TÜBİTAK

Turkish Journal of Agriculture and Forestry

http://journals.tubitak.gov.tr/agriculture/

Research Article

Turk J Agric For
(2015) 39: 730-741

doi:10.3906/tar-1405-93

Evaluation of pyrosequencing for large-scale identification of plant species (grasses as a model)

Nadia HAIDER*
Department of Molecular Biology and Biotechnology, Atomic Energy Commission of Syria, Damascus, Syria

Received: 22.05.2014 • **Accepted/Published Online:** 11.12.2014 • **Printed:** 30.09.2015

Abstract: Grasses (Poaceae) are among the most important and widely cultivated plants. Identification of grass species using morphological characters is a difficult task for nonspecialists, and it is not always an accurate method. DNA-based molecular markers have been widely used as an alternative towards more accurate identification of plant species. The molecular method that can be effective for large-scale identification of grass species, however, must satisfy requirements for specific amplification of DNA and reveal enough variability to distinguish species. The goal of this study was to evaluate the ability of pyrosequencing to detect SNP polymorphisms for an optimal discrimination of grass species. PCR amplifications and the sequencing of 19 fragments of chloroplast genes were performed using a pooled DNA template of 32 British native grass species to be identified and chloroplast DNA-specific universal primers. Based on amplicon size and sequence variance, the 6 most variable loci (*psbE&F*, *rbcL*.1, *ndhF*.1, *ndhF*.2, *ndhF*.3, and *clpP&rps12*) were selected to be targeted for pyrosequencing of the species analyzed. The pyrosequenced regions contained sufficient polymorphisms to allow the complete diagnosis of all species analyzed, except for *Elymus caninus* and *Elytrigia repens*, which had identical sequences unique from the remaining species. When targeted loci were pyrosequenced in representative species of the 2 economically important grass genera *Festuca* L. (8 species) and *Poa* L. (11 species), all species in each genus could be separated. Should the SNP markers developed here proved to be species-specific, they can provide a valuable tool for reliable diagnosis of grasses and applications (e.g., forensic botany) moving forward.

Key words: Grasses, identification, pyrosequencing, species-specific markers

1. Introduction

Of all crops, 70% are grasses (Poaceae, Graminaceae) (Constable, 1985). This makes them among the most important and widely cultivated plants on earth. Grasses include 4 of the 5 most important food crops (all cereals) and indirectly provide the primary food source for most of the world's livestock (forage grasses). They are also used for turf and amenity purposes and play a significant role in environmental protection (http://www.stabroeknews.com/2010/features/05/16/the-grass-family/).

The huge economic and ecological importance of the grass family, comprising about 10,000 species (Ward et al., 2009), coupled with the wide diversity of species it contains, means that there is a pressing need for a system allowing accurate species identification so that desirable and important species can be distinguished (Tzvelev, 1989). Applications such as the breeding of economically important cereals and forage grasses or lawn and turf management rest on the fundamental need for accurate species identification. Grass-specific molecular markers may also have value for a small number of forensic cases, where grass samples (or even grass stains on clothing) are used to tie a suspect or victim to the crime scene (Ward et al., 2009).

The grass morphology (leaf sheaths, leaf blades, ligules, the inflorescence, spikelets, glumes, lemmas, and anthers) can be useful for grass identification (Wheeler et al., 1990). Grasses are most easily identified when they are in flower, using a combination of vegetative and floral features. The structure of the flower heads is, however, quite complex and must be clearly understood (http://www.countrysideinfo.co.uk/grass_id/index.htm). This problem has been exacerbated by repeated hybridization and polyploidy during the evolution of the group (Stebbins, 1956).

Identification keys (e.g., Krishnan et al., 2000; http://www.landcareresearch.co.nz/resources/identification/plants/grass-key) and supporting computer systems (e.g., Fermanian and Michalski, 1989; Ward et al., 2009) and programs (Aiken et al., 1996) have been developed to assess biodiversity and support recognition of various grass

* Correspondence: ascientific3@aec.org.sy

species. However, existing identification keys are based upon artificial classifications and are incomplete. This also means that diagnosis of grass species using morphological characters is a difficult task for nonspecialists, as it can be carried out only in flowering time and it is not always accurate. Hence, molecular markers are an alternative strategy towards more accurate identification of grass species, since they are more stable within species (Haider and Nabulsi, 2008).

Pyrosequencing technology (Ronaghi et al., 1998) is sequencing-by-synthesis technology that employs 4 enzymatic reactions in a single-tube assay to measure DNA synthesis and quantitatively monitor nucleotide incorporation in real time (Salk et al., 2006). It is a unique sequencing method that was developed as an alternative to classical DNA sequencing for short- to medium-read applications (Novais and Thorstenson, 2011). The technique is built on 4-enzyme real-time monitoring of DNA synthesis by bioluminescence using a cascade that, upon nucleotide incorporation, ends in a detectable light signal (bioluminescence) (Ahmadian et al., 2006). It has been argued that pyrosequencing is a solution that offers high discriminatory power in mitochondrial DNA (mtDNA) analysis and provides a fast, real-time read-out. It is also a reliable and robust method that reduces turnaround times. In the present study, the applicability of pyrosequencing in order to genotype single nucleotide polymorphisms (SNPs, single-base variations in the genetic code) in grasses was evaluated by using universal primers that amplify very short but informative DNA fragments of the chloroplast DNA (cpDNA) for rapid molecular identification of grasses. Loci selected should amplify a wide range of grasses and should possess sufficient sequence variability to enable diagnosis of species targeted. The developed SNP markers for targeted species could work in conjunction with the morphological identification.

2. Materials and methods

2.1. Collection of plant material, identity validation, and DNA extraction

Fresh or dry leaf material of 32 British native grass species, as described by Stace (1997) and representing 32 genera, was collected to test the efficiency of pyrosequencing using universal primers for identification of targeted species. Table 1 lists names, sources, and years of collection (for fresh samples) or preservation (for dry samples) of all targeted grasses.

Dry leaf samples of 10 species of *Poa* L. and 8 species of *Festuca* L. (Table 2) were also collected from the herbarium

Table 1. Names of British native grass species representing 32 genera used in this study, their sources, the year of their preservation, and the name of their collector.

Species name	Source, collector and year of preservation
Arrhenatherum elatius (L.) P.Beauv. ex J. & C.Presl.	The field and greenhouses of the School of Biological Sciences, University of Reading, Reading, UK
Phalaris arundinacea L.	
Brachypodium sylvaticum (Huds.) P.Beauv.	Bradfield, Berkshire, UK
Dactylis glomerata L.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 17711 Flora of France, G. Certa (1995)
Polypogon monspeliensis (L.) Desf.	RNG, H. J. M. Bowen 10098 (1999)
Leymus arenarius Hochst.	RNG, PMD Etherington 88108 (1988)
Hordelymus europaeus (L.) Jess. ex Harz	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 15810, France (1991)
Corynephorus canescens (L.) P.Beauv.	Spain CA 5171, S-VIII (1999)
Milium effusum L.	Italy, Optima Itter VIII 1100 (1997)
Festuca rubra L. 'Trump'	The field and greenhouses of the School of Biological Sciences, University of Reading
Lolium perenne L.	cultivar AberImp (2005)
Vulpia unilateralis (L.)	Flora of Spain, det. HJM Bowen et al., 1795 (1988)
Puccinellia maritima (Huds.)	Flora of the British Isles, D. Doogue 8062 (1991)
Briza media L.	Macedonia, A.J. Richards et al., 116 (1999)
Poa pratensis L.	Sports Turf Research Institute (STRI) cultivar Cesar (2005)
Glyceria maxima (Hartm.) Holmb.	Bradfield, Birkshire, UK
Sesleria caerulea (L.) Ard.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 18772 (1998)
Melica uniflora Retz.	Flora of Italy, Optima Itter VIII 379 (1997)
Helictotrichon pubescens (Huds.) Pilg.	Flora of Slovinia, HJM Bowen 9531, det. Humphry Bowen (1997)
Koeleria vallesiana (Honck.)	Flora of the British Isles, coll. RM Payne v. c. 6 (1984)
Deschampsia cespitosa (L.) P.Beauv.	Cultivar Barcampsia (2005)
Holcus mollis L.	The field and greenhouses of the School of Biological Sciences, University of Reading
Hierochloe odorata (L.) P.Beauv.	Flora of Suecica, LIV 38 (1974)
Agrostis gigantea Roth	Flora of Morocco, S. L. Jury 11240 (1993)
Calamagrostis purpurea (Trin.) Trin.	Finland, Raino Lampinen 9708 (1990)
Gastridium ventricosum (Gouan) Schinz & Thell.	Spain, Optima Itter 1496 VI, Canti F (1994)

Table 2. Names of *Festuca* and *Poa* species used in this study, their sources, the year of their preservation, and the name of their collector.

Species name	Source, collector, and year of preservation
Festuca pratensis Huds.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 4052, France, JP. Theurillat (1986)
F. altissima All.	Italy, Optima Itter VIII 1257 (1997)
F. arenaria Osbeck.	Suffolk, det. CA Stace (1991)
F. ovina L.	STRI cultivar Quatro (2005)
F. vivipara (L.) Sm.	Mid Pearthshire v. c. 88, Coll. RM Payne, det. CA Stace (1984)
F. filiformis Pourr.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 10991, France (1983)
F. longifolia Thuill.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 14794 (1991)
F. rubra L. 'Trump'	The field and greenhouses of the School of Biological Sciences, University of Reading
Poa infirma Kunth.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 81/17, leg. et det. L. Lambinon (1983)
P. annua L.	STRI sample, 102 (2005)
P. pratensis L.	STRI cultivar Cesar (2005)
P. angustifolia L.	Societe Pour L'echange Des Plantes Vasculaires De L'europe Et Du Bassin Mediterraneen no 8878, R. Lampinen (1993)
P. flexuosa Sm.	Flora Norvegica, leg. J Ax Namfelett (1934) (the only year collection)
P. compressa L.	Italy, Optima Itter VIII 1454 (1997)
P. glauca Vahl.	Flora of Canada, SP. Thornton-Wood 115 (1991)
P. nemoralis L.	Italy, Optima Itter VIII 1936 (1997)
P. bulbosa L.	Optima Itter VIII 132 (1997)
P. alpina L.	Italy, Optima Itter VIII 1921 (1997)
P. trivilis L. 'Solo'	The field and greenhouses of the School of Biological Sciences, University of Reading

of the School of Biological Sciences, University of Reading, to detect the level of SNP variation within targeted regions among species of each of the 2 genera, including the representative samples of the 2 genera (i.e. *P. pratensis* and *F. rubra*, Table 1). Great care was taken to ensure the authenticity of the identity of all specimens, which were verified using the diagnostic key given by Stace (1997) for the family Poaceae.

DNA was isolated from all leaf samples using the DNeasyTM Plant Mini Kit (QIAGEN, UK) according to the manufacturer's instructions. Recovered DNA pellets were dried under laminar flow and then resuspended in 150 µL of double-distilled and sterilized water. DNA was quantified using Gene Quant Spectrometer (Amersham Biosciences, UK) and the concentration of each DNA template was set at 10 ng/µL. A pooled mix of all DNA templates was prepared for the initial sequencing screen for locus variability (Haider and Wilkinson, 2011).

2.2. Selection of variable loci of the chloroplast genome for the design of pyrosequencing primers

2.2.1. PCR and sequencing

PCR amplification of 19 fragments of chloroplast genes (or gene clusters) (Table 3) was performed using the pooled DNA template and cpDNA-specific universal primers developed for grasses by Haider (2003). The highly variable chloroplast noncoding region *trnL-F* (Taberlet et al., 1991) was also amplified for comparative purposes (Haider and Wilkinson, 2011). The PCR protocol described by Haider and Wilkinson (2011) for preparation of PCR reactions and PCR temperature cycling was followed. Amplification products were visualized on 1.5% agarose gel that was subjected to electrophoresis at 120 V for 30 min. PCR products generated were purified using either the NucleoSpin Kit or the QIAGEN Kit and were then sequenced with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit (produced by PE Applied Biosystems) as mentioned by Haider and Wilkinson (2011).

2.2.2. Design of grass-universal primer sets for PCR and pyrosequencing

Based on amplicon size and sequence variance, the 8 most variable loci were selected to be targeted for pyrosequencing. These were *rps4*, *psbE&F*, *rbcL*.1, *ndhF*.1, *ndhF*.2, *ndhF*.3, *23S,4.5S&5S*, and *clpP&rps12*. Sequences generated from the pool sample for only 2 of those loci (*psbE&F* and *23S,4.5S&5S*) were used for the design of grass-universal pyrosequencing primers, comprising 3 primers for each locus to be pyrosequenced. Primers targeting the other 6 loci were designed using alignments of homologous sequences, which were retrieved through a BLASTn search for available grasses in the NCBI database (http://blast.ncbi.nlm.nih.gov/) for each locus. Sequences recovered for each locus were aligned using Clustal X

Table 3. DNA sequences, annealing T °C, and expected product size of grass-universal primers (Haider, 2003) used to amplify the selected 19 chloroplast regions. Numbers 1, 2, and 3 refer to the portion of locus.

Locus name	Primer sequence/5`-3`	Annealing T/°C	Expected product size (bp)
rps4.F	atggtctgacagaacgac	52	383
rps4.R	accgacccacttcctatc		
psbE&F.F	gggatcaatatcagcaatgc	55	508
psbE&F.R	cctttccgctctaatcag		
rpoA.2.F	ggaagttatcctatagatgctg	53.5	489
rpoA.2.R	gatgcttctctagagtgtcc		
rbcL.1.F	atatcttggcagcattccgag	55	524
rbcL.1.R	tctctccagcgcataaatgg		
rbcL.2.F	accatttatgcgctggaga	55	284
rbcL.2.R	ctgtctatcaataactgcatgc		
ndhF.1.F	ttccttagaatttggatacttg	52	298
ndhF.1.R	agaccaaaatcccctacacg		
ndhF.2.F	ctactttagctcttgctcaga	55	316
ndhF.2.R	gaccagaagcaagcaagagg		
ndhF.3.F	ctcttgcttgccttctggtc	55	393
ndhF.3.R	gtgtttttagccccaaagg		
23S,4.5S&5S.F	tctcctccgacttccctag	52	393
23S,4.5S&5S.R	accatgaacgaggaaaggc		
rpl2&trnH.1.F	acaaaacacctatcccgagc	55	272
rpl2&trnH.1.R	atccccatagtgtatgagac		
rpl2&trnH.2.F	tcataacatgagcgattcgg	55	564
rpl2&trnH.2.R	ttccaagcgcaggataacc		
psbE-psbF-orf38-orf40.F	tgacaatctcacaaggatga	55	365
psbE-psbF-orf38-orf40.R	gcctcctaaaaggatctac		
clpP&rps12.F	aggggatgaaatgcttgacg	55	608
clpP&rps12.R	agctgaccctgttagtccg		
psbK-psbI-trnS.1.F	tgtgtttgggaaagtccga	55	534
psbK-psbI-trnS.1.R	ggttaggtggttaggtattg		
rpl23&rpl2.2.F	atccactcatgtggtacttc	55	353
rpl23&rpl2.2.R	tacccacttgtccgactg		
psbC.F	gaacctatttgaagtggccc	55	487
psbC.R	acccaaccatacatgtccac		
orf62.F	catgattaaaccccctcctac	58.5	414
orf62.R	gagaagagctagcaattcg		
clpP.1.F	cacgataccaaggcaaacc	55	389
clpP.1.R	gagaaagaaagaatccgcag		
clpP.2.F	ccggagatgaagaagcgac	55	434
clpP.2.R	acccttgtgatcatttcgcg		

software, and conserved regions in each alignment were displayed using GeneDoc software (Version 2.7.000, NRBSC, USA). Conserved regions within alignments were identified and used to design pyrosequencing primer binding sites that flank more variable regions. Even though some binding sites showed some variation among the aligned sequences, no degeneracy was permitted in primer design. In such cases, the most common base identity at the variable position was used for primer design (Haider and Wilkinson, 2011).

To assess the applicability of pyrosequencing for specific SNP detection in grass species, locus-specific PCR and pyrosequencing primer sets were designed using the Pyrosequencing Assay Design Software (Biotage). Pyrosequencing assays were composed of 3 primers. Two were designed to allow PCR amplification of the targeted region, with one of these primers modified using 5' biotin labeling to enable preparation of a single-stranded template in further stages of the pyrosequencing analysis. PCR primer pairs were designed to flank the SNP site to be targeted in the locus and to generate amplicons between 150 and 200 bp in length. In addition to these, a sequencing primer targeting the immediate adjacent region (at its 3' OH end) of the first SNP variant within the sequence flanked by the forward and reverse primers was also designed. This primer was modeled on the same strand as the nonbiotinylated primer and was used to generate pyrosequences. One set of PCR primers was designed for each of 7/8 sequenced regions that flank variable SNPs. These were *psbE&F*, *23S,4.5S&5S*, *ndhF*.1, *ndhF*.2, *ndhF*.3, *rps4*, and *clpP&rps12*. For *rbcL*.1, two sets of primers were designed that amplify different portions of the locus (referred to as portion 1 and portion 2).

Beside the criteria that were mentioned by Haider and Wilkinson (2011) for PCR primer design, the 2 other crucial criteria that were considered for the choice of PCR primers were: 1) the distance between PCR primer annealing sites and the SNP, which should be large enough (as an amplicon) to be evaluated by agarose gel electrophoresis and to ensure high product yield; and 2) the region flanked by the primers, which should allow for the selection of the optimal genomic flanking sequence for annealing of the sequencing primer (Ringquist et al., 2005). After the selection of the PCR primer pairs for each locus, one primer that would be used to initiate sequencing of the amplification product was designed. Sequencing primers were chosen on either side of the polymorphism, and the binding site of the sequencing primer was selected by placing the 3' end of the primer as near to the polymorphic residue (SNP) as possible (Ringquist et al., 2005). One sequencing primer was designed for each locus except for *ndhF*.3 (3 primers). All primers including those 5' end biotinylated were purchased from Biomers.net (Germany). Sequences of all primers designed are given in Table 4.

2.3. Testing the efficiency of the designed primers

2.3.1. PCR

Primers that were developed for amplification of the 8 chloroplast loci to be pyrosequenced in DNA samples of all targeted grass species were tested in gradient PCRs to define the optimum annealing temperature. Those loci were then PCR-amplified in a final reaction volume of 50 µL as described by Haider and Wilkinson (2011). Cycling conditions mentioned earlier were applied using an annealing temperature of 54.9 °C for all loci except for *clpP&rps12* (51.9 °C). Generated amplicons were visualized on 2% agarose gel that was subjected to electrophoresis at 100 V for 30 min. The samples were then stored at –20 °C or 4 °C prior to pyrosequencing.

Table 4. DNA sequences of pyrosequencing primers developed and names of loci they target. F, R, and S refer to the forward, reverse, and sequencing primers, respectively.

Chloroplast locus name	Primer sequence 5'-3'
23S,4.5S&5S	F- agacagcgacgggttctcc R-biotin gtctcgccgctaccttatcct S- ctgtcgctccatccc
psbE&F	F- tcttatcccatttggaaggatacc R1-biotin tcaagtggtcatgctagagacaa S- cccatttggaaggata
ndhF.1	F- gcaactatggtagctgcggg R-biotin cctaagaagagtgttattgtacc S- cttcttcctcttttcata
ndhF.3	F-biotin gttaaccctattagcaaagtgg R- cgaaccacgcttatatgacca
(portion 1)	S1- cgaaccacgcttatatgacc
(portion 2)	S2- gaagctatatgctatgaata
(portion 3)	S3- aagaattcgaactttcctgg
rps4	F-biotin cgaaggataaccaacgatca R- tcaaccgtcaaatgctttgg S1- tcaaccgtcaaatgctttg
clpP&rps12	F- atgtttaaggattggtagtg R-biotin gatgattcctcatcatgaag S- atgtttaaggattggtagt
rbcL.1	F- cgatgctatcacatcgagcc R-biotin gaaccctcttcaaataggtc
(portion 1)	S- tgagcc tgtt gctgggga
rbcL.1	F- tttgtgttttgtgccgaagc R-biotin cattcaagtaatgccccttg
(portion 2)	S- ttgtgttttgtgccgaag
ndhF.2	F- tttactgcaggattaaccgc R-biotin tgaacacgcaggtacccatc S- tactgcaggattaaccg

2.3.2. Detection of SNP variation in targeted loci by pyrosequencing

PCR products were prepared for pyrosequencing analysis following the manufacturer's instructions (Pyrosequencing Biotage). One of the primers used for PCR was biotinylated to allow immobilization of the PCR product on streptavidin-coated beads and preparation of the single-stranded DNA fragment required for pyrosequencing analysis. Sample preparation was done in 96-well format with the use of a Vacuum Prep handset tool (Biotage, Sweden), according to the instructions of the manufacturer, by using streptavidin-coated beads (Amersham Biosciences, USA). Primer annealing was done at 80 °C for 2 min according to the instructions of the manufacturer. Pyrosequencing was performed at 25 °C in a total volume of 50 µL in an automated 96-well pyrosequencer according to the instructions of the manufacturer on a Pyrosequencing Analyser PSQ96MA 2.1 and using SQA Reagent Kits (Pyrosequencing Biotage). Data analysis was performed with PyroMark ID software. Recovered sequences were edited and aligned and variable SNPs among samples for each locus analyzed were displayed using GeneDoc software.

2.3.3. Detection of SNPs variation within two grass genera

Loci *psbE&F*, *rbcL*.1 (portion 2), *ndhF*.1, *ndhF*.3 (portion 2), and *clpP&rps12*, which allowed the identification of the majority of the targeted species, were amplified and pyrosequenced as described above using the primers designed and DNA samples that represented all species of *Festuca* (8 species) and *Poa* (11 species) (Table 2) for which plant material was available at the herbarium of the School of Biological Sciences at the University of Reading, Reading, UK.

3. Results

3.1. PCR, sequencing, and development of pyrosequencing grass-universal

All primer pairs used to amplify the targeted 20 loci generated distinct and strong band amplicons of the appropriate size (Haider, 2003) (Table 3) when applied to the pool template of DNAs from the 32 species analyzed. It was also possible to generate sequences of good quality for all PCR products. These sequences showed the highest BLASTn match to the corresponding site on the chloroplast genome in grasses. For example, the sequence generated for *reps4* exhibited 96% homology to the equivalent site on the *D. giganteus* (Z29235.1).

Based on amplicon size and sequence variance (Haider and Wilkinson, 2011), the most variable loci were *rps4*, *psbE&F*, *rbcL*.1, *ndhF*.1, *ndhF*.2, *ndhF*.3, *23S,4.5S&5S*, and *clpP&rps12*. When the sequences generated for those loci were checked to select pyrosequencing primers' binding sites that flank variable SNPs, only loci *psbE&F* and *23S,4.5S&5S* had the same SNP variants in both forward and reverse sequences. Therefore, primers to be developed for the pyrosequencing of these 2 loci were designed based on sequences generated for them. As for the remaining 6 selected loci, alignments of cpDNA sequences found in the NCBI database for grasses, known to be homologous to those loci, were used for primer design. The number of species involved in these alignments varied between 5 and 35 and typically included representatives of the grass family.

3.2. PCR and pyrosequencing

All PCR primer pairs designed for amplification of loci to be pyrosequenced amplified a single and clear amplicon (Figure 1). As seen using agarose gel electrophoresis, there was no size variation among species for any of the targeted regions except for locus *clpP&rps12*. It was possible to achieve universal amplification across all 8 chloroplast regions targeted using only subtle changes to annealing temperature, with 54.9 °C applied for all loci except for *clpP&rps12* (51.9 °C).

When the obtained PCR products were pyrosequenced for the 32 grass species, sequencing primers that target region *23S,4.5S&5S* did not give any pyrosequencing signals, and pyrosequencing profiles generated for *rps4* had poor quality. Sequencing primers that were developed for pyrosequencing of the remaining 6 loci (*psbE&F*, *rbcL*.1 (portions 1 and 2), *ndhF*.1, *ndhF*.2, *ndhF*.3 (only portions 1 and 2), and *clpP&rps12*) worked efficiently and generated very clear pyrograms. Very little variation among targeted species was observed for portion 1 of *rbcL*.1, *ndhF*.2, and portion 1 of *ndhF*.3. Therefore, SNP variation analysis for identification of these species was based on *psbE&F*, *rbcL*.1 (portion 2), *ndhF*.1, *ndhF*.3 (portion 2), and *clpP&rps12*.

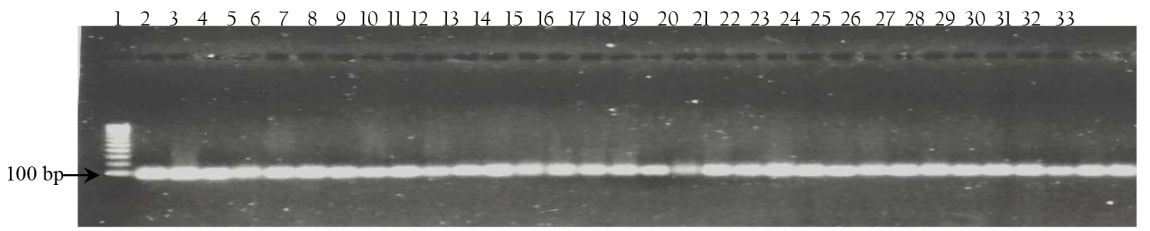

Figure 1. Amplification products of *psbE&F* in the 32 species analyzed (lanes 2–33). Lane 1 is a 100-bp DNA ladder.

Generally, the sizes of these 5 loci ranged from 18 bp (*psbE&F*) to 34 bp (*ndhF*.1). Sizes of the other loci were 28 bp (*rbcL*.1, portion 2), 32 bp (*ndhF*.3, portion 2), and 21 bp (*clpP&rps12*). Size variation observed among species for amplicons of *clpP&rps12* was confirmed by pyrosequencing since some species yielded pyrograms that were 11 bases longer than those of the other species. The most variable locus was *ndhF*.1 and the least was *rbcL*.1 (portion 2). The majority of SNP polymorphisms were detected in the same positions of predicted SNPs in the sequence used for the design of pyrosequencing primers.

The 34-bp pyrosequence for *ndhF*.1 contained sufficient polymorphisms (nucleotide substitutions and indels) to allow complete diagnosis of 13 of the 32 species analyzed. These were *B. sylvaticum*, *H. europaeus*, *C. canescens*, *M. effusum*, *G. maxima*, *M. uniflora*, *H. pubescens*, *D. cespitosa*, *H. mollis*, *P. maritima*, *H. murinum*, *P. pratensis*, and *B. ramosa*. The remaining 19 species that were not identified were grouped into 4 groups as follows: 1) *S. caerulea*, *D. glomerata*, *L. perenne*, *F. rubra*, *V. unilateralis*, and *Ph. pratense*; 2) *L. arenarius*, *E. caninus*, and *E. repens*; 3) *H. odorata*, *K. vallesiana*, *Ph. arundinacea*, *A. elatius*, *G. ventricosum*, *C. purpurea*, and *B. media*; and 4) *A. gigantea* and *P. monspeliensis*. Species in each of these groups were identical to each other but different from all remaining species. Analysis of pyrosequences of the other 4 loci (*psbE&F*, *rbcL*.1 (portion 2, Fig 2), *clpP&rps12*, and *ndhF*.3 (portion 2)) allowed the separation of species in each of these groups apart from *E. aninus* and *E. repens*.

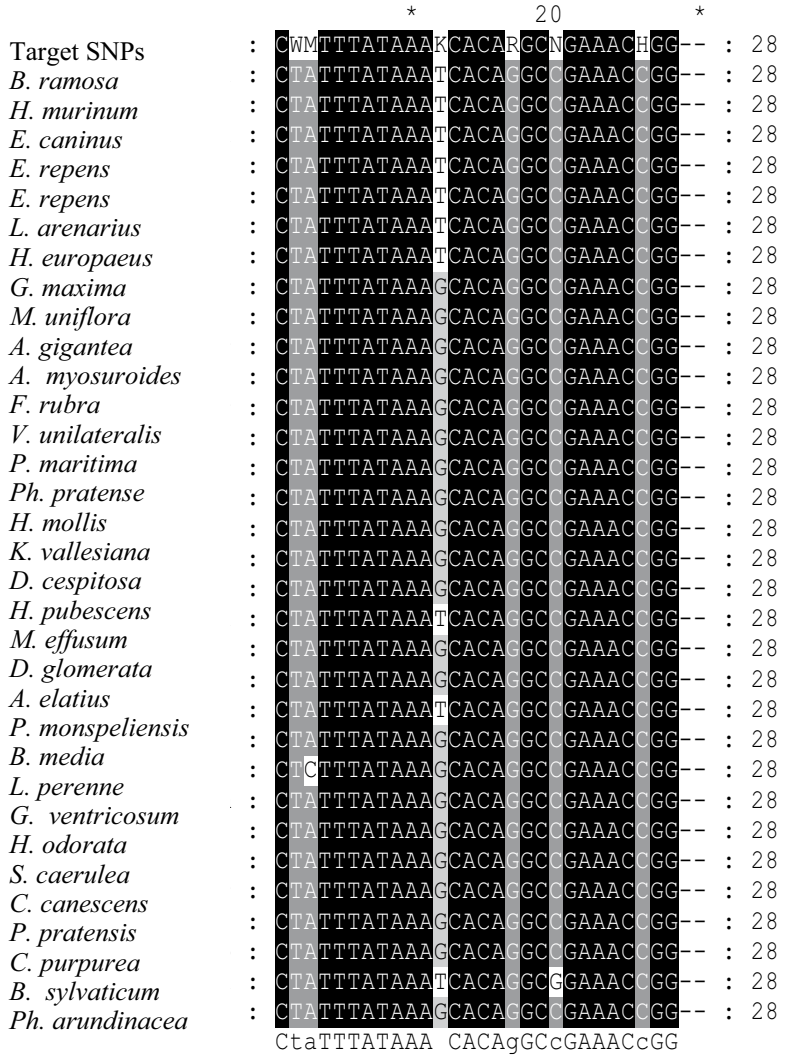

Figure 2. A portion of pile up of pyrosequences generated for *rbcL*.1 (portion 2) in 32 grass species (names are on the left side). Each of the numbers on the right side of the pile up refers to the position of the last nucleotide that appeared at the end of the portion aligned in the pyrosequences generated for the locus in corresponding species.

3.3. Variation within Festuca and Poa

When the 5 loci *psbE&F*, *rbcL*.1 (portion 2), *ndhF*.1, *ndhF*.3 (portion 2), and *clpP&rps12* were amplified in the targeted species of *Festuca* and *Poa* to detect SNP variation for the diagnosis of species in each of the 2 genera, the used primer sets amplified single and clear amplicons of the expected sizes in all template DNA samples. SNP variation was detected among species of each of the 2 genera. Based on SNP variation in loci *psbE&F*, *rbcL*.1 (portion 2), *ndhF*.1, and *clpP&rps12*, all 11 species of *Poa* could be identified. When *clpP&rps12* was pyrosequenced in those species, length variation was detected. As for *Festuca* species, all 8 species could be separated based on pyrosequences of loci *rbcL*.1 (portion 2), *ndhF*.1, and *ndhF*.3 (portion 2). It was also possible in both genera to observe the level of allele conservation among congeneric species.

4. Discussion

In Poaceae, molecular-based techniques have been employed mainly to study the origin and diversity of species and genera (e.g., Li and Ge, 2001; Kumar et al., 2009) and evaluate the genetic relationships (e.g., Liu et al., 2010). Few attempts, however, have been made for developing distinctive species-specific markers for the identification of grass species. Some of these markers were based on nuclear DNA (nDNA) using inter-SSR (Paäakinskiene et al., 2000), microsatellites (Studer et al., 2006), RAPD (Chandra and Dubey, 2010), and SCAR (Scheef et al., 2003). Generally, the analysis of nDNA has been so far limited to a few grass species, mainly cereals such as wheat, rye, rice, barley, and maize (e.g., Gealy et al., 2002), and their wild relatives (Haider et al., 2010b).

Studies based on the cpDNA of grasses, which has been proved to be the most suited for plant species identification (Ford et al., 2009), have similarly centered on addressing systematic problems at all levels (e.g., Gillespie and Soreng, 2005), with the most effort being focused on important cereal crops (e.g., Haider and Nabulsi, 2008). Hence, few studies have exploited cpDNA polymorphisms for diagnostic purposes, such as those of Parani et al. (2001, PCR-RFLP of *trnS-psbC*), Ridgway et al. (2003, PCR-RFLP of the *trnL* intron), Zapiola et al. (2010, microsatellite markers), and Wallinger et al. (2012, PCR amplification of *trnT-F* region).

The molecular method can be effective for large-scale identification of grass species; however, it must satisfy requirements for specific amplification of DNA, reveal enough variability to distinguish species (Ridgway et al., 2003), and, at the same time, generate highly conserved markers within species. It should also be a fast, reliable, cost-effective, and high-throughput method. Methods based on the analysis of SNPs, which are the most abundant molecular markers and the most common type of sequence variation in plant genomes (Ching et al., 2002), are among these methods. Traditionally, SNPs in cpDNA have been analyzed by conventional DNA sequencing of certain DNA loci (DNA barcoding) (Ford et al., 2009). Although DNA sequence analysis using the conventional Sanger sequencing (chain termination sequencing) is reliable and very robust, it is time-consuming and labor-intensive (Andréasson et al., 2002). In contrast, SNP analysis of cpDNA based on pyrosequencing technology (Langaee and Ronaghi, 2005), which has been developed to allow for high-quality sequencing of short stretches of DNA (Alexander et al., 2005), is faster, provides more quantitative allele ratio information, and does not require labeled primers or labeled nucleotides or electrophoresis (Salk et al., 2006). However, the SNP must be known in advance. Added to that, pyrosequencing is simpler to implement and requires significantly less DNA than Sanger sequencing (Hastings, 2004). It also offers higher specificity and sensitivity in analysis of genomic DNA (Wang et al., 2007), much higher throughput, and lower cost than Sanger sequencing (Weber et al., 2007). Additional advantages include flexibility, high accuracy, and ability to automate sample preparation (Novais and Thorstenson, 2011). Furthermore, no biological cloning is required (Weber et al., 2007). One of the many unique advantages of pyrosequencing over other DNA sequencing technologies is that this technique can be automated for large-scale screenings (Gharizadeh et al., 2003). Fakruddin and Chowdhury (2012) reported that by increasing the read length and shortening the sequence reaction time per base calling, pyrosequencing may take over many broad areas of DNA sequencing applications. The authors performed a comprehensive review of the literature on the principles, applications, challenges, and prospects of this technology.

Advances in pyrosequencing methodology, including multiplex and universal primer applications, have reduced assay cost and improved throughput (Marsh, 2007). Pyrosequencing has been developed to generate hundreds of thousands of sequences at an exceptional speed; the latest pyrosequencing platform by 454 Life Sciences can generate 400 Mb in a 10-h run with a single machine (Nilsson et al., 2009). Alexander et al. (2005) presented software that can efficiently design PCR and pyrosequencing primers for large numbers of SNPs.

Since the invention of pyrosequencing, hundreds of articles have been published describing its different applications, most notably for DNA structure variation and microbial detection (Ronaghi et al., 2007; Siqueira et al., 2012). Pyrosequencing has been applied narrowly to plants for purposes such as SNP genotyping (Silvar et al., 2011), rapid sequencing of complete chloroplast genomes (e.g., *Magnolia officinalis*) (Li et al., 2012), and

description of spatial and temporal patterns of algal diversity (Steven et al., 2012). Few attempts, however, have used pyrosequencing for the identification of plant species in small-scale research projects with a limited number of species. Examples are those of Leem et al. (2005) for the identification of 2 ginseng species (*Panax ginseng* and *P. quinquefolius*), Soininen et al. (2009) for deciphering the composition of complex plant mixtures in the diet of small herbivores, and Han et al. (2005) to identify the *Akebia* and *Aristolochia* species (*Akebia quinata* and *Aristolochia manshuriensis*). This is the first report that reveals the power of pyrosequencing for large-scale identification of plant species (grasses).

Ronaghi and Elahi (2002) tested pyrosequencing technology for analysis of mtDNA polymorphisms with regard to reproducibility and sensitivity when applied to control samples and actual casework materials. The results showed that the method is sensitive and very accurate and that the data that it generates are easily interpreted. The authors concluded that using pyrosequencing for detection of mtDNA polymorphisms resulted in an optimal discrimination power in relation to the number of bases determined. In this study, the efficiency of pyrosequencing has been evaluated for the first time for the detection of SNP variation in the chloroplast genome among British native grasses representing 32 genera; it is not feasible to sample all known grass species. It is worth noting here that the primary objective of this study is to validate the efficiency of pyrosequencing for identification of plant species rather than to develop species-specific SNP markers for the grasses targeted.

The pyrosequencing-based identification system presented here was designed such that the PCR assays targeted informative variable sequences of the cpDNA. Hence, 19 chloroplast genes (or gene clusters) that were proved by Haider (2003) to be useful for grass species identification were targeted in the pooled DNAs of the grasses examined. These primer pairs were able to generate locus-specific amplicons of the appropriate locus and size in the 32 species analyzed using the same conditions suggested by the author. When the sequences of those amplicons were analyzed, the most variable loci were *rps4*, *psbE&F*, *rbcL*.1, *ndhF*.1, *ndhF*.2, *ndhF*.3, *23S,4.5S&5S*, and *clpP&rps12* (Haider, 2003). In previous studies, one or more of these loci proved useful for the identification of different plant species such as those of *Aegilops* and *Triticum aestivum* L. (*rps4* and *psbE&F*) (Haider and Nabulsi, 2008), *Vicia* subgenus *Vicia* (Fabaceae) (*23S,4.5S&5S*) (Haider et al., 2012), Orchidaceae (*23S,4.5S&5S*) (Haider et al., 2010a), and the B-genome donor of *T. aestivum* (*psbE&F*, *ndhF*.1, and *ndhF*.2) (Haider, 2012).

SNP markers observed in the 6 chloroplast loci pyrosequenced collectively generated sufficient polymorphisms to allow the diagnosis of the vast majority of samples used in the study.

To assess the utility of the pyrosequencing-based identification system developed here for diagnosis of closely related grass species, PCR and pyrosequencing were performed on these loci in representative species of each of the 2 grass genera *Festuca* and *Poa*. The 2 genera were selected because they include important grasses. *Festuca* contains many fescues that are grown as lawn and golf-course grasses and for pasture and hay. Fescue species are also valuable breeding material, but the breeding of these species is complicated by morphological similarity among species and high heterogeneity within species (Armonienė et al., 2010). Identification and classification of numerous *Festuca* species is still a difficult problem due to their close morphological resemblance. The most difficult fescues to identify belong to the *F. ovina* aggregate, which is the largest group in the genus (Galli et al., 2006). Stace et al. (1992) also reported that *F. ovina* and *F. rubra* aggregates are very variable and very important from both ecological and economic points of view. They argued that it is vital that plants of such importance be identified correctly and that characters often used to separate the 2 aggregates are unreliable, therefore frequently causing misidentification (Stace et al., 1992). In this study, it was possible to identify all 8 *Festuca* species targeted based on pyrosequencing of chloroplast coding loci *rbcL*.1 (portion 2), *ndhF*.1, and *ndhF*.3 (portion 2), which are more likely to be species-specific than markers used in previous studies (Galli et al., 2006; Armonienė et al., 2010).

As for *Poa* (meadow grasses or bluegrasses), which is the largest grass genus (Soreng et al., 2010), many species of this genus are important pasture plants that are used extensively by grazing livestock. *P. pratensis* (Kentucky bluegrass) is a premier species for lawns, sports fields, and golf courses (Beard, 1973). Bor (1952) asserted that the genus is evolutionarily complex and taxonomically difficult, and Kellogg (1985) stated that morphological characters used for classification of species belonging to the genus are overlapping. Hybridization and polyploidization have also played a prominent role in the evolution of this complex genus (Patterson et al., 2005). Therefore, most studies focused on the classification and phylogeny of the group based on cpDNA (Soreng, 1990; Gillespie and Boles, 2001) or combination of cpDNA and nDNA (Patterson et al., 2005; Soreng et al., 2010). This study conducted the first attempt, however, to identify the targeted species of *Poa*. Based on SNP variations in coding loci *psbE&F*, *rbcL*.1 (portion 2), *ndhF*.1, and *clpP&rps12*, all 11 species of *Poa* could be differentiated. If this finding is applied generally, it can be reasoned that these markers have huge potential for large-scale and relatively low-cost identification of grass species.

Although Ward et al. (2009) developed a grass molecular identification system for forensic botany, it was based on the analysis of noncoding chloroplast and mitochondrial loci, which are less likely to generate species-specific markers compared to the coding loci targeted here. Wallinger et al. (2012) also revealed that an overall reliable sequence alignment of all studied species, including grasses, was impossible due to the great length differences in noncoding regions even within closely related taxa. Should the SNP markers developed here prove to be species-specific, they would provide a valuable tool for reliable diagnosis of grasses. Such grass-specific markers can provide significant evidence during criminal investigations (Ferri et al., 2009) since very short DNA fragments, like those that remain in degraded DNA samples, were targeted in this study. They can also be useful for studies on the detection of chloroplast capture, which is usually caused by several crosses between 2 different species, by determining the mode of the cpDNA inheritance (Gulsen et al., 2005), the maternal parent of the hybrid and polyploid species (Mason-Gamer, 2004), and maintenance of germplasm collections. These markers could also have direct and immediate applications in selection and breeding studies (Kindiger and Conley, 2009).

Due to the high level of conservation of sequences of cpDNA coding regions among the different plant species, the pyrosequencing primers that proved useful here for the identification of grass species may also prove more generally useful in the distinction of plant species in other plant families.

Acknowledgments

This research was conducted at and funded by the School of Biological Sciences, University of Reading, Reading, UK. Many thanks to the Director General of the AEC of Syria and the Head of the Department for their support. I would also like to thank Prof MJ Wilkinson (School of Agriculture, Food, and Wine, University of Adelaide, Australia) and Dr J Allainguillaume (Institute of Biological, Environmental, and Rural Sciences, Aberystwyth University, Aberystwyth, UK) for their advice and help.

References

Ahmadian A, Ehn M, Hober S (2006). Pyrosequencing: history, biochemistry and future. Clin Chim Acta 363: 83–94.

Aiken SG, Consaul LL, Dallwitz MJ (1996). Grasses of the Canadian Arctic Archipelago: a DELTA database for interactive identification and illustrated information retrieval. Can J Bot 74: 1812–1825.

Alexander AM, Pecoraro C, Styche A, Rudert WA, Benos PV, Ringquist S, Trucco M (2005). SOP3: A web-based tool for selection of oligonucleotide primers for single nucleotide polymorphism analysis by pyrosequencing. Biotechniques 38: 87–94.

Andréasson H, Asp A, Alderborn A, Gyllensten U, Allen M (2002). Mitochondrial sequence analysis for forensic identification using pyrosequencing technology. Biotechniques 32: 124–126, 128, 130–133.

Armonienė R, Stukonis V, Paplauskienė V, Brazauskas G (2010). The genetic diversity of fine-leaved fescue (*Festuca* L.) species in Lithuania. In: Huyghe C, editor. Sustainable Use of Genetic Diversity in Forage and Turf Breeding. Amsterdam, the Netherlands: Springer, pp. 41–45.

Beard JB (1973). Turfgrass: Science and Culture. Englewood Cliffs, NJ, USA: Prentice Hall.

Bor NL (1952). The genus *Poa* L. in India. Part I. J Bombay Nat Hist Soc 50: 787–783.

Chandra A, Dubey A (2010). Identification of species-specific RAPD markers in genus *Cenchrus*. J Environ Biol 31: 403–407.

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002). SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3: 19.

Constable G (1985). Grasslands and Tundra: Planet Earth. New York, NY, USA: Time Life Books.

Fakruddin MD, Chowdhury A (2012). Pyrosequencing - an alternative to traditional Sanger sequencing. Am J Biochem Biotechnol 8: 14–20.

Fermanian TW, Michalski RS (1989). Weeder - an advisory system for the identification of grasses in turf. Agronomy J 81: 312–316.

Ferri G, Alù M, Corradini B, Beduschi G (2009). Forensic botany: species identification of botanical trace evidence using a multigene barcoding approach. Int J Legal Med 123: 395–401.

Ford CS, Ayres KL, Toomey N, Haider N, Stahl JVA, Kelly L, Wilstöm N, Hollingsworth P, Duff RJ, Hoot SB et al. (2009). Selection of candidate coding DNA barcoding regions for use on land plants. Bot J Linn Soc 159: 1–11.

Galli Z, Penksza K, Kiss E, Sági L, Heszky LE (2006). Low variability of internal transcribed spacer rDNA and trnL (UAA) intron sequences of several taxa in the *Festuca ovina* aggregate (Poaceae). Acta Biol Hung 57: 57–69.

Gealy DR, Tai TH, Sneller CH (2002). Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Sci 50: 333–339.

Gharizadeh B, Ghaderi M, Donnelly D, Amini B, Wallin KL, Nyrén P (2003). Multiple-primer DNA sequencing method. Electrophoresis 24: 1145–1151.

Gillespie LJ, Boles R (2001). Phylogenetic relationships and infraspecific variation in Canadian Arctic *Poa* based on chloroplast DNA restriction site data. Can J Bot 79: 679–701.

Gillespie LJ, Soreng RJ (2005). A phylogenetic analysis of the bluegrass genus *Poa* based on cpDNA restriction site data. Syst Bot 30: 84–105.

Gulsen O, Shearman RC, Vogel KP, Lee DJ, Heng-Moss T (2005). Organelle DNA diversity among buffalo grasses from the Great Plains of North America determined by cpDNA and mtDNA RFLPs. Crop Sci 45: 186–192.

Haider N (2003). Development and use of universal primers in plants. PhD, University of Reading, Reading, UK.

Haider N (2012). Evidence for the origin of the B genome of bread wheat based on chloroplast DNA. Turk J Agric For 36: 13–25.

Haider N, Nabulsi I (2008). Identification of *Aegilops* L. species and *Triticum aestivum* L. based on chloroplast DNA. Genet Resour Crop Evol 55: 537–549.

Haider N, Nabulsi I, Kamary Y (2010a). Identification of *Orchidaceae* species of northern west of Syria based on chloroplast DNA. Russ J Genet 46: 1067–1078.

Haider N, Nabulsi I, MirAli N (2010b). Comparison of the efficiency of A-PAGE and SDS-PAGE, ISSRs and RAPDs in resolving genetic relationships among *Triticum* and *Aegilops* species. Genet Resour Crop Evol 57: 1023–1039.

Haider N, Nabulsi I, MirAli N (2012). Identification of species of *Vicia* subgenus *Vicia* (Fabaceae) using chloroplast DNA data. Turk J Agric For 36: 297–308.

Haider N, Wilkinson MJ (2011). A set of plastid DNA-specific universal primers for flowering plants. Russ J Genet 47: 1204–1215.

Han S, Kim S, Yang CH, Seo J (2005). Genetical identification of *Akebia* and *Aristolochia* species by using Pyrosequencing. J Biotechnol 120: 360–363.

Hastings PAS (2004). Mitochondrial DNA analysis by pyrosequencing. MSc, University of Central Florida, Orlando, FL, USA.

Kellogg EA (1985). A biosystematic study of the *Poa secunda* complex. J Arnold Arboretum 66: 201–242.

Kindiger BK, Conley T (2009). Utilizing single primers as molecular markers in *Poa* spp. Jap Soc Grassland Sci 55: 206–215.

Krishnan S, Samson NP, Ravichandran P, Narasimhan D, Dayanandan P (2000). Phytoliths of Indian grasses and their potential use in identification. Bot J Linn Soc 132: 241–252.

Kumar J, Verma V, Goyal A, Shahi AK, Sparoo R, Sangwan RS, Qazi GN (2009). Genetic diversity analysis in *Cymbopogon* species using DNA markers. Plant Omics J 2: 20–29.

Langaee T, Ronaghi M (2005). Genetic variation analyses by pyrosequencing. Mutat Res 573: 96–102.

Leem K, Kim SC, Yang CH, Seo J (2005). Genetic identification of *Panax ginseng* and *Panax quinquefolius* by pyrosequencing methods. Biosci Biotechnol Biochem 69: 1771–1773.

Li A, Ge S (2001). Genetic variation and clonal diversity of *Psammochloa villosa* (Poaceae) detected by ISSR markers. Ann Bot 87: 585–590.

Li D, Rothballer M, Engel M, Hoser J, Schmidt T, Kuttler C, Schmid M, Schloter M, Hartmann A (2012). Phenotypic variation in *Acidovorax radicis* N35 influences plant growth promotion. FEMS Microbiol Ecol 79: 751–762.

Liu Q, Li L, Zhang N, Liu JIE, Shi Y (2010). Development and characterization of chloroplast microsatellite markers in *Pseudoroegneria* and *Leymus* (Poaceae: Triticeae). J Genet 89: e11–e22.

Marsh S (2007). Pyrosequencing applications. Methods Mol Biol 373: 15–24.

Mason-Gamer RJ (2004). Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. Syst Biol 53: 25–37.

Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E (2009). The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiol Lett 296: 97–101.

Novais RC, Thorstenson YR (2011). The evolution of Pyrosequencing® for microbiology: from genes to genomes. J Microbiol Methods 86: 1–7.

Paäakinskiene I, Griffiths CM, Bettany AJE, Paplauskiene V, Humphreys MW (2000). Anchored simple-sequence repeats as primers to generate species-specific DNA markers in *Lolium* and *Festuca* grasses. Theor Appl Genet 100: 384–390.

Parani M, Rajesh K, Lakshmi M, Parducci L, Szmidt AE, Parida A (2001). Species identification in seven small millet species using polymerase chain reaction-restriction fragment length polymorphism of *trn*S-*psb*C gene region. Genome 44: 495–499.

Patterson JT, Larson SR, Johnson PG (2005). Genome relationships in polyploid *Poa pratensis* and other *Poa* species inferred from phylogenetic analysis of nuclear and chloroplast DNA sequences. Genome 48: 76–87.

Ridgway KP, Duck JM, Young JPW (2003). Identification of roots from grass swards using PCR-RFLP and FFLP of the plastid *trn*L (UAA) intron. BMC Ecol 3: 8.

Ringquist S, Pecoraro C, Gilchrist CMS, Styche A, Rudert WA, Benos PV, Trucco M (2005). SOP³v2: web-based selection of oligonucleotide primer trios for genotyping of human and mouse polymorphisms. Nucleic Acids Res 33: W548–W552.

Ronaghi M, Elahi E (2002). Discovery of single nucleotide polymorphisms and mutations by pyrosequencing. Comp Funct Genomics 3: 51–56.

Ronaghi M, Shokralla S, Gharizadeh B (2007). Pyrosequencing for discovery and analysis of DNA sequence variations. Pharmacogenomics 8: 1437–1441.

Ronaghi M, Uhlen M, Nyren P (1998). A sequencing method based on real-time pyrophosphate. Science 281: 365–363.

Salk JJ, Sanchez JA, Pierce KE, Rice JE, Soares KC, Wangh LJ (2006). Direct amplification of single-stranded DNA for pyrosequencing using linear-after-the-exponential (LATE)-PCR. Anal Biochem 353: 124–132.

Scheef EA, Casler MD, Jung G (2003). Development of species-specific SCAR markers in bentgrass. Crop Sci 43: 345–349.

Silvar C, Perovic D, Casas AM, Igartua E, Ordon F (2011). Development of a cost-effective pyrosequencing approach for SNP genotyping in barley. Plant Breeding 130: 394–397.

Siqueira JF Jr, Fouad AF, Rôças IN (2012). Pyrosequencing as a tool for better understanding of human microbiomes. J Oral Microbiol 4: 10743.

Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, Brochmann C, Brysting AK, Sønstebø JH, Ims RA, Yoccoz NG et al. (2009). Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6: 16.

Soreng RJ (1990). Chloroplast-DNA phylogenetics and biogeography in a reticulating group: Study in *Poa* (Poaceae). Am J Bot 77: 1383–1400.

Soreng RJ, Bull RD, Gillespie LJ (2010). Phylogeny and reticulation in *Poa* based on plastid *trn*TLF and nrITS sequences with attention to diploids. In: Seberg O, Peterson G, Barfod A, Davis JI, editors. Diversity, Phylogeny, and Evolution in the Monocotyledons. Aarhus, Denmark: Aarhus University Press, pp. 619–643.

Stace CA, Al-Bermani AKKA, Wilkinson MJ (1992). The distinction between the *Festuca ovina* L. and *Festuca rubra* L. aggregates in the British Isles. Watsonia 19: 107–112.

Stace CA (1997). New Flora of the British Isles. 2nd ed. Cambridge, UK: Cambridge University Press.

Stebbins GL (1956). Cytogenetics and evolution of the grass family. Am J Bot 43: 890–905.

Steven B, McCann S, Ward NL (2012). Pyrosequencing of plastid 23S rRNA genes reveals diverse and dynamic cyanobacterial and algal populations in two eutrophic lakes. FEMS Microbiol Ecol 82: 607–615.

Studer B, Widmer F, Enkerli J, Kölliker R (2006). Development of novel microsatellite markers for the grassland species *Lolium multiflorum*, *L. perenne* and *Festuca pratensis*. Mol Ecol Notes 6: 1108–1110.

Taberlet P, Gielly L, Pautou G, Bouvet J (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biology 17: 1105–1109.

Tzvelev NN (1989). The system of grasses (Poaceae) and their evolution. Bot Rev 55: 141–203.

Wallinger C, Juen A, Staudacher K, Schallhart N, Mitterrutzner E, Steiner EM, Thalinger B, Traugott M (2012). Rapid plant identification using species- and group-specific primers targeting chloroplast DNA. PLoS One 7: e29473.

Wang L, Luhm R, Lei M (2007). SNP and mutation analysis. Adv Exp Med Biol 593: 105–116.

Ward J, Gilmore SR, Robertson J, Peakall R (2009). A grass molecular identification system for forensic botany: a critical evaluation of the strengths and limitations. J Forensic Sci 54: 1254–1260.

Weber AP, Weber KL, Carr K, Wilkerson C, Ohlrogge JB (2007). Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiol 144: 32–42.

Wheeler DJB, Jacobs SWL, Norton BE (1990). Grasses of New South Wales. 2nd ed. Armidale, Australia: University of New England Monographs.

Zapiola ML, Cronn RC, Mallory-Smith CA (2010). Development of novel chloroplast microsatellite markers to identify species in the *Agrostis* complex (Poaceae) and related genera. Mol Ecol Resour 10: 738–740.