (potassium phosphate monobasic, e.g. Fisher P285) in 120 mL of dH2O. Bring the volume to 200 mL with dH2O.

- Mix 188 mL of 1 M K2HPO4 solution with 10 mL of 1 M KH2PO4 solution. Adjust to pH 8 by adding approximately 2 mL of the KH2PO4 solution. This is a 1 M potassium phosphate buffer pH 8.
- Storage: In a bottle at room temperature for years.
- Mix 48 mL of 1 M potassium phosphate buffer with 152 mL of dH2O. This is the 240 mM potassium phosphate buffer pH 8, Solution A.2.
- Storage: In a bottle at room temperature for years.
- Solution A:
- Mix equal volumes of solution A.1 and A.2. This is Solution A.
- Storage: In a bottle at room temperature for years.
- Before use, under the fume hood, add 10 µL of β-mercaptoethanol (e.g. Sigma M3148) to 1 mL of Solution A. This can be saved overnight at room temperature prior to use.

2.4 **Solution B – Phenol : chloroform : isoamyl alcohol (25:24:1) (e.g. Fisher BP17521-400)**

- Handle only under the fume hood
- Need to be equilibrated with the supplied Tris buffer before use.
- Under the fume hood, pour the whole bottle of the supplied Tris into the bottle of phenol-chloroform-isoamyl alcohol, invert several time to mix.
- Let sit in the fridge overnight. It separates into two phases; the upper phase is the Tris buffer, the lower phase the phenol-chloroform-isoamyl alcohol.
- To use, transfer the lower phase to a separate brown bottle by aspiration. Handle only under the fume hood.
- Storage: In a brown bottle at 4°C for years.

2.5 **Solution C – Chloroform : isoamyl alcohol (24:1) (e.g. Sigma C0549)**

- Handle only under the fume hood

2.6 **Solution D – 30% PEG solution in 1.6 M NaCl**

- In a 250 mL Erlenmeyer flask with a stir bar, add 50 mL of dH2O and 64 mL of 5M NaCl solution. Add 60.0 g of PEG (polyethylene glycol 8000, e.g. Fisher BP233) slowly, in little amount at a time while stirring. To accelerate, more dH2O can be added while stirring, but keep the total volume below 180 mL. The solution is very thick. After all dissolved, transfer the solution to a measuring cylinder and bring the volume to 200 mL with dH2O. This is Solution D.
- Storage: In a bottle at room temperature for years. Keep a stir bar inside the bottle. The solution may recrystallize after long storage. Stir the solution on a stir plate will re-dissolve the crystals.

2.7 **Solution E – 70% ethanol**

- Ice cold, keep in the -20°C freezer.

3. **Procedure**

1. Put about 0.2 g of dried ground mineral soil (or 0.05 g of dried ground peat soil) in 2 mL safe-lock microcentrifuge tube
2. Set centrifuge at 4°C
3. (Under the fume hood) Add 0.5 mL of Solution A + 0.5 mL of Solution B
4. Homogenize by vortex at 2000 rpm for 30 s at room temperature
5. Centrifuge at 16,000 g for 5 min at 4°C
6. (Under the fume hood) Use a 100 µL pipet, transfer 400 µL of the upper phase to a new 2 mL tube, add 400 µL of Solution C, mix well by vortex

7. Centrifuge at 16,000 g for 5 min at 4°C
8. Repeat (3.6) & (3.7)
9. (Under the fume hood) Use a 100 µL pipet, transfer 400 µL of the upper phase to a new 2 mL tube, add 800 µL of Solution D, invert 10 times to mix
10. Incubate at 4°C for 2 h
11. Centrifuge at 18,000 g for 10 min at 4°C
12. Pour out and discard the supernatant. Centrifuge again at 18,000 g for 2 min at 4°C, remove all supernatant using the 100 µL pipet. Be careful not to suck away the pellet, it is very tiny.
13. Add 1 mL of ice cold 70% ethanol (Solution E). Wash the pellet by inverting the tube 5 times.
14. Centrifuge at 16,000 g for 10 min at 4°C
15. Remove and discard the supernatant by pipet. Be careful not to suck away the pellet, it is very tiny. Use the 1000 µL pipet to suck away the major volume, then remove the last bit by a 100 µL pipet.
16. Dry the pellet in a Speedvac for 20 min, or leave it on the working bench at room temperature for 1 h with a Kimwipe cover.
17. Dissolve the pellet in 50 µL sterilized dH2O or TE buffer.
18. Measure the DNA yield and quality using the Nanodrop.
19. (OPTIONAL) If the DNA quality measured in (3.18) is not good enough, clean the DNA extract by running 50 µL of it through steps (3.3) to (3.18), but skip step (3.8).
20. Save the DNA extract in the freezer.

III. Randomized Table for Root Genetic Material Decay Experiment

NORTH

S3D2R2	S5D1R5	S6D3R1	S6D3R10	S6D1R24	S7D1R10	S5D1R24	S4D1R16	S5D2R24	S4D1R6	S3D2R14	S5D1R3	S6D3R13	S2D1R14	S6D3R9	S5D2R21	S4D1R23	S6D1R3	S4D1R5
S7D1R12	S8D1R14	S8D1R12	S1D1R17	S3D2R16	S7D1R16	S5D3R7	S3D2R15	S2D2R2	S5D3R2	S6D2R21	S6D3R14	S4D1R10	S8D1R25	S3D1R25	S1D1R6	S3D3R10	S3D1R21	S5D3R17
S6D2R1	S3D1R20	S2D2R19	S3D3R23	S1D1R22	S2D2R13	S2D1R24	S8D1R20	S2D1R20	S3D1R11	S5D2R17	S3D3R13	S5D3R8	S2D2R18	S5D1R4	S3D1R14	S5D2R18	S2D1R6	S2D1R7
S2D1R8	S5D3R4	S6D2R13	S2D3R23	S3D1R19	S3D3R9	S3D2R23	S5D3R11	S7D1R19	S7D1R14	S7D1R11	S3D1R15	S2D2R1	S2D1R16	S5D1R15	S3D1R9	S1D1R23	S5D1R18	S7D1R20
S2D2R7	S6D1R23	S5D1R2	S8D1R1	S3D3R15	S2D2R4	S7D1R5	S3D1R7	S6D1R9	S1D1R11	S1D1R8	S5D3R6	S6D1R8	S3D1R5	S3D2R2	S2D2R20	S5D1R5	S3D1R10	S5D2R10
S2D1R3	S8D1R23	S2D2R8	S8D1R10	S6D1R11	S3D1R2	S5D1R17	S3D3R21	S4D1R22	S7D1R3	S5D2R6	S5D2R15	S4D1R1	S8D1R6	S5D1R23	S2D2R9	S5D1R9	S6D2R16	S2D2R12
S5D1R14	S5D2R22	S6D1R22	S2D2R22	S3D1R17	S6D2R4	S5D2R19	S2D2R11	S1D1R9	S7D1R13	S2D3R21	S3D1R18	S2D2R6	S6D2R15	S2D2R25	S8D1R22	S1D1R13	S5D3R19	S5D3R21
S5D2R13	S3D3R25	S2D1R15	S3D2R10	S2D3R5	S6D1R5	S4D1R17	S3D3R19	S4D1R7	S3D1R24	S1D1R10	S5D2R4	S5D1R11	S1D1R14	S5D2R11	S3D3R22	S2D1R10	S7D1R17	S6D1R12
S1D1R20	S2D3R14	S5D3R14	S6D1R6	S5D2R3	S2D3R16	S6D2R10	S5D3R5		S2D2R21	S6D3R2	S7D1R9	S4D1R12	S3D2R20	S6D3R20	S2D1R9	S1D1R21	S6D2R24	S8D1R17
S5D1R25	S3D2R8	S5D1R7	S3D2R18	S6D2R18	S5D1R1	S7D1R6	S3D2R22	S6D2R7	S5D1R19	S4D1R8	S2D3R18	S7D1R25	S2D2R23	S5D1R20	S2D1R2	S2D2R3	S6D2R5	S6D1R20
S4D1R9	S3D1R23	S3D2R5	S2D1R22	S4D1R19	S3D3R8	S6D1R2	S4D1R21	S3D1R22	S6D2R14	S3D1R13	S5D3R9	S5D1R22	S2D1R18	S2D2R5	S5D1R16	S2D1R5	S6D1R16	S1D1R4
S5D1R21	S5D1R6	S4D1R4	S6D3R6	S1D1R7	S7D1R8	S4D1R20	S5D2R20	S7D1R18	S3D1R16	S4D1R2	S5D2R5	S2D1R17	S3D3R14	S1D1R3	S6D1R19	S5D1R12	S7D1R22	S6D2R3
S5D3R25	S1D1R25	S1D1R16	S5D2R23	S8D1R16	S4D1R25	S3D3R6	S3D2R19	S2D3R9	S5D2R2	S8D1R21	S3D2R6	S3D2R3	S5D1R13	S3D2R17	S8D1R8	S7D1R2	S6D3R23	S3D3R2
S5D2R9	S6D1R4	S6D1R1	S1D1R2	S2D3R3	S2D3R22	S6D3R17	S8D1R19	S3D1R12	S3D3R24	S2D2R14	S2D1R1	S5D3R1	S2D2R10	S2D1R23	S5D2R14	S2D2R17	S3D3R11	S4D1R15
S3D3R5	S5D2R16	S3D2R21	S8D1R7	S4D1R13	S2D2R24	S6D3R24	S6D1R15	S3D3R3	S5D2R8	S3D3R4	S7D1R21	S3D1R8	S3D2R7	S3D1R3	S6D1R25	S8D1R11	S4D1R24	S4D1R3
S8D1R18	S7D1R4	S6D2R2	S8D1R2	S2D3R17	S3D3R16	S2D1R12	S1D1R12	S3D3R7	S7D1R15	S5D3R23	S6D1R21	S8D1R13	S2D1R11	S6D3R3	S2D1R4	S5D2R25	S6D1R14	S2D3R11
S2D3R2	S6D3R16	S3D1R1	S1D1R15	S2D1R19	S6D2R22	S2D2R16	S8D1R15	S7D1R24	S2D3R15	S2D3R20	S2D1R2	S1D1R19	S6D3R21	S3D3R1	S2D3R19	S8D1R9	S6D3R15	S6D1R18
S3D1R8	S7D1R1	S2D3R6	S8D1R4	S6D3R19	S3D3R20	S6D2R12	S6D1R13	S1D1R18	S5D2R1	S2D1R25	S5D1R10	S8D1R24	S6D1R10	S2D2R15	S3D1R4	S5D1R8	S7D1R23	S8D1R3
S1D1R1	S2D1R13	S6D2R23	S4D1R11	S6D2R20	S4D1R18	S8D1R5	S3D2R12	S5D2R7	S6D1R7	S6D2R19	S4D1R14	S3D2R24	S6D2R11	S6D1R17	S5D2R12	S6D2R6	S7D1R7	S1D1R24

S: Species Number (1 *Cornus canadensis,* 2 *Rosa acicularis,* 3 *Populus tremuloides,* 4 *Linnaea borealis,* 5 *Viburnum edule,*6 *Picea glauca,* 7 *Maianthemum canadense,* 8 *Poa pratensis*)

D: Diameter Class (1 <2mm, 2 2-5mm, 3 >5mm)
R: Replicate number (Numbers that are the same within a species come from the same individual, except in S2D1, where that information was lost)