Molecular Characterization of the **St** Genome in *Pseudoroegneria* Species and Their Relationships to the **Y** Genome in *Elymus* Species, and Origin of *Elymus caninus*

By
Chi Yan

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.

August, 2011, Halifax, Nova Scotia

Copyright Chi Yan, 2011

Approved:	Dr. Genlou Sun, Supervisor
Approved:	Dr. Tim Frasier, Supervisor committee
Approved:	Dr. Pierre Jutras, Supervisor committee
Approved:	Dr. Mirwais Qaderi, External examiner
Date:	August 22th, 2011

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-80945-7
Our file Notre référence
ISBN: 978-0-494-80945-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Molecular Characterization of the **St** Genome in *Pseudoroegneria* Species and Their Relationships to the **Y** Genome in *Elymus* Species, and Origin of *Elymus caninus*

by Chi Yan

Abstract

Although five basic genomes (**St**, **H**, **Y**, **P** and **W**) have so far been identified in species of the genus *Elymus*, the origin of the **Y** genome in species with an **StY** genome is still unknown and under debate. Previous studies suggested that the **St** and **Y** genomes may share a common progenitor genome. To test this hypothesis and to avoid under-sampling biases, we analyzed extensive accessions of diploid *Pseudoroegneria* (**St**) species using molecular markers. The results showed that diploid **St**-genome species have a high level of nucleotide variation, especially in *P. spicata*, and rejected the suggestion of the same origin of the **St** and **Y** genomes. In addition, the study on the origin of *Elymus caninus* clearly confirmed the **StH** genome constitution in *E. caninus* and indicated that both the **St** and **H** genomes in this species have multiple origins.

August 22th, 2011

Acknowledgements

I would like to express my first and foremost gratitude to my supervisor Dr. Genlou Sun. This thesis would not have been possible without his consistent and illuminating instruction, encouragement and financial support.

I would like to express my heartfelt appreciation to my committee members, Dr. Tim Frasier and Dr. Pierre Jutras for their supportive suggestions and comments in this project.

I thank other faculty and staff members of the Biology Department at Saint Mary's University for their help: Dr. Doug Strongmen, Dr. Zhongmin Dong, Janet White, Jing Yang, Carmen Cranley and Heidi de Boer.

Last my thanks would go to my family, especially my parents, Keqin Yan and Changwen Ma, my girlfriend Qianni Hu, and my fellow classmates, especially Jonathon Forest and Dollie Campbell, for their sincere supports and concerns on my study and life.

Table of Contents

Content of Tables

Content of Figures

Part 1. Molecular characterization of the St Genome in *Pseudoroegneria* species and their relationships to the Y genome in *Elymus* species

1. Introduction

In order to understand evolutionary patterns and history, the estimation of relationships among the hierarchy of individuals, populations and species is required. Plant systematics combines molecular and morphological data to explore the taxonomy of plants and the associated evolutionary process by revealing their phylogenetic relationships at all levels. Phylogenetic analyses in plants are routinely applied to test these evolutionary questions, and tracing the origin of polyploidy based on interspecific data. These analyses generally assume that intraspecific variation is smaller than interspecific variation, and require that within and between species, sample sizes are sufficiently large to capture variation at both levels (Garamszegi and Møller, 2010). Systematists deal with intraspecific variation in many different ways. For sequence analysis, it is common for scientists to sample a single individual per species or treat each individual or haplotype as a separate terminal taxon (Wiens and Servedio, 1998). This treatment may result in the potential risks of biases.

1.1. Hybridization and polyploidization

Hybridization and polyploidization have played an important role in the history of plant evolution, and contribute greatly to speciation (Cui *et al.*, 2006). Actually, polyploidization is a very common process, especially in plants. For instance, 70% of all angiosperms have experienced one or more episodes of polyploidization (Masterson, 1994; Wendel, 2000). Not only many economically important crops such as wheat, cotton and potatoes are polyploids; but also some insects, fish and amphibians (Mangus and Bengt, 2007).

Previous studies have reported that hybridization and chromosome doubling would create genetic shocks to which the newly formed allopolyploids may burst in vast intra- and inter-genomic changes. Many of these widespread genome-wide alterations in allopolyploids may arise from rapid loss and recombination of low-copy DNA, retrotransposon activation, DNA methylation pattern change and epigenetic gene silencing during or after polyploidization (Chen and Pikaard, 1997; Comai *et al.*, 2000; Lee and Chen, 2001).

Recent data suggest that these rapid genomic changes may lead to genetic asymmetry evolution resulting in conformity and convergent effects, which are caused by the inter-genome invasion of chromatin segments occurring in only one pair of chromosomes (chromosome-specific sequences) or in several chromosome pairs of one genome (genome-specific sequences) to the other in the alloploids (Shaked *et al.*, 2001; Ozkan *et al.*, 2001, 2002; Salina *et al.*, 2004; Han *et al.*, 2003, 2005). At the phenotype level, these changes may result in full fertility and stabilization of the hybrid condition and therefore

assist in forming a stable taxonomic entity in nature (Feldman and Levy, 2009). However, to expand such investigations of the evolutionary dynamics of genome-wide duplication, a clear and appropriate identification of phylogenetic relationships among taxa and genes, as well as genomic elements, are needed.

1.2. *Elymus* and its genome constitutions

1.2.1. Poaceae, Triticeae, *Elymus*

Tribes of the grass family Poaceae are often used as research materials in the studies of evolution, phylogeny, or taxonomy, since they include a number of plants of major economic importance and they have a complex evolutionary history worldwide.

Triticeae is one of these valuable tribes in Poaceae and contains the world's most economically important grasses, such as wheat and barley, as well as forage crops. While most crop plants are annuals, the majority of species within Triticeae are perennials, such as species in the *Elymus* and *Pseudoroegneria* genera, as well as the partially perennial genus *Hordeum*, which includes both perennials and annuals. The tribe combines a wide variety of biological mechanisms and genetic systems which make it an excellent group for research in evolution, and speciation in plants (Bothmer and Salomon, 1994).

The allopolyploid genus *Elymus*, which is the largest genus in tribe Triticeae, contains approximately 150 species that are widely distributed all over the world (Löve, 1984). Moreover, *Elymus* has its origins from other exclusively allopolyploid genera and is closely related to other genera in the Triticeae. More than half of the all known *Elymus*

species (approximately 80) have their origins in Asia, which is regarded as a center of high diversity. North America is another prominent habitat of *Elymus* species, where about 50 species are found with low barriers to interspecific hybridization (Mason-Gamer *et al.*, 2002; Zhang *et al.*, 2000). Due to its worldwide distribution, great economic value and complex genetic composition, *Elymus* is an ideal candidate for studying the evolution and polyploidy in plants.

1.2.2. Genome contents of *Elymus*

Cytological analyses have identified five basic genomes (**St**, **H**, **Y**, **P**, and **W**) (Genome symbols follow Wang *et al.*, 1994) in this genus. The **St** genome, found in all *Elymus* species, supposedly originated from the genus *Pseudoroegneria* (Nevski) Á. Löve. The **H**, **P**, and **W** genomes are derived from the genera *Hordeum* L., *Agropyron* Gaertn., and *Australopyrum* (Tzvelev) Á. Löve, respectively (Dewey, 1984; Jensen, 1990; Torabinejad and Mueller, 1993; Jensen and Salomon, 1995). Presently, no species that is diploid for the **Y** genome has been found, and the origin of the Y genome remains unknown, although it is a common *Elymus* genome in Central and East Asia (Dewey, 1971; Jensen, 1990; Torabinejad and Mueller, 1993).

1.3. *Pseudoroegneria* species from the St genome

Pseudoroegneria (Nevski) Á. Löve is a genus in Triticeae (Poaceae) that contains valuable forage and turf grasses of excellent quality. It is of importance due to its desirable characters such as its resistance to drought and salt, as well as its palatability to animals for rangeland seeding (Daubenmire, 1942; Dewey, 1984). Based on Dewey's classification system (Dewey, 1984), about 15 species are included in this genus which range throughout open rocky hillsides from the Middle East and Transcaucasia, across Central Asia and Northern China to western North America (Dewey, 1984).

Morphologically, the species in this genus exhibit caespitose growth, are long-anthered, can cross-pollinate and live in perennial habitats. Particularly, *P. strigosa* has long awns and equal glumes, *P. spicata* has slender awns with unequal glumes, while *P. tauri* and *P. libanotica* have no awns with unequal glumes, *P. stipifolia* has rough rachis densely covered by pricklets (Yen *et al.*, 2007). As a diploid genus, *Pseudoroegneria* is designated as the **St** genome and is the proposed donor genome of the **St** genome to the polyploid genera *Elymus*, *Elytrigia* and *Pascopyrum* (Dewey, 1984; Löve, 1984). Due to its importance, *Pseudoroegneria* has been the focus of vast phylogenetic and evolutionary research (Dewey, 1984; Löve, 1984; Hsiao *et al.*, 1995). However, though Larson *et al.* (2004) investigated the population structure and genetic diversity in *P. spicata* using AFLP data, little effort has yet been committed to the study of intra- and inter-species nucleotide diversity and chloroplast haplotype in this genus.

1.4. Debate of the Y genome origin

Polyploidy in *Elymus* includes tetraploidy (**StH**, **StY**, **StP**) and hexaploidy (**StHP**, **StYP**, **StStH**, **StHY**, *etc.*) (Dubcovsky *et al.*, 1997). Previous studies have revealed the molecular evolution and phylogeny of some *Elymus* species (Lu and Liu, 2005; Liu *et al.*, 2006; Mason-Gamer *et al.*, 2005, 2010; Sun *et al.*, 2008, Sun and Komatsuda, 2010). However, it is still unknown where the **Y** genome originated, although it is a common *Elymus* genome in Central and East Asia (Dewey, 1971; Jensen, 1990; Torabinejad and Mueller, 1993). Internal transcribed spacer (ITS) sequence data suggests that **Y** and **St** may share the same progenitor genome (Lu and Liu, 2005; Liu *et al.*, 2006). However, ITS, as a member of a multi-copy gene family, often fails as a marker in this respect due to the fact that it may suffer from sequence homogenization, which is caused by concerted evolution (Li and Zhang, 2002; Kovařík *et al.*, 2003). Therefore, single or low copy nuclear genes are the best choice for phylogenetic studies since they do not undergo concerted evolution. Analyses of single- and low-copy nuclear genes including the phosphoenolpyruvate carboxylase (*PepC*), *β-amylase*, the granule-bound starch synthase I (*GBSSI*), the second largest subunit of RNA polymerase II (*RPB2*) and the translation elongation factor G (*EF-G*), have rejected the suggestion of the same origin of **St** and **Y** genome, but support Dewey's hypothesis that the **Y** genome has an independent origin from a **Y** diploid species, which is now extinct or undiscovered (Mason-Gamer *et al.*, 2005, 2010; Sun *et al.*, 2008, 2010).

More recently, Okito *et al.* (2009) used a random amplified polymorphic DNA (RAPD) based sequence tagged site (STS) as a marker to investigate the origin of the **Y**

genome with relatively large samples of *P. spicata* (**St** genome). The data show that one accession of *P. spicata* (PI 232134) may be the donor of the **Y** genome and the prime candidate as the origin of the **Y** genome to *E. longearistatus* (**StY**). This supports the suggestion of Liu *et al.* (2006). Therefore, since no study on the nucleotide diversity of the **St** genome species has been committed yet, whether the controversy on the origin of the **Y** genome in the **StY** *Elymus* species is caused by ignoring the possible high level of intraspecific variation within the *Pseudoroegneria* diploid species during sampling, needs to be examined.

1.5. The purpose of this study

In the present study, in order to avoid the possible biases when assessing the **Y** genome origin debate due to overlooking the intraspecific variation within *Pseudoroegneria* species, which may be caused by under sampling of diploid *Pseudoroegneria* species from the **St** genome, I first characterized nucleotide diversity and chloroplast haplotype for 32 accessions from 7 *Pseudoroegneria* species with the **St** genome using two single copy nuclear genes: *RPB2* and *EF-G,* along with one non-coding chloroplast DNA region, *TrnD/T.* Secondly, I further sampled sequences from 58 accessions of *Pseudoroegneria* (**St**), together with those from *Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), *Lophopyrum* ($\mathbf{E^e}$), *Thinopyrum* ($\mathbf{E^a}$), *Thinopyrum* ($\mathbf{E^b}$), and *Dasypyrum* (**V**), as well as 15 allotetraploids representing the **StY** genome in the phylogenetic analyses, still using the two single copy nuclear genes *RPB2* and *EF-G.* The objectives of this study were: (1) to examine the levels of intra- and inter-species

nucleotide polymorphism in *Pseudoroegneria* species with the **St** genome; (2) to infer the chloroplast haplotype relationships among *Pseudoroegneria* species; (3) to test the hypothesis that the **St** and **Y** share a common progenitor genome; (4) to explore whether intraspecific variation during sampling would affect the result on identification of the origin of the **Y** genome in allotetraploid **StY** species; (5) to investigate the evolution of the *RPB2* and *EF-G* genes and identify the potential donor species in allopolyploid **StY** species.

2. Materials and Methods

2.1. Plant materials and DNA extraction

Twenty-five accessions of 8 diploid species representing the **St** genome were sequenced for the three genes (*RPB2*, *EF-G* and *TrnD/T*). The seeds were provided by the USDA (United States Department of Agriculture). Germinated seeds were transplanted to a sand-peat mixture, and the plants were maintained in a greenhouse. DNA was extracted from fresh young leaf tissues using the method of Junghans and Metzlaff (1990). *RPB2* and *EF-G* sequences for 15 polyploid Triticeae species representing the **StY** genome, together with 17 accessions representing the **St**, **H**, **W**, **P**, **E** and **V** genomes, along with *Bromus sterilis* as the outgroup were obtained from published data (Sun *et al.*, 2008; Sun and Komatsuda, 2010), and included in the phylogenetic analyses. Plant material with accession numbers, genomic constitutions, geographical origins, and GenBank identification numbers are presented in Table 1.

2.2. DNA amplification and sequencing

The single-copy nuclear genes *RPB2* and *EF-G,* and the cpDNA gene *TrnD/T* were amplified by polymerase chain reaction (PCR) using the primers P6F and P6FR (Sun *et al.*, 2007), cMWG699T3-2 and cMWG699T7-2 (Komatsuda *et al.*, 1999), TrnD and TrnT (Sun, 2002), respectively. The amplification profile for the *RPB2* gene is as follows: an initial denaturation at 95 °C for 4 min and 35-40 cycles of 95 °C for 40 sec, 51 °C for 30 sec, 72 °C for 90 sec. The cycling ended with 72 °C for 10 min. The PCR profile for amplifying *EF-G* gene was based on Komatsuda *et al.* (1999) except with a modified annealing temperature of 49 °C. The PCR protocol for *TrnD/T* followed Sun (2002). PCR products were purified using the QIAquickTM PCR purification kit (QIAGEN Inc.) according to the manufacturer's instructions.

PCR products were commercially sequenced by MACROGEN (Seoul, Korea). To confirm sequence validity, both forward and reverse strands were sequenced independently. To avoid any error, which would be induced by *Taq* DNA polymerase during PCR amplification, each sample was independently amplified two times and sequenced since *Taq* errors that cause substitutions are mainly random, and it is unlikely that any two sequences would share identical *Taq* errors to create a false synapomorphy.

Table 1. Taxa from *Bromus*, *Elymus*, *Hordeum*, *Pseudoroegneria*, *Lophopyrum*, *Thinopyrum*, *Agropyron*, *Australopyrum* and *Dasypyrum* used in this study

Species	**Accession No.**	**Genome***	**Origin**	***RPB2***	***EF-G***	***TrnD/T***
Ag. cristatum (L.) Gaertn.	PI 383534	P	Kars, Turkey	EU18743	GU982325	-
Aust. retrofractum (Vickery) Á. Löve	PI 533014	W	New South Wales, Australia	EU187482	GU982345	-
	PI 547363	W	New South Wales, Australia	EU187470	GU982347	-
	PI 531553	W	New South Wales, Australia	HQ231849	-	-
B. sterilis	PI 229595		Iran	HQ231839	-	+
	55777			-	AY836187	-
D. villosum (L.) P. Candargy	PI 368886	V	Gaziemir, Turkey	EU187471	-	-
H. bogdanii Wilensky	PI 499498	H	Inner Mongolia, China	EF596768	GU982334	-
	PI 499645	H	Xinjiang, China	EU18747	GU982335	-
H. stenostachys Godr.	H6439	H	Argentina	EU187473	-	-
P. ferganensis (Nevski) Á. Löve	H10248	St	Gissar mtns, Tajikistan	-	GU982369	-
P. libanotica (Hack.) D. R. Dewey	PI 330688	St	Sirak-Sar, Iran	EF596751	HQ231866	+
	PI 330687	St	Kandavan Pass, Iran	EF596753	-	-
	PI 401274	St	Saqqez, Iran	EF596752	-	-
	PI 228389	St	Iran	HQ231837	-	+
	PI 228390	St	Iran	HQ231838	HQ231862	+
P. spicata (Pursh) Á. Löve	PI 506274	St	Washington, United States	EF596746	GU982338	-
	PI 610986	St	Utah, United States	EF596747	GU982354	-
	PI 232128	St	Idaho, United States	HQ231840	HQ231863	+
	PI 232134	St	Whoming, United States	HQ231841	HQ231864	-
	PI 236669	St	British Columbia, Canada	HQ231842	-	-
	PI 286198	St	Washington, United States	HQ231843	HQ231865	-
	PI 516184	St	Oregon, United States	HQ231848	HQ231868	-
	PI 537379	St	Oregon, United States	HQ231851	HQ231870	-
	PI 537389	St	Washington, United States	HQ231852	HQ231871	-

Table 1. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*EF-G*	*TrnD/T*
	PI 539873	St	Idaho, United States	HQ231853	HQ231872	-
	PI 547154	St	Idaho, United States	HQ231854	HQ231873	-
	PI 547162	St	Oregon, United States	HQ231855	-	-
	PI 563869	St	Oregon, United States	HQ231856	HQ231874	+
	PI 563872	St	Montana, United States	HQ231857	HQ231875	+
	PI 598822	St	Colorado, United States	HQ231858	HQ231876	+
	PI 619445	St	Nevada, United States	HQ231859	-	+
	PI 232140	St	Montana, United States	-	-	+
P. stipifolia (Czern. ex Nevski) Á. Löve	PI 325181	St	Russian Federation	EF596748	GU982324	+
	PI 440095	St	Russian Federation	+	-	-
P. strigosa	PI 531752	St	Estonia	HQ231850	HQ231869	+
	W6 14049	St	Russian Federation	HQ231836	HQ231861	+
P. strigosa subsp. *aegilopoides*	W6 13089	St	Xinjiang, China	HQ231835	HQ231860	+
P. gracillima	PI 420842	St	Former Soviet Union	HQ231846	GU982329	+
	PI 440000	St	Russian Federation	HQ231847	HQ231867	+
P. tauri	PI 401324	St	Iran	HQ231844	-	+
	PI 401326	St	Iran	HQ231845	-	+
	PI 401330	St	Toward Ahar, Iran	-	GU982328	-
L. elongatum (Host) Á. Löve	PI 142012	E^e	Russian Federation	EU187439	-	-
T. bessarabicum (Savul. & Rayss) Á. Löve	PI 531712	E^b	Estonia	EU187474	GU982344	-
E. abolinii (Drobow) Tzvelev	PI 531554	StY	Xinjiang, China	EU187443	GU982339	-
				EU187444	GU982340	-
E. ciliaris (Trin.) Tzvelev	PI 564917	StY	Soviet Far East	EF596749	-	-
				EU187483	-	-
E. semicostatus (Nees ex Steud.) Melderis	PI 207452	StY	Afghanistan	EU187445	GU982318	-
				EU187446	GU982319	-

Table 1. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*EF-G*	*TrnD/T*
E. longearistatus (Boiss.) Tzvelev	PI 401280	StY	North of Tehran, Iran	EU187447	GU982326	-
				EU187448	GU982327	-
E. canaliculatus (Nevski) Tzvelev	H4123	StY	Swat District, Pakistan	EU187449	-	-
				EU187450	-	-
E. gmelinii (Ledeb.) Tzvelev	PI 610898	StY	Xinjiang, China	EU187451	GU982352	-
				EU187452	GU982353	-
E. caucasicus (Koch) Tzvelev	PI 531573	StY	Estonia	EU187453	GU982342	-
				EU187454	GU982343	-
E. pendulinus (Nevski) Tzvelev	H8986	StY	Tibet, China	EU187455	GU982367	-
				EU187456	-	-
E. praeruptus Tzvelev	H10218	StY	Gissar Mt., Tajikistan	EU187459	-	-
				EU187460	-	-
E. strictus (Keng) Á. Löve	PI 499476	StY	Lanzhou, China	EU187457	GU982330	-
				EU187458	GU982331	-
E. validus (Meld.) B. Salomon	H4100	StY	Hazara, Pakistan	EU187461	-	-
				EU187462	-	-
E. antiquus (Nevski) Tzvelev	PI 619528	StY	Sichuan, China	EU187463	GU982355	-
				EU187464	GU982356	-
E. barbicallus (Ohwi) S.L.Chen	PI 504441	StY	Shaanxi, China	EU187465	-	-
				EU187466	-	-
E. fedtschenkoi Tzvelev	PI 564927	StY	Alma Ata, Kazakhstan	EU187467	GU982349	-
				EU187468	-	-
E. tibeticus (Meld.) G. Singh	PI 639828	StY	Xizang, China	EU187469	-	-
				EU187481	-	-

*Note: The genome designations are according to Wang *et al.* (1994).

2.3. Data analysis

The chromatographs of automated sequence results were compared visually. Multiple sequence alignments were conducted using ClustalX with default parameters and additional manual edits to minimize gaps (Thompson *et al.*, 1997). Nucleotide variations of diploid *Pseudoroegneria* species were estimated by Tajima's (1989) π and Watterson's (1975) θ statistics. The former 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. Tests of neutral evolution were performed as described by Tajima (1989), and Fu and Li (1993). The above calculations were performed using the software program DnaSP 4.0 (Rozas *et al.*, 2005). A diploid *Pseudoroegneria* species haplotype network based on mutational steps of *TrnD/T* sequence data was created to evaluate possible genetic relationships between haplotypes using TCS (Clement *et al.*, 2000).

Phylogenetic relationships were estimated using maximum-parsimony (MP), maximum-likelihood (ML) and Bayesian analyses. Each analysis was performed twice: one only included sequence data from the diploid *Pseudoroegneria* species; the other included all sequences as listed in Table 1 for assessing the **Y** genome origin debate. Phylogenetic analysis using the MP method was performed with the computer program PAUP* ver. 4 beta 10 (Swofford, 2003). All characters were specified as unweighted and unordered, and gaps were excluded in the analyses. Most-parsimonious trees were obtained by performing a heuristic search using the Tree Bisection-Reconnection (TBR) option with MulTrees selected, and ten replications of random addition sequences with

the stepwise addition option. Multiple parsimonious trees were combined to form a strict consensus tree. Overall character congruence was estimated by the consistency index (CI), and the retention index (RI). In order to infer the robustness of clades, bootstrap values with 1000 replications (Felsenstein, 1985a) were calculated by performing a heuristic search using the TBR option with MulTrees on.

In addition to maximum parsimony (MP) analysis, maximum-likelihood (ML) and Bayesian analyses were performed. For ML analysis, 8 nested models of sequence evolution were tested for all the *RPB2*, *EF-G* and *TrnD/T* data sets using PhyML 3.0 (Guindon and Gascuel, 2003). For each data set, the general time-reversible (GTR) (Lanave *et al.*, 1984) substitution model led to the largest ML score compared to the other 7 substitution models: JC69 (Jukes and Cantor, 1969), K80 (Kimura, 1980), F81 (Felsenstein, 1981), F84 (Felsenstein, 1993), HKY85 (Hasegawa *et al.*, 1985), TN93 (Tamura and Nei, 1993) and custom (data not shown). As a result, the GTR model was used in the Bayesian analysis using MrBayes 3.1 (Ronquist and Huelsenbeck, 2005). Default uniform priors were used for all model parameters (six substitution rates, four base frequencies, proportion of invariable sites, and alpha value of gamma distribution). One cold and three incrementally heated Markov Chain Monte Carlo (mcmc) chains were run simultaneously, each for both the two sequence data with the default heating value (0.2). In order to make the standard deviation of split frequencies fall below 0.01 so that the occurrence of convergence could be certain, 4,200,000 generations for *RPB2* data and 500,000 generations for *EF-G*, together with 1,000,000 generations for *TrnD/T* were run for phylogenetic analysis of diploid *Pseudoroegneria* species. For the study including

the data from the **Y** genome, 4,750,000 generations for *RPB2* data and 600,000 generations for *EF-G* were run. Samples were taken every 1000 generations under the GTR model with gamma-distributed rate variation across sites and a proportion of invariable sites. For all analyses, the first 25% of samples from each run were discarded as burn in to ensure the stationarity of the chains. Bayesian posterior probability (PP) values were obtained from a majority rule consensus tree generated from the remaining sampled trees.

3. Results

3.1. Sequence variation

The amplified products from each diploid species showed a single band for both the *RPB2* and *EF-G* sequences with a size of approximately 800~1000 bp and 1300 bp for *TrnD/T*, which corresponds well to previous findings (Sun *et al.*, 2008; Sun and Komatsuda, 2010; Sun 2002). Not only have extensive sequence variations been detected between the sequences from the **St** and **Y** genomes, but also among the sequences from the diploid **St** genome species.

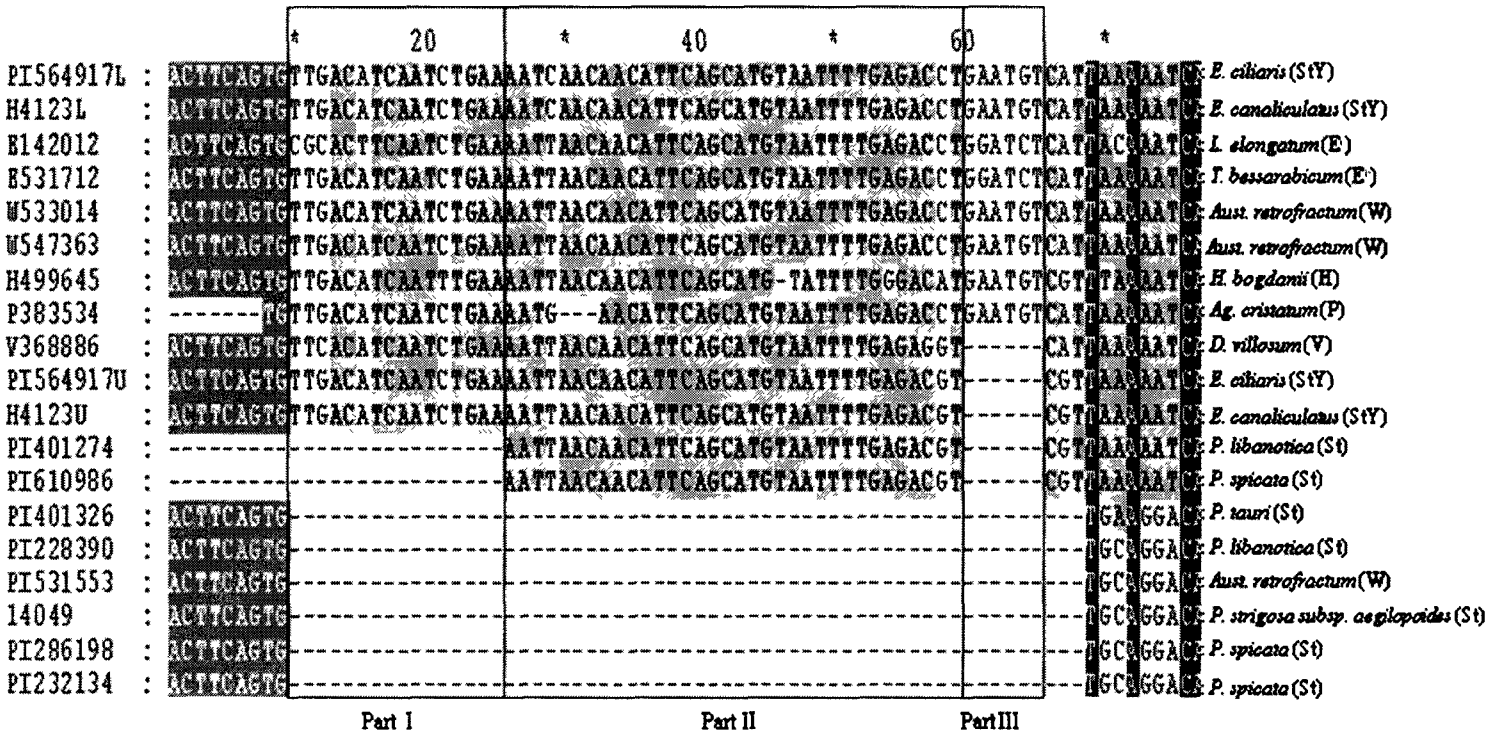
Notably, the sequence alignment showed a larger insertion/deletion (indel) in the *RPB2* data (Fig. 1) which occurred at position 10. However, some of the sequences were downloaded from Genbank and shorter than the others, so they were excluded for the comparison of this indel. In order to better compare the insertion/deletion among different species, we subdivided it into three parts as shown in the boxes (Fig. 1). The

sequences in part I were deleted in all diploid **St** genome sequences and two sequences from the **W** genome (PI 533014 and PI 547363) together with one sequence from the **H** genome (PI 499645), compared to the **St** and **Y** sequences from allotetraploid **StY** species and all the sequences from **P**, $\mathbf{E}^{b}$, $\mathbf{E}^{e}$ and **V** genomes. As for part II, all sequences from diploid **St** species except the three *P. libanotica* accessions (PI 330687, PI 330688 and PI401274) and one *P. spicata* accession (PI 610986) have 31 bp deletions, while the sequences from **St** and **Y** genomes in tetraploid *Elymus* species, and the sequences from the **H**, **W** (except PI 531553), **P**, $\mathbf{E}^{b}$, $\mathbf{E}^{e}$, and **V** genome species did not have this deletion. Part III is a 6 bp insertion that occurred in the **Y**, **H**, **W** (except PI 531553), **P**, and **E** genome sequences. None of the **St** genome contained the sequence GAATGT in this region.

In the *EF-G* sequence data, species-specific indels were found. The first is an 8 bp insertion (ATTTTCTC) at position 350 in all the sequences except those from the **Y**, **W** and $\mathbf{E}^{b}$ genomes. The second indel with a 7 bp insertion (GACTATT) at position 430 was only detected in sequences of *P. spicata*. Notably, a 36 bp deletion at position 660 was only found in one *P. libanotica* accession (PI 228390).

As for the 18 accessions of *TrnD/T* sequences, a 5 bp TATAT insertion/deletion was found at position 128 for three accessions of *P. spicata* (PI 232128, PI 563869, and PI 598818).

Fig. 1 Partial alignment of the amplified sequences of *RPB2* from *Elymus* **StY** species and their putative diploid donor species. Three uninterrupted indels, shown in different boxes: (I) A 16 bp insertion were deleted in all diploid **St** genome sequences and two sequences from **W** genome (PI 533014 and PI 547363) together with one sequence from **H** genome (PI 499645), compared to the **St** and **Y** sequences from allotetraploid **StY** species and all the sequences from **P**, $\mathbf{E}^{b}$, $\mathbf{E}^{e}$ and **V** genome. (II) All sequences from diploid **St** species, except for the three *P. libanotica* accessions (PI 330687, PI 330688 and PI 401274) and one *P. spicata* accession (PI 610986) have the 31 bp deletion compared to the sequences from **H**, **W** (except PI 531553), **P**, $\mathbf{E}^{b}$, $\mathbf{E}^{e}$, **V** and **StY** genome species. (III) A 6 bp insertion occurred in the **Y**, **H**, **W** (except PI 531553), **P**, and **E** genome sequences. None of the **St** genome contained the sequence (GAATGT) in this region.


* 20 * 40 * 60 *
PI564917L : E. ciliaris (StY)
H4123L : E. canaliculatus (StY)
B142012 : L. elongatum (E)
B531712 : T. bessarabicum (E')
W533014 : Aust. retrofractum (W)
W547363 : Aust. retrofractum (W)
H499645 : H. bogdanii (H)
P383534 : Ag. cristatum (P)
V368886 : D. villosum (V)
PI564917U : E. ciliaris (StY)
H4123U : E. canaliculatus (StY)
PI401274 : P. libanotica (St)
PI610986 : P. spicata (St)
PI401326 : P. tauri (St)
PI228390 : P. libanotica (St)
PI531553 : Aust. retrofractum (W)
14049 : P. strigosa subsp. aegilopoides (St)
PI286198 : P. spicata (St)
PI232134 : P. spicata (St)
Part I
Part II
PartIII

Estimates of nucleotide variations, π and θ_w, are shown in Table 2. The estimates of nucleotide variations in the **St** genome of all *Pseudoroegneria* species studied were θ_w = 0.15086 (*RPB2*), 0.10488 (*EF-G*), 0.04149 (*TrnD/T*); π = 0.13478 (*RPB2*), 0.04908 (*EF-G*), 0.02019 (*TrnD/T*). Among the *Pseudoroegneria* species studied, *P. spicata* exhibited the highest nucleotide variation in all three gene regions (θ_w = 0.11493, 0.09739, 0.04811; π = 0.10667, 0.06174, 0.03743 as for *RPB2*, *EF-G* and *TrnD/T*, respectively), while *P. tauri* showed the lowest (θ_w = 0.04577, 0.00183; π = 0.04577, 0.00183 for *RPB2* and *TrnD/T*, respectively). The Tajima (1989) and Fu and Li's (1993) tests were conducted on each data set. For all the three data sets, both Tajima's D and Fu and Li's D values were negative, which indicated an excess of low-frequency polymorphisms. The estimates of Tajima's D of *RPB2* data set were -1.28561 for *P. spicata*, and -0.67001 for *P. libanotica*. However, the Tajima's D and Fu and Li's D values of the *EF-G* gene and non-coding region *TrnD/T* data set were statistically significant, which suggests directional selection or bottleneck during evolution (Tajima, 1989).

Table 2. Estimates of nucleotide diversity and test statistics for selection at *RPB2*, *EF-G* and *TrnD/T* in diploid *Pseudoroegneria* from the **St** genome.

Species	**Gene**	**n**	**s**	π	θ_w	**Fu & Li's D**	**Tajima's D**
Pseudoroegneria	*RPB2*	396	251	0.13478	0.15086	-1.92579	-1.43206
	EF-G	682	264	0.04908	0.10488	-3.45064*	-2.44274*
	TrnD/T	1037	148	0.02019	0.04149	-3.26627*	-2.25217*
P. spicata	*RPB2*	620	263	0.10667	0.11493	-1.38561	-1.28561
	EF-G	761	230	0.06174	0.09739	-2.41473*	-2.10707*
	TrnD/T	1052	124	0.03743	0.04811	-1.70986*	-1.67077*
P. libanotica	*RPB2*	720	109	0.06931	0.07267	-0.57270	-0.67001
	EF-G	913	48	0.05257	0.05257		
	TrnD/T	1094	13	0.00823	0.00792		
P. tauri	*RPB2*	721	33	0.04577	0.04577		
	TrnD/T	1094	2	0.00183	0.00183		
P. stipifolia	*RPB2*	472	46	0.06568	0.06497		
P. gracillima	*RPB2*	788	100	0.12690	0.12690		
	EF-G	843	14	0.01661	0.01661		
	TrnD/T	1095	27	0.02466	0.02466		
P. strigosa	*RPB2*	790	54	0.06835	0.06835		
	EF-G	860	84	0.09767	0.09767		
	TrnD/T	1107	3	0.00271	0.00271		

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

*: Statistical significance

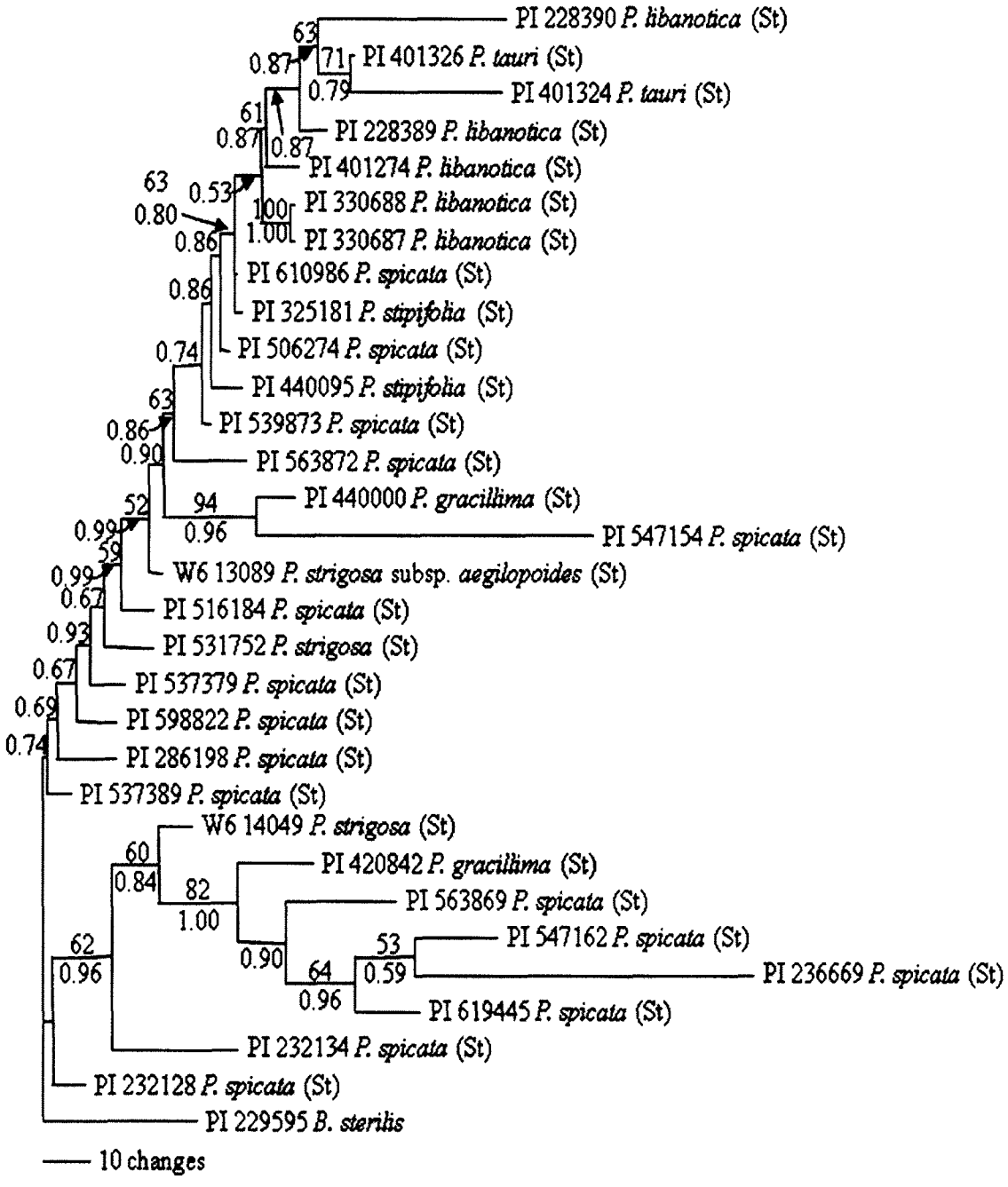
3.2. *Pseudoroegneria* species phylogeny and haplotype network analyses

3.2.1. Phylogenetic analyses of *Pseudoroegneria* species on the *RPB2* gene

The partial *RPB2* sequences that were obtained from the **St** genome were approximately 750~900 bp in length. The data matrix contained 646 characters, of which 186 were parsimoniously informative. Parsimony analysis produced 737 equally parsimonious trees with a consistency index (CI) of 0.643 and a retention index (RI) of 0.736. The separate Bayesian analyses using the GTR model resulted in identical trees with mean log-likelihood values -4556.23 and -4623.69 (data not shown). The tree topologies were almost identical in both ML and Bayesian trees and similar to those generated by MP, but only one of the most parsimonious trees with Bayesian PP and maximum parsimony bootstrap (BS) (1000 replicates) value is shown (Fig. 2).

Phylogenetic analyses separated the two accessions of *P. gracillima* from each other, PI 440000 formed a subclade (BS = 94%, PP = 0.96) with one accession of *P. spicata* (PI 547154), while the other accession of *P. gracillima* (PI 420842) also grouped with some accessions of *P. spicata* within well supported clade (BS = 82%, PP = 1.00). Five *P. libanotica* and two *P. tuari* accessions were grouped into a subclade (PP = 0.53).

Fig. 2. One of the 737 parsimonious trees derived from *RPB2* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.643, retention index (RI) = 0.736.

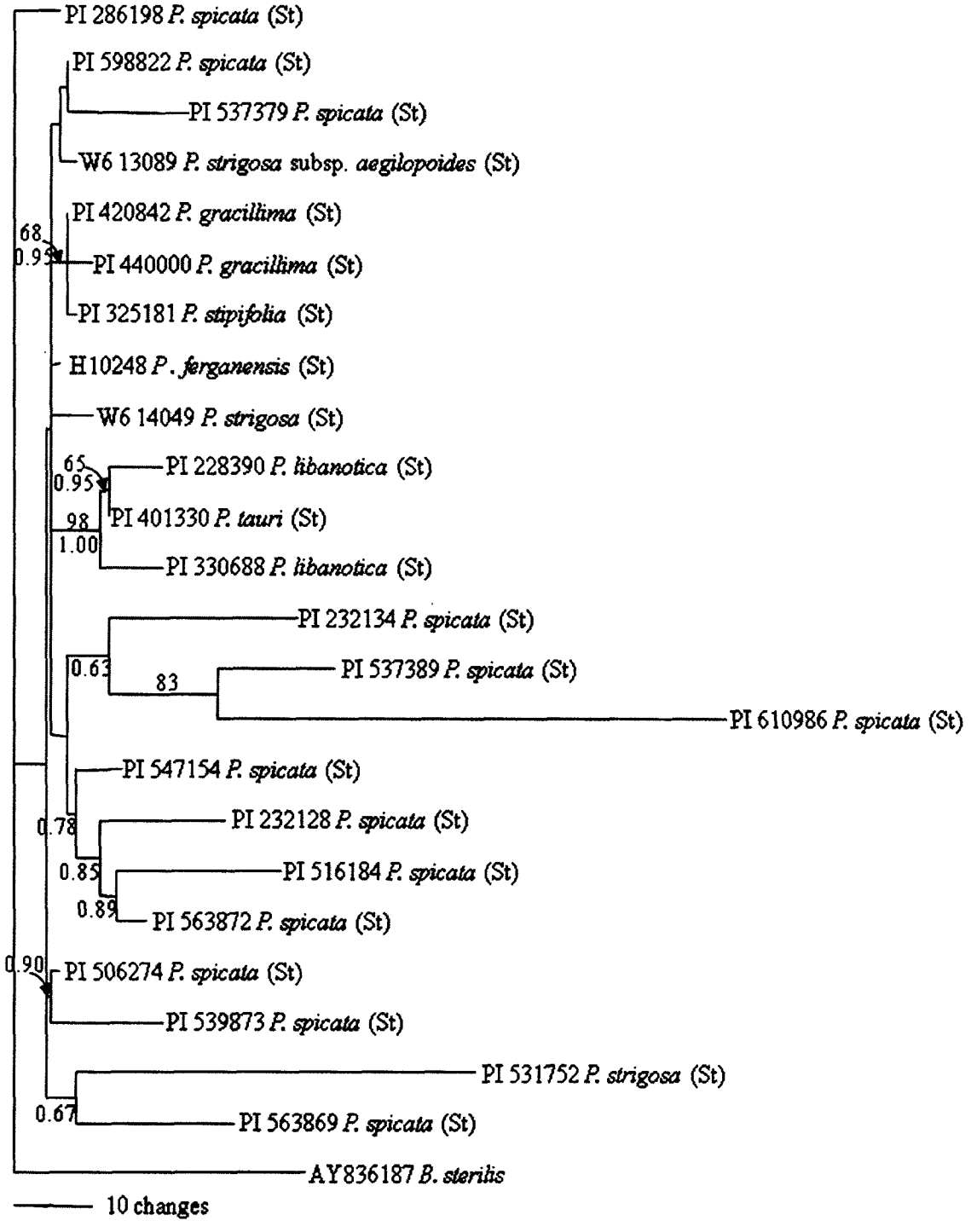

PI 228390 P. libanotica (St)
PI 401326 P. tauri (St)
PI 401324 P. tauri (St)
PI 228389 P. libanotica (St)
PI 401274 P. libanotica (St)
PI 330688 P. libanotica (St)
PI 330687 P. libanotica (St)
PI 610986 P. spicata (St)
PI 325181 P. stipifolia (St)
PI 506274 P. spicata (St)
PI 440095 P. stipifolia (St)
PI 539873 P. spicata (St)
PI 563872 P. spicata (St)
PI 440000 P. gracillima (St)
PI 547154 P. spicata (St)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 516184 P. spicata (St)
PI 531752 P. strigosa (St)
PI 537379 P. spicata (St)
PI 598822 P. spicata (St)
PI 286198 P. spicata (St)
PI 537389 P. spicata (St)
W6 14049 P. strigosa (St)
PI 420842 P. gracillima (St)
PI 563869 P. spicata (St)
PI 547162 P. spicata (St)
PI 236669 P. spicata (St)
PI 619445 P. spicata (St)
PI 232134 P. spicata (St)
PI 232128 P. spicata (St)
PI 229595 B. sterilis
10 changes

3.2.2. Phylogenetic analyses of *Pseudoroegneria* species on the *EF-G* gene

The partial *EF-G* sequences that we obtained from the **St** genome species were approximately 950~1100 bp in length. The data matrix contained 812 characters, of which 63 were parsimoniously informative. The parsimony analysis resulted in 433 most parsimonious tress (CI excluding uninformative characters = 0.894; RI = 0.511). The separated Bayesian analyses using the GTR model resulted in identical trees with mean log-likelihood values -3334.30 and -3345.14 (data not shown).

As shown in Figure 3, different from the *RPB2* sequences, MP and Bayesian analyses based on *EF-G* sequences clearly grouped the sequences from *P. gracillima* and *P. stipifolia* into one clade (BS = 68%, PP = 0.95). However, in the ML tree, one accession *P. gracillima* (PI 440000) was placed outside the *P. gracillima* + *P. stipifolia* subclade. Additionally, instead of being grouped with accession PI 537389 of *P. spicata* with 83% BS in the MP and Bayesian trees, the accession PI 610986 of *P. spicata* formed a subclade with the sequence of H10248 from *P. ferganensis* in the ML tree. Notably, all MP, ML and Bayesian trees of *EF-G* data set grouped *P. libanotica* + *P. tauri* together with well supported values (BS = 98%, PP = 1.00).

Fig. 3. One of the 433 parsimonious trees derived from *EF-G* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.894, retention index (RI) = 0.511.

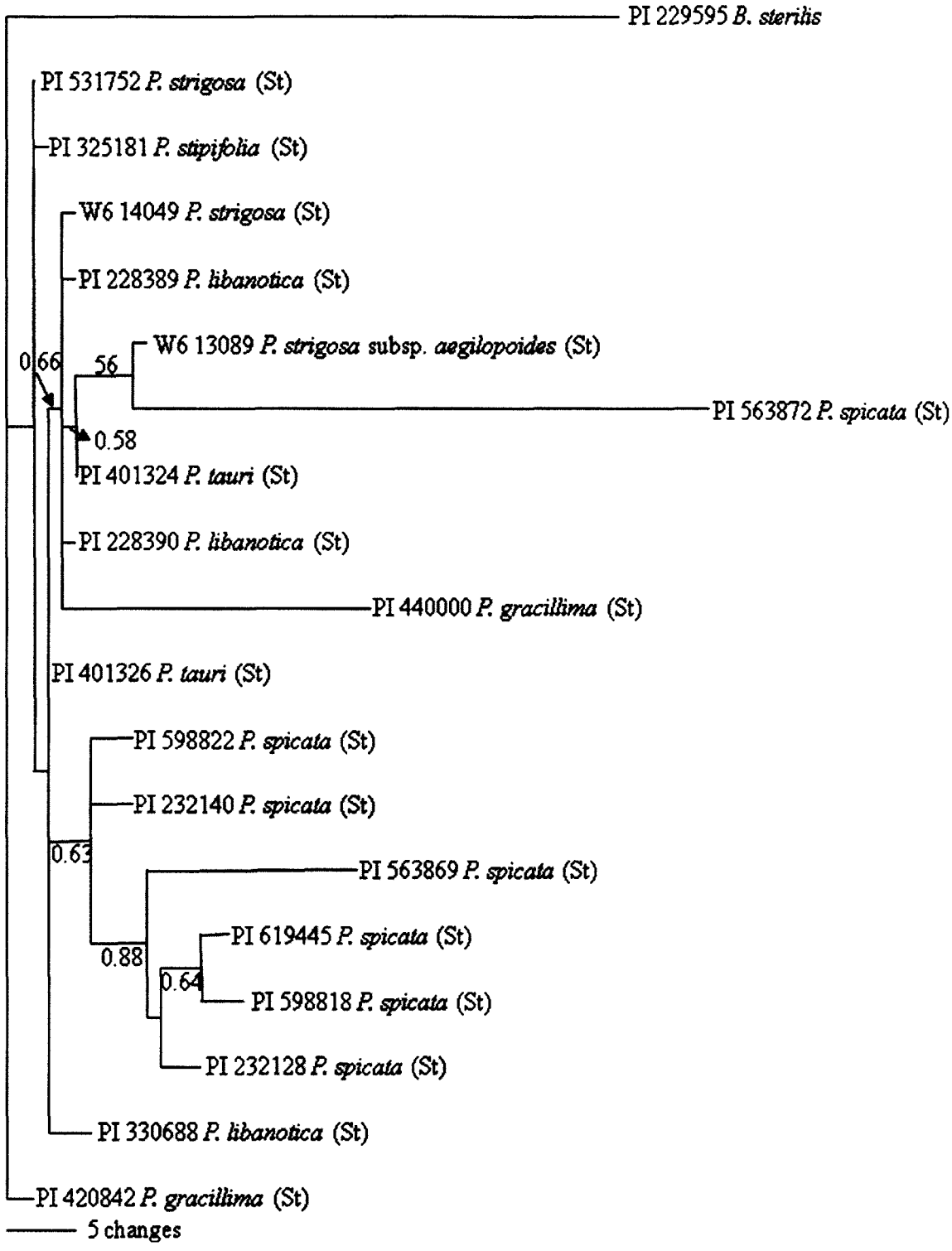

PI 286198 P. spicata (St)
PI 598822 P. spicata (St)
PI 537379 P. spicata (St)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 420842 P. gracillima (St)
68
0.99
PI 440000 P. gracillima (St)
PI 325181 P. stipifolia (St)
H10248 P. ferganensis (St)
W6 14049 P. strigosa (St)
65
0.95
PI 228390 P. libanotica (St)
98
1.00
PI 401330 P. tauri (St)
PI 330688 P. libanotica (St)
PI 232134 P. spicata (St)
0.63
83
PI 537389 P. spicata (St)
PI 610986 P. spicata (St)
PI 547154 P. spicata (St)
0.78
PI 232128 P. spicata (St)
0.85
PI 516184 P. spicata (St)
0.89
PI 563872 P. spicata (St)
0.90
PI 506274 P. spicata (St)
PI 539873 P. spicata (St)
PI 531752 P. strigosa (St)
0.67
PI 563869 P. spicata (St)
AY836187 B. sterilis
10 changes

3.2.3. Phylogenetic analyses of *Pseudoroegneria* species on the *TrnD/T* gene

The partial *TrnD/T* sequences that we obtained from the **St** genome were approximately 1100~1250 bp in length. The data matrix contained 1079 characters, of which 22 were parsimoniously informative. Parsimony analysis produced 208 equally parsimonious trees (CI excluding uninformative characters = 0.880; RI = 0.576). The separate Bayesian analyses using the GTR model resulted in identical trees with mean log-likelihood values -2624.84 and -2632.84 (data not shown). Since the tree topology results of MP, ML and Bayesian analysis were similar to each other, only one of the most parsimonious trees with BS and PP values is shown (Fig. 4).

Based on *TrnD/T* sequences data, all but one of the *P. spicata* accessions formed a fairly supported clade (PP = 0.63). The only one accession of *P. spicata* (PI 563872) outside this clade was grouped with *P. strigosa* subsp. *aegilopoides* (W6 13089). However, contrary to the *RPB2* and *EF-G* sequences, instead of forming a separated subclade, the sequences from *P. libanotica* and *P. tuari* did not group together.

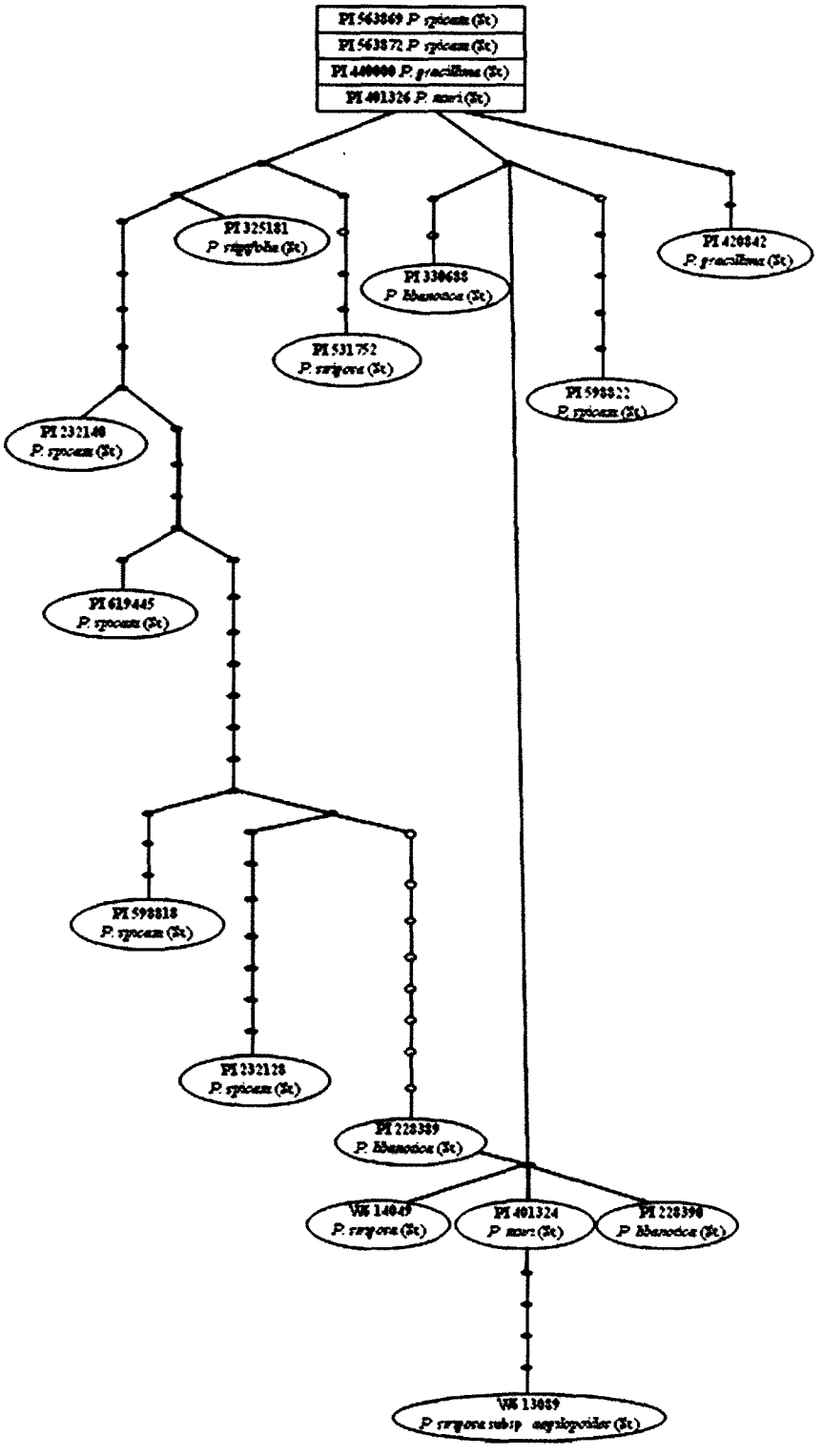
Fig. 4. One of the 208 parsimonious trees derived from *TrnD/T* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.880, retention index (RI) = 0.576.


PI 229595 B. sterilis
PI 531752 P. strigosa (St)
PI 325181 P. stipifolia (St)
W6 14049 P. strigosa (St)
PI 228389 P. libanotica (St)
W6 13089 P. strigosa subsp. aegilopoides (St)
0.66
56
PI 563872 P. spicata (St)
0.58
PI 401324 P. tauri (St)
PI 228390 P. libanotica (St)
PI 440000 P. gracillima (St)
PI 401326 P. tauri (St)
PI 598822 P. spicata (St)
PI 232140 P. spicata (St)
0.63
PI 563869 P. spicata (St)
PI 619445 P. spicata (St)
0.88
0.64
PI 598818 P. spicata (St)
PI 232128 P. spicata (St)
PI 330688 P. libanotica (St)
PI 420842 P. gracillima (St)
5 changes

3.2.4. Haplotype network analyses of *Pseudoroegneria* species on the *TrnD/T* gene

The TCS haplotype network (Fig. 5) which was based on mutational steps of 18 accessions from *Pseudoroegneria* species showed a clear grouping pattern and was similar to the MP tree based on *TrnD/T* data. The length of the branches between two nodes was meaningless here and they always imply one nucleotide difference. One difference between TCS haplotype network and the MP tree from *TrnD/T* data was that the haplotype diversity network indicated a much closer and clearer relationship of the *P. libanotica* + *P. tauri* group members. Moreover, one accession of *P. strigosa* (W6 14049) showed a similar haplotype with the *P. libanotica* + *P. tauri* group. As indicated in Table 4, based on the *TrnD/T* data set, the nucleotide polymorphism within *P. spicata* (θ_w = 0.04811) was even higher than the whole *Pseudoroegneria* genus (θ_w = 0.04149). As a result, *P. spicata* are widely distributed in the haplotype network and show a close link with *P. gracillima* and *P. strigosa*.

Fig. 5. TCS network of the *Pseudoroegneria* species haplotypes based on mutational steps from *TrnD/T* sequence data.


PI 563869 P. cyricum (St)
PI 563872 P. cyricum (St)
PI 440000 P. gracilima (St)
PI 401326 P. mori (St)
PI 325181
P. nigrifolia (St)
PI 330688
P. lhanotica (St)
PI 420842
P. gracilima (St)
PI 531752
P. sargova (St)
PI 598822
P. cyricum (St)
PI 232140
P. cyricum (St)
PI 619445
P. cyricum (St)
PI 598818
P. cyricum (St)
PI 232128
P. cyricum (St)
PI 228389
P. lhanotica (St)
VR 14045
P. sargova (St)
PI 401324
P. mori (St)
PI 228390
P. lhanotica (St)
VR 13059
P. sargova subsp. aegilopoides (St)

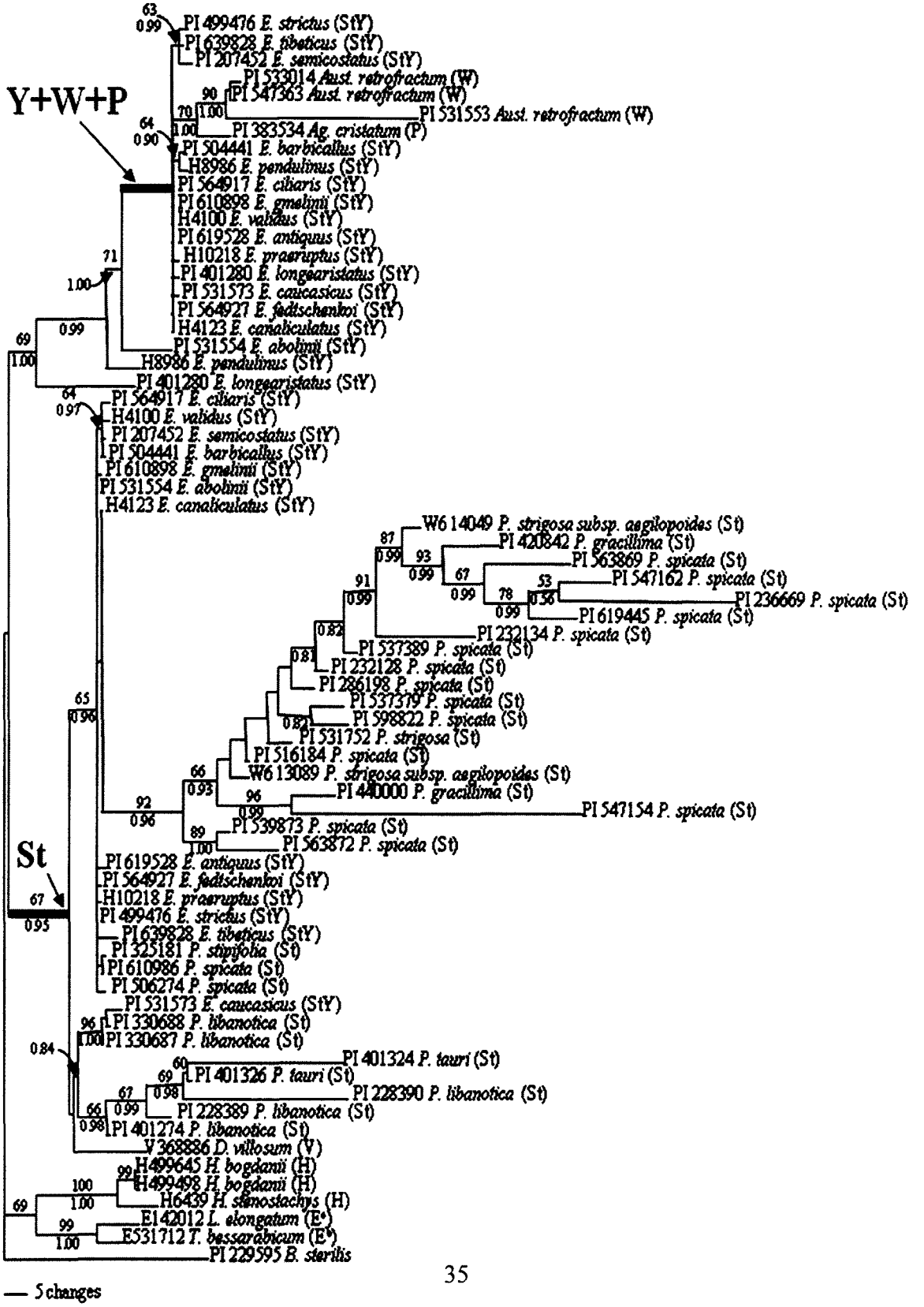
3.3. Phylogenetic analyses for the study of the Y genome origin

Since a high level of nucleotide variations, as well as complex phylogenetic relationships and haplotype networks were detected in the diploid *Pseudoroegneria* species, especially in *P. spicata*, our further phylogenetic study on the **Y** genome origin included all these diploid **St** genome species as potential genome donors in order to avoid under-sampling biases.

3.3.1. Phylogenetic analyses of *RPB2* sequences

Maximum parsimony analysis was conducted using *Bromus sterilis* as the outgroup. The parsimony analysis resulted in 826 equally most parsimonious trees (CI excluding uninformative characters = 0.650; RI = 0.836). The separate Bayesian analyses using the GTR model resulted in identical trees with mean log-likelihood values -5318.82 and -5490.62 (data not shown). The tree topologies were almost identical in both ML and Bayesian trees and similar to those generated by MP, but only one of the most parsimonious trees with Bayesian PP and maximum parsimony bootstrap (1000 replicates) value is shown (Fig. 6).

Fig. 6. One of the 826 parsimonious trees derived from *RPB2* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. Heavy internal branches are retained in the strict consensus tree. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.650, retention index (RI) = 0.836, rescale consistency index (RCI) = 0.543.


Y+W+P
St
PI 499476 E. strictus (StY)
PI 639828 E. tibeticus (StY)
PI 207452 E. semicostatus (StY)
PI 533014 Aust. retrofractum (W)
PI 547363 Aust. retrofractum (W)
PI 531553 Aust. retrofractum (W)
PI 383534 Ag. cristatum (P)
PI 504441 E. barbicallus (StY)
H8986 E. pendulinus (StY)
PI 564917 E. ciliaris (StY)
PI 610898 E. gmelinii (StY)
H4100 E. validus (StY)
PI 619528 E. antiquus (StY)
H10218 E. praeruptus (StY)
PI 401280 E. longearistatus (StY)
PI 531573 E. caucasicus (StY)
PI 564927 E. fedtschenkoi (StY)
H4123 E. canaliculatus (StY)
PI 531554 E. abolinii (StY)
W6 14049 P. strigosa subsp. aegilopoides (St)
PI 420842 P. gracillima (St)
PI 563869 P. spicata (St)
PI 547162 P. spicata (St)
PI 236669 P. spicata (St)
PI 619445 P. spicata (St)
PI 232134 P. spicata (St)
PI 537389 P. spicata (St)
PI 232128 P. spicata (St)
PI 286198 P. spicata (St)
PI 537379 P. spicata (St)
PI 598822 P. spicata (St)
PI 531752 P. strigosa (St)
PI 516184 P. spicata (St)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 440000 P. gracillima (St)
PI 547154 P. spicata (St)
PI 539873 P. spicata (St)
PI 563872 P. spicata (St)
PI 325181 P. stipifolia (St)
PI 610986 P. spicata (St)
PI 506274 P. spicata (St)
PI 330688 P. libanotica (St)
PI 330687 P. libanotica (St)
PI 401324 P. tauri (St)
PI 401326 P. tauri (St)
PI 228390 P. libanotica (St)
PI 228389 P. libanotica (St)
PI 401274 P. libanotica (St)
V368886 D. villosum (V)
H499645 H. bogdanii (H)
H499498 H. bogdanii (H)
H6439 H. stenostachys (H)
E142012 L. elongatum (E)
E531712 T. bessarabicum (E)
PI 229595 B. sterilis
5 changes

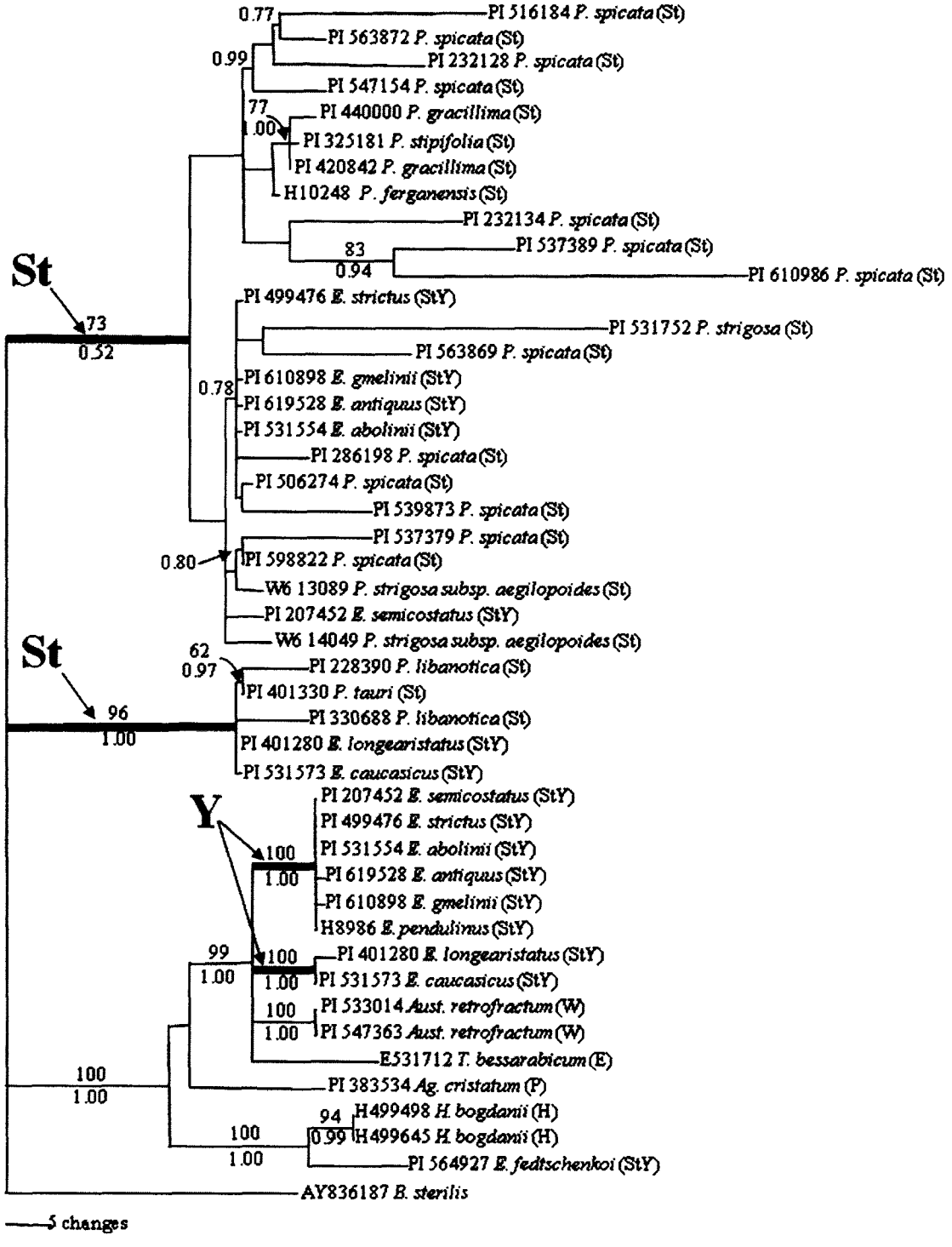
Phylogenetic analyses separated the sequences into three clades. All diploid species from the **St** genome and one sequence of *D. villosum* representing the **V** genome were grouped together in 67% BS (PP = 0.95). Included in the clade are the **St** genome sequences from tetraploid *Elymus* species (Fig. 6), while the **Y** copy sequences from tetraploid **StY** *Elymus* species formed a clade with the **W** and **P** genome species in 71% BS (PP = 1.00). The two copies of sequences from each *Elymus* species except *E. pendulinus* (H8986) and *E. longearistatus* (PI 401280) were well separated into two different clades, while the one each sequence from *E. pendulinus* (H8986) and *E. longearistatus* (PI 401280) were grouped with the presumed **Y**+**W**+**P** genome clade. The only difference among the MP, ML and Bayesian trees is that the **Y** copy sequence of *E. semicostatus* (PI 207452) was separated from the subclade with two other **Y** copy accessions (PI 639828 and PI 499476) in the ML tree while the three sequences were grouped together and fairly supported in the Bayesian and MP trees (63% BS, PP = 0.99).

Within the **St** (*Pseudoroegneria* + *Elymus*) clade, all *P. spicata* accessions except PI 610986 and PI 506274 formed a well supported subclade (92% BS, PP = 0.96) within which *P. strigosa* and *P. gracillima* were nested. The *P. spicata* sequences from accession PI 610986 and PI 506274, along with the **St** copy from **StY** tetraploid species were placed outside this subclade. The sequences from *P. tauri*, *P. libanotica* and *E. caucasicus* were grouped into a separate subclade.

3.3.2. Phylogenetic analyses of the *EF-G* sequences

The parsimony analysis resulted in 556 most parsimonious tress (CI excluding uninformative characters = 0.863; RI = 0.906). The Bayesian analyses using GTR model resulted in identical trees with mean log-likelihood values -4132.80 and -4508.50 (data not shown). The tree topologies generated by ML, MP and Bayesian analyses were similar to each other, but only one of the most parsimonious trees with BS and PP values is shown in Figure 7.

Fig. 7. One of the 556 parsimonious trees derived from *EF-G* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. Heavy internal branches are retained in the strict consensus tree. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.863, retention index (RI) = 0.906, rescale consistency index (RCI) = 0.783.


0.77
PI 516184 P. spicata (St)
PI 563872 P. spicata (St)
0.99
PI 232128 P. spicata (St)
PI 547154 P. spicata (St)
77
1.00
PI 440000 P. gracillima (St)
PI 325181 P. stipifolia (St)
PI 420842 P. gracillima (St)
H10248 P. ferganensis (St)
PI 232134 P. spicata (St)
83
0.94
PI 537389 P. spicata (St)
PI 610986 P. spicata (St)
St
73
0.52
PI 499476 E. strictus (StY)
PI 531752 P. strigosa (St)
PI 563869 P. spicata (St)
0.78
PI 610898 E. gmelinii (StY)
PI 619528 E. antiquus (StY)
PI 531554 E. abolinii (StY)
PI 286198 P. spicata (St)
PI 506274 P. spicata (St)
PI 539873 P. spicata (St)
PI 537379 P. spicata (St)
0.80
PI 598822 P. spicata (St)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 207452 E. semicostatus (StY)
W6 14049 P. strigosa subsp. aegilopoides (St)
St
62
0.97
PI 228390 P. libanotica (St)
PI 401330 P. tauri (St)
96
1.00
PI 330688 P. libanotica (St)
PI 401280 E. longearistatus (StY)
PI 531573 E. caucasicus (StY)
Y
PI 207452 E. semicostatus (StY)
PI 499476 E. strictus (StY)
100
1.00
PI 531554 E. abolinii (StY)
PI 619528 E. antiquus (StY)
PI 610898 E. gmelinii (StY)
H8986 E. pendulinus (StY)
99
1.00
100
1.00
PI 401280 E. longearistatus (StY)
PI 531573 E. caucasicus (StY)
100
1.00
PI 533014 Aust. retrofractum (W)
PI 547363 Aust. retrofractum (W)
E531712 T. bessarabicum (E)
100
1.00
PI 383534 Ag. cristatum (P)
94
0.99
H499498 H. bogdanii (H)
100
1.00
H499645 H. bogdanii (H)
PI 564927 E. fedtschenkoi (StY)
AY836187 B. sterilis
5 changes

Different from the *RPB2* sequences, Maximum Parsimony and Bayesian analyses based on *EF-G* sequences clearly separated the sequences from the **St**-genome species into two distinct clades (Fig. 7), one containing *P. tauri* and *P. libanotica*, as well as *E. longearistatus* and *E. caucasicus* with 96% BS support (PP = 1.00); another containing *P. spicata*, *P. gracillima*, *P. stipifolia*, *P. ferganensis*, *P. strigosa* and *Elymus* species. However, the ML tree based on *EF-G* sequences still combined these two subclades into one large clade. According to Figure 7, within the larger *Pseudoroegneria* + *Elymus* clade (BS = 73%), it contains a heterogeneous group of sequences. Moreover, the *P. spicata* accession PI 232134 that was suggested to be the **Y** genome donor to *E. longearistatus* (Okito *et al.*, 2009) is sister to the subclade (83% BS) of the other two *P. spicata* accessions (PI 537389 and PI 610986), and included in this *Pseudoroegneria* + *Elymus* clade.

All but one of the **Y** containing taxa plus the **E** and **W** genomes formed a well supported clade (BS = 99%, PP = 1.00). The **Y** genome clade was further divided into two subclades. The first contained *E. semicostatus*, *E. strictus*, *E. abolinii*, *E. antiquus*, *E. pendulinus*, and *E. gmelinii* with 100% bootstrap support (PP = 1.0). The second group comprised of *E. longearistatus* and *E. caucasicus* with 100% bootstrap support (PP = 1.0). The **Y** copy sequence from *E. fedtschenkoi* (PI 564927) did not group into the **Y** genome clade and instead appeared as a sister to the **H** genome sequences.

4. Discussion

4.1. Nucleotide divergence in *Pseudoroegneria* species

Although *Pseudoroegneria* species have long been the subject of extensive studies, little investigation has been conducted on nucleotide variation of species-wide samples of the diploid species using sequence data of single copy nuclear and chloroplast genes. In previous sequence analysis studies, less than 3 accessions for each of two or three *Pseudoroegneria* species were used to represent the **St** genome (e.g., Sun *et al.*, 2007; Zeng *et al.*, 2010). In the present study, we sampled 32 accessions from 7 diploid species within the **St** genome. Due to the increase of sample size, our data exhibits a significantly higher level of nucleotide variation compared to previous findings. Sun *et al.* (2007) used the *RPB2* gene to investigate nucleotide divergence of tetraploid *Elymus* **St**, **H** genomes and the diploid **St** genome, the estimates for the three *Pseudoroegneria* species (θ_w = 0.01640, π = 0.01711) were approximately one tenth of our present results (θ_w = 0.15086, π = 0.13478). In the study of Zeng *et al.* (2010), the *COX II* intron was used to estimate the nucleotide diversity for the genus *Kengyilia* (Triticeae: Poaceae) and its related species. However, since only one accession from each of the three *Pseudoroegneria* species was used, the estimates (θ_w = 0.00082, π = 0.00082) were considerably lower than those of our data (Table 2). The present data does, however, corresponded well with the study on nucleotide diversity of *P. spicata*, which showed relatively high rates of overall nucleotide variation (Larson *et al.*, 2000).

Based on this study, a high level of nucleotide polymorphism was detected in *Pseudoroegneria* species, especially among the different accessions of *P. spicata*. Among the 7 species from the **St** genome analyzed, *P. spicata* and *P. gracillima* exhibit a 2-fold higher level of sequence variation than *P. libanotica*, *P. strigosa* and *P. tauri*. Moreover, for all the *Pseudoroegneria* species, our data clearly show a higher polymorphism of single copy nuclear genes than the chloroplast DNA region.

4.2. Genetic relationships within and among *Pseudoroegneria* species

Our data could not clearly separate *P. spicata*, *P. gracillima* and *P. stipifolia* from each other. The MP, ML and Bayesian analyses based on the *RPB2* data set indicated a notable sequence separation among different accessions of *P. spicata* (Fig. 2). Therefore, this sequence dissimilarity may be the reason why only using one gene *RPB2* to amplify the seven accession sequences was successful. Comparatively, sequences from *P. strigosa* and *P. strigosa* subsp. *aegilopoides* show fairly close relationships with different *P. spicata* accessions. However, the sequences from these two species are not placed together in any of our phylogentic trees, which indicate the presence of nucleotide differences between each other and support the current subspecies definition. Notably, a closer relationship between *P. spicata* and *P. strigosa* sequences has been reported based on the *GBSSI* gene data by Mason-Gamer *et al.* (2010), which were interpreted as resulting from either past gene exchange or from the maintenance of ancestral polymorphism as a result of incomplete lineage sorting. *P. libanotica* and *P. tauri* exhibited high levels of sequence homology, and as a result, both the *RPB2* and *EF-G*

sequence data clearly separated the *P. libanotica* + *P. tauri* group from other **St**-genome species, which indicate the differentiation between the *P. libanotica* + *P. tauri* group with other diploid species from the **St** genome. The ITS data have also been used to evaluate the phylogenetic relationships of species in *Pseudoroegneria* and suggest that *P. libanotica*, *P. tauri*, *P. strigosa* subsp. *aegilopoides* and *P. spicata* are in the same group (Yu *et al.*, 2008). Since only one accession from *P. spicata* was used by Yu *et al.* (2008), considering the high level of nucleotide polymorphism of *P. spicata*, their results may not reflect the true genetic relationships of these species. Moreover, the separation of *P. libanotica* from *P. spicata* is expected since previous AFLP analysis indicated great differences between the two, and the low level of inter-specific similarity coefficients between *P. libanotica* and *P. spicata* (Larson *et al.*, 2004). This separation between sequences from the *P. libanotica* + *P. tauri* group and *P. spicata* could also be supported by their different morphological divergences (Yen *et al.*, 2007) and geographic distributions since *P. spicata* are Western North American species, while *P. libanotica* and *P. tauri* originate from the Middle East.

Based on our *TrnD/T* data set, no clear *P. libanotica* + *P. tauri* subclade was found. The incongruence in the phylogenies from cpDNA *TrnD/T* and the two nuclear gene (*RPB2* and *EF-G*) data sets, especially with respect to the positions of the *P. libanotica* + *P. tauri* group and *P. spicata*, may be due to biparentally inherited nuclear genes that have a higher evolutionary rate than maternally inherited cpDNA genes.

4.3. Possible origin of the Y genome

Since a high level of nucleotide variation was detected in the diploid *Pseudoroegneria* species, especially in *P. spicata*, our further study on the **Y** genome origin included all these diploid **St**-genome species in order to avoid under-sampling biases. The phylogenetic analyses of **StY** *Elymus* species and as many as 32 accessions from 7 diploid **St**-genome species provided additional support for the distinct origin of the **Y** genome in polyploid **StY** species. Both *RPB2* and *EF-G* phylogenetic trees clearly separate the **Y** genome from the **St** genome. These results are in accordance with the previous findings by Mason-Gamer *et al.* (2005, 2010) and Sun *et al.* (2008), and support Dewey's hypothesis that there is a **Y** diploid species from which the **Y** genome originated. These results conflict with the suggestion that the **St** and **Y** genomes have the same origin, which has been made based on ITS data by Liu *et al.* (2006). Recently, Okito *et al.* (2009) analyzed many accessions of diploid species *P. spicata* (**St** genome), and found that one accession of *P. spicata* (PI 232134) may be the donor of the **Y** genome and the prime candidate as the origin of the **Y** genome to *E. longearistatus* (**StY**). In our study, the accession of *P. spicata* (PI 232134) was included, both *RPB2* and *EF-G* phylogenetic trees have placed the accession of *P. spicata* (PI 232134) in the **St** genome together with other *Pseudoroegneria* species, indicating that there is no close link between the **St** genome in *P. spicata* and the **Y** genome in *E. longearistatus* as well as other **StY**-genome species. Since the **Y** genome was grouped with the **W** genome sequences in both the *RPB2* and *EF-G* trees, this indicates that the **W** genome is closely related to the **Y** genome.

Our result in the *RPB2* tree is that the presumed **St** genome copy of *E. longearistatus* (**StY**) was placed closely to the **Y** genome clade with a 69% BS, instead of being grouped with other **St**-genome sequences from tetraploid **StY** species. In addition, in the *EF-G* tree, the presumed **St**-genome copy of *E. longearistatus* (**StY**) has again showed a difference with other larger **St**-genome clade sequences and formed another well-supported clade (96% BS) with two *P. libanotica* individuals and *P. tauri*. This relative discrepancy within the sequence data between the **St** copy of *E. longearistatus* and the **St** genome clade sequences may be the reason why Okito *et al.* (2009) suggest that one accession of *P. spicata* (PI 232134) may be the candidate donor of the **Y** genome to *E. longearistatus* (**StY**). However, since the sequences in the **St** genome of *E. longearistatus* are quite different from the other **St** genome sequences, revealing that the phylogenetic relationship of this species clearly requires further study on a more comprehensive sample of this species. Moreover, in both the trees of *RPB2* and *EF-G* sequences, *P. libanotica* have grouped with the **St** copy of *E. caucasicus* (PI 531573) which are separated from the **St** genome containing other diploid **St** genome species. These data indicated that *P. libanotica* may be the donor of the **St** genome in allotetraploid *E. caucasicus* species.

4.4. Dense sampling of intraspecies accession does not affect the identification of genome origin in polyploids

Intraspecific variation is abundant in all types of systematic characters that could cause the risk of biases in phylogenetic analysis (Harmon and Losos, 2005). Systematists use different ways to deal with intraspecific variation including sampling a single individual per species, and treating each individual or haplotype as a separate terminal taxon for DNA data (Wiens and Servedio, 1998). There has been considerable debate as to which of the methods for directly analyzing polymorphic data is superior (e.g., Felsenstein, 1985b; Swofford *et al.*, 1996; Wiens and Servedio, 1997). A simulation study of phylogenetic method performance found that, overall, the most accurate methods were likelihood, the additive distance methods, and the frequency of parsimony method (Wiens and Servedio, 1998).

Okito *et al.* (2009) used relatively large samples from *P. spicata* (**St** genome) to investigate the origin of the **Y** genome, and suggested that one accession of *P. spicata* (PI 232134) may be the donor of the **Y** genome, which conflicts the previous findings of Mason-Gamer *et al.* (2005, 2010), Sun *et al.* (2008) and Sun and Komatsuda (2010) who used a few samples from the **St** genomes species. In most previous phylogenetic studies, only one or two accessions have been used to represent entire species data, however, this neglects the change of intraspecific variation, and may result in a biased conclusion. This is the first time we used as many as 31 accessions from 7 **St** genome species in the phylogenetic analysis to evaluate whether intraspecific variation could affect the phylogenetic result of the **Y** genome origin. MP, ML and Bayesian analyses reached the

same conclusion that the **St** and **Y** genomes have distinct origins. Although fairly large intraspecific variations have been detected in *P. spicata* (Larson *et al.*, 2004) and other diploid species, they do not influence the identification of the **Y** genome origin. It has been shown in previous studies that effective taxon sampling would be beneficial when analyzing the relationships across various levels of biological organization (e.g., genes, genomes, individuals, populations, species, or clades) due to poor taxon sampling leading to an increase in the apparent rate of variation which results in the overrepresentation of older nodes in the phylogenetic trees, and therefore, the bias caused by incomplete species sampling must be considered when using phylogenies to test hypotheses about species diversity (e.g., Hillis, 1996, 1998; Hedtke *et al.*, 2006; Heath *et al.*, 2008; McCormack *et al.*, 2009). However, since the genome-wide recombination would have a much greater variation than the ones of intraspecies accessions, the sample size of each species would not affect our investigation of the inter-genome questions, such as genome origination, if the minimal requirement (two or more representative accessions per species) is reached.

4.5. Genome evolution in allopolyploid StY species

The process of polyploidy occurs in cells and organisms when there are more than two paired (homologous) sets of chromosomes. During or after the process of allopolyploidization, rapid sequence elimination and restructuring of low-copy DNA, cytosine methylation, as well as the changes of transposable element activation and epigenetic gene silencing in allopolyploids shape the genomes in plants (Chen *et al.*,

1997; Comai *et al.*, 2000; Lee and Chen, 2001). In this study, extensive nucleotide changes and genome-wide indels have been found between the sequences of the diploid **St**-genome species and the allopolyploid **StY** species. The present study shows a 47 bp insertion in all the **St** copies of the *RPB2* sequences from allotetraploid **StY** species and all the diploid **St** species lack this insertion (Fig. 1 Part I and II) except the three *P. libanotica* accessions (PI 330687, PI 330688 and PI401274) and one *P. spicata* accession (PI 610986). One possible scenario is that the **St** genome in tetraploid **StY** species was donated by these four accessions. However, considering that the geographical distribution of *P. spicata* does not overlap with **StY** *Elymus* species, and phylogenetic analyses did not provide convincing evidence of that the accession PI 610986 of *P. spicata* and *P. libanotica* are the **St** donor species to all allotetraploid **StY** species analyzed here (Figs. 2 and 3), this explanation does not appear to be the case.

Another possible scenario is that the **St** genome in *Elymus* species acquired this part of the sequence by the inter-genome invasion of chromatin segments from the **Y** genome to the **St** genome and abundant genome-wide recombination following the fusion of **St** and **Y** gametes, before or after the process of polyploidization. It is also possible that the noted indels are homopasious, however, since the *RPB2* gene data from other genomic diploid species contain the same insertion as well, we could not rule out the possibility that this is due to the gene introgression between the **St** genome and other diploid species from the **W**, **H** or **E** genomes. Nevertheless, this theory of gene introgression is not consistent with some other data in this study. For instance, as the *RPB2* sequence data indicated, the tetraploid species *E. pendulinus* (H8986) and *E. longearistatus* (PI 401280)

representing the **StY** genome exhibited high similarity between the two copies from **St** and **Y** genomes. As the result, the **St** copies of *E. pendulinus* (H8986) and *E. longearistatus* (PI 401280) are sister with the presumed **Y**-genome clade (69% BS) in the MP, ML and Bayesian trees. One argument could be that the **St** and **Y** genome have a common origin, however, the majority of 13 other tetraploid **StY** species sequences contain distinct **St** and **Y** genomes. The similarities between the *RPB2* sequences in the **St** and **Y** genome in *E. pendulinus* (H8986) and *E. longearistatus* (PI 401280) could not be explained by the same origin of the **St** and **Y** genomes. Also, only one copy of *EF-G* sequence from *E. pendulinus* (H8986) was found even though more than ten clones were screened. Assuming no bias in cloning or PCR amplification, this gives a 99.9% chance of obtaining at least one copy of each of the two ancestral allelic types for the allotetraploid (Jakobsson *et al.*, 2006). The gene introgression event between the **St** and **Y** genome species and other diploid species could not explain this occurrence. Another explanation could be a vast genome-wide recombination between the **St** and **Y** genomes resulting in the two genome sequences at this location being identical to the extent that we could not distinguish one from the other, depending on the region being analyzed. Therefore, the explanation of genetic asymmetry evolution between the two parental genomes following polyploidization seems to be more likely in this study.

Previous research has shown genome-wide recombination of allotetraploids between their two constituent genomes in wheat species *T. turgidum* subsp. *dicoccoides* (Koern.) Thell., indicating that allopolyploid organisms can make "intelligent" choices about "selecting" the most efficient gene combination from one genome to control a set of

related traits (Feldman and Levy, 2009). Moreover, comparative chromosomal studies using genetic mapping and fluorescence in situ hybridization (FISH) have demonstrated that inter-genome invasion of chromatin segments can occur from the **B** genome into the **A** genome (e.g., Belyayev *et al.*, 2000). In the multiple independent synthetic lines of *Brassica napus* allotetraploids, genetic asymmetry evolution was reported by Gaeta *et al.* (2007). These authors found that in the allopolyploid group, convergent evolution of the two parental genomes would be exhibited due to the vast genomic combination. Additionally, in contrast to the multiple copy genes, such as ITS, single-copy genes may more easily suffer from the loci loss due to random events after polyploidization (Mahelka and Kopecky, 2010). The present study finds support for the genome convergent evolution in allopolyploid **StY** genome species.

4.6. Summary

Among the *Pseudoroegneria* species studied, the estimates of nucleotide diversity (π) ranged from 0.04577 (*RPB2*) and 0.00183 (*TrnD/T*) for *P. tauri* to 0.10667 (*RPB2*), 0.06174 (*EF-G*) and 0.03743 (*TrnD/T*) for *P. spicata*. The highest nucleotide diversity of the *RPB2* data set was found for *P. spicata* among the taxa analyzed. The phylogenetic analysis also suggests that *P. spicata*, *P. gracillima* and *P. stipifolia* have a closer relationship than the other species within *Pseudoroegneria*. *P. libanotica* and *P. tauri* were also found to exhibit a high level of sequence homology, however, only nuclear gene data (*RPB2* and *EF-G*) clearly indicated the differentiation between the *P. libanotica* + *P. tauri* group with other **St** genome species. The incongruence of the

phylogenies between the nuclear genes (*RPB2* and *EF-G*) and cpDNA *TrnD/T* data may be due to the lower evolutionary rate of cpDNA *TrnD/T*.

Since a high level nucleotide variation of *Pseudoroegneria* species were revealed, our further study on the **Y** genome origin included all these diploid accessions to avoid undersampling biases. Our results rejected the suggestion that the **St** and **Y** genomes share a common progenitor genome. However, based on the two nuclear gene sequence data, we still could not determine the exact origin and the clear location of the **Y** genome in the tribe Triticeae, even its close relationship with the **W** genome has been confirmed. As extensive accessions were sampled in the present study, we concluded that intraspecific variation during the sampling does not affect the identification of genome origin in polyploids, if the minimal requirement (two or more representative accessions per species) is reached. However, since a more sequence variability has been found in the sequences of *E. pendulinus* and *E. longearistatus*, to reveal the phylogenetic relationships and further investigate the genome convergent evolution of these species, clearly requires further study on more comprehensive widespread accessions.

Part 2. Gene duplication dynamics and origin of tetraploid StStHH *Elymus caninus* from various geographic regions

1. Introduction

1.1. Gene duplication in plants

Plant genomes have been shaped to a large extent by gene duplications. The duplication process and variation of existing genes play a critical role in the creation of new genes (Lynch and Conery, 2000; Zhang, 2003). Complex mechanisms have been proposed to explain the gene duplication and mutation processes for specific genes in cereal crops, such as in maize, sorghum, and rice genomes (e.g., Ilic *et al.*, 2003). One possible scenario to account for gene duplication processes is hybridization and polyploidization, which are very common in plants (Cui *et al.*, 2006). Alternatively, tandem duplication can also occur by an unequal crossover in repeated sequences flanking a single copy gene. In addition, this unequal crossover process always involves the movement of transposable elements, which are responsible for inducing rearrangement of the gene flanking regions (Habu *et al.*, 1997; Zhang and Gaut, 2003). For instance, two newly discovered transposons, Helitrons and Pack-MULEs, have been found to acquire and fuse fragments of plant genes and create the raw materials for the evolution of new genes and new genetic function (Morgante *et al.*, 2005; Ning *et al.*, 2004).

However, in both these two element types, the acquired gene fragments are all significantly smaller than most intact genes. Theoretically, it is improbable that a complete gene fragment acquisition will occur, and no such case has yet been found (Bennetzen, 2005). Therefore, these local sequence repeats should all be caused by so-called "segmental duplications". In general, by whatever scenario it occurs, there is no doubt that segmental gene duplication can have a fundamentally important role in evolution. In order to expand investigations of the evolutionary dynamics of gene duplication, clear and appropriate identification of phylogenetic relationships among taxa and genes, as well as genomic elements are needed.

1.2. *Elymus caninus* (L.) L.

Elymus caninus (L.) L. is a perennial, self-pollinating and allotetraploid species (2n= 4x = 28) that combines the genome of *Pseudoroegneria* (**St**) and *Hordeum* (**H**) (Dewey, 1968). It has a wide distribution, ranging from Iceland and the British Isles in the west to southern Siberia in the east and from the subarctic in the north to the Mediterranean in the south. It shows considerable variation in morphology, isozyme, prolamine and DNA level (Díaz *et al.*, 1998; Kostina *et al.*, 1998; Sun *et al.*, 1999, 2001). This makes *E. caninus* a suitable candidate for studying evolutionary dynamics and speciation.

1.3. The purpose of this study

In order to explore genome evolutionary dynamics and the origin of tetraploid *E. caninus*, 21 accessions of *E. caninus* from wide range of geographic regions were analyzed, together with 29 accessions of 16 diploid species from *Hordeum* (**H**), 38 accessions of 6 diploid species from *Pseudoroegneria* (**St**) and 43 accessions from other diploid Triticeae species. Two single copy nuclear genes: the second largest subunit of RNA polymerase II (*RPB2*) and the phosphoenolpyruvate carboxylase (*PepC*), along with one non-coding chloroplast DNA region *TrnD/T* were used. The objectives of this study were to: (1) explore whether the origin of **StStHH** genome in *E. caninus* is attributable to allopolyploidization alone; (2) identify which *Hordeum* and *Pseudoroegneria* species are possible progenitor species of *E. caninus*; and (3) investigate the possible relationship between species sequence variation and their different geographic regions.

2. Materials and Methods

2.1. Plant materials and DNA extraction

In the present study, DNA was extracted from fresh young leaf tissues using the method of Junghans and Metzlaff (1990). The accessions used were the same as those used by Sun *et al.* (1999), and Sun *et al.* (2006). Two nuclear gene (*RPB2* and *PepC*) and chloroplast *TrnD/T* sequences from different accessions of *E. caninus* were amplified and

sequenced. *RPB2* and *PepC* sequences for some diploid Triticeae species representing the **St**, **H**, **I**, **Xa**, **Xu**, **W**, **P**, **E** and **V** genomes along with *Bromus sterilis* were obtained from published data (Sun *et al.*, 2008; Helfgott and Mason-Gamer, 2004), and included in the analyses. Plant material with accession numbers, genomic constitutions, geographical origins, and GenBank identification numbers are presented in Table 3.

Table 3. Taxa from *Bromus*, *Aegilops*, *Eremopyrum*, *Heteranthelium*, *Psathyrostachys*, *Secale*, *Taeniatherum*, *Agropyron*, *Australopyrum*, *Dasypyrum*, *Thinopyrum*, *Triticum*, *Pseudoroegneria*, *Hordeum* and *Elymus* used in this study

Species	**Accession No.**	**Genome***	**Origin**	***RPB2***	***PepC***	***TrnD/T***
B. sterilis	PI 229595		Iran	HQ231839	-	+
B. tectorum	Kellogg s.n.			-	AY553239	-
Aegilops comosa Sibth. and Smith	PI 551032	M		-	-	+
	G 602	M		-	AY553236	-
Aegilops longissima Schweinf. & Muschl.	PI 542196	S^I		-	-	+
Aegilops sharonensis Eig	PI 542237	S^I		-	-	+
Aegilops speltoides Tausch	PI 499261	S		-	-	+
Eremopyrum bonaepartis (Spreng.) Nevski	PI 203442	F		-	-	+
Eremopyrum orientale (L.) Jaub. & Spach	PI 203440	F		-	-	+
	H 5555	F		-	AY553254	-
Heteranthelium piliferum (Banks & Sol.) Hochst.	PI 402352	Q		-	AY553255	-
Psathyrostachys juncea (Fischer) Nevski	PI 406469	Ns		-	-	+
Secale cereale L.	PI 573710	R		-	-	+
	Kellogg s.n.	R		-	AY553266	-
Taeniatherum caput-medusae	RJMG 189	Ta		-	AY553268	-
Ag. fragile (Roth) P. Candargy	PI 598674	P		-	-	+
Ag. mongolicum Keng	PI 598460	P		-	-	+
Ag. cristatum (L.) Gaertn.	PI 383534	P	Kars, Turkey	EU187438	-	-
	PI 279802	P		-	AY553237	-
Aust. retrofractum (Vickery) Á. Löve	PI 533013	W	New South Wales, Australia	-	-	+
	PI 533014	W	New South Wales, Australia	EU187482	-	+
	PI 547363	W	New South Wales, Australia	EU187470	-	+
	PI 531553	W	Austr. Capital, Australia	HQ231849	-	+
Aust. velutinum (Nees) B.K.Smion	D 2873-2878	W		-	AY553238	-
D. villosum (L.) P. Candargy	PI 368886	V	Gaziemir, Turkey	EU187471	-	-
	D 2990	V		-	AY553240	-
Thinopyrum elongatum (Host) D.R. Dewey	PI 142012	E^e	Odessa, Russian Federation	EU187439	-	-

Table 3. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*PepC*	*TrnD/T*
Thinopyrum elongatum	RJMG 113	E^e		-	AY553269	-
Thinopyrum bessarabicum (Savul. & Rayss) Á. Löve	PI 531712	E^b	Estonia	EU187474	-	-
Triticum monococcum L.		A^M		-	AJ007705	-
H. vulgare	RJMG 107	I		-	AY553260	+
H. vulgare ssp. *vulgare*	H 7514A	I	China	+	-	-
H. vulgare ssp. *spontaneous* (K.Koch) Thell	H 3140A	I	Cyprus	+	-	-
H. bulbosum L.	H 3878	I	Italy	+	-	-
	PI 440417	I		-	EU282294	-
				-	EU282295	-
				-	EU282296	-
H. marinum Huds.	PI 304346	Xa		-	AY553258	-
H. marinum	PI 304347	Xa		-	EU282298	-
H. marinum Huds. ssp. *marinum*	H 121	Xa	Greece	+	-	-
H. marinum Huds. ssp. *gussoneanum* (Parl.) Thell.	H 581	Xa	Greece	+	-	-
H. murinum L.	PI 247054	Xu		-	EU282299	-
				-	EU282300	-
H. murinum	CIho 15683	Xu		-	AY553259	-
H. murinum L. ssp. *glaucum* (Steud.) Tzvel.	H 74	Xu	Egypt	+	-	-
	H 52	Xu	Jordan	+	-	-
H. pusillum Nutt.	H 2024	H	U. S. A.	+	-	-
	CIho 15684	H		-	EU282301	-
H. intercedens Nevski	H 1941	H	U. S. A.	+	-	-
H. euclaston Steud.	H 2148	H	Uruguay	+	-	-
H. muticum J. Presl.	H 6479	H	Argentina	+	-	-
H. erectifolium Bothmer, Jacobsen and Jørgensen	H 1150	H	Argentina	+	-	-
H. stenostachys Godr.	H 1780	H	Argentina	+	-	-
	PI 531791	H		-	EU282302	-

Table 3. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*PepC*	*TrnD/T*
H. cordobense Bothmer, Jacobsen and Nicora	H 6460	H	Argentina	+	-	-
H. flexuosum Steud.	H 2127	H	Uruguay	+	-	-
H. comosum Presl.	H 1181	H	Argentina	+	-	-
H. pubiflorum Hook. F.	H 1236	H	Argentina	+	-	-
H. bogdanii Wilensky	PI 499498	H	Inner Mongolia, China	EF596768	-	-
	PI 499645	H	Xinjiang, China	EU18747	-	-
	H4014	H	Pakistan	+	-	-
	PI 531762	H	Tajikistan	-	-	+
H. bogdanii	PI 531760	H		-	EU282293	-
H. stenostachys Godr.	H 6439	H	Argentina	EU187473	-	+
H. roshevitzii Bowden	H 9152	H	China	+	-	-
	H 10070	H		-	-	+
	H 7754	H		-	-	+
H. chilense Roem. and Schult.	H 1816	H	Chile	+	-	-
	PI 531781	H		-	EU282297	-
H. patagonicum (Haumann) Covas ssp. *patagonicum*	H 6052	H	Argentina	+	-	-
H. patagonicum (Haumann) Covas ssp. *setifolium* (Parodi and Nicora) Bothm. *et al.*	H 1352	H	Argentina	+	-	-
H. patagonicum (Haumann) Covas ssp. *santacrucense* (Parodi and Nicora) Bothm. *et al.*	H 1353	H	Argentina	+	-	-
H. patagonicum (Haumann) Covas ssp. *magellanicum* (Parodi and Nicora) Bothm. *et al.*	H 1342	H	Argentina	+	-	-
H. patagonicum (Haumann) Covas ssp. *mustersii* (Nicora) Bothm. *et al.*	H 1358	H	Argentina	+	-	-
H. brachyantherum Nevski ssp. *californicum* (Covas and Stebbins) Bothm. *et al.*	H 3317	H	U. S. A.	+	-	-
P. libanotica (Hack.) D. R. Dewey	PI 330688	St	Sirak-Sar, Iran	EF596751	-	+
	PI 228389	St	Iran	HQ231837	-	+

Table 3. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*PepC*	*TrnD/T*
	PI 228390	St	Iran	HQ231838	-	+
	PI 228391	St		-	EU282304	-
P. libanotica	PI 282392	St		-	EU282305	-
P. spicata (Pursh) Á. Löve	PI 506274	St	Washington, United States	EF596746	-	-
	PI 610986	St	Utah, United States	EF596747	AY553263	-
	PI 232128	St	Idaho, United States	HQ231840	-	+
	PI 563869	St	Oregon, United States	HQ231856	-	+
	PI 563872	St	Montana, United States	HQ231857	-	+
	PI 598822	St	Colorado, United States	HQ231858	-	+
	PI 619445	St	Nevada, United States	HQ231859	-	+
	D 2844	St		-	AY553264	-
	PI 232140	St	U. S. A.	-	-	+
	PI 537377	St	Nevada, United States	+	-	-
	PI 595194	St	Montana, United States	+	-	-
	PI 598818	St	Oregon, United States	+	+	+
	PI 628742	St	Oregon, United States	+	-	-
	PI 628743	St	Alberta, Canada	+	-	-
	PI 628744	St	Alberta, Canada	+	-	-
	PI 636611	St	Alberta, Canada	+	-	-
P. stipifolia (Czern. ex Nevski) Á. Löve	PI 325181	St	Russian Federation	EF596748	-	+
	PI 440095	St	Russian Federation	+	-	-
	W6 21759	St	Krym, Ukraine	+	-	-
	PI 313960	St		-	EU282306	-
P. stipifolia	PI 531751	St		-	EU282307	-
				-	EU282308	-
P. strigosa	PI 531752	St	Estonia	HQ231850	-	+
	W6 14049	St	Russian Federation	HQ231836	-	+
P. strigosa (M.Bieb.) Á. Löve	PI 499637	St		-	EU282309	-

Table 3. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*PepC*	*TrnD/T*
				-	EU282310	-
P. strigosa subsp. *aegilopoides*	W6 13089	St	Xinjiang, China	HQ231835	-	+
P. strigosa subsp. *aegilopoides* (Drobow) Á. Löve	PI 531755	St		-	EU282311	-
P. gracillima	PI 420842	St	Former Soviet Union	HQ231846	-	+
	PI 440000	St	Stavro, Russian Federation	HQ231847	-	+
P. tauri (Boiss. & Balansa) Á. Löve	PI 380652	St		-	EU282312	-
P. tauri	PI 401324	St	Iran	HQ231844	-	+
	PI 401326	St	Iran	HQ231845	-	+
	PI 401319	St		-	EU282313	-
	PI 380644	St		-	EU282314	-
				-	EU282315	-
E. caninus (L.)L.	H 3169	StH	Västmanland, Sweden	St_1, H	-	+
	H 2009	StH	Fyledalen, Sweden	St_1, St_2, H	St, H	+
	H 7550-2	StH	China	-	-	+
	H 7550-3	StH	China	St_1, H	-	+
	H 7516	StH	China	-	-	+
	H 8753	StH	China	St_2, H	St, H	+
	H 8743	StH	China	-	-	+
	H 8745	StH	China	-	-	+
	H 10096	StH	Altai, Russia	St_2, H	-	+
	H 10404	StH	Siberia, Russia	St_1, St_2, H	St, H	+
	H 3858	StH	Siberia, Russia	-	-	+
	H 3857	StH	Siberia, Russia	St_1, H	-	+
	H 10115	StH	USSR, Russia	St_1, H	-	+
	H 10323	StH	KiL, Finland	-	-	+
	H 10314	StH	KiL, Finland	St_1, H	-	+
	H 4111	StH	Pakistan	-	St, H	+
	H 3915	StH	Slovakia	St_1, H	-	+

Table 3. (Continued)

Species	Accession No.	Genome*	Origin	*RPB2*	*PepC*	*TrnD/T*
	H 3698	StH	Italy	-	-	+
	H 10359	StH	Iceland	St_2, H	St, H	+
	H 10353	StH	Norway	-	-	+
	H 3339	StH	Kazakhstan	-	-	+

*Note: The genome designations are according to Wang *et al.* (1994).

2.2. DNA amplification and sequencing

The single and low copy nuclear genes *RPB2*, *PepC* and cpDNA gene *TrnD/T* sequences were amplified by polymerase chain reaction (PCR) using the primers P6F and P6FR (Sun *et al.*, 2007), *PEPC*-F and *PEPC*-R (Helfgott and Mason-Gamer, 2004), TrnD and TrnT (Sun, 2002), respectively. The amplification profiles for the *RPB2* gene and cpDNA gene *TrnD/T* were the same as Part 1. The PCR profile for amplifying the *PepC* gene was based on Helfgott and Mason-Gamer (2004). PCR products were purified using the QIAquickTM PCR purification kit (QIAGEN Inc.) according to the manufacturer's instructions.

Each PCR product of *E. caninus* amplified by the two nuclear genes was cloned into pGEM-easy T vector (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. 10-20 colonies for each accession were randomly selected for screening. Each was transferred to 10μL of LB broth with 0.1mg/ml antibiotics (AMP). These solutions were incubated at room temperature for 30 minutes before using 2 μL for PCR to check for the presence of an insert using the same primers that were used for original PCR amplification. For the solution that was confirmed to contain the insert, the remaining 8 μL of solution was transferred to 5 mL LB broth (with antibiotics) and incubated at 37°C overnight. Plasmid DNA was isolated using Promega Wizard ® *Plus* Minipreps DNA Purification System (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. The PCR products amplified by cpDNA primer *TrnD/T* were purified and then directly sequenced. Both the PCR products and Plasmid DNAs were commercially sequenced by MACROGEN (Seoul, Korea).

2.3. Data analysis

The chromatographs from the automated sequencing were compared visually. Multiple sequence alignments were made using ClustalX with default parameters (Thompson *et al.*, 1997). Phylogenetic analysis using maximum-parsimony (MP) was performed with the computer program PAUP* ver. 4 beta 10 (Swofford, 2003). In addition to maximum parsimony analysis, maximum-likelihood (ML) and Bayesian analyses were performed. All the detailed parameters and options of the three analyses were the same as Part 1. In the Bayesian analyses, in order to make the standard deviation of split frequencies fall below 0.01 so that the occurrence of convergence could be certain, 2,000,000 generations for *RPB2* data and 341,000 generations for *PepC*, together with 2,210,000 generations for *TrnD/T* were ran.

3. Results

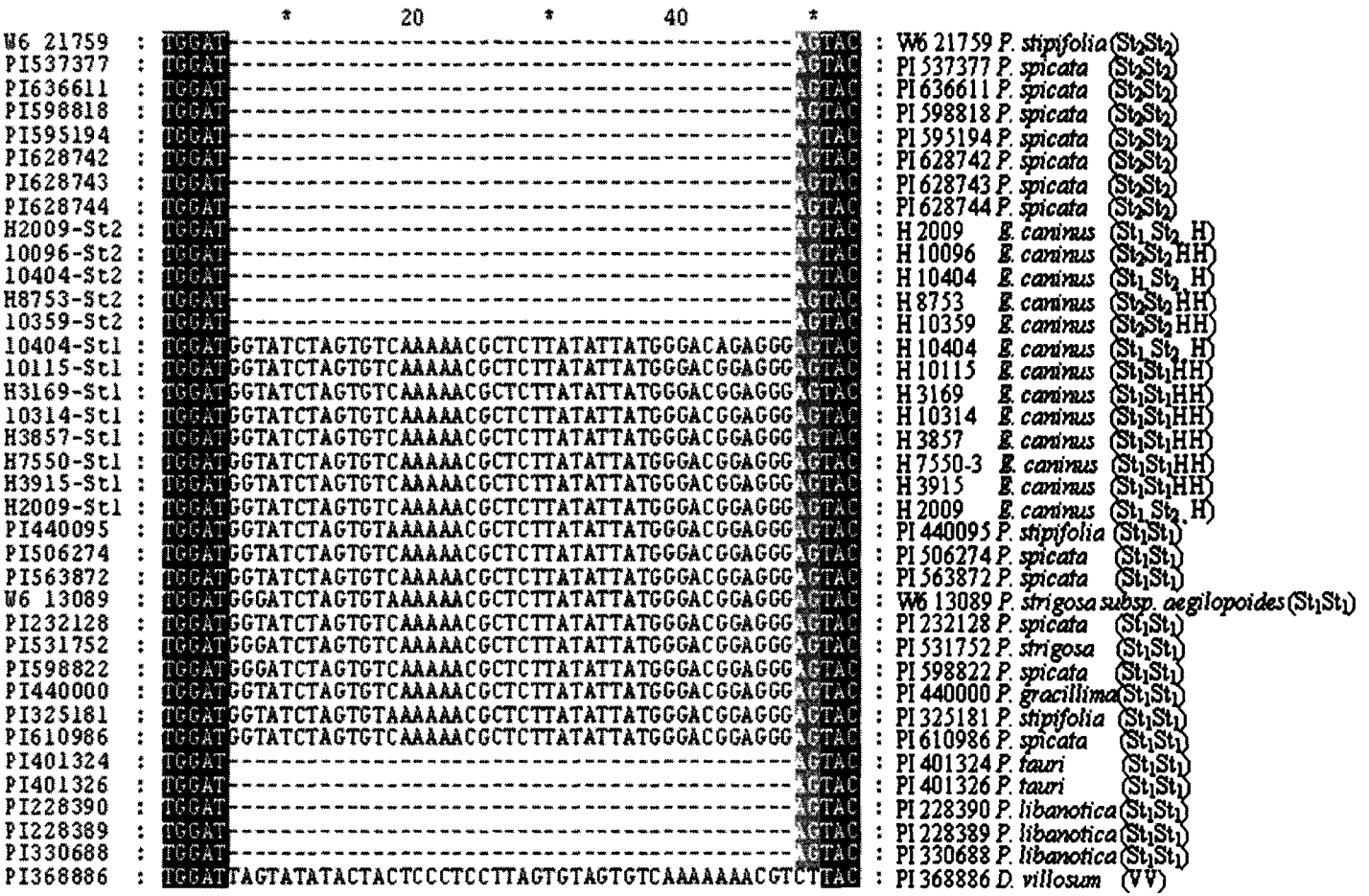
3.1. Sequence variation

The amplified products from each diploid species showed a single band for each of the three genes (*RPB2*, *PepC,* and *TrnD/T*) with sizes of approximately 1000, 2000 and 1300 bp, respectively, which correspond well with previous findings (Sun *et al.*, 2008; Helfgott and Mason-Gamer, 2004; Sun *et al.*, 2002). Not only have extensive sequence variations been detected between the sequences from the **St** and **H** genomes, but also among the sequences within them. The DNA from some accessions of *P. spicata* could

not be amplified regardless of PCR condition optimization or repeated DNA isolation. In order to better interpret sequence variations within the **St** genome, we further refer to the **St** genome sequences as $\mathbf{St_1}$ and $\mathbf{St_2}$ clade sequences based on phylogenetic analysis of the *RPB2* (see Fig. 10 in next section). Notably, though 8 accessions of *E. caninus* with $\mathbf{St_2}$ sequences were amplified at the *RPB2* gene, we could only succeed in amplifying one of them (PI 598818) using the primer set *PEPC*-F/*PEPC*-R and TrnD/TrnT. Furthermore, the *PepC* and *TrnD/T* genes were successfully amplified from the accession H4111 of *E. caninus*, but the *RPB2* gene could not be amplified from this accession.

Complete alignment of the 88 *RPB2* sequences confirmed a 43 bp insertion/deletion reported by Sun *et al.* (2007). The 5' flanking sequences of the 43 bp insertion were GGTA. The 5' flanking sequences for the sequences without 43 bp insertion were AGTA (Fig. 8). This 43 bp insertion occurred in all $\mathbf{St_1}$ genome sequences, except accessions from *P. libanotica* and *P. tauri*. None of the other genome ($\mathbf{St_2}$, **H**, **W**, **P**, **E**) sequences contained the insertion. The accession PI 368886 from the **V** genome had a completely different insertion in this region. In addition, a 9 bp insertion (ACAAAT/ATAT) has been deleted in all the American *Hordeum* diploid species with the **H** genome and the sequences from 6 *E. caninus* accessions (H3857, H7550-3, H10359, H2009, H10404 and H8753) (Fig. 10).

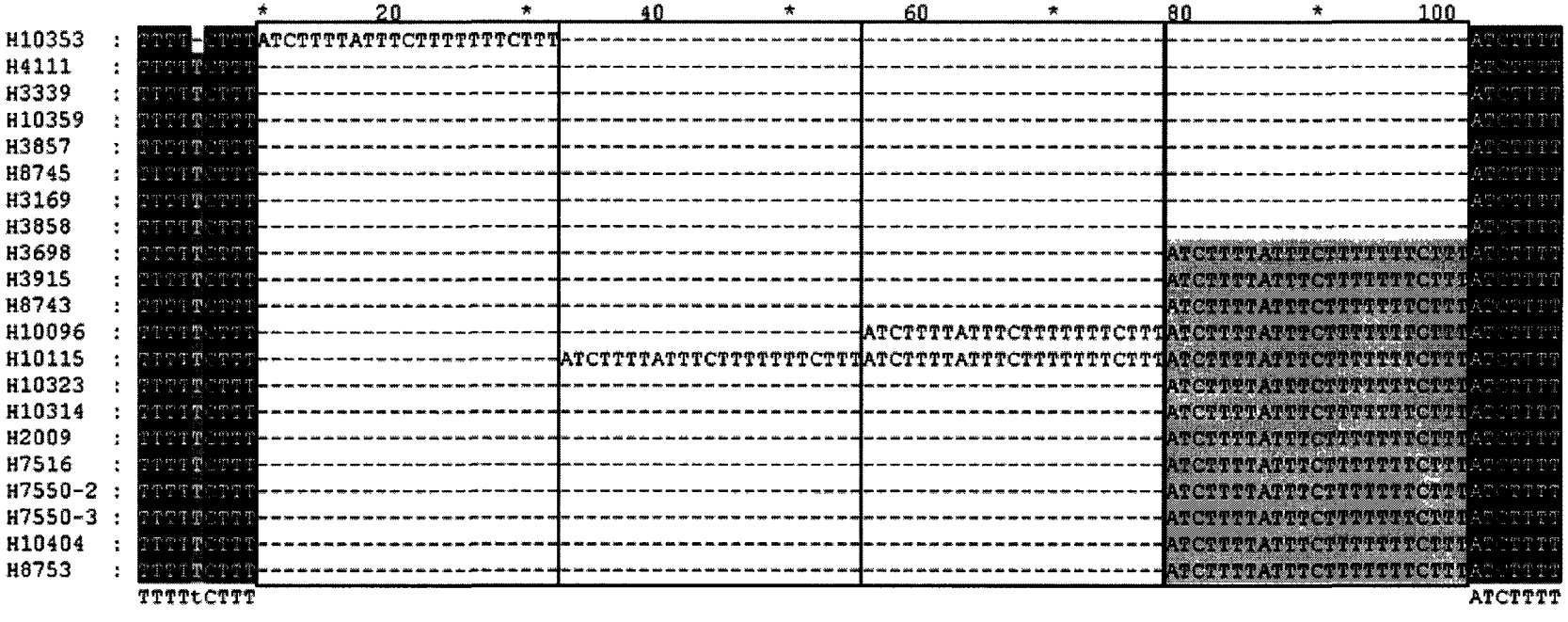
Fig. 8 Partial alignment of the amplified sequences of *RPB2* from *E. caninus* and their putative diploid donor species. A 43 bp insertion were only occurred in all $\mathbf{St_1}$ genome sequences, except accessions from *P. libanotica* and *P. tauri*. None of the other genome ($\mathbf{St_2}$, **H**, **W**, **P**, **E** etc.) sequences contained the insertion, but the accession PI 368886 from the **V** genome had a totally different insertion in this region.


* 20 * 40 *
W6 21759 : TGGAT----------GTAC : W6 21759 P. stipifolia (St2St2)
PI537377 : TGGAT----------GTAC : PI 537377 P. spicata (St2St2)
PI636611 : TGGAT----------GTAC : PI 636611 P. spicata (St2St2)
PI598818 : TGGAT----------GTAC : PI 598818 P. spicata (St2St2)
PI595194 : TGGAT----------GTAC : PI 595194 P. spicata (St2St2)
PI628742 : TGGAT----------GTAC : PI 628742 P. spicata (St2St2)
PI628743 : TGGAT----------GTAC : PI 628743 P. spicata (St2St2)
PI628744 : TGGAT----------GTAC : PI 628744 P. spicata (St2St2)
H2009-St2 : TGGAT----------GTAC : H 2009 E. caninus (St1 St2 H)
10096-St2 : TGGAT----------GTAC : H 10096 E. caninus (St2St2HH)
10404-St2 : TGGAT----------GTAC : H 10404 E. caninus (St1 St2 H)
H8753-St2 : TGGAT----------GTAC : H 8753 E. caninus (St2St2HH)
10359-St2 : TGGAT----------GTAC : H 10359 E. caninus (St2St2HH)
10404-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACAGAGGGGTAC : H 10404 E. caninus (St1 St2 H)
10115-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 10115 E. caninus (St1St1HH)
H3169-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 3169 E. caninus (St1St1HH)
10314-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 10314 E. caninus (St1St1HH)
H3857-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 3857 E. caninus (St1St1HH)
H7550-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 7550-3 E. caninus (St1St1HH)
H3915-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 3915 E. caninus (St1St1HH)
H2009-St1 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : H 2009 E. caninus (St1 St2 H)
PI440095 : TGGATGGTATCTAGTGTAAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 440095 P. stipifolia (St1St1)
PI506274 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 506274 P. spicata (St1St1)
PI563872 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 563872 P. spicata (St1St1)
W6 13089 : TGGATGGGATCTAGTGTAAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : W6 13089 P. strigosa subsp. aegilopoides (St1St1)
PI232128 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 232128 P. spicata (St1St1)
PI531752 : TGGATGGGATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 531752 P. strigosa (St1St1)
PI598822 : TGGATGGGATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 598822 P. spicata (St1St1)
PI440000 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 440000 P. gracillima (St1St1)
PI325181 : TGGATGGTATCTAGTGTAAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 325181 P. stipifolia (St1St1)
PI610986 : TGGATGGTATCTAGTGTCAAAAACGCTCTTATATTATGGGACGGAGGGGTAC : PI 610986 P. spicata (St1St1)
PI401324 : TGGAT----------GTAC : PI 401324 P. tauri (St1St1)
PI401326 : TGGAT----------GTAC : PI 401326 P. tauri (St1St1)
PI228390 : TGGAT----------GTAC : PI 228390 P. libanotica (St1St1)
PI228389 : TGGAT----------GTAC : PI 228389 P. libanotica (St1St1)
PI330688 : TGGAT----------GTAC : PI 330688 P. libanotica (St1St1)
PI368886 : TGGATTAGTATATACTACTCCCTCCTTAGTGTAGTGTCAAAAAAACGTCTTAC : PI 368886 D. villosum (VV)

In the *PepC* sequence data matrix, 49 sequences were analyzed, no large indel was detected within the **St** genome, however, a 17 bp insertion (CATTCATAATTGAATTA) at position 1800 has been found in all the sequences from the **St**, **P** and **F** genomes relative to other sequences analyzed in this study.

For the 59 *TrnD/T* sequences, a 23 bp tandem insertion (ATCTTTATTTCTTTTTTTTCTTT) occurred at position 32 in some accessions of *E. caninus* with repeat numbers varying from 1 to 3 (Fig. 9). Seven accessions of *E. caninus* (H4111, H3339, H10359, H3857, H8745, H3169 and H3858) and none of the other species sequences contained the insertion in this region.

Fig. 9 Partial alignment of the amplified sequences of *TrnD/T* from 21 *E. caninus* accessions. A 23 bp tandem insertion (ATCTTTATTTCTTTTTTTTCTTT) occurred in some accessions of *E. caninus* with copy number varing from 1 to 3. Seven accessions of *E. caninus* (H4111, H3339, H10359, H3857, H8745, H3169 and H3858) and none of the other species sequences contained the insertion in this region.

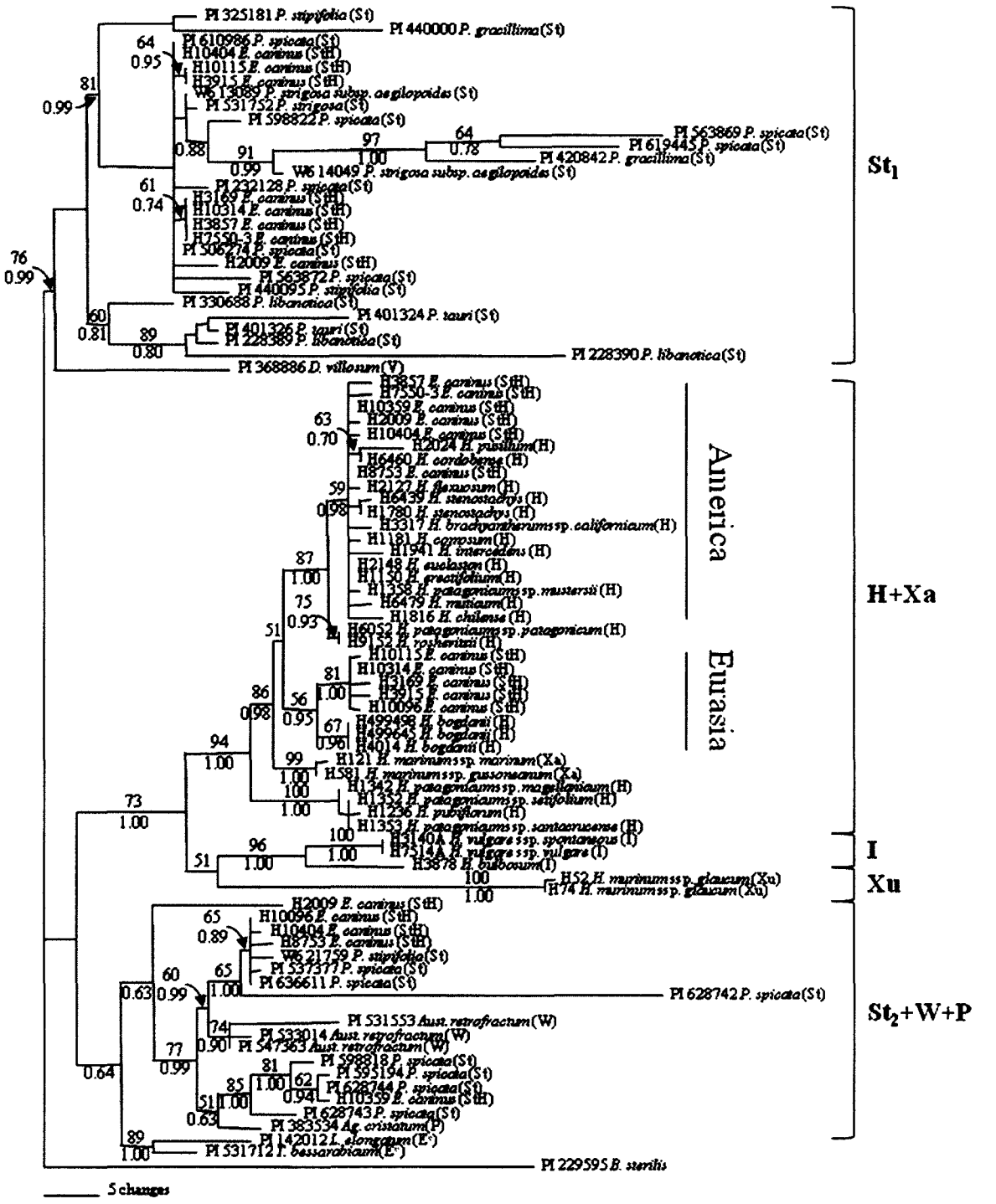

* 20 * 40 * 60 * 80 * 100
H10353 : ATCTTTTATTTCTTTTTTTCTTT
H4111 :
H3339 :
H10359 :
H3857 :
H8745 :
H3169 :
H3858 :
H3698 :
H3915 :
H8743 :
H10096 : ATCTTTTATTTCTTTTTTTCTTT
H10115 : ATCTTTTATTTCTTTTTTTCTTT ATCTTTTATTTCTTTTTTTCTTT
H10323 :
H10314 :
H2009 :
H7516 :
H7550-2 :
H7550-3 :
H10404 :
H8753 :
TTTTtCTTT
ATCTTTT

3.2. Phylogenetic analyses of *RPB2* sequences

Maximum Parsimony analysis using *Bromus sterilis* as the outgroup was conducted (152 parsimony-informative characters, 551 equally most parsimonious trees, consistency index, CI= 0.731; retention index, RI= 0.906). The separated Bayesian analyses using GTR model resulted in identical trees with mean log-likelihood values -4140.20 and -4364.95 (data not shown). The tree topologies were almost identical in both ML and Bayesian trees and similar to those generated by MP with minor difference. One of the most parsimonious trees with Bayesian PP and maximum parsimony bootstrap (1000 replicates) value is shown (Fig. 10).

Phylogenetic analyses clearly separated the sequences from *Hordeum* species (**H**, **Xa**, **I** and **Xu** genomes) into a clade (BS = 73%, PP = 1.00). While the sequences from the species with the **St** genome were separated into two clades that were named $\mathbf{St_1}$ and $\mathbf{St_2}$ clades (Fig. 10). The ML and Bayesian trees did not group the sequences from *Hordeum* (**H**, **Xa**, **I, Xu**) and the sequences from $\mathbf{St_2}$, **W**, **P** together (data not shown).

Fig. 10. One of the 551 parsimonious trees derived from *RPB2* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.731, retention index (RI) = 0.906, rescale consistency index (RCI) = 0.663.


PI 325181 P. stipifolia (St)
PI 440000 P. gracillima (St)
PI 610986 P. spicata (St)
H10404 E. caninus (StH)
H10115 E. caninus (StH)
H3915 E. caninus (StH)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 531752 P. strigosa (St)
PI 598822 P. spicata (St)
PI 563869 P. spicata (St)
PI 619445 P. spicata (St)
PI 420842 P. gracillima (St)
W6 14049 P. strigosa subsp. aegilopoides (St)
PI 232128 P. spicata (St)
H3169 E. caninus (StH)
H10314 E. caninus (StH)
H3857 E. caninus (StH)
H7550-3 E. caninus (StH)
PI 506274 P. spicata (St)
H2009 E. caninus (StH)
PI 563872 P. spicata (St)
PI 440095 P. stipifolia (St)
PI 330688 P. libanotica (St)
PI 401324 P. tauri (St)
PI 401326 P. tauri (St)
PI 228389 P. libanotica (St)
PI 228390 P. libanotica (St)
PI 368886 D. villosum (V)
St1
H3857 E. caninus (StH)
H7550-3 E. caninus (StH)
H10359 E. caninus (StH)
H2009 E. caninus (StH)
H10404 E. caninus (StH)
H2024 H. pusillum (H)
H6460 H. cordobense (H)
H8753 E. caninus (StH)
H2127 H. flexuosum (H)
H6439 H. stenostachys (H)
H1780 H. stenostachys (H)
H3317 H. brachyantherum ssp. californicum (H)
H1181 H. compsum (H)
H1941 H. intercedens (H)
H2148 H. euclasion (H)
H1150 H. gracilifolium (H)
H1358 H. patagonicum ssp. mustersii (H)
H6479 H. rusticum (H)
H1816 H. chilense (H)
H6052 H. patagonicum ssp. patagonicum (H)
H9152 H. roshevitzii (H)
America
Eurasia
H+Xa
H10115 E. caninus (StH)
H10314 E. caninus (StH)
H3169 E. caninus (StH)
H3915 E. caninus (StH)
H10096 E. caninus (StH)
H499498 H. bogdanii (H)
H499645 H. bogdanii (H)
H4014 H. bogdanii (H)
H121 H. marinum ssp. marinum (Xa)
H581 H. marinum ssp. gussoneanum (Xa)
H1342 H. patagonicum ssp. magellanicum (H)
H1334 H. patagonicum ssp. setifolium (H)
H1236 H. pubiflorum (H)
H1353 H. patagonicum ssp. santacruzense (H)
H5140A H. vulgare ssp. spontaneus (I)
H7514A H. vulgare ssp. vulgare (I)
H3878 H. bulbosum (I)
I
H52 H. marinum ssp. glaucum (Xu)
H74 H. marinum ssp. glaucum (Xu)
Xu
H2009 E. caninus (StH)
H10096 E. caninus (StH)
H10404 E. caninus (StH)
H8753 E. caninus (StH)
W6 21759 P. stipifolia (St)
PI 537377 P. spicata (St)
PI 636611 P. spicata (St)
PI 628742 P. spicata (St)
PI 531553 Aust. retrofractum (W)
PI 533014 Aust. retrofractum (W)
PI 547363 Aust. retrofractum (W)
PI 598818 P. spicata (St)
PI 595194 P. spicata (St)
PI 628744 P. spicata (St)
H10359 E. caninus (StH)
PI 628743 P. spicata (St)
PI 383534 Ag. cristatum (P)
PI 142012 L. elongatum (E)
PI 531712 L. bessarabicum (E)
St2+W+P
PI 229595 B. sterilis
5 changes

Two distinct copies of sequences were obtained for 9 out of 11 tetraploid *E. caninus* accessions of **StStHH** genomes that were amplified and sequenced. Phylogenetic analyses well separated the two copies of sequences from each accession into two different clades, one in the **H** genome clade, another in the **St** genome clade (either $\mathbf{St_1}$ or $\mathbf{St_2}$). However, three copies of sequences each were found for the accession H10404 and H2009. Phylogenetic analysis placed the three copies of sequences from each accession into $\mathbf{St_1}$, $\mathbf{St_2}$ and **H** clades, respectively. The **H** genome from 11 *E. caninus* accessions were clearly separated into two groups, one containing five *E. caninus* accessions (H10115, H10314, H3169, H3915 and H10096) (BS = 81% and PP = 1.00), which was sister to the Eurasian *H. bogdanii*; another containing 6 accessions of *E. caninus*, which were grouped together with several American **H** genome species. The **St** copy sequences from eleven *E. caninus* accessions were separated into two clades, 8 sequences in $\mathbf{St_1}$ and 5 sequences in $\mathbf{St_2}$ clades.

3.3. Phylogenetic analyses of the *PepC* sequences

To confirm whether accession H10404 and H2009 contain three copies of sequences in other genes, these two accessions, the accessions H10096, H8753 and H10359, and another accession of *E. caninus* H4111 (which was not amplified by the *RPB2* gene), were analyzed using primer for the *PepC* gene. However, regardless of PCR condition optimization, the *PepC* gene was not amplified for DNA from the accession H10096. Two copies of sequences were detected for each accession of *E. caninus* amplified.

The maximum parsimony analysis of 49 sequences resulted in 921 most parsimonious trees (374 parsimony-informative characters, CI excluding uninformative characters = 0.760; RI = 0.895). The Bayesian analyses using GTR model resulted in identical trees with mean log-likelihood values -6468.31 and -6973.56 (data not shown). The tree topologies generated by ML, MP and Bayesian analyses were similar to each other, but only one of the most parsimonious trees with BS and PP values is shown in Figure 11.

Two copies of sequences from each accession of tetraploid *E. caninus* were separated into two different clades, one into the **St** genome clade and the other into the **H** genome clade. All five **St** genome copies formed a well supported subclade with one accession PI 598818 from *P. spicata* (BS = 100, PP =1.00). All **H** copy sequences from *E. caninus* were grouped with the sequence from *H. bogdanii* in 98% BS and PP = 0.76.

3.4. Phylogenetic analyses of the *TrnD/T* sequences

The *TrnD/T* data matrix of 59 sequences contained 973 characters, of which 133 were parsimoniously informative. Maximum Parsimony analysis produced 542 equally parsimonious trees (CI excluding uninformative characters = 0.777; RI = 0.754). The separated Bayesian analyses using GTR model resulted in identical trees with mean log-likelihood values -4678.54 and -4707.78 (data not shown). Since the tree topologies of MP, ML and Bayesian analysis were similar to each other, only one of the most parsimonious trees with BS and PP values is shown (Fig. 12).

Fig. 11. One of the 921 parsimonious trees derived from *PepC* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.760, retention index (RI) = 0.895.

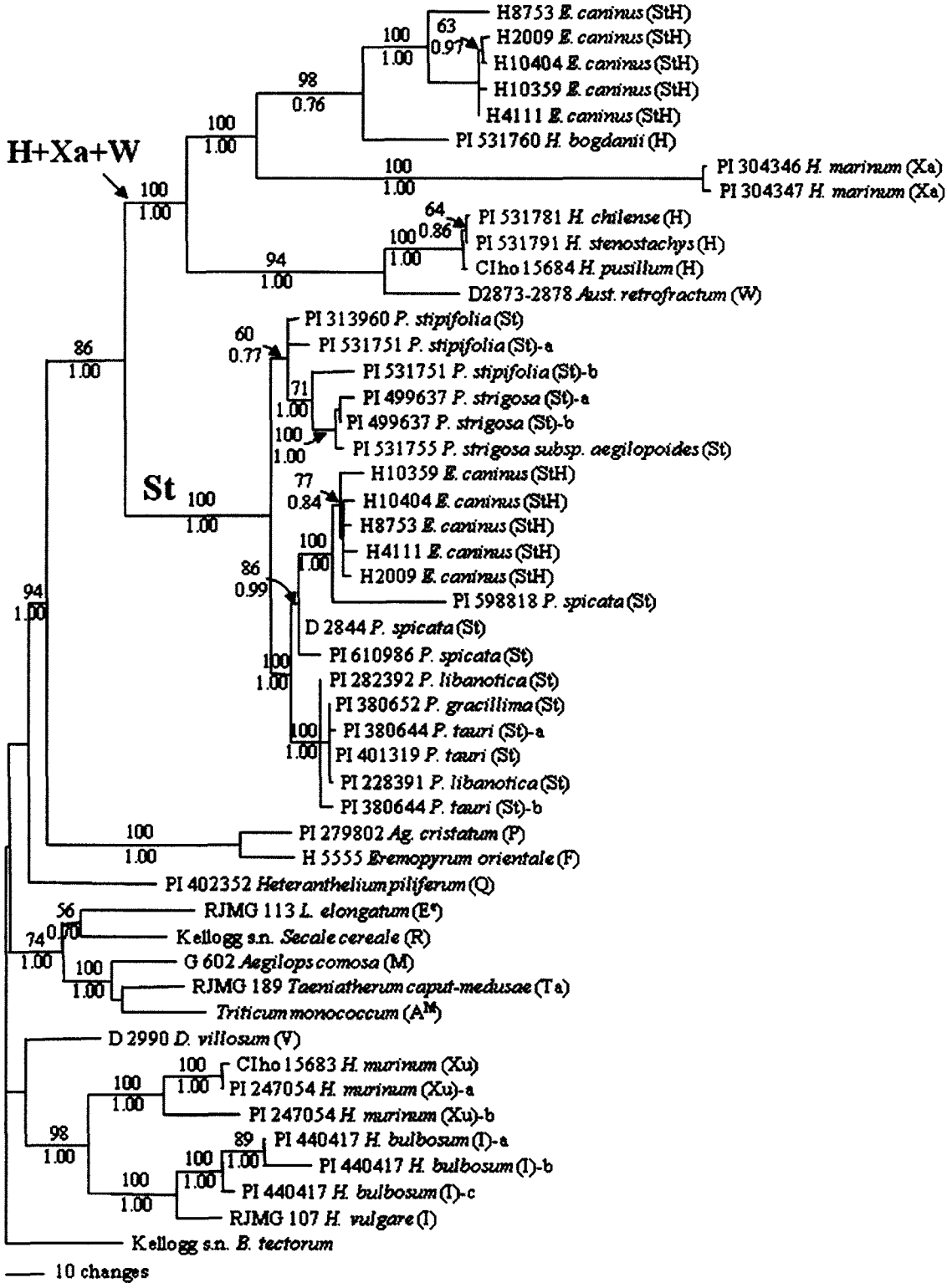
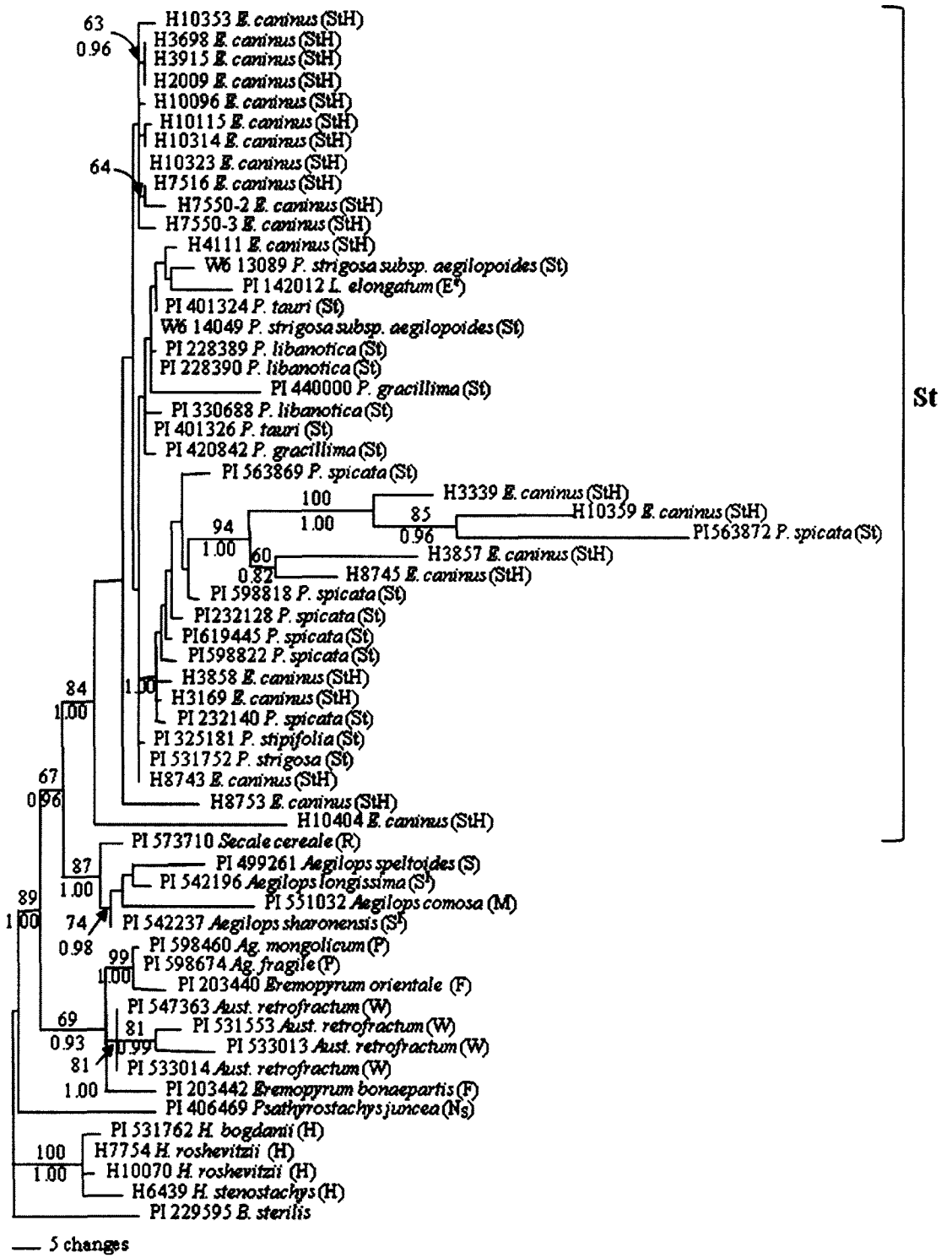

H+Xa+W
St
H8753 E. caninus (StH)
H2009 E. caninus (StH)
H10404 E. caninus (StH)
H10359 E. caninus (StH)
H4111 E. caninus (StH)
PI 531760 H. bogdanii (H)
PI 304346 H. marinum (Xa)
PI 304347 H. marinum (Xa)
PI 531781 H. chilense (H)
PI 531791 H. stenostachys (H)
CIho 15684 H. pusillum (H)
D2873-2878 Aust. retrofractum (W)
PI 313960 P. stipifolia (St)
PI 531751 P. stipifolia (St)-a
PI 531751 P. stipifolia (St)-b
PI 499637 P. strigosa (St)-a
PI 499637 P. strigosa (St)-b
PI 531755 P. strigosa subsp. aegilopoides (St)
H10359 E. caninus (StH)
H10404 E. caninus (StH)
H8753 E. caninus (StH)
H4111 E. caninus (StH)
H2009 E. caninus (StH)
PI 598818 P. spicata (St)
D 2844 P. spicata (St)
PI 610986 P. spicata (St)
PI 282392 P. libanotica (St)
PI 380652 P. gracillima (St)
PI 380644 P. tauri (St)-a
PI 401319 P. tauri (St)
PI 228391 P. libanotica (St)
PI 380644 P. tauri (St)-b
PI 279802 Ag. cristatum (P)
H 5555 Eremopyrum orientale (F)
PI 402352 Heteranthelium piliferum (Q)
RJMG 113 L. elongatum (E^e)
Kellogg s.n. Secale cereale (R)
G 602 Aegilops comosa (M)
RJMG 189 Taeniatherum caput-medusae (Ta)
Triticum monococcum (A^M)
D 2990 D. villosum (V)
CIho 15683 H. murinum (Xu)
PI 247054 H. murinum (Xu)-a
PI 247054 H. murinum (Xu)-b
PI 440417 H. bulbosum (I)-a
PI 440417 H. bulbosum (I)-b
PI 440417 H. bulbosum (I)-c
RJMG 107 H. vulgare (I)
Kellogg s.n. B. tectorum
10 changes

Fig. 12. One of the 542 parsimonious trees derived from *TrnD/T* sequence data was conducted using heuristic search with TBR branch swapping. Numbers above and below branches are bootstrap values and Bayesian posterior probability (PP) values, respectively. *Bromus sterilis* was used as an outgroup. Consistency index (CI) = 0.777, retention index (RI) = 0.754.


H10353 E. caninus (StH)
H3698 E. caninus (StH)
H3915 E. caninus (StH)
H2009 E. caninus (StH)
H10096 E. caninus (StH)
H10115 E. caninus (StH)
H10314 E. caninus (StH)
H10323 E. caninus (StH)
H7516 E. caninus (StH)
H7550-2 E. caninus (StH)
H7550-3 E. caninus (StH)
H4111 E. caninus (StH)
W6 13089 P. strigosa subsp. aegilopoides (St)
PI 142012 L. elongatum (E^e)
PI 401324 P. tauri (St)
W6 14049 P. strigosa subsp. aegilopoides (St)
PI 228389 P. libanotica (St)
PI 228390 P. libanotica (St)
PI 440000 P. gracillima (St)
PI 330688 P. libanotica (St)
PI 401326 P. tauri (St)
PI 420842 P. gracillima (St)
PI 563869 P. spicata (St)
H3339 E. caninus (StH)
H10359 E. caninus (StH)
PI563872 P. spicata (St)
H3857 E. caninus (StH)
H8745 E. caninus (StH)
PI 598818 P. spicata (St)
PI232128 P. spicata (St)
PI619445 P. spicata (St)
PI598822 P. spicata (St)
H3858 E. caninus (StH)
H3169 E. caninus (StH)
PI 232140 P. spicata (St)
PI 325181 P. stipifolia (St)
PI 531752 P. strigosa (St)
H8743 E. caninus (StH)
H8753 E. caninus (StH)
H10404 E. caninus (StH)
PI 573710 Secale cereale (R)
PI 499261 Aegilops speltoides (S)
PI 542196 Aegilops longissima (S^l)
PI 531032 Aegilops comosa (M)
PI 542237 Aegilops sharonensis (S^l)
PI 598460 Ag. mongolicum (P)
PI 598674 Ag. fragile (P)
PI 203440 Eremopyrum orientale (F)
PI 547363 Aust. retrofractum (W)
PI 531553 Aust. retrofractum (W)
PI 533013 Aust. retrofractum (W)
PI 533014 Aust. retrofractum (W)
PI 203442 Eremopyrum bonaepartis (F)
PI 406469 Psathyrostachys juncea (N_s)
PI 531762 H. bogdanii (H)
H7754 H. roshevitzii (H)
H10070 H. roshevitzii (H)
H6439 H. stenostachys (H)
PI 229595 E. sterilis
5 changes
St

Phylogenetic analyses based on *TrnD/T* sequence data grouped all sequences from *E. caninus* into the **St** genome clade (BS = 84, PP = 1.00). Within this clade, four accessions of *E. caninus* (H3339, H10359, H3857 and H8745) formed a subclade with the accession PI 563872 of *P. spicata* in 94% BS and PP = 1.00. The accession H4111 of *E. caninus* grouped with one accession of *P. strigosa* subsp. *aegilopoides* (W6 13089).

4. Discussion

4.1. On the origin of *E. caninus*

Our results indicated that *E. caninus* has experienced a very complex evolutionary history which has involved multiple hybridizations and polyploidization. It has been proposed that *Pseudoroegneria* (**St**) and *Hordeum* (**H**) species were genome donors of the allopolyploid *E. caninus* (Dewey, 1968). The present study clearly confirmed the **StH** genome constitution of *E. caninus*. Furthermore, the cpDNA sequence data indicated that the maternal origin of *E. caninus* is clearly the **St** genome species, which is consistent with the previous suggestion that *Pseudoroegneria* (**St**) is the maternal parent of some other *Elymus* species (Redinbaugh *et al.*, 2000; Mason-Gamer, 2001; Mason-Gamer *et al.*, 2002; McMillan and Sun, 2004; Xu and Ban, 2004; Liu *et al.*, 2006, Ni *et al.*, 2011). Additionally, the present study further extends the understanding of its particular origins within both the **St** and **H** genomes.

Of the 11 accessions of *E. caninus* from distinct geographic origins analyzed, 9 showed a genome constitution of either **St_1St_1HH** (H10115, H3915, H3169, H10314, H3857 and H7550-3) or **St_2St_2HH** (H10096, H8753 and H10359); however, two accessions H10404 and H2009 had both the **St_1** and **St_2** copies, and one **H** copy sequence. To verify whether these two accessions contain three copies of genomes, these accessions were analyzed using *PepC* gene. The results showed that each of these two accessions contains two copies of *PepC* sequences, confirming that these two accessions of *E. caninus* are tetraploid with **StH** genome constitution. The reason why three copies of *RPB2* sequences were found in these two accessions will be discussed later. The results confirmed that the **St** genome in *E. caninus* has two distinct origins in either **St_1** and/or **St_2**, and that *P. spicata* and *P. stipifolia* are the most likely donors of the **St_2** genome copies.

In a study of allohexaploid *Elymus repens* by Mahelka and Kopecký (2010), the *GBSSI* gene data indicated that besides the "original" *Pseudoroegneria* clade (**St**) and *Hordeum* **(H)** clade species, another "unknown" *Pseudoroegneria*-like clade species may represent the third genome lineage within hexaploid *E. repens*. Since only a few accessions of *Pseudoroegneria* species were included in the study, this result is still not conclusive (Mahelka and Kopecký, 2010). In the present study, phylogenetic analyses based on *RPB2* gene data clearly divided the **St** genome sequences into two distinct sub-genomes, **St_1** and **St_2** (Fig. 10). Eight accessions of *Psuedoroegneria* species formed a **St_2** genome group, including 7 accessions of *P. spicata* (PI 537377, PI 595194, PI 598818, PI 628742, PI 628743, PI 628744 and PI 636611), together with one accession of *P.*

stipifolia (W6 21759). Remaining accessions of *Pseudoroegneria* species formed a $\mathbf{St_1}$ group. The data clearly indicated that the "original" *Pseudoroegneria* and the "unknown" *Pseudoroegneria*-like genomes that Mahelka and Kopecký (2010) referred to were actually the $\mathbf{St_1}$ and $\mathbf{St_2}$ genome defined in the present study (Fig. 10).

Furthermore, this study clearly indicated multiple origins of the **H** genome in *E. caninus*. As the *RPB2* phylogenetic tree indicated, all **H** sequences from *E. caninus* fell into two different subclades (Fig. 10). One of these two subclades constituted of five **H** copies from *E. caninus* (H10115, H10314, H3169, H3915 and H10096), together with three Eurasian *H. bogdanii* accessions (BS = 56%, PP = 0.95), which indicated that these five accessions of *E. caninus* may have their **H**-donor from Eurasian native *H. bogdanii.* Another **H** subclade included the sequences from six accessions of *E. caninus* (H3857, H7550-3, H10359, H2009, H10404 and H8753) and several American **H**-genome species (BS = 59%, PP = 0.98), suggesting that the **H** genome in these accessions of *E. caninus* might have been donated by American **H** genome *Hordeum* species. *PepC* sequence data also supported Eurasian **H** genome species *H. bogdanii* as the **H** genome donor to the five accessions of *E. caninus* (H8753, H2009, H10404, H10359 and H4111). Both data clearly indicated a close phylogenetic relationship between *H. bogdanii* and **H** sequences in some accessions of *E. caninus*, thus, it is more likely that *H. bogdanii* is one of the major donors of the **H** copy in *E. caninus*.

4.2. Segmental gene duplication and introgression shaped the reticulate origin of StStHH *E. caninus*

During the evolutionary process, plants might have increased the DNA content of their nuclear genomes by polyploidization, segmental duplication and transposon amplification (Li and Graur, 1991).

Our *RPB2* data revealed that accession H10404 and H2009 of *E. caninus* possess both $\mathbf{St_1}$ and $\mathbf{St_2}$ copies, and an **H** genome copy sequences. Since *RPB2* is a single-copy nuclear gene, and if it is true that *E. caninus* (2n= 4x = 28) only contains two different genomes (**St** and **H**) as pointed out by Dewey (1968) and confirmed by our *Pepc* sequence data, then there should be mechanisms other than allopolyploidization, which are responsible for *RPB2* gene duplication and the mutation process from $\mathbf{St_1}$ (or $\mathbf{St_2}$) – like sequence to $\mathbf{St_2}$ (or $\mathbf{St_1}$) – like sequence. Since segmental duplication and transposon amplification could only insert portion of the original gene (Bennetzen, 2005), the full-length *RPB2* gene duplication in one chromosome should involve multiple instances of segmental duplication and gene capture process, which were possibly induced by transposable elements (TEs). One reasonable mechanism is that transposon amplification completed the first segmental gene duplication process, and then, due to further TEs-induced unequal crossover or gene transfer process, the duplicated partial gene copy gradually captured the full length gene sequence with different **St**-copy ($\mathbf{St_1}$ or $\mathbf{St_2}$) sequence patterns compared to the intact gene. Previous studies of the molecular evolution and genome divergence at the *RPB2* gene of the **St** and **H** genomes in *Elymus* species provided direct evidence that there are three TEs-like elements in the *RPB2* gene,

including a MITE stowaway element and two other elements with part of the sequence matching with CACTA transposable elements, and part of the sequence matching with *copia*, or *gypsy* (Sun *et al.*, 2007). Therefore, although the initial step of *RPB2* gene duplication in the **St** genome is unclear, transposable elements may have been involved in inducing rearrangements of gene flanking regions (Habu *et al.*, 1997). Actually, tandem duplication cooperating with TEs have also been found in other studies. For instance, in maize the formation of inverted duplicates of the S component (for seed pigmentation) of the *R-r* complex was associated with a *dopia* transposable element (Walker *et al.*, 1995). More recently, Ning *et al.* (2004) found more than 3000 Mutator-like DNA elements in rice genome sequences and named them as 'Pack-MULEs'. The small genomic fragments (averaging 325 bp, with a range of 47~986 bp) had been acquired at the DNA level, with both exons and introns present. Gene fragments from multiple genomic loci were found together in 20% of identified Pack-MULEs, and at least 5% of Pack-MULEs were found to be expressed. Hence, these results demonstrated the potential of transposable elements to create novel genes through the duplication, rearrangement and fusion of different genomic sequences. In maize, another newly discovered transposon *Helitons* was also reported to produce new nonautonomous elements, which are responsible for the duplicative insertion of multiple partial copies of gene segments into new locations (pseudogenes) (Morgante *et al.*, 2005).

Moreover, similar phenomena of nuclear gene duplication without chromosome doubling have also been found by previous studies on other polyploid wheatgrass in the tribe Triticeae. For example, Mason-Gamer *et al.* (2009) used three nuclear genes to

investigate the phylogeny of *Elymus* **StStHH** allotetraploids. All three gene data confirmed their tetraploid **StStHH** genome constitution, however, the *GBSSI* gene data also indicated that some accessions of *E. lanceolatus* and *E. wawawaiensis* each have two different gene copies both from the **St** genome, including one *P. spicata* – like sequence and another *P. strigosa* – like sequence. Another similar case was found in a more recent study on the genome constitution and evolution of allohexaploid *Thinopyrum intermedium* (Triticeae: Poaceae) (Mahelka *et al.*, 2011). Although two genomes from *Pseudoroegneria* and *Dasypyrum* in the hexaploid have been determined, the identity of the third genome remained unsatisfactorily resolved due to the presence of five lineages, which was not consistent with hexaploidy, in *Th. intermedium*. As stated by Mason-Gamer *et al.* (2009) and Mahelka *et al.* (2011), these unexpected dual or multiple placements of several individuals in different genome groups may reflect the maintenance of ancestral polymorphism as a result of hybridization and/or introgression. Furthermore, these hybridization and introgression processes may be caused by incomplete lineage sorting or segmental gene duplication. However, since extensive sampling of Triticeae diploids were used in the present study, and no diploid *Pseudoroegneria* which simultaneously possess both $\mathbf{St_1}$ and $\mathbf{St_2}$ copies of *RPB2* gene sequences were found, segmental gene duplication seems more likely to be the cause here.

Theoretically, these segmental gene duplication and gene capture processes could occur either before or after the allopolyploidization. In other words, there are three possible hypothetical scenarios for the accessions H10404 and H2009 of *E. caninus*

origin. One scenario is that these two accessions evolved through fusion of reduced gametes of a diploid *Pseudoroegneria* (**St**), containing both $\mathbf{St_1}$ and $\mathbf{St_2}$ copies of the *RPB2* gene, with other reduced gametes of a diploid *Hordeum* (**H**), followed by chromosome duplication. In a second scenario, these two accessions might have evolved from a tetraploid **StH** *E. caninus* (either $\mathbf{St_1St_1HH}$ or $\mathbf{St_2St_2HH}$), through additional sequence exchanges by segmental gene duplication and gene capture process from another species containing the different **St** copy ($\mathbf{St_1St_1HH}$ from $\mathbf{St_2}$-containing species or $\mathbf{St_2St_2HH}$ from $\mathbf{St_1}$-containing species). However, as stated above, since extensive diploid **St** genome species were sampled in the present study and none of them were shown to possess both the two different **St** copies ($\mathbf{St_1}$ and $\mathbf{St_2}$) of the *RPB2* gene, it is inferred that the second scenario may be more reasonable in this case. One argument may be that some species in *Pseudoroegneria* are autotetraploid, the accession H10404 and H2009 of *E. caninus* might originate from an autotetraploid *Pseudoroegneria* (possessing both $\mathbf{St_1}$ and $\mathbf{St_2}$ copies) and a diploid *Hordeum* species. If this is the case, we should expect to find three copies *PepC* sequences in these two accessions. However, only two distinct copies of *PepC* sequences were found in these two accessions, which did not support this possibility.

4.3. Relationship among *E. caninus* accessions from different geographic regions

With respect to the geographical origin of the *E. caninus* accessions considered in the present study, the accessions investigated were generally collected from two distant regions: Europe and Asia (Table 3). Previous genetic diversity study by Sun *et al.* (1999) on *E. caninus* using isozyme, RAPD and microsatellite markers indicated no clear correlation between the clusters and geographic regions. Geographically distant accessions can be remarkably similar to each other, while neighboring accessions can differ greatly. However, one particular geographic group, the Icelandic accessions, displayed a single, compact and distinct group from the other accessions in the first dimension of PCA analysis. It suggested that geographic isolation strongly influenced the evolution of the Icelandic population of *E. caninus*. This suggestion matched well with earlier study on endosperm protein patterns in *E. caninus* populations by Kostina *et al.* (1998); their data indicated that Icelandic material possessed a deviating protein composition from all other *E. caninus* populations studied. Moreover, another genetic diversity study on the Nordic *Elymus* species by Díaz (1999) pointed out that the Nordic accessions of *E. caninus* are more variable than the other accessions based on allozymic and morphological results. Further studies on the molecular genetic similarity between two morphological similar species, *E. trachycaulus* and the Eurasian *E. caninus*, revealed a close genetic similarity between two *E. caninus* accessions H8745 and H10096, which both originate from the Altai Mountain region (Sun *et al.*, 2006).

In the present study, the Icelandic, Nordic and Altaic accessions, as well as other Eurasian *E. caninus* were sampled and analyzed, however, both the two nuclear gene data and the cpDNA data could not provide clear links between the sequence placement of *E. caninus* in the phylogenetic trees and their species' geographic origins. Moreover, based on our *TrnD/T* gene data, no strong link was found between the two Altai Mountain accessions of *E. caninus* H8745 and H10096. Nevertheless, a close relationship among three Asian accessions (H8745, H3339 and H3857) was revealed by the cpDNA *TrnD/T* data, since they formed a well supported subclade with the Icelandic accession H10359, along with one accession of *P. spicata* PI 563872 (BS = 94, PP = 1.00, see Fig. 12). One possible explanation for the weak link of both the **St** and **H** genome copies between sequence variation and their geographic origins is that *E. caninus* inhabits diverse environments. Thus, geographically close habitats can be ecologically quite distinct, and conversely, habitats that are geographically distant from one another can be very similar in their environmental conditions. Therefore, gene sequence variation, if under natural selection, could be expected to display low correlation between accessions separated by geographical distance.

4.4. Summary

In order to explore genome evolutionary dynamics and origin of tetraploid *Elymus caninus*, twenty-one accessions of *E. caninus* from wide geographic distribution regions were analyzed using two single-copy nuclear genes, the second largest subunit of RNA polymerase II (*RPB2*) and the phosphoenolpyruvate carboxylase (*PepC*), along with one

non-coding chloroplast DNA *TrnD/T* region. The *RPB2* data revealed that accessions H10404 and H2009 of *E. caninus* possessed three distinct copies of sequences, $\mathbf{St_1}$- like, $\mathbf{St_2}$- like, and **H** genome sequences, while *PepC* data showed that these two accessions each contains two copies of *PepC* sequences, **St** and **H** genome sequences. This suggested that the *RPB2* gene might have been duplicated without chromosome doubling during its evolutionary history. The segmental duplication and gene capture process was possibly induced by transposable elements. The data clearly confirmed the **StH** genome constitution of *E. caninus* and indicated that the **St** genome in *E. caninus* has two distinct origins in either $\mathbf{St_1}$ and/or $\mathbf{St_2}$, and that *Pseudoroegneria spicata* and *P. stipifolia* are the most likely donors of the $\mathbf{St_2}$ genome copies. Moreover, our study clearly indicated multiple origins of the **H** genome in *E. caninus,* and *Hordeum bogdanii* is one of the major donors of the **H** copy in *E. caninus*. The cpDNA sequence data indicated that the maternal origin of *E. caninus* is clearly the **St** genome species. The sequences in *E. caninus* accessions are not differentiated in accordance with their geographical origins.

Part 3. References

Belyayev A, Raskina O, Korol A, Nevo E. 2000. Coevolution of A and B genomes in allotetraploid *Triticum dicoccoides*. *Genome* 43: 1021–1026.

Bennetzen JL. 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics & Development* 15: 621–627.

Bothmer R. von, Salomon B. 1994. Triticeae: a tribe for food, feed and fun. In: Wang RRC, Jensen KB, Jaussi C, eds. *Proceedings of the 2nd International Triticeae Symposium*, Utah, USA: Logan Press, 1–12.

Chen ZJ, Pikaard CS. 1997. Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in *Brassica*. *Proceedings of the National Academy of Sciences, USA* 94: 3442–3447.

Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

Comai L, Tyaqi AP, Winter K, Holmes-Davis R, Reynolds SH. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12: 1551–1567.

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, dePamphilis CW. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Research* 16: 738–749.

Daubenmire RF. 1942. An ecological study of the vegetation of southeastern Washington and adjacent Idaho. *Ecological Monographs* 12: 53–79.

Dewey DR. 1968. Synthetic *Agropyron-Elymus* hybrids: III. *Elymus canadensis* x *Agropyron caninum*, *A. trachycaulum*, and *A. striatum*. *American Journal of Botany* 55: 1133–1139.

Dewey DR. 1971. Synthetic hybrids of *Hordeum bogdanii* with *Elymus canadensis* and *Sitanion hystrix*. *American Journal of Botany* 58: 902–908.

Dewey DR. 1984. The genomic system of classification. A guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP, ed. *Gene manipulation in Plant Improvement*. New York, NY, USA: Plenum Press, 209–280.

Díaz O, Salomon B, von Bothmer R, 1998. Description of isozyme polymorphisms in *Elymus* species using starch gel electrophoresis, In: Jaradat AA, ed. *Triticeae III.* New Hampshire, USA: Science Publishers, 199–208.

Díaz O. 1999. Genetic diversity in *Elymus* species (Triticeae) with emphasis on the Nordic region. PhD Thesis. Svalöv, Sweden: Swedish University of Agricultural Sciences.

Dubcovsky J, Schlatter AR, Echaide M. 1997. Genome analysis of South American *Elymus* (Triticeae) and *Leymus* (Triticeae) species based on variation in repeated nucleotide sequences. *Genome* 40: 505–520.

Feldman M, Levy AA. 2009. Genome evolution in allopolyploid wheat – a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics* 36: 511–518.

Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.

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

Felsenstein J. 1985b. Phylogenies and the comparative method. *The American Naturalist* 125: 1–15.

Felsenstein J. 1993. *PHYLIP (Phylogeny Inference Package) version 3.6a2.* Distributed by the author. Seattle, WA, USA: Department of Genetics, University of Washington.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 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: 3403–3417.

Garamszegi LZ, Møller AP. 2010. Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. *Biological Reviews* 85: 797–805.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.

Habu Y, Peyachoknagul S, Sakata Y, Fukasawa K, Ohno T. 1997. Evolution of a multigene family that encodes the Kunitz chymotrypsin inhibitor in winged bean: A possible intermediate in the generation of a new gene with a distinct pattern of expression. *Molecular and General Genetics* 254: 73–80.

Han FP, Fedak G, Quellet T, Liu B. 2003. Rapid genomic exchanges in interspecific and intergeneric hybrids and allopolyploids of Triticeae. *Genome* 46: 716–723.

Han FP, Fedak G, Guo WL, Liu B. 2005. Rapid and repeatable elimination of a parental genome-specific DNA repeat (pGc1R-1a) in newly synthesized wheat allopolyploids. *Genetics* 170: 1239–1245.

Harmon LJ, Losos JB. 2005. The effect of intraspecific sample size on type I and type II error rates in comparative studies. *Evolution* 59: 2705–2710.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the Human-Ape splitting by a molecular clock of mitochondrial-DNA. *Journal of Molecular Evolution* 22: 160–174.

Heath TA, Hedtke SM, Hillis DM. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hedtke SM, Townsend TM, Hillis DM. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* 55: 522–529.

Helfgott DM, Mason-Gamer RJ. 2004. The evolution of North American *Elymus* (Triticeae, Poaceae) allotetraploids: evidence from phosphoenolpyruvate carboxylase gene sequences. *Systematic Botany* 29: 850–861.

Hillis DM. 1996. Inferring complex phylogenies. *Nature* 383: 130–131.

Hillis DM. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47: 3–8.

Hsiao C, Chatterton NJ, Asay KH, Jensen KB. 1995. Phylogenetic relationships of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome* 38: 221–223.

Ilic K, SanMiguel PJ, Bennetzen JL. 2003. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proceedings of the National Academy of Sciences, USA* 100: 12265–12270.

Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M. 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Molecular Biology and Evolution* 23: 1217–1231.

Jensen KB. 1990. Cytology, fertility, and morphology of *Elymus kengii* (Keng) Tzvelev and *E. grandiglumis* (Keng) Á. Löve (Triticeae: Poaceae). *Genome* 33: 563–570.

Jensen KB, Salomon B. 1995. Cytogenetics and morphology of *Elymus panormitanus var. heterophyllus* (Keng) Á. Löve and its relationship to *Elymus panormitanus* (Poaceae: Triticeae). *International Journal of Plant Sciences* 156: 731–739.

Jukes T, Cantor C. 1969. Evolution of protein molecules. In: Munro H, ed. *Mammalian Protein Metabolism*, New York, NY, USA: Academic Press. vol. III, 24: 21–132.

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

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.

Komatsuda T, Tanno K, Salomon B, Bryngelsson T, von Bothmer R. 1999. Phylogeny in the genus *Hordeum* based on nucleotide sequences closely linked to the *vrs1* locus (row number of spikelets). *Genome* 42: 973–981.

Kostina EV, Agafonov AV, Salomon B. 1998. Electrophoretic properties and variability of endosperm proteins in *Elymus caninus* (L.). In: Jaradat AA, ed. *Triticeae III.* New Hampshire, USA: Science Publishers, 265–272.

Kovařík SJ, Goo HJ, Kim NS. 2003. Sequence variation of ITS1, 5.8S rDNA, and ITS2 regions in 45S rRNA genes and species relationship among the *Hordeum* species. *Korean Journal of Genetics* 25: 83–93.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* 20: 86–93.

Larson SR, Jones TA, Hu ZM, McCracken CL, Palazzo A. 2000. Genetic diversity of bluebunch wheatgrass cultivars and a multiple-origin polycross. *Crop science* 40: 1142–1147.

Larson SR, Jones TA, Jensen KB. 2004. Population structure in *Pseudoroegneria spicata* (Poaceae: Triticeae) modeled by Bayesian clustering of AFLP genotypes. *American Journal of Botany* 91: 1789–1801.

Lee HS, Chen ZJ. 2001. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. *Proceedings of the National Academy of Sciences, USA* 98: 6753–6758.

Li DY, Zhang XY. 2002. Physical localization of the 18S-5.8S-26S rDNA and sequence analysis of ITS regions in *Thinopyrum ponticum* (Poaceae: Triticeae): implications for concerted evolution. *Annals of Botany* 90: 445–452.

Li W, Graur D. 1991. Fundamentals of Molecular Evolution. Sunderland, MA: Sinauer Associates.

Liu Q, Ge S, Tang H, Zhang X, Zhu G, Lu BR. 2006. Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *New Phytologist* 170: 411–420.

Löve Á. 1984. Conspectus of the Triticeae. *Feddes Repertorium* 95: 425–521.

Lu BR, Liu Q. 2005. The possible origin of the "StY" genome *Elymus*: a new mechanism of allopolyploidy in plants. *Czech Journal of Genetics and Plant Breeding* 41: 58.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151–1155.

Magnus P, Bengt O. 2007. Origin and evolution of North American polyploid *Silene* (Carryophyllaceae). *American Journal of Botany* 94: 330–349.

Mahelka V, Kopecky D. 2010. Gene capture from across the grass family in the allohexaploid *Elymus repens* (L.) Gould (Poaceae, Triceae) as evidenced by ITS, *GBSSI* and molecular cytogenetics. *Molecular Biology and Evolution* 27: 1370–1390.

Mason-Gamer RJ. 2001. Origin of North America *Elymus* (Poaceae: Triticeae) allotetraploids based on granule-bound starch synthase gene sequences. *Systematic Botany* 26: 757–758.

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.

Mason-Gamer RJ, Burns MM, Naum M. 2005. Polyploidy, introgression, and complex phylogenetic patterns within *Elymus*. *Czech Journal of Genetics and Plant Breeding* 41: 21–26.

Mason-Gamer RJ, Burns MM, Naum M. 2010. Phylogenetic relationships and reticulation among Asian *Elymus* (Poaceae) allotetraploids: Analyses of three nuclear gene trees. *Molecular Phylogenetics and Evolution* 54: 10–22.

Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.

McCormack JE, Huang H, Knowles LL. 2009. Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Systematic Biology* 58: 501–508.

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. *Theoretical and Applied Genetics* 108: 535–542.

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A. 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nature Genetics* 37: 997–1002.

Ni Y, Asamoah-Odei, Sun GL. 2011. Maternal origin, genome constitution and evolutionary relationships of polyploid *Elymus* species and *Hordelymus europaeus*. *Biologia Plantarum* 55: 68–74.

Ning J, Bao ZR, Zhang XY, Eddy SR, Wessler SR. 2004. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431: 569–573.

Okito P, Mott IW, Wu Y, Wang RR. 2009. A Y genome specific STS marker in *Pseudoroegneria* and *Elymus* species (Triticeae: Gramineae). *Genome* 52: 391–400.

Ozkan H, Levy AA, Feldman M. 2001. Allopolyploidyinduced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. *Plant Cell* 13: 1735–1747.

Ozkan H, Levy AA, Feldman M. 2002. Rapid differentiation of homoeologous chromosomes in newly formed allopolyploid wheat. *Israel Journal of Plant Sciences* 50: 65–76.

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

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2005. DNA sequence polymorphism, version 4.10.4. DNAsp4 computer software. Barcelona, Spain: Barcelona University.

Ronquist F, Huelsenbeck JP. 2005. Bayesian analysis of molecular evolution using MrBayes. In: Nielsen R, ed. *Statistical methods in molecular evolution.* Springer-Verlag Press, 183–232.

Salina EA, Numerova OM, Ozkan H, Feldman M. 2004. Alterations in subtelomeric tandem repeats during early stages of allopolyploidy in wheat. *Genome* 47: 860–867.

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: 1749–1759.

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

Sun GL, Díaz O, Salomon B, von Bothmer R. 2001. Genetic diversity and structure in a natural *Elymus caninus* population from Denmark based on microsatellite and isozyme analysis. *Plant Systematics and Evolution* 227: 235–244.

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

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

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 Molecular Biology* 64: 645–655.

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. *Molecular Phylogenetics and Evolution* 46: 897–907.

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. *Molecular Phylogenetics and Evolution* 56: 727–733.

Swofford DL. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA, USA: Sinauer Associates.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogeny reconstruction. In: Hillis DM, Moritz C, Mable BK, eds. *Molecular* systematics, 2nd edition. Massachusetts, MA, USA: Sinauer Associates, 407–414.

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

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* 10: 512–526.

Thompson JD, Gibson TJ, Plewniak F, Jeanmouguin F, Higgins DG. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

Torabinejad J, Mueller RJ. 1993. Genome Constitution of the Australian hexaploid grass *Elymus scabrus* (Poaceae: Triticeae). *Genome* 36: 147–151.

Walker EL, Robbins TP, Bureau TE, Kermicle J, Dellaporta SL. 1995. Transposon-mediated chromosome rearrangements and gene duplications in the formation of the maize *R-r* complex. *EMBO Journal* 14: 2350–2363.

Wang RR–C, Bothmer R von, Dvorak J, Fedak G, Linde-Laursen I, Muramatsu M. 1994. Genome symbols in the Triticeae. In: Wang RRC, Jensen KB, Jaussi C, eds. *Proceeding of the 2nd International Triticeae Symposium*, Utah, USA: Logan, 29–34.

Watterson GA. 1975. On the number of segregation sites in genetical models without recombination. *Theoretical Population Biology* 7: 256–276.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Molecular Biology* 42: 225–249.

Wiens JJ, Servedio MR. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. *Systematic Biology* 46: 332–345.

Wiens JJ, Servedio MR. 1998. Phylogenetic analysis and intraspecific variation: performance of parsimony, likelihood, and distance methods. *Systematic Biology* 47: 228–253.

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. *Theoretical and Applied Genetics* 108: 1443–1448.

Yen C, Yang JL, Baum BR. 2007. Systematics of Triticeae. Vol. 4. Beijing, China: Agriculture Press.

Yu HQ, Fan X, Zhang C, Ding CB, Wang XL, Zhou YH. 2008. Phylogenetic relationships of species in *Pseudoroegneria* (Poaceae: Triticeae) and related genera inferred from nuclear rDNA ITS (internal transcribed spacer) sequences. *Biologia* 63: 498–505.

Zeng J, Fan X, Li Z, Wang XL, Zhang HQ, Kang HY, Zhou YH. 2010. Molecular phylogeny and maternal progenitor implication in the genus *Kengyilia* (Triticeae:

Poaceae): Evidence from *COXII* intron sequences. *Biochemical Systematics and Ecology* 38: 202–209.

Zhang J. 2003. Evolution by gene duplication: An update. *Trends in Ecology & Evolution* 18: 292–298.

Zhang L, Gaut BS. 2003. Does recombination shape the distribution and evolution of tandemly arrayed genes (TAGs) in the *Arabidopsis thaliana* genome? *Genome Research* 13: 2533–2540.

Zhang XQ, Salomon B, von Bothmer R, Díaz O. 2000. Patterns and levels of genetic differentiation in North American populations of the Alaskan wheatgrass complex. *Heredity* 133: 123–132.