Molecular phylogeny of the *RPB2* gene in five *Elymus* species with the StH genome

By
Katherine Irene Purvis

A Thesis Submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Bachelor of Science – Honours Biology

April, 2020, Halifax, Nova Scotia

Approved: _____
Dr. Genlou Sun
Supervisor

Approved: _____
Dr. Kevin Vessey
Examiner

Date Submitted: April 27th, 2020

Molecular phylogeny of the *RPB2* gene in five *Elymus* species with the StH genome

Katherine Irene Purvis

Abstract

Elymus is a polyploid genus within the grass tribe (Triticeae) and is thought to have originated from the hybridization between species in other Triticeae genera. As *Elymus* originated from processes such as hybridization and polyploidization, it is an ideal model species for studying how these processes lead to plant diversification and speciation. However, due to the taxonomic complexity of *Elymus,* many of the relationships within this genus are still under investigation. This study investigates allotetraploid *Elymus* species with the genome from the hybridization between *Pseudoroegneria* (**St)** and *Hordeum* (**H**). Specifically, the phylogenetic relationship among five *Elymus* species of the **StH** genome, *E. caninus, E. mutabilis, E. fibrosus, E. alaskanus* and *E. trachycaulus,* and their genome donor genera, *Pseudoroegneria* and *Hordeum,* were analyzed using the second-largest subunit of RNA polymerase II (*RPB2*) gene. Phylogenetic analysis revealed two distinct clades. One clade was formed with the **St** genome sequences of *Elymus* and sequences from *Pseudoroegneria.* A second clade was formed with the **H** genome sequences of *Elymus* and sequences from *Hordeum.* Phylogenetic analysis also revealed a close association between *E. mutabilis* and *E. fibrosus,* as well as with *H. bogdanii;* indicating *H. bogdanii* as a possible genome donor species. Furthermore, high nucleotide diversity was found within *E. caninus* and *E. trachycaulus.* The lowest nucleotide diversity was detected within *E. mutabilis.* Overall, the phylogenetic analysis proved a complex and diverse evolutionary history among the five *Elymus* species. The study demonstrates how the hybridization between two genetically distinct genera can lead to the production of a new genus encompassing numerous species with varying levels of genetic diversity through mutations and possibly subsequent hybridizations.

April 27th, 2020

Acknowledgments

I would like to express my utmost appreciation to my supervisor Dr. Genlou Sun for the opportunity to conduct an honours project under his support and guidance. I would like to thank the members of Sun Lab for their help and support, specifically to Ms. Ning Chen and Ms. Maggie Scarrow. I would also like to thank the Honours Class of 2020, Dr. David Chiasson, Dr. Colleen Barber, and Dr. Kevin Vessey, for their help, feedback, and support throughout this thesis. Lastly, I would like to thank SMUWorks for the funding of this study.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

RPB2: second-largest subunit of RNA polymerase II gene

SSR: simple sequence repeat

RAPD: random amplified polymorphic DNA

PepC: phosphoenolpyruvate carboxylase

cpDNA: chloroplast DNA

rDNA: ribosomal DNA

mRNA: messenger RNA

bp: base pairs

NJ: neighbor-joining

BS: bootstrap

Tris-HCl: Tris hydrochloride

NaCl: sodium chloride

EDTA: ethylenediaminetetraacetic

SDS: sodium dodecyl sulfate

PCR: polymerase chain reaction

dNTPs: deoxynucleotides

TBE: Tris/Borate/EDTA

H_2O: water

SOC: Super Optimal broth with Catabolite repression

LB: Lysogeny Broth

CI: consistency index

RCI: rescaled consistency index

RI: retention index

1. Introduction

1.1 Hybridization and polyploidization

Polyploidization and hybridization are two interconnected processes (Albertin and Marullo 2012). Hybridization refers to the crossbreeding of two genetically distinct species (Harrison and Larson 2014). Polyploidization refers to organisms with greater than two chromosome sets (Woodhouse et al. 2009). There are two main forms of polyploids: autopolyploids and allopolyploids (Alix et al. 2017). Autopolyploids are organisms with more than two sets of chromosomes, in which all sets are from the same genome. Allopolyploids have more than two sets of chromosomes as well; however, the chromosome sets are derived from various genomes (Shaked et al. 2001). Allopolyploids are formed when two or more genetically distinct species crossbred with one another to form a hybrid (Kawashima 2019).

Hybridization and polyploidization are of importance as they can lead to diversification and speciation (Wu et al. 2015). When an allopolyploid is formed, there are numerous challenges that the organisms must overcome in order to maintain successful functioning with multiple distinct genomes (Feldman and Levy 2009). For example, the organism must ensure intragenomic pairing at meiosis in order to provide viable offspring and conduct intergenomic gene expression and DNA replication (Levy and Feldman 2004). As a result, the genetic contents of the newly formed allopolyploid organism undergo immediate alterations. Previous studies have demonstrated that when an allopolyploid is produced, numerous genomic changes may arise. These genetic changes may include rapid elimination of specific low-copy DNA, retrotransposon activation, intergenomic conversion, and epigenetic modifications (Feldman et al. 1997; Comai et al. 2000; Lee and Chen 2001; Shaked et al. 2001; Guo et al. 2014).

The rapid genetic alterations must occur to increase the fitness of the allopolyploid to guarantee its survival in nature (Feldman and Levy 2012). However, the rapid and diverse genomic changes that can result from allopolyploidization can quickly lead to speciation. In addition, allopolyploidization also leads to distinctive evolutionary dynamics, which can cause genetic asymmetry evolution. As a result of genetic asymmetry evolution, conformity and convergent effects may be observable (Feldman et al. 2012).

1.2 Triticeae

Triticeae Dumort. is a tribe within the grass family Poaceae (Bothmer and Salomon 1994). Triticeae is a significant tribe within Poaceae as it contains some of the world's most important cereals such as wheat, barley, rye, and triticale (Lu and Ellstrand 2014). In addition to cereal crops, Triticeae also contains essential grains for animal fodder and grazing, as well as problematic weeds (Sun and Li 2006). Previously, morphological traits were used to classify the genera within Triticeae which posed issues as numerous genera within Triticeae are highly morphologically variable (Dewey 1984). Löve (1984) and Dewey (1984) proposed a classification system for Triticeae based on genomic relationships. Since this classification system was proposed, it has remained the predominant accepted classification system of Triticeae (Helfgott and Mason-Gamer 2004). Genomic analysis has revealed that Triticeae is highly variable in the genetic systems that exist within the tribe. Triticeae consists of diploids, allopolyploids, autopolyploids, and hybrids (Eilam et al. 2010). The proper classification of genera within Triticeae is important as it will increase our understanding of how the biological mechanisms and genetic systems within Triticeae lead to speciation in plants (Bothmer and Salomon 1994). The proper classification of Triticeae is also significant as it will

increase the ability to use the genetically diverse genera within Triticeae as genetic resources for domesticated cereal crops (Bothmer and Salomon 1994).

1.3 *Elymus*

The genus *Elymus* L., commonly known as wheatgrass or wild rye, belongs to the Triticeae tribe. The origin of *Elymus* is thought to have arisen from the hybridization of other Triticeae genera (Díaz 1999). *Elymus* is globally distributed and contains over 150 different species, making it the largest and most widely spread genus in Triticeae (Sun and Li 2006). *Elymus* can be found in nearly every continent, including Europe, Asia, North America, South America, and Australia. However, it grows prominently in the northern temperate climates of Asia and North America (Dewey 1984). *Elymus* occupies numerous different habitats, such as grasslands, semi-deserts, mountain slopes, valleys, and along forest edges (Sun and Li 2006).

Previously, *Elymus* has been used as a "taxonomic wastebasket" for Triticeae; meaning that undetermined polyploids without distinct morphological characteristics that could not be included into other Triticeae genera often were placed into *Elymus* (Lu 1993; Svitashev 1997). As a result, *Elymus* is a large and extremely diverse genus. The taxonomic classification of *Elymus* has proven to be complicated due to the high morphological variations, and the recurrent hybridizations between species that exist within this genus (Tzvelev 1976; Love 1984; Barkworth 1992; Díaz 1999).

The classification of *Elymus* is important as it will allow for the appropriate construction of conservation strategies. Conservation strategies can be improved by proper taxonomy as classification provides knowledge into the species' life history, species' threats, and different abiotic and biotic factors that affect the species (Falk 1992;

Given 1994). Determining the phylogenetic relationships of *Elymus* is also important for proper conservation as it will allow for the protection of the evolutionary potential within *Elymus* and species related to *Elymus* (Barrett and Kohn 1991).

The appropriate classification of *Elymus* will also improve the ability to use *Elymus* as a genetic resource for other Triticeae taxa. The species within *Elymus* have a close relationship with various Triticeae species and as a result can be used for interbreeding for improvements in genetic diversity as well as for the genes within *Elymus* for disease resistance, stress tolerance, wide adaptation, high protein and lysine contents, and high productivity under environmental stresses (McGuire and Dvorak 1981; Dewey 1984; Crane and Carman 1987; Dong et al. 1992; Liu et al. 1994; Lawrence 2011). Since the domestication of crops within Triticeae has led to a loss in genetic diversity due to the bottleneck effect, methods to enhance crops' diversity are a highly valuable field of research (Kilian et al. 2010). The phylogenetic analysis of *Elymus* is also beneficial as *Elymus* serves as a valuable model species for studying how hybridization and polyploidization lead to plant diversification and speciation (Wu et al. 2016).

Elymus can be genetically defined as an allopolyploid in which at least one genome set is derived from *Pseudoroegneria*, designated as the **St** genome. In addition to the **St** genome, there are four other genome donor genera to *Elymus: Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), and an unknown donor (**Y**) (Mason-Gamer 2013). *Elymus* comprises tetraploids, hexaploids, and octoploids (Dong et al. 2015). The genus is primarily tetraploids, as they comprise 75% of the genus. Fewer *Elymus* species are hexaploids and octoploids, as they comprise 20% and 5% of the genus, respectively (Dewey 1984; Löve 1984). The tetraploids within *Elymus* have the genome combination of either **StStHH** or **StStYY** (Dong et al. 2015). This study will be focusing on the

tetraploid *Elymus* species with the **StStHH** genome, which can be abbreviated as the **StH** genome. Specifically, we will be analyzing *E. caninus, E. trachycaulus, E. fibrosus, E. alaskanus* and *E. mutabilis.*

It has been under investigation as to whether the **StH** species of *Elymus* resulted from a single origin or from multiple origin events. A single origin event describes one parental hybrid species which differentiated to form all other *Elymus* species. Multiple origin events refer to more than one hybridization event that leads to the creation of different *Elymus* species. Molecular phylogenetic analysis can reveal the mode of species origination. Evidence for a single origin can be provided if the **St** genome sequences for each *Elymus* species appear similar, and the **H** genome sequences for each *Elymus* species appear similar. If at least one species has an **H** genome sequence or a **St** genome sequence that significantly diverges from the rest, then it is likely that species of *Elymus* has descended from multiple hybridizations (Helfgott and Mason-Gamer 2004). Current phylogenetic analysis suggests that multiple origins have occurred within *Elymus* (Sun et al. 2008; Mason-Gamer and Naum 2010; Sun and Komatsuda 2010; Sun and Zhang 2011).

1.4 *Elymus alaskanus*

Elymus alaskanus (Scribn. and Merr.) Á. Löve is commonly known as Alaskan wheatgrass, is an Artic-alpine species that is naturally distributed across the Circumpolar North (Zhang et al. 2002). Specifically, it is found in the northern parts of the former Russian Federation, Greenland, United States of America, and Canada, and as well as in Norden, Siberia, and Alaska (Sun and Salomon 2003). *E. alaskanus* is perennial, self-fertilizing, and typically grows in environments with minimal competition. *E. alaskanus*

can be found on limestone outcrops, screes, moraines, and dry meadows (Sun and Salomon 2003), typically displays a clumped dispersion pattern with inconsistent population sizes (Díaz et al. 1999a). Morphologically, *E. alaskanus* is highly diverse. However, among species populations, *E. alaskanus* typically appears homogenous (Sun and Salomon 2003). As a result, several arguments for the classification of *E. alaskanus* into different taxa have been proposed based on morphological and cytological studies (Tzvelev 1976; Melderis 1978; Löve 1984, Baum et al. 1991; Barkworth 1994; Cody 1996). Currently, 15 taxa have been identified in the *E. alaskanus* complex (Sun and Salomon 2003).

Although the genetic diversity within *E. alaskanus* has been widely investigated, the origin and phylogeny of *E. alaskanus* is still not well understood (Díaz et al. 1999a; Sun et al. 2002; Zhang et al. 2002; Sun and Salomon 2003; Gaudett et al. 2005; Stevens et al. 2007). A study by Alobeid (2017) attempted to analyze the origins of *E. alaskanus* using the second-largest subunit of RNA polymerase II (*RPB2*) gene and two chloroplasts genes (RPS16 and RP0A). However, no conclusions could be made as to whether *E. alaskanus* resulted from single or multiple origins, and as to which species within the genome donor genera *Pseudoroegneria* and *Hordeum* were responsible for the hybridization that lead to *E. alaskanus.* These findings were due to the lack of sequence data for the *RPB2* gene analyzed for *E. alaskanus.* Using random amplified polymorphic DNA (RAPD) markers, *E. alaskanus* has been found to have close relations with *E. mutabilis* (Sun et al. 1997) and *E. trachycaulus* (Gaudett et al. 2005).

1.5 *Elymus fibrosus*

Elymus fibrosus (Schrenk) Tzvelev is perennial and self-pollinating. *E. fibrosus* can be found across the northern and central regions of the former Russian Federation, and in northern Scandinavia (Tzvelev 1989). The habitats of *E. fibrosus* include damp meadows, riverside pebbles and sand, and between shrubs. *E. fibrosus* typically distribute either alone or sympatrically alongside *E. caninus, E. sibiricus* and *E. repens* (Díaz et al. 2000).

E. fibrosus has low genetic diversity; it has been speculated this could potentially be caused by the bottleneck effect (Sun et al. 1999; Díaz et al. 2000). A study by Wu et al. (2016) suggested that the **St** and **H** genomes within *E. fibrosus* are likely derived from a single origin. In addition, the single origin for the *E. fibrosus* genome is a probable explanation for its low genetic diversity when compared to other *Elymus* specieswhich have comparable ecological preferences and breeding systems to *E. fibrosus* (Wu et al. 2016).

1.6 *Elymus mutabilis*

Elymus mutabilis (Drobov) is a caespitose, meaning it forms dense tufts. *E. mutabilis* is also perennial and self-fertilizing (Díaz et al. 2004). *E. mutabilis* is found in northern Europe, and the northern and central regions of Asia (Tzvelev 1976). The populations of *E. mutabilis* typically have clumped distributions with irregular sizes. The typical habitats of *E. mutabilis* include meadows and birch forests, and among willows. *E. mutabilis* either grows in isolation or sympatrically with *E. caninus*, *E. alaskanus,* and *E. fibrosus* (Díaz et al. 2004).

In a study by Díaz et al. (2004), genetic variation within *E. mutabilis* was analyzed using allozymes. The study revealed that at both the species and population

levels of *E. mutabilis,* genetic variations were absent (Díaz et al. 2004), which did not agree with the predictions based on the comparisons towards other self-pollinating plants, as well as other species of *Elymus* (Sanders et al. 1979; Hamrick and Godt 1990; 1997; Knapp and Rice 1996; Díaz et al. 1998, 1999a,b). The phenomenon was rationalized as the species-level analysis of *E. mutabilis* had a small sampling distribution, which could have led to a lack of genetic diversity to be represented within the species. At the population level, Díaz et al. (2004) proposed that perhaps the absence of genetic variation observed in *E. mutabilis* could be due to the bottleneck effect or from genetic drift.

Díaz (1999b) and Sun et al. (2001) both observed a large number of hybrids between *E. caninus* and *E. mutabilis* or *E. fibrosus.* The interspecific hybridization between *E. caninus* and *E. mutabilis* would indicate an increase in gene flow (Díaz 1999b; Sun et al. 2001). The degree of gene flow between *E. mutabilis* and other members of *Elymus,* including *E. alaskanus*, *E. caninus*, and *E. fibrosus,* was determined by Wu et al. (2015). Within this study, simple sequence repeat (SSR) markers were used to analyze the level of gene flow between these species. The results revealed that genetic diversity within *E. mutabilis* was considered to be relatively low. *E. mutabilis* had the lowest genetic diversity when compared to *E. alaskanus*, *E. caninus*, and *E. fibrosus* (Wu et al. 2015).

1.7 *Elymus caninus*

Elymus caninus (L.) L. is perennial, self-pollinating and can be found throughout Iceland, the British Isles, southern Siberia, Subartic and Mediterranean (Wu et al. 2016). *E. caninus* grows in deciduous forests, forest glades, between shrubs, and occasionally within the meadows in the subalpine regions. The populations of *E. caninus* have a patchy

distribution pattern (Díaz et al. 1999b). Throughout previous studies of *E. caninus,* it has been found to be highly genetically and structurally variable. Variations within *E. caninus* have been found in morphology, prolamine and glutelines quantities, and polymorphisms within isozymes, RAPD, and microsatellite (Díaz et al. 1998; Kostina et al. 1998; Sun et al. 1998).

A study by Sun and Yan (2012) investigated the origins of *E. caninus* using the *RPB2* gene, phosphoenolpyruvate carboxylase (PepC), and the TrnD/T region within chloroplast DNA (cpDNA). Three distinct sequences of the *RPB2* gene were found within two accessions of *E. caninus*; one for the **H** genome sequence and two different sequences for the **St** genome, $\mathbf{St_1}$ and $\mathbf{St_2}$. The article concludes that the most likely explanation for this finding is that *E. caninus* with the original genome of **StH,** likely acquired additional sequence exchanges from another species with a different **St** genome copy. As a result, a $\mathbf{St_1St_2HH}$ species of *E. caninus* was created. The PepC gene sequence analysis confirmed that the **St** genome in *E. caninus* has two origins and that the most probable $\mathbf{St_2}$ genome donors are either *P. spicata* or *P. stipifolia*. The **H** genome sequences also indicated multiple origins of the *E. caninus.*

1.8 *Elymus trachycaulus*

Elymus trachycaulus (Link) Gould ex Shinners is perennial, self-pollinating, and is the most variable of all of the North American *Elymus* species in both morphology and distribution (Dewey 1982). *E. trachycaulus* is distributed across Eurasia and extends to Alaska, Newfoundland, and Mexico. The typical habitat of *E. trachycaulus* ranges from coastal to alpine regions, as well as from dry hillsides to damp meadows (Sun and Li 2006). The high morphological and geographical variability in *E. trachycaulus* has led

taxonomists to separate *E. trachycaulus* into subspecies, as well as to describe new species that were previously classified as *E. trachycaulus* (Tzvelev 1976, 1977). Currently, there are 3-6 subspecies within the *E. trachycaulus* complex; however, there previously was a significantly larger number of intraspecific taxa that have been included (Barkworth 1994).

Another issue adding to the complexity of the *E. trachycaulus* complex is that the subspecies can produce hybrids with other subspecies of *E. trachycaulus,* as well as with other tribe members of Triticeae (Jozwik 1966; Sun and Li 2006). As a result, *E. trachycaulus* often has close associations with other species of *Elymus* and Triticeae. Morphologically, *E. trachycaulus* is very similar to *E. alaskanus, E. glaucus,* and *E. caninus* (Sun and Li 2006). Whether or not *E. trachycaulus* should be considered a member of *E. caninus* has been debated by several taxonomists in the past (Malte 1932; Hitchcock and Chase 1950; Hitchcock 1969; Dewey 1975). However, in the present date, *E. trachycaulus* and *E. caninus* remain as separate species.

Sun and Li (2006) used RAPD to investigate the genetic diversity and the relationships among species within the *E. trachycaulus* complex and compared to other species of *Elymus*. *E. trachycaulus* was found to have the highest genetic diversity, even more so than the previously reported findings for *E. caninus* (Sun et al. 1999). Interestingly, *E. trachycaulus* showed more genetic similarity to *E. caninus, E. alaskanus,* and *E. mutabilis,* than it did within the species complex of *E. trachycaulus.* The authors also stated that although previous studies have found a genetic distinction between *E. caninus* and *E. trachycaulus,* further research on the relationship between the two species should be completed as the studies only included a few accessions of *E. caninus* (Dewey 1975; Larson et al. 2003; McMillan and Sun 2004).

A study by Zuo et al. (2015) examined the origins of *E. trachycaulus* using the *RPB2* and PepC genes and the TrnD/T region within cpDNA. Prior to this study, the origins of *E. trachycaulus* had been minimally explored. Multiple sources were found within the **H** genome sequences of the PepC gene from *E. trachycaulus.* It was then concluded that this finding is likely due to multiple origins or from successive hybridizations within *E. trachycaulus*. Multiple accessions of *E. trachycaulus* did not have **St** genome sequences that were recovered from the PepC gene sequences suggesting that *E. trachycaulus* may have undergone convergent evolution. Overall, it was concluded that *E. trachycaulus* has a very complicated evolutionary history.

1.9 Molecular phylogeny

Molecular phylogeny refers to the comparison of DNA sequences to determine evolutionary relationships (Brown 2002). Molecular phylogeny functions on the principle that mutations within DNA increase over time (Brown 2002). Therefore, variations within DNA can be used to determine evolutionary relationships. A recent divergence from an ancestor would have little variations when the genomes are compared, whereas an ancient divergence would have significantly greater nuclear variations (Brown 2002).

The selection of the DNA type or region to conduct the phylogenetic analysis is dependent on the goal of the phylogenetic study (Brown 2002). For intraspecific studies, the same members of a species are extremely genetically similar (Brown 2002). The highly conserved region of genome makes phylogenetic analysis difficult as typically no genetic variations can be detected, which leads to a lack of phylogenetic information (Brown 2002). However, relatively fast-evolving portions of the genome can be used for intraspecific analysis as a higher amount of DNA mutations occur within these regions

(Brown 2002). Examples of highly variable portions of the genome include repeated sequences, microsatellites, multiallelic genes, and intergenic regions of gene sequences (Lindsay 1995; Brown 2002; Small et al. 2004).

For analysis of interspecific phylogenetic relationships, highly conserved regions of the genome are typically used. Chloroplast DNA has been widely used within plant molecular phylogenetic studies (Small et al. 2004). Chloroplast DNA is highly conserved as it is haploid and does not undergo recombination (Small et al. 2004). However, cpDNA is donated solely by the maternal species. This possesses an issue for studying allopolyploids as the genome is a result of two genetically distinct parental species. Thus, the use of cpDNA within analysis of allopolyploids will only contain genetic information about the maternal species and will lack the paternal genome (Small et al. 2004).

In addition to cpDNA, nuclear ribosomal DNA (rDNA) has also commonly been used when determining interspecific phylogeny (Small et al. 2004). Ribosomal DNA contains both slow- and fast-evolving regions. However, rDNA is highly susceptible to concerted evolution in polyploids (Small et al. 2004). Concerted evolution refers to the homogenization of a repetitive multigene family (Liao 1999). Homogenization of a repetitive multigene family within a species requires the genetic exchange among homologous or nonhomologous chromosomes (Liao 1999). The most common method of genetic exchange for chromosomes is through gene conversion (Liao et al. 1997). Gene conversion occurs when similar sequences of DNA on chromosomes non-reciprocally transfer genetic information. As a result, the DNA sequences become more homogenized to one another (Liao 1999). This poses issues in the use of rDNA within allopolyploids, as nonhomologous chromosomes are present from each donor parent genome (Sang 2002). Concerted evolution can lead to the homogenization of nonhomologous

chromosomes, causing the chromosome pair to become more similar to one another, and more distinct from their parent donor species (Sang 2002). As a result, rDNA can be problematic when studying interspecific relationships within allopolyploids (Sang 2002). Another issue with the use of rDNA within allopolyploids is that concerted evolution is more probable within polyploids, and the chromosomes become increasingly more concerted as the number of genome sets increases (Small et al. 2004). This solution to studying interspecific relationships within allopolyploids is likely low-copy nuclear genes. Low-copy nuclear genes are less likely to produce concerted evolution and are still maternally and paternally inherited (Sang 2002).

1.10 *RPB2* gene

The *RPB2* gene is a low-copy nuclear gene that encodes for the second-largest subunit within RNA polymerase II (Kolodziej et al. 1990). RNA polymerase II is an enzyme responsible for the catalysis of messenger RNA (mRNA) synthesis within eukaryotes (Sun et al. 2007). According to a study on yeast by Ishiguro et al. (1998), the second-largest subunit of RNA polymerase II was found to interact with the largest subunit, along with other smaller subunits, to form RNA polymerase II.

The *RPB2* gene was first fully sequenced by Larkin and Guilfoyle (1993) in *Arabidopsis thaliana.* Larkin and Guilfoyle (1993) reported that the *RPB2* gene was 3,563 base pair (bp) long and had 24 introns and 25 exons, which vary in lengths. Several studies have reported that in eukaryotes, the *RPB2* gene has numerous regions that are exceptionally conserved (Sweetser et al. 1987; Kolodziej et al. 1990; Denton et al. 1998). Gibson and Spring (1998) postulated that the multiple interactions between the subunits of an enzyme lead to the genes encoding for the subunits to remain conserved. If a

mutation occurred at gene encoding for a domain within an enzyme, it would alter the interactions between the domain and the other subunits within the enzyme complex. Altering the interactions within the enzyme complex could lead to a deficiency in the functioning of the enzyme. A deficient enzyme would decrease the fitness of an organism (Gibson and Spring 1998). If the comments of Gibson and Spring (1998) were to be applied to the *RPB2* gene, a mutation would lead to a deficiency in RNA polymerase II. As a result, there would be a decrease in the effective synthesis of messenger RNA (mRNA), and thus would result in an organism that is less evolutionary fit. As a result, individuals with mutations within the *RPB2* gene should not be naturally selected.

The *RPB2* gene is biparentally inherited and is a single and low-copy gene. These characteristics of the *RPB2* gene lead it to be ideal for analyzing phylogenetic relationships within allopolyploids and their parental genome donors (Small et al. 2004). Sun et al. (2007) were the first to demonstrate the use of the *RPB2* gene for the analysis of the phylogenetic relationships within the **StH** genome species of *Elymus.* Within this study, the *RPB2* gene was proven to be exceptional for examining the phylogeny, evolutionary dynamics, speciation, nucleotide polymorphism, and type of polyploidy within *Elymus* (Sun et al. 2007).

Allotetraploid species of *Elymus* with the **StH** genome are expected to have copies of the *RPB2* gene. The *RPB2* gene should be present in the **St** haplomes and the **H** haplomes. The five *Elymus* species studied are primarily self-pollinators. Self-pollinating plants typically contain homozygous alleles for a gene. As a result, there should be minimal genetic variation between the haplomes within each of the genomes (Hartl and Clark 1997). The *RPB2* gene sequences for the **St** and **H** genomes should differ significantly as the genomes originate from genetically distinct sources (Wendel 2000).

1.11 Phylogenetic trees and analysis

A phylogenetic tree is a geometric representation of evolutionary relationships (Brown 2002). A clade on a phylogenetic tree refers to a branch which groups together organisms from a common ancestor (Baum 2008). A subclade is a branch extending from a member of a clade that further groups together organisms (Baum 2008). An outgroup on a phylogenetic tree is a species that is thought to be the most genetically unrelated to the ingroup of study and functions as a genetic reference point for the phylogenetic tree (Maddison et al. 1984).

A phylogenetic tree is constructed by first aligning the compared sequences of the chosen region of DNA (Brown 2002). After the alignment is performed, comparative data of the sequences is obtained (Brown 2002). The comparative data is used to construct the phylogenetic tree to represent the genetic relatedness of the compared sequences (Brown 2002).

Saitou and Nei (1987) created the algorithm for the NJ tree. The NJ tree implements the distance matrix method to determine the position of each sequence on the tree (Brown 2002). The distance matrix refers to a matrix that is created using calculated nucleotide differences, also known as polymorphisms, between all possible pairs of the sequences analyzed. The calculated nucleotide differences are used to determine the evolutionary distances among the compared sequences. The evolutionary distances are represented on the NJ tree in the length of the branches connecting the two sequences. The NJ tree uses one starting point from which all of the branches form. All possible pairs are determined by choosing a pair of sequences at a time and placing them at a second internal node. The total branch length of the second internal node for this pair is calculated using the distance matrix (Brown 2002). After all of the possible pairings have

been analyzed, the tree is constructed with the pairing that produced the shortest total branch length at the second internal node. This decision is based on the principle that the more genetically related two organisms are, the fewer genetic mutations should be observed. The pair of sequences that are chosen are also known as 'neighbours' to one another on the tree. This process is then repeated for the consecutive internal nodes until all of the sequences analyzed have been placed on the tree (Brown 2002).

The next step within the construction of a phylogenetic tree is to determine bootstrap (BS) values for each branch. Bootstrap values function as confidence limits to indicate the likeliness of each component of the reconstructed tree (Halliburton 2004). Within the determination of BS values, 1000 new gene alignments are created, in which 1000 phylogenetic trees are reconstructed. The BS value indicates how many times out of the 1000 tree reconstructions, each branch present on the final tree was incorporated (Brown 2002). The BS values are represented as percentages out of 100. A BS value of less than 70% are considered low, whereas a BS value greater than 80% is considered high (Daley 2006).

1.12 Study objectives

The objectives of the study were to: 1) determine the molecular phylogeny among five *Elymus* species with the **StH** genome: *E. caninus, E. mutabilis, E. fibrosus, E. alaskanus,* and *E. trachycaulus* using the *RPB2* gene sequences; 2) identify their genome donor from *Pseudoroegneria* and *Hordeum;* 3) determine and compare the nucleotide diversity among the *Elymus* species studied, as well as to their genome donor genera.

2. Methods

2.1 Plant materials

The species used in this study were: *Elymus alaskanus*, *E. caninus*, *E. fibrosus, E. mutabilis, E. trachycaulus, Hordeum bogdanii, H. chilense, H. intercedens, H. stenostachys, H. euclaston, H. pusillum, H. roshevitzii, Pseudoroegneria spicata, P. stipifolia, P. libanotica, Agropyron cristatum, Lophopyrum elongatum, and Bromus catharticus* (Table 1). The origins of the seeds used in this study were from various locations globally (Figure 1). The seeds were obtained from the Swedish University of Agricultural Sciences. The seeds were germinated and grown within a greenhouse at Saint Mary's University, Halifax, Nova Scotia.

Table 1: The species names, accession numbers, origin, and genome constitution of the *Elymus, Hordeum, Pseudoroegneria, Agropyron, Lophopyrum,* and *Bromus* species used in this study.

Species	Accession Number	Origin	Genome
E. caninus	SV9714-15	Jamtland, Sweden	StH
	SV9714-56	Jamtland, Sweden	StH
	FI9303-3	Pera-Pohjanmaa, Finland	StH
	FI9303-7	Pera-Pohjanmaa, Finland	StH
	H10096	Korgonskij, Altai, Russia	StH
	H3915	Sobostiste, Slovakia	StH
	DK9604-21	Sjoelland, Denmark	StH
	DK9604-29	Sjoelland, Denmark	StH
	H10404	Siberia, Russia	StH
	H2009	Fyledalen, Sweden	StH
	H8753	China	StH
	H3857	Siberia, Russian Federation	StH
	H10125	USSR, Russian Federation	StH
	H10359	Iceland	StH
	FI9310-5	Kittilan, Finland	StH
	FI9310-13	Kittilan, Finland	StH
	NO0308-15	Finnmark, Norway	StH
	NO0308-2	Finnmark, Norway	StH
	H3169	Vastmanland, Sweden	StH
	H10314	Finland	StH
	H7550	China	StH
E. fibrosus	PI 531609	Germany	StH
	PI 406467	Russian Federation	StH
	PI 564933	Kazakhstan	StH
	PI 564930	Russian Federation	StH
	PI 598465	Russian Federation	StH
	PI 406448	Russian Federation	StH
	FI9306-2	Pera-Pohjanmaa, Finland	StH
	FI9306-4	Pera-Pohjanmaa, Finland	StH
	PI 439999	Russian Federation	StH
	PI 564932	Russian Federation	StH
	US9601	Washington, United States	StH

Table 1 continued

Species	Accession Number	Origin	Genome
E. mutabilis	SV9319-3	Norrbotten, Sweden	StH
	SV9318-4	Norrbotten, Sweden	StH
	SV9318-5	Norrbotten, Sweden	StH
	NO0305-3	Finnmark, Norway	StH
	NO0305-51	Finnmark, Norway	StH
E. trachycaulus	H10140	Altai, Russian Federation	StH
	H10665	United States	StH
	H4228	Lincoln County, Utah, USA	StH
	PI 232150	United States	StH
	PI 440101	Shorthandy, Kazakhstan	StH
	H3995	Rich County, Utah, USA	StH
	PI 387895	Beaverlodge, Alberta, Canada	StH
	PI 232147	United States	StH
	H3526	Nerungri, Russian Federation	StH
	PI 440098	Tselinograd, Kazakhstan	StH
E. alaskanus	Gr9720	Strømfjord, Greenland	StH
	Gr9718	Julianehab, Greenland	StH
H. bogdanii	H4014	Pakistan	H
H. chilense	H1816	Chile	H
H. intercedens	H1941	United States	H
H. stenostachys	H1780	Argentina	H
H. euclaston	H2148	Uruguay	H
H. pusillum	H2024	United States	H
H. roshevitzii	H9152	China	H
P. spicata	PI 506274	Washington, United States	St
	PI 610986	Washington, United States	St
P. stipifolia	PI 325181	Russia Federation	St
P. libanotica	PI 330688	Sirak-Sar, Iran	St
Ag. cristatum	PI 383534	Kars, Turkey	P
L. elongatum	PI 142012	Russian Federation	E
B. catharticus	CN32048	Germany	N/a

Figure 1: Global map depicting the origins of the *Elymus, Hordeum, Pseudoroegneria, Agropyron, Lophopyrum,* and *Bromus* species used in this study. Image of map was obtained from Google Maps (2020).

2.2 DNA isolation

Ms. Panpan Wu and Mr. Yi Xu performed the DNA isolation. To perform the DNA isolation, first, liquid nitrogen was used to snap freeze 300 mg of a leaf sample. The leaf sample was crushed into a fine powder using a mortar and pestle. The powder sample was transferred into a 2.0 mL centrifuge tube. Seven hundred fifty µL of lysis buffer was added to the tube. The lysis buffer consisted of 50mM Tris hydrochloride (Tris-HCl) (pH 7.6), 100mM sodium chloride (NaCl), 50 mM ethylenediaminetetraacetic (EDTA), 0.5 % sodium dodecyl sulfate (SDS) and 10 mM β-mecaptoethanol. The tube then incubated for 10 minutes at room temperature. Four hundred fifty µL of phenol-tris-chloroform (pH 7.5) was then added to the tube. The tube was inverted several times to mix the solution. Four hundred fifty µL of (24:1) chloroform-isoamyl alcohol was added, and the tube was inverted several times again. The tube was then centrifuged for 5 minutes at 1,300 RPM. The supernatant was transferred to a new 2.0 mL centrifuge tube. Another 450 µL aliquot of chloroform-isoamyl alcohol was added, followed by centrifuging the tube for five-minutes at 1,300 RPM. The supernatant was placed into a 1.5 mL centrifuge tube. One mL of chilled isopropanol was then placed in the tube. The tube then incubated at -20°C for 20 minutes. The tube was then centrifuged for 20 minutes at 1,300 RPM and 4°C. The supernatant was removed, and the pellet was washed. The pellet was washed twice with cold 70% ethanol. The pellet was left to dry to 10 minutes. Four hundred µL of TE buffer was then added to resuspend the pellet. The TE buffer consisted of 10 mM of Tris-HCl (pH 7.5) and 1.0 mM EDTA (pH 8.0). The tube was then left to incubate for 24 hours. Fifty µg/mL of RNase was then added to the test tube and incubated for 30 minutes at 37°C. Four hundred fifty µL of phenol-tris-chloroform (pH 7.5) was then added to the tube. Four hundred fifty µL of (24:1) chloroform-isoamyl alcohol was added and the tube

was centrifuged for 5 minutes at 1,300 RPM. The supernatant was transferred to a 2.0 mL centrifuge tube. Another 450 µL aliquot of chloroform-isoamyl alcohol was added, followed by centrifuging the tube for 5 minutes at 1300 RPM. The supernatant was placed into a 1.5 mL centrifuge tube. One mL of chilled isopropanol was then placed in the tube. The tube then incubated at -20°C for 20 minutes. The tube was then centrifuged 20 minutes at 1,300 RPM and 4°C. The supernatant was removed, and the pellet was washed twice using cold 70% ethanol. The pellet dried for 10 minutes at room temperature. The pellet was resuspended in 100 µL of TE buffer and stored at -20°C until further use.

2.3 DNA amplification of the *RPB2* gene with PCR

The DNA amplification of the *RPB2* gene was performed by Ms. Panpan Wu and Mr. Yi Xu. The primers used in this study were synthesized by Operon Technologies Inc. The primers used were: P6F (5'-TGGGGAATGATGTGTCCTGC-3') and F6FR: (5'-CGAACCACACCAACTTCAGTGT-3') (Denton et al. 1998). The primers were used to amplify the *RPB2* gene through Polymerase Chain Reaction (PCR). A Biorad iCycler® Thermal Cycler was used to conduct PCR. *Taq* DNA polymerase (M0267L) and reaction buffer (B90046S) were both obtained from New England Biolabs Ltd. The master mix for amplifying the *RPB2* gene contained: 3.00 µL of 1.5 ng/µL DNA, 2.00 µL of 0.2 µM P6F primer, 2.00 µL of 0.2 µM F6FR primer, 2.00 µL of the reaction buffer, 4.00 µL of 0.2 mM deoxynucleotides (dNTPs), 0.15 µL of 0.75 U *Taq* DNA polymerase, and 6.85 µL of autoclaved water. The samples were placed into the PCR. The PCR program consisted of an initial denaturing of the sample at 95°C for 5 minutes. Then the following three steps occurred for 35 cycles: denaturing of the sample at 95°C for 1 minute, annealing at 52°C

for 2 minutes, extension at 72°C for 2 minutes. After the cycles were complete, a final extension occurred at 72°C for 7 minutes. The samples were then incubated at 4°C.

2.4 Gel electrophoresis

The gel for electrophoresis was created using 1.5% agarose gel and 150 mL of 1x Tris/Borate/0.5 M EDTA (TBE) buffer (pH 8.0). The agarose gel and the TBE buffer was placed into an Erlenmeyer flask and microwaved for 2 minutes. The gel was placed into a 15-well gel comb mould. Once the gel hardened, the gel was positioned in the electrophoresis chamber. Five µL of the DNA sample was mixed with 2 µL of bromophenol blue loading dye and loaded into each well. The gel ran at a consistent 180 volts for 30 minutes. The electrophoresis gel was stained for 30 minutes in 30 µL ethidium bromide in 200 mL of distilled water (H_2O). SynGene photographic equipment and a UV light was then used to view the PCR products.

2.5 Cloning and sequencing

The PCR products were cloned using a pCR8/GW/TOPO® TA Cloning kit (Cat. 45-0642, Invitrogen, Carlsbad, CA). The procedure for cloning was adapted from the User Manual (Invitrogen 2012). In brief, to create the TOPO® Cloning reaction, 4 µL of the PCR product was incubated with 1 µL of TOPO® vector and 1 µL of salt solution and kept at room temperature for 10 minutes. The solution was then placed on ice. Two µL of the TOPO® Cloning reaction was placed in a vial of One Shot® of *Escherichia coli,* and gently mixed. The vial was then left to incubate on ice for 30 minutes, then the vial was heat-shocked by placing the mixture into a bath for 30 seconds at 42°C and then immediately back on ice. Two hundred fifty µL of Super Optimal broth with Catabolite

repression (SOC) medium was then added to the vial. The vial was placed into a bath for 1 hour with shaking at 2,000 RPM and a temperature of 37°C. Fifty µL of the mixture from the vial was spread on a pre-warmed Lysogeny Broth (LB) with 100 µg/mL spectinomycin and x-gal, and incubated overnight at 37°C.

The colonies were first cultured for 20 minutes in LB medium with 100 µg/mL spectinomycin. Two µL of the sample was then obtained and placed in the PCR. The primers for the PCR reaction were obtained from the TOPO® kit. For the PCR products that tested positive for the insert, 8 µL of the remaining solution was placed into a 5 mL LB broth with 100 µg/mL spectinomycin. The broth was left to incubate overnight in a bath at 37°C.

To isolate the plasmid DNA, a Promega Wizard ® *Plus* Minipreps DNA Purification System (Promega Corporation, Madison, WI, USA- Cat # A7500) was used. The method for plasmid DNA isolation was adapted from the user manual (Promega 2011). Five µL of the bacterial culture was centrifuged for 10 minutes at 10,000g. The supernatant was discarded, and the tube was blotted upside-down on a paper towel to eliminate any additional solution. The cell pellet was resuspended in 300 µL of Cell Resuspension Solution. Three hundred µL of Cell Lysis Solution and 300 µL of Neutralization Solution were then pipetted into the centrifuge tube. The centrifuge tube was then centrifuged for 5 minutes at 10,000g. The Wizard® Miniprep Column was prepared by placing a Syringe Barrel on the Luer-Lok® extension of the Minicolumn. One mL of the resuspended resin was pipetted in the Syringe Barrel. The supernatant from the centrifuge tube was then pipetted into the Syringe Barrel. The resin/lysate mixture was then slowly pulled through the Minicolumn by applying a vacuum. Two mL of Column Wash Solution was added to the Syringe Barrel and was pulled through the

Minicolumn by the vacuum. The Syringe Barrel was removed, and the Minicolumn was then transferred to a 1.5 mL centrifuge tube. The tube was centrifuged for 2 minutes at 10,000g to remove the Column Wash Solution. The Minicolumn was moved to a new 1.5 mL centrifuge tube. Fifty μL of deionized H_2O was added to the Minicolumn. After 1 minute, the tube was centrifuged for 10,000g for 20 seconds to elute the DNA. To ensure that the required DNA segment was properly isolated, a small sample was then placed into the electrophoresis apparatus with 1% agarose gel. The plasmid DNA was sequenced commercially.

2.6 Data analysis

The sequences were aligned using the default parameters of ClustalX (1.81) (Thompson et al. 2003). The sequence alignment was imported to the program GeneDoc (2.7) for the removal of large gaps, followed by a manual realignment (Nicholas and Nicholas 1997). The PAUP* (4.0b10) program was used to construct the phylogenetic tree from the sequence alignment, as well as to determine a maximum parsimony score (Swofford 2002). *Bromus catharticus* was chosen as the outgroup for the phylogenetic analysis (Sun and Ma 2009). An NJ tree was selected to represent the phylogenetic relationships after several trees had been constructed. In order to determine the robustness of the branches, BS values with 1000 replicates were calculated (Felsenstein 1985). The maximum parsimony score was calculated to indicate the character congruency by determining the consistency index (CI), rescaled consistency index (RCI), and the retention index (RI). The CI is a mathematical algorithm that represents the smallest amount of homoplasy inferred by the tree (Farris 1989). The closer the CI value is to one, the more reliable the tree is (Nei and Kumar 2000). The RI describes how parsimonious

the tree is. The RI is determined by calculating the (maximum possible number of changes on the tree and subtracting the number of changes on the tree), and then by dividing this number by the (maximum number of changes on the tree subtracted by the number of the smallest changes within the data set) (Farris 1989). The RCI is determined by multiplying the CI value by the RI value. The RCI value allows for a CI with a rescaled minimal possible value of zero.

The sequence alignments were then imported into the software DnaSP5 (Librado and Rozas 2009) to determine the nucleotide diversity of *Elymus,* as well as within each *Elymus* species studied. The nucleotide variation was not determined within *E. alaskanus* as the study did not contain enough sequences for this species. DNAsp5 was used to calculate the number of sites without missing data (*n*), the number of segregating sites (s), Tajima's statistics (π), and Watterson's θ (θ_w). The π and θ_w values were selected as parameters of nucleotide diversity. Watterson's θ is the diversity based on the number of segregating sites (Waterson 1975). Tajima's statistic represents the average pairwise diversity (Tajima 1989). The Fu and Li's D test and the Tajima's D test were then calculated. The Fu and Li's D test statistic is based on how many mutations appeared only once compared to the total mutations present in the data set (Fu and Li 1993). The Tajima's D statistic is based on the difference between the mean number of pairwise differences and the number of segregating sites (Tajima 1989). Fu and Li's D tests and Tajima's D tests were considered statistically significant when the p-values were < 0.05.

3. Results

3.1 Sequence alignment

The aligned sequences of the *RPB2* gene for the *Elymus, Hordeum,* and *Pseudoroegneria* species studied had a size of approximately 800-1000 bp. This finding agrees with previously published data (Sun and Komatsuda 2010). Remarkably, the sequence alignment showed a large deletion (indel) of 70 bp at the 50 position from the **H** genome (Figure 2). For the **St** genome, a 4 bp deletion was found at the 37 position.

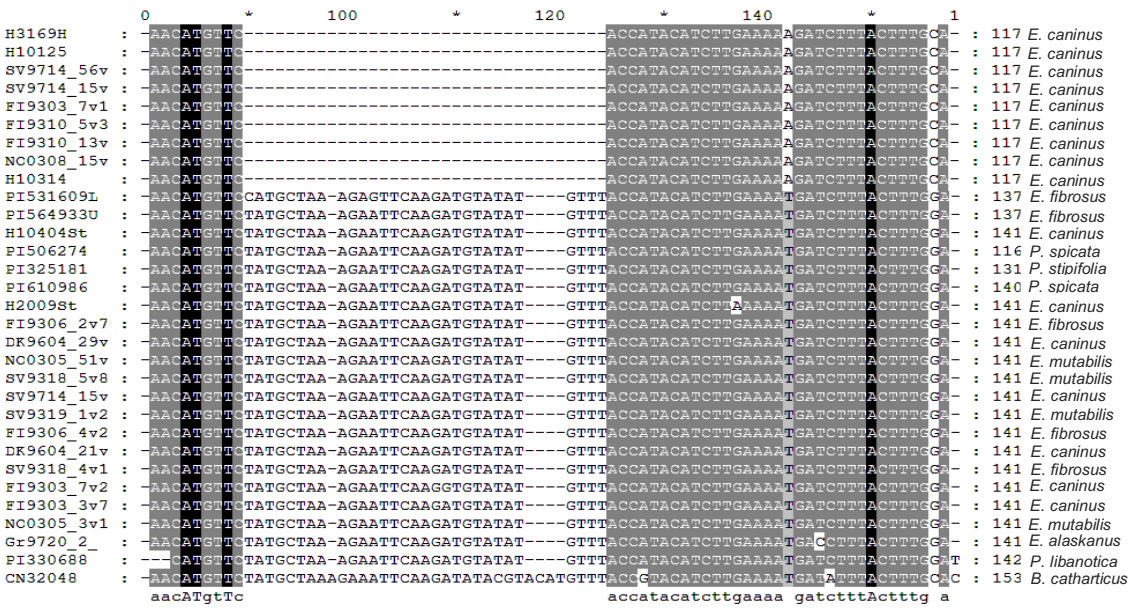


Figure 2: ClustalX sequence alignment of the *RPB2* obtained from five *Elymus* StH species. The sequence alignment showed a large deletion (indel) of 70 bp at the 50 position from the **H** genome. Black highlighting represented a nucleotide base position in which no difference was found among sequences. Dark grey highlighting represented a nucleotide base position in which one difference was found among sequences. Light grey highlighting indicated a nucleotide base position in which numerous differences were found.

3.2 Nucleotide variation

DnaSP 5.0 was used to analyze nucleotide diversity within the *RPB2* sequences from the **H** and **St** genomes of the five *Elymus* species studied (Table 2). The average pairwise diversity (π) and the diversity based on the number of segregating sites (θ_w) were used as estimates of nucleotide variations and were compared among sequences. When comparing the π and θ_w values of each *Elymus* species, higher nucleotide variation was found within the **St** genome sequences when compared to the **H** genome in *E. caninus*, *E. fibrosus* and *E. mutabilis.* In *E. trachycaulus,* the **H** genome was found to have higher π and θ_w values than the **St** genome, indicating higher nucleotide polymorphism in the **H** genome in *E. trachycaulus*. The *Elymus* species with the lowest nucleotide variation was found to be *E. mutabilis* (θ_w = 0.00121, 0.00062; π = 0.00101, 0.00047 for **St** and **H**, respectively). The *Elymus* species with the highest nucleotide variation within the **St** genome found to be *E. caninus* (θ_w = 0.02085; π = 0.00898). The *Elymus* species with the highest nucleotide variation within the **H** genome found to be *E. trachycaulus* (θ_w = 0.01458; π = 0.01458). Tajima (1989) and Fu and Li's (1993) tests were then performed on *RPB2* gene sequences for the *Elymus* species. Each species of *Elymus* had a Tajima's and Fu and Li's D values that were negative; signifying an excess of low-frequency polymorphisms. The Tajima's and Fu and Li's D values were found to be significant in both genomes of *E. fibrosus* and the **St** genome of *E. caninus,* suggesting for the directional selection or population expansion after a bottleneck for these species (Tajima 1989). The Tajima's and Fu and Li's D values were found to be not significant in both genomes of *E. mutabilis* and *E. trachycaulus,* as well as within the **H** genome of *E. caninus.*

Previous studies on the *RPB2* gene nucleotide diversity of the **H** and **St** genomes in the *Elymus* species showed that the θ_w values for the **St** and **H** genome within *Elymus* were similar, with values of 0.02423 and 0.02359, respectively (Table 2). The π value for the **H** genome within *Elymus* was found to be higher than **St** value (0.013180 and 0.00727, respectively). The nucleotide polymorphism within the **St** genome of *Pseudoroegneria* was obtained from previously published data (Sun et al. 2008). The π and θ_w values for *Pseudoroegneria* were found to 0.01711 and 0.01640, respectively. The π value for *Pseudoroegneria* was determined to be over 2.35-fold higher than for π value for *Elymus.* However, the θ_w value was found to be over 1.47-fold higher for *Elymus* when compared to *Pseudoroegneria.* A Fu and Li's test was then performed for the **St** and **H** genome sequences within *Elymus*. The Fu and Li's D value for the **St** genome was found to be -4.3121 and statistically significant ($p<0.02$). The Fu and Li's D value for the **H** genome was found to be -3.4712 and statistically significant ($p<0.02$). Negative Fu and Li's D values indicate that there is an excess of mutations in the external branches to the *RPB2* sequences of the **H** and **St** genomes of the *Elymus* species studied (Fu and Li 1993). A Tajima's test was then performed on both genome types. For the **H** and **St** genomes, the Tajima's D value was determined to be not statistically significant ($p<0$; $p>0.10$, respectively).

Table 2: Estimates of nucleotide polymorphism and test statistics for the *RPB2* gene within each *Elymus* species of a St and H genome. *n* represents the number of sites without missing data, s is the number of segregating sites, π is the average pairwise diversity, θ_w is the diversity based on the number of segregating sites.

Species	Genome	*n*	s	π	θ_w	Fu and Li's D	Tajima's D
Elymus	St	648	69	0.00727	0.02423	-4.3121*	-2.4988
	H	549	58	0.01380	0.02359	-3.4712*	-1.4434
Pseudoroegneria#	St	721	27	0.01711	0.01640	-0.0332	0.0364
E. caninus	St	741	54	0.00898	0.02085	-3.3396*	-2.3232*
	H	695	26	0.01005	0.01054	-0.9290	-0.4490
E. fibrosus	St	795	28	0.00799	0.01412	-2.4581*	-2.0785*
	H	705	16	0.00545	0.00975	-2.2777*	-1.9649*
E. mutabilis	St	795	2	0.00101	0.00121	-0.9726	-0.9726
	H	705	1	0.00047	0.00062	-0.9502	-0.9330
E. trachycaulus	St	796	16	0.00560	0.00712	-0.4176	-0.9915
	H	704	40	0.01458	0.02008	-1.2643	-1.3297

#: Data for *Pseudoroegneria* was obtained using previously published data (Sun et al. 2007).
*: Indicates statistical significance

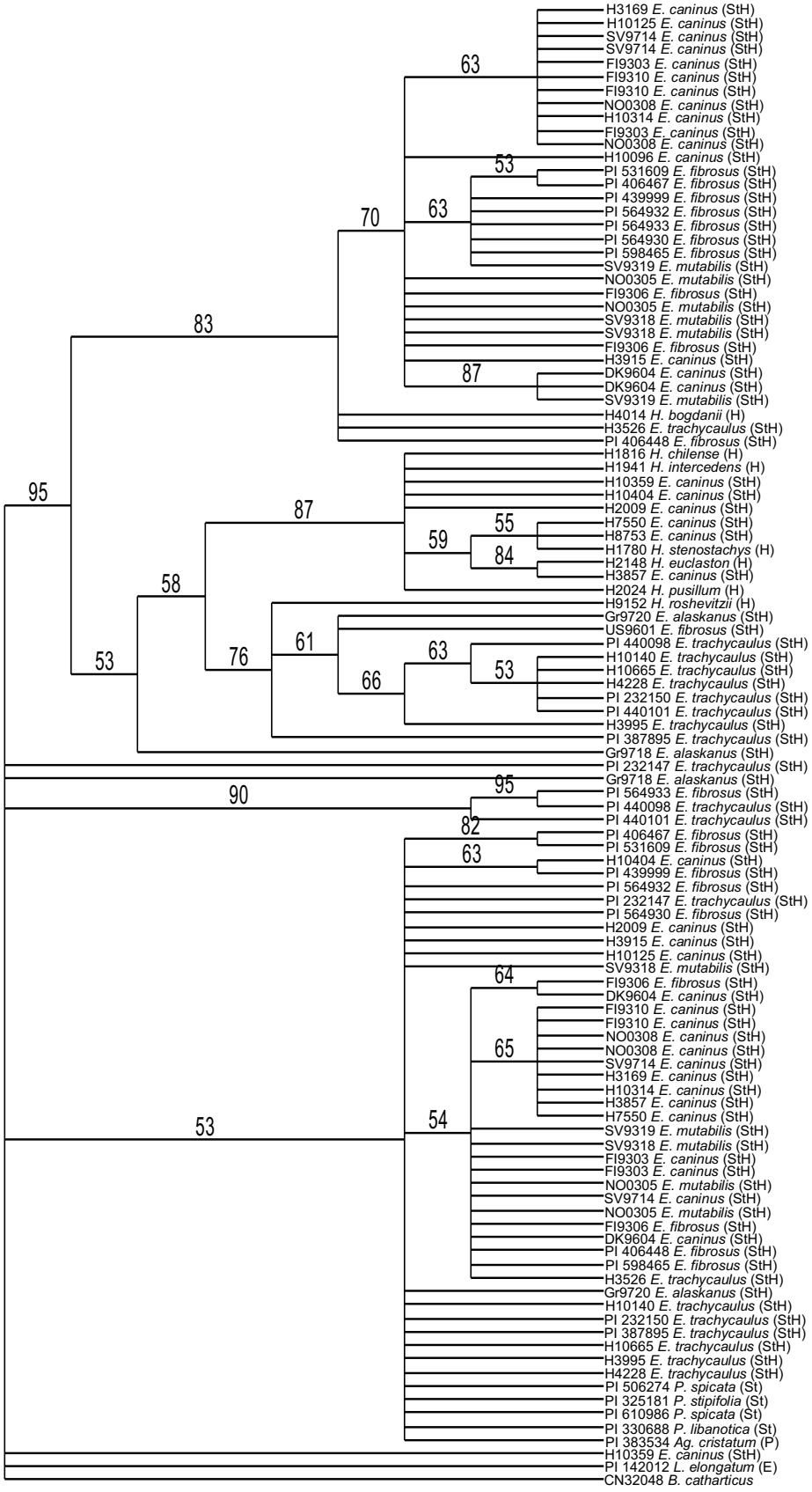
3.3 Phylogenetic analysis

Phylogenetic analysis was conducted on 110 sequences of the *RPB2* gene using *Bromus catharticus* as an outgroup. The data matrix consisted of 826 characters, 89 of which were parsimony-informative characters. Parsimony analysis of the data produced a CI of 0.754, a RI of 0.965, and an RCI of 0.582.

An NJ tree was constructed and shown in Figure 3 with BS values above branches. The two distinct copies of sequences from each *Elymus* species were separated into two clades; one clade contained **St** genome sequences from *Elymus* and all *Pseudoroegneria* (**St**) with a 53% BS value. Several subclades were formed within the **St**

(*Pseudoroegneria* + *Elymus*) clade. The sequences from the **H** genome of *Elymus* and all the *Hordeum* (**H**) sequences were placed in a clade with a 95% BS value. **H** genome sequences from *E. caninus, E. alaskanus,* and *E. trachycaulus* were placed in a subclade with all *Hordeum* sequences, excluding *H. bogdanii* (53% BS). *H. bogdanii* was found in a subclade with the **H** genome sequences from *E. caninus, E. fibrosus,* and *E. mutabilis* (83% BS).

Figure 3: Neighbor-joining (NJ) tree derived from the sequenced data of the *RPB2* gene in *Elymus, Pseudoroegneria, Agropyron,* and *Lophopyrum.* Numbers above the branches were bootstrap (BS) values. *Bromus catharticus* was used as an outgroup. Consistency index (CI) of 0.754, a retention index (RI) of 0.965 and a rescaled consistency index (RCI) of 0.582.


H3169 E. caninus (StH)
H10125 E. caninus (StH)
SV9714 E. caninus (StH)
SV9714 E. caninus (StH)
FI9303 E. caninus (StH)
FI9310 E. caninus (StH)
FI9310 E. caninus (StH)
NO0308 E. caninus (StH)
H10314 E. caninus (StH)
FI9303 E. caninus (StH)
NO0308 E. caninus (StH)
H10096 E. caninus (StH)
PI 531609 E. fibrosus (StH)
PI 406467 E. fibrosus (StH)
PI 439999 E. fibrosus (StH)
PI 564932 E. fibrosus (StH)
PI 564933 E. fibrosus (StH)
PI 564930 E. fibrosus (StH)
PI 598465 E. fibrosus (StH)
SV9319 E. mutabilis (StH)
NO0305 E. mutabilis (StH)
FI9306 E. fibrosus (StH)
NO0305 E. mutabilis (StH)
SV9318 E. mutabilis (StH)
SV9318 E. mutabilis (StH)
FI9306 E. fibrosus (StH)
H3915 E. caninus (StH)
DK9604 E. caninus (StH)
DK9604 E. caninus (StH)
SV9319 E. mutabilis (StH)
H4014 H. bogdanii (H)
H3526 E. trachycaulus (StH)
PI 406448 E. fibrosus (StH)
H1816 H. chilense (H)
H1941 H. intercedens (H)
H10359 E. caninus (StH)
H10404 E. caninus (StH)
H2009 E. caninus (StH)
H7550 E. caninus (StH)
H8753 E. caninus (StH)
H1780 H. stenostachys (H)
H2148 H. euclaston (H)
H3857 E. caninus (StH)
H2024 H. pusillum (H)
H9152 H. roshevitzii (H)
Gr9720 E. alaskanus (StH)
US9601 E. fibrosus (StH)
PI 440098 E. trachycaulus (StH)
H10140 E. trachycaulus (StH)
H10665 E. trachycaulus (StH)
H4228 E. trachycaulus (StH)
PI 232150 E. trachycaulus (StH)
PI 440101 E. trachycaulus (StH)
H3995 E. trachycaulus (StH)
PI 387895 E. trachycaulus (StH)
Gr9718 E. alaskanus (StH)
PI 232147 E. trachycaulus (StH)
Gr9718 E. alaskanus (StH)
PI 564933 E. fibrosus (StH)
PI 440098 E. trachycaulus (StH)
PI 440101 E. trachycaulus (StH)
PI 406467 E. fibrosus (StH)
PI 531609 E. fibrosus (StH)
H10404 E. caninus (StH)
PI 439999 E. fibrosus (StH)
PI 564932 E. fibrosus (StH)
PI 232147 E. trachycaulus (StH)
PI 564930 E. fibrosus (StH)
H2009 E. caninus (StH)
H3915 E. caninus (StH)
H10125 E. caninus (StH)
SV9318 E. mutabilis (StH)
FI9306 E. fibrosus (StH)
DK9604 E. caninus (StH)
FI9310 E. caninus (StH)
FI9310 E. caninus (StH)
NO0308 E. caninus (StH)
NO0308 E. caninus (StH)
SV9714 E. caninus (StH)
H3169 E. caninus (StH)
H10314 E. caninus (StH)
H3857 E. caninus (StH)
H7550 E. caninus (StH)
SV9319 E. mutabilis (StH)
SV9318 E. mutabilis (StH)
FI9303 E. caninus (StH)
FI9303 E. caninus (StH)
NO0305 E. mutabilis (StH)
SV9714 E. caninus (StH)
NO0305 E. mutabilis (StH)
FI9306 E. fibrosus (StH)
DK9604 E. caninus (StH)
PI 406448 E. fibrosus (StH)
PI 598465 E. fibrosus (StH)
H3526 E. trachycaulus (StH)
Gr9720 E. alaskanus (StH)
H10140 E. trachycaulus (StH)
PI 232150 E. trachycaulus (StH)
PI 387895 E. trachycaulus (StH)
H10665 E. trachycaulus (StH)
H3995 E. trachycaulus (StH)
H4228 E. trachycaulus (StH)
PI 506274 P. spicata (St)
PI 325181 P. stipifolia (St)
PI 610986 P. spicata (St)
PI 330688 P. libanotica (St)
PI 383534 Ag. cristatum (P)
H10359 E. caninus (StH)
PI 142012 L. elongatum (E)
CN32048 B. catharticus
63
53
63
70
83
87
95
87
55
59
84
58
53
76
61
63
66
53
90
95
82
63
64
65
53
54

4. Discussion

4.1 Nucleotide variation

The *Elymus* species with the lowest nucleotide variation was found to be *E. mutabilis.* This finding agrees with the results of Wu et al. (2015), in which *E. mutabilis* had the lowest genetic diversity when compared to *E. alaskanus*, *E. caninus*, and *E. fibrosus* using SSR markers. A possible explanation for the lack of genetic diversity within *E. mutabilis* is that the species could have experienced the bottleneck effect, genetic drift, or a high degree of inbreeding, as hypothesized by Díaz et al. (1999).

The *Elymus* species with the highest nucleotide variation within the **St** genome was found to be *E. caninus.* High nucleotide diversity has been reported within *E. caninus* in several studies (Sun et al. 1998; Sun and Ma 2009; Sun et al. 2006). The *Elymus* species with the highest nucleotide variation within the **H** genome was found to be *E. trachycaulus*. This finding agrees with the previous results of Sun and Li (2006), in which *E. trachycaulus* was found to have the highest genetic diversity when compared to *E. caninus, E. alaskanus,* and *E. mutabilis* using RAPD. High genetic diversity within *E. trachycaulus* was also found by Díaz et al. (2000) and Zhang et al. (2002), in which *E. trachycaulus* was compared to *E. fibrosus* and *E. alaskanus.*

When the π values for *Elymus* were compared to the π values for *Pseudoroegneria* determined by Sun et al. (2007), *Pseudoroegneria* had over twice the π value. A higher genetic diversity in *Pseudoroegneria* does not agree with current research (Barrier et al. 2001; Sun et al. 2007). Sun et al. (2007) suggested that *Elymus* is likely more genetically diverse than the diploid ancestor *Pseudoroegneria* as polyploids evolve faster than diploids. In addition, a polyphyletic origin within *Elymus* would lead to the species to evolve multiple times and thus increase diversity. A possible explanation for

the differences in genetic diversity within *Elymus* could be due to the differences in experimental design. In our study, five *Elymus* species were selected, and multiple accessions within each species were analyzed as it seeks to analyze both intra- and inter-specific relationships of *Elymus*. In the study of Sun et al. (2007), 20 species of *Elymus* were examined, with single accessions used for each species. Therefore, a decrease in genetic diversity should be expected as genetic diversity is higher among species than within species.

4.2 Phylogenetic analysis

The **St** genome within *Elymus* is significant as several molecular studies have stated that *Pseudoroegneria* was the maternal genome donor during the speciation of *Elymus* (Jones et al. 2000; Mason-Gamer et al. 2002; Redinbaugh et al. 2000; McMillan and Sun 2004; Xu and Ban 2004; Liu et al. 2006a). Our phylogenetic study formed a clade with the **St** genome sequences of the *Elymus* species studied and *Pseudoroegneria,* providing support for *Pseudoroegneria* being a genome donor genus to *Elymus.* All *Pseudoroegneria* species sequenced were placed outside of the subclades within the **St** clade. It is difficult to figure out which *Pseudoroegneria* species examined here is likely the direct ancestor to the studied *Elymus* species. *Agropyron cristatum* was found inside of the **St** clade, along with *Pseudoroegneria*, and the **St** genome sequences of *Elymus.* This finding is likely due to the close association between *Pseudoroegneria* and *Agropyron* (Wang et al. 1985). Wang et al. (1985) first reported successful chromosome pairing from the cross-fertilization between *Agropyron* and *Pseudoroegneria.* Subsequently, introgression between the two genera is possible, which can lead to an increase in genetic similarities to develop. As a result of introgression, a loss of

phylogenetic distinctions is experienced between the two species, despite their different evolutionary histories (Rhymer and Simberloff 1996).

The second major clade within our phylogenetic analysis was between the **H** genome sequences of *Elymus* and *Hordeum,* providing evidence that *Hordeum* is a genome donor genus to *Elymus*. In a subclade formed in the **H** clade, *H. bogdanii* was found with the **H** genome sequences from some *E. caninus,* and all of *E. fibrosus and E. mutabilis*. A close association between members of *Elymus* and *H. bogdanii* has been reported several times within literature (Dewey 1971; Yan 2007; Yan and Sun 2012; Wu et al. 2015). The phylogenetic analysis here suggests *H. bogdanii* as a possible genome donor species of *E. fibrosus* and *E. mutabilis* and some accessions of *E. caninus.* The phylogenetic analysis also indicates a close association between *E. fibrosus* and *E. mutabilis,* as they likely share the same **H** genome donor.

Several accessions were found outside of the **H** clade, including PI 232147, Gr9718, PI 564933, PI 440098, and PI 440101. All of the listed accessions were either members of *E. alaskanus* or *E. trachycaulus.* A possible explanation for that finding could be due to the introgression between the **H** and the **St** chromosomes within *Elymus.* As a result, the **H** genome within these accessions could have become genetically distinct from their genome donor genera. Several studies on *Brassica napus* reported this phenomenon (Pires et al. 2004; Udall et al. 2005; Leflon et al. 2006; Liu et al. 2006b; Nicolas et al. 2007). *B. napus* is an allopolyploid, and it is reported that the genetic exchange among the homoeologous chromosomes may result in genetic asymmetry and convergent evolution of the two genetically distinct parental genomes (Gaeta et al. 2007). Other possible explanations could include that a *Hordeum*-like species donate **H** genome within *E. trachycaulus* and *E. alaskanus*, or that the **H** genome within these accessions is a result of

multiple or subsequent hybridizations, leading to an **H** genome radically different from any genome donor genera.

4.3 Conclusions

The phylogenetic relationship among the five *Elymus* species *E. caninus, E. fibrosus, E. mutabilis, E. trachycaulus* and *E. alaskanus* was successfully analyzed using the *RPB2* gene. The phylogenetic analysis proved a complex and diverse evolutionary history among these five *Elymus* species. The nucleotide diversity of the *Elymus* species, as well as their genome donor genera, was successfully determined. Further evidence was provided that *Pseudoroegneria* (**St**) and *Hordeum* (**H**) are genome donors of the five *Elymus* species studied. A close association between *E. fibrosus* and *E. mutabilis* was found within the **H** genome. The phylogenetic tree also suggests *H. bogdanii* as a potential **H** genome donor for *E. caninus, E. fibrosus, and E. mutabilis.* However, further investigation of the relationships among and between these species is required to form more consolidated conclusions.

Future investigations of the **St** and **H** genome sequences of the *Elymus* species studied at other nuclear genes should be analyzed to support the findings of this study. Different *Pseudoroegneria* species than those used in this study should be analyzed to determine the origin of the **St** genome within the *Elymus* species studied. Further research is required to identify possible **H** genome donor species for *E. alaskanus, E. trachycaulus,* and *E. caninus.*

5. References

Albertin W, Marullo P. 2012. Polyploidy in fungi: evolution after whole-genome duplication. Proc Biol Sci. 279(1738):2497-2509.

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP. 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot. 120(2):183-194.

Alobeid AK. 2017. Molecular phylogeny of *Elymus alaskanus* complex [honors thesis]. [Halifax (NS)]: Saint Mary's University.

Barkworth ME. 1992. Taxonomy of the Triticeae: a historical perspective. Hereditas. 116:1-14.

Barkworth ME. 1994. The *Elymus trachycaulus* complex in North America: more question than answer. In: Wang RRC, Jensen LB, Jaussi C, editors. Proceedings of the 2nd international Triticeae symposium. Utah: Logan Press. p. 189-198.

Barrett SCH, Kohn JR. 1991. Genetic and evolutionary consequences of small population size in plants: implications for conservation. In: Falk DA, Holsinger KE, editors. Genetics and conservation of rare plants. New York (NY): Oxford University Press. p. 3-30.

Barrier M, Robichaux RH, Purugganan MD. 2001. Accelerated regulatory gene evolution in an adaptive radiation. Proc Natl Acad Sci U S A. 98(18):10208-10213.

Baum BR, Yen C, Yang JL. 1991. *Roegneria*: its generic limits and justification for its recognition. Can J Bot. 69:282-294.

Baum D. 2008. Reading a phylogenetic tree: the meaning of monophyletic groups. Nat Educ. 1(1).

Bothmer RV, Salomon B. 1994. Triticeae: a tribe for food, feed and fun. In: Wang RRC, Jensen KB, Jaussi C, editors. Proceedings of the 2nd international Triticeae symposium. Utah: Logan Press. p. 1-12.

Brown TA. 2002. Genomes. 2nd ed. Oxford: Wiley-Liss.

Cody WJ. 1996. Flora of the Yukon Territory. Ottawa (ON): NRC Research Press.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell. 12(9):1551-1567.

Crane CF, Carman JG. 1987. Mechanisms of apomixis in *Elymus rectisetus* from east Australia and New Zealand. Amer J Bot. 74(4):447-496.

Daley T. 2006. *RPB2* gene: a useful molecular marker for phylogenetic study of *Elymus* species [honours thesis]. [Halifax (NS)]: Saint Mary's University.

Denton AL, McConaughy BL, Hall BD. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. Mol Biol Evol. 15(8):1082-1085.

Dewey DR. 1984. The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP, editor. Gene manipulation in plant improvement. Boston (MA): Springer. p. 209-279.

Dewey DR. 1982. Genomic and phylogenetic relationships among North American perennial Triticeae grasses. In: Estes JE, Tyrl RJ, Brunken JN, editors. Grasses and grasslands: systematics and ecology. Oklahoma: University of Oklahoma Press. p. 51-88.

Dewey DR. 1975. Introgression between *Agropyron dasystachyum* and *A. trachycaulum.* Bot Gaz. 136(1):122-128.

Dewey DR. 1971. Synthetic hybrids of *Hordeum bogdanii* with *Elymus canadensis* and *Sitanion hystrix*. Amer J Bot. 58(10):908-1971.

Díaz O. 1999. Genetic diversity in *Elymus* species (Triticeae) with emphasis on the Nordic region [dissertation]. Sweden: Swedish University of Agricultural Sciences.

Díaz O, Salomon B, Bothmer RV. 1998. Description of isozyme polymorphisms in *Elymus* species using starch gel electrophoresis. In: Jaradat AA, editor. Triticeae III. Enfield (NH): Science Publishers. p. 199-207.

Díaz O, Salomon B, Bothmer RV. 1999a. Genetic variation and differentiation in Nordic populations of *Elymus alaskanus* (Scrib. ex Merr.) Löve (Poaceae). TAG. 99(1/2):210-217.

Díaz O, Salomon B, Bothmer RV. 1999b. Genetic diversity and structure in populations of *Elymus caninus* (L.) L. (Poaceae). Hereditas. 131(1):63-74.

Díaz O, Salomon B, Bothmer RV. 2004. Absence of genetic variation in Scandinavian populations of *Elymus mutabilis* (Drob.) Tzel. (Poaceae). Hereditas. 131(1):83-86.

Díaz O, Sun GL, Salomon B, Bothmer RV. 2000. Levels and distribution of allozyme and RAPD variation in populations of *Elymus fibrosus* (Poaceae). Genet Resour Crop Evol. 47:11-24.

Dong YS, Zhou RH, Xu SJ, Li LH, Cauderon Y, Wang RR-C. 1992. Desirable characteristics in perennial Triticeae collected in China for wheat improvement. Hereditas. 116:175-178.

Dong ZZ, Fan X, Sha L-N, Wang Y, Zeng J, Kang H-Y, Zhang H-Q, Wang X-L, Zhang L, Ding C-B, et al. 2015. Phylogeny and differentiation of the St genome in *Elymus* L. sensu lato (Triticeae; Poaceae) based on one nuclear DNA and two chloroplast genes. BMC Plant Biol. 15(1):179-193.

Eilam T, Anikster Y, Millet E, Manisterski J, Feldman M. 2010. Genome size in diploids, allopolyploids, and autopolyploids of Mediterranean Triticeae. J Bot. 2010:1-12.

Falk DA. 1992. From conservation biology to conservation practice: strategies for protecting plant diversity. In: Fiedler PL, Jain SK, editors. Conservation biology. Boston (MA): Springer. p. 397-431.

Farris JS. 1989. The retention index and the rescaled consistency index. Cladistics. 5(4):417-419.

Feldman M, Levy AA. 2009. Genome evolution in allopolyploid wheat-a revolutionary reprogramming followed by gradual changes. J Genet Genomics. 36(9):511-518.

Feldman M, Levy AA. 2012. Genome evolution due to allopolyploidization in wheat. Genetics. 192(3):763-774.

Feldman M, Levy AA, Fahima T, Korol A. 2012. Genomic asymmetry in allopolyploid plants: wheat as a model. J Exp Bot. 63(14):5045-5059.

Feldman M, Liu B, Segal G, Abbo S, Levy AA, Vega JM. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics. 147(3):1381-1387.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39(4):783-791.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. Genetics. 133(3):693-709.

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007. Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell. 19(11):3403-3417.

Gaudett M, Salomon B, Sun G. 2005. Molecular variation and population structure in *Elymus trachycaulus* and comparison with its morphologically similar *E. alaskanus*. Plant Syst Evol. 250(1/2):81-91.

Gibson TJ, Spring J. 1998. Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. Trends Genet. 14(2):46-49.

Given DR. 1994. Principles and practice of plant conservation. London (UK): Chapman and Hall.

Guo H, Wang X, Gundlach H, Mayer KFX, Peterson DG, Scheffler BE, Chee PW, Paterson AH. 2014. Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). Genetics. 197(4):1153-1163.

Halliburton R. 2004. Introduction to population genetics. Upper Saddle River (NJ): Prentice Hall.

Hamrick JL, Godt MJW. 1990. Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS, editors. Plant population genetics, breeding and genetic resources. Sunderland (MA): Sinauer Associates Inc. p. 43-63.

Hamrick JL, Godt MJW. 1997. Allozyme diversity in cultivated crops. Crop Sci. 37(1):26-39.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. J Hered. 105:795-809.

Hartl DL, Clark AG. 1997. Principles of population genetics. 4th ed. Sunderland (MA): Sinauer Associates Inc.

Helfgott M, Mason-Gamer RJ. 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. Syst Bot. 29(4):850-861.

Hitchcock AS. 1969. Gramineae. In: Hitchcock CL, Cronquist A, Ownbey M, Thompson JW, editors. Vascular plants of the Pacific Northwest. Seattle (WA): University of Washington Press.

Hitchcock AS, Chase A. 1950. Manual of the grasses of the United States. 2nd ed. Washington: Government Printing Office.

Invitrogen. 2012. pCR™8/GW/TOPO® TA Cloning® Kit User Guide. Burlington (ON): Life Technologies.

Jones TA, Redinbaugh MG, Zhang Y. 2000. The western wheatgrass chloroplast genome originates in *Pseudoroegneria*. Crop Sci. 40:43-47.

Jozwik FX. 1966. A biosystematics analysis of the slender wheatgrass complex [dissertation]. [Laramie (WY)]: University of Wyoming.

Kawashima T. 2019. Comparative and evolutionary genomics. In: Ranganathan S, Nakai K, Schönbach C, Gribskov M, editors. Encyclopedia of bioinformatics and computational biology. Netherlands: Elsevier Inc. p. 257-267.

Kilian B, Martin W, Salamini F. 2010. Genetic diversity, evolution and domestication of wheat and barley in the Fertile Crescent. In: Glaubrecht M, editor. Evolution in action. Berlin: Springer. p. 137-166.

Knapp EE, Rice KJ. 1996. Genetic structure and gene flow in *Elymus glaucus* (blue wildrye): Implications for native grassland and management. Restor Ecol. 4(1):1-10.

Kolodziej PA, Woychik N, Liao SM, Young RA. 1990. RNA polymerase II subunit composition, stoichiometry, and phosphorylation. Mol Cell Biol. 10(5):1915-1920

Kostina EV, Agafonov AV, Salomon B. 1998. Electrophoretic properties and variability of endosperm proteins in *Elymus caninus* (L.). In: Jaradat AA, editor. Triticeae III. Enfield (NH): Science Publishers. p. 265-272.

Larkin R, Guilfoyle T. 1993. The second largest subunit of RNA polymerase II from

Arabidopsis thaliana. Nucl Acids Res. 21(4):1038.

Larson SR. Jones TA, McCracken CL. Jensen KB. 2003. Amplified fragment length polymorphism in *Elymus elymoides, Elymus multosetus,* and other *Elymus* taxa. Can J Bot. 81:789-804.

Lawrence T. 2011. Inheritance of a dwarf character in Russian wildrye grass, *Elymus junceus*. Genome. 9(1):126-128.

Lee HS, Chen ZJ. 2001. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc Natl Acad Sci. 98(12):6753-6758.

Leflon M, Eber F, Letanneur JC, Chelysheva L, Coriton O, Huteau V, Ryder CD, Barker G, Jenczewski E, Chèvre AM. 2006. Pairing and recombination at meiosis of *Brassica rapa* (AA) x *Brassica napus* (AACC) hybrids. Theor Appl Genet. 113(8):1467-1480.

Levy AA, Feldman M. 2004. Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization: reprogramming of the wheat genome. Biol J Linn Soc. 82(4):607-613.

Liao D. 1999. Concerted evolution: molecular mechanism and biological implications. Am J Hum Genet. 64(1):24-30.

Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM. 1997. Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. EMBO J. 16(3):588-598.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25(11):1451-1452.

Lindsay GC. 1995. DNA content of nuclei of *Actinidia deliciosa* var. *deliciosa* 'Hayward' (kiwifruit) analysed by image cytometry. New Zeal J Bot. 33(2):243-255.

Liu WZ, Wang RR-C, Carman JG. 1994. Hybrids and backcross progenies between wheat (*Triticum aestivum* L.) and apomictic Australian wheatgrass [*Elymus rectisetus* (Nees in Lehm.) A. Löve & Connor]: karyotypic and genomic analyses. Theor Appl Genet. 89(5):599-605.

Liu Q, Ge S, Tang H, Zhang X, Zhu G, Lu BR. 2006a. Phylogenetic relationships in

Elymus (Poaceae: Triticeae) based on the nuclear ribosomal transcribed spacer and chloroplast trnL-F sequences. New Phytol. 170(2):411-420.

Liu ZQ, Adamczyk K, Manzanares-Dauleux M, Eber F, Lucas MO, Delourme R, Chèvre AM, Jenczewski E. 2006b. Mapping PrBn and other quantitative trait loci responsible for the control of homoeologous chromosome pairing in oilseed rape (*Brassica napus* L.) haploids. Genetics. 174(3):1583-1596.

Löve A. 1984. Conspectus of the Triticeae. Feddes Repert. 95:425-521.

Lu BR. 1993. Biosystematic investigations of Asiatic wheatgrasses: *Elymus* L. (Triticeae: Poaceae) [dissertation]. [Sweden]: Swedish University of Agricultural Sciences.

Lu BR, Ellstrand N. 2014. World food security and the tribe Triticeae (Poaceae): Genetic resources of cultivated, wild, and weedy taxa for crop improvement. J Syst Evol. 52(6):661-666.

Maddison WP, Donoghue MJ, Maddison DR. 1984. Outgroup analysis and parsimony. Syst Zool. 33(1):83-103.

Malte MO. 1932. The so-called *Agropyron caninum* (L.) Beauv. of North America. Bull Nat Mus Canada. 68:27-57.

Mason-Gamer RJ. 2013. Phylogeny of a genomically diverse group of *Elymus* (Poaceae) allopolyploids reveals multiple levels of reticulation. PLoS One. 8(11):e78449.

Mason-Gamer RJ, Burns MM, Naum M. 2010. Reticulate evolutionary history of a complex group of grasses: phylogeny of *Elymus* StStHH allotetraploids based on three nuclear genes. PLoS One. 5(6):e10989.

Mason-Gamer RJ, Orme NL, Anderson CM. 2002. Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. Genome. 45:991-1002.

McGuire PE, Dvôrák J. 1981. High salt-tolerance potential in wheat grasses. Crop Sci. 21(5):702-705.

McMillan E, Sun GL. 2004. Genetic relationships of tetraploid *Elymus* species and their

genomic donor species inferred from polymerase chain reaction-restriction length polymorphism analysis of chloroplast gene regions. Theor Appl Genet. 108(3):535-542.

Melderis A. 1978. Taxonomic notes on the tribe Triticeae (Gramineae), with special reference to the genera *Elymus* L. sensu lato, and *Agropyron* Gaertner sensu lato. Bot J Linn Soc. 76:368-384.

Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. 1st ed. Oxford (UK): Oxford University Press.

Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme R, Chalhoub B, et al. 2007. Homoeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of *Brassica napus* haploids. Genetics. 175(2):487-503.

Nicholas KB, Nicholas HB. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors.

Pires JC, Zhao JW, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC. 2004. Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). Biol J Linn Soc. 82(4):675-688.

Promega. 2011. Wizard® Plus Minipreps DNA Purification System Technical Bulletin. Madison (WI): Promega Corporation.

Redinbaugh MG, Jones TA, Zhang Y. 2000. Ubiquity of the St chloroplast genome in St-containing Triticeae polyploids. Genome. 43(5):846-852.

Rhymer JM, Simberloff D. 1996. Extinction by hybridization and introgression. Annu Rev Ecol Syst. 27(1):83-109.

Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4):406-425.

Sanders TB, Hamrick JL and Holden LR. 1979. Allozyme variation in *Elymus canadensis* from the tallgrass prairie region: Geographic variation. Am Midl Nat. 101(1):1-12.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Crit Rev Biochem Mol Biol. 37(3):121-147.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell. 13(8):1749-1759.

Small RL, Cronn RC, Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Aust Syst Bot. 16:145-170.

Stevens L, Salomon B, Sun GL. 2007. Microsatellite variability and heterozygote excess in *Elymus trachycaulus* populations from British Columbia in Canada. Biochem Syst Ecol. 35(11):725-736.

Sun GL. 2002. Interspecific polymorphism at non-coding regions of chloroplast, mitochondrial DNA and rRNA *IGS* region in *Elymus* species. Hereditas. 137(2):119-124.

Sun GL, Daley T, Ni Y. 2007. Molecular evolution and genome divergence at *RPB2* gene of the St and H genome in *Elymus* species. Plant Mol Biol. 64(6):645-655.

Sun GL, Díaz O, Salomon B, Bothmer RV. 1999. Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD and microsatellite markers. Genome 42:420-431.

Sun GL, Díaz O, Salomon B, Bothmer RV. 2001. Genetic diversity and structure in a natural *Elymus caninus* population from Denmark based on microsatellite and isozyme analyses. Plant Syst Evol. 227(3):235-244.

Sun GL, Komatsuda T. 2010. Origin of the Y genome in *Elymus* and its relationship to other genomes in Triticeae based on evidence from elongation factor G (*EF-G*) gene sequences. Mol Phylogenet Evol. 56(2):727-733.

Sun GL, Li W. 2006. Molecular diversity of species of the *Elymus trachycaulus* species complex and their relationships to non-North American taxa. Plant Syst Evol. 256(1):179-191.

Sun GL, Ma X. 2009. Nucleotide diversity and minisatellite in chloroplast Asp(GUC)-Thr(GGU) region in *Elymus trachycaulus* complex, *Elymus alaskanus* and *Elymus caninus*. Biochem Syst Ecol. 37(2):67-75.

Sun GL, Ni Y, Daley T. 2008. Molecular phylogeny of *RPB2* gene reveals multiple

origin, geographic differentiation of H genome, and the relationship of the Y genome to other genomes in *Elymus* species. Mol Phylogenet Evol. 46(3):897-907.

Sun GL, Salomon B. 2003. Microsatellite variability and heterozygote deficiency in the artic-alpine Alaskan wheatgrass (*Elymus alaskanus)* complex. Genome. 46(5):729-737.

Sun GL, Salomon B, Bothmer RV. 2002. Microsatellite polymorphism and genetic differentiation in three Norwegian populations of *Elymus alaskanus* (Poaceae). Plant Syst Evol. 234:101-110.

Sun GL, Salomon B, Bothmer RV. 1997. Analysis of tetraploid *Elymus* species using wheat microsatellite markers and RAPD markers. Genome. 40(6):806-814.

Sun GL, Salomon B, Bothmer RV. 1998. Characterization and analysis of microsatellite loci in *Elymus caninus* (Triticeae: Poaceae). Theor Appl Genet: 96:676-682.

Sun GL, Tang H, Salomon B. 2006. Molecular diversity and relationships of North American *Elymus trachycaulus* and the Eurasian *E. caninus* species. Genetica. 127:55-64.

Sun GL, Zhang X. 2011. Origin of the H genome in StH-genome *Elymus* species based on single copy nuclear gene *DMC1*. Genome. 54(8):655-662.

Svitashev S. 1997. Analysis of Triticeae genomes using DNA markers. Acta Univ Agric Suec. 61.

Sweetster D, Nonet M, Young RA. 1987. Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci U S A. 84(5):1192-1196.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other materials) V. 40b10. Sunderland (MA): Sinauer Associates.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3):585-595.

Thompson JD, Gibson TJ, Higgins DG. 2003. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 00(1):2.3.1-2.3.22.

Tzvelev NN. 1976. Tribe 3. Triticeae Dumort. In: Fedorov AA, editor. Poaceae USSR.

Russia: Nauka Publishing House. p. 147-181.

Tzvelev NN. 1977. O nekorotih adventinih rastenijah Lemingradskoy oblasti [about some adventive plants of the Leningrad region]. Novit Syst Pl Vase. 14:224-255.

Tzvelev NN. 1989. The system of grasses (Poaceae) and their evolution. Bot Rev. 55(3):141-204.

Udall JA, Quijada PA, Osborn TC. 2005. Detection of chromosomal rearrangements derived from homoeologous recombination in four mapping populations of *Brassica napus* L. Genetics. 169(2):967-979.

Wang RR-C, Dewey DR, Hsiao C. 1985. Intergeneric hybrids of *Agropyron* and *Pseudoroegneria*. Bot Gaz. 146(2):268-274.

Watterson GA. 1975. On the number of segregation sites in genetical models without recombination. Theor Popul Biol. 7(2):256-276.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol Biol. 42(1):225-249.

Woodhouse M, Burkart-Waco D, Comai L. 2009. Polyploidy. Nat Educ. 2(1):1.

Wu DC, He DM, Gu HL, Wu PP, Yi X, Wang WJ, Shi HF, Wu D-X, Sun G. 2016. Origin and evolution of allopolyploid wheatgrass *Elymus fibrosus* (Schrenk) Tzvelev (Poaceae:Triticeae) reveals the effect of its origination on genetic diversity. PLoS One. 11(12):e0167795.

Wu P-P, Zuo H-W, Sun G, Wu D-X, Salomon B, Hu Q-W, Dong Z-R. 2015. Comparison of gene flow among species that occur within the same geographic locations versus gene flow among populations within species reveals introgression among several *Elymus* species. J Syt Evol. 54(2):152-161.

Xu DH, Ban T. 2004. Phylogenetic and evolutionary relationships between *Elymus humidus* and other *Elymus* species based on sequencing of non-coding regions of cpDNA and AFLP of unclear DNA. Theor Appl Genet. 108(8):1443-1448.

Yan N. 2007. Molecular diversity and relationship of Elymus species [master's thesis]. [Halifax (NS)]: Saint Mary's University.

Zhang X-Q, Salomon B, Bothmer RV. 2002. Application of random amplified

polymorphic DNA markers to evaluate intraspecific genetic variation in the *Elymus alaskanus* complex (Poaceae). Genet Resour Crop Ev. 49(4):399–409.

Zuo H, Wu P, Wu D, Sun G. 2015. Origin and reticulate evolutionary process of wheatgrass *Elymus trachycaulus* (Trticeae: Poaceae). PLoS One. 10(5):e0125417.