Gene 533 (2014) 142–148

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Molecular evolution and nucleotide diversity of nuclear plastid phosphoglycerate kinase (PGK) gene in Triticeae (Poaceae)

Shawn Adderley, Genlou Sun *

Biology Department, Saint Mary's University, Halifax, NS, Canada, B3H 3C3

ARTICLE INFO

Article history:
Accepted 26 September 2013
Available online 11 October 2013

Keywords:
Nucleotide diversity
Plastid phosphoglycerate kinase (PGK) gene
Evolution
Triticeae

ABSTRACT

Levels of nucleotide divergence provide key evidence in the evolution of polyploids. The nucleotide diversity of 226 sequences of *pgk1* gene in Triticeae species was characterized. Phylogenetic analyses based on the *pgk1* gene 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. Sequences from the Ns genome represented the highest nucleotide diversity values for both polyploid and diploid species with $\pi = 0.03343$ and $\theta = 0.03536$ for polyploid Ns genome sequences and $\pi = 0.03886$ and $\theta = 0.03886$ for diploid *Psathyrostachys* sequences, while *Triticum urartu* represented the lowest diversity among diploid species at $\pi = 0.0011$ and $\theta = 0.0011$. Nucleotide variation of diploid *Aegilops speltoides* ($\pi = 0.2441$, presumed the B genome donor of *Triticum* species) is five times higher than that ($\pi = 0.00483$) of B genome in polyploid species. Significant negative Tajima's *D* values for the St, A^u, and D genomes along with high rates of polymorphisms and low sequence diversity were observed. Origins of the A^u, B, and D genomes were linked to *T. urartu*, *A. speltoides*, and *A. tauschii*, respectively. Putative St genome donor was *Pseudoroegneria*, while Ns and P donors were *Psathyrostachys* and *Agropyron*. H genome diploid donor is *Hordeum*.

1. Introduction

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., 2002a). 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; Blattner, 2009; Escobar et al., 2011; Huang et al., 2002a, 2002b; Kellogg, 1996; Petersen and Seberg, 1996). 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).

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 et al., 2002a). 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 urarta* (AA) and proposed BB genome donor *Aegilops speltoides* (current designation is SS) (Gu et al., 2004; Huang et al., 2002a, 2002b; Kilian et al., 2007). Within the genera *Elymus* (StStHHYY) and *Kengyilia* (StStYYPP), common origin for each genera's St genus is understood to be the diploid donor *Pseudoroegneria* (Fan et al., 2012; Sun and Komatsuda, 2010), while the H genome of *Elymus* is believed to derive from an unknown diploid *Hordeum* (Sun et al., 2008), and the P genome of *Kengyilia* from diploid *Agropyron* (Fan et al., 2009). The Y genome origin of each genus as of yet remains unknown (Fan et al., 2009; Sun et al., 2008). In allotetraploid *Leymus* (NsNsXmXm), its Ns genome origin has been traced to *Psathyrostachys* (Sun et al., 1995; 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.

Polyploidization provides a pool of duplicate genes as building blocks for potential evolutionary advancements (Fan et al., 2009, 2012). 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,

---

Abbreviations: PGK, plastid phosphoglycerate kinase; MP, maximum-parsimony; NJ, neighbor-joining; TBR, tree bisection–reconnection.

* Corresponding author. Tel.: +1 902 420 5642; fax: +1 902 420 5046.
E-mail address: genlou.sun@smu.ca (G. Sun).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.09.103

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 (Gu et al., 2004; Han et al., 2003). 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 (Gu et al., 2004; Han et al., 2003). 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.

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 during glycolysis in the Calvin Cycle by catalyzing the conversion of 1,3-diphosphoglycerate to 3-phosphogycerate (Fan et al., 2012; Huang et al., 2002a). The *pgk1* gene contains ~4 introns and 5 exons and is thought to have originated via duplication of a prokaryotic gene (Huang et al., 2002a).

In the present study, we investigated the evolution of the *pgk1* gene in the tribe Triticeae and compare the nucleotides of the *pgk1* gene among different genomes, and re-examined the current suggestions on the origin of polyploid species within the Triticeae and their relationships.

2. Materials and methods

2.1. Taxon sampling

Sixty-one 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* sequences were downloaded from GenBank (Fan et al., 2009, 2012, 2013; Huang et al., 2002a; Kilian et al., 2007), 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 sequences, 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 ClustalW (Larkin et al., 2007). 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 that the sequences that would provide the best represented tree were selected. The representative sequences from *Triticum*/*Aegilops* and "Others" groups were combined into a complete data set containing 96 sequences.

Multiple alignments of the 96 sequences were completed using ClustalW via default settings. Manual edits were conducted following alignment to minimize gaps (Thompson et al., 1997) (supplement material). 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

Table 1
Triticeae species used for phylogenetic analyses, genomic designations, and GenBank 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		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^uBD	JQ32706.1, JQ327063.1, JQ327125.1
Triticum carthlicum	A^uB	JQ327065.1, JQ327066.1, JQ327073.1
Triticum compactum	A^uBD	JQ327070.1, JQ327071.1, JQ327072.1
Triticum durum	A^uB	JQ327116.1, JQ327115.1
Triticum monococcum subsp. *aegilopoides*	A^m	DQ290665.1
Triticum sphaerococcum	A^uBD	JQ327094.1, JQ327117.7, JQ327118.1JQ327119.1
Triticum timopheevii subsp. *armeniacum*	A^uG	DQ364906.1, DQ364895.1
Triticum turgidum subsp. *dicoccum*	A^uB	DQ290692.1, DQ290736.1
Triticum urartu	A^u	DQ290659.1

All genomic designations as per Wang et al., 1995.