Evolution and Origin of Polyploid Triticeae (Poaceae) Revealed through the Nuclear *pgk1* Gene

By
Shawn James Adderley

A Thesis Submitted to

Saint Mary's University, Halifax, Nova Scotia

in Partial Fulfillment of the Requirements for

The Degree of Bachelor of Science Honours.

March, 2012, Halifax, Nova Scotia

Copyright Shawn James Adderley, 2012

Approved: Dr. G. Sun

Signature: _____

Supervisor

Approved: Dr. C. Shamputa

Signature: _____

Reviser

Approved: Dr. Z. Dong

Signature: _____

Reviser

Date: April 10, 2013

Abstract

Despite many studies on the taxonomy of the Triticeae, no general consensus has yet been agreed upon as to their phylogenetics. This is attributed to extremely high rates of hybridization and polyploidization. Levels of nucleotide divergence provide key evidence in to the evolution of polyploids. The *pgk1* gene was used to investigate the potential evolutionary dynamics of the Triticeae, and phylogenetic analyses (Maximum Parsimony and Neighbor-Joining) were carried out to determine the diploid origin of polyploids within the tribe in relation to their A^u, B, D, St, Ns, P, and H haplomes. Origins of the A^u, B, and D genomes were linked to *Triticum urartu*, *Aegilops speltoides*, and *Aegilops tauschii* respectively. Putative St genome donor was *Pseudoroegneria*, while Ns and P donors were *Psathyrostachys* and *Agropyron*. H genome diploid donor is *Hordeum*. Highly negative and significant Tajima's D values for the St, A^u, and D genomes along with high rates of polymorphisms and low sequence diversity correlate with purifying selection pressures within the polyploid members of the tribe, acting to conserving homologous copies of *pgk1* from each diploid origin. Nucleotide diversity values varied, indicating different selective pressures on each species within the tribe, and fortifying their identification as separate species.

Acknowledgements

I would like to initially thank my supervisor Dr. Genlou Sun for all his help and support throughout the many ups and downs throughout the year, without his guidance this would not have been possible. As well, I would like to say a special thank you to the students in the Sun lab for their continuous patience and help. I would like to thank my professors at Saint Mary's University over the course of my bachelor's degree for their insight and knowledge that allowed me to get to where I am today. Lastly, I would like to recognize my family and girlfriend Heather for their constant support and words of encouragement each and every day.

Table of Contents

List of Tables and Figures

Tables

Figures

1. Introduction

(1.1) The Triticeae

The tribe Triticeae lies within the family Poaceae. It is estimated such tribes radiated following the divergence of monocot and dichotomous plants, placing their origin some 50-70 mya (Huang et. al, 2002). Today the Triticeae contains many of the world's most economically important forage crops such as barley (*Hordeum*), wheat (*Triticum*), and rye (*Secale*) (Barkworth, 1992; Kellogg, 1996; Huang et. al, 2002; Petersen and Seberg, 1996; Blattner, 2009; Escobar et al., 2011). In addition to these vastly important cultivated crops, there are many other wild members of the tribe, whose values should not be underplayed. Each species represents a potentially important germplasm, a source of genetic material that may contain valuable genes that can be isolated and transferred into the genome of cultivated crops (Kellogg, 1996). For example, Singh (2009) found that species belonging to the genus *Thinopyrum* and *Psathyrostachys* respectively show high resistance to the Root-Knot nematode (*M. chitwood*).

Cytological analyses have revealed that the Triticeae contains a basic chromosome number of n=7, while having diploid (2n=14), and varying levels of polyploid states (von Bothmer et al., 1986; Komatsuda et al. 1999; Huang et al., 2002; Nishikawa et al., 2002; Petersen and Seberg, 2003; Blattner, 2004; Jakob and Blattner, 2006; Petersen and Seberg, 1996; Blattner, 2009; Sun et al., 2009; Wang et al., 2011).

Despite the extensive research on the Triticeae, brought about by its economic importance, its intricate levels of hybridization within its genera compound the difficulties in achieving a consensus on the tribes phylogenetic history (von Bothmer et

al., 1986; Kellogg, 1996; Doebley et al. 1992; Komatsuda et al., 1999; Nishikawa et al., 2002; Jakob and Blattner, 2006). In looking at the number of genera within the tribe that has been proposed over the years, it is easy to picture the complexities within the Triticeae; Love (1984) recognized 38 genera, Clayton and Renvoize (1986) recognized 18 genera, while Tzvelev (1989) recognized 24 genera. Today, the accepted genera lie as an intermediate of these numbers. Stebbins (1956), a prominent figure in the science of the evolution of plants, had once even proposed that the entire tribe be placed in a single genus due to the fact that morphological, cytological, and biochemical differences within the tribe were ones that would be expected below the genus level (Kellogg, 1989). In Barkworth's (1992) paper entitled *Taxonomy of the Triticeae: a historical perspective*, she stated that "…in an ideal group ….no feature would evolve more than once; obvious morphological differences would mark species group having different ancestors: species, once differentiated would either not hybridize or would produce completely sterile hybrids; and there would be no polyploidy. The Triticeae does not follow any of these criteria."

(1.2) Polyploidization within the Tribe

When observing the diploid genera of the Triticeae, one can easily note the distinct genomes, many of which are unable to produce viable offspring if inter-bred, re-enforcing their designations as species in accordance with the biological species concept (Kellogg, 1996; Huang, 2002). It has been observed in many species, such as in the genera *Hordeum* that the success of polyploid species in generating viable seed sets, and fitness, is much higher than that of diploid species (Bothmer and Komatsuda, 2011**)**. Due to the high success of the polyploid species within the tribe, studies have shown they may prove

to serve as bridges to carry over genes between diploid species that may not have high crossability on their own (Escobar et al., 2011).

Duplication events are a vital part of genomic architecture. Genome duplication (polyploidy) provides a pool of duplicate genes as building blocks for potential evolutionary advancements (Fan et al., 2009; 2012). These polyploids are categorized by their nature of formation. Autopolyploid species result from the doubling of the genome of a single species, while allopolyploids (much more prominent within Triticeae) are formed when two or more sets of chromosomes from different species are brought into the same nucleus to undergo chromosome doubling (Stebbins, 1947; Gu et al., 2004).

In addition to the complex nature of allopolyploidy within the genus, introgression has been noted as well. Introgression can be explained as the gene flow resulting from repeated backcrossing of interspecific hybrids with one of their parent species (Anderson, 1949). Studies on allopolyploidization and introgression shed light on the complexity within the tribe. Allopolyploidization is particularly successful in nature and in the Triticeae. It has also been shown to create rapid and extensive changes in the genome of plants (Han et al., 2003). These changes have been known to include losses, or disappearances of parental DNA fragments, the appearance of novel fragments, the activation and silencing of specific genes, and the reactivation of retrotransposons (Han et al, 2003; Gu et al., 2004). These changes are believed to represent mechanisms for novel genomic and phenotypic variation within species, and may help explain the evolutionary success of the polyploids in over their diploid predecessors (Han et al., 2003; Gu et al., 2004). As if this were not enough, studies recognize that each novel gene studied within a

family, tribe, or genus, represents its own unique evolutionary history (Kellogg, 1996), adding to the list of confounds in determining a static phylogeny.

It has long been proposed that there will never be a clear consensus on the taxonomy of the Triticeae due to its highly reticulate evolutionary history, and high levels of hybridization between its genera (Barkworth, 1992). Despite the lack of phylogenetic congruence, there is much support on the origin of the polyploids within the Triticeae. As was touched upon, polyploid species get their genetic composition from two or more related species. As such, their genomic designations can accurately reflect their evolutionary history (Huang, 2002). The hexaploid bread wheat (*Triticum aestivum*) is a vastly important crop, and an excellent example of allopolyploidization in the tribe. Its genome consists of the three haplomes A, B, and D. It is believed to have originated via the hybridization of *Triticum turgidum* (AABB), and *Aegilops tauchii* (DD). While *T.turgidum* is believed to be the outcome of a cross between the diploid *Triticum urartu* (AA), and proposed BB genome donor *Aegilops speltoides* (current designation is SS) (Huang et al., 2002a; 2002b; Gu et al., 2004; Kilian et al., 2007). Within the genera *Elymus* (StStHH) and *Kengyilia* (StStYYPP), common origin for each genera's St genus is understood to be the diploid donor *Pseudoroegneria* (Sun and Komatsuda, 2009; Fan et al., 2012), while the H genome of *Elymus* is believed to derive from an unknown diploid *Hordeum* (Sun et al., 2011), and the P genome of *Kengyilia* from diploid *Agropyron* (Fan et al., 2011). The Y genome origin of each genus as of yet remains unknown (Fan et al., 2011; Sun et al., 2011). In allotetraploid *Leymus* (NsNsXmXm), its Ns genome origin has been traced to *Psathyrostachys* (Zhang and Dvorak, 1991). Despite support for the

genomic origins, as no common tree has been recognized, additional support is needed to elucidate upon the relationships in the tribe.

(1.3) Pgk1 Gene

Classical studies based on morphological and physiological traits have been advanced to analysis at the DNA sequence level with the advent of more recent genomic techniques (Huang et al., 2002). Such techniques provide a reliable information as long as it is based on sequences that are well understood in origin, copy number, and structure/function, as well as being sufficiently variable in characters (Huang et al., 2002).

The nuclear *pgk1* in plants is responsible for coding the plastid phosphoglycerate kinase (PGK) isoenzymes. These PGK's are small monomeric proteins that facilitate the formation of ATP in the glycolytic pathway by catalyzing the conversion of 1,3-diphosphoglycerate to 3-phosphogycerate (Huang et al., 2002; Fan et al., 2012). The *pgk1* gene contains ~4 introns and 5 exons and is thought to have originated via duplication of a prokaryotic gene (Huang et al., 2002). The *pgk1* gene is ideal for phylogenetic study within the Triticeae. Firstly, it is a nuclear gene, which is associated with higher character content, biparental inheritance, and higher variability at the species and genus level than chloroplast sequences (Kellogg, 1989). Second, and perhaps most importantly, it is present in single-copy within the diploid species of the Triticeae. Its single-copy presence makes it ideal for inferring orthologous relationships between the species. It ensures us that the genes we are comparing between species and genera, are the same genes, and not genetic duplicates (paralogues), from which artifactural evidence of relatedness and

dissimilarities could easily result (Huang et al., 2002; Escobar et al., 2002; Fan et al., 2012)

In the present study, we aim to re-examine the current suggestions on the origin of polyploid species within the Triticeae, and their relationships using the nuclear, single-copy *pgk1* gene. In addition, we investigate the evolution of the *pgk1* gene throughout the tribe and compare the nucleotides of the *pgk1* gene among different genomes. As previously suggested, we expect the origins of the A, B, and D genomes within the *Triticum* to be from *Triticum urartu*, *Aegilops speltoides*, and *Aegilops tauchii* respectively. Additionally, we predict the Ns genome of *Leymus* to be donated from the diploid *Psathyrostachys*, the St genome of *Elymus* and *Kenygilia* to be from *Pseudoroegneria*, and the H genome from Elymus do be from *Hordeum*. We expect that the P genome in *Kengyilia* originated from *Agropyron*. These relationships should be quite clear based on the genomic groupings from our phylogenetic trees. It should be noted that Kellogg (1989) found that hybrids do not always appear as sister taxon, or group with their parents, and often result in incongruent intersecting data, which may make the hierarchy unclear.

2. Materials and Methods

2.1 Taxon Sampling

Sixty-two Triticeae species were analyzed in this study, including 32 diploid species from 21 basic genomes within the tribe (Table 1). Three hundred and twenty two *pgk1* accessions were originally downloaded from GenBank, along with one *pgk1* sequence of *Bromus inermis* that was used as the outgroup for phylogenetic analysis.

2.2 Data Analysis

Due to the large number of acquired accessions, genera were first split into two categories for preliminary analysis. First *Triticum*/*Aegilops* was examined, and then all other 19 genera (this group will be referred to as "*Others*"). In each group, a preliminary phylogeny was made; Sequences were aligned using Clustal W. Any gaps in aligned sequences were removed during the alignment phase, and initial trees were created using the Maximum likelihood method with 1000 bootstrap replicates. This was all facilitated by the program Mega 5.10 (Tamura et al., 2011).

After phylogenies were created, the trees were carefully examined to establish which sequences were redundant within the tree. That is, any sequences of the same species that formed a monophyletic group were removed from the data set, leaving one representative of the species. This process was repeated with both the *Triticum*/*Aegilops* group, and the "*Others"* group twice to ensure the sequences that would provide the best represented tree were selected. Once all unnecessary sequences were removed, the sequences from *Triticum*/*Aegilops* and "*Others"* groups were combined into a complete data set containing 96 sequences (Figure 1).

Multiple alignments of final 96 sequences were completed using Clustal X via default settings. Manual edits were conducted following alignment to minimize gaps (Thompson et al., 1997). PAUP 4.0 was used to create phylogenetic analysis using the maximum-parsimony (MP) and Neighbor-Joining (NJ) methods (Swofford, 2003). Most-parsimonious trees were created using the heuristic search method with the Tree Bisection-Reconnection (TBR) option and MulTrees selected, and ten replications of random addition sequences with the stepwise addition option. A strict-consensus tree was created by combining 511 most parsimonious trees. Overall character congruence was calculated using the consistency index (CI=0.578), and the retention index (RI=0.876). Bootstrap values with 1000 replications were calculated to determine the robustness of formed clades (Felsenstein, 1985) using a heuristic search with the TBR option with MulTrees on and max trees set to 100. Neighbor-Joining trees were created unrooted and without correcting distances measured. Bootstrap values were calculated as per the Maximum Parsimony method with 1,000 replications.

DnaSP 5.10 (Librado and Rozas, 2009) was used to assess the divergence and genetic relationships between polyploidy Triticeae and its diploid predecessors. Nucleotide diversity was estimated using Tajima's π (Tajima, 1989) and Watterson's θ (Watterson 1975), as well as the number of polymorphic segregating sites (S) (Wakeley and Hey, 1997). Neutrality was tested following Tajima's D (Tajima, 1989), as well as Fu and Li's D (Fu and Li, 1997) statistics. For all statistics, our full data set (N=322) was used to accurately represent all sequences in the tribe.

Table 1: Triticeae species used for phylogenetic analyses, Genomic Designations, and Gen Bank Accession numbers.

Species	**Genomic Designation**	**GenBank Accession**
Aegilops bicornis	S^b	DQ290763.1
Aegilops longissima	S^l	DQ290764.1
Aegilops searsii	S^s	DQ290761.1
Aegilops sharonensis	S^l	DQ290767.1
Aegilops speltoides	B	DQ290768.1
Aegilops tauschii	D	DQ290771.1
Agropyron cristatum	P	JF965626.1 JF965633.1
Agropyron mongolicum	P	FJ711024.1
Australopyrum retrofractum	W	FJ711025.1
Bromus inermis	-Outgroup-	FJ711014.1
Crithopsis delileana	K	FJ711026.1
Dasypyrum villosum	V	FJ711027.1
Elymus canadensis	StH	FJ711008.1 FJ711009.1
Elymus caninus	StH	FJ711039.1
Elymus hystrix	StH	FJ711012.1 FJ711013.1
Elymus sibiricus	StH	FJ711048.1
Elymus wawawaiensis	StH	FJ711010.1 FJ711011.1
Eremopyrum distans	F	FJ711018.1
Eremopyrum triticeum	F	FJ711028.1
Henrardia persica	O	FJ711029.1
Heteranthelium piliferum	Q	FJ711030.1
Hordeum bogdanii	H	FJ711020.1
Hordeum brevisubulatum	H	FJ11019.1
Hordeum chilense	H	FJ711017.1
Hordeum vulgare	I	DQ290773.1
Hystrix duthiei	NsXm	FJ711043.1
Kengyilia alatavica	StYP	JF965594.1 JF965595.1 JF965596.1
Kengyilia batalinii	StYP	JF965580.1 JF965579.1 JF965586.1
Kengyilia gobicola	StYP	JF965583.1 JF965588.1 JF965607.1
Kengyilia hirsuta	StYP	JF965576.1 JF965577.1 JF965578.1
Kengyilia kaschgarica	StYP	JF965584.1 JF965600.1 JF965601.1
Kengyilia kokonorica	StYP	JF965578.1 JF965587.1 JF965615.1
Kengyilia longiglumis	StYP	JF965604.1 JF965603.1 JF965611.1
Kengyilia melanthera	StYP	JF965592.1 JF965593.1 JF965614.1
Kengyilia rigidula	StYP	JF965574.1 JF965606.1 JF965616.1
Kengyilia tahelacana	StYP	JF965590.1 JF965591.1 JF965605.1
Leymus akmolinensis	NsXm	FJ711038.1
Leymus arenarius	NsXm	FJ711046.1
Leymus pseudoracemosus	NsXm	FJ711040.1
Leymus secalinus	NsXm	FJ711041.1
Leymus triticoides	NsXm	FJ711042.1
Lophophyrum elongatum	E^e	FJ711035.1
Peridictyon sanctum	Xp	FJ711037.1
Psathyrostachys fragilis	Ns	FJ711016.1

Psathyrostachys juncea	Ns	FJ711031.1
Pseudoroegneria libanotica	St	FJ711032.1
Pseudoroegneria spicata	St	FJ711015.1
Pseudoroegneria stipifolia	St	FJ711033.1
Pseudoroegneria strigosa	St	FJ711034.1
Secale cereale	R	AF343493.1
Taeniatherum caput-medusae	Ta	FJ711021.1
Thinopyrum bessarabicum	E^{b}	FJ711036.1
Triticum aestivum subsp. tibeticum	$A^{u}BD$	JQ327064.1 JQ327063.1 JQ327125.1
Triticum carthlicum	$A^{u}B$	JQ327065.1 JQ327066.1 JQ327073.1
Triticum compactum	$A^{u}BD$	JQ327070.1 JQ327071.1 JQ327072.1
Triticum durum	$A^{u}B$	JQ327116.1 JQ327115.1
Triticum monococcum subsp. aegilopoides	A^{m}	DQ290665.1
Triticum sphaerococcum	$A^{u}BD$	JQ327094.1 JQ327117.7 JQ327118.1 JQ327119.1
Triticum timopheevii subsp. armeniacum	$A^{u}G$	DQ364906.1 DQ364895.1
Triticum turgidum subsp. dicoccum	$A^{u}B$	DQ290692.1 DQ290736.1
Triticum urartu	A^{u}	DQ290659.1

All genomic designations as per Wang et al. 1995

3. Results

3.1 Sequence Analysis

Following our preliminary analysis of redundant sequences, 96 final representative sequences were selected from 60 species spanning 21 genera (excluding outgroup). Multiple distinct homologs of *pgk1* sequences were obtained for most polyploid species (i.e. one copy from each haplome present in the genome), except for *Elymus caninus*, and *Elymus sibiricus* wherein only one accession from each was available. Additionally in *Leymus*, separate distinct types of *pgk1* genes were unavailable and were represented by one accession for each species. Multiple types of the gene would have been uninformative due to the lack of knowledge regarding their Xm genome (Fan et al., 2009).

Our *pgk1* matrix included 669 aligned characters of which 14.5% (97/669) were variable (parsimony uninformative) and 22.30% (149/669) were parsimony informative. Our phylogenetic analyses resulted in many equally parsimonious trees of length 511. Sequence alignment showed there were 3 major indels within the *pgk1* loci in the Triticeae. The first was a 25bp insertion from first to the 25th position shown in the S' genome species *Aegilops searsii* (DQ290761.1), *Aegilops speltoides,* (DQ290768.1), *Aegilops sharonensis* (DQ290767.1), *Aegilops longissima* (DQ290764.1), and *Aegilops bicornis* (DQ290763.1). This indel also occurred in the I genome of *Hordeum vulgare* (DQ290773.1), and A' genome species *Triticum monococcum subsp. aegilopoides* (DQ290665.1), and *Triticum urartu* (DQ290659.I), as well as the D genome species *Aegilops tauschii* (DQ290771.1). Second was a 6bp insertion from positions 93-98

present in all nine members of the A^u group of *Triticum*. Lastly, an 8bp insertion in sites 511-518 was present in diploid species *Agropyron cristatum* (JF965633.1) (figure 1).

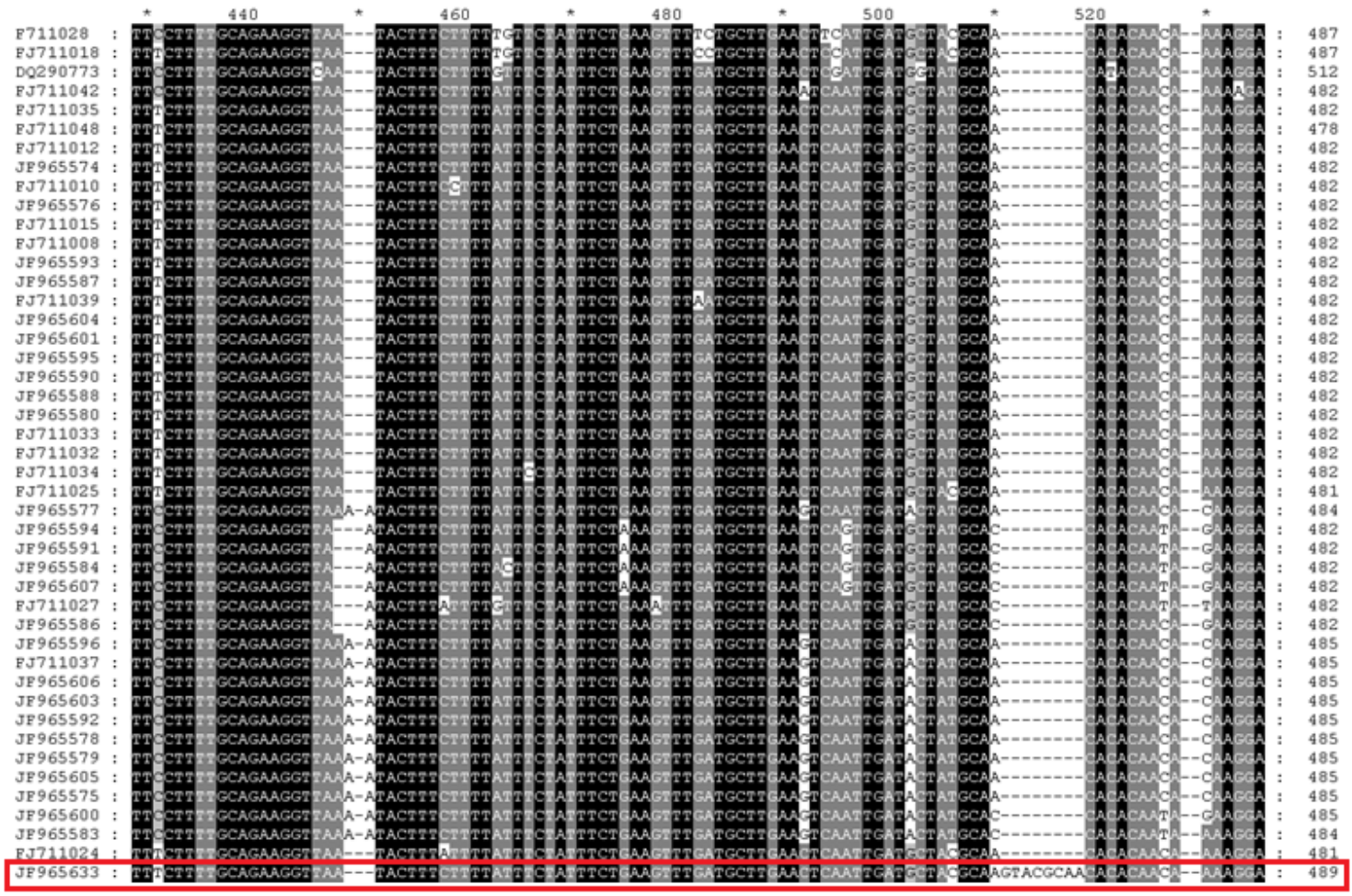


Figure 1: Example of indel. Partial alignment of *pgk1* accesssions from Triticeae showing distinct 8bp insertion at the 511-518bp site in species *Agropyron cristatum* (JF965633). Insertion outlined in red.

3.2 Phylogenetic Analysis

Maximum Parsimony and Neighbor-Joining analyses yielded five distinct clades; three highly supported clades and two clades with bootstrap values less than 50% (Figure 2 and Figure 3). The five clades are representative of each St, H, A^u, B, D, P, and Ns genomes present in polyploid Triticeae. Genomes A, B and D all belong to the *Triticum*/*Aegilops* complex which, formed a monophyly, and were grouped together into clade III. Bootstrap (BS) values are labeled underneath branches. Tree topology among both trees is fairly constant, but varies with the least supported clade IV; occurring on the basal region of the MP tree (Figure 2), while occurring higher up, and not as a monophyly in the NJ tree (Figure 3). This variation in topology of clade IV can be seen to displace clades II and III lower on the NJ tree than in the MP tree.

Clade I (95% BS) was composed of the St genomic sequences from polyploid *Kengylia* and *Elymus*, and diploid *Pseudoroegneria*. *Lophophyrum elongatum* (E^eE^e) also groups in this clade (76% BS). Clade II (100% BS) represents the sequences from H genome *Elymus* and *Hordeum*. Clade III (100% BS) was split into three subclades. Each subclade respectively represented A^u-type (100% BS) *Triticum* species, B-type (97% BS), and D-type (99% BS) *Triticum*/*Aegilops* species. *Aegilops speltoides* (BB) forms a poorly supported (<50% BS) monophyly with *Aegilops searsii* (S^sS^s), sister to the B and D-genome groups. *Triticum timopheevii ssp. armeniacum* (DQ364906.1) can be seen splitting the B and D genome groups (<50% BS). Also included within clade III are *Taeniatherum caput-medusae* (TaTa), *Crithopsis delileana* (KK), *Heteranthelium piliferum* (QQ), *Thinopyrum bessarabicum* (E^bE^b), *Secale cereal* (RR), and *Henrardia persica* (OO), all with lower bootstrap support. Within the NJ tree, *Psathyrostachys*

fragilis (NsNs) is included within clade III, while sister to *Eremopyrum* (FF) in our MP tree, with lower support in each case. Clade IV (<50% BS) is split into two weakly supported subclades. The top clade represents Y-genome *Kengyilia*, as well as *Peridictyon sanctum* (XpXp). The bottom clade represents the P-genome and is itself made up of two sub-clades, each with much stronger shallow nodes (for example, *K. alatavica*, *K. tahelacana*, *K. kaschgarica*, and *K. gobicola* all group with 100% BS). *Agropyron* P-type species group with *Kengyilia* (69% BS) in the lower clade, while *Dasypyrum villosum* (VV) forms a monophyly with the upper clade. Clade IV is split up in the NJ tree, grouping the *Agropyron*/*Kengyilia* clade sister to the monophyly containing clade I, *Eremopyrum* (FF), and *Australopyrum retrofractum* (WW), while grouping the *Kengyilia*/*Dasypyrum* clade as Y-genome species. Clade V is poorly supported, although its two sub-clades are highly supported, showing 96% BS and 100%BS respectively. Its members represent the Ns-genome groups of *Hystrix duthiei*, *Leymus*, and *Psathyrostachys* species. The well supported (100% BS) clade containing *Leymus arenarius* and *Psathyrostachys juncea* does not fall in clade V in the NJ tree, but instead forms its own monophyly basal to clades I, II and IV. In general, it should be noted that deeper clades show lower bootstrap values than shallow clades.

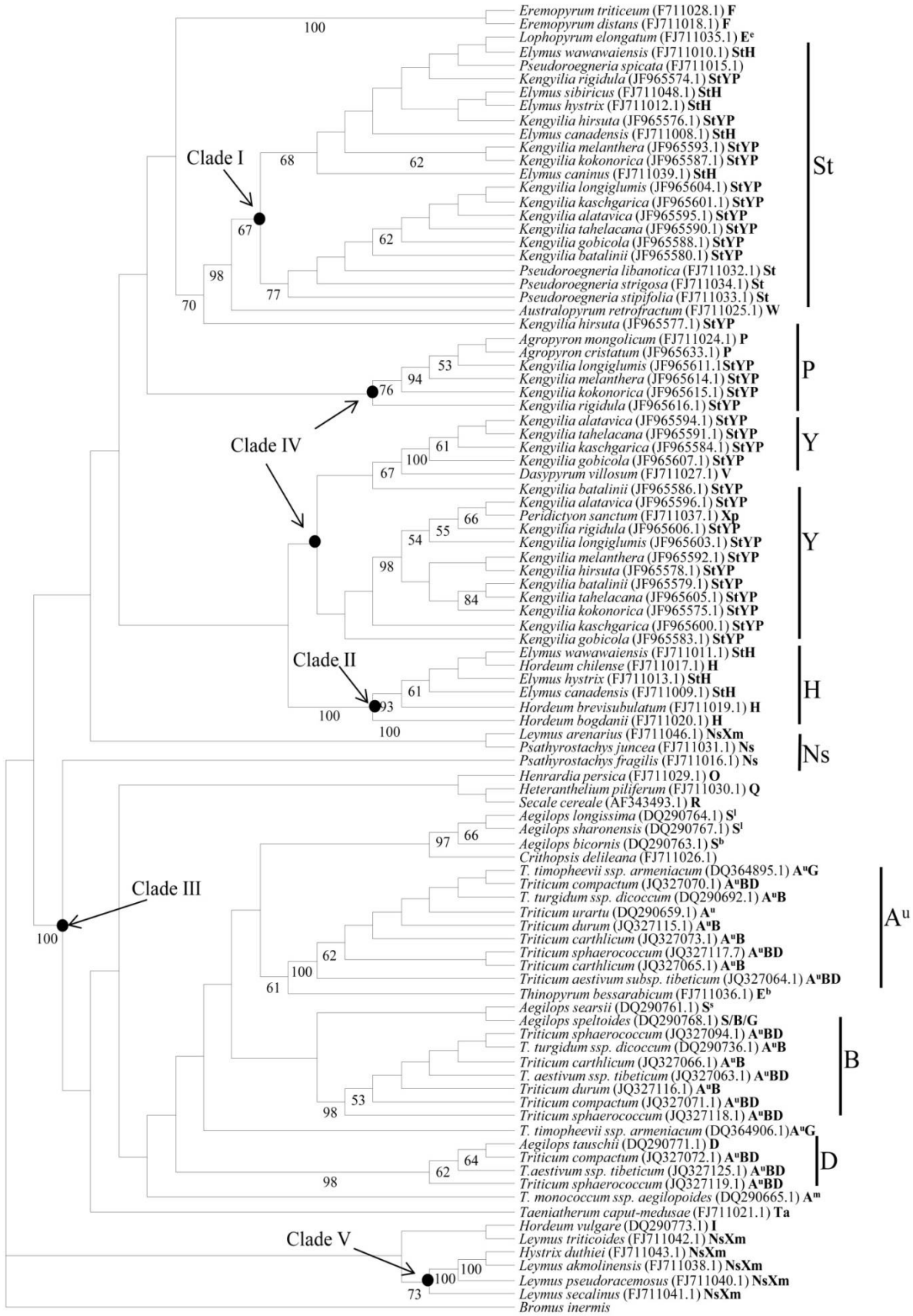
To properly reveal the diploid genome donors in polyploid Triticeae, adequate numbers of representatives were chosen during the preliminary sequence analyses as outlined in materials and methods. As previously noted, multiple homologs of *pgk1* sequences were chosen to represent each single-copy gene from each distinct genome within the polyploid Triticeae. Accordingly, homologous *pgk1* sequences from distinct genomes grouped separately in all polyploid Triticeae species represented with multiple sequences; all

polyploid *Kengyilia* species split into St-, P-, and Y- genomes, with the exception of *Kengyilia gobicola* and *Kengyilia kaschgarica* in the MP tree, which each grouped one sequence (JF965588.1, and JF965601.1 respectively) among the St-group of clade I, and two accessions (JF965583.1, JF965607.1, and JF965584.1, JF965600.1 respectively) within the P-genome group of clade VI. All *Kengyilia* accessions split into separate haplome groups in the NJ tree. *Elymus* sequences split accordingly between the H and St genomes, with the exception of *Elymus caninus* and *Elymus sibiricus* of which only one sequence from each species was available, both of which grouped with the St-genome. All sequences from tetraploid *Triticum* species split into different clades according to their haplomes homologs, as well as all hexaploid *Triticum* sequences.

Figure 2: A most parsimonious tree based on *pgk1* sequences. Bootstrap support values (>50%) are shown below nodes. Accession numbers, species name, and genome are labeled on each branch. Genome groupings are indicated to right of tree in bold.


Eremopyrum triticeum (F711028.1) F
Eremopyrum distans (FJ711018.1) F
Psathyrostachys fragilis (FJ11016.1) Ns
Lophopyrum elongatum (FJ711035.1) E^e
Kengyilia rigidula (JF965574.1) StYP
Elymus hystrix (FJ711012.1) StH
Kengyilia melanthera (JF965593.1) StYP
Kengyilia kokonorica (JF965587.1) StYP
Elymus caninus (FJ711039.1) StH
Kengyilia hirsuta (JF965576.1) StYP
Elymus canadensis (FJ711008.1) StH
Elymus wawawaiensis (FJ711010.1) StH
Elymus sibiricus (FJ711048.1) StH
Pseudoroegneria spicata (FJ711015.1) St
Pseudoroegneria stipifolia (FJ711033.1) St
Kengyilia longiglumis (JF965604.1) StYP
Kengyilia kaschgarica (JF965601.1) StYP
Kengyilia alatavica (JF965595.1) StYP
Kengyilia tahelacana (JF965590.1) StYP
Kengyilia gobicola (JF965588.1) StYP
Kengyilia batalinii (JF965580.1) StYP
Pseudoroegneria strigosa (FJ711034.1) St
Pseudoroegneria libanotica (FJ711032.1) St
Australopyrum retrofractum (FJ711025.1) W
Hordeum vulgare (DQ290773.1) I
Elymus wawawaiensis (FJ711011.1) StH
Elymus hystrix (FJ711013.1) StH
Elymus canadensis (FJ711009.1) StH
Hordeum chilense (FJ711017.1) H
Hordeum brevisubulatum (FJ711019.1) H
Hordeum bogdanii (FJ711020.1) H
Henrardia persica (FJ711029.1) O
Secale cereale (AF343493.1) R
Aegilops searsii (DQ290761.1) S^s
A. speltoides (DQ290768.1) S/B/G
T. sphaerococcum (JQ327094.1) A^uBD
T.turgidum ssp. dicoccum (DQ290736.1) A^uB
Triticum carthlicum (JQ327066.1) A^uB
T.aestivum ssp. tibeticum (JQ327063.1) A^uBD
T.durum (JQ327116.1) A^uB
T. compactum (JQ327071.1) A^uBD
T. sphaerococcum (JQ327118.1) A^uBD
T.timopheevii ssp. armeniacum (DQ364906.1) A^uG
Aegilops tauschii (DQ290771.1) D
Triticum compactum (JQ327072.1) A^uBD
T.aestivum ssp. tibeticum (JQ327125.1) A^uBD
Triticum sphaerococcum (JQ327119.1) A^uBD
T. timopheevii ssp. armeniacum (DQ364895.1) A^uG
T.turgidum ssp. dicoccum (DQ290692.1) A^uB
Triticum urartu (DQ290659.1) A^u
Triticum durum (JQ327115.1) A^uB
Triticum carthlicum (JQ327073.1) A^uB
Triticum compactum (JQ327070.1) A^uBD
Triticum sphaerococcum (JQ327117.7) A^uBD
Triticum carthlicum (JQ327065.1) A^uB
T. aestivum ssp. tibeticum (JQ327064.1) A^uBD
Thinopyrum bessarabicum (FJ711036.1) E^b
Heteranthelium piliferum (FJ711030.1) Q
Aegilops longissima (DQ290764.1) S^l
Aegilops bicornis (DQ290763.1) S^b
Aegilops sharonensis (DQ290767.1) S^l
Crithopsis delileana (FJ711026.1) K
T. monococcum ssp. aegilopoides (DQ290665.1) A^m
Taeniatherum caput-medusae (FJ711021.1) Ta
Kengyilia hirsuta (JF965577.1) StYP
Kengyilia alatavica (JF965596.1) StYP
Peridictyon sanctum (FJ711037.1) Xp
Kengyilia rigidula (JF965606.1) StYP
Kengyilia longiglumis (JF965603.1) StYP
Kengyilia melanthera (JF965592.1) StYP
Kengyilia kokonorica (JF965575.1) StYP
Kengyilia hirsuta (JF965578.1) StYP
Kengyilia batalinii (JF965579.1) StYP
Kengyilia tahelacana (JF965605.1) StYP
Kengyilia alatavica (JF965594.1) StYP
Kengyilia tahelacana (JF965591.1) StYP
Kengyilia kaschgarica (JF965584.1) StYP
Kengyilia gobicola (JF965607.1) StYP
Dasypyrum villosum (FJ11027.1) V
Kengyilia kaschgarica (JF965600.1) StYP
Kengyilia gobicola (JF965583.1) StYP
Kengyilia batalinii (JF965586.1) StYP
Agropyron mongolicum (FJ711024.1) P
Agropyron cristatum (JF965633.1) P
Kengyilia longiglumis (JF965611.1) StYP
Kengyilia melanthera (JF965614.1) StYP
Kengyilia kokonorica (JF965615.1) StYP
Kengyilia rigidula (JF965616.1) StYP
Hystrix duthiei (FJ711043.1) NsXm
Leymus akmolinensis (FJ711038.1) NsXm
Leymus pseudoracemosus (FJ711040.1) NsXm
Leymus arenarius (FJ711046.1) NsXm
Psathyrostachys juncea (FJ711031.1) Ns
Leymus triticoides (FJ711042.1) NsXm
Leymus secalinus (FJ711041.1) NsXm
Bromus inermisFJ11014
Clade I
Clade II
Clade III
Clade IV
Clade V
St
H
B
D
A^u
P
P
Ns
5 changes

Figure 3: Phylogenetic tree obtained by Neighbor-Joining method. Bootstrap support values (>50%) are shown below nodes. Accession numbers, species name, and genome are labeled on each branch. Genome groupings are indicated to right of tree in bold.


Eremopyrum triticeum (F711028.1) F
Eremopyrum distans (FJ711018.1) F
Lophopyrum elongatum (FJ711035.1) E^e
Elymus wawawaiensis (FJ711010.1) StH
Pseudoroegneria spicata (FJ711015.1)
Kengyilia rigidula (JF965574.1) StYP
Elymus sibiricus (FJ711048.1) StH
Elymus hystrix (FJ711012.1) StH
Kengyilia hirsuta (JF965576.1) StYP
Elymus canadensis (FJ711008.1) StH
Kengyilia melanthera (JF965593.1) StYP
Kengyilia kokonorica (JF965587.1) StYP
Elymus caninus (FJ711039.1) StH
Kengyilia longiglumis (JF965604.1) StYP
Kengyilia kaschgarica (JF965601.1) StYP
Kengyilia alatavica (JF965595.1) StYP
Kengyilia tahelacana (JF965590.1) StYP
Kengyilia gobicola (JF965588.1) StYP
Kengyilia batalinii (JF965580.1) StYP
Pseudoroegneria libanotica (FJ711032.1) St
Pseudoroegneria strigosa (FJ711034.1) St
Pseudoroegneria stipifolia (FJ711033.1) St
Australopyrum retrofractum (FJ711025.1) W
Kengyilia hirsuta (JF965577.1) StYP
Agropyron mongolicum (FJ711024.1) P
Agropyron cristatum (JF965633.1) P
Kengyilia longiglumis (JF965611.1StYP
Kengyilia melanthera (JF965614.1) StYP
Kengyilia kokonorica (JF965615.1) StYP
Kengyilia rigidula (JF965616.1) StYP
Kengyilia alatavica (JF965594.1) StYP
Kengyilia tahelacana (JF965591.1) StYP
Kengyilia kaschgarica (JF965584.1) StYP
Kengyilia gobicola (JF965607.1) StYP
Dasypyrum villosum (FJ711027.1) V
Kengyilia batalinii (JF965586.1) StYP
Kengyilia alatavica (JF965596.1) StYP
Peridictyon sanctum (FJ711037.1) Xp
Kengyilia rigidula (JF965606.1) StYP
Kengyilia longiglumis (JF965603.1) StYP
Kengyilia melanthera (JF965592.1) StYP
Kengyilia hirsuta (JF965578.1) StYP
Kengyilia batalinii (JF965579.1) StYP
Kengyilia tahelacana (JF965605.1) StYP
Kengyilia kokonorica (JF965575.1) StYP
Kengyilia kaschgarica (JF965600.1) StYP
Kengyilia gobicola (JF965583.1) StYP
Elymus wawawaiensis (FJ711011.1) StH
Hordeum chilense (FJ711017.1) H
Elymus hystrix (FJ711013.1) StH
Elymus canadensis (FJ711009.1) StH
Hordeum brevisubulatum (FJ711019.1) H
Hordeum bogdanii (FJ711020.1) H
Leymus arenarius (FJ711046.1) NsXm
Psathyrostachys juncea (FJ711031.1) Ns
Psathyrostachys fragilis (FJ711016.1) Ns
Henrardia persica (FJ711029.1) O
Heteranthelium piliferum (FJ711030.1) Q
Secale cereale (AF343493.1) R
Aegilops longissima (DQ290764.1) S^l
Aegilops sharonensis (DQ290767.1) S^l
Aegilops bicornis (DQ290763.1) S^b
Crithopsis delileana (FJ711026.1)
T. timopheevii ssp. armeniacum (DQ364895.1) A^uG
Triticum compactum (JQ327070.1) A^uBD
T. turgidum ssp. dicoccum (DQ290692.1) A^uB
Triticum urartu (DQ290659.1) A^u
Triticum durum (JQ327115.1) A^uB
Triticum carthlicum (JQ327073.1) A^uB
Triticum sphaerococcum (JQ327117.7) A^uBD
Triticum carthlicum (JQ327065.1) A^uB
Triticum aestivum subsp. tibeticum (JQ327064.1) A^uBD
Thinopyrum bessarabicum (FJ711036.1) E^b
Aegilops searsii (DQ290761.1) S^s
Aegilops speltoides (DQ290768.1) S/B/G
Triticum sphaerococcum (JQ327094.1) A^uBD
T. turgidum ssp. dicoccum (DQ290736.1) A^uB
Triticum carthlicum (JQ327066.1) A^uB
T. aestivum ssp. tibeticum (JQ327063.1) A^uBD
Triticum durum (JQ327116.1) A^uB
Triticum compactum (JQ327071.1) A^uBD
Triticum sphaerococcum (JQ327118.1) A^uBD
T. timopheevii ssp. armeniacum (DQ364906.1)A^uG
Aegilops tauschii (DQ290771.1) D
Triticum compactum (JQ327072.1) A^uBD
T.aestivum ssp. tibeticum (JQ327125.1) A^uBD
Triticum sphaerococcum (JQ327119.1) A^uBD
T. monococcum ssp. aegilopoides (DQ290665.1) A^m
Taeniatherum caput-medusae (FJ711021.1) Ta
Hordeum vulgare (DQ290773.1) I
Leymus triticoides (FJ711042.1) NsXm
Hystrix duthiei (FJ711043.1) NsXm
Leymus akmolinensis (FJ711038.1) NsXm
Leymus pseudoracemosus (FJ711040.1) NsXm
Leymus secalinus (FJ711041.1) NsXm
Bromus inermis
Clade I
Clade IV
Clade II
Clade III
Clade V
St
P
Y
Y
H
Ns
A^u
B
D

3.3 Nucleotide Diversity and Normality

Tajima's π and Watterson's θ were used determine nucleotide diversity within the tribe. Separate analyses were carried out to compare diversity between polyploid and diploid species within a specific genome group, as well as between polyploid genera where possible (ie. Within the St and Ns genomes; all other genomes had single genera representing polyploid sequences) (Table 2). Clade V (Ns) represented the highest nucleotide diversity values for both polyploid and diploid species with π= 0.03343 and θ= 0.03536 for polyploid Ns genome sequences (π= 0.04103 and θ= 0.04158 for *Leymus*, and π= 0.00148 and θ= 0.00148 for *Hystrix* respectively), and π= 0.03886 and θ= 0.03886 for diploid *Psathyrostachys* sequences. *Hystrix* represented the lowest values for polyploid species, but was also represented by the fewest accessions, while *Triticum urartu* represented the lowest diversity among diploid species at π= 0.0011 and θ= 0.0011. The number of polymorphic sites varied in each group, with diploid *Agropyron* (B) representing the highest abundance of polymorphisms with S=69, and diploid *Triticum urartu* (A^u) and polyploid *Hystrix* (Ns) tying for the lowest (S=1). Tajima (1989) and Fu and Li's (1993) D statistics were calculated for polyploid and diploid members of each genomic group (groups represented by fewer than 4 sequences were excluded from this analysis) (Figure 2). Tajima's D values within polyploid St genome species (-1.90096), polyploid A^u genome (-2.42928) and D genome *Triticum* (-1.82540) sequences were all significant. Fu and Li's D values for St genome *Kengyilia* (-2.4447), A^u genome *Triticum* (-4.27290), and polyploid D genome sequences (-2.21165) were also significant (Table 2).

Table 2: Tajima's π, Watterson's θ, and Sequence site polymorphisms (S) for all taxa. For species represented with at least 4 members, Tajima's and Fu and Li's D values are expressed, significant values are in red. N= number of sequences.

Genome	N	π	θ	S	Tajima's D	Fu and Li's D
St						
Polyploid	19	0.0160	0.0298	68	-1.90096	-2.10028
Elymus	6	0.0221	0.0287	44	-1.04607	-1.49227
Kengyilia	13	0.0130	0.0222	45	-1.85902	-2.44447
Diploid						
Pseudoroegneria	4	0.0081	0.0089	11	-0.83741	-0.83741
H						
Polyploid						
Elymus	3	0.0030	0.0030	3		
Diploid						
Hordeum	3	0.0119	0.0119	12		
A^u						
Polyploid						
Triticum	30	0.0043	0.0130	28	-2.42928	-4.27290
Diploid						
T. urartu	3	0.0011	0.0011	1		
B						
Polyploid						
Triticum	88	0.00483	0.01088	32	-1.73391	-1.60532
Diploid						
A. speltoides	22	0.02441	0.02941	58	-0.67723	-1.21404
D						
Polyploid						
Triticum	11	0.00468	0.00799	14	-1.82540	-2.21165
Diploid						
A. tauschii	3	0.00223	0.00223	2		
P						
Polyploid						
Kengyilia	13	0.03150	0.02665	53	0.81745	0.53340
Diploid						
Agropyron	18	0.02814	0.03186	69	-0.48636	-1.46968
Ns						
Polyploid	7	0.03343	0.03536	54	-0.31521	-0.17684
Leymus	5	0.04103	0.04158	54	-0.09946	0.01105
Hystrix	2	0.00148	0.00148	1		
Diploid						
Psathyrostachys	2	0.03886	0.03886	26		

Discussion

4.1 Origin of Polyploid Triticeae

Despite substantial research on the Triticeae, a common taxonomy has yet to be agreed upon. Kellogg (1996) proposed two possible reasons for this. The first being that all studies are faulty due to differing methodological shortcomings and sampling biases, and the second that each study is correct in its own way; each gene representing a different evolutionary history and pattern of hybridization and introgression within an extremely complex tribe. Before accepting either of these possibilities, this study re-examined the history of polyploid Triticeae, as well as the evolution of the 3-phosphoglycerate kinase encoding nuclear *pgk1* gene within the tribe. This analysis was performed with a larger data set, and represented a comprehensive analysis of all of major genomes in the tribe, contrary to many previous studies focusing specifically within a common genus or group of genera (Hsiao et al., 1995; Huang et al., 2001; 2002; Mason-Gamer et al., 2002; Sun, 2006; Kilian et al., 2007; Fan et al., 2012).

Our Phylogenetic analyses (Maximum Parsimony and Neighbor-Joining) yielded some well supported clades (Clades I, and II and III in particular). Deeper tree branches are generally shorter, and less supported than shallow nodes within clades IV and V in particular. Escobar et al. (2011) propose that this pattern could be an indication of rapid species radiation, and patterns of incomplete lineage sorting.

Our clade I grouping of diploid St genome *Pseudoroegneria* with polyploid St type *Elymus* and *Kengyilia* is extremely well supported in the literature. In specific our phylogenetic analysis supports findings by Hodge et al. (2010) whom used data from the

chloroplast rps16 gene, and Mason-Gamer (2010) through analyses of the nuclear starch synthase gene, both linking *Pseudoroegneria* as the donor of the St genome in *Elymus*. In relation to the St genome *Kengyilia* species, our findings are congruent with that of Fan et al. (2012). In direct relation to high bootstrap support in both our MP (95% BS) and NJ trees (67% BS; 98% BS if *Australopyrum retrofractum* is included), and congruence among literature, we agree that diploid *Pseudoroegneria* is the donor of the St genome both within StH polyploid *Elymus* species, and StYP *Kengyilia*.

Polyploid *Elymus pgk1* haplotypes also grouped among diploid *Hordeum* within clade II, representing the origin of the H genome in the StH *Elymus*. Through analysis of the RPB2 gene, Sun et al. (2008) found strong geographic origin-dependant grouping among *Elymus* and *Hordeum*. While the present study did not take geographical origin into consideration, the grouping of *H. bogdanii* with *E. hystrix* and *E. wawawaiensis* is well supported in our study (100% BS for both MJ and NJ). In addition, Liu et al. (2005) also found that StH polyploid *Elymus* resulted through hybridization between *Pseudoroegneria* (St) and an unspecified *Hordeum* (H) species. These data, along with our findings lead us to support *Hordeum* as the donor of the H haplome in at least some (most likely North American) StH *Elymus* species.

Clade III was representative of the *Triticum*/*Aegilops* complex. Results on the origins of the A^u, B, and D genomes respectively within the *Triticeae* follow the findings of Petersen et al. (2006). In particular, our data best fits with their analyses of the nuclear DMC1 and EF-G genes. Our A^u genome clade is well supported (100% BS), sharing in particular the grouping of *Triticum urartu* (A^uA^u) and common wheat (*T. aestivum*). A study by Dvorak et al. (1988) investigating the origin of the polyploid A^u genome and its

relationship between *T. urartu* and *T. monococcum* ssp. *aegilopoides* found that diploid *T.m.* ssp. *monococcum* was domesticated from *T. monococcum* ssp. *aegilopoides*, while the A^u genome of *T. urartu* was synonymous to the A^u genome of polyploid wheat, as is supported by our study. Kilian et al. (2007) found through 65 AFLPs specific to the *T.aestivum* B genome, that *Aegilops speltoides* is a relative of the extant B genome in polyploid *Triticeae*. These findings are consistent with Petersen et al. (2006) whom repeatedly found *A. speltoides* to group sister to the monophyly containing *T. turgidum* and *T. aestivum*. This finding was mirrored in our study, although differing in that *A. speltoides* formed a monophyly with *A. searsii* (S^s) in both MP and NJ trees. This may reflect the relationship between *Aegilops speltoides* and its current S genome designation. Its forming of a sister clade to extant B genome species may emphasize its past relationship with the B genome of polyploid wheat. Again concurrent with Petersen et al. *Aegilops tauschii* grouped as the D genome donor to polyploid wheat with high affinity in both MJ and NJ trees (99% BS and 98%BS respectively). With 100% BS values for both our MP and NJ trees, and accordance with literature, we assume with great support the classification of *T. urartu*, *A. speltoides*, and *A. tauschii* as the donors of the A^u, B, and D genomes respectively to modern polyploid wheat.

Clade IV is our most unsupported clade, showing different topology within both the MJ and NJ trees. Although not directly covered in this study, we attribute this to the current unresolved origin of the Y genome in extant *Kengyilia* (Wang et al. 1999). Sun et al. (2008) found that in polyploid *Elymus* StY genome species that the Y genome was distinctly separate from the St genome, suggesting its own diploid origin. This finding was supported by Fan et al. (2012) in StYP *Kengyilia*. Our data correlates with their

findings that Y genome *Kengyilia* group with the Xp genome species *Peridictyon sanctum* and is distinctly separate from both St and P genome *Kengyilia*. Additionally Sha et al. (unpublished data, noted by Fan et al., 2012) found that Y genome Triticeae grouped along with *Dasypyrum* (V) species, which is supported by grouping in our NJ tree. Although merely speculation, this data supports the possibility that ancient Y genome species may be related to the extant Xp or V genomes of *P. sanctum* and *Dasypyrum villosum* respectively. *Agropyron* as the putative P genome donor occurs with relatively high probability in both our MP and NJ trees (75% BS and 76% BS respectively). This correlates with findings from Fan et al. (2012) through nuclear analysis, and is mirrored by Zhang et al. (2009) through analysis of the chloroplast *trnL-F* gene.

Clade V has low support, however the clade of *Psathyrostachys juncea* and *Leymus triticoides* show a very high affinity (100% BS) in our MP tree. This suggests that *Psathyrostachys juncea* is the donor of the Ns genome to *Leymus* NsXm species. This result agreed with Wang et al. (2005), whom placed *Psathyrostachys juncea* as the likely donor of the Ns genome over *Psathyrostachys fragilis* based upon the fact that all *Leymus* species have plants showing at least 2 pLrTaiI-1 sites, while *Psathyrostachys fragilis* lacks this site completely. In our study, *Psathyrostachys fragilis* does not group amongst *Leymus* in either phylogenetic analysis, supporting this finding.

After analyses, we support that the respective putative donors of the A^u, B, D, St, H, P, and Ns haplomes in polyploid Triticeae are from diploid *T. urartu, A. speltoides* (although current designation is S genome), *A. tauschii*, *Pseudoroegneria*, *Hordeum*, *Agropyron*

and *Psathyrostachys juncea* respectively. These findings confirm our predictions based upon cytogenetic data.

4.2 Evolution of pgk1 in the Triticeae

Initial phylogenetic analyses of the Triticeae resulted in the successful isolation of divergent *pgk1* homologous loci in both tetraploid and hexaploid species. This supports research by Fan et al. (2012) in polyploid *Kengyilia* whom isolated triplicate homoeologues of the *pgk1* gene in all hexaploid species. These data suggest that the *pgk1* single copies from diploid progenitors are conserved in duplicate, or triplicate (depending on ploidy levels) within the genome of polyploid Triticeae.

Analyses of nucleotide diversity of the *pgk1* genes (π) in the Triticeae gave interesting results. Among polyploid *Triticum* belonging to the A^u, B, and D genomes, rates of variation were all very similar (0.0043, 0.00483, and 0.00468 respectively), and showed little diversity among sites. This could be attributed to higher rates of hybridization and introgression within the species, over time promoting sequence homology at the *pgk1* locus (Gu et al., 2004). A better reason for this pattern, which we attribute to the nucleotide variation in polyploid wheat in our study, is the narrowing of the genetic diversity in polyploid *Triticum* species due to the extensive amount of cultivation that the species has undergone in the past 10,000 years (Caldwell et al., 2004).

Expected patterns within cultivated species would be much lower sequence variability in modern polyploid species compared to their diploid progenitors. Caldwell et al. (2004) attribute this again to the prevalence of genome narrowing as a result of gene selection by cultivars desiring certain phenotypic traits in their crops. In our samples, we see this

pattern most prominently among the B genome polyploid and diploid *Triticum* and *Aegilops speltoides* (π= 0.00483 and π=0.02441) with nucleotide variation among diploid species at five times higher than that of polyploid species (table 2). These differences in variability however are far less than Caldwell et al. (2004) observed; *Aegilops tauschii* was shown to have a 30-fold rate of variability compared to that of D genome hexaploid wheat. The present study noted a roughly two-fold increase in the variability of polyploid wheat versus diploid *A. tauschii*; however *A. tauschii* was only represented by 3 sequences, while Caldwell et al. (2004) used 198 representative sequences. This pattern of lower polyploid sequence variability was also found in polyploid Ns species and diploid *Psathyrostachys*. While variation in *Leymus* nucleotides was more than that of its Ns donor, *Psathyrostachys* had a rate of variation of roughly 30-times that of *Hystrix*. These species were all represented by a low number of sequences, and a study including more would be desirable to observe the differences and eliminate possible errors due to sample size. Additionally, the high amount of nucleotide variability between *Psathyrostachys juncea* and *Psathyrostachys fragilis* further supports the findings by Wang et al. (2005) in genome differences between the two species previously touched upon, even at homologous loci.

Contrary to expected results, we observe increased nucleotide variability in polyploid species in particular within the St, A^u, D and P genomes. This may be linked to higher rates of evolution observed in the *pgk1* gene within polyploid members of these genomes (Fan et al., 2009). Fan et al. (2009) additionally proposed increased variability in nucleotides as a direct result of separate incidences of hybridization between different populations, or even different species of the same genus. This is fairly well supported

within the St, and P genomes, wherein no specific species has yet been linked as the putative donors to the genomes in each case.

Tajima's D values and Fu and Li's D values were significant for polyploid St *Kengyilia* (-1.85902, -2.4447), A^u (-2.42928, -4.27290), and D genome (-1.82540, -2.21165) *Triticum*. In all significant cases, highly negative Tajima's D values can be seen with relatively high rates of polymorphisms (S), but low rates of sequence diversity (Table 2). Holsinger (2012) correlates this trend with incidence of purifying selection on the genome. From our case, this can be explained as relatively high amounts of mutations in 'silent' regions of the *pgk1* gene, that do not proliferate in the genome (ie. remain rare). Fan et al. (2012) found through comparison of the P genome between polyploid and diploid *Kengyilia* evidence of purifying selection in the *pgk1* gene. Although not all significant, the prevalence of negative Tajima's D and Fu and Li's values provides substantiation that there are large forces of purifying selection acting on the *pgk1* gene within the polyploid Triticeae, effectively conserving its function within the polyploid genomes.

Variations in the nucleotide diversity and levels of polymorphism (S) of the *pgk1* gene between diploid species within the Triticeae may be attributed to the large differences in evolutionary pressures between each genome, supporting their genomic distinctions from one another. Future studies investigating the evolution of the *pgk1* gene within the tribe should increase the sample size of many of the diploid species (ex. *Hordeum* and *Triticum urartu* were only represented by 3 sequences, while *Psathyrostachys* was only represented by 2 sequences), as larger sample size is observed to give more legitimate differences between polyploid and diploid species (Caldwell et al., 2004)

As for the investigation of polyploid origins in the Triticeae, this study provided excellent justification to currently accepted suggestions. Future studies should further investigate the relationships of the Y genome in *Elymus* and *Kengyilia* species to establish its origin and relation to the Xp genome of *Peridictyon sanctum*, as well as the origin to the Xm genome of *Leymus* and *Hystrix*.

References

Anderson, E., 1949. Introgressive Hybridization. Biological Reviews. 28(3). 280-307.

Barkworth, M.E. 1992. Taxonomy of the Triticeae: a historical perspective. *Hereditas* 116: 1-14.

Blattner, F.R. 2009. Progress in Phylogenetic Analysis and a new Infrageneric Classification of the barley genus *Hordeum* (Poaceae: Triticeae). Breeding Sciences 59: 471-480.

Caldwell, K.S., Dvorak, J., Lagudah, E.S., Akhunov, E., Luo, M-C., Wolters, P., Powell,W. 2004. Sequence Polymorphism in Polyploid Wheat and Their D-Genome Diploid Ancestor. Genetics Society of America. DOI: 10.1534/genetics.103.016303

Clayton, D.W., Renvoize, S.A. 1986. Genera graminum: Grasses of the World. edited by T.A. Cope.

Doebley, J., von Bothmer, R., Larson, S. 1992. Chloroplast DNA variation and the phylogeny of hordeum (Poaceae). American Journal of Botany 79(5): 576-584.

Dvorak, J., McGuire, P., Cassidy, B.1988. Apparent sources of A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30: 630-689

Escobar, J., Scornavacca, C., Cenci, A., Guilhaumon, C., Santoni, S., Douzery, E., Ranwez, V., Glemin, S., David, J. 2011. Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). BMC Evolutionary Biology 11:181.

Fan, X., Sha, L-N, Zeng, J., Kang, H-Y, Zhang, H-Q, Wang, X-L, Zhang, L., Yang, R-W, Ding, C-B, Zheng, Y-L, Zhou, Y-H. 2012. Evolutionary Dynamics of the *Pgk1* Gene in the Polyploid Genus *Kengyilia* (Triticeae: Poaceae) and Its Diploid Relatives. PLoS ONE 7(2). doi:10.1371/journal.pone.0031122

Fan, X., Zhang, H.-Q., Sha, L.-N., Zhang, L., Yang, R.-W., Ding, C.-B., Zhou, Y.-H. 2009. Phylogenetic analysis among *Hystrix*, *Leymus*, and its affinitive genera (Poaceae: Triticeae) based on the sequence of a gene encoding plastid acetyl-CoA carboxylase. Plant Science 172(4): 701-707.

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

Fu, Y.X., Li, W.H. 1997. Estimating ancestral population parameters. Genetics 133: 693-709.

Gu, Y.G., Coleman-Derr, D., Kong, X., Anderson, O.D. 2004. Rapid Genome Evolution Revealed by Comparative Sequence Analysis of Orthologous Regoins from Four Triticeae Genomes. Plant Physiol. 135: 459-470.

Han, F.P., Fedak, G., Oullet, T., Liu, B. 2003. Rapid genomic changes in interspecific and intergeneric hybrids and alloplyploids of Triticeae. Genome 46: 716-723.

Hodge, C.D.,Wang, H., Sun, G. Phylogenetic analysis of the maternal genome of tetraploid StStYY *Elymus* (Tritciceae: Poaceae) species and the monogenomic Triticeae based on rps16 sequence data. Plant Science 178(5): 463-468.

Holsinger, K.E.2012. Tajima's D, Fu's FS, Fay and Wu's H, and Zeng et al.'s E. http://darwin.eeb.uconn.edu/eeb348/lecturenotes/molevol-tajima.pdf. Accessed 22/03/2013.

Hsiao, C., Chatterton, N.J., Asay, K.H., Jensen, K.B. 1995. Phylogenetic relationship of the monogenomic species of the wheat tribe, Triticeae (Poaceae), inferred from nuclear rDNA (internal transcribed spacer) sequences. Genome 38(2): 211-223.

Huang, S., Sirikhachornkit, A., Faris, J., Su, X., Gill, B., Haselkorn, R., Gornicki, P. 2002. Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Molecular Biology 48: 805-820.

Huang, S., Sirikhachornkit, A., Faris, J., Su, X., Gill, B.S., Haselkorn, R., Gornicki, P. 2001. Phylogenetic analysis of acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Molecular Biology 48: 805-820.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., Gornicki, P. 2002.Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. PNAS 99 (12): 8133-8138.

Jakob, S., Blattner, F. 2006. A Chloroplast Genealogy of *Hordeum* (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequences for Phylogenetic Inference. Molecular Biology and Evolution. 23(8): 1602-1612.

Kellogg, E. 1989. Comments on genomic genera in the Triticeae (Poaceae). American Journal of Botany, 76 (6): 796-805.

Kellogg, E. 1996. When The Genes Tell Different Stories: the Diploid Genera of Triticeae (Gramineae). Systematic Botany 21(3): pp. 321-347

Kilian, B., Ozkan, H., Deusch, O., Effgen, S., Brandolini, A., Khol, J., Martin W., Salamini, F. 2007. Independent Wheat B and G Genome Origins in Outcrossing Aegilops Progenitor Haplotypes. Mol. Biol. Evol. 24(1): 217-227.

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.

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

Liu, Q., Ge, S., Tang, H., Zhang, Z., Zhu, G., Lu, B.-R. 2005. Phylogenetic relationships in *Elymus* (Poaceae: Triticeae) based on the nuclear ribosomal internal transcribed spacer and chloroplast *trntL-F* sequences. New Phytologist 170: 411-420.

Löve, Á. (1984) Conspectus of the Triticeae. Feddes Repertorium 95: 468-473.

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

Nishikawa, T., Salomon, B., Komatsuda, T., von Bothmer, R., Kadowaki, K. 2002. Molecular phylogeny of the genus *Hordeum* using three chloroplast DNA sequences. http://genome.nrc.ca. DOI: 10.1139/G02-088 *Hordeum* (Poaceae). American Journal of Botany. 79(5): 576-584.

Petersen, G., Seberg, O. 1996. Phylogenetic Analysis of the Triticeae (Poaceae) Based on *rpoA* Sequence Data. Molecular Phylogenetics and Evolution 7(2): 217-230.

Petersen, G., Seberg, O. 2003. Phylogenetic Analyses of the Diploid Species of *Hordeum* (Poaceae) and a Revised Classification of the Genus. Systematic Botany. 28(2): 293-306.

Petersen, G., Seberg, O., Yde, M., Berthelsen, K. 2006. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the A, B, and D genomes of common wheat (*Triticum aestivum*). Molecular Phylogenetics and Evolution 39: 70-82.

Singh, R.J. 2009. Genetic Resources, Chromosome Engineering, and Crop Improvement Volume 5. Taylor Francis Group, LLC. Boca Raton, Florida, USA.

Stebbins, G.L. 1947. Types of Polyploids: Their classification and significance. Advances in Genetics Volume 1. Academic Press Inc. 111 Fifth Avenue, New York, NY 10003.

Stebbins, G.L. 1956. Taxonomy and the evolution of genera with special reference to the family Graminae.

Sun, G.L. 2007. Genetic diversity of *rbcL* gene in *Elymus trachycaulus* complex and their relationship to several Triticeae species. Genet. Resour. Crop Evol. 54: 1737-1746.

Sun, G., Pourkheirandish, M., Komatsuda, T. 2009. Molecular evolution and phylogeny of the *RPB2* gene in the genus *Hordeum*. Annals of Botany 103: 975-983.

Sun, G.L., Ni, Y., Dayley, T. 2008. Molecular phylogeny of *RPB2* gene reveals multiple origin, geographic differentiation of H genome, and the relationship of the Y genome to other genomes in *Elymus* species. Mol. Phylogenet. Evol. 46: 897-907.

Swofford, D.L. 2003. PAUP beta version 4.0b2. Phylogenetic analysis using parsimony-Macintosh version. Sineaur Associates, Sunderland, M.A.

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

Tamura, K., Peterson, D., Stecher, G., Nei, M., Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731-2739.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmouguin, F., Higgins, D.G. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.

Tzvelev, N.N. 1989. The system of grasses (Poaceae) and their evolution. The Botanical Review 55(3): 141-203.

von Bothmer, R., Flink, J., Landström, T. 1986. Meiosis in interspecific *Hordeum* hybrids. I. Diploid combinations. Can. J. Genet. Cytol. 28: 525-535.

Wakeley,J., Hey, J. 1997. Estimating Ancestral Population Parameters. Genetics 145: 847- 855. Wang et al. 1999

Wang, R.-C., von Bothmer, R., Dvorak, J., Fedak, G., Linde-Laursen, I., & Muramatsu, M. (1995). Genome symbols in the Triticeae (Poaceae). In R.-C. Wang, K. Jensen,& C. Jaussi (Eds.), Proc. Second International Triticeae Symposium (pp. 29- 34). Logan, Utah: Forage and Range Laboratory, U.S.D.A.-A.R.S.

Wang, H., Sun, D., Sun, G. 2011. Molecular phylogeny of diploid *Hordeum* species and incongruence between chloroplast and nuclear datasets. www.nrcresearchpress.com/gen. DOI: 10.1139/G11-063.

Wang, R.R.-C., Zhang, J.-Y., Lee, B.S., Jensen, K.B., Kishii, M., Tsujimoto, H. 2005. Variations in abundance of 2 repetitive sequences in *Leymus* and *Psathyrostachys* species. Genome 49: 511-519.

Watterson, G.A. 1975. On the number of segregation sites in genetical models without recombination. Theor. Popul. Biol. 7: 256-276

Zhang et al., 2009. Different maternal genome donor to *Kengyilia* species inferred from chloroplast *trnL-F* sequences. Biologia Plantarum 53(4): 759-763.