Theor Appl Genet (2003) 107:1113–1122
DOI 10.1007/s00122-003-1308-y

Patricia Lem · Joëlle Lallemand

Grass consensus STS markers: an efficient approach for detecting polymorphism in *Lolium*

Received: 18 October 2002 / Accepted: 24 February 2003 / Published online: 14 June 2003

Abstract For ryegrass and many forage crops, characterization of varieties is often difficult for two reasons: few of discriminant morphological traits and a great within-varieties variation. Futhermore, few molecular markers are publicly accessible. In this paper we describe two approaches for the development of 42 sequence-tagged-site (STS) markers. Firstly, 14 STS markers were developed from *Lolium* sequences found in data bases. Secondly, 28 STS markers were developed from sequences found in related species of *Gramineae*. Out of 42 STS markers developed, 85.8% yielded successfull amplification and 62% revealed a high level of polymorphism with an average of five alleles per locus. The analysis of amplicons reveals a high STS marker specificity, a high conservation in gene structure and a strong intron sequence homology between allelic forms. Moreover, the majority of the STS markers can be considered as "universal markers" because 81% of these STS markers amplified successfully across 20 related grass species. These results permit us to consider the use of these markers in synteny studies.

Keywords *Lolium* · STS · Consensus marker · Grass species · Intronic polymorphism

Introduction

Ryegrass belongs to the genus *Lolium* that contains eight species. Two of them are major fodder crops: perennial ryegrass (*Lolium perenne* L.) with turf and forage varieties, and Italian ryegrass (*Lolium multiflorum* L.) that is used for hay and silage production.

For ryegrass and many forage crops, characterization of varieties is often difficult due to a lack of reliable morphological traits and a high degree of intra-varietal variation. At present, isoenzyme markers have been proposed for varietal description of ryegrass (Hayward and McAdam 1977; Greneche et al. 1991) and the description of genetic resources (Balfourier and Charmet 1994). However, these markers are of limited utility relative to the increasing number of cultivars. Moreover, presently available molecular markers (RFLP, RAPD, AFLP) for forage and turf species do not seem convenient for routine description of varieties. It is the same for microsatellites, which are likely the most interesting markers, but their development is expensive and at present only few sequences are publicly accessible (Kubik et al. 1999, 2001; Jones et al. 2001).

Facing the small number of reliable molecular markers we are steering towards sequence tagged site (STS) marker research. Our previous work (Lallemand et al. 1998) has shown a good level of polymorphism of STS markers in *Lolium* and encouraged us to continue this work. Moreover the studies of Bert et al. (1999) and Jones et al. (2002) have shown that the STS markers can be useful in genetic map construction: indeed five of these STS loci were mapped in *L. perenne*. However, fodder grasses have been less intensively studied than other members of the *Poaceae* and few DNA sequences are available in the data bases. Therefore, in this research, we present firstly STS markers developed from *Lolium* sequences and secondly, a STS marker development strategy using consensus sequences from related species of *Gramineae*. Primer pairs were designed in order to amplify the intronic regions. Thus, the polymorphism detected was based on intronic length polymorphism, which does not require digestion of the amplification products. Lastly, these STS markers were analysed to test their polymorphism in *Lolium* and their amplification-capacity in related grass species.

Communicated by G. Wenzel

P. Lem · J. Lallemand (✉)
BioGEVES, INRA du Magneraud, BP 52, 17700 Surgères, France
e-mail: joelle.lallemand@geves.fr
Fax: +33-05-46-68-31-00

Materials and methods

Plant materials and isolation of genomic DNA

Plant materials and the methods used for extraction of genomic DNA are summarized in Table 1. Young leaves were collected from 20 individuals for each of the *Lolium* cultivars used for polymorphism evaluation of STS markers and analysis of PCR products. STS markers were screened on five standards samples: one individual of each cv "Aramo", "Tribune" and "Repell" and two individuals of a wild population. In order to evaluate cross-species amplification, 23 cultivars representing 18 species of *Gramineae*, were analysed with 24 STS markers.

DNA concentration and quality was checked by comparing to a standard series of λ-DNA on 0.8% agarose gels after electrophoresis.

Strategy used to design the STS primers

Searches of DNA sequences were performed in the GenBank/EMBL database. We have searched for designing primers pairs in exonic sequences, supposed to be conserved among related species, in order to amplify the target intronic region, supposed to be polymorphic. However, all genes derived from *Lolium* sequences, and most of those derived from related species, were cDNA or mRNA with no indication on the intronic positions. Therefore, alignments with genes from more-distantly related species were necessary to define the intronic positions. Multi-alignments using BLASTN were also necessary to find consensus zones in exons flanking the target intron. Primers sequences were designed using the OLIGO software (MedProbe).

In short, DNA sequences were chosen with the following charateristics: (1) a minimum of one intron, (2) intron size should not exceed 1,000 bp, and (3) exonic regions should have characteristics allowing primer design.

Table 1 Plant materials and DNA extraction methods used for a: STS polymorphism evaluation, and b: amplification capacity study of STS primers in other *Gramineae* species

Species	Cultivar	Plant material	DNA extraction
a			
Lolium multiflorum	Aramo	20 individual plants/cultivar (ground in liquid nitrogen)	Cheung et al. 1993
	Tribune		
Lolium perenne	Repell Blazer Pacage Yatsyn Barpolo Concile Wild Population		
b			
Festuca rubra	Ludvina	Pool of ten individual plants (ground in liquid nitrogen)	Cheung et al. 1993
Festuca arundinacea var. *arundinacea*	Lubrette		
Festuca pratensis	Bundy		
Festuca ovina	Biljard		
Festuca arundinacea var. *glaucescens*	Wild Population		
Bromus catharticum	Luprime		
Bromus sitchensii	Lubro		
Phleum pratense	Pecera		
Poa pratensis	Pool of Amazone and Ampella		
Hordeum vulgare	Okney	Pool of five germs/cultivar (freeze-dried)	Dneasy Mini Kit (Qiagen France)
	Cadix		
Triticum aestivum	Autan Ariza		
Triticum durum	Corlit		
Sorghum sudanense	Alpilles		
Sorghum bicolor	Fulgus		
S.sudanense × *S.bicolor*	Soleillade		
Zea mays	F 2 F 252		
Oryza sativa	Nipponbare	One individual genotype	CTAB (Rogers and Bendich 1988)
	IR 72		
Saccharum officinarum	R 570 SES 14		

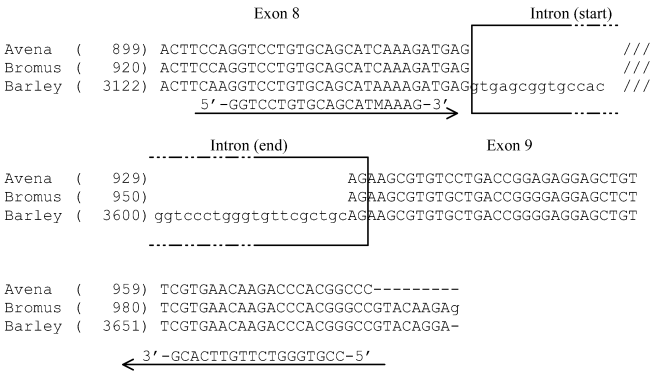

Fig. 1 Example of STS marker development from Aldose reductase genes. Primers (*black arrows*) are designed in exons from consensus sequences between related species of *Gramineae*. In this case, the primers permit to amplify one intron (*boxed sequence*). Exon numbering is derived from the barley gene. The Genbank accession number of each sequence is given in Table 3

Figure 1 gives an example to show how the consensus primers have been designed. In this case, the use of a degenerate nucleotide in the primer was necessary.

Polymerase chain reaction (PCR) and electrophoresis

PCR reaction was performed in a 25-μl volume containing 10 mM of Tris-HCL (pH 8.3), 50 mM of KCL, 3.5 mM of $MgCl_2$, 125 μM dNTP mix, 0.5 μM of each primers, 40 ng of total genomic DNA and 0.625 U of Ampli*Taq* Gold (AB Biosystems). Cycling conditions were as follow: an initial denaturation step of 10 min at 94 °C, followed by 40 cycles of 30-s at 94 °C, 30-s at the annealing temperature and 1 min extension at 72 °C. After cycling, the reactions were incubated for 10 min at 72 °C.

PCR products were separated on 1.4% agarose/TBE gels stained by BET for ADH and CAT STSs. Other STSs were run on 5% acrylamide/TBE gels, stained by silver (Chalhoub et al. 1997) or on sequencing gels for the shortest products.

Sequencing and analysis of *Lolium* PCR products

Bands of interest were extracted from the gel as follow: a puncture was made in the band using a pipette tip and the gel thus extracted was placed in 10 μl of sterile water for 10–20 min. Two microliters of the suspension were transferred into a PCR tube. The amplification was performed using the initial conditions. The purity of PCR products was checked on acrylamide gel (5%). PCR products were sent to the Qiagen society (Germany) for sequencing; 22 STSs showing polymorphism were sequenced. Except for LP1 and TRX STSs, only one allele per STS was extracted and re-amplified. Each allele was chosen from one individual of perennial ryegrass or Italian ryegrass, presenting when possible only one band. For STSs showing complex profiles, the band of expected size was chosen for sequencing.

BLASTN (with manual correction) has permitted us to check the identity of the amplified sequence and to locate the intron(s) in *Lolium*.

The different alleles of *Lolium* sequences have been studied for two STS loci (LP1 and TRX) and compared using CLUSTALW.

Results

STS development and polymorphism evaluation

In the present research, two approaches were described. Firstly, *Lolium* specific STSs were developed from *Lolium* sequences available in the databases. Secondly, we have developed consensus STSs, for which the primers were designed in grass consensus exonic sequences. We have looked for intronic length polymorphism by amplifying specifically the target intronic sequence, supposed to be polymorphic. The only exception was the LP1 STS, which does not contain an intron.

After preliminary work on sequence alignments and the study of domain architecture to avoid redundant information (using DART), there remained 42 STSs. From those, 14 derive from *Lolium* and 28 from related species of *Gramineae*, such as oat, fescue, maize, rice, barley and wheat. These STSs were screened on a subset of five *Lolium* samples: one individual of cv "Aramo", "Tribune" and "Repell", and two individuals of the wild population. Although, the intron size is overall conserved, the variations are sufficient on acrylamide or agarose gels to use these STSs as markers. Out of the 42 STSs tested, 6 did not amplify (14.2%), 10 gave monomorphic patterns (23.8%) (Table 2) and 26 showed intronic length polymorphism (62%) (Table 2). The polymorphism detected is high (the number of alleles ranged from 2 to 9) and is probably underestimated seing the small number of individuals tested. Figure 2 shows the polymorphism observed on MZE, ADP, LHAB and LP1 STSs. Table 3 gives a selection of primer sequences designed.

These results showed good efficiency for primers designed from consensus exonic sequences. STS markers developed from related species of *Gramineae* gave the best results. Indeed 67.9% of the STS markers derived from related species of *Gramineae*, showed length polymorphism, by comparison to 50% of the STSs derived from *Lolium*. Moreover only 10.7% of the STS markers derived from related species of *Gramineae* gave

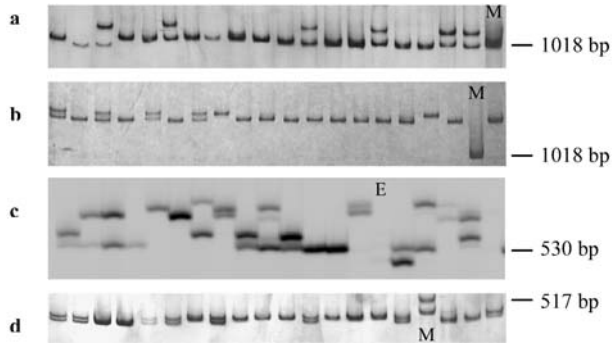

Fig. 2 PCR products amplified by **a**: *MZE*, **b**: *ADP*, **c**: *LHAB*, and **d**: *LP1* markers. MZE, ADP and LP1 STSs were run on 5% acrylamide/TBE gels revealed with silver staining. LHAB was run on the sequencing gel. *M*: size markers. *E*: empty lane

Table 2 Description of STSs showing amplification success in *Lolium*

STS	Gene	Source[a]	Exp size (bp)[b]	Obs size (bp)[c]	P/M[d]	STS	Gene	Source[a]	Exp size (bp)[b]	Obs size (bp)[c]	P/M[d]
ADH	Alcohol dehydrogenase	**Maize** Barley Rice	648	700	p	NDPK	Nucleoside diphosphate kinase	Ryegrass ***Arabidopsis***	183	180	p
ADP	ADP Glucose phosphorylase	**Rice** Barley Wheat	1,003	1,300	p	OSE	LEA = Late Embryogenesis Abundant	**Rice** Barley Wheat	311	298	p
ALDR	Aldose reductase	Oat **Barley** Bromus	540	548	p	OSO	Cystatin	**Rice** Maize	595	600	p
ASP	Aspartic protease	**Rice** Barley	506	520	p	OSRB	α-Amylase 3	**Rice** Barley Wheat	759	680	p
ATP 9	ATP 9	Ryegrass ***Arabidopsis***	593	520	M	OSW	ADP-Glucose glycosyl transferase	**Rice** Maize Barley Sorghum	414	410	p
CAF	Caffeic acid/ 5-hydroxyferulic acid 3/5-O-methyl transferase	Ryegrass **Maize**	993	400	p	PAL	Phenylalanine ammonialyase	**Wheat** Barley Rice	980	800	p
CAT	Catalase	Barley **Maize** Rice Wheat Rye	962	900	p	PCR	Protochlorophyllide reductase	Oat Wheat Barley ***Arabidopsis***	348	265	M
CCR	Cinnamoyl CoA reductase	Ryegrass ***Arabidopsis***	730	1,000	p	PGK	3-Phosphoglycerate kinase	Oat **Wheat**	201	209	p
CYP	Cysteine protease	Ryegrass **Barley**	228	232	p	PHOS	Phospholipase D	**Rice** Maize	665	650	p
EXP	Expansin	Fescue **Rice**	257	225	p	PHY	Phytochrome (AP3)	**Oat** Rice Sorghum	500	515	M
GLU	Glutelin	Oat **Rice**	380	370	p	PROFIL	Profilin 4	Maize Phleum ***Arabidopsis***	696	750	p
HVM	Myb3	**Barley** Rice Maize	306	280	M	RIB	α-amylase 1	**Rice** Barley Wheat	143	154	M
INV	Invertase	Ryegrass ***Arabidopsis***	215	220	M	RIT	Endochitinase	**Rice** Barley Wheat	430	370	p
LHAB	Chlorophyll a/b binding protein	Ryegrass ***Arabidopsis*** Rice	411	540	p	SER	Serine carboxypeptidase	**Rice** Barley	322	400	p
LP1	Pollen allergen LOLP I	**Ryegrass**	526	480	p	TRX	Thioredoxin	**Ryegrass** Rice	650	670	p
MADS 1	*MADS box protein 1 (transcription factor)*	Ryegrass **Rice**	237	222	M	VP1	Transcription factor	Oat **Rice** Barley	258	300	p

Table 2 (continued)

STS	Gene	Source[a]	Exp size (bp)[b]	Obs size (bp)[c]	P/M[d]	STS	Gene	Source[a]	Exp size (bp)[b]	Obs size (bp)[c]	P/M[d]
MADS 2	*MADS box protein 2 (transcription factor)*	Ryegrass	194	190	M	ZEA	Actin 1	**Maize**	558	510	M
		Rice						Rice			
MZE	Triosephosphate isomerase	**Maize**	1,202	1,000	p	ZMS	Sucrose synthase	**Maize**	431	400	M
		Rice						Rice			
		Rye						Barley			
		Barley									

[a] Source: sequence(s) permitting to design primers. In boldface, our reference sequence containing information on the intron size
[b] Exp. size: size of fragment deduced from reference sequence
[c] Obs. size: average size of pattern observed
[d] P: polymorph, M: monomorph

Table 3 A selection of primer sequences designed for STS loci that yielded amplification products with the specific TM. The sequence permitting to determine the expected size (noted in Table 2) is indicated in bold

STS	Source Genbank accession number	Primer sequence	TM (°C)
ADH	**Maize X04049** Barley AF253472 Rice AF148598	5′-GCGTCAAGATCCTCTTCACC-3′ 5′-CRCCCTCTCCAACACTCTCY-3′	56
ADP	**Rice M31616** Barley Z48563 Wheat X66080	5′-CCTCCGTGAACAATTTCCTG-3′ 5′-TCCAATACGAGCATTCTTGT-3′	50
ALDR	Oat U21747 **Barley Z48360** Bromus L12042	5′-GGTCCTGTGCAGCATMAAAG-3′ 5′-CCGTGGGTCTTGTTCACG-3′	55
CAF	Ryegrass: AF010291/AF033538/ AF033540 **Maize M73235**	5′-CGCTCATGAACCAGGACAAG-3′ 5′-GGGATGCCGCCGTCAAGGAC-3′	58
LHAB	Ryegrass X79277 ***Arabidopsis* AL132962** Rice AF010321	5′-GCGCCCGAGAGGCCCATCTGGTTCC-3′ 5′-GCCCAGCCGATGAGGATGAGCTC-3′	64
LP1	Ryegrass X57678	5′-CACCAAGCCGACATTCCAC-3′ 5′-CACCGTGCGAGCAAAGAAAG-3′	58
MZE	**Maize D00012** Rice L04967 Rye Z26875 Barley U83414	5′-CTAAAGGTCATTGCATGT-3′ 5′-CNGNGTTGATGATGTCGATGAA-3′	55
OSE	**Rice X63126** Barley X77157 Wheat X52103	5′-CGTCGTCCCCGGCGGCACCG-3′ 5′-TTGGACTCGTCGATGTCGAT-3′	60
OSRB	**Rice X56337** Barley J04202 Wheat X05809	5′-GACAGCCGCCTCGACTGGGG-3′ 5′-GATCTCCTGCTTCAGGTTCC-3′	60
OSW	**Rice X58228** Maize X03935 Barley X07931 Sorghum U23945	5′-TTCTGCATCCACAACATCTCCTA-3′ 5′-CTGACGTCCATGCCGTTGACGAT-3′	58
RIT	**Rice D16222** Barley L34211 Wheat X76041	5′-CAGACCTCCCACGAGACCAC-3′ 5′-CTGCGGCGTCATCCAGAACC-3′	62
TRX	Ryegrass AF159387/**AF290448** Rice AB024702	5′-GCCGTGTCATTGCACCTGTTTATGC-3′ 5′-CGGGGTTGCACGGATGTCCCATG-3′	56

Table 4 Intron size defined from the *Lolium* PCR product isolated[(a)] and genes available in genbank[(b)]. NA: data not available

STS	[a] Origin of *Lolium* PCR product	[b] Source Genbank accession number	Intron(s) size (bp)
ADH	*L. perenne*		102
		Maize X04049	97
		Barley AF253472	104
ASP	*L. perenne*		NA and 163
		Rice D32165	83 and 160
CYP	*L. perenne*		85
		Barley X05167	95
EXP	*L. perenne*		76
		Rice AF394554	117
GLU	*L. multiflorum*		92
		Oat J05485	126
		Rice D00584	103
LHAB	*L. perenne*		86 and 116
		Arabidopsis AL132962	0 and 104
MZE	*L. perenne*		456
		Maize D00012	457
		Rice L04967	481
NDPK	*L. multiflorum*		94
		Arabidopsis AL117386	120
		Rice AC084763	89
OSE	*L. multiflorum*		96
		Rice X63126	107
		Barley X77157	99
		Wheat X52103	71
		Secale AJ011948	117
PAL	*L. perenne*		86
		Wheat X99705	119
		Rice X87946	117
PGK	*L. perenne*		95
		Wheat X73528	102
PHOS	*L. multiflorum*		421
		Rice AB001920	478
RIT	*L. perenne*		89
		Rice D16222	130
SER	*L. multiflorum*		162 and 105
		Rice D17586	86 and 78

no amplification, by comparison to 21.4% of the STSs derived from *Lolium*.

Analysis of *Lolium* PCR products

Twenty two out of the 26 polymorphic STS were sequenced. The alleles were isolated either from *L. perenne* or from *L. multiflorum*. The sequencing quality of OSRB and CCR did not allow analysis. ADP, VP1, PROFIL and OSO were not sequenced. So, 20 STSs were analysed and the sequence analyses showed 80% to 100% homology among differents species, which suggests that we amplified the expected gene. Sequences searches in databases and alignments have also shown a high level of conservation of gene structure, such as the position of intron/exon junctions, exon size and low variation of intron length among the species. This has permitted us to approximately position intron/exon junctions in *Lolium* PCR sequences and to determine approximate intron size (Table 4). Except for the CAF locus, the observed size of PCR products is near to the expected size. The CAF marker revealed the amplification of two loci at 500 bp and 400 bp, whereas the expected size, deduced from maize OMT gene, was of 915 bp.

Out of 20 PCR products analysed, one case has not revealed the expected structure. It concerns the LHAB locus, which has been determined from *Lolium temulentum*, barley and rice sequences. The intron/exon junctions were derived from the *Arabidopsis thaliana* Lhca 2 gene (Table 3), constituted of four exons and three introns. Alignments have permitted us to establish the position of potential introns in the *L. temulentum* sequence and to design primer pairs in order to amplify one intron. After sequencing, LHAB STS seems to contain two putatives introns: 86 bp and 116 bp. Alignment of the *L. perenne* sequence (isolated from cv "Pacage") with the *L. temulentum* initial sequence reveals an homology of 95% to 100% and demonstrates the position of an other intron, that is not present in the *Arabidopsis* gene. Using BLASTX, amino-acid sequences deduced from the *Lolium* PCR product, presents a homology of 98% with the *L. temulentum* LhcI protein, 89% with the *Arabidopsis* Lhca2 protein and 93% with the tomato cab-7 protein (photosystem I Chlorophyll a/b binding protein). The tomato cab-7 gene (gb/M20241) is constituted of 5 exons and 4 introns contrary to the *Arabidopsis* gene. In

conclusion, two introns are present in *L. perenne* as in tomato and contrary to *Arabidopsis*. In this case, there has been an intron loss or an intron gain as has been observed in gene evolution, for example the Catalase (Frugoli et al. 1998) and ADH (Mitchell et al. 1989) genes.

Analysis of allelic forms in the *Lolium* PCR product

The differents allelic forms of LP1 and TRX loci were sequenced. The genotyping study on ten varieties of perennial ryegrass (data not shown) has revealed two alleles for LP1 and seven alleles for TRX.

The LP1 STS marker was developed from the *L. perenne* Lolp I gene (pollen allergen protein). No related sequence was found to locate a potential intron. The two allele have been isolated from cv "Repell" and the wild population. Alignments of these allele sequences with the initial sequence demonstrate that the LP1 STS has no intron. Alignment between these two alleles shows one substitution and an insertion of ten bases located outside the cDNA sequence. Compared to the original sequence (gb/X 57678), we have observed five other nucleotide substitutions. So six SNPs were counted on 434 bp (an average of one SNP every 72 bases). Five out of six SNPs are localised in the cDNA. These substitutions change the codons without changing the amino acids in the protein sequence.

The TRX STS marker was developed from *L. perenne* cDNA and a complete rice sequence (the thioredoxin gene, encoding redox proteins known to be present in all eukaryotic and prokaryotic organisms). Their comparison has permitted us to determine the location of one hypothetical intron in the *Lolium* sequence (gb/AF290448). The seven allelic forms were isolated from *L. perenne* and sequenced. Figure 3 presents the alignment of six alleles of the TRX locus (the quality for sequencing it one among the seven allelic sequences did not permit us to analyse it). Sequences analysis confirmed the presence of an intron and revealed a high similarity of *L. perenne* coding regions (from 91% to 97%) with rye, phalaris, barley and rice sequences coding the thioredoxin-like protein. Alignments of the six different allelic forms have been shown; firstly, complete homology in coding regions except for two nucleotide substitutions on 86 bases, and secondly, a strong homology (87%) in intronic sequences with the presence of insertion/deletion (indel) zones (Fig. 3). The indel frequency was 1/44 bp. Of the 11 indels identified, single-nucleotide indels accounted for half. The largest deletion observed was 18 nucleotides. Moreover, 28 single nucleotide polymorphisms have been recorded on an average of 487 bases of the intron sequence (an average of one SNP every 17 bp). We observe a complete concordance between the size of the allelic products sequenced and the allele order observed on the gel (Fig. 3).

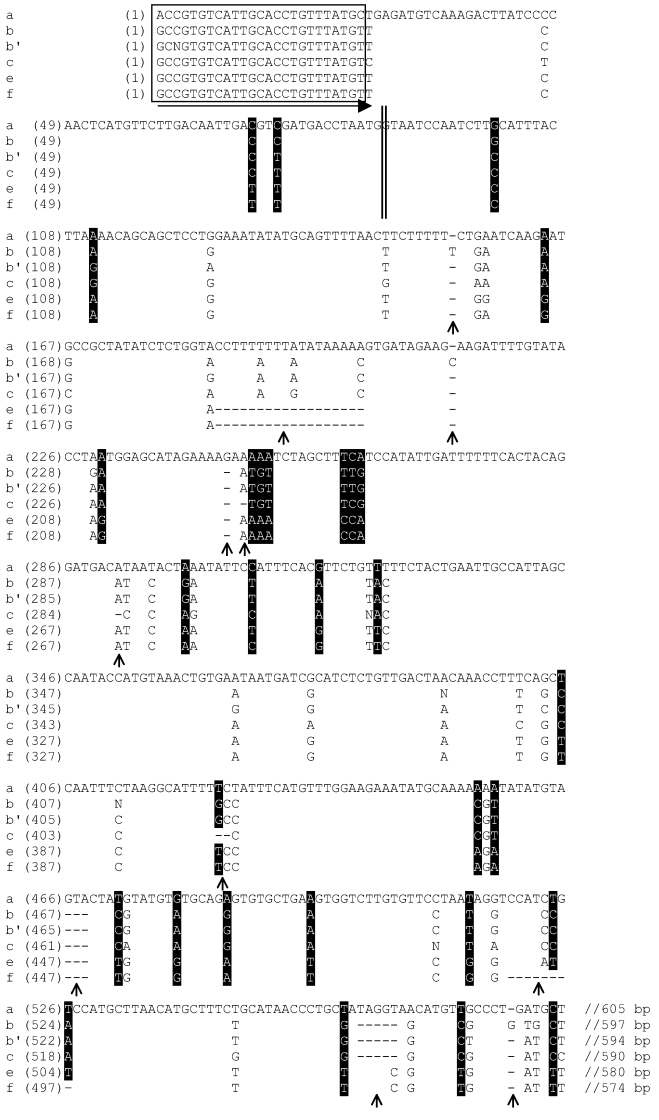

Fig. 3 Alignments of six alleles from the TRX STS locus. The *horizontal black arrow* indicates the forward primer. The *vertical double line* signals the intron start. Only the differences are indicated, using the allele-a (largest allele) as a reference. The following are signalled: the deletions are little drawn (–); the indels are by a *vertical arrow* and the SNPs in *white on a black background*. The size of the alleles sequenced is indicated at the end of each sequence

Amplification of the STSs in other grass species

From the strategy used, two groups of STS markers were defined: STSs derived from sequences of *Lolium* and STSs derived from species related to *Lolium*. In order to evaluate the capacity of cross-species amplification by the STS primers, 21 polymorphic STSs were tested on 18 *Gramineae* species represented by 23 cultivars (Table 1); ADP and two monomorphic STSs (ZEA and ZMS) were tested on 11 species belonging to the *Festuca, Phleum, Poa, Bromus* and *Lolium* genus. The average amplification success was 81% per STS. As expected, STSs developed from consensus sequences give a best output

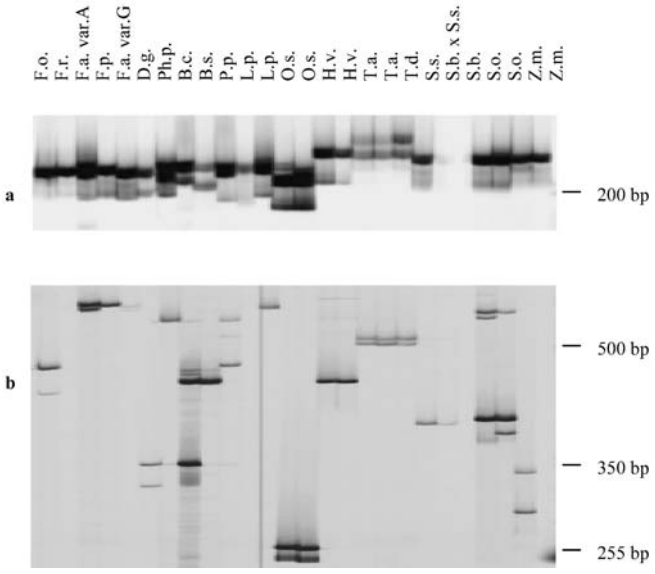

Fig. 4 STS amplification products from **a**: *PGK* STS, and **b**: *TRX* STS in related species of *Gramineae*. F.o.: *Festuca ovina*, F.r.: *F. rubra*, F.a. var. A: *F. arundinacea* var. *arundinacea*, F.p.: *F. pratense*, F.a. var. G: *F. arundinacea* var. *glaucescens*, D.g.: *Dactylis glamerata*, Ph.p.: *Phleum pratense*, B.c.: *Bromus catharticum*, B.s.: *B. sitchensii*, P.p.: *Poa pratensis*, L.p.: *Lolium perenne*, O.s.: *Oryza sativa*, H.v.: *Hordeum vulgare*. T.a.: *Triticum aestivum*; T.d.: *Triticum durum*, S.s.: *Sorghum sudanense*, S.b.: *S. bicolor*, S.o.: *Saccharum officinarum*, Z.m.: *Zea mays*

with an efficiency of 86% contrary to 67.5% for STSs developed directly from *Lolium* sequences. Neverthless the NDPK STS marker developed from a *Lolium* sequence amplified on all the species tested. Amplification success revealed by species was 81.7% on average, with an efficiency of 90.5% in *Hordeum*, 87.5% in *Bromus*, 86% in *Oryza* and 85% in *Festuca* species. However, these STSs are less efficient in *Poa* (67%) and *Sorghum* (63%) species, which were less or not used in the consensus zone research. PCR product size was overall conserved throughout the species, although we have observed an inter-specific polymorphism. The most variable STSs were TRX (Fig. 4), CAF and ALDR, which showed a strong variation in PCR product size ranging from 200 to 700 bp. The highest size homology was found between *Festuca* and *Lolium*.

Discussion

The length polymorphism detected (62%) is specially related to the strategy used. Polymorphism was looked for in intronic regions. This approach did not require digestion to reveal polymorphism in the fragment amplified, contrary to other studies on wheat (Chen et al. 1994; Talbert et al. 1994), barley (Tragoonrung et al. 1992), rice (Ghareyazie et al. 1995) and pea (Gilpin et al. 1997) where it was most often necessary. It was the same for another study carried on *Lolium* by Taylor et al. (2001), where STSs derived from wheat and barley, were used to amplify homologous loci in *L. perenne*. PCR products were monorphic for all primer pairs used and had consequently to be digested. The development of the STSs described in this paper, is easier and less expensive than that of microsatellite markers. This type of marker could be used for genotyping, in *Lolium*; indeed, these STSs showed a high discriminant power on ten perennial ryegrass cultivars (data not shown) but could also be used in species little studied, such as some vegetable, forest or ornamental species.

The nomenclature of genes being sometimes complicated, we have been careful to avoid redundancy which could have arisen from different names given to the same gene. Conversely, identical domains can belong to different genes. In the case of Lolp1 and expansin, for example, protein sequences revealed a similar domain architecture (using DART): recently, group-1 allergens from grass pollen, was identified as a second family of expansins (Cosgrove et al. 1997). These proteins are called β-expansins to distinguish them from the original class of expansins, referred as α-expansins. These proteins are shown to be structurally related to expansins, which are able to induce extension of plant cell walls (Cosgrove et al. 1997). So primers pairs of Exp and LP1 STS were designed in order to amplify specifically expansin and the allergen gene and to avoid redundant information.

The observed size of PCR products corresponds most often to the expected size, although the intron size was determined from that of related species of *Gramineae*, and even of less-related species. For example, for CCR, LHAB and NDPK loci, the position of the potential intron(s) and the expected size of PCR products were deduced by sequence information from the *A. thaliana* sequence (Table 2). In only one case (CAF) we did not obtain the expected size. The primers amplifying the CAF marker were designed directly from the *L. perenne* Caffeic acid/5-hydroxyferulic acid O-methyltransferase sequence (gb/AF010291). The position of a potential intron was determined by alignment with a maize gene (OMT: O-methyltransferase), which has two exons. The expected size, based on the intron size (915 bp) of the maize sequence, was not observed. Moreover, these STS primers amplify two loci, one at 500 bp and another at 400 bp. The primers designed could have allowed the amplification of two *L. perenne* isoforms, which are OMT1 (gb/AF033538) and OMT3 (gb/AF033540). The similarity of the nucleotide sequence (gb/AF010291) with OMT1 and OMT3 isoforms is 83% to 98%. The 400-bp locus has been sequenced and shows 100% of homology with OMT1 and OMT3. It is very probable that the sequencing of the locus at 500 bp would show also a strong homology with these two isoforms. It seems likely that in spite of a strong sequence homology, the *Lolium* gene (Caffeic acid/5-hydroxyferulic acid O-methyltransferase) is not homologous to the maize gene (OMT: O-methyltransferase). As a matter of fact the maize enzyme does not seem to have the same substrate specificity.

Analysis of *Lolium* PCR products has shown a high level of conservation in gene structure (exon and intron size, position of intron/exon junctions). Analysis of allelic forms in *Lolium* PCR products has revealed a strong homology in intronic sequence (LP1 and TRX STS loci) with the presence of SNP and indel zones. It is very probable that single-nucleotide indels and polymorphisms are overestimated. Indeed, sequence polymorphisms can be due to sequencing errors or PCR-amplification errors. Saiki et al. (1988) have observed a cumulative error frequency of about 0.25% after 30 cycles of PCR. Therefore, for sequence analysis of TRX alleles, we have recorded as SNP, at a given location, only the substitutions encountered at least twice among the six alleles (Fig. 3). So, the SNP frequency was 1/43 bp in the exon and 1/17 bp in the intron sequence (Fig. 3). Another study carried on a sequence comparison of four alleles of the SER locus found in *Lolium* (Virginie Lauvergeat, personnal communication) has shown similar results: on one hand, three SNPs occurred on 164 bp (1/54 bp) in the exons and, on the other hand, the introns show also a strong sequence homology with an average of one SNP occuring every 12 bp.

Although intron sequences are not similar between species, their size was approximately the same. This intron size similarity has been observed in grass ADH sequences (Gaut et al. 1999) but also in others organisms, such as *Drosophila* (Clark et al. 1996). Moreover, in *L. perenne*, a strong intron sequence homology between alleles has been observed. Knowing the ability of intron sequence to form secondary structures (Kirby et al. 1995; Leicht et al. 1995), these results assume that introns could play an important role in gene evolution (Gaut et al. 1999).

The strategy of STS development described in this paper has permitted us to develop consensus STS markers. Thus, these markers have presented a high amplification success in the species related to *Lolium*, with an average success of 81%. The STSs showed a better result than the *L. perenne* SSR described by Jones et al. (2001), which allowed an average amplification efficiency of around 55% in related species. Moreover, the species studied were restricted to the Aveneae and Poeae tribes. Considering only the nine species belonging to Aveneae and Poeae tribes (Table 1), our study shows an average amplification success of 87%. These markers can be used in synteny studies in addition to RFLP markers, which are more expensive and more limited. Indeed Jones et al. (2002), in their comparative map relationships study, have mapped the OSRB STS locus in the perennial ryegrass and Triticeae maps, and observed conserved synteny. Comparative mapping could allow the location of target genes across species, and in this case our STS development approach could be an interesting alternative. Moreover, these STSs are an interesting source for SNP development. In the case of ryegrass, our sequencing data indicate that SNPs seem to be numerous.

Acknowledgements This project was partly supported by financial assistance from CTPS (Comité Technique Permanent de la Sélection) and CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le développement. France). We also thank the CIRAD for providing the rice and sugarcane DNAs. We are very grateful to Virginie Lauvergeat for her STS sequencing results and her valuable comments on this manuscript.

References

Balfourier F, Charmet G (1994) Geographical patterns of isozyme variation in Mediterranean populations of perennial ryegrass. Heredity 72:55–63

Bert PF, Charmet G, Sourdille P, Hayward MD, Balfourier F (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. Theor Appl Genet 99:445–452

Chalhoub BA, Thibault S, Laucou V, Rameau C, Hofte H, Cousin R (1997) Silver staining and recovery of AFLP amplification products on large denaturing polyacrylamide gels. Biotechniques 22:216–220

Chen HB, Martin JM, Lanvin M, Talbert LE (1994) Genetic diversity in hart red spring wheat based on sequence-tagged-site PCR markers. Crop Sci 34:1628–1632

Cheung WY, Hubert N, Landry BS (1993) A simple and rapid DNA microextraction method for plant, animal and insects suitable for RAPD and other PCR analyses. PCR Methods Appl 3:69–70

Clark AG, Muse SV, Leicht BG (1996) Length variation and secondary structure of introns in the *Mlc*1 gene in six species of *Drosophila*. Mol Biol Evol 13:471–482

Cosgrove DJ, Bedinger P, Durachko DM (1997) Group I allergens of grass pollen as cell wall-loosening agents. Proc Natl Acad Sci USA 94:6559–6564

Frugoli JA, McPeek MA, Thomas TL, McClung CR (1998) Intron loss and gain during evolution of the catalase gene family in angiosperms. Genetics 149:355–365

Gaut BS, Peek AS, Morton BR, Clegg MT (1999) Patterns of genetic diversification within the *Adh* gene family in the grasses (*Poaceae*). Mol Biol Evol 16:1086–1097

Ghareyazie B, Huang N, Second G, Bennett J, Khush GS (1995) Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP. Theor Appl Genet 91:218–227

Gilpin BJ, McCallum JA, Frew TJ, Timmerman-Vaughan GM (1997) A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). Theor Appl Genet 95:1289–1299

Greneche M, Lallemand J, Michaud O (1991) Comparaison of different enzyme loci as a means of distinguishing ryegrass varieties by electrophoresis. Seed Sci Technol 19:147–158

Hayward MD, McAdam NJ (1977) Isozyme polymorphism as a measure of distinctness and stability in cultivars of *L. perenne*. Z PflanzenZücht 79:59–68

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterisation of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Jones ES, Mahoney NL, Hayward MD, Armstead IP, Jones JG, Humphreys MO, King IP, Kishida T, Yamada T, Balfourier F, Charmet G, Forster JW (2002) An enhanced molecular marker-based genetic map of perennial ryegrass (*Lolium perenne*) reveals comparative relationships with other *Poaceae* genomes. Genome 45:282–295

Kirby DA, Muse SV, Stephan W (1995) Maintenance of pre-mRNA secondary structure by epistatic selection. Proc Natl Acad Sci USA 92:9047–9051

Kubik C, Meyer WA, Gaut BS (1999) Assessing the abundance and polymorphism of simple sequence repeats in perennial Ryegrass. Crop Sci 39:1136–1141

Kubik C, Sawkins M, Meyer WA, Gaut BS (2001) Genetic diversity in seven perennial ryegrass (*Lolium perenne* L.) cultivars based on SSR markers. Crop Sci 41:1565–1572

Lallemand J, Lem P, Ghesquière M, Charmet G, Balfourier F (1998) Potentiality of STSs for variety distinction in ryegrass. In: UPOV (ed) Working group on biochemical and molecular techniques and DNA-profiling in particular. Session 5, Betsville, USA, September 28–30, 1998, BMT/S11

Leicht BG, Muse SV, Hanczyc M, Clark AG (1995) Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila*. Genetics 139:299–308

Mitchell LE, Dennis ES, Peacock WJ (1989) Molecular analysis of an alcohol dehydrogenase (*Adh*) gene from chromosome 1 of wheat. Genome 32:349–358

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. Plant Mol Biol A6:1–10

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Talbert LE, Blake NK, Chee PW, Blake TK, Magyar GM (1994) Evaluation of "sequence-tagged-site" PCR products as molecular markers in wheat. Theor Appl Genet 87:789–794

Taylor C, Madsen K, Borg S, Moller MG, Boelt B, Holm PB (2001) The development of sequence-tagged sites (STSs) in *Lolium perenne* L.: the application of primer sets derived from other genera. Theor Appl Genet 103:648–658

Tragoonrung S, Kanazin V, Hayes PM, Blake TK (1992) Sequence-tagged-site-facilitated PCR for barley genome mapping. Theor Appl Genet 84:1002–1008