SHORT COMMUNICATION

Sequence variation at the three chloroplast loci (*matK*, *rbcL*, *trnH-psbA*) in the Triticeae tribe (Poaceae): comments on the relationships and utility in DNA barcoding of selected species

Wojciech Bieniek · Marta Mizianty · Marek Szklarczyk

Abstract This study presents a phylogenetic analysis of the selected representatives of the Triticeae tribe and the assessment of possibilities of species identification on the basis of sequence variation in the three regions from the chloroplast DNA (*matK*, *rbcL* and *trnH-psbA*). The analysis included diploid and polyploid species of the following genomic compositions: E^b, E^e, **H, I, Ns, St, Ta, Xu, NsXm, XoXr, XuXu**, E^bE^e, $E^bE^bE^e$, E^bE^eSt, **StH, StStH**. The obtained *matK* and *rbcL* sequences revealed three well-supported clades: (1) *Elymus, Lopophyrum, Pseudoroegneria, Thinopyrum* (genomes E^b, E^e, **St, StH**, E^bE^e, $E^bE^bE^e$, E^bE^eSt, **StStH**); (2) *Leymus, Hordelymus, Psathyrostachys* (genomes **Ns, NsXm, XoXr**); (3) *Hordeum* (genomes **H, I, Xu**). This results support the view that the analyzed *Elymus* species (*E. caninus, E. repens*) acquired the *Pseudoroegneria* **St** genome maternally. They also indicate that *Hordelymus europaeus* carries the **Ns** genome from *Psathyrostachys* and point at the maternal origin of this component. Moreover, the presented results confirm the contribution of *Psathyrostachys* in the origin of *Leymus* as well as indicate the maternal character of this input. Differentiation observed within the *Hordeum* clade aligns with the alleged genome composition of the analyzed species. The DNA barcoding potential of the analyzed sequences was low. Nevertheless, several informative polymorphisms were found like e.g. the *matK* substitution specific for *Thinopyrum bessarabicum*, the *trnH-psbA* substitution differentiating *Leymus arenarius* from *Psathyrostachys juncea* and the group of species-specific deletions within *trnH-psbA*. In the DNA barcoding context we have also demonstrated negative importance of the high frequency cpDNA inversion leading to intraspecific variation.

Introduction

Triticeae, one of the 41 tribes belonging to the Poaceae family, is economically very important. The most significant representatives of the Triticeae are crops: wheat (*Triticum aestivum*), rye (*Secale cereale*), barley (*Hordeum vulgare*) and also many forage species. On the other hand, some Triticeae are harmful weeds (e.g. *Elymus repens*). Such an important taxonomic group is very intensely and universally examined. The number of recognized Triticeae taxa depends on morphological, ecological, cytological, karyological or genetical taxonomic criteria and there has been no consensus on this issue so far. The ease of interspecific or even intergeneric hybrid formation within the Triticeae results in a high number of the described taxa. Among these taxa both diploids and polyploids (either auto- or allopolyploids) can be found. The diploids and polyploids are linked with genetic relationships that form complex reticulate evolution patterns. Moreover, in the Triticeae, genomes of unknown origin were found, distinct from other genomes in the tribe. All issues mentioned above complicate the

W. Bieniek (✉) · M. Mizianty
Department of Vascular Plant Systematics and Phytogeography, W. Szafer Institute of Botany, Polish Academy of Sciences, ul. Lubicz 46, 31-512 Kraków, Poland
e-mail: w.bieniek@botany.pl

M. Szklarczyk
Unit of Genetics, Plant Breeding and Seed Science, Institute of Plant Biology and Biotechnology, Agricultural University of Krakow, al. 29 Listopada 54, 31-425 Kraków, Poland

Springer

phylogenetic studies as well as species defining. Thus the number of the Triticeae genera varies from 18 (based on morphology; Clayton and Renvoize 1986) to 38 (based on the genomic criterion; Löve 1984) depending on the adopted taxonomic criteria. Currently, the widely applied generic classification, based mainly on the genome composition (Dewey 1984; Löve 1984), does not correspond to morphological divisions and causes delimitation of monotypic genera. However, it is more precise in reflecting the phylogenetic relationships and the evolutionary history.

The chloroplast DNA (cpDNA) is a rich resource of phylogenetic information for the land plants. As the chloroplast genome is mostly maternally inherited among angiosperms, its sequence can help to identify maternal forms of the polyploid species. Thus the cpDNA sequence variation has been used to track maternal genealogy of the allopolyploid angiosperm taxa for circa 30 years. The cpDNA analysis has changed over time: starting from RFLP (restriction fragment length polymorphism) analysis of the whole chloroplast genome, through PCR–RFLP of selected loci, then sequencing of selected loci, to the recently applied whole chloroplast genome sequencing. The majority of phylogenetic and populational studies conducted on the chloroplast DNA so far has been based on the analysis of one or a few relatively short (1–2 kb) genomic fragments (e.g. Shaw et al. 2007).

A relatively new trend in exploitation of chloroplast sequences is DNA barcoding—the use of short DNA sequences (from a few dozens to a few hundreds bp) for species identification. For animals, DNA barcoding was developed on the basis of species specificity in the sequence of the mitochondrial CO1 gene (Hebert et al. 2003). Early studies on the CO1 gene helped to define the main requirements for an optimal DNA barcoding marker: length up to 500–700 bp (ease of amplification and sequencing), high interspecific sequence variation for a wide spectrum of organisms (e.g. for all land plants), low intraspecific sequence variation, ease of amplification with universal primers and large copy number in cells. These features allow species determination using standard procedures of molecular genetics with only minute tissue amounts. Unfortunately, for the land plants no DNA barcoding standard has been discovered so far which would meet all above criteria as ideally as CO1 does for animals. This is related to the fact that plant mitochondrial genes appeared to be unsuitable markers for DNA barcoding (Kress et al. 2005). Moreover, there is probably no single marker in the plant genome that exhibits all features required for DNA barcoding (Hollingsworth et al. 2011). Therefore, two chloroplast loci that meet the requirements for DNA barcoding only partially were chosen for species identification among the land plants, these are: the fast-evolving *matK* (about 700 bp) and the slowly evolving *rbcL* (about 500 bp) (CBOL Plant Working Group 2009). The full-length sequences of *matK* and *rbcL* (both of about 1.5 kb) were successfully applied to analyze the interspecific phylogenetic relationships among the land plants, including members of the Poaceae family (e.g. Hilu et al. 1999).

It was also suggested that other species-identifying markers should be used if resolution of *matK* and *rbcL* appeared to be insufficient. One of the most commonly proposed additional loci is the chloroplast *trnH-psbA* intergenic spacer (e.g. Kress et al. 2005). The aim of this study was to estimate the value of the phylogenetic information provided by the mentioned DNA barcoding markers (*matK*, *rbcL*) and by the candidate marker (*trnH-psbA*) as well as to evaluate the species-identification effectiveness of these plastid sequences. The di-, tetra- and hexaploid Triticeae taxa occurring in Eurasia have been used as a model group.

Materials and methods

Plant material

A total of 28 Triticeae accessions and 1 outgroup accession [*Brachypodium distachyon* (L.) Beauv.] were analyzed in this study. Eight polyoploid Triticeae species, represented by 14 specimens were analyzed: *Elymus caninus* L., *Elymus repens* (L.) Gould, *Hordelymus europaeus* (L.) Harz, *Hordeum murinum* L., *Leymus arenarius* (L.) Hochst., *Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey, *Thinopyrum junceiforme* (Á. Löve & D. Löve) Á. Löve, *Thinopyrum junceum* (L.) Á. Löve. These accessions came from Central Europe as well as from other areas such as Afghanistan, China and Greece. The nucleotide sequences obtained for the alloploid species were compared to the nucleotide sequences obtained from the eight diploid Triticeae species (14 specimens): *Hordeum bogdani* Wilensky, *Hordeum glaucum* Steud., *Hordeum bulbosum* L., *Lophopyrum elongatum* (Host) Á. Löve, *Psathyrostachys juncea* (Fisch.) Nevski, *Pseudoroegneria strigosa* (M. Bieb.) Á. Löve, *Taeniatherum caput-medusae* (L.) Nevskii, *Thinopyrum bessarabicum* (Savul. & Rayss) Á. Löve. These eight species represent different potential parental genomes of the polyploid species. 17 specimens were obtained from germplasm repositories (American National Plant Germplasm System, Pullman, Washington, USA; Nordic Gene Bank, Alnarp, Sweden), 9 were collected by the authors of the study from the natural habitats, in case of 3 accessions the nucleotide sequences were obtained from GenBank (Table 1).

Table 1 The list of plant specimens and sequence accessions included in the study

Species	Genome/s	Origin	Plant accession no.	Sequence accession no.		
				matK	*rbcL*	*trnH-psbA*
Elymus caninus 1	**StH**	China	PI 595129	KF277159	KF728574	KF872133
Elymus caninus 2	**StH**	Poland	ECA2[b]	KF277160	KF728575	KF872134
Elymus repens	**StStH**	Poland	ERE3[b]	KF277158	KF728576	KF872135
Hordelymus europaeus 1	**XoXr**	Sweden	NGB15820	KF277156	KF728585	KF872127
Hordelymus europaeus 2	**XoXr**	Poland	HEU3[b]	KF277157	KF728586	KF872128
Hordeum bogdani 1	**H**	Iran	NGB90046	KF277172	KF728591	KF872136
Hordeum bogdani 2	**H**	China	NGB90049	KF277173	KF728592	KF872137
Hordeum bulbosum	**I**	Italy	H3878[a]	AB078095.1	AY137454.1	EU118529.1
Hordeum glaucum	**Xu**	Tunisia	NGB6872	KF277174	KF728593	KF872138
Hordeum murinum 1	**XuXu**	Germany	HMU1[b]	KF277176	KF728595	KF872139
Hordeum murinum 2	**XuXu**	Poland	HMU2[b]	KF277175	KF728594	KF872140
Leymus arenarius 1	**NsXm**	Lithuania	PI 531800	KF277154	KF728589	KF872129
Leymus arenarius 2	**NsXm**	Poland	LAR3[b]	KF277155	KF728590	KF872130
Lophopyrum elongatum 1	**E**[e]	France	PI 469212	KF277166	KF728581	KF872141
Lophopyrum elongatum 2	**E**[e]	China	PI 595139	KF277167	KF728582	KF872142
Psathyrostachys juncea 1	**Ns**	Turkey	PI 206684	KF277152	KF728587	KF872131
Psathyrostachys juncea 2	**Ns**	Russian Federation	PI 565059	KF277153	KF728588	KF872132
Pseudoroegneria strigosa 1	**St**	Estonia	PI 531752[a]	HQ652698.1	HQ652778.1	HQ652797.1
Pseudoroegneria strigosa 2	**St**	China	PI 499638	KF277177	KF728596	KF872143
Taeniatherum caput-medusae 1	**Ta**	Spain	NGB7426	KF277170	KF728583	KF872151
Taeniatherum caput-medusae 2	**Ta**	Afghanistan	PI 220591	KF277171	KF728584	KF872152
Thinopyrum bessarabicum 1	**E**[b]	Ukraine	PI 531711	KF277169	KF728580	KF872144
Thinopyrum bessarabicum 2	**E**[b]	Ukraine	W6 10232	KF277168	KF728579	KF872145
Thinopyrum intermedium 1	**E**[b] **E**[e] **St**	Afghanistan	PI 220498	KF277163	KF728573	KF872148
Thinopyrum intermedium 2	**E**[b] **E**[e] **St**	Romania	EHI2[b]	KF277164	KF728571	KF872149
Thinopyrum intermedium 3	**E**[b] **E**[e] **St**	Poland	EHI3[b]	KF277165	KF728572	KF872150
Thinopyrum junceiforme	**E**[b] **E**[e]	Poland	EFA2[b]	KF277162	KF728578	KF872147
Thinopyrum junceum	**E**[b] **E**[b] **E**[e]	Greece	PI 531729	KF277161	KF728577	KF872146
Brachypodium distachyon	Outgroup	Not specified	Bd21[a]	EU325680.1	EU325680.1	EU325680.1

Genome designations are given according to Wang et al. (1995)

Specimens with the PI/W6-containing numbers were provided by the American National Plant Germplasm System (Pullman, Washington, USA). Specimens with the NGB-containing numbers were provided by the Nordic Genetic Resource Center (NordGen) (Alnarp, Sweden)

[a] Nucleotide sequences retrieved from GenBank

[b] Specimens collected and analyzed by the authors

DNA sequencing

Amplification of DNA fragments was conducted without DNA extraction using the Phire Plant Direct PCR Kit (Thermo Scientific) according to the dilution protocol. The PCR was carried out in a total volume of 20 µl. Both reagents' concentrations and thermal cycling parameters were applied according to the manufacturer's protocol. The primer sequences were taken from Yu et al. (2011) (*matK*, Tm = 48 °C), Erickson (www.barcoding.si.edu/pdf/informationonbarcodeloci.pdf, *rbcL*, Tm = 55 °C) and Shaw et al. (2005) (*trnH-psbA*, Tm = 56 °C). The PCR products were purified using the *GeneJet PCR Purification Kit* (Thermo Scientific) according to the manufacturer's instruction. The purified PCR products were sequenced in both directions with the primers used for DNA amplification. Sequencing was performed either using custom sequencing services or in the authors' laboratory with the use of the GenomeLab DTCS Quick Start Kit (Beckman-Coulter) and the GenomeLab GeXP Genetic Analysis System (Beckman-Coulter). The resulting DNA sequences were checked and aligned with the BioEdit software (Hall 1999). Multiple sequence alignments are available from the corresponding author on request.

Data analysis

The analyzed dataset consisted of 87 nucleotide sequences: 78 resulted from the DNA sequencing experiments performed during this study and 9 were retrieved from GenBank (Table 1). *Brachypodium distachyon* (L.) Beauv was used as an outgroup in the phylogenetic analysis. The BioEdit software was used to perform multiple sequence alignments. Prior to the phylogenetic analysis all indels, poly(T) tract of variable length and the inversion found within the *trnH-psbA* spacer were manually removed from the alignment.

With the MEGA 5 software (Tamura et al. 2011): (1) the optimal substitution model was determined for each locus (*matK*: Tamura 3-parameter + Gamma distribution, *rbcL*: Tamura 3-parameter + Gamma distribution, *trnH-psbA*: Tamura 3-parameter), (2) the minimum-evolution trees were constructed and (3) the reliability of branching was tested using the bootstrap method with 1,000 replications.

Results

Twenty-nine nucleotide sequences were obtained for each of the analyzed loci—*matK*, *rbcL* and *trnH-psbA*. The lengths of the sequences were: 741 nt for *matK*, 553 nt for *rbcL*, and from 618 (*Leymus*) to 630 nt (*Elymus, Lophopyrum, Pseudoroegneria*) for *trnH-psbA*. The average GC contents were: 34.1 % for *matK*, 43.0 % for *rbcL*, 36.6 % for *trnH-psbA*. The estimated Transition/Transversion biases (R) were: 1.07, 1.71 and 1.08 for *matK*, *rbcL* and *trnH-psbA*, respectively. In *matK* and *rbcL* the only mutations found were the substitution mutations, *matK* contained 34 (4.59 %) variable sites while *rbcLa*—15 (2.71 %) (Table 2).

The length differences among the *trnH-psbA* fragments were caused by indel mutations and mononucleotide repeat (poly(T) tract) of variable length (Fig. 1). All of the three observed deletions were species specific: 9 nt indel at position 94–102 in *T. caput-medusae*, 4 nt deletion at positions 96–99 in *P. juncea* and 1 nt deletion at position 109 in *T. bessarabicum*. Moreover, within *trnH-psbA* the inversion mutation of 6 bp (position 117–122) was discovered. The same orientation of this sequence was observed for *E. caninus, H. bulbosum, H. glaucum, H. murinum, L. arenarius, L. elongatum, P. juncea, T. caput-medusae, T. bessarabicum, T. junceum* and *B. distachyon*. The opposed (inversed) orientation was found in *E. repens* and *T. junceiforme*. Interestingly, in the remaining species—*H. europaeus, H. bogdani, P. strogosa* and *T. intermedium* both orientations were noted among the analyzed specimens. To perform the minimum-evolution analysis above mentioned indels, inversion and

poly(T) tract were removed from the multiple alignment. Thus the *trnH-psbA* fragments of 597 nucleotides including 9 (1.51 %) variable sites were used for the construction of phylogenetic trees (Table 2).

For each of the three analyzed loci minimum-evolution phylogenetic trees were constructed (Fig. 2). Two well-supported clades were present at each of the three resulting trees: (1) the **E**$^\text{b}$, **E**$^\text{e}$ and **St** genome clade, genetically almost uniform, containing genera *Elymus, Lophopyrum, Pseudoroegneria* and *Thinopyrum*; (2) the **H**, **I** and **Xu** clade, genetically differentiated, containing all of the analyzed *Hordeum* taxa. Another significant clade was formed by the **Ns** genome-containing taxa (*Leymus, Psathyrostachys*) that share the same sequences of *matK* and *rbcL*, but differ at the *trnH-psbA* locus. The specimens of *H. europaeus* (genome **XoXr**) showed similarity with the **Ns** genome-carrying taxa at *matK* and *trnH-psbA*. *Taeniatherum caput-medusae* (**Ta** genome) formed a separate clade that was close to the **St/E** genome group at *matK* and *rbcL*, but the *trnH-psbA* sequences of the **Ta** and **St/E** representatives were identical.

Discussion

Phylogenetic relationships

The major clades previously defined within the Triticeae on the basis of DNA sequences are the following: (1) *Psathyrostachys,* (2) *Hordeum,* (3) *Pseudoroegneria,* (4) *Aegilops-Triticum-Secale-Taeniatherum* and (5) *Eremopyrum-Agropyron-Australopyrum* (Adderley and Sun 2014; Escobar et al. 2011; Petersen and Seberg 1997). All of the here analyzed loci have revealed the *Hordeum* and *Pseudoroegneria* clades. However, only *matK* is variable enough to distinguish four well bootstrap-supported clades: *Psathyrostachys, Hordeum, Pseudoroegneria* and *Taeniatherum*. Moreover, the minimum-evolution dendrogram obtained on the basis of the *matK* sequences (Fig. 2a) revealed a close relationship between the *Pseudoroegneria* and the *Taeniatherum* clades, as well as a clear distinction of the *Psathyrostachys* and the *Hordeum* clades, which is consistent with the previous molecular studies (Fan et al. 2013; Escobar et al. 2011; Petersen and Seberg 1997). Thus, out of the three analyzed loci, the *matK* reflects intergeneric phylogenetic relationships of the Triticeae in the most complete manner.

Elymus s.l. For all three analyzed loci the largest and predominantly homogeneous clade was formed by representatives of the **E** and the **St** genomes. This clade contained diploids [*P. strigosa* (**St**), *L. elongatum* (**E**$^\text{e}$) and *T. bessarabicum* (**E**$^\text{b}$)], tetraploids [*E. caninus* (**StH**), *T. junceiforme* (**E**$^\text{b}$**E**$^\text{e}$)] as well as hexaploids [*E. repens*

Table 2 Substitutions in the DNA sequences of *matK*, *rbcL* and *trnH-psbA* among analyzed Triticeae specimens

Locus/nucleotide position Specimen	matK																												
	105	107	108	110	111	112	117	138	141	142	182	226	312	361	383	385	395	398	444	445	467	507	512	537	578	617	623	625	651
Elymus caninus 1	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Elymus caninus 2	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Elymus repens	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Hordelymus europaeus 1	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	A	G	C	A	C	C	C	C
Hordelymus europaeus 2	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	A	G	C	A	C	C	C	C
Hordeum bogdani 1	A	T	T	T	T	A	C	T	A	A	C	C	T	A	C	G	G	A	T	T	A	A	A	C	C	C	T	T	C
Hordeum bogdani 2	A	T	T	T	T	T	C	G	A	A	C	C	T	A	C	G	G	G	T	T	A	A	A	C	C	C	T	T	C
Hordeum bulbosum	T	T	T	T	T	T	C	G	A	A	C	A	T	A	A	G	G	A	T	T	A	A	A	C	C	C	T	T	C
Hordeum glaucum	A	T	T	T	T	T	C	T	A	A	C	C	T	A	C	G	G	G	T	T	A	A	A	C	C	C	T	T	C
Hordeum murinum 1	A	T	T	T	T	T	C	T	A	A	C	C	T	A	C	G	G	G	T	T	A	A	A	C	C	C	T	T	C
Hordeum murinum 2	A	T	T	T	T	T	C	T	A	A	C	C	T	A	C	G	G	G	T	T	A	A	A	C	C	C	T	T	C
Leymus arenarius 1	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	C	G	C	A	T	T	C	G
Leymus arenarius 2	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	C	G	C	A	T	T	C	G
Lophopyrum elongatum 1	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Lophopyrum elongatum 2	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Psathyrostachys juncea 1	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	C	G	C	A	T	T	C	G
Psathyrostachys juncea 2	A	T	T	T	C	T	A	T	C	G	C	C	C	A	C	C	G	G	T	G	G	C	G	C	A	T	T	C	G
Pseudoroegneria strigosa 1	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	A	A	A	C	T	T	C
Pseudoroegneria strigosa 2	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	A	A	A	C	T	T	C
Taeniatherum caput-medusae 1	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	T	G	A	A	G	T	A	C	T	T	C
Taeniatherum caput-medusae 2	A	C	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	T	G	A	A	G	T	A	C	T	T	C
Thinopyrum bessarabicum 1	A	T	T	A	T	T	C	T	C	G	T	C	C	T	C	C	G	G	C	G	A	A	A	A	A	C	T	T	C

Table 2 continued

Locus/nucleotide position Specimen	matK																												
	105	107	108	110	111	112	117	138	141	142	182	226	312	361	383	385	395	398	444	445	467	507	512	537	578	617	623	625	651
Thinopyrum bessarabicum 2	A	T	T	A	T	T	C	T	C	G	T	C	C	T	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Thinopyrum intermedium 1	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	A	G	C	G	A	A	G	A	A	C	T	T	C
Thinopyrum intermedium 2	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Thinopyrum intermedium 3	A	T	T	T	T	T	C	T	C	G	T	C	C	A	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Thinopyrum junceiforme	A	T	T	T	T	T	C	T	C	G	T	C	C	T	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C
Thinopyrum junceum	A	T	C	T	T	T	C	T	C	G	T	C	C	T	C	C	G	G	C	G	A	A	G	A	A	C	T	T	C

Locus/nucleotide position Specimen	matK					rbcL																					trnH-psbA								
	668	698	705	708	710	14	112	117	138	141	142	182	226	245	256	257	258	259	264	265	301	337	352	355	398	401	7	90	167	235	259	328	337	556	578
Elymus caninus 1	G	G	T	G	T	C	G	G	C	C	A	A	C	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	G	G	C	G	G	T	A
Elymus caninus 2	G	G	T	G	T	C	G	G	C	C	A	A	C	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Elymus repens	G	G	T	G	T	C	G	A	C	C	A	A	C	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Hordelymus europaeus 1	G	A	T	G	C	C	G	G	C	G	G	A	G	C	G	G	A	G	G	G	A	G	T	A	A	G	T	A	T	G	C	G	G	C	A
Hordelymus europaeus 2	G	A	T	G	C	C	G	G	C	G	G	A	G	C	G	G	A	G	G	G	A	G	T	A	A	G	T	G	T	G	C	G	G	C	A
Hordeum bogdani 1	G	G	T	G	T	G	G	G	G	C	A	A	C	G	C	A	A	C	G	G	G	G	G	C	C	A	T	G	G	A	T	A	G	C	C
Hordeum bogdani 2	G	A	T	G	T	G	A	G	G	C	A	A	C	G	C	A	A	C	G	G	G	G	G	C	C	A	T	G	T	A	T	A	G	C	C
Hordeum bulbosum	G	G	T	G	T	G	G	G	G	C	A	G	C	G	C	A	G	C	G	G	G	G	G	C	C	A	T	A	A	A	T	A	G	C	A
Hordeum glaucum	G	G	T	G	T	A	G	G	C	A	A	A	C	C	C	A	A	C	G	G	G	G	G	C	C	A	T	G	T	A	T	A	G	C	A
Hordeum murinum 1	G	G	T	G	T	A	G	G	C	A	A	A	C	C	C	A	A	C	G	G	G	G	G	C	C	A	T	G	T	A	T	A	G	C	A
Hordeum murinum 2	G	G	T	G	T	A	G	G	C	A	A	A	C	C	C	A	A	C	G	G	G	G	G	C	C	A	T	G	T	A	T	A	G	C	A
Leymus arenarius 1	A	A	T	G	C	C	C	G	C	C	A	A	A	C	A	A	A	C	G	G	A	G	T	C	C	A	T	G	T	G	C	G	T	T	A
Leymus arenarius 2	A	A	T	G	C	C	C	G	C	C	A	A	A	C	A	A	A	C	G	G	A	G	T	C	C	A	T	G	T	G	C	G	T	T	A
Lophopyrum elongatum 1	G	G	T	G	T	C	G	G	C	C	A	A	C	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Lophopyrum elongatum 2	G	G	T	G	T	C	G	G	C	C	A	A	C	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Psathyrostachys juncea 1	A	A	T	G	C	C	G	G	C	C	A	A	G	C	G	A	G	C	G	G	A	G	T	C	C	A	C	G	G	G	C	G	G	T	A
Psathyrostachys juncea 2	A	A	T	G	C	C	G	G	C	C	A	A	G	C	G	A	G	C	G	G	A	G	T	C	C	A	C	G	G	G	C	G	G	T	A

Table 2 continued

Locus/nucleotide position	matK					rbcL															trnH-psbA								
Specimen	668	698	705	708	710	14	238	245	256	257	258	259	264	265	301	337	352	355	398	401	7	90	167	235	259	328	337	556	578
Pseudoroegneria strigosa 1	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Pseudoroegneria strigosa 2	G	G	T	A	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Taeniatherum caput-medusae 1	G	A	T	G	T	C	G	G	C	A	A	C	G	G	A	A	T	C	C	A	T	G	T	G	C	G	G	T	A
Taeniatherum caput-medusae 2	G	A	T	G	T	C	G	G	C	A	A	C	G	G	A	A	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum bessarabicum 1	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum bessarabicum 2	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum intermedium 1	G	G	G	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum intermedium 2	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum intermedium 3	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum junceiforme	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	T	C	A	T	G	T	G	C	G	G	T	A
Thinopyrum junceum	G	G	T	G	T	C	G	C	C	A	A	C	T	T	A	G	T	C	C	A	T	G	T	G	C	G	G	T	A
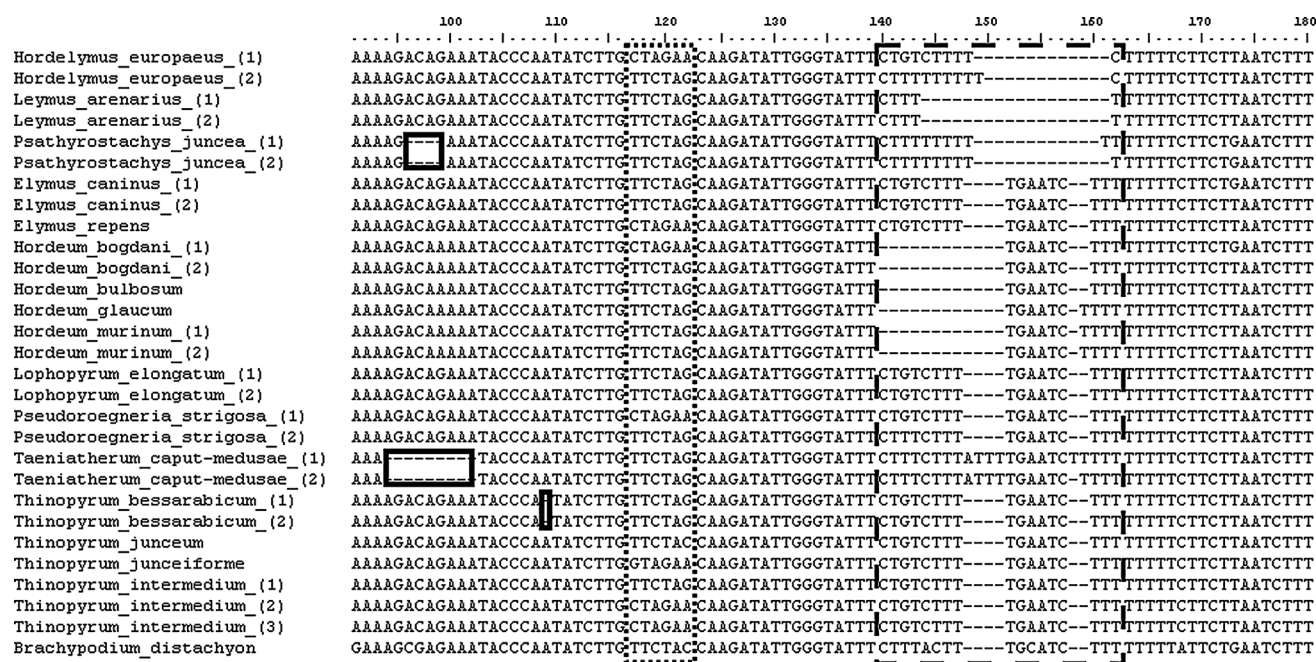
Fig. 1 Multiple alignment of the highly variable part of the *trnH-psbA* region from the analyzed Triticeae species. Marked sequence polymorphisms: indels (*unbroken line*), inversion (*dotted line*) and poly(T) track (*dashed line*)

(**StStH**), *T. intermedium* (**E**b**E**e**St**) and *T. junceum* (**E**b**E**b-**E**e)]. The nucleotide sequences obtained for the specimens from these taxa were almost identical. This group of genome representatives corresponds to the genus *Elymus* sensu lato (Melderis 1980). The previous cpDNA studies on the Triticeae (e.g. Adderley and Sun 2014; Mahelka et al. 2011; Mason-Gamer 2013; Redinbaugh et al. 2000; Yan and Sun 2012) have shown that during the allopolyploid formation the chloroplast genomes were inherited from the parent carrying the **St** (nuclear) genome. Hence, the allopolyploid **St**-containing genera, delimited on the basis of genomic composition, probably share the same (or very closely related) chloroplast genomes that originate from *Pseudoroegneria*. In case of *E. repens* and *E. caninus* our data support this view since their plastotypes were very similar to that of *Pseudoroegneria* and clearly distinct from the species carrying genome **H**. In case of the here analyzed, *Thinopyrum* species such a conclusion is not justified (although it cannot be excluded) since the plastotypes of the diploid species carrying genomes **St**, **E**b and **E**e were hardly distinguishable.

Phylogenetic relationships between the **E**e and **E**b genomes remain unclear. The cytogenetic studies led their authors to opposite conclusions i.e. Wang (1985) treated **E**e and **E**b (marked as **J**E and **J,** respectively) as two forms of the same genome while Jauhar (1988), on the contrary, considered them to be distinct (with designations **J** and **E,** respectively). The nucleotide sequences of the diploid **E**e, **E**b and **St** taxa analyzed in this study are

almost the same. Only one substitution within *matK* differentiates *T. bessarabicum* (**E**b) from the **E**e and **St** diploid taxa. Sequencing of the chloroplast *rpoA* gene showed the same result: *Lophopyrum* and *Pseudoroegneria* shared identical sequences, whereas *T. bessarabicum* slightly diverged (Petersen and Seberg 1997). Our results confirm the very close relationship among the **E**e, **E**b and **St** taxa as well as support the inclusion of *Lophopyrum* in the *Pseudoroegneria* genus (Petersen and Seberg 1997).

Hordelymus-Leymus-Psathyrostachys Generally the obtained data indicate a close relationship between *P. juncea* (**Ns**), *L. arenarius* (**NsXm**) and *H. europaeus* (**XoXr**), with *H. europaeus* being slightly distinct from the two others. *Hordelymus europaeus* represents the monotypic genus of unclear origin. According to Löve (1984) it is an allotetraploid species containing the **H** (*Hordeum*) and **Ta** (*Taeniatherum* Nevski.) genomes. von Bothmer et al. (1994) excluded the presence of the **H** genome in *Hordelymus* and discovered similarity of its genomes to **Ta** and **Ns**. As this similarity was uncertain Wang et al. (1995) proposed to denote temporarily the *Hordelymus* genomes as **XoXr**. The close similarity to the **Ns** genome was later confirmed by Petersen and Seberg (2008). Moreover, the southern and fluorescence in situ hybridization indicated that *H. europaeus* is an autotetraploid carrying only the **Ns** genome (Ellneskog-Staam et al. 2006). The chloroplast non-coding *trnS-psbC* sequences of *Hordelymus* are similar to those of *T. caput-medusae* subsp. *caput-medusae* (**Ta**),
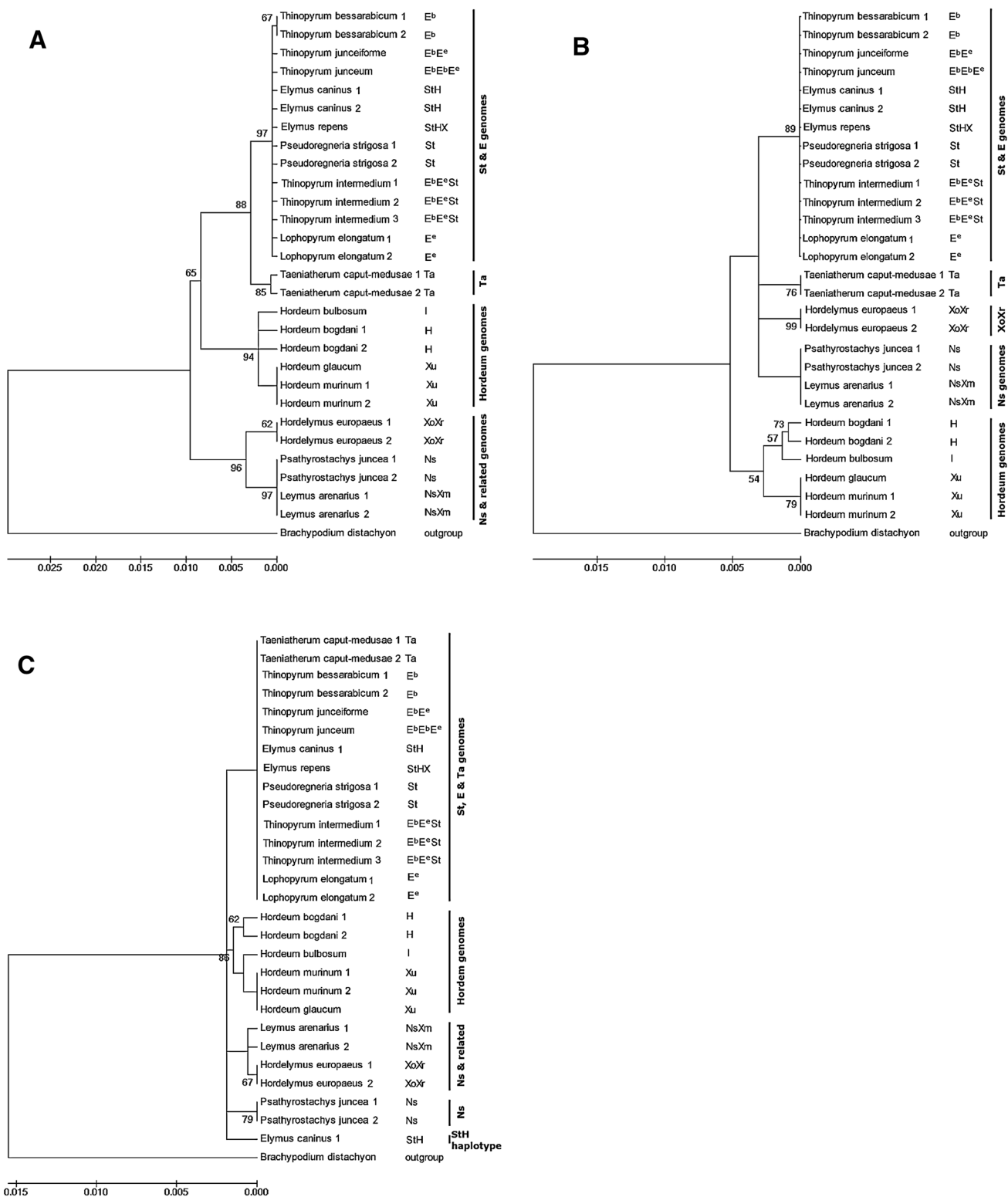
Fig. 2 The minimum-evolution dendrograms obtained for the analyzed Triticeae species on the basis of the three chloroplast sequences: *matK* (**a**), *rbcL* (**b**) and *trnH-psbA* (**c**). *Numbers at nodes* represent the proportion (%) of 1,000 bootstrap replicates supporting each node. Bootstrap values of less than 50 % are not shown. For the GenBank accession numbers refer to Table 1 in "Materials and methods"

P. juncea (**Ns**) and also *H. bogdani* (**H**) (Ni et al. [2011]). The *matK* and *trnH-psbA* sequences analyzed during this study revealed a close affinity between the chloroplast genomes of *Hordelymus* and the **Ns**-carrying taxa (*Psathyrostachys*, *Leymus*). Moreover, none of the examined sequences confirmed the close relationship of *Hordelymus* and *Taeniatherum* (**Ta**) or *Hordeum* (**H**). Therefore, our data indicate that the maternal component of *H. europaeus* likely represented genus *Psathyrostachys* and accordingly they support presence of the **Ns**-related genome in this species.

Psathyrostachys Nevski, a genus with about ten mainly diploid species carrying the **Ns** genome, is also considered as one of the progenitors of the allopolyploid *Leymus* genus (Adderley and Sun [2014]; Anamthawat-Jónsson and Bödvarsdóttir [2001]; Culumber et al. [2011]; Mizianty et al. [1999]). Identity of the second *Leymus* progenitor, responsible for the **Xm** genome component, remains unclear (Guo et al. [2014]). The results of this study showed a close affinity between cpDNA of *Psathyrostachys* and *Leymus*, with only a single substitution mutation in the *trnH-psbA* locus differentiating these genera. Therefore, the presented data fully support the contribution of *Psathyrostachys* in the origin of *Leymus* as well as indicate maternal character of this contribution.

Hordeum. All of the analyzed loci proved *Hordeum* to be an independent group. Moreover, the analysis separated the representatives of the **H** (diploid *H. bogdani*), **I** (diploid or tetraploid *H. bulbosum*) and **Xu** (diploid *H. glaucum*, teraploid *H. murinum*) genomes. This division corresponds to sections *Campestria* Anderson, *Cerealia* Anderson and *Trichostachys* Dumortier respectively (Yen et al. [2005]). Furthermore, the obtained results are consistent with the previous studies of chloroplast or nuclear DNA sequences (e.g. Blattner [2009], Naghavi et al. [2013]). Within each of the analyzed loci the tetraploid *H. murinum* specimens from Central Europe were the same as their probable progenitor—the diploid *H. glaucum*. Therefore, in the present study a close relationship of the **Xu**-genome taxa is evident although some differences between *H. murinum* and *H. glaucum* were found within the *trnL-trnF* locus by Jakob and Blattner ([2010]).

DNA barcoding

DNA barcoding aims at identification of organisms to the level of species. None of the here analyzed loci allows identification of species or even genera carrying genomes **E^b**, **E^e** and **St** (*Elymus*, *Lophopyrum*, *Pseudoroegneria*, *Thinopyrum*). The only exception is one substitution mutation within *matK* that is typical for the diploid *T. bessarabicum* (**E^b**). Such species-specific mutations were already reported for the chloroplast genome. Mahelka et al.

([2011]) separated *L. elongatum* and *T. bessarabicum* from *P. strigosa* on the basis of the *trnL-trnF* sequences. Mason-Gamer ([2013]) found chloroplast mutations that separated *E. caninus* and *E. repens* from *L. elongatum*, *T. bessarabicum* and *P. strigosa*. Ni et al. ([2011]) differentiated *L. elongatum* from *T. bessarabicum* and the **St** clade on the basis of the *trnT-psbC* sequence.

Similarly, the analyzed in this study specimens of *P. juncea* and *L. arenarius* showed identity within the *matK* and *rbcL* sequences. Nevertheless, some differences between genera *Psathyrostachys* and *Leymus* were found in other cpDNA loci—e.g. *trnH-psbA* and *rps16-trnK* (Culumber et al. [2011]).

The analyzed markers can be used to identify representatives of the various *Hordeum* genomes, i.e. **I** (*H. bulbosum*), **H** (*H. bogdani*) and **Xu** (*H. murinum*, *H. glaucum*). None of these markers allowed distinguishing between diploid *H. glaucum* and tetraploid *H. murinum* within the **Xu** group, although earlier mutations differentiating these taxa were found at the *trnL-trnF* locus (Jakob and Blattner [2010]). Both *Hordelymus* and *Taeniatherum* exhibited unique polymorphisms within the *matK* and *rbcL* loci, while with respect to the *trnH-psbA* sequence, *Hordelymus* remained independent from other taxa and *Taeniatherum* was indistinguishable within the **E-St** clade.

The *trnH-psbA* locus in DNA barcoding of Triticeae

The density of substitutions identified within the *trnH-psbA* spacer was the lowest among the three analyzed loci. The major contribution of the *trnH-psbA* analysis to the here-reported species identification was separation of *L. arenarius* from *P. juncea* on the basis of a single substitution mutation. It is also worthwhile to mention the species-specific deletions found in the *trnH-psbA* sequence of *P. juncea*, *T. caput-medusae* and *T. bessarabicum*. Although these polymorphisms were removed from the alignment, as routinely they are not used in phylogenetic analysis—they can serve as very useful diagnostic tools in the DNA barcoding context. Another interesting feature of the examined *trnH-psbA* sequence was the presence of a minute (6 bp) inversion mutation. This inversion is located in between the inverted repeats of approx. 20 bp and shows intra-specific variation in five of the analyzed species: *H. europaeus*, *H. bogdani*, *P. strigosa*, *T. junceiforme* and *T. intermedium*. According to Kim and Lee ([2005]) such small inversions are very common in the chloroplast genomes of land plants and likely result from intra-molecular recombination of the surrounding inverted sequences. These authors also reported that within *Jasminum elegans* both orientations of such inversion were present. Another inversion of this type was found by Whitlock et al. ([2010]) at the *trnH-psbA* spacer in Gentianaceae. Our results show plasticity of a similar polymorphism

across the whole range of taxa indicating that cpDNA sequence data must be carefully inspected for the presence of small inversions to avoid species misidentification.

References

Adderley S, Sun G (2014) Molecular evolution and nucleotide diversity of nuclear plastid phosphoglycerate kinase (PGK) gene in Triticeae (Poaceae). Gene 533:142–148

Anamthawat-Jónsson K, Bödvarsdóttir SK (2001) Genomic and genetic relationships among species of *Leymus* (Poaceae: Triticeae) inferred from 18S-26S ribosomal genes. Amer J Bot 88:553–559

Blattner FR (2009) Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). Breed Sci 59:471–480

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106:12794–12797

Clayton WD, Renvoize SA (1986) Genera graminum: grasses of the world. H.M.S.O, London

Culumber CM, Larson SR, Jensen KB, Jones TA (2011) Genetic structure of Eurasian and North American *Leymus* (Triticeae) wildryes assessed by chloroplast DNA sequences and AFLP profiles. Pl Syst Evol 294:207–225

Dewey DR (1984) The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Publ. Corp, New York, pp 209–279

Ellneskog-Staam P, Taketa S, Salomon B, Anamthawat-Jónsson K, von Bothmer R (2006) Identifying the genome of wood barley *Hordelymus europaeus* (Poaceae: Triticeae). Hereditas 143:103–112

Escobar JS, Scornavacca C, Cenci A, Guilhaumon C, Santoni S, Douzery EJP, Ranwez V, Glémin S, David J (2011) Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). BMC Evol Biol 11:181

Fan X, Sha LN, Yu SB, Wu DD, Chen XH, Zhuo XF, Zhang HQ, Kang HY, Wang Y, Zheng YL, Zhou YH (2013) Phylogenetic reconstruction and diversification of the Triticeae (Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene data. Biochem Syst Ecol 50:346–360

Guo GY, Yang RW, Ding CB, Fan X, Zhang L, Zhou YH (2014) Phylogenetic relationships among *Leymus* and related diploid genera (Triticeae: Poaceae) based on chloroplast *trnQ–rps16* sequences. Nordic J Bot. doi:10.1111/j.1756-1051.2013.00233.x

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc Roy Soc London Ser B Biol Sci 270:313–321

Hilu KW, Alice LA, Liang H (1999) Phylogeny of Poaceae inferred from *matK* sequences. Ann Missouri Bot Gard 86:835–851

Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS One 6:e19254. doi:10.1371/journal.pone.0019254

Jakob SS, Blattner FR (2010) Two extinct diploid progenitors were involved in allopolyploid formation in the *Hordeum murinum* (Poaceae: Triticeae) taxon complex. Mol Phyl Evol 55:650–659

Jauhar PP (1988) A reassessment of genome relationships between *Thinopyrum bessarabicum* and *T. elongatum* of the Triticeae. Genome 30:903–914

Kim KJ, Lee H-L (2005) Widespread occurrence of small inversions in the chloroplast genomes of land plants. Mol Cells 19:104–113

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374

Löve A (1984) Conspectus of Triticeae. Feddes Repert 95:425–521

Mahelka V, Kopecky D, Paštová L (2011) On the genome constitution andevolution of intermediate wheatgrass (*Thinopyrum intermedium*: Poaceae, Triticeae). BMC Evol Biol 11:127. doi:10.1186/1471-2148-11-127

Mason-Gamer RJ (2013) Phylogeny of a Genomically Diverse Group of *Elymus* (Poaceae) Allopolyploids Reveals Multiple Levels of Reticulation. PLoS One 8:e78449. doi:10.1371/journal.pone.0078449

Melderis A (1980) *Elymus*. In: Tuti TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA (eds) Flora Europaea 5. Cambridge University Press, Cambridge, pp 192–198

Mizianty M, Frey L, Szczepaniak M (1999) The *Agropyron-Elymus* complex (Poaceae) in Poland: nomenclatural problems. Fragm Flor Geobot 44:3–33

Naghavi MR, Rad MB, Riahi M, Taleie A (2013) Phylogenetic analysis in some *Hordeum* species (Triticeae; Poaceae) based on two single-copy nuclear genes encoding acetyl-CoA carboxylase. Biochem Syst Ecol 47:148–155

Ni Y, Asamoah-Odei N, Sun G (2011) Maternal origin, genome constitution and evolutionary relationships of polyploid *Elymus* species and *Hordelymus europaeus*. Biol Plant 55:68–74

Petersen G, Seberg O (1997) Phylogenetic analysis of the Triticeae (Poaceae) based on *rpo*A sequence data. Mol Phyl Evol 7:217–230

Petersen G, Seberg O (2008) Phylogenetic relationships of allotetraploid *Hordelymus europaeus* (L.) Harz (Poaceae: Triticeae). Pl Syst Evol 273:87–95

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

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Amer J Bot 92:142–166

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Amer J Bot 94:275–288

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

von Bothmer R, Lu B-R, Linde-Laursen I (1994) Intergeneric hybridization and C-banding patterns in *Hordelymus* (Triticeae, Poaceae). Pl Syst Evol 189:259–266

Wang RR-C (1985) Genome analysis of *Thinopyrum bessarabicum* and *T. elongatum*. Canad J Genet Cytol 27:722–728

Wang RR-C, von Bothmer R, Dvorak J, Fedak G, Linde-Laursen I, Muramatsu M (1995) Genome symbols in the Triticeae (Poaceae). In: Wang RR-C, Jensen KB, Jaussi C (eds) Proceedings of the Second International Triticeae Symposium. Forage and Range Laboratory, USDA-ARS, Logan, pp 29–34

Whitlock BA, Hale AM, Groff PA (2010) Intraspecific inversions pose a challenge for the *trnH-psbA* plant DNA barcode. PLoS One 5:e11533. doi:10.1371/journal.pone.0011533

Yan C, Sun G (2012) Multiple origins of allopolyploid wheatgrass *Elymus caninus* revealed by *RPB2*, *PepC* and *TrnD/T* genes. Mol Phyl Evol 64:441–451

Yen C, Yang J, Yen Y (2005) Hitoshi Kihara, Áskell Löve and the modern genetic concept of the genera in the tribe Triticeae (Poaceae). Acta Phytotaxon Sin 43:82–93

Yu J, Xue JH, Zhou SL (2011) New universal *matK* primers for DNA barcoding angiosperms. J Syst Evol 49:176–181