Original Article

DNA Barcoding in Selected Species and Subspecies of Rye (*Secale*) Using Three Chloroplast *Loci* (*matK, rbcL, trnH-psbA*)

Lidia SKUZA[1,2]*, Izabela SZUĆKO[1,2], Ewa FILIP[1,2], Anastazja ADAMCZYK[1]

[1]*University of Szczecin, Faculty of Biology, Institute for Research on Biodiversity, Department of Cell Biology, Wąska 13, 71-415 Szczecin, Poland; lidia.skuza@usz.edu.pl (*corresponding author); izabela.szucko@usz.edu.pl; ewa.filip@usz.edu.pl; anastazja.ada@gmail.com*
[2]*Centre for Molecular Biology and Biotechnology, Wąska 13, 71-415 Szczecin, Poland*

Abstract

DNA barcoding is a relatively new method of identifying plant species using short sequences of chloroplast DNA. Although there is a large number of studies using barcoding on various plant species, there are no such studies in the genus *Secale*. In this study the plant material consisted of 10 cultivated and non-cultivated species and subspecies of rye genus. Three chloroplast DNA regions (*rbcL, matK, trnH-psbA*) were tested for their suitability as DNA barcoding regions. Universal primers were used, and sequenced products were analyzed using Neighbor Joining and the Maximum Likelihood in the MEGA 7.1 program. We did not observe high variability in nucleotide sequences within the *matK* and *rbcL* regions. Only 2.2% of the sequences showed polymorphism in the *rbcL* region, while 6.5% in the *matK* region. The most variable *trnH-psbA* (15.6%) intergenic region was the most useful for rye barcoding. Individual application of the studied regions did not provide the expected results. None of the regions used in the study allowed the division of rye species and subspecies according to the adopted classification of the genus *Secale*. The results confirm that the use of *matK* and *rbcL* is insufficient for DNA barcoding in rye species, and better discrimination within the genus *Secale* can be obtained only in combination with the non-coding *trnH-psbA* sequence. Our results also indicate the necessity of using a different region. All of the new sequences have been deposited in Genbank.

Keywords: barcoding DNA; rye; cpDNA; intergenic spacer

Introduction

Rye (*Secale cereale* L.) belongs to the tribe *Triticeae* from the *Poaceae* (grass) family and is related to bread wheat (*Triticum aestivum* L.) and barley (*Hordeum vulgare* L.). It has the largest genome (~7.9 Gbp) among all diploid *Triticeae*, containing over 90% of repetitive sequences (Bartoš *et al.*, 2008; Bauer *et al.*, 2017). Rye crop (*Secale cereale* L.) is a rich and important source of valuable genes encoding, e.g., high protein content, resistance to diseases as well as morphological and biochemical traits that increase triticale value (×*Triticosecale* Wittmack) (Kubicka *et al.*, 2006). Rye, as compared to other cereals, is distinguished by its exceptional cold tolerance and higher yields from wheat and barley in poor and moderate soils and under drought stress conditions (Schittenhelm *et al.*, 2014). Translocations from rye genome are present in many cultivars of wheat grown all over the world, thanks to which wheat is characterized by better stress tolerance caused by both abiotic and biotic factors (Rabinovich, 1998). In addition, rye is a difficult object of genetic and breeding studies. The reason is the open-pollination, self-incompatibility and the relationship between heterozygosity and productivity, which arises as a result of inter-chromosomal gene interactions (Schlegel, 2006).

Recently, the species of cultivated rye has been fully sequenced (Bauer *et al.*, 2017), which will certainly have a positive effect on large-scale functional analyses as well as rye genetic modification for sustainable plant production. However, the task of modern cereal breeding is still to obtain new, better yielding cultivars, characterized by high resistance to diseases, pathogens and unfavourable abiotic conditions. Progress in rye breeding has been unfortunately significantly slowed down and limited, because the cultivars used in cultivation are characterized by limited variability due to continuous selection and attempts to use old cultivars proved to be ineffective. The wild rye species and

subspecies are an excellent starting material for research aimed at expanding the recombination variability in the *Secale cereale* L. species. They are, due to their genetic distinctiveness and high trait expression, a valuable source of genes, in which our cultivars are poor (Rzepka-Plevneś, 1990).

Disease-resistance genes, e.g., to downy mildew or brown rust, resistance to pre-harvest sprouting (Rzepka-Plevneś, 1993; Rzepka-Plevneś and Tomczak, 1993) and lodging (Rzepka-Plevneś, 1993) are sought in wild species (*S. vavilovii*, *S. sylvestre*, *S. montanum*, ssp. *montanum* and *S. kuprijanovii*) (Rzepka-Plevneś, 1990). *S. anatholicum* is used for breeding cultivars with increased grain protein content and cultivars for the green mass. Wild rye species are also a source of male sterility genes (Stracke, 2003).

Generation of interspecific hybrids is currently often successful, however, their yield and quality pose problems (Rzepka-Plevneś, 1993). In addition, hybrids produced are not suitable for cultivation, they require many years of backcrossing with cultivated rye to restore functional traits. These difficulties cause that despite many years of research, wild rye species are still under-utilized as a source of desired genes and there are no reports on the genetic structure of these species in the world literature.

The introduction of DNA barcoding was a breakthrough in species identification methods. The basis of this technique is the use of a very short, defined genomic sequence that allows obtaining a DNA barcode – an image of base pair sequence in the DNA fragment that can be compared to determine individual species classification (Ajmal *et al.* 2014; Skuza *et al.*, 2015). The gene encoding cytochrome oxidase (*COI*, *coxI*) subunit of 648 bp, located in the mitochondrial genome is the best gene used for barcoding in animals (Stoeckle and Thaler 2014). The *COI* gene has also been shown to be effective in identifying birds, fish, butterflies, flies, bats and many other animal groups. However, among plants, the mitochondrial genome could not be used due to the different evolution of this genome in plants as well as the possibility of plant interbreeding – the possible presence of mitochondria from different species in one plant (Hollingsworth *et al.*, 2011). Research conducted by the group working on plant barcoding (CBOL Plant Working Group, 2009), which compared several different sets of genes potentially useful for barcoding, found that the best and most reliable results are obtained for chloroplast genes: *matK*, encoding maturase, and *rbcL*, encoding the large subunit of RuBisCO (Hollingsworth *et al.*, 2011). As a result, various fragments of the chloroplast genome have been proposed as plant barcodes. They were selected from 4 coding regions: *matK*, *rbcL*, *rpoB*, *rpoC1* and from the pool of non-coding fragments: *atpF-atpH*, *trnH-psbA*, *psbK-psbI* (Hollingsworth *et al.*, 2009), *trnL*, *trnL-trnF* and *trnK* intron/*matK* (Bellstedt *et al.*, 2001; Ge *et al.*, 2002; Klak *et al.*, 2003; Muellner *et al.*, 2003; Samuel *et al.*, 2003).

Eventually, the group of potential plant barcodes was narrowed down to *matK* and *rbcL* genes and to the non-coding *trnH-psbA* region. Thorough research on the effectiveness of species identification and the ease of obtaining the sequence of each potential barcode was carried out. None of these regions alone met all the requirements of DNA barcoding, therefore, in contrast to animals in which the barcode region consists of only one locus (*coxI*), it was decided to use a barcode consisting of two loci in plants, i.e., *matK* and *rbcL* (Hollingsworth *et al.*, 2009). The *rbcL* and *trnH-psbA* pair provided very good results, but due to difficulties in amplification and subsequent assembly of the *trnH-psbA* intergenic region sequence, a pair consisting of two coding regions was selected for the official plant DNA barcode: *rbcL* i *matK* (Kress and Erickson, 2007; Hollingsworth *et al.* 2009).

Plastid DNA is widely used as a marker of choice in phylogenetic and phylogeographical studies, however, little is known about its usefulness in analyzing the relationships between closely related species. The slow rate of cpDNA - specific evolution hinders taxonomic analyses at lower levels, especially at the population level. In addition, studies clearly indicate that the utility in phylogenetic analyses of different cpDNA non-coding regions within a given taxonomic group can vary enormously (Sang *et al.*, 1997; Xu *et al.*, 2000; Hartmann *et al.*, 2002; Hamilton 2003; Sakai *et al.*, 2003), and the selection of the appropriate cpDNA region is often difficult due to the lack of information about the rate of evolution between different non-coding cpDNA regions.

Given the above, the aim of our research was to: i) investigate whether the three cpDNA regions, which were previously proposed as barcoding tools for various angiosperms: *rbcL* and *matK* genes, and the intergenic *trnH-psbA* region, can be used as barcodes to distinguish representatives of rye species and subspecies and ii) assess the value of phylogenetic information provided by these markers.

Materials and Methods

Plant material

The plant material consisted of 10 cultivated and non-cultivated species and subspecies of rye genus, obtained from several world collections (Center for Biological Diversity Conservation in Powsin-Warsaw, Poland; United States Department of Agriculture - Agricultural Research Service, USA; Nordic Genetic Resource Center, Sweden) (Table 1).

DNA extraction, PCR amplification and sequencing

The probes of genomic DNA were isolated from 10 randomly chosen fresh leaves of 6 to 7-day-old etiolated plants. The leaves were ground with liquid nitrogen, producing ~100 mg of fine powder. The isolation was performed using FastDNA® Green SPIN Kit (DNAeasy Plant Mini Kit-Wizard® Genomic DNA, Promega). Both quality and concentration of the DNA were assessed by agarose gel electrophoresis and spectrophotometry (NanoDrop 2000; Thermo Scientific).

The reaction was carried out in duplicate by way of PCR analyses which were performed in a T100™ Thermal Cycler (Bio-Rad) in the final volume of 20 µl. The single PCR reaction mixture contained: 1x DreamTaq Buffer, 0.2 mM dNTP, 0.1 µM of each primer, 50 ng genomic DNA and 1 U DreamTaq DNA Polymerase (Thermo Scientific).

The primers used for amplification of *rbcL* were rbcL1f: ATGTCACCACAAACAGAAAC and rbcL724r: TCG

56

CATGTACCCTGCAGTAGC; for *matK* were matK 390F:CGATCTATTCATTCAATATTTC, matK1326 R:TCTAGCACACGAAAGTCGAAGT, while the primers used for amplification of *trnH-psbA* were psbA3'f: GTTATGCATGAACGTAATGCTC, trnHf05: CGCG CATGGTGGATTCACAATCC (Parmentier *et al.*, 2013). The primers were synthesized in the Laboratory of DNA Sequencing and Synthesis of IBB PAN Genomed S.A. (Warsaw).

The following thermal reaction profile was used to amplify the *rbcL* and *trnH-psbA* regions: initial denaturation at 95 °C for 3 min followed by 33 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, extension of the primer at 72 °C for 1 min and a final extension of 72 °C for 10 min. The following thermal reaction profile was used to amplify the *matK* region: initial denaturation at 95 °C for 3 min, then 40 cycles including denaturation at 95 °C for 30 s, annealing at 49 °C for 30 s, primer extension at 72 °C for 1 min and final extension at 72 °C for 10 min. The conditions and profiles of PCR reactions have been optimized accordingly.

PCR products were checked by electrophoresis in 1.5% (m v^{-1}) agarose gel containing ethidium bromide and a TBE buffer (pH 8.0); the gels were visualized under UV. The gel was analyzed and archived using the Molecular Imager® GelDoc™XR software. Bands were scored and analyzed with the Quantity One software (Bio-Rad). The size of the products was determined by comparison with a DNA ladder (MassRuler, Thermo Scientific). The purified PCR products were sequenced on both strands by Genomed (Poland) using the PCR primers. The sequences reported in this paper have been deposited in the GenBank nucleotide sequence database with the accession numbers MG905722 - MG905751 (Table1).

Barcoding analyses

The analyzed dataset consisted of 36 nucleotide sequences: 30 resulted from DNA sequencing experiments performed during this study and 6 were received from GenBank. At first the forward and reverse sequences were edited and consensus sequences were obtained using Basic Local Alignment Tool software. ClustalW and Mega7.1 software were used to perform multiple sequence alignments.

The genetic variability of each marker was described by the total alignment length (bp); the number of monomorphic sites; the number of polymorphic sites; the percentage of polymorphic and monomorphic sites; the number of singleton variable sites; the number of parsimony informative sites (PIC), nucleotide diversity (Pi); the number of haplotypes and the average G+C contents in each region using DnaSP6.10.01.

The barcoding analyses were conducted separately for each region and combined in the following arrangement: coding region (*matK+rbcL*) and plastid genome regions (*matK+rbcL+ trnH-psbA*).

The resolution of each locus was evaluated by the Neighbor Joining (NJ) trees and the Maximum Likelihood (ML) trees which were built with Mega 7.1 software. Tamura-3 parameter model was determined for each locus. The reliability branching was tested using the bootstrap method with 1000 replications in the NJ and ML analyses.

The outgroup consisted of the chloroplast genome sequences obtained from *Avena sativa* (NC_027468.1) and the in-group consisted of sequences obtained from *Triticum turgidum* (KJ614402.1).

Results

As a result of PCR reactions and sequencing, the number of *matK* and *rbcL* gene sequences as well as the *trnH-psbA* intergenic region sequences, which are used to refine the barcoding analysis, were obtained. All the new sequences have been deposited in GenBank under accession numbers. Twelve nucleotide sequences were analyzed for each of the loci: *matK*, *rbcL* and *trnH-psbA*. The *matK* gene had the longest sequence (832 bp) and the *trnH-psbA* intergenic region the shortest (589 bp) based on the multiple alignments of all sequences obtained from the analyzed regions (Table 2). The average GC content was 33.6% for *matK*, 43.8% for *rbcL* and 35.7% for *trnH-psbA*. The *trnH-psbA* region was the most variable sequence, as it was characterized by variability at the level of 81.5%, while the least variable was the *rbcL* region -97.8%. The *rbcL* region was characterized by a small number of haplotypes (3) and no parsimonious informative sites were observed due to its high level of monomorphism. Similarly, the *trnH-psbA* region was characterized by the greatest

Table 1. The list of plant species, origin, accession number, type, life cycle and sequence accessions number included in the study

Species	Origin	Accession number	Type	Life cycle	Sequence accession no.		
					matK	*rbcL*	*trnH-psbA*
S. cereale ssp. *afghanicum*	Armenia	PI 618662	We	A	MG905723	MG905722	MG905724
S. cereale ssp. *ancestrale*	USA	PI 445976	We	A	MG905732	MG905731	MG905733
S. cereale ssp. *cereale*	Canada	PI 590948	C	A	MG905735	MG905734	MG905736
S. cereale ssp. *segetale*	Azerbaijan	PI 267102	We	A	MG905741	MG905740	MG905742
S. cereale ssp. *rigidum*	Turkey	PI 618669	We	A	MG905738	MG905737	MG905739
S. strictum ssp. *africanum*	Poland	6063	Wi	P	MG905726	MG905725	MG905727
S. strictum ssp. *anatolicum*	USA	PI 445973	Wi	P	MG905729	MG905728	MG905730
S. strictum ssp. *strictum*	Iraq	PI 253956	Wi	P	MG905744	MG905743	MG905745
S. vavilovii	Hungary	PI 28842	Wi	A	MG905750	MG905749	MG905751
S. sylvestre	Poland	PI 618676	Wi	A	MG905747	MG905746	MG905748

C: cultivated; Wi: wild; We: weedy; A: annual; P: perennial.

polymorphism, hence the number of haplotypes was the highest (6) and 5 PICs were observed. The intergenic region was also characterized by the highest Pi coefficient -0.03, and the lowest one was found for the *rbcL* gene sequence - 0.004 (Table 2).

Phylogenetic NJ and ML trees were constructed based on sequence analysis of 3 regions and their combinations. No differences were observed in the topology of trees constructed using the aforementioned algorithms. Their degree of sampling (bootstrap) was the only difference.

There was no high variability in nucleotide sequences within the *matK* and *rbcL* regions. Only 2.2% of the sequences showed polymorphism in the *rbcL* region, while 6.5% in the *matK* region. Considering the above, the *rbcL* gene was characterized by insufficient sequence variation to be used to distinguish rye species and subspecies analyzed by the authors (Fig. 2). In turn, it could be observed that *Secale cereale* spp. *ancestrale* was slightly different from the other analyzed species/subspecies of rye, thanks to the analysis trees' topology of *matK* sequences (Fig. 1). This species clustered on a separate branch. It could also be observed that the taxon *Secale cereale* sp. *ancestrale* is distant from the others in the trees obtained using the *matK+rbcL* combination (Fig. 4). A different length of the *Secale*

sylvestre branch was also recorded, which might have been caused by a genetic change between the remaining rye species, but not sufficiently different that it could be separated from the other 8 rye species or subspecies.

The *trnH-psbA* region was characterized by the highest polymorphism -15.6%, thus the trees obtained based on the sequences of this region also most highly discriminated the material analyzed by the authors. The analyzed species/subspecies were divided into two similarity groups. Within the first group, two subgroups were distinguished, clustered on separate branches: i) *Secale cereale* ssp. *afghanicum* and the second more numerous: ii) *Secale cereale* ssp. *ancestrale, Secale cereale* ssp. *cereale, Secale cereale* ssp. *rigidum, Secale cereale* ssp. *segetale* and *Secale strictum* ssp. *africanum*. Two subgroups were additionally distinguished on separate branches in the second group: i) *Secale sylvestre* and *Secale strictum* ssp. *anatolicum* and ii) *Secale vavilovii* and *Secale strictum* ssp. *strictum* (Fig. 3).

The use of the *matK* and *rbcL* combination with *trnH-psbA* increased the efficiency of the barcode analysis. This was due to the fact that the *trnH-psbA* region was characterized by a high genetic variability in closely related taxa analyzed in the above experiments. The analyzed species/subspecies were divided into two similarity groups,
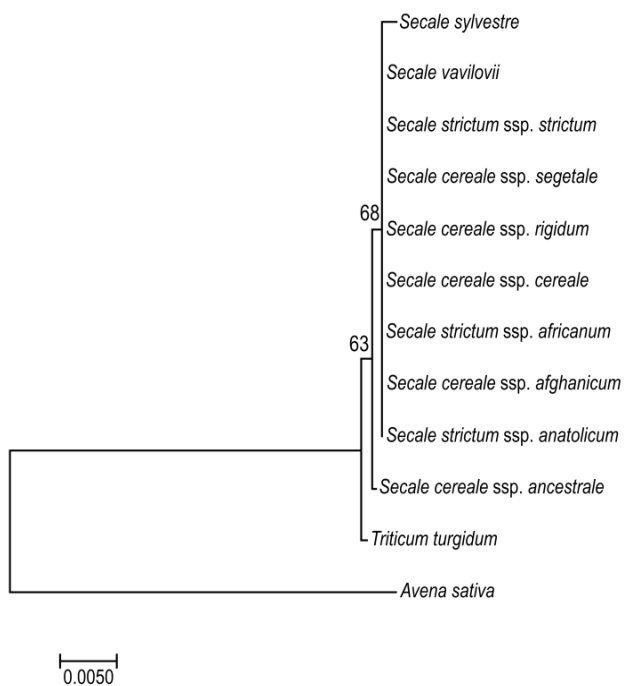

Fig. 1. Cladogram for *matK* sequences for *Secale* species and subspecies generated by the Neighbor-Joining method's. The bootstrap values are shown under the branches
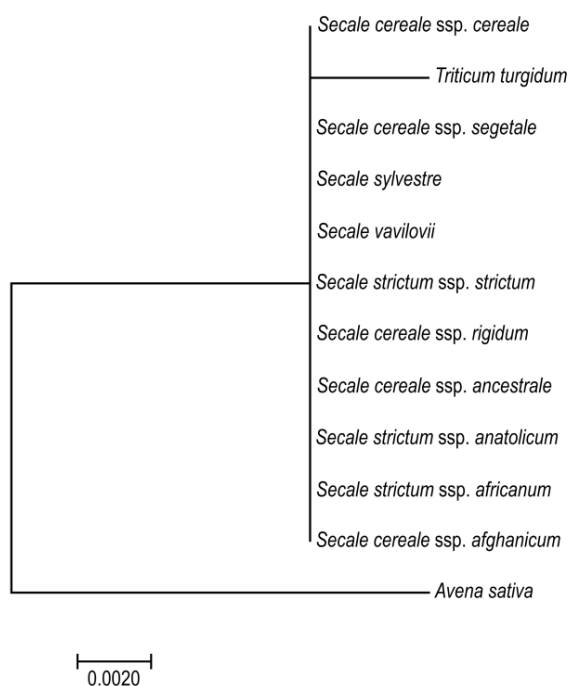
Fig. 2. Cladogram for *rbcL* sequences for *Secale* species and subspecies generated by the Neighbor-Joining method's

Table 2. Molecular characteristic of the three chloroplast *loci* evaluated for genus *Secale*

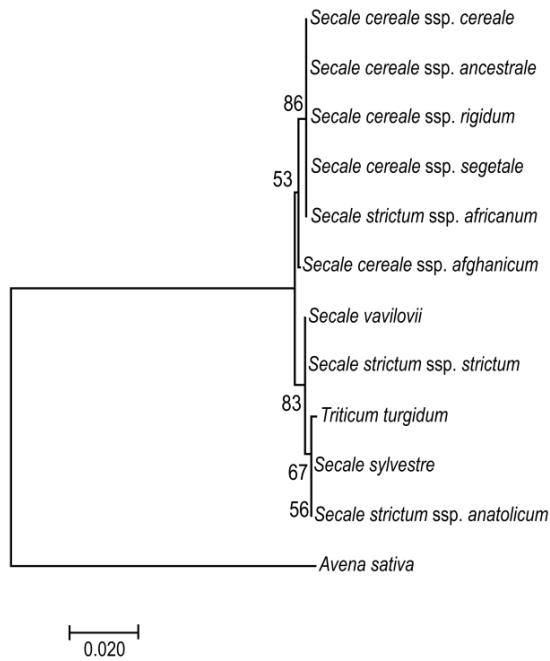| Barcode	Total alignment length (bp)	Number of monomorphic sites	Number of polymorphic sites	Number of parsimony informative sites (PIC)	Number of singleton variable sites	Number of haplotypes	Nucleotide diversity (Pi)
matK	832	778	54	2	52	5	0.011
rbcL	624	610	14	0	14	3	0.004
trnH-psbA	589	480	92	5	87	6	0.030

Fig. 3. Cladogram for *trnH-psbA* for *Secale* species and subspecies generated by the Neighbor-Joining method's. The bootstrap values are shown under the branches
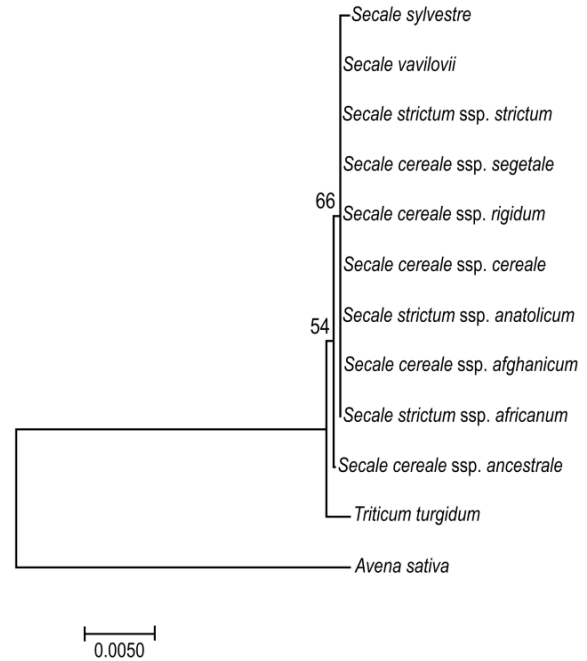

Fig. 4. Cladogram for *matK+rbcL* for *Secale* species and subspecies generated by the Neighbor-Joining method's. The bootstrap values are shown under the branches
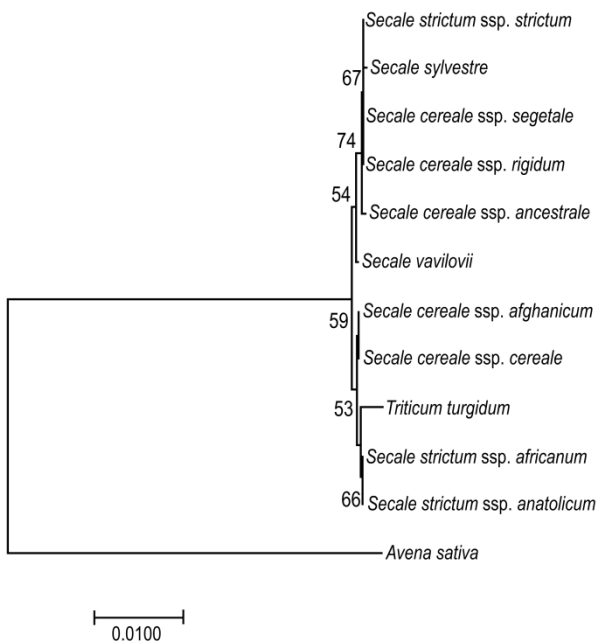

Fig. 5. Cladogram for *matK+rbcL+trnH-psbA* for *Secale* species and subspecies generated by the Neighbor-Joining method's. The bootstrap values are shown under the branches

in which subgroups were distinguished. Three similarity subgroups were identified in the first group: i) *Secale vavilovii*; ii) *Secale cereale* spp. *ancestrale*; iii) *Secale sylvestre*, *Secale cereale* ssp. *rigidum*, *Secale strictum* ssp. *strictum* and *Secale cereale* ssp. *segetale*. Two subgroups were identified in the first group: i) *Secale cereale* ssp. *afghanicum* and *Secale cereale* ssp. *cereale*; ii) *Secale strictum* ssp. *africanum* and *Secale strictum* ssp. *anatolicum* (Fig. 5).

Discussion

According to the CBOL plant working group, an ideal DNA barcode needs to have the following features: capacity of amplification with universal primers, high amplification and sequencing efficiency, and genetic variation that is sufficiently high to distinguish sequences at the species level, but also sufficiently conservative among individuals of the same species (Hebert *et al.*, 2003; Cowan *et al.*, 2006; CBOL Plant Working Group, 2009).

Evaluation of universal applicability by PCR quantification and sequencing success is the first step in determining the suitability of a given DNA fragment as a barcode.

In this respect, all analyzed regions (*matK, rbcL* and *trnH-psbA*) amplified effectively, which allowed for simple and high-quality sequencing.

The amplification of the *trnH-psbA* region was also successful, despite the fact that many authors reject this region as a barcode because of its length (>1000 bp) and difficulties in bi-directional sequencing (CBOL Plant Working Group, 2009; Hollingsworth *et al.*, 2009). The amplicons obtained in our experiments were shorter (about 600 bp), which allowed for effective sequencing. Similar results were obtained for other groups of terrestrial plants, where the amplification of the *trnH-psbA* region and the sequencing quality was sufficiently high to consider it a barcode (Kress *et al.*, 2009; Tripathi *et al.*, 2013; Bieniek *et al.*, 2015; Su *et al.*, 2016).

In turn, many studies have indicated that *matK* is a key marker discriminating specific groups (Newmaster *et al.*, 2009; De Mattia *et al.*, 2011), although many authors questioned the usefulness of this gene as a barcode due to poor amplification and sequencing efficiency and problems

related to primers' universality (Sass *et al.*, 2007; Roy *et al.*, 2010; Kelly *et al.*, 2010; Du *et al.*, 2011; Yan *et al.*, 2011; Theodoridis *et al.*, 2012). The research presented in the study indicates that despite PCR and sequencing efficiency, unfortunately, this region can not be considered as an effective rye barcode. Analyses involving this sequence showed only 6.5% polymorphism in the studied taxa.

However, in terms of molecular variability, *rbcL* was the most conservative sequence among the three analyzed regions, as indicated by the lowest number of polymorphic sites and the obtained haplotypes (Fig. 2). This was also confirmed by other authors (Fazekas *et al.*, 2008; Zimmermann *et al.*, 2013; Bolson *et al.*, 2015; Bieniek *et al.*, 2015; Gamache and Sun, 2015).

Phylogenetic analysis is one of the most effective methods to determine the suitability of a DNA region as a barcode, because it should detect species-specific clusters. Unfortunately, it is complicated in rye, because the phylogenetic relationships between *Secale* species remain unclear, despite the large number of analyses. A division of the genus *Secale* even into 15 different species has been adopted (Delipavlov 1962), while Frederiksen and Petersen (1998) recognized only three *Secale* species: *Secale sylvestre*, *Secale strictum* and *S. cereale*. The classification system of the American Germplasm Resources Information Network (GRIN, http://www.arsgrin.gov) currently includes four species in the genus *Secale*: annual *S. cereale* L., annual *S. sylvestre* Host and *S. vavilovii* Grossh and perennial *S. strictum* (Presl.) Presl. (syn. *S. montanum*) (Spencer and Hawkes 1980; DeBustos and Jouve 2002). Moreover, *S. cereale* also comprises 8 subspecies, *S. strictum* - 5, and *S. cereale ssp. cereale* is the only cultivated species.

The differences in the classification of the genus *Secale* result from the use of various experimental methods, such as isoenzyme analyzes (Vences *et al.* 1987a,b), RAPD (Del Pozo *et al.*, 1995), ITS rDNA analyses (Reddy *et al.* 1990; Cuadrado and Jouve 2002), thin-layer chromatography (Dedio *et al.* 1969), ITS rDNA region length analyses (Reddy *et al.* 1990), restriction fragment length polymorphism (RFLP) of the chloroplast genome (Murai *et al.* 1989) and cpDNA variation analyses (Petersen and Doebley 1993), microsatellite sequence analyses (Shang *et al.*, 2006), but also traditional morphological and cytogenetic methods.

None of the regions used in the study allowed the division of rye species and subspecies according to the adopted classification of the genus *Secale*. The *rbcL* region did not differentiate the analyzed taxa (Fig. 2), because the obtained sequences were very similar, with only 14 polymorphic sites (Table 2).

Similar results were obtained by Gamache and Sun (2015), who identified species from the genus *Pseudoroegneria*. As regards the genus *Panicum*, the *rbcL* gene alone was also insufficient to identify individual species, similarly as the *matK* gene (Zimmermann *et al.*, 2013). Only the combination of results of these two regions was sufficient for analysis. In turn, Zhang *et al.* (2011) showed that individual plant species can be distinguish by analyzing this region (including *Arabidopsis thaliana*, *Oryza sativa* subsp. *japonica* or *Zea mays*). This region also demonstrated reasonably good effectiveness at lower taxonomic levels in *Hordeum* (Bieniek *et al.*, 2015; Gamache and Sun, 2015). Bieniek *et al.* (2015) identified *Hordeum bulbosum* or *H. bogdani* using the *rbcL* region.

Our research shows that the *matK* gene sequences are also highly similar in the analyzed taxa (54 polymorphic sites have been identified) and allow only the identification of *Secale cereale ssp. ancestrale*. Bieniek *et al.* (2015) obtained different results, demonstrating high species identification capacity, but also for the genus, using the *matK* gene alone in the genera *Elymus, Loptiopyrum, Pseudoroegneria* and *Thinopyrum*. Similarly, the identification of species of the genus *Panicum* using the *rbcL* and *matK* genes individually discriminated species, despite the low number of SNPs (Hunt *et al.*, 2014). These results are in contradiction with the study of Zimmerman *et al.* (2013) in relation to the genus *Panicum*. This might result from a larger number of species selected for analysis – 9 (Zimmermann *et al.*, 2013) and 24 (Hunt *et al.*, 2014), respectively.

The intergenic *trnH-psbA* region demonstrated the highest species identification capacity in our study among all 3 regions used autonomously. Six haplotypes were distinguished, however, sequence analysis of this region allowed to identify rye only at the species level. This region was insufficient for the identification of rye to subspecies among *S. cereale* and *S. strictum* species (Fig. 3).

Only the combination of *matK* and *rbcL* with *trnH-psbA* increased the efficiency of barcode analysis, although in this case there were also some discrepancies with the adopted classification (Fig. 5).

S. cereale ssp. *cereale* species were dispersed within both groups. *S. vavilovii* species was in the *S. cereale, S. strictum* and *S. sylvestre* species group (NJ), or *S. cereale* and *S. africanum, Secale strictum* ssp. *anatolicum*. The result of our analysis was partly consistent with the classification of Frederiksen and Petersen (1998), who identified only three species within the genus *Secale*: *S. sylvestre, S. strictum* and *S. cereale* and included *S. vavilovii* to *S. cereale*. Similarly, Bolibok-Brągoszewska *et al.* (2014) classified *S. vavilovii* as a subspecies of *S. cereale*. Shang and *et al.* (2006) reached similar conclusions, indicating high similarity between these species.

S. sylvestre is highly similar to *S. cereale* ssp. *segetale*. This was confirmed by previous results (Skuza *et al.*, 2007) obtained in the RFLP analysis of mitochondrial genes. However, the obtained results were not consistent with the current classification of the genus *Secale* based on many nuclear molecular markers. Although Ren *et al.* (2011) did not classify *S. sylvestre* as a separate group, nevertheless, they claimed, on the basis of their research, that it was more related to *S. strictum* ssp. *africanum* and *S. strictum* ssp. *anatolicum*.

In turn, Bolibok-Brągoszewska *et al.* (2014) classified *S. sylvestre* to a separate taxon. Ren *et al.* (2011) obtained different results based on microsatellite analysis. They showed similarity of *S. sylvestre* to *S. strictum* ssp. *africanum* and *anatolicum*. Skuza *et al.* (2007) in turn classified *S. sylvestre* together with *S. cereale* ssp. *segetale* based on mtDNA analysis.

S. sylvestre along with *S. vavilovii* are the only species that do not generate hybrids, although both are annual and self-pollinating (Singh, 1975). These results would support the suggestion of Khush (1962) that *S. sylvestre* should be placed in a separate silvestria section. However, the research carried out in the present work suggests that this species should be included together with *S. segetale*.

Our results partially confirmed the very close relationship between *S. sylvestre* and *S. segetale* species and also supported the exclusion of *S. vavilovii* as a separate species.

The *strictum* species group is heterogeneous and shows similarity to *S. cereale ssp. ancestrale* similarly to the work of Ren (2011) and to *S. afghanicum*. The analysis showed low similarity of *S. strictum* ssp. *africanum* and *S. strictum* ssp. *strictum* species, contrary to the currently adopted classification. However, they are consistent with the ISSR analyses, indicating a close relationship between *S. strictum* ssp. *africanum* and *S. strictum* ssp. *anatolicum* (Ren *et al.*, 2011). Genetic diversity in the evolutionary process was lower in the *strictum* group than between perennial and annual forms and species. In addition, it has been shown that perennial forms are morphologically similar and cross easily to form hybrids (Spencer and Hawkes 1980).

Conclusions

The present study is the first to analyze selected rye species and subspecies, in which the usefulness of the combinations of the plastid *rbcL* and *matK* coding regions and intergenic *trnH-psbA* region for DNA barcoding was assessed. The results confirm that the use of *matK* and *rbcL* is insufficient for DNA barcoding in rye species, and better discrimination within the genus *Secale* can be obtained only in combination with the non-coding *trnH-psbA* sequence. Our results also indicate the need to use a different region, e.g., the previously proposed ITS2 supported by the intergenic *trnH-psbA* region, in order to correctly identify rye species (Chen *et al.*, 2010; Roy *et al.*, 2010).

References

Ajmal AM, Gyulai G, Hidwegi N, Kerti B, Al Hemaid F, Pandey AK, Lee J (2014). The changing epitome of species identification – DNA barcoding. Saudi Journal of Biological Sciences 21:204-231.

Bartoš J, Paux E, Kofler R, Havránková M, Kopecký D, Suchánková P, … Doležel J (2008). A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biology 8:1-12.

Bauer E, Schmutzer T, Bauer E, Schmutzer T, Barilar I, Mascher M, … Scholz U (2017). Towards a whole-genome sequence for rye (*Secale cereale* L.). The Plant Journal 89(5):853-869.

Bellstedt DU, Linder HP, Harley EH (2001). Phylogenetic relationships in *Disa* based on noncoding *trnL-trnF* chloroplast sequences: evidence of numerous repeat regions. American Journal of Botany 88:2088-2100.

Bieniek W, Mizianty M, Szklarczyk M (2015). 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. Plant Systematics and Evolution 301:1275-1286.

Bolibok-Brągoszewska H, Targońska M, Bolibok L, Kilian A, Rakoczy-Trojanowska M (2014). Genome-wide characterization of genetic diversity and population structure in *Secale*. BMC Plant Biology 14:184.

Bolson M, de Camargo SE, Brotto MA, Silva-Pereira V (2015). ITS and *trnH-psbA* as Efficient DNA Barcodes to Identify Threatened Commercial Woody Angiosperms from Southern Brazilian Atlantic Rainforests. PLoS One 10(12):e0143049.

CBOL Plant Working Group (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106:12794-12797.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, … Leon C (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal. PLoS One 5(1):e8613.

Cowan RS, Chase MW, Kress WJ, Savolainen V (2006). 300,000 species to identify: problems, progress and prospects in DNA barcoding of land plants. Taxon 55:611-616.

Cuadrado A, Jouve N (2002). Evolutionary trends of different repetitive DNA sequences during speciation in the genus *Secale*. Journal of Heredity 93(5):339-345.

DeBustos A, Jouve N (2002). Phylogenetic relationships of the genus *Secale* based on the characterization of rDNA ITS sequences. Plant Systematics and Evolution 235:147-154.

Dedio W, Kaltsikes PJ, Larter EN (1969). Numerical chemotaxonomy in the genus *Secale*. Canadian Journal of Botany 47:1175-1180.

Del Pozo JC, Figueiras AM, Benito C, De La Pena A (1995). PCR derived molecular markers and phylogenetic relationships in the *Secale* genus. Biologia Plantarum 37:481-489.

Delipavlov D (1962). *Secale rhodopaeum* Delipavlov - A new species of rye from the Rhodope Mountains. Doklady Bolgarskoi Akademii Nauk 15:407-411.

DeMattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M (2011). A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. Food Research International 44:693-702.

Du ZY, Qimike A, Yang CF, Chen JM, Wang QF (2011). Testing four barcoding markers for species identification of Potamogetonaceae. Journal of Systematic and Evolution 49:246-251.

Fazekas AJ, Burgess KS, Kesanakurit PR, Graham SW, Newmaster SG, Husband BC, … Barrett SCH (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS One 3(7):e2802.

Frederiksen S, Petersen G (1998). A taxonomy revision of *Secale* (Triticeae, Poceae). Nordic Journal of Botany 18:399-420.

Gamache J, Sun G (2015). Phylogenetic analysis of the genus Pseudoroegneria and the Triticeae tribe using the *rbcL* gene. Biochemical Systematics and Ecology 62:73-81.

Ge SA, Li B, Lu R, Zhang SZ, Hong DY (2002). A phylogeny of the rice tribe Oryzeae (Poaceae) based on *matK* sequence data. American Journal of Botany 89:1967-1972.

Hamilton MB, Braverman JM, Soria-Hernanz DF (2003). Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. Molecular Biology and Evolution 20:1710-1721.

Hartmann S, Nason JD, Bhattacharya D (2002). Phylogenetic origins of *Lophocereus* (Cactaceae) and the senita cactus-senita moth pollination mutualism. American Journal of Botany 89:1085-1092.

Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003). Biological identification through DNA barcodes. Proceedings of the Royal Society of London 270:313-322.

Hollingsworth ML, Clark A, Forrest LL, Richardson J, Pennington RT, Long DG, … Hollingsworth PM (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Resources 9:439-457.

Hollingsworth PM, Graham SW, Little DP (2011). Choosing and using a plant DNA barcode. PLoS One 6(5):e19254.

Hunt HV, Badakshi F, Romanova O, Howe CJ, Jones MK, Heslop-Harrison JSP (2014). Reticulate evolution in *Panicum* (Poaceae): the origin of tetraploid broomcorn millet, *P. miliaceum*. Journal of Experimental Biology 65(12):3165-3175.

Kelly LJ, Ameka GK, Chase MW (2010). DNA Barcoding of African Podostemaceae (river-weeds): A test of proposed barcode regions. Taxon 59:251-260.

Khush GS (1962). Cytogenetic and evolutionary studies in *Secale*. II. Interrelationships of the wild species. Evolution 16:484-496.

Klak C, Hedderson TA, Linder P (2003). A molecular systematic study of the *Lampranthus* group (Aizoaceae) based on the chloroplast *trnT-trnF* and nuclear ITS and 5S NTS sequence data. Systematic Botany 28:70-85.

Kress WJ, Erickson DL (2007). A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLOS ONE 2(6):e508.

Kress WJ, Erickson DL, Andrew JF, Swenson NG, Perez R, Sanjur O, Bermingham E (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proceedings of the National Academy of Sciences of the United States of America 106:18621-18626.

Kubicka H, Puchalski J, Niedzielski M, Łuczak W, Martyniszyn A (2006). Gromadzenie i ocena zasobów genowych żyta [Collection and evaluation of rye gene resources]. Biulletin of Plant Breeding Acclimatization Instituite IHAR 40(241):141-149.

Muellner AN, Samuel R, Johnson SA, Cheek M, Pennington TD, Chase MW (2003). Molecular phylogenetics of *Meliaceae* (Sapindales) based on nuclear and plastid DNA sequences. American Jouranl of Botany 90:471-480.

Murai K, Naiyu X, Tsunewaki K (1989). Studies on the origin of crop species by restriction endonuclease analysis of organellar DNA. III. Chloroplast DNA variation and interspecific relationships in the genus *Secale*. The Japanese Journal of Genetics 64:36-47.

Newmaster SG, Ragupathy S (2009). Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). Molecular Ecology Resources 9:172-180.

Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, … Hardy OJ (2013). How effective are DNA barcodes in the identification of African rainforest trees?. PLoS One 8(4):e54921.

Petersen G, Doebley F (1993). Chloroplast DNA variation in the genus *Secale* (Poaceae). Plant Systematics and Evolution 187:115-125.

Rabinovich SV (1998). Importance of wheat-rye translocations for breeding modern cultivar of *Triticum aestivum* L. Euphytica 100:323-340.

Reddy P, Appels R, Baum BR (1990). Ribosomal DNA spacer-length variation in *Secale* spp. (Poaceae). Plant Systematics and Evolution 171:203-220.

Ren TH, Chen F, Zou YT, Jia YH, Zhang HQ, Yan BJ, Ben ZL (2011). Evolutionary trends of microsatellites during the speciation process and phylogenetic relationships within the genus *Secale*. Genome 54(4):316-326.

Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, … Tuli R (2010). Universal plant DNA barcode loci may not work in complex groups: A case study with Indian Berberis species. PLoS One 5:e13674.

Rzepka-Plevneś D (1990). Właściwości międzygatunkowych mieszańców żyta (*Secale* sp.) [Properties of cross-species hybrid rye (*Secale* sp.)]. Bulletin of Plant Breeding Acclimatization Institute IHAR 32/3(4):27-36.

Rzepka-Plevneś D (1993). Badania nad mieszańcami *S. cereale × S. vavilovii* Gross. w aspekcie ich przydatności w hodowli odmian żyta odpornych na porastanie. Część I [Utility properties of hybrids *S. cereale × S. vavilovii* Gross. in terms of their suitability in growing rye varieties resistant to sprouting, Part I]. Bulletin of Plant Breeding Acclimatization Institute IHAR 37/5(6):69-79.

Rzepka-Plevneś D, Tomczak P (1993). Badania nad mieszańcami *S. cereale × S. vavilovii* Gross. w aspekcie ich przydatności w hodowli odmian żyta odpornych na porastanie. Część II [Utility properties of hybrids *S. cereale × S. vavilovii* Gross. in terms of their suitability in growing rye varieties resistant to sprouting, Part II]. Bulletin of Plant Breeding Acclimatization Institute IHAR 37/5(6):81-91.

Sakai MA, Kanazawa A, Fujii F, Thseng S, Abe J, Shimamoto Y (2003). Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer sequences. Plant Systematics and Evolution 239:29-54.

Samuel RT, Stuessy F, Tremetsberger K, Baeza CM, Siljak-Yakovlev S (2003). Phylogenetic relationships among species of *Hypochaeris* (Asteraceae, Cichorieae) based on ITS, plastid *trnL* intron, *trnL-F* spacer, and *matK* sequences. American Journal of Botany 90:496-507.

Sang TD, Crawford J, Stuessy TF (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). American Journal of Botany 84:1120-1136.

Sass C, Little DP, Stevenson DW, Specht CD (2007). DNA Barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. PLoS One 2:e1154.

Schittenhelm S, Kraft M, Wittich KP (2014). Performance of winter cereals grown on field-stored soil moisture only. European Journal of Agronomy 52(B):247-258.

Schlegel R (2006). Rye (*Secale cereale* L.): A younger crop plant with a bright future. Singh RJ, Jauhar PP (Eds). In: Genetic Resources, Chromosome

62

Engineering, and Crop Improvement: Cereals. Vol 2. CRC Press, US pp 365-394.

Shang H-Y, Wei Y-M, Wang X-R, Zheng Y-L (2006). Genetic diversity and phylogenetic relationships in the rye genus *Secale* L. (rye) based on *Secale cereale* microsatellite markers. Genetics and Molecular Biology 29:685-691.

Singh R, Robbelen G (1975). Comparison of somatic Giemza banding pattern in several species of rye. IG Pfalnzenzuchtg 75:270-285.

Skuza L, Filip E, Szućko I (2015) Intergenic spacer length variability in cultivated, weedy and wild rye species. Open Life Science 10(1):175-181.

Skuza L, Rogalska SM, Bocianowski J (2007). RFLP analysis of mitochondrial DNA in the genus *Secale*. Acta Biologica Cracoviensia Series Botanica 49(1):77-87.

Spencer HA, Hawkes JG (1980). On the origin of cultivated rye. Biological Journal of the Linnean Society 13:299-313.

Stoeckle MY, Thalerd S (2014). DNA barcoding works in practice but not in (neutral) theory. PLoS One 9:e100755.

Stracke S, Schilling AG, Forster J, Weiss C, Glass C, Miedaner T, Geiger HH (2003). Development of PCR-based markers linked to dominant genes for male-fertility restoration in Pampa CMS of rye (*Secale cereale* L.). Theoretical and Applied Genetics 106:1184-1190.

Su X, Liu YP, Chen Z, Chen KL (2016). Evaluation of candidate barcoding markers in *Orinus* (Poaceae). Genetic Molecular Research 15(2):gmr.15027714.

Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE (2012). DNA barcoding in native plants of the *Labiatae* (Lamiaceae) family from Chios Island (Greece) and the adjacent Çesme-Karaburun Peninsula (Turkey). Molecular Ecology Resources 12:620-633.

Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, Roy S (2013). The Internal Transcribed Spacer (ITS) region and *trnH-psbA* are suitable candidate loci for DNA barcoding of tropical tree species of India. PLoS One 8:e57934.

Vences FJ, Vaquero F, de la Vega MP (1987a) Phylogenetic relationships in *Secale*: An isozymatic study. Plant Systematics and Evolution 157:33-47.

Vences FJ, Vaquero F, Garcia P, de la Vega MP (1987b). Further studies on phylogenetic relationships in *Secale*: On the origin of its species. Plant Breeding 98:281-291.

Xu DH, Sakai AJ, Kanazawa M, Shimamoto A, Shimamoto Y (2000). Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. Theoretical and Applied Genetics 101:724-732.

Yan HF, Hao G, Hu CM, Ge XJ (2011). DNA Barcoding in closely related species: A case study of *Primula* L. sect. Proliferae Pax (Primulaceae) in China. Journal of Systematics and Evolution 49:225-236.

Zhang B, Luo L, Zhang X, Li R, Song Y, Zhang D, … Wei Y (2011). Bioinformatics analysis on ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunits in different plant. Molecular Plant Breeding 2(15):101-108.

Zimmermann T, Bocksberger G, Brüggemann G, Berberich W (2013). Phylogenetic relationship and molecular taxonomy of African grasses of the genus *Panicum* inferred from four chloroplast DNA-barcodes and nuclear gene sequences. Journal of Plant Research 126(3):363-371.