Molecular diversity and relationship of *Elymus* species

Yan Ni

This thesis has been submitted to the Saint Mary's University

Department of Biology in partial fulfillment of the requirements

for a Degree of Master of Science

Saint Mary's University, Halifax, Nova Scotia, Canada

October 1, 2007

Your file Votre référence
ISBN: 978-0-494-35773-6
Our file Notre référence
ISBN: 978-0-494-35773-6

Certification

Molecular Diversity and Relationship of *Elymus* Species

by

Yan Ni

A Thesis Submitted to Saint Mary's University, Halifax, Nova Scotia,
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Applied Science

September 28, 2007

Examining Committee:

Approved: Dr. Gavin Kernaghan, External Examiner
Department of Biology, Mount Saint Vincent University

Approved: Dr. Genlou Sun, Senior Supervisor
Department of Biology

Approved: Dr. Zhongmin Dong, Supervisory Committee
Department of Biology

Approved: Dr. Harry Murray, Supervisory Committee
Adjunct Professor

Approved: Dr. Hugh Broders, Program Co-ordinator Rep

Approved: Dr. Pawan Lingras, Graduate Studies Rep

SIGNATURES OF THE EXAMINING COMMITTEE

This thesis has been examined and approved by the following members of The Saint Mary's University Department of Biology:

Dr. G. Sun: _____

Dr. Z. Dong: _____

Dr. H. Murray: _____

Molecular diversity and relationship of Elymus species

By

Yan Ni

October 1, 2007

ABSTRACT

Classification of the tribe *Triticeae* has been a longstanding problem, especially the treatment of the genus *Elymus*. The taxonomy of *Elymus* is extremely complex because of the huge morphological variation within and between species. Previous studies suggested that the **StH** genome Elymus species originated from *Pseudoroegneria* (**St**) and *Hordeum* (**H**). In this study, we have conducted the phylogenetic analysis of *Elymus* species with one nuclear and two chloroplast DNA sequence data sets. Chloroplast data further supports that *Pseudoroegneria* as the maternal donor of the tetraploid *Elymus* **St** genome. This paper is also the first report microsatellites in the chloroplast *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region, which is highly conserved (TGAAAGAAA).

The nuclear RPB2 gene was used to examine the nuclear diversity of tetraploid *Elymus* and their diploid donor. The results indicated that the RPB2 gene may evolve faster in the polyploid species than in diploid. This study found a 39bp MITE *stowaway* element insertion in the region of the RPB2 gene for all tetraploid *Elymus* **St** genome species and also the diploid *Pseudoroegneria spicata* and *P. stipifolia* **St** genomes. The sequences on the 3'- end are highly conserved, with AGTA in all sequences but 10339 (*E. fibrosis*), in which the AGTA is replaced with AGAA. The 5'-end sequence of GGTA is changed to AGTA or deleted, resulting in Stowaway excition in the H genome of *Elymus* species. The three transposable element indels have occurred prior to polyploidization, and shaped the homoelogous RPB2 loci in **St** and **H** genome of Elymus species.

ACKNOWLEDGEMENTS

I would like to express my sincerest thanks to Dr. Genlou Sun who has been a supervisor, a constant support and a relative for me. Without his financial support, I would not able to pursue my graduate study; without his time, knowledge and encouragement, this research would not have been possible. He taught me the molecular techniques hand by hand and he always discusses the problems with me with great patience. I also acknowledge my supervisor committee members, Dr. Zhongming Dong, Dr. Harry Murray for their excellent academic guidance and expertise.

I would like to thank the staff members of the Biology Department at Saint Mary's University for their generous help with both my study and life: Carmen Cranley, Heidi de Boer and Janet White. I would like to express my gratitude to Russell Easy (NRC) Tracy Delay, Michael Foote, Stanley King and Amber Golding for their friendly support.

Finally, I thank my family for their continued support, especially my mother; her belief and confidence in me helped me through all the difficulties.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. Introduction

It is difficult to interpret allopolyploid taxa within phylogenies of their diploid relatives. There is no common agreement as to where to place hybrid taxa within the bifurcating trees, although many studies have been done (Mason-Gamer et al. 2002; Mason-Gamer 2004). Plant systematics, which combines traditional phylogenetic methods and molecular data, provides both an intellectual framework and tools to understand the relationships and circumscriptions at all levels of the taxonomic hierarchy (Small et al. 2004).

Cytogenetic studies of genome-pairing data were the milestone of the classification of Triticeae (Dewey 1984, Love 1984). Cytogenetic data in *Triticeae* have been widely used in the delimitation of genera and in the determination of the origins of the many polyploids in the tribe, and have a huge influence on the classification within the tribe. However, its utility is limited because it cannot provide insights into the hierarchical relationships among the taxa examined (Petersen and Seberg 1997). Furthermore, the ability of cytogenetic data to uncover phylogenetic history remains controversial and has been questioned on both philosophical and methodological aspect (Seberg and Petersen 1998).

Significant developments in establishing relationships between plant families have occurred in the use of chloroplast DNA (Vedel et al. 1976) and ribosomal RNA (Hamby and Zimmer 1988), which are considered as a starting point for molecular studies that revealed the reticulate histories of several polyploidy species on groups. In the Triticeae, data from both chloroplast DNA (Petersen and Seberg 1998;

Mason-Gamer et al. 2002), nuclear rDNA (Hsiao et al. 1995) and nuclear DNA (Petersen and Seberg 2002) have greatly increased overall understanding of the evolution of the tribe.

In this study, two set of chloroplast DNA sequence data sets and one nuclear DNA sequence set were used to investigate the phylogenetic relationships between tetraploid **StH** genome *Elymus* species. This paper briefly reviewed the historical and current situation of tribe *Triticeae* and *Elymus* species; then focused on the **StH** genome *Elymus* species, describing its phylogenetic evolution. Finally, it reviewed the use of chloroplast DNA and nuclear DNA in phylogenetic studies, especially the use of *RPB2* gene.

1.1 The definition of tribe *Triticeae*

Tribe *Triticeae*, combining a wide variety of biological mechanisms and genetic systems, is of the utmost interest for basic research, especially studies of evolution, genetics, genetic diversity, taxonomy, and speciation. *Triticeae*, one of the most important tribes in *Poaceae*, contains many economically important cereal crops and forage grasses, such as wheat, barley and rye. The *Triticeae* genera, such as *Elymus*, *Pseudoroegneria* and *Leymus*, as well as the partially perennial genus *Hordeum*, provide an important source of new genetic variation to improve both cereal crops and forage grasses through inter-generic and inter-specific hybridization (Lu 1993). Some species possessing agronomically useful characteristics such as pest resistance, salt and alkalinity tolerance, and stress adaptation are potentially promising sources for

upgrading breeds (Knott 1961, 1989; Tosa and Sasai 1991; Lu 1993). A number of resistance genes from wild relatives of wheat have been transferred into useful cultivars. For example, a mildew-resistance gene from *Hordeum bulbosum L.* was incorporated into barley (*Hordeum vulgare L.*) through hybridization and backcross.

Classification of the tribe *Triticeae* has been a longstanding problem, and there has been no sign of a universally acceptable solution until now. Inter-generic hybridization, along with polyploid origin and ambiguous morphological characters, has caused taxonomic confusion. Moreover, environmental factors and different human perspective also affect classification of *Triticeae*. Various classification systems have been proposed by different taxonomists with different criteria, geographical and historical perspectives. Therefore, to fully understand *Triticeae,* its classification must first be clearly defined.

1.2 *Elymus* and genomic constitution of the species

The present studies mainly focus on *Elymus*, an entirely polyploid genus of *Triticeae* that contains about 150 species, which occur abundantly in Asia and North America. According to the genomic classification of Löve (1984) and Dewey (1984), *Elymus* species are perennial wheat grasses and it is the largest genus of the tribe *Triticeae*. It is also the most widespread genus, found on most continents: Europe, Asia, North America, South America, New Zealand and Australia (Helfgott and Mason-Gamer 2004). Most *Elymus* species are acervate, rarely rhizomatous, small-anthered, and self-pollinating (Lu 1993). Although predominately a northern

temperate genus, *Elymus* species occur from the Arctic and temperate to subtropical regions. These species inhabit various ecological niches, including grasslands, semi-desert, mountain slopes and valleys among bushes, and both inside and alongside the edges of forests.

Various classification systems have been proposed by different taxonomists with divergent concepts. In North America, the traditional system of classification is based on morphological characteristics (Hitchcock 1951), in which *Elymus* is treated as a genus including all *Leymus* and *Taeniatherum* species. However, the cytogenetic classification of *Elymus* (Dewey 1984; Löve 1984) is very different from the morphology-based circumscription and is one in which *Elymus* is defined as those allopolyploid species containing one set of **St**-genome from *Pseudoroegneria* (Nevski 1934) and one or more other sets of genomes from other diploid donors. *Elymus* species are explicitly heterogeneous as they usually may combine the **St** genome with the **H** genome from *Hordeum L.*, the **P** genome from *Agropyron Gaertner*, the **W** genome from *Australopyrum* and/or **Y** from an unknown donor, or even with a combination of these genomes in various tetraploid, hexaploid, and, on rare occasions, octoploid configurations (Dewey 1984, Helfgott and Mason-Gamer 2004, Mason-Gamer et al. 2002). However, the utility of cytogenetic tests is also limited because it cannot provide insights into the hierarchical relationships among the taxa examined.

The taxonomy of *Elymus* is extremely complex because of the huge morphological variation within and between species. Levels and patterns of genetic

diversity vary greatly within and among populations and species. This variation reflects the interplay of a myriad of historical factors and evolutionary forces, including external forces and internal genomic and genetic factors. External forces involve natural selection, population size and history, gene flow, and breeding system; Internal factors, include recombination, mutation rate, and gene conversion (Aquadro and Begun 1993; Tajima 1993; Moriyama and Powell 1996).

Molecular techniques create an effective way to study relationships and understand mechanisms that create and maintain genetic differences among evolutionary lineages.

1.3 The StH group of *Elymus* and its phylogenetic evolution

The **StH** genome species are found from the Arctic to the subtropics, from the seacoast to above 5000m in the Himalayas, and their morphology is likewise diverse. Hence, the **StH** group of *Elymus* species, displaying an enormously wide range of adaptation to various climates and habitats, makes it an excellent group for research. Previous studies have shown that despite several **StH** genome species having similar ecological preferences and breeding systems, their genetic structure variation highly deviate and each species possesses a unique pattern of genetic variation. This group of *Elymus* consists of approximately 50 species. Cytogenetic studies of North American *Elymus* suggest that the genus is an allopolyploid derivative of *Pseudoroegneria* and *Hordeum*. The **St** haplome donor, *Pseudoroegneria* contains approximately 16 diploid or tetraploid species distributed in Middle East, central Asia,

northern China and western North America (Watson and Dallwitz 1992). The **H** haplome donor, *Hordeum*, contains approximately 32 species and is distributed in temperate areas of both the northern and on the southern hemispheres.

Several authors have speculated as to whether the *Elymus* tetraplod species with the **StH** genome is a monophyletic or polyphyletic genus (Wang et al. 1986; Svitashev et al. 1998; Sun et al. 1997). In evolutionary events, this corresponds to the hypothesis that the **StH** genomic species arose at once by an ancient hybridization between the parental genera *Pseudoroegneria* and *Hordeum* and then differentiated into a number of species, or the hypothesis that hybridizations have occurred repeatedly at different times and places during evolutionary history. There have been several studies suggesting that the **H** haplome in *Elymus* could be from different sources. Genome (Wang et al. 1986) and karyotype analyses (Linde-Laursen et al. 1994) suggested that the Old and New World taxa may be of separate origin. These studies seem to favour the hypothesis of a multiple origin of the **H** haplome possessing species in the genus, but did not provide strong undisputable evidence to support their claims. In this study, *RPB2* sequence data was used to explore the origin of **StH** genome tetraploid *Elymus* species and its phylogeny, which help us to gain a better understanding of the genus.

1.4 The use of Chloroplast DNA

Chloroplast DNA (cpDNA) is the most widely used source of data in plant molecular phylogenetic analyses. The primary advantages of cpDNA as a molecular tool are its relatively simple genetics. Chloroplast DNA is generally uniparentally

inherited and characterised as structurally stable, haploid, and non-recombinant. Structural stability across large evolutionary scale has been demonstrated by comparative cpDNA mapping and sequencing, and numerous studies have shown that cpDNA molecules are highly conserved with respect to gene content and arrangement, especially among closely related species (Olmstead and Palmer 1994). The haploid nature of cpDNA, which serves to reduce the amount of intraspecific and intrapopulation variation, simplifies analyses. It also reduces the chance of participating in recombination and variation between alleles. The content of cpDNA tends to be highly conserved, thus universal primers has been widely used to amplify DNA fragments.

Following the pioneering work of Vedel et al. (1976) who made the first restriction endonuclease analysis is now called restriction fragment length polymorphism (RFLP); analysis of the chloroplast DNA, the molecular study of the chloroplast genome became more and more popular. Many taxonomists used cpDNA to study evolutional relationship of species. For example, Mason-Gamer (2004) analysed three sets of cpDNA data and identified that the potential maternal genome donors of *Elymus* were likely to be *Pseudorogneria*. This finding was consistent with previous cytogenetic result (Dewey 1984).

Unfortunately, because of its characteristics, cpDNA analysis of hybrid or polyploidy plants may incorrectly identify them as belonging to a clade of one of the two parents, without revealing the hybrid history (Mason-Gamer 2004). If hybridisation is followed by introgression and subsequent fixation of the alien cpDNA,

then the phylogeny may be accurately resolved with regard to the maternal history; however, cpDNA fails to identify the phylogenetic conflict arising from a hybrid ancestry.

Slow rates of sequence divergence of cpDNA have also limited its phylogenetic utilities to high taxonomic levels. Even the non-coding regions of cpDNA, which are more prone to mutation than the coding regions, often cannot distinguish differences at the species level (Small et al. 2004).

1.5 The use of nuclear DNA in phylogenetic study

Nuclear genes have, recently, been used in phylogenetic analysis because of their elevated rate of sequence evolution relative to organellar genes. This elevated evolutionary rate yields a greater efficiency from sequencing effort, since more variation is detected per unit of sequence than in organellar genes (Small et al. 2004).

Biparentally inherited nuclear markers are needed for the reconstruction of allopolyploidization. However, the process of concerted evolution, array expansion/contraction and the presence of paralogous sequence can make isolation of both parental copies difficult. For instance, in tetraploid **StH** genome *Elymus* species, the complete genome consists of two copies of the **St** haplome and two copies of the **H** haplome. Taking two copies on **St** haplome referred to A1 and A2 in species. If concerted evolution occurs which means genes evolve together, A1 and A2 change in sequence together, and they will be the same in phylogenetic studies; if concerted evolution is incomplete or does not occur, A1 and A2 will likely evolve independently.

In this case, A1 might diverge from A2 enough that they can be considered as two copies. Isolation of only one of them can be difficult. If we accidentally isolate A1 in one species and A2 in another species, then the comparison is no longer accurate (Small et al. 2004).

Low-copy and single copy genes are ideal candidates for identifying parental donors of polyploids since they are less frequently subject to concerted evolution (Small et al. 2004), and they have been used to identify the origin of hybrid or polyploidy taxa in a number species (e.g. Doyle et al. 1999; Mason-Gamer 2001).

1.6 The nuclear *RPB2* gene

Nuclear RNA polymerases in eukaryotes belong to three different classes, referred to as RNA polymerase I, II, and III which are responsible for the synthesis of rRNA, mRNA and small stable RNAs. RNA polymerase II (*RPB2*) catalyzes mRNA synthesis in eukaryotic cell nuclei. It is functionally and structurally analogous to the "β subunit" in the RNA polymerase enzyme in prokaryotes. This β subunit has binding sites for the mRNA chain and DNA template that is to be transcribed (Kolodziej et al. 1990).

The second largest subunit of the nuclear RNA polymerase II (*RPB2*) gene provides useful sequence information for phylogenetic analyses. Genomic Southern blotting experiments suggest only one copy of *RPB2* in *Arabidopsis* (Larkin and Guifoyle 1993); Warrilow and Symons (1996) cloned and sequenced cDNA of the *RPB2* gene in tomato and testified that a single-copy gene was transcribed into mRNA;

Denton et al. (1998) confirmed the presence of only one copy in 10 plant species, including the species *Hordeum vulgare*. However, Oxelman and Bremer (2000) demonstrated duplication of the *RPB2* gene copies (*RPB2-d* and *RPB2-I*) occurring in *Mostuea* (Gentianales) and *Kopsia* (Gentianales). The survey of Oxelman et al. (2004) gave evidence that *RPB2-d* is expressed in all the species in asteroid group, while *RPB2-I* is expressed in *Rhododendron*. Although they did find some dicots that had two copies, the monocot that they observed contained only one copy.

2. The purpose of this study

In this study, chloroplast and nuclear DNA sequence data were used to investigate the phylogenetic relationships between tetraploid **StH** genome *Elymus* species. Although many studies have been performed on *Elymus* (For example, Sun et al. 2006; McMillan and Sun 2004; Helfgott and Mason-Gamer 2004; Mason-Gamer et al. 2002), they either focus on the various genome combinations or the relationship within certain inter/intra species. This paper is primarily focused on the **StH** genome species and it is the first study that systematically identifies the relationships of StH genome species and their diploid genome donors. Mason-Gamer et al. (2002) indicated *Pseudoroegneria* was the maternal genome donor of *Elymus* species, but further explorations are still needed to confirm this result since there are many *Pseudoroegneria* species and some of them may not be the donor species.

The purposes of this study were as follows: (1) to confirm and identify the maternal parents of tetraploid **StH** genome *Elymus* species. (2) to analyze the

nuclear diversity and evolution process of *Elymus* species and diploid donors. (3) to generate phylogenetic trees and analyse the molecular relationship of these species.

3. Materials and Methods

3.1 Plant materials

A total of 41 Triticeae accessions were used in this study, including 26 *Elymus* accessions with most of them having **StH** genomic combinations, and 15 accessions of diploid Triticeae species from *Pseudoroegneria* (**St**), *Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), *Lophopyrum* ($\mathbf{E^e}$), *Thinopyrum* ($\mathbf{E^b}$) and *Dasypyrum* (**V**) (Table 1).

All the seeds were provided by the Swedish University of Agricultural Sciences and the Germplasm Resources Information Network (GRIN), and were germinated on absorbent filter paper in Petri dishes. Germinated seeds were transplanted to a sand-peat mixture, and plants maintained in a greenhouse.

Table 1. Taxa used in this study

Species	**Accesion No.**	**Genome**	**Origin**	**GeneBank Accseion No.**	***RPB2* gene**	***TrnT/TrnW* gene**	***TrnS/PsbC* gene**
Ag. cristatum (L.) Gaertn.	PI 383524	P	Kars, Turkey		*	/	/
Aus. retrofractum (Vickery) Á. Löve	PI 533014	W	New South Wales, Australia		*	/	/
	PI 547363	W	New South Wales, Australia		*	*	/
D. villusum (L.) P. Candargy	PI 368886	V	Gaziemir, Turkey		*	/	/
H. bogdanii Wilensky	PI 499498	H	Inner Mongolia, China		*	/	/
	PI 499645	H	Xinjiang, China		*	/	/
H. stenostachys Godr.	H6439	H	Argentina		*	/	/
P. libanotica (Hack.) D. R. Dewey	PI 330688	St	Sirak-Sar, Iran		*	/	/
	PI 330687	St	Kandavan Pass, Iran		*	/	*
	PI 401274	St	Saqqez, Iran		*	/	/
P. spicata (Pursh) Á. Löve	PI 506274	St	Washington, United States		*	/	/
	PI 610986	St	Utah, United States		*	/	/

P. spicata (Pursh) Á. Löve	PI 232140		Montana, USA	/	*	/
	PI506273	St				*
P. stipifolia (Czern. Ex Nevski) Á. Löve	PI 325181	St	Stavropol, Russian Federation	*	/	/
L. elongatum (Host) Á. Löve	PI 142012	E^e	Odessa, Russian Federation	*	/	/
T. bessarabicum (Savul. & Rayss) Á. Löve	PI 531712	E^b	Estonia	*	/	/
E. bakeri (E. E. Nelson) Á. Löve	PI 610985	StH	Colorado, United States	*	/	/
E. coreanus Honda	PI 531578	X_mN_s	Russian Federation	*	/	*
E. lanceolatus (Scribn. & J. G. Sm) Gould	PI 236663	StH	Maryland, United States	*	*	/
E. dentatus (Hook.f.) Tzvelev	PI 628702	StH	Altay, Russian Federation	*	*	*
E. mutabilis (Drobow) Tzvelev	PI 564949	StH	Altay, Russison Federation	*	*	/
E. sibiricus L.	PI 499461	StH	Qilien Mts, Lanzhou, China	*	/	*
E. submuticus (Keng) Á. Löve	PI 499480	??	Qilien Mts., China	*	*	*
E. transbaicalensis (Nevski) TZvelev	H10391	StH	Siberia, Russia	*	*	/
E. wiegandii Fernald	PI 531708	StH	Aylwin, Quebec, Canada	*	/	*
E. trachycaulus (Link) Gould ex Shinners	H3526	StH	Nerungri, Russia	*	*	*
	PI 537323	StH	Utah, United States	*	*	/
E. virescens Piper	H10584	StH	Julianehåb, Greenland	*	*	/

E. virescens Piper	H10586	StH	Julianehåb, Greenland	/	*	/
E. wawawaiensis ined.	PI506262	StH	Washington, United States	*	*	*
E. caninus (L.)L.	H3169	StH	Västmanland, Sweden	*	*	*
E. hystrix L.	H5495	StH	Canada	*	*	*
E. latiglumis J. G. Sm	H10476	StH	Washington, United States	*	/	/
E. glaucus Buckley	PI 232258	StH	United States	*	*	/
E. fibrosis (Schrenk) Tzvelev	H10339	StH	Pelkosniemi, Finland	*	*	/
E. virginicus L.	PI 436946	StH	Oklahoma, United States	*	/	/
E. multisetus (J. G. Sm.) Burtt Davy	W6-20963	StH	California, United States	*	/	/
E. violaceus (Hornem.) Feilberg	H10588	StH	Julianehåb, Greenland	*	*	*
E. scabriglumis (Hack.) Á. Löve	PI 331168	StHH	Argentina	*	/	/
E. abolinii (Drobow) Tzvelev	PI 531554	StY	Xinjiang, China	*	*	*
E. ciliaris (Trin.) Tzvelev	PI 564917	StY	Vladivostock, Soviet Far East	*	*	*
E. semicostatus (Nees ex Steud.) Melderis	PI 207452	StY	Afghanistan	*	*	/
E. tschimganicus (Drobow) Tzvelev	PI 499481	StStY	Inner Mongolia, China	*	/	/
H. bogdanii	PI 531762	H	Ishkashim, Tajikistan	/	*	/
E.drovovii	PI 564926			/	*	/

E.mutisetus	PI 619552			/	*	/
E. batalinii	PI 314623	StPY	Alma Ata, Former Soviet Union	/	*	*
	PI 531620			/	*	/
E. subsecundus	PI 537322		Heber City, Utah, USA	/	*	/
E. alaskanus	H10358		Akureyri, Eyjafjardarsysla, Iceland	/	*	/
E. hyperarcticus	H10613		Tuktoyaktuk airport, Canada (Ka9902)	/	*	/
E. antarcticus	PI 636671	StH		/	/	*
E. dahuricus	PI 628674	StHY	Xinjiang, China	/	/	*
E. interuptus	PI 531617			/	/	*
E. villosus	PI 531703			/	/	*
E. tsukushiensis	PI 531698	StHY	Tokaraskya, Osaka, Japan	/	/	*

*The genome designations are according to wang et al. (1994). The genome constitution in each species refers to Dewey (1982), Jensen (1993), Jensen et al. (1994), and Agafonov (2004).

* PI: Accession number used in Germplasm Resources Information Network (GRIN) of the United States Department of Agriculture (USDA); H: Accession number used in the Swedish University of Agricultural Sciences

3.2 DNA isolation

DNA was extracted from young freeze-dried leaf tissue collected from 5-10 plants of each accession using the method of Junghans and Metzlaff (1990) with some modifications according to Sun et al. (1997). After freezing in liquid nitrogen, leaves were ground with a mortar and pestle. The powder was distributed to 2ml tube and 750µl of lysis buffer (50mM Tris-HCl (pH 7.6), 100mM NaCl, 50mM EDTA (ethylenediaminetetraacetic acid), 0.5% SDS (sodium dodecyl sulphate) and 10mM β-mercaptoethanol) was added. After 10 minutes of lysis, 450µl phenol-tris-chloroform (pH 7.5) was added. Tubes were inverted, 450µl of (24:1) chloroform-isoamyl alcohol was added, inverted again, and centrifuged at 13000RPM for 5 minutes. The upper phase was transferred to a new 2mL tube. Another aliquot of chloroform-isoamyl alcohol was added, followed by centrifugation. The upper phase was transferred to a 1.5 mL tube and to that; about 1mL of cold isopropanol was added. The tube was incubated for 20 minutes at -20°C. Tubes were centrifuged at 13,000RPM for 20 minutes at 4°C. The pellet was washed twice with 500µl aliquots of 70% cold ethanol. The pellet was air dried for 10 minutes and dissolved in 400µl of TE buffer (10mM Tris-HCl (pH7.5), 1mM EDTA (pH 8.0)). After incubation at room temperature for 24 hours, 50µg/ml RNase was added to the tube. The tube was incubated at 37°C for 30 minutes. Then the steps were repeated again, from the stage of addition of phenol-tris-chloroform through to resuspension in TE Buffer. Only 100µl of TE were added in this second stage.

3.3 DNA amplification

The *RPB2* gene, was amplified with the primer pair P6F/P6FR (Denton et al. 1998; Sun et al. 2007) (Table 2.). Amplification of DNA was carried out in 20 μL reaction mixture containing 30 ng template DNA, 0.2μM of each primer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1 unit *Taq* DNA polymerase (Biolabs, New England) and distilled deionized water to the final volume 20μL. The reaction mixture was amplified using Techne Genius thermal cycler with following protocol: one cycle of 5 min at 95°C, 35 cycles of 1 min at 95°C , 2 min at 52°C, 2 min at 72°C, followed by 10 min at 72°C.

Asp (GUC)-The (GGU) intergenic region were amplified with the primer pair TrnD/TrnTW (Sun et al. 2002). Reaction mixture was the same as described above. PCR reactions were carried using Techne Genius thermal cycler with following protocol: one cycle of 4 min at 95°C, 40 cycles of 30 s at 94°C , 30 s min at 54°C, 1 min 45 s at 72°C, followed by 10 min at 72°C.

The *trnS* gene [tRNA – Ser (UGA)] and the adjacent *psbC* gene (PS II 44kDa) were amplified with the primer pair *trn*S/*psb*C (Demesure et al. 1995). The reaction mixture was the same as described above and amplified using the Techne Genius thermal cycler with following protocol: one cycle of 4 min at 94°C, 35 cycles of 1 min at 94°C , 2 min at 62°C, 2 min at 72°C, followed by 10 min at 72°C.

Table 2. Primer detail

Amplified gene	Primer name	Nucleotide sequence	Reference
RPB2 gene	P6F	5'-TGGGGAATGATGTGTCCTGC-3'	Denton et al. 1998
	P6FR	5'-CGAACCACACCAACTTCAGTGT-3'	Sun et al. 2007
Asp(GUC)-The (GGU)	*trn*D	5'-ACCAATTGAACTACAATCCC-3'	Sun, 2002
region	*trn*T	5'-CCCTTTTAACTCAGTGGTAG-3'	
trnS[tRNA-ser	*trn*S	5'-GGT TCG AAT CCC TCT CTC TC-3'	Demesure et al. 1995
(UGA)]-psbC[spII44kd	*psb*C	5'-GGT CGT GAC CAA GAA ACC AC-3'	

3.4 PCR purification

In order to promote the cloning efficiency, PCR purification was completed using the QIAquick PCR Purification Kit (Qiagen Inc., Mississauga, Ont - Cat. No. 28104). The PCR product was mixed with 100μL of the included buffer, placed in the spin column and centrifuged at 13,000 RPM for 1 minute. Wash buffer (0.75ml)was added to the spin column tube and centrifuged. After discarding flow-through, the tube and spin column were centrifuged again to remove wash buffer residue. The column was transferred to a clean tube and DNA was eluted by adding 20μL distilled deionized water and centrifuging. The purification products are stored at -20°C.

3.5 Cloning and sequencing

The purified PCR products were cloned by the pGEM-T Easy Vector System II kit (Promega Corporation, Madison, WI, USA). The cloning reaction consisted of 2.2μL of the PCR product, 2.5μL reaction buffer, 0.5μL vector and 0.8μL ligase. This reaction mixture was incubated at room temperature for about 1 hour and then placed on ice. Then, all the reaction mixture was added to a tube that contained 35μL of competent cell, stirred and incubated on ice for about 30 minutes. Then, it was placed in a 42°C water bath for 40 seconds. It was transferred, immediately, back to ice. Then, 250μL of S.O.C. medium was added. The tubes were placed in a 37°C water bath for 1 hour, with shaking at 200RPM. About 200μL of the mixture was spread on a pre-warmed plate and incubated at 37°C overnight.

Six to eight white colonies from each cloned PCR product were randomly selected

for screening. Each was transferred to 20μL LB broth with 100μg/ml ampicillin. These solutions were incubated at room temperature for 1 hour. PCR was performed with its amplification primers, using 1μL of solution as template to check if the DNA fragment is successfully insert into plasmid. For those solutions that were confirmed to contain the insert, the rest of the 19μL solution was transferred into 4 mL of LB broth (with 100μg/ml ampicillin) and incubated at 37 °C overnight.

Plasmid DNA was isolated using the GenElute TM Plasmid Miniprep Kit (Sigma) followed the manufactures instructions. DNA was sequenced commercially at Macrogen Inc. (Seoul, Korea) with universal primer pair T7/SP6.

3.6 Data analysis

Automated sequence outputs were corrected visually by comparison with chromatographs. Sequences similar to *RPB2, Asp(GUC)-The(GGU), trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region were identified by BLAST search (http://www.ncbi.nlm.nih.gov/blast/). BLAST 2 SEQUENCES and All-in-One Seq Analyzer (http://www-personal.umich.edu/~ino/blast.html) were used to generate full sequence for each clone.

Multiple nucleotide sequence alignments were carried out using ClustalX (Thompson et al. 2003). Nucleotide diversity was estimated by Tajima's (1989) π and Watterson's (1975) θ statistics. The former measure quantifies the mean percentage of nucleotide differences among all pairwise comparisons for a set of sequences, whereas the latter is simply an index of the number of segregating (polymorphic) sites. The

above calculations were conducted by the software program DnaSP 4.0 (Rozas et al. 2003).

Phylogenetic trees based on these sequences were constructed using the parsimony analysis default setting of the program PAUP* version 4 beta 10 Win (Swofford 2003) and different sequence were taken as an out-group, to be more specific, *Dasypyrum villosum* (accession PI 368886) for phylogenetic trees derived from *RPB2* gene; *Aust. retrofractum* (accession PI 547363) for phylogenetic trees generated from *Asp (GUC)-The (GGU)* intergenetic region and flanking region; *E. coreanus* (accession PI 531578) which is now grouped as *Leymus*, consists of XN genome combination and has been used for phylogenetic trees that generated from *trnS[tRNA-ser(UGA)] – psbC[spII44kd protein]* region. The most parsimonious trees were obtained by performing a heuristic search using the Tree Bisection-Reconnection (TBR) option with MulTrees on, and ten replications of random addition sequence with the stepwise addition option. All characters were specified as unweighted. Multiple parsimonious trees were combined to form a strict consensus tree. Overall character congruence was estimated by the consistency index (CI), rescaled consistency index (RC) and then the retention index (RI). Bootstrap values with 1000 replications (Felsenstein 1985) were calculated by performing a heuristic search using the TBR option with MulTrees off. Parsimony methods try to minimize the number of substitutions, irrespective of branch lengths on the tree. To complement the parsimonious analysis, the neighbour-joining (NJ) method (Saitou and Nei 1987) Tajima-Nei distance was also used for phylogenetic reconstruction.

4. Results

4.1 Results of *RPB2* gene

4.1.1 Electrophoresis banding patterns

A total of 41 accessions with various genomic combinations, were amplified with P6F/P6FR primer combination. For some species, more than one accession was used. Figure 1 shows the example of amplified patterns of different species. The **StH** species showed two distinguished bands corresponding in length to those of *P. spicata*, *P. stipifolia* and *H. bogdanii*. For **StY** species, *E. ciliaris*, *E. abolinni* and *E. semicostatus* and hexaploid **StStY** species *E. tschimganicus*, the patterns appeared to display two bands but they were very close to each other and it is difficult to resolve them, thus it appears to have only one band.

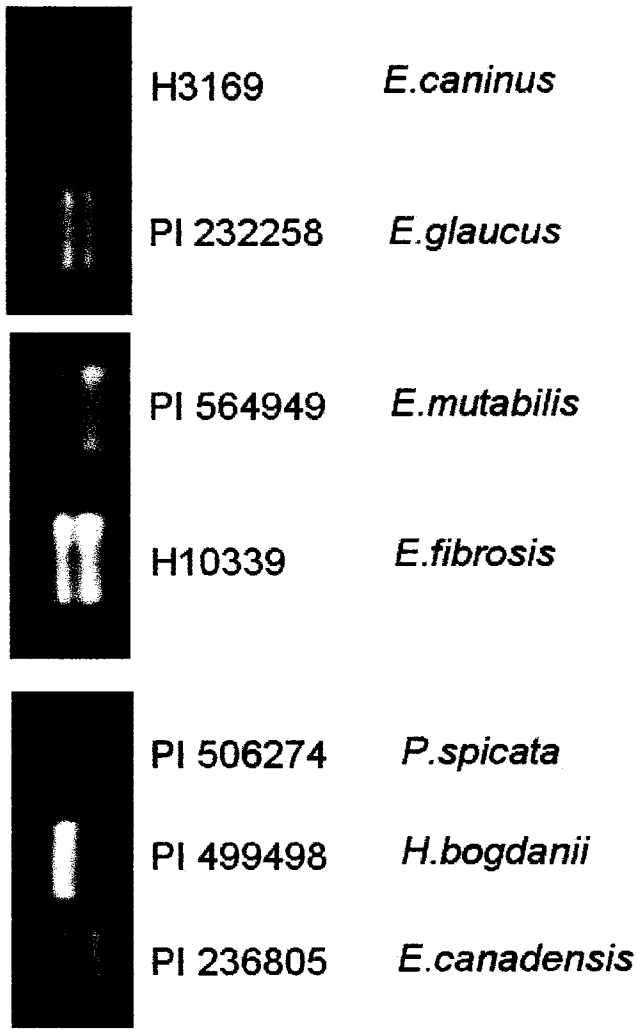


Figure 1: Electrophoresis banding patterns of amplification of *RPB2*. *Pseudoroegneria spicada*, *Hordeum bogdanii* and several *Elymus* species with **StH** genome combinations

4.1.2 Sequence results and their variation

The *RPB2* genes were sequenced using primer pair T7/SP6. The length of resulted sequence varied from 699bp (accession PI 236663) to 1049bp (accession PI 368886). Sequence alignment showed three larger transposable-like element insertion/deletions (indels) (Sun, et al. 2007). The first indel occurred at position of 80 bp. All **H** genomes have a deletion in this region. Inverted repeated sequences were detected for the sequences from **St**, **Y**, **W**, $\mathbf{E^b}$, **V** and **P** genomes, but the repeat motif in P genome was short (see Figure 2). All sequences from **St** genome except for *P. libanotica* have a 38-39 bp MITE stowaway element insertion at the position of 220 bp. The longest MITE stowaway insertion (~100 bp) in this region was found for the V genome of *D. villosum* (Figure 3). A third large indel with about a 27-29 bp insertion appeared in the sequences from all genomes except for H genome. None of H genome contained the sequence (GTTTTTATAAAATGGAGTGCTCCCAATCT) in this region (see Figure 4).

Figure 2 Partial alignment of the amplified sequences of *RPB2* from *Elymus* and their putative St and H genome donor *Pseudoroegneria* and *Hordeum* species. The indel I are transposable-like element deletions. The boxed region shows conserved TIRs element.

Figure 3: Partial alignment of the amplified sequences of *RPB2* from *Elymus* and their putative St and H genome donor *Pseudoroegneria* and *Hordeum* species. (Indel II)

Figure 4: Partial alignment of the amplified sequences of *RPB2* from *Elymus* and their putative St and H genome donor *Pseudoroegneria* and *Hordeum* species. (Indel III)

4.1.3 Nucleotide diversity analysis

Nucleotide polymorphisms, π and θ were calculated separately from **St** genome of *Elymus* species, *Pseudoroegneria* and **H** genome from *Elymus* species. The number of polymorphic sites in tetraploid *Elymus* **St** genome was around 2-folds higher than in its diploid donor species.

For **St** *Elymus* sequences alignment, 7 species were included. 753 nucleotides (excluding sites with missing data and gaps) encompassed 59 polymorphic sites. The nucleotide diversity (π) was 0.02489. The proportion of polymorphic nucleotide sites (θ) based on the number of segregating site was 0.02883.

For *Pseudoroegneria* **St** species, *P. libanotica*, *P. spicata*, and *P. stipifolia* were involved in the analysis. A total number of 721 nucleotides, excluding gaps (indels) and missing data, were determined by DNAsp4; in which, 694 were invariant sites and 27 were variant (polymorphic) sites. The estimates of nucleotide diversity in the **St** genome of *Pseudoroegneria* were $\pi = 0.01711$ and $\theta = 0.01640$.

H genomes *Elymus* sequences were used to analyze the nuclear diversity. When gaps and missing data were excluded, 682 nucleotides were included in analysis. Among which, 620 were invariant sites and 62 were variant sites. The standard deviation of haplotype diversity was 0.014. For nuclear diversity, π was 0.01874 and θ was 0.02643.

Table 3: Estimates of nucleotide diversity for selection at *RPB2* in tetraploid *Elymus* **St** and **H** genome and diploid *Pseudoroegneria* **St** genome

	n	S	π	θ_w
Elymus genome				
St	753	59	0.02489	0.02883
H	682	62	0.01874	0.02643
Pseudoroegnera				
St	721	27	0.01711	0.01711

The n is the number of the sites (excluding sites with gaps/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.

4.1.4 Phylogenetic analysis of the *RPB2* sequences

In order to understand the genomic relationships of the polyploidy *Elymus*, a total of 52 unique sequences were obtained and used for phylogenetic analysis, including *RPB2* sequences of all the polyploidy species, together with those of diploid from Triticeae.

Maximum parsimony analysis was performed using *D. villosum* as outgroup. The data matrix contains 842 characters, of which 629 characters were constant. A hundred and three characters were parsimony informative. Parsimony analysis resulted in 313 trees that were equally parsimonious. Each of the trees were 331 steps with a consistency index (CI) of 0.722, a retention index (RI) of 0.904, and rescaled consistency index (RCI) of 0.653. A strict consensus tree was constructed from these trees and this is shown in (Figure 5). The bootstrap (1000 replicates) values are shown on each branch. All the homoeologous *RPB2* sequences from polyploidy StH accessions grouped with those of the diploid parental clades are expected from cytological studies. The **H** genome of the *Elymus* species together with *Hordeum* species formed a clade with 100% bootstrap support. The **St** genome sequences of *Elymus* species with *Pseudoroegneria* species formed a clade with 61% bootstrap support, but with a 92% bootstrap support when grouped with *P. spicata* and *P. stipifolia.* The **H** genome clade was further divided into two groups. The first group has 73% bootstrap support and contains *H. stenostachys*, all **StH** *Elymus* species from America and one Eurasian species *E. mutabilis*. The second group with 86% bootstrap support comprised of *H. bogdanii* and **StH** *Elymus* species from Eurasia. Two

accessions of *E. trachycaulus*, one from Eurasia (H3526) and one from North America (PI 537323) were included in this study, and it was interesting that the grouping of the **H** genome from *E. trachycaulus* reflected their geographical origin. The sequences from **StY** *Elymus* species were separated into two clades, some were put into the St-clade, another formed a well supported clade with **W** and **P** genome (70% bootstrap value) (**W+P+Y** clade). *E. coreanus* is the only *Elymus* species that does not group with *Hordeum* or *Pseudoroegneria*, or the **Y** genome. Neighbor-Joining analysis based on Tajima-Nei distance generated a similar topology with minor variation in bootstrap value (Figure 6)

Figure 5: Strict consensus parsimonious tree derived from *RPB2* sequence data was conducted using a heuristic search with TBR branch swapping. Numbers on branches are bootstrap value. *Dasypyrum villosum* was used as outgroup. Consistency Index (CI) =0.724, retention index (RI) =0.910, rescale consistency index (RCI) =0.659.

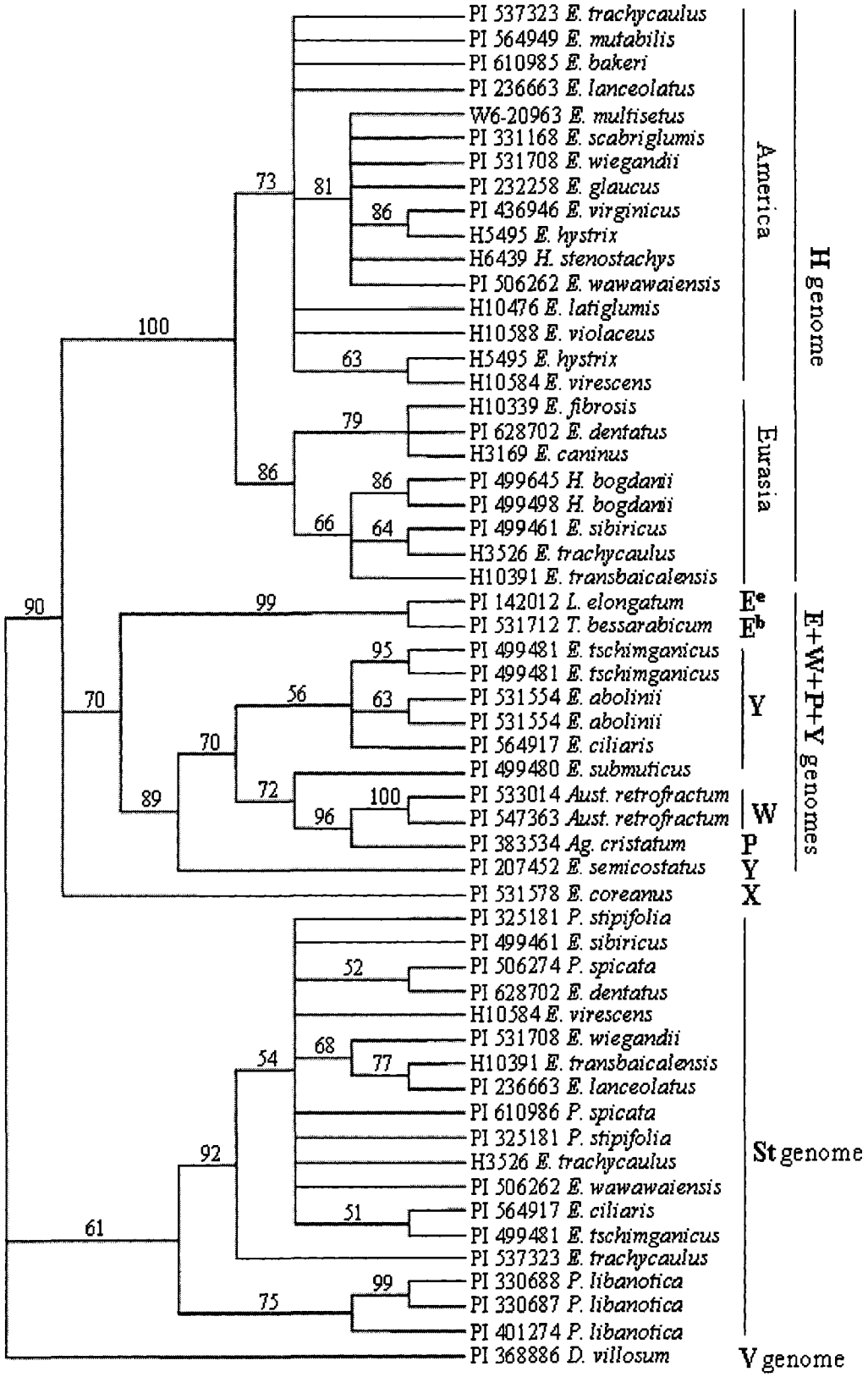
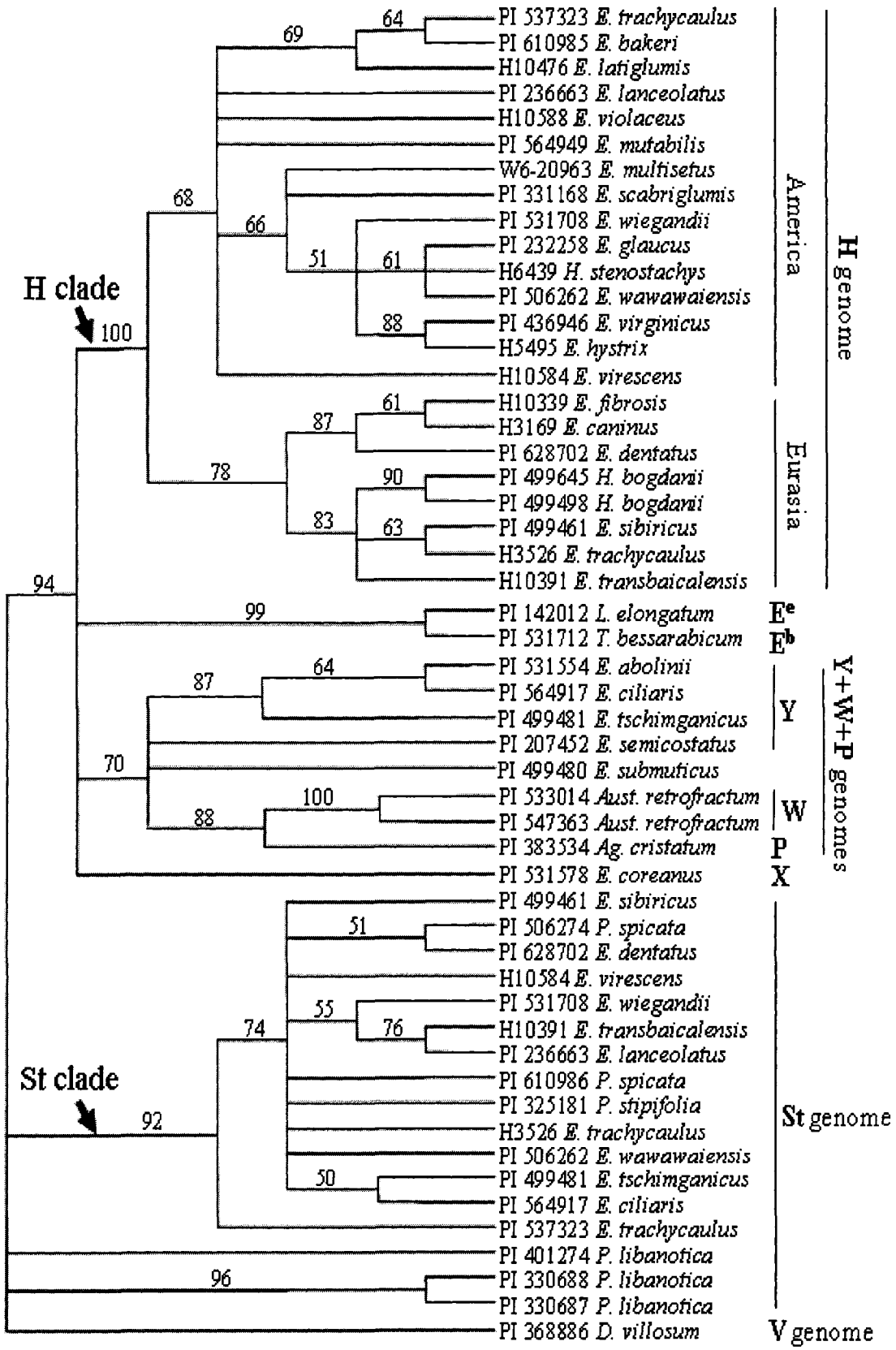


Figure 6: The phylogenetic tree and outgroup taxa obtained by the NJ method based on Tajima-Nei distance. Bootstrap supports (%) above 50% with 1000 replicates are given along the branches. *D. villosum* was used as outgroup.



4.2 Results of *Asp (GUC)-The (GGU)* intergenic region and flanking region

4.2.1 Sequence results and their variation

Thirty-six samples were successfully amplified by PCR using the TrnT/TrnD primer pair, resulting in an approximately 1.5k bp band for each sample. Sequence results revealed the observed polymorphisms. At position 50-60, AAATA repeat insertion was found in *E. caninus* (accession H3169), *E. fibrosis* (accession H10339), *E. dentatus* (accession PI 628702) and *E. multabilis* (PI 564949). High sequence variation was observed at position 260-380. Microsatellites and minisatellites were found between the repeated sequences TAGA (Fig 7). The microsatellite is a mononucleotide repeat represented as (A)n. The minisatellite is observed in *E. transbaicalensis* (accession H10391), *E. caninus* (accession H10404) with the form of $(TTTTTTCTTTATCTTTATTTCT)_2$. One *E. trachycaulus* (accession H3526) shows $(TTTTTTCTTTATCTTTATTTCT)_3$ repeat. For *E. hyperarcticus* species, the $(TTTTTTCTTTATCTTTATTTCT)_2$ minisatellite is only observed in accession H10613. At position 710-720, accession H10613 also shows a six nucleotide ATCATA insertion, but accession H10621 shows no insertion. These two accession of *E. hyperarcticus* sampled from Canada were distinguished by these insertions. At position 810-820, one accession of *H. bogdanii* (accession PI 531762) and one accession of *E. wawawaiensis* (accession PI 506262) have CATATA insertions. Another microsatellite locus with $(T)_n$ was observed at position 1000-1010.

4.2.2 Nucleotide diversity analysis

A total of twenty-five species, and thirty-six accessions were amplified with TrnT/TrnW primer combination. We have successfully sequenced all clones from 34 accessions of **StH** species, their genome donor species and one *Aust. retrofractum* species. Seven species with multiple accessions have been used in the analysis as following: four accessions of *E. trachycaulus*, three accessions of *E. subscundus*, *E. virescenus* and two accessions of *E. hyperacticus*, *E. caninus*, *E. wawawaiensis*, *E. botalinii*. DNAsp4 determined that there are a total of 1150 nucleotides when gaps and missing data were excluded. 137 of them are polymorphic (segregating) sites. A total number of 36 haplotypes are used, including genera *Aust. retrofractum* (accession PI 547363). All species have the haplotype (gene) diversity value of 1.000. The variance of haplotype diversity is 0.00004 and its standard deviation is 0.007. The nucleotide diversity (π) is 0.00889. The proportion of polymorphic nucleotide site (θ) is 0.02873.

4.2.3 Phylogenetic analysis of *Asp (GUC) –The (GGU)* region

After excluding the indels, the data matrix contains 1284 characters, of which 1120 characters are constant. Thirty two characters are parsimony informative. Parsimony analysis was conducted to reveal the relationships of 22 *Elymus* species together with their putative St and H genome donor *Pseudoroegneria* and *Hordeum*. Parsimony analysis produced 186 equally parsimonious trees with a consistency index (CI) of 0.735, a retention index (RI) of 0.523, and rescaled index (RC) 0.385 (Figure 9). *Aust. retrofractum* (accession PI 547363) is used as outgroup for analysis.

The 35 accessions formed two clades, one clade with a 98% bootstrap support value containing 34 accessions of 22 *Elymus* taxa and *P. spicata* taxa; the other clade was formed by the single taxa *H. bogdanii.* Two *E. subseundus* accessions (accession PI 236685, accession PI 537322) and *E. drovovii* (accession PI 564926) are grouped together with 61% bootstrap support; the other accession of *E. subseundus* (accession PI 232147) and *E. wawawaiensis* (accession PI 506262) are grouped together with 62% bootstrap support. Three accessions of *E. trachycaulus* (accession PI 537323, accession PI 440099 and accession PI 276711) and *E. hyperaticus* (accession H 10621) were grouped together with 83% bootstrap.

The neighbor joining (NJ) analysis was performed. The tree reconstructed by the NJ method is shown in Figure 8. The tree obtained was similar to that of parsimony method. Two distinguished major clades were formed. The *Elymus* species and their **St** genome donor species *P. spicata* were well separated from the **H** genome donor species *H. bogdanii. P. spicata* (accession PI 232140) grouped together with other eight *Elymus* species and formed a subclade with *E. mutalibis* (accession PI 564949), *E. fibrosis* (accession H10339), *E. caninus* (accession H3169), *E. violaceus* (accession H10588), *E. latiglume* (H10476), *E. mutisetus* (accession PI 619552), *E. denatus* (accession PI 628702) and *E. botalinii* (accession PI 531620).

Figure 7: Partial alignments of the amplified sequences of *Asp (GUC) – The (GGU)* from *Elymus* and their putative **St** and **H** genome donor *Pseudoroegneria* and *Hordeum* species. One *Aust. retrofractum* species is also included. The boxed regions refer to the repeated sequences TAGA. The microsatellite region has been marked and it is represented as $(A)_n$. Minisatellite region was also highlighted in the figure.

Figure 8: The phylogenetic tree and outgroup taxa obtained by the Neighbor Joining (NJ) method. Branch lengths were shown above each branch. *Aust. retrofractum* (accession PI 547363) was used as outgroup.

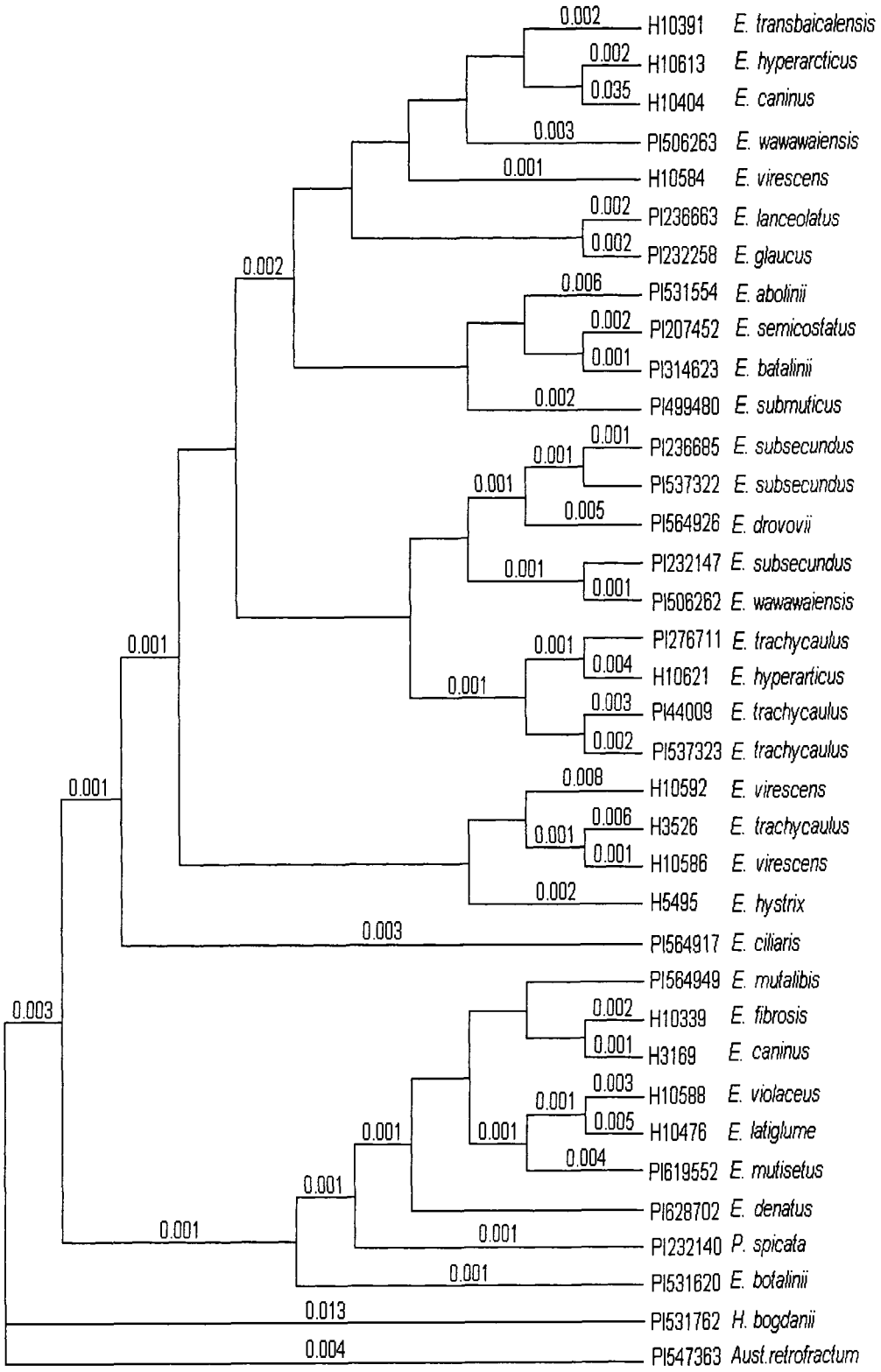
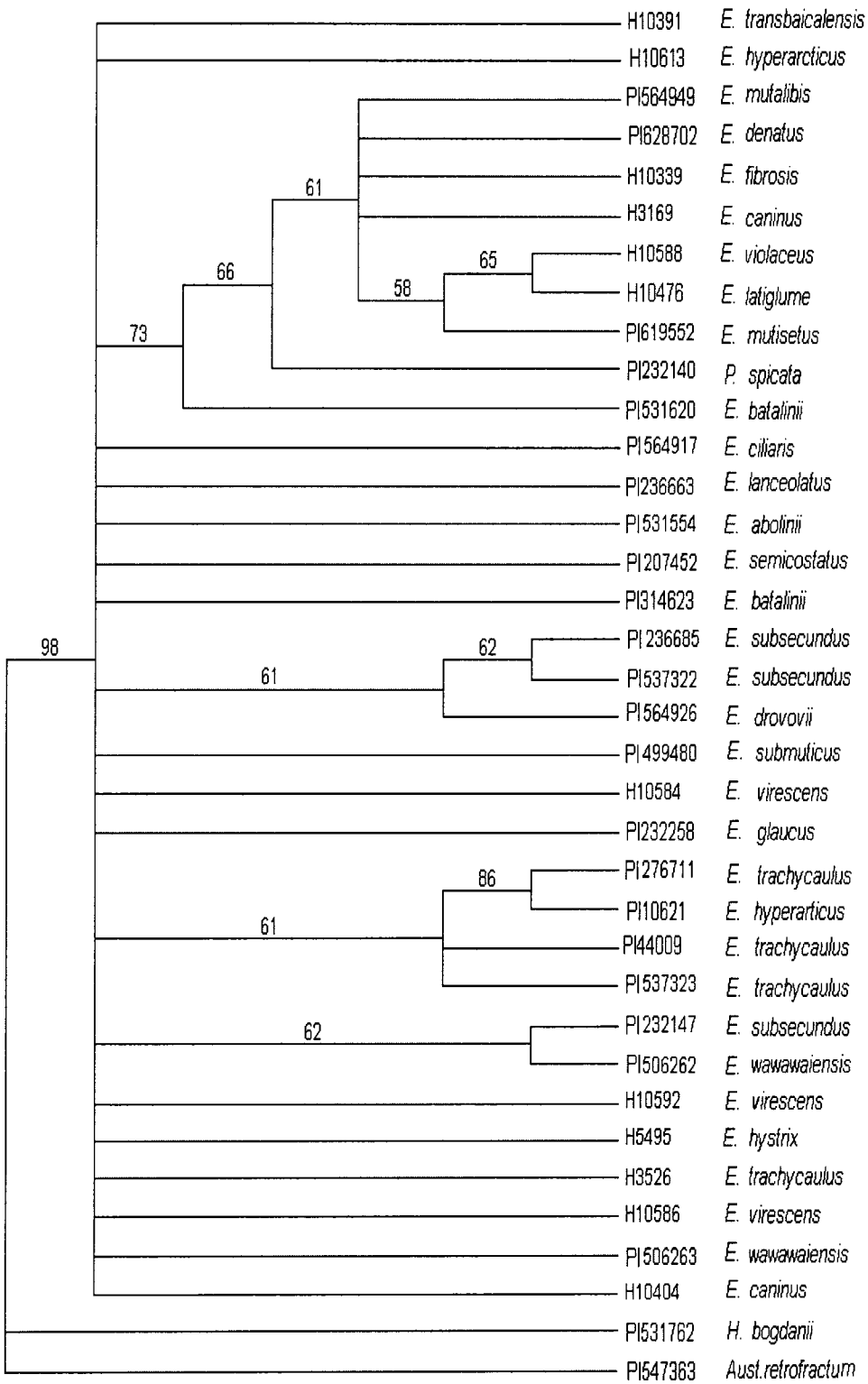


Figure 9: Maximum Parsimony analysis of *P. spicata*, *H. bogdanii* and *Elymus species*. Outgroup: *Aust.retrofractum*. Bootstrap support (%) above 50% with 1000 replicates is given along the branches. Parsimony analysis produced 186 equally parsimonious trees with a consistency index (CI) of 0.735, a retention index (RI) of 0.523, and rescaled index (RC) 0.385



4.3 Results of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region

4.3.1 Sequence results and their variation

Twenty samples were successfully amplified by PCR using TrnS/PsbC primer pair due to the time limitation and quality of difference among DNA samples, since some of the DNA samples were really old and their concentration were low. Therefore, 18 *Elymus* species and 2 species from *Pseudoroegneria* were used in this study. The *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region was sequenced using universal primer pair T7/SP6 combination. The length of sequences was approximately 1.5 kb. After alignment, the main variation of repeated nucleotide "A" was detected between 1490 bp to 1530 bp. This flanking region of the microsatellite $(A)_n$ was highly conserved with sequence (TGAAAGAAA) (Figure 10).

Figure 10: Partial alignment of the amplified sequences of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region from *Elymus* species, their putative **St** and **H** genome donor *Pseudoroegneria* and *Hordeum*. The dark grey boxes show the conservative TGAAAGAAA sequences. Microsatellite has also been marked.

```
                    *      1480        *      1500        *      1520        *
PI314623  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAACAA-----A ... ...CAAACGTATTAA : E.batalinii
PI499480  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAA-AA-----A ... ...CAAACGTATTAA : E.submuticus
PI330687  : CCGAGCCATTCATTCCTTTTTA ... ...AATAAGGGACAGAATAAAAAAAA-----A ... ...CAAACATATTCA : P.libanotica
PI531578  : CCGAGCCATTCGTTCCTTTTTA ... ...GATAAGGGACAGAATAAAAAAAAGAAAAA ... ...CAAACGTATTCA : E.coreanus
PI628702  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA-----A ... ...CAAACGTATTAA : E.dentatus
PI499461  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA-----  ... ...CAAACGTATTAA : E.sibiricus
PI531554  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA----A ... ...CAAACGTATTAA : E.abolinii
PI628674  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAA------A ... ...CAAACGTATTAA : E.dahuricus
PI531617  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAGA-----A ... ...CAAACGTATTAA : E.interruptus
H10588    : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAA-------A ... ...CAAACGTATTAA : E.violaceus
PI531698  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAA-A----A ... ...CAAACGTATTAA : E.tsukushiensis
PI531708  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAA--------  ... ...CAAACGTATTAA : E.wiegandii
PI636671  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA---AAA ... ...CAAACGTATTAA : E.antarticus
H3526     : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAA-----A ... ...CAAACGTATTAA : E.trachycaulus
H5495     : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA----A ... ...CAAACGTACTAA : E.hystrix
PI564917  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA----A ... ...CAAACGTA-TAA : E.ciliaris
PI531703  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAA-----  ... ...CAAACGTATTAA : E.villosus
H3169     : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAAA---A ... ...CAAACGTATTAA : E.caninus
PI506262  : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAAAA---AA ... ...CAAACGTATTAA : E.wawawaiensis
PI5062733 : CCGAGCCATTCATTCCTTTTTA ... ...GATAAGGGACAGAAAAAAAAAA------  ... ...CAAACGTATTAA : P.spicata
            CCGAGCCATTCaTTCCTTTTTA ... ...gATAAGGGACAGAAaAAAAAa a     ... ...CAAACgTAtTaA
```

microsatellite

4.3.2 Nucleotide diversity analysis

Estimates of nucleotide diversity (θ, π; gaps treated as missing data) were calculated based on sequence data. Twenty species were used with one accession for each species. There were 200 polymorphic sites. Among these variable sites, 27 were parsimony informative. A total of 20 haplome types were detected among the 20 sequences analyzed, including two *Pseudoroegneria* species, *P. libanotica* (accession PI 330687) and *P. spicata* (accession PI 506273). The variance of haplotype diversity is 0.00025 and its standard deviation is 0.016. A number of statistical tests were used to determine the nucleotide diversity proportion of polymorphic nucleotide site. Based on Tajima (1989), the nucleotide diversity π is 0.01673 and θ is 0.03677.

4.3.3 Phylogenetic analysis of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region

Excluding the indels, the data matrix contains 1598 characters, of which 1370 characters are constant. One hundred and eighty seven variable characters were parsimony-uninformative, and 29 were parsimony informative. Gaps were treated as missing data and multistate taxa interpreted as uncertainty. Parsimony analysis produced 277 equally parsimonious trees with a consistency index (CI) of 0.907, a relationship index (RI) 0.770, and rescaled consistency index (RCI) 0.698. A strict consensus tree based on maximum parsimony analysis was constructed and shown in Figure 11. The bootstrap (1000 replicates) values were shown on each branch. The tree reconstructed by the Neighbor Joining method was shown in Figure 12.

The species were well separated into two major groups and supported by both MP and NJ analysis. All *Elymus* species were grouped together with *P. spicata* (accession PI 506273) with 53% bootstrap confidence (clade I); *P. libanotica* (accession PI 330687) formed a monotypic branch. In clade I, *E. villosus* (accession PI 531703), *E. caninus* (accession H3169) and *E. wawawaiensis* (accession PI 506262) formed a subclade with 100% bootstrap confidence which is consistent with the results revealed by NJ analysis (Figure 12). *E. wiegandii* (accession PI 531708), *E. antarcticus* (PI 636671) and *E. trachycaulus* (accession H3526) formed another subclade with 76% bootstrap value and supported by NJ analysis. *E. dahuricus* (accession PI 628674), *E. interuptus* (accession PI 531674) and *E. violaceus* (accession H 10588) were also grouped together with 51% bootstrap value which was also supported by NJ analysis. However, NJ analysis revealed that *E. interuptus* and *E. violaceus* had closer relationship than *E. dahuricus*.

The tree reconstructed by the Neighbor Joining method revealed well solved interspecific relationships within *Elymus* species and *P. spicata* as shown in clade I (Figure 12). Two subclades, subclade I and subclade II were formed. Subclade II included two species, *E. dentatus* (accession PI 628702) and *E. sibiricus* (accession PI 499461). All other *Elymus* species were grouped together with *P. spicata* and generated subclade I.

Figure 11: Maximum Parsimony (MP) analysis of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region. The species *E. coreanus* with **XN** genome (accession PI 531578) was used as outgroup. Bootstrap support (%) above 50% with 1000 replicates is given along the branches.

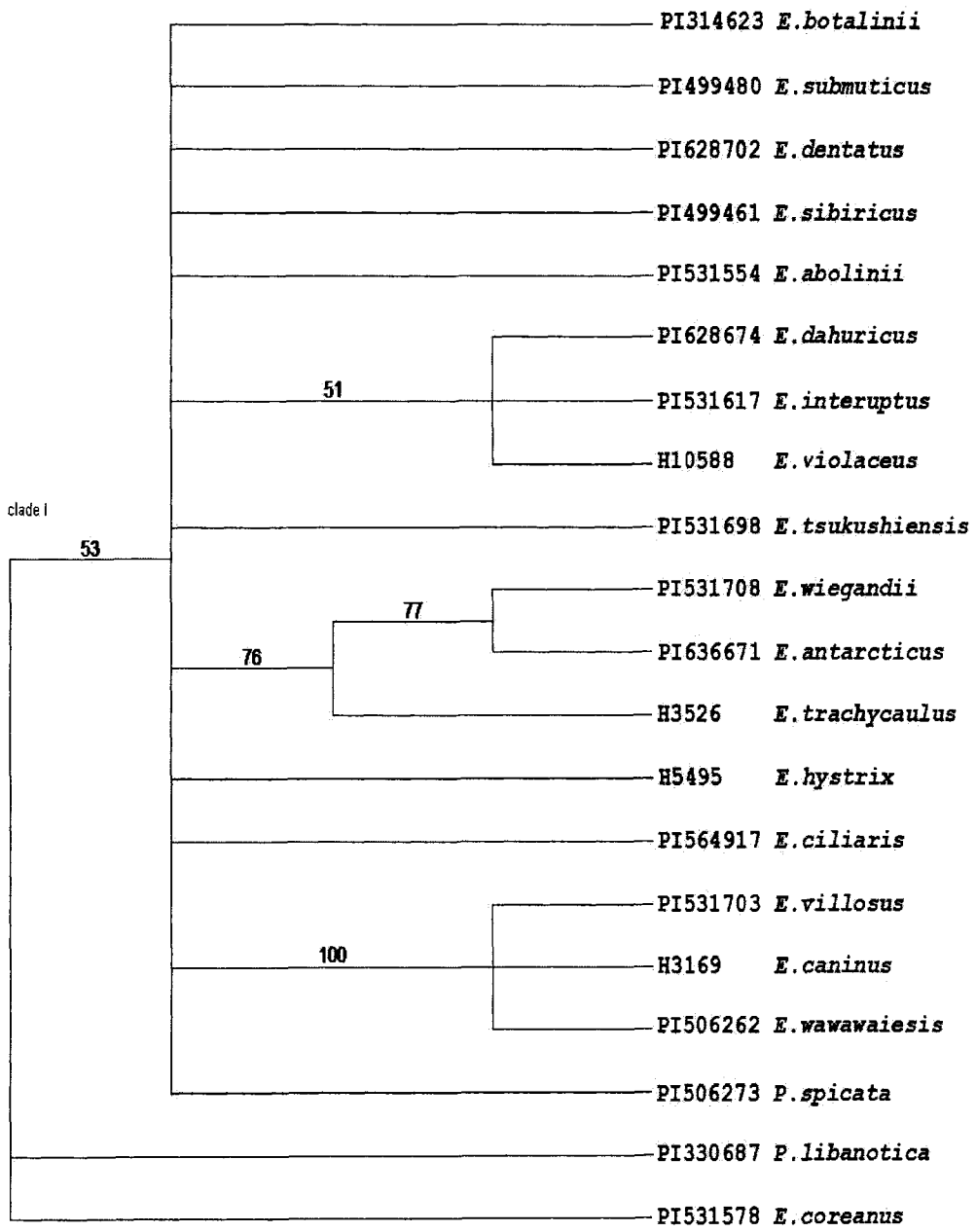
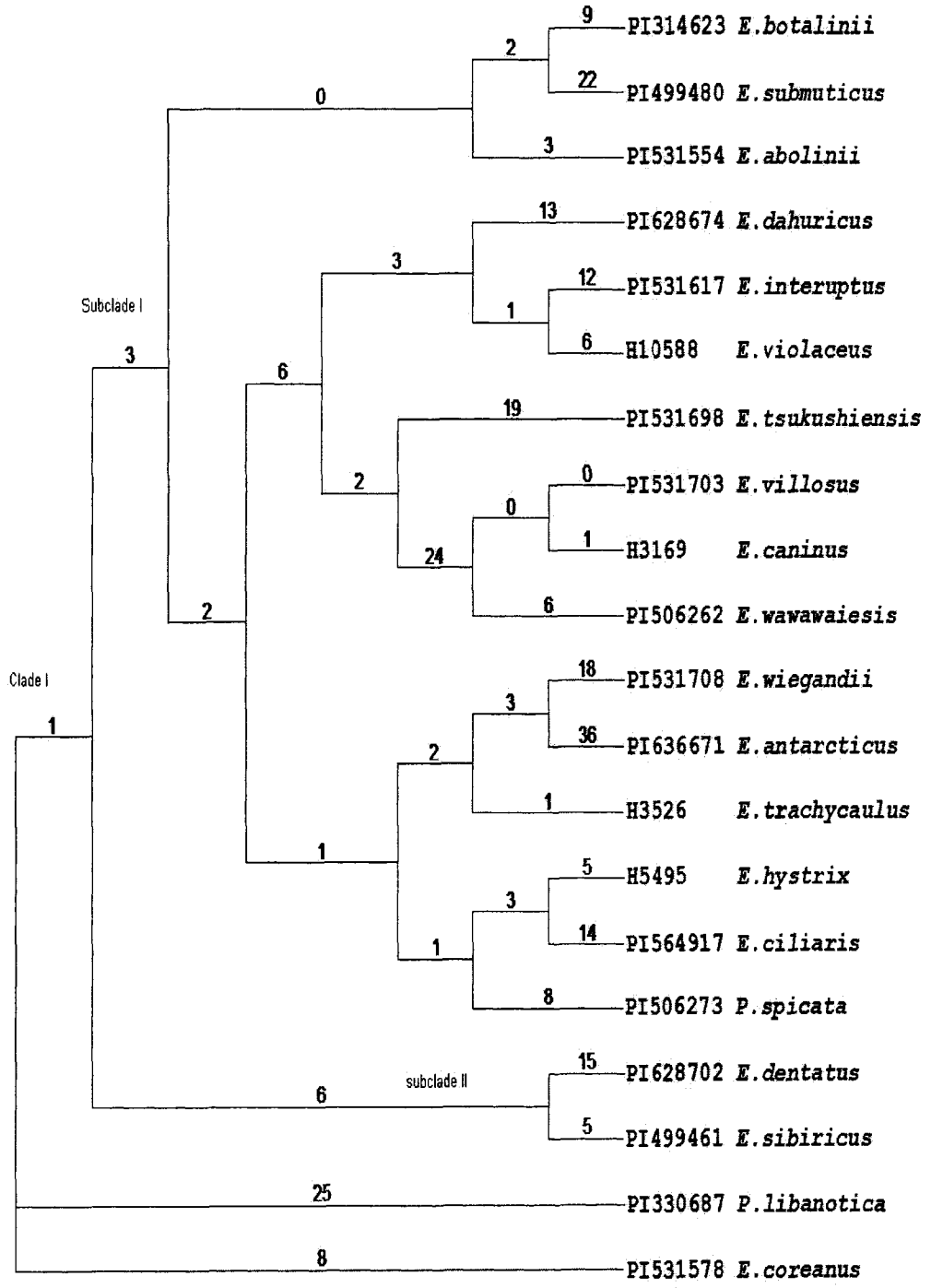


Figure 12: The Neighbor Joining (NJ) analysis of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region. *E. coreanus* (accession PI 531578) was used as outgroup. Branch lengths were included above each branch.



5. Discussion

5.1 The confirmation of maternal donor of StH *Elymus* species.

The accurate reconstruction of genetic trees has long been difficult in phylogenetics. Comparison of cpDNA sequences is a useful way to identify the female parent in interspecific hybridization (Mason-Gamer and Kellogg 1996). The chloroplast genome is generally uniparentally inherited and thus can be used to identify the maternal genome donor of a given polyploid. Redingbaugh et al. (2000) found that the cpDNA sequences were highly similar among species containing the **St** genome. The chloroplast DNA analysis of North American *Elymus* indicated that *Pseudoroegneria* is the potential chloroplast donor to these *Elymus* species (Mason-Gamer et al. 2002). Studies of cpDNA in tetraploid and hexaploid *Elymus* species containing **StH**, **StHY** or **StY** genome obtained the same conclusion (McMillan and Sun 2004, Xu and Ban 2004).

In this study, all **St** genome *Elymus* species were grouped together with *P. spicata*, which were supported by analyses of two cpDNA regions. The tree reconstructed by the MP method for *Asp(GUS) – The(GGU)* intergenic region formed a subclade with 8 *Elymus* species; The tree reconstructed by the NJ method for *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region generated a subclade containing *P. spicata* and 12 *Elymus* species. *E. caninus* and *E. violaceus* were grouped together with *P. spicata* in both analyses. Our cpDNA sequence results suggested that *Pseudoroegneria* is the maternal genome donor to tetraploid *Elymus*

StH genome species.

5.2 Multiple origins of tetraploid StH genome species

Polyploidy, the presence of more than two genomes per cell, is an important evolutionary force in plants. In evolutionary events of tetraploid **StH** genome *Elymus*, this corresponds to the hypothesis of whether they arose once by an ancient hybridization between the parental genera *Pseudoroegneria* and *Hordeum* and then differentiated into a number of species; or whether hybridizations have happened repeatedly at different times and places during the evolutionary history (figure 13).

Many researchers have addressed the origin of tetraploid StH genome species and try to find out whether they are monophyletic or polyphyletic. Cytological data suggested that the genome differentiation exists among the *Pseudoroegneria* species (Wang et al. 1986). The *waxy* exon data significantly rejects monophyly in the genus *Pseudoroegneria* (Mason-Gamer and Kellogg 2000). The polyphyletic origins were confirmed in *E. repens* (Mason-Gamer 2004). In this study, phylogenetic analysis of *RPB2* sequences revealed several independent groups for both **H** and **St** Sequences from *Elymus* (figure5, figure6). This molecular evidence indicated that **St** genome in *Elymus* species originated from different ancestors, and supports the multiple origins of **StH** species in *Elymus*. Our *RPB2* sequence data also showed that **St** genome in *P. libanotica* is highly differentiated from the **St** genome in *P. spicata* and *P. stipifolia*. The result indicated that unlike *P. spicata* and *P. stipifolia*, *P. libanotica* was not the **St** genome donor of the *Elymus* species studied here. This conclusion did not conflict

with the previous studies (Mason-Gamer et al. 2002; Mason-Gamer and Kellogg 2000).

Isoenzyme analyses suggested that Eurasian and American **StH** genome tetraploid species may independently originate from different **H** genome donors (Jaaska 1992). The C- and N-banded karyotypes of *Elymus* species also suggested that **H** genome *Elymus* species may be derived from different diploid **H** donors (Linde-Laursen et al. 1994). In the present analysis of the *RPB2* sequence data, the **H** genome *Elymus* species were separated and grouped together by their geographical origin: Eurasian and North/South American. Our data strongly supported that the **H** genome in *Elymus* species is highly differentiated, reflecting its geographical origin. Therefore, the Eurasian and American **StH** genome species have independent alloploid origins with different **H** genome donors. This result is consistent with the isoenzymic and cytological data (Jaaska 1992; Linde-Laursen et al. 1994).

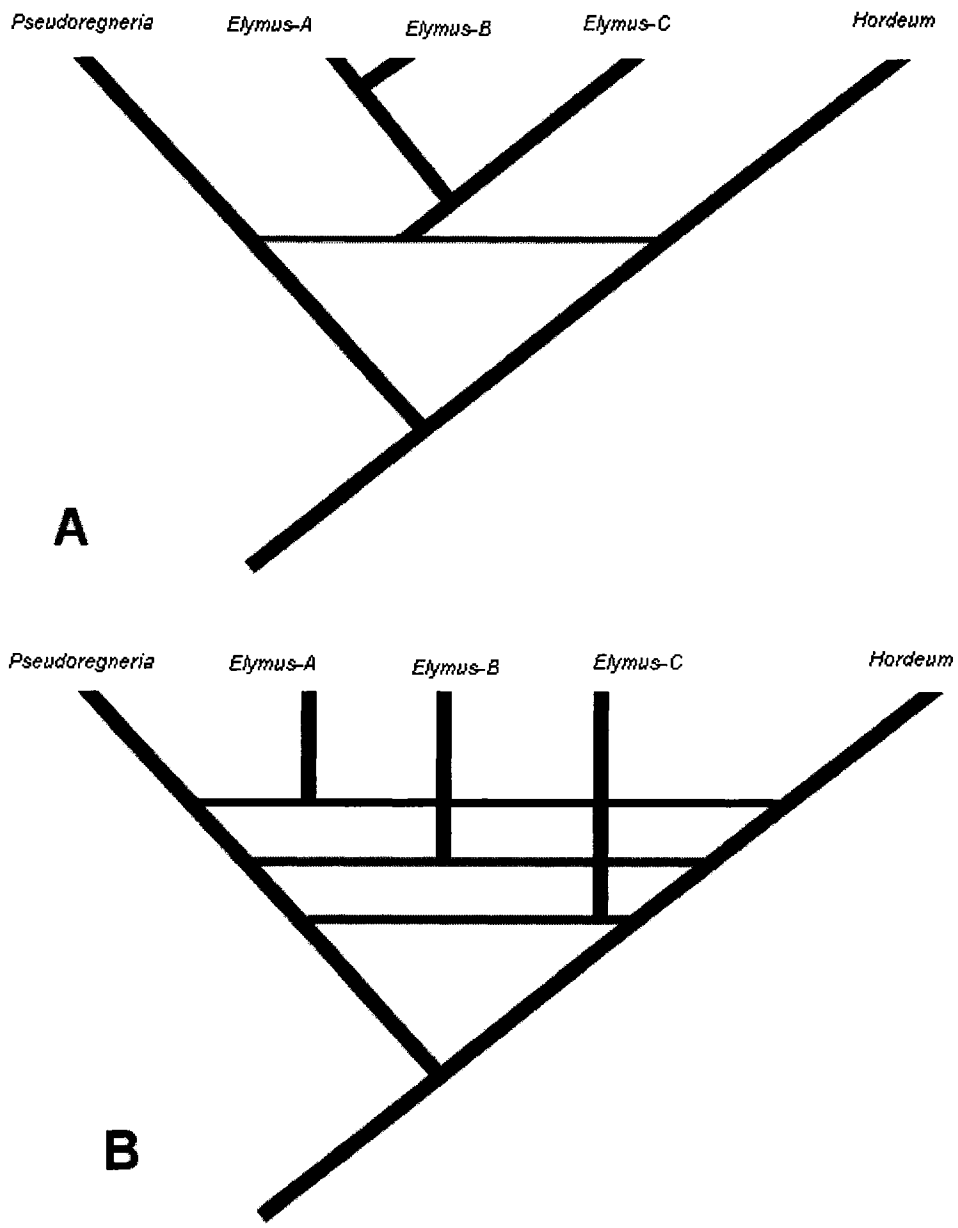


Figure 13: Simplified phylogenetic trees illustrating the two hypotheses. A. *Elymus* is of monophyletic origin; B. *Elymus* is of polyphyletic origin.

5.3 Molecular evolution of *RPB2* in *Elymus* species and their diploid genome donor species

Resent studies have focused on the evolutionary dynamics and the rate of molecular evolution in polyploids compared to their diploid relatives. Phylogenies of low-copy nuclear genes have proven effective in determining the rates of molecular evolution in polyploidy compared to their diploid relatives (Soltis et al. 2003). Alcohol dehygrogenase (*Adh*) gene family, such as *AdhC*, the most rapidly evolving of the *Gossypium Adh* gene family, was examined in allotetraploid and diploid species of cotton (*Gossypium*) (Small et al.1998, Small and Wendel 2000, 2002). *AdhC* genes from the D-(sub) genomes were accumulating substitutions at a significantly faster rate than they were in A-(sub) genomes (Small et al. 1998, 1999). Variable rates of evolution between the loci (AdhA-AdhE) were also found in diploid and allotetraploid species (Small and Wendel 2000). Cronn et al. (1999) sequenced 16 loci from allotetraploid species *Gossypium hirsutum* and found that most duplicated genes in allopolyploid cotton evolve independently of each other and at the same rate as those of their diploid progenitors. MYB genes, which encode nuclear proteins that function as transcriptional transactivators, also suggest that they evolve independently and at relatively equal rates in *G. hirsutum* and the diploid relatives, *G raimondii* and *G. arboreum* (Cedroni et al. 2003).

To the contrary, the floral regulatory gene *APETALA1 (ASAP1)* and *APETALA3 (APETALA3/TM6)* are evolving much faster in the polyploid species than in the diploids (Barrier et al. 2001). The analysis of *Xwye838* locus (encoding

ADP-glucopyrophosphorylase) and *Gss* locus (encoding granule-bound starch synthase) in polyploidy wheat D genome and diploid *Aegilops tauschii* D genome indicated that genetic diversity of *Gss* locus on the D genome of hexaploid wheat was greatly reduced compared to that of *Ae. tauschii* (Caldwell *et al.* 2004). In present study, nucleotide sequence diversity (π) of the **St** genome of tetraploid *Elymus* was higher than in the diploid *Pseudoroegneria* St genome. The results did not contradict with the findings of Barrier et al. (2001), and suggests that the *RPB2* sequence may evolve faster in the polyploidy species than in the diploids. Another possible explanation is that *Elymus* is likely polyphyletic, and StH species might evolved several times.

5.4 transposable elements indels in *RPB2* loci

As described by Feschotte *et al.* (2002): "Transposable elements (TEs) are fragments of DNA that can insert into new chromosomal locations and often make duplicate copies of themselves in the process". Transposable elements, which are the single largest component of the genetic material of most eukaryotes, play an important role in plant genome evolution and polyploid evolution (Feschotte et al.2002; Soltis et al. 2003). There are three classes of TEs: class I, elements move via an RNA intermediate; class II, elements move via a DNA intermediate; class III, elements move with unknown mode (Capy et al. 1997).

Miniature inverted-repeat transposable elements (MITEs) are particular class-II TEs (Bureau and Wessler 1994). Peterson and Seberg (2000) defined MITEs as short

(~60-700bp) fragments with no coding capacity; they have conserved, usually short, terminal inverted repeats (TIR), and they potentially are able to form a hairpin-like secondary structure; they also show target site preference.

Based on the similarity of the terminal inverted repeats (TIRS) and target site duplications (TSDs), MITEs have been classified into two superfamilies: *Tourist*-like and *Stowaway*-like (Feschotte et al. 2002). *Stowaway* elements vary considerably in length, from 39bp (*RPB2* gene loci in this study), to 323bp (Bureau and Wessler 1994). Recent studies suggested that the MITE *stowaway* element is common in Triticeae species (Petersen and Seberg 2000, Wicker et al. 2003). Our study first found a 39bp MITE *stowaway* element insertion in the genic region of *RPB2* gene for all tetraploid *Elymus* species **St** genome and diploid *P. spicata* and *P. stipifolia* **St** genome (Sun et al. 2007). Comparison of the sequences from the **St** and **H** genome indicated that transposable element indels have shaped the homoeologous *RPB2* gene loci in **St** and **H** genome of *Elymus* species.

5.5 Genetic relationship of *Elymus* species revealed by two cpDNA regions

Many cpDNA noncoding regions (introns and intergeneic spacers) have been characterized either by direct sequencing, or by restriction site analysis of polymerase chain reaction (PCR)-amplified products (Mason-Gamer et al. 2002; Sun et al. 2006; McMillan and Sun 2004; Larson et al. 2003). PCR-RFLP analysis of a chloroplast DNA region consisting of coding and intergeneic spacer sequences of *trnS* and the adjacent *psbC* gene was used to detect the variation in gymnosperms and angiosperms

(Ziegenhagen and Fladung 1997).

The *E. trachycaulus* species has often been noticed because of their morphological variability and geographical diversity (Dewey 1982). *E. trachycaulus* is often distributed from Eurasia through Alaska to Newfoundland, south to Mexico along the Rocky Mountains, and it usually grows in open forests and along roadsides (Hitchcock 1951). Phylogenetic relationships among four *E. trachycaulus* accessions were inferred based on the sequence of *Asp(GUS)-The(GGU)* intergeneic region. Three accessions (accession PI 276711, PI 440099 and PI 537323) of *E. trachycaulus* that originated from Hungary, Kazakhstan, and the USA were grouped together; however accession H3526 which originated from Russia was not included in the clade. This result is consistent with the previous conclusion that there was high genetic variation within *Elymus trachycaulus* species and no clear association between geographic origin and genetic grouping (Sun and Li 2005).

Despite the high level of conservation of the chloroplast gene among *Elymus* species, our results are in good agreement with previous studies. The tree reconstructed by the NJ method for *Asp(GUS)-The(GGU)* intergeneic region, grouped *E. mutabilis* and *E. caninus* together with the other 7 *Elymus* species which is in accordance with the close relationship revealed by PCR-RFLP between *E. mutabilis* and *E. caninus* (McMillan and Sun 2004). The NJ analyses showed close relationship between *E. caninus* and *E. wawawaiensis* and it was consistent with previous parsimonious trees (Mason-Gamer et al. 2002). However, our result grouped *E. villosus* with these two species which have not been addressed so far.

Sun et al. (2006*b*) determined the molecular diversity and relationships of *Elymus* species through amplified fragment length polymorphism (AFLP). They found that *E. virescens* species was distinguished from *E. trachycaulus*, *E. subsecundus*, and *E. violaceus*. The maximum parsimony (MP) analysis of *Asp(GUS)-The(GGU)* intergeneic region separated *E. violaceus* and *E. virescens* into different clades, which did not contradict with previous study (Zhang et al. 2001; Sun et al. 2006*b*). However two accessions of *E. virescens* were not grouped together as before (Sun et al. 2006*b*). The accession H10621 is typical *hyperarcticus* type in morphology, and different from the tall *hyperarcticus* type accession H10613 (Sun and Salomon 2003). Minisatellite insertion/deletion was observed between these two accessions, and MP analysis of *Asp(GUS)-The(GGU)* intergeneic region showed that, unlike H10613, accession H10621 is close to *E. trachycaulus*. Therefore, these two accessions were highly differentiated.

P. libanotica formed a monotypic branch in both MP and NJ analyses of *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region. A similar result was revealed by *RPB2* sequence data in this study, which indicated that *Elymus* **St** genome species did not likely originate from *P. libanotica*. *E. dentatus* and *E. sibiricus* were grouped together in NJ analyses. This result was consistent with *RPB2* sequence data, which placed *E. dentatus* and *E. sibiricus* in the same clade with 90 bootstrap values.

5.6 The microsatellites and minisatellites in chloroplast DNA

Tandem repeat sequences, genome regions in which a single base or a short sequence is tandemly repeated, have gained increasing attention over the last decade (Debrauwere et al. 1997). Based on the number of bases in the repeated sequence, tandem repeat sequences are divided into two classes: microsatellite (usually <10bp units) and minisatellite (>10 and <100bp) (Avise 1994).

Microsatellites and minisatellites were widely used in plant research. Barreneche et al. (1998) applied microsatellites and minisatellite as markers to do a genetic map of Pedunculate oak; micro- and minisatellite were also used to identify the genotypes of rice (Ramakrishna et al. 1995); microsatellite markers was also useful in discriminating the barley, soybean, and chickpea accessions (Kraic et al. 2002).

The chloroplast genome has been shown to contain hypervariable regions similar to those found in nuclear genome (Powell et al. 1995a). Chloroplast microsatellites are predominately mononucleotide repeats (Powell et al. 1995b). In our study, the microsatellite of (T)n was observed in *Asp(GUC)-The(GGU)* region and (A)n was observed in *trnS[tRNA-ser (UGA)]-psbC[spII44kd protein]* region. Minisatellites were found in the inverted repeat region, intergeneic spacer, introns and non-coding sequences (King and Ferris 2002). The minisatellite identified in this study is the first report of an *Elymus Asp(GUS)-The(GGU)* intergeneic chloroplast region.

The copies of the chloroplast DNA molecule (within the same chloroplast or slipped-strand mispairing) could cause unequal cross-over which might be the origin of the repeat sequences (King and Ferris 2002). The origin slipped-stand misparing

sequences typically have the following features: (1) the presence of tandemly repeated sequences; (2) the motif which differs slightly in sequence or size is close to the variable region; (3) the occurrence of polypurine or polypyrimidine motif. The tandem repeated sequence appeared in *Asp(GUS)-The(GGU)* intergeneic chloroplast region has all these features. The *Elymus* minisatellite may generate from slipped-strand mispairing.

6. References Cited

Aquadro, C.F. and Begun D.J. 1993. Evidence for and implications of genetic hitchkiking in the Drosophila genome. In: Takahata N, Clark AG, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass

Avise, J.C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York, N.Y.

Barrier, M., Robichaux, R.H. and Purugganan, M.D. 2001. Accelerated regulatory gene evolution in an adaptive radiation. *Proceedings of the National Academy of Sciences USA* 98: 10208-10213

Barreneche, T., Bodenesl, C., Lexer, C., Trontin, J.F., Fluch, S., Streiff, R., Plomion, C., Roussel, G., Steinkellner, H., Burg, K., Favre, J.M., Glössl, J. and Kremer, A. 1998. A genetic linkage map of Quercus robur L. (pedunculate oak) based on RAPD, SCAR, microsatellite, minisatellite, isozyme and 5S rDNA markers. *Theoretical and Applied Genetics* 97(7): 1090-1103

Bureau, T.E. and Wessler, S.R. 1994. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant cell* 6: 907-916

Caldwell, K.S., Dvorak, J., Lagudah, E.S., Akhunov, E., Luo, M.C., Wolters, P. and Powell, W. 2004. Sequence polymorphism in polypoid wheat and their D-genome diploid ancestor. *Genetics* 167: 941-947

Capy, P., Bazin, C., Higuet, D. and Langin, T. 1997. Dynamic and Evolution of Transposable Elements. Landes Company, Texas, USA

Cedroni, M., Cronn, R.C., Adams, K.L., Wilkins, T.A. and Wendel, J.F. 2003. Evolution and expression of MYB genes in diploid and polyploidy cotton. *Plant Molecular Biology* 51: 313-325

Cronn, R., Small, R.L., and Wendel, J.F. 1999. Duplicated genes evolve independently following polyploid formation in cotton. *Proceedings of the National Academy of Sciences* 96: 14 406 – 14 411

Debrauwere, H., Gendrel, C.G., Lechat, S. and Dutreix, M. 1997. Differences and similarities between various tandem repeat sequence: minisatellites and microsatellites. *Biochimie* 79 (9): 577-586

Demesure, B., Sodzi, N. and Petit, R.J. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129-131

Denton, A.L., McConaughy, B.L. and Hall, B.D. 1998. Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution* 15: 1082-1085

Dewey, D.R. 1982. Genomic and phylogenetic relationships among North American perennial Triticeae, pp. 51-88 Grasses and Grasslandin edited by J.E. Estes. University of Oklahoma Press, Norman

Dewey, D.R. 1984. The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticea. pp. 209-280 in *Gene Manipulation in Plant Improvement*, 16th Stadler Genetics Symposium, edited by J.P. Gustafson, Plenum Publishing Crop., New York.

Doyle, J.J., Doyle, J.L. and Brown, A.H.D. 1999. Incongruence in the diploid B-genome species complex of *Glycine* (Leguminosae) revisited: histone H3-D alleles versus chloroplast haplotypes. *Molecular Biology and Evolution* 16: 354-362

Feschotte, C., Jiang, N. and Wessler, S.R. 2002. Plant transposable elements: where genetics meet genomics. *Genetics* 3: 329-341

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Hamby, R.K. and Zimmer, E.A. 1988. Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). *Plant Systematics and Evolution* 160: 29-37

Helfgott, D.M. and Mason-Gamer, R.J. 2004. The evolution of North American *Elymus* (*Triticeae*, Poaceae) allotetrploids: Evidence from phosphoenoleyruvate carboxylase gene sequences. *Systematic Botany.* 29(4): 850-861

Hitchcock, A. S. 1951. Manual of the grasses of the United States, ed. 2, revised by A. Chase. USDA Miscellaneous Publications 200. Washington, D. C.: U.S. Government Printing Office.

Hsiao, C., Chatterton, N.J., Asay, K.H. and Jensen, K.B. 1995. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome*, 38: 211-223

Jaaska, V. 1992. Isoenzyme variation in the grass genus *Elymus* (Poaceae). *Hereditas*

117: 11-22

Junghans, H. and Metzlaff, M. 1990. A simple and rapid method for the preparation of total plant DNA. *Biotechniques* 8: 176

King, R.A. and Ferris, C. 2002. A variable minisatellite sequence in chloroplast genome of Sorbus L. (Rosaceae:Maloideae). *Genome* 45: 570-576

Kraic, J., Gregová, E., Jomová, K. and Hudcovicová, M. 2002. Microsatellite markers discriminating accessions within collections of plant genetic resources. *Cell and molecular biology letters* 7(2B):745-51.

Knott, D. R. 1961. The inheritance of rust resistance. VI. The transfer of stem rust resistance from Agropyron elongatum to common wheat. --- Can. J. Pl. Sci. 41:109-123

Knott, D. R. 1989. The effect of transfer of alien genes for leaf rust resistance on the agronomic and quality characteristics of wheat. *Euphytica* 44: 65-72

Kolodziej, P.A., Woychik, N., Liao, S. and Young, R.A. 1990. RNA polymerase II subunit composition, stoichiometry, and phosphorylation. *Molecular and Cellular Biology* 10: 1915-1920

Larkin, R. and Guilfoyle, T. 1993. The second largest subunit of RNA polymerase II from *Arabibopsis thaliana. Nucleic Acid Research* 21: 1038

Larson, S.R., Jones, T.A., McCracken, C.L. and Jensen K.B. 2003. Amplified fragment length polymorphism in *Elymus elymoides, Elymus mutisetus,* and other *Elymus* taxa. *Canada journal of Botany* 81: 789-804

Linde-Laursen, I., Seberg, O. and Salomon, B. 1994. Comparison of the Giemsa

C-banded and N-banded karyotypes of two Elymus species, E. dentatus and E. glaucescens (Poaceae: Triticeae). *Plant Systematic Evolution* 192: 165-176

Löve, A. 1984. Conspectus of the Triticeae. *Feddes Repertorium*. 95: 425-521

Lu, B.R. 1993. Biosystematic Inverstigations of Asiatic Wheatgrasses – *Elymus L.* (*Triticeae*: Poaceae) – Dissertation, The Swedish University of Agricultural Sciences, Department of Plant Breeding Research, Svalöve, Sweden.

Mason-Gamer, R.J. and Kellogg, E.A. 1996. Chloroplast DNA analysis of the monogenomic *Triticeae*: Phylogenetic implications and genome-specific markers. In methods of genome Analysis in Plants. Edited by P.P. Jauhar. CRC Press, New York. pp. 301-324

Mason-Gamer, R.J. and Kellogg, E.A. 2000. Phylogenetic analysis of the Triticeae using the starch synthase gene, and a preliminary analysis of some North American *Elymus* species. In: Jacobs SWL, Everett, J, Editors. Grasses: *systematics and evolution.* Melbourne: CSIRO. pp. 102-109

Mason-Gamer, R.J., Orme, N.L. and Anderson, C.M. 2002. Phylogenetic analysis of North American *Elymus* and the monogenomic Triticeae (Poaceae) using three chloroplast DNA data sets. *Genome* 45: 991-1002

Mason-Gamer, R.J. 2001. Origin of North America *Elymus* (Poaceae: Triticeae) allotetraploids based of granule-bound starch synthase gene sequences. *Systematic botany* 26: 757-758

Mason-Gamer, R.J. 2004. Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Systematic biology* 53 (1): 25-37

McMillan, E. and Sun, G.L. 2004. Genetic relationship of tetraploid Elymus species and their genomic donor species inferred from polymerase chain reaction-restriction length polymorphism analysis of chloroplast gene regions. *Theoretical and applied genetic* 108 535-542

Moriyama, E.N, Powell, J.R. 1996. Intraspecific nuclear DNA variation in Drosophila. *Molecular Biology and Evolution* 13: 261-277

Nevski, S.A., 1934: Hordeae Benth. In: Komarov, V. L. Roshevits, R. Y. and Shishkin, B. K. (eds.), Flora URSS. II: Leningrad, pp. 590-728

Olmstead, R.G and Palmer, J.D. 1994: Chloroplast DNA Systematics: A Review of Methods and Data Analysis. *American Journal of Botany*, 81 (9): 1205-1224

Oxelman, B. and Bremer, B. 2000. Discovery of paralogous nuclear gene sequences coding for the second-largest subunit of RNA Polymerase II (RPB2) and their phylogenetic utility in gentianales of the asterids. *Molecular Biology and Evolution* 17: 1131-1145

Oxelman, B., Yoshikawa, N., McConaughy, B.L., Luo, J., Denton, A.L. and Hall, B.D. 2004. RPB2 gene phylogeny in phylogeny in flowering plants, with particular emphasis on asterids. *Molecular Phylogenetics and Evolution* 32: 462-479

Petersen, G and Seberg, O. 1997. Phylogenetic Analysis of the Triticeae (Poaceae) Based on rpo A Sequence Data. *Molecular Phylogenetics and Evolution* 7: 217-230

Petersen, G and Seberg, O. 1998. Molecular characterization and sequence polymorphism of the alcohol dehydrogenase 1 gene in *Hordeum vulgare L.*

Euphytica Genetics 102:57-63

Petersen, G. and Seberg, O. 2000. Phylogenetic evidence for excision of *Stowaway* miniature interted-repeat transposable elements in Triticeae (Poaceae). *Molecular biology and evolution* 17(11): 1589-1596

Peterson, G. and Seberg, O. 2002. Molecular evolution and phylogenetic application of DMC1. *Molecular Phylogenetics and Evolution* 22(1): 43-50

Powell, W., Morgante, M., Andre, C., McNicol, J.W., Marchray, G.C., Doyle, J.J., Tingey, S.V. and Rafalski, J.A. 1995a. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology* 5: 1023-1029

Powell, W., Morgante, M., McDevitt, R., Vendrami, G.G. and Rafalski, J.A. 1995b. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to population genetics of pines. *Proceedings of the National Academy of Sciences*, USA 92: 7759-7763

Ramakrishna, W., Chowdari, K.V., Lagu, M.D., Gupta, V.S. and Ranjekar, P.K. 1995. DNA finger printing to detect genetic variation in rice using hypervariable DNA sequences. *Theoretical and Applied Genetics* 90: 1000-1006

Redinbaugh, M.G., Jones, T.A. and Zhang, Y. 2000. Ubiquity of the St chloroplast genome in St-containing Triticeae polyploids. *Genome* 43: 846-852

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497

Saitou, M. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425

Seberg, O. and Petersen, G. 1998. A critical review of concepts and methods used in classical genome analysis. *Botany Review* 64: 373-414

Small, R.L., Cronn, R.C. and Wendel, J.F. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australia Systematic Botany* 17: 145-170

Small, R.L. and Wendel, J.F. 2000. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium* L., Malvaceae). *Molecular phylogenet Evolution* 16: 73-84

Small, R.L. and Wendel, J.F. 2002. Differential evolutionary dynamics of duplicated paralogous *Adh* loci allotetraploid cotton (*Gossypium*). *Molecular Biology and Evolution* 19: 597-607

Small, R.L., Rybum, J.A., Cronn, R.C., Seelanan, T. and Wendel, J.F. 1998. The tortoise and the hare. Choosing between noncoding plastome and unclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301-1315

Small, R.L., Rybum, J.A. and Wendel J.F. 1999. Low levels of nucleotide diversity at homoeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). *Molecular Biology and Evolution* 16: 491-501

Soltis, D.E., Soltis, P.S. and Tate, J.A. 2003. Advances in the study of polyploidy since plant specisation. *New phytologist* 161: 173-191

Sun, G.L., Tang, H.M. and Salomon, B. 2006. Molecular diversity and relationship of

North American *Elymus trachycaulus* and the Eurasian *E. caninus* species. *Genetica* 127: 55-64

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

Sun, G.L., Shee, J. and Salomon, B. 2006b. Molecular diversity and relationship among *Elymus trachycaulus, E. subsecundus, E. virescens, E. violaceus,* and *E. hyperarcticus* (Poaceae: Triticeae) as determined by amplified fragment length polymorphism. *Genome* 49: 1160-1169

Sun, G.L., Salomom, B. and von Bothmer, R. 1997. Analysis of tetraploid *Elymus* species using wheat microsatellite markers and RAPD markers. *Genome* 40: 806-814

Sun, G.L., Daley, T. and Ni, Y. 2007. Molecular evolution and genome divergence at *RPB2* gene of the St and H genome in *Elymus* species. *Molecular biology and evolution* (in press)

Svitashev, S., Bryngelsson, T., Li, X. and Wang R.R.C. 1998. Genome-specific repetitive DNA and RAPD markers for genome identification in *Elymus* and *Hordelymus*. *Genome* 41: 120-128

Swofford, D.L. 2003. *PAUP: Phylogenetic analysis using parsimony, version 4 beta 10 win.* Sinauer Press, Sunderland, MA, USA.

Tajima, F. 1989. Statistical method for testing the neutral mutation of hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tajima, F. 1993. Simple Methods for Testing the Molecular Evolutionary Clock Hypothesis. *Genetics* 135: 599-607

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmouguin, F. and Higgins, D.G. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool. *Nucleic Acid Research* 25: 4876-4882

Tosa, Y. and Sasai, K. 1991: Analysis of resistance of Aegilops squarrosa on the wheat grass mildew fungus by using the gene- for-gene relationship. *Theoretical and applied genetics* 81: 735-739.

Vedel, F., Quetier, F. and Bayen, M. 1976. Specific cleavage of chloroplast DNA from higher plants by Eco RI restriction nuclease. *Nature* 263: 440-442

Wang, R.R.C., Dewey, D.R. and Hsiao, C. 1986. Genomic analysis of the tetraploid *Pseudoroegneria tauri*. *Crop Science* 26: 723-727

Warrilow, D. and Symons, R.H. 1996. Sequence analysis of the second-largest subunit of tomato RNA polymerase II. *Plant molecular Biology* 30: 337-342

Watterson, G.A. 1975. On the number of segregation sites in genetical models without recombination. *Theoretical population biology* 7: 256-276

Watson, L. and Dallwitz, M.J. 1992. The grass genera of the world. CAB International. Wallingford.

Wicker, T., Guyot, R., Yahiaoui, N. and Keller, B. 2003. CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. *Plant Physiology* 132:52-63

Xu, D.H. and 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 nuclear DNA. *Theoretical and applied genetics* 108: 1443-1448

Zhang, X.Q., Salomon, B., von Bothermer R. and Díaz, O. 2001. Patters and levels of genetic differentiation in North American populations of the Alaskan wheatgrass complex. *Hereditas* 133: 123-132

Ziegenhagen, B. and Fladung, M. 1997. Variation in the psbC gene region of gymnosperms and angiosperms as detected by a single restriction site polymorphism. *Theoretical and Applied Genetics* 94 (8): 1065-1071