Theor Appl Genet (2001) 102:416–424 © Springer-Verlag 2001

R. Kölliker · E.S. Jones · M.C. Drayton · M.P. Dupal
J.W. Forster

Development and characterisation of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.)

Received: 3 April 2000 / Accepted: 12 May 2000

Abstract Highly informative molecular markers, such as simple sequence repeats (SSRs), can greatly accelerate breeding programs. The aim of this study was to develop and characterise a comprehensive set of SSR markers for white clover (*Trifolium repens* L.), which can be used to tag genes and quantitative trait loci controlling traits of agronomic interest. Sequence analysis of 1123 clones from genomic libraries enriched for $(CA)_n$ repeats yielded 793 clones containing SSR loci. The majority of SSRs consisted of perfect dinucleotide repeats, only 7% being trinucleotide repeats. After exclusion of redundant sequences and SSR loci with less than 25 bp of flanking sequence, 397 potentially useful SSRs remained. Primer pairs were designed for 117 SSR loci and PCR products in the expected size range were amplified from 101 loci. These markers were highly polymorphic, 88% detecting polymorphism across seven white clover genotypes with an average allele number of 4.8. Four primer pairs were tested in an F_2 population revealing Mendelian segregation. Successful cross-species amplification was achieved in at least one out of eight legume species for 46 of 54 primer pairs. The rate of successful amplification was significantly higher for *Trifolium* species when compared to species of other genera. The markers developed in this study not only provide valuable tools for molecular breeding of white clover but may also have applications in related taxa.

Communicated by G. Wenzel

R. Kölliker · E.S. Jones · M.C. Drayton · M.P. Dupal
J.W. Forster (✉)
Plant Biotechnology Centre, Agriculture Victoria,
La Trobe University, Bundoora, Victoria 3083, Australia
e-mail: john.forster@nre.vic.gov.au
Tel.: +61-3-9479 2995, Fax: +61-3-9479 3618

Present address:
R. Kölliker,
Swiss Federal Research Station for Agroecology and Agriculture,
Reckenholzstrasse 191, 8046 Zurich, Switzerland

Keywords Simple sequence repeats · SSR-enriched libraries · Molecular marker · *Trifolium repens* · Cross-species amplification

Introduction

White clover (*Trifolium repens* L.) is one of the most important legumes grown in temperate pastures, due to its excellent nutritive value and ability to fix atmospheric nitrogen. It is an outbreeding, allotetraploid (2n=4*x*=32) species with a high level of genetic heterogeneity (Williams 1987). White clover cultivars which are well adapted to specific environments and management conditions may significantly enhance pasture production systems. Breeding aims for this species include improved persistence, increased dry matter yield, improved competitive ability, higher stolon density, extended climatic range, improved disease resistance, higher digestibility and improved seed yield (Mather et al. 1996). More than six decades of plant breeding have resulted in significant genetic improvement of white clover (Woodfield and Caradus 1994), but most of the target traits are complex, making breeding costly and progress slow.

Molecular markers allow for the selection of desired traits based on genotype rather than phenotype and can therefore complement and accelerate plant breeding programs. They can also be used for the early selection of traits which are not expressed during the juvenile phase, such as persistence, competitive ability and seed yield. Molecular markers have been successfully used for the construction of genetic linkage maps and for the identification and tagging of economically important genes and quantitative trait loci (QTLs) in a large number of plant species (Rafalski et al. 1996; Staub et al. 1996; Mohan et al. 1997; Kumar 1999). Forage crops have gained attention more recently (Jenczewski et al. 1997; Chen et al. 1998; Hayward et al. 1998; Barcaccia et al. 1999; Bert et al. 1999; Brouwer and Osborn 1999), but agronomically important genes and QTLs have only been characterised in a few species (Faville et al. 1994; Hayward et al. 1994; Kidwell et al. 1994).

A basic pre-requisite for any molecular breeding program is a robust set of polymorphic markers for the species under investigation. Among the large variety of marker systems available (Sweeney and Danneberger 1995; Rafalski et al. 1996; Weising et al. 1998), simple sequence repeats (SSRs, also called microsatellites), which are short tandem repeat units of between 1 and 6 bp in length (Tautz 1989), offer a number of advantages which have made them increasingly popular in plant and animal studies. SSRs are highly polymorphic due to variation in repeat number and are co-dominantly inherited (Rafalski et al. 1996). Their detection is based on the polymerase chain reaction (PCR), requiring only small amounts of DNA and proving suitable for automation. They are ubiquitous in eukaryotic genomes and have been found to occur every 21 to 65 kb in plant genomes (Morgante and Olivieri 1993; Wang et al. 1994). Consequently, SSRs are ideal markers for a broad range of applications such as genome mapping, trait mapping and marker-assisted selection.

The main disadvantage of SSR markers is the high cost of development. SSRs may be identified by screening DNA databases (Morgante and Olivieri 1993; Wang et al. 1994), but such information is not available for most forage crop species. SSR discovery in species with little or no DNA sequence information usually involves the construction and screening of partial genomic libraries and the sequencing of SSR-positive clones (Rafalski et al. 1996). This is a labour intensive process, but the screening effort may be greatly reduced by the construction of genomic libraries highly enriched for SSRs (Karagyozov et al. 1993; Kijas et al. 1994; Edwards et al. 1996).

In this paper we report the development and the characterisation of several hundred SSR markers for white clover based on genomic libraries enriched for $(CA)_n$ repeats. We have evaluated a substantial number of these markers for their ability to detect polymorphism across a panel of seven diverse genotypes, confirmed the Mendelian inheritance of some markers in an F_2 population, and investigated their cross-transferability to other legume species. This set of markers provides a valuable tool for genome mapping and marker-assisted selection in white clover.

Materials and methods

Plant material and DNA extraction

Enriched libraries were constructed from the inbred *T. repens* genotype I7J (Joyce et al. 1999). SSR primers were screened for polymorphism across a set of seven *T. repens* genotypes; a single genotype from the cultivar Dusi, two genotypes each from the cultivars Haifa and Prop and two inbred genotypes, I5J and I4R (Michaelson-Yeates et al. 1997). Segregation of SSR loci was assayed using three F_1 and 20 F_2 progeny individuals from a cross between I5J×I4R (Michaelson-Yeates and Abberton, unpublished). Cross-species amplification of *T. repens* SSRs was tested using single genotypes from the following species: *Trifolium pratense* L., *Trifolium subterraneum* L., *Trifolium ambiguum* Bieb., *Trifolium nigrescens* Viv., *Glycine max* L. (Merr.), *Medicago sativa* L., *Lotus corniculatus* L. and *Melilotus alba* Medik.

Genomic DNA from inbred genotypes and segregating progeny was provided by the Institute of Grassland and Environmental Research, IGER, Aberyswyth, UK, and extracted using a 2×CTAB method (Doyle and Doyle 1987). DNA of all the other genotypes was extracted using a 1×CTAB method (Fulton et al. 1995).

Construction of SSR-enriched libraries and sequencing

Six SSR-enriched libraries of *T. repens* were constructed using the procedure of Edwards et al. (1996), with modifications as described below. Different restriction enzymes (*Alu*I, *Dra*I, *Eco*RV, *Hae*III, *Rsa*I or *Ssp*I) were used to restrict DNA prior to enrichment for each library. Enrichment was carried out with $(CA)_{20}$ synthetic oligonucleotides bound to the selection filter. $(CA)_n$ selection was employed as it proved to be superior to other repeat motifs for the enrichment of partial genomic libraries from perennial ryegrass (*Lolium perenne* L.; Jones et al., unpublished). Hybridisation was carried out at 50°C for 16 h followed by three washes in 2×SSC (30 mM sodium citrate, 300 mM NaCl, pH 7.0) and 0.1% (w/v) sodium dodecyl sulphate (SDS) at 50°C and five washes in 0.5×SSC and 0.1% (w/v) SDS. Eluted fragments were run through a MicroSpin S-300 HR column (Pharmacia Biotech) and cloned into the *Bss*HI site of a modified pUC19 vector (pJV1), kindly provided by Dr. K.J. Edwards, IACR-Long Ashton Research Station, Bristol, UK.

Plasmids were transformed into Max Efficiency DH5α competent cells (Life Technologies), plated onto LB agar plates (Sambrook et al. 1989), containing 50 µg/ml of ampicillin, 50 µg/ml of X-galactosidase and 0.5 mM of IPTG, and incubated for 16 h at 37°C. DNA from white colonies was extracted using the Wizard Plus SV (Promega Co., Madison, Wis., USA) or the QIAprep turbo (Qiagen) purification kit and sequenced on an ABI automated sequencer (PE Applied Biosystems) using the M13 forward primer and the BigDye terminator (PE Applied Biosystems) cycle-sequencing kit.

After sequencing ten clones from each library, the two libraries with the highest level of enrichment were selected for further SSR development.

Classification of SSRs and primer design

Sequences containing at least five uninterrupted dinucleotide repeats, four uninterrupted trinucleotide repeats or three uninterrupted tetranucleotide (or larger) repeats were classified into perfect, imperfect and compound SSRs as suggested by Weber (1990). Primers were designed to the flanking regions of the SSRs using the PrimerPremier 4 (Premier Biosoft International, Palo Alto, Calif., USA) software, based on criteria of GC content, melting temperature and the lack of secondary structure. Primers were designed in the 18–27 bp range to yield amplification products of 80–350 bp. They were synthesised by PacificOligos (Lismore, NSW, Australia).

PCR amplification and product electrophoresis

PCR amplifications were performed in a 20-µl vol. containing 25 ng of genomic DNA, 1×PCR buffer (Finnzyme, Espoo, Finland), 0.2 mM of each dNTP, 0.5 µM of each forward and reverse primer and 0.4 U of Dynazyme II (Finnzyme) DNA polymerase. The forward primer was end-labelled with γ-^{33}P-ATP (400 Ci/mmol, Geneworks, Adelaide, S.A., Australia). PCR was performed in a MJ PT-200 (MJ Research Inc., Waltham, Mass., USA) thermocycler using one of the following touch-down profiles, depending on the T_m value of the primer pairs: (1) 10 cycles of 60 s at 94°C, 30 s at 65°C, 60 s at 72°C with a reduction of the annealing temperature of 1°C by every cycle, followed by 20 cycles of 30 s at 94°C, 30 s at 55°C and 30 s at 72°C; (2) a similar profile as (1) with an initial annealing temperature of 60°C and a final annealing temperature of 50°C; (3) a similar profile as (1) with an initial annealing tempera-

ture of 55°C and a final annealing temperature of 45°C, and (4) a similar profile as (1) with an initial annealing temperature of 50°C and a final annealing temperature of 40°C. The PCR products were denatured by adding 15 µl of denaturing gel loading buffer (Sambrook et al. 1989) and heating at 94°C for 5 min. SSR alleles were separated by running PCR products on a denaturing 6% (w/v) acrylamide gel (19:1 acrylamide:bis-acrylamide, Amresco, Solon, Ohio, USA) in 1× TBE (Sambrook et al. 1989) at 80 W for 5000 Vh using a Biomax STS 45i DNA sequencing unit (Kodak). A 100-bp size ladder (Promega Co., Madison, Wis., USA) was included on each gel and the size of amplification products was estimated by extrapolation. Gels were transferred onto Whatman 3 MM paper and dried in a gel dryer (BioRad 583) at 80°C for 45 min. Banding patterns were visualised using a Phosphorimager 400b (Molecular Dynamics, Sunnyvale Calif., USA) or by exposing gels for 48 to 72 h to X-ray film (Biomax MR, Kodak).

Primer evaluation

All primer pairs were screened on the set of seven diverse genotypes for their ability to yield an amplification product of the expected size and to detect polymorphism. For primers that detected polymorphism, the number of alleles and the polymorphism information content (PIC) of the SSR was calculated as described by Saal and Wricke (1999), based on expected heterozygosity (Hedrick 1985):

$$\mathrm{PIC} = H = 1 - \sum_{i=1}^{k} p_i^2 \quad (1)$$

where p_i is the frequency of the i-th allele out of the total number of alleles and k is the number of different alleles in the sample.

Cloning and sequencing of PCR amplification products

To verify the presence of SSR loci, products of successful cross-species amplification with primer pairs TRSSRA02B08, TRSSRA02B09 and TRSSRA02C04 were cloned and sequenced. Two microliters of unlabelled PCR product were cloned into the pGEM-T Easy Vector (Promega), transformed using XL 10-Gold ultracompetent cells (Stratagene, La Jolla, Calif., USA), plated on LB agar plates (Sambrook et al. 1989), containing 50 µg/ml of ampicillin, 50 µg/ml of X-galactosidase and 0.5 mM of IPTG, and incubated for 16 h at 37°C. DNA from three to six white colonies of each PCR reaction was extracted and sequenced as described above. Sequences containing SSRs were compared to each other using the multiple sequence alignment procedure of Clustal W (Thompson et al. 1994).

Results

Preliminary screening of enriched libraries

Fifty nine clones were sequenced from across six libraries, which were constructed using different restriction enzymes, in order to identify those libraries most suitable for large-scale SSR discovery. The level of SSR enrichment ranged from 30% in library TRSSRF to 90% in TRSSRA (Table 1). All but one of the SSRs consisted of dinucleotide motifs of the type used for enrichment [(AC),(CA)/(GT),(TG)] and had an average repeat number of 11, with a range from 5 to 28 (data not shown). The two libraries showing the highest enrichment for SSRs (TRSSRA and TRSSRB) were selected for large-scale SSR discovery in *T. repens*.

Large-scale SSR discovery and characterisation of SSR loci

Readable DNA sequences were obtained from a total of 1123 clones (Table 2). Seventy one percent of these clones contained SSR loci (Table 2). Approximately one tenth of all SSR clones were redundant, i.e. they consisted of the same SSR locus (or a variant of it) and showed more than 95% similarity in flanking sequences. Most redundant SSRs were found to be repeated only once across all libraries. However, one redundant SSR was found in 28 clones, consisting of four variants of the SSR motif [$(AC)_5GG(AC)_6$, $(AC)_5GG(AC)_5$, $(AC)_5GG(AC)_7$, $(AC)_5GG(AC)_5CGAC$]. The flanking sequences of these clones were almost identical, with only six clones showing single base-pair mutations. After eliminating redundant sequences, 696 unique SSR clones remained, which corresponded to an enrichment rate of 62% across all libraries (Table 2). Almost 60% of all unique SSR clones were non-truncated, i.e. they had more than 25 bp of flanking sequence at both ends of the SSR, so that successful primer design was highly probable. Total SSR enrichment was 10% higher for the TRSSRA library when compared to the TRSSRB library (Table 2). However, the TRSSRA library contained more redundant and truncated SSR clones than TRSSRB, so that enrichment for unique non-truncated SSR clones was almost the same for both libraries (Table 2).

The majority (93%) of unique SSRs consisted of dinucleotide repeats with 7% being trinucleotide repeats (Table 3). There were only two tetra-, two penta- and one hepta-nucleotide repeat. Across all categories, more than 80% of all SSRs were perfect or compound repeats. There was no difference in the SSR type observed among different libraries. The length of dinucleotide SSRs ranged from 5 to 54 repeats with an average of 12.3, and that of trinucleotide SSRs from 4 to 118 with an average of 29 repeats (data not shown). Most SSRs consisted of less than 21 repeats with only 5% being larger than 30 repeats. The average repeat length was slightly lower for TRSSRA (12.8) when

Table 1 Preliminary screening for SSR frequency in enriched *T. repens* libraries constructed using different restriction enzymes

Item	TRSSRA	TRSSRB	TRSSRC	TRSSRD	TRSSRF	TRSSRG
Restriction enzyme	*Alu*I	*Dra*I	*Eco*R V	*Rsa*I	*Ssp*I	*Hae*III
Number of clones sequenced	10	10	10	10	10	9
Number of clones containing SSR loci	9	7	4	6	3	4
Percent library enrichment	90	70	40	60	30	44

Table 2 Large scale SSR discovery in different libraries of *T. repens*

Item	Library[a]			Total
	TRSSRA	TRSSRB	Other	
Number of clones sequenced[b]	583	498	42	1123
Clones containing SSR loci[c] (percentage of clones sequenced)	445 (76)	329 (66)	19 (45)	793 (71)
Redundant SSR clones[d] (percentage of clones containing SSRSs)	61 (14)	32 (10)	4 (10)	97 (12)
Unique SSR clones (percentage of clones sequenced)	384 (66)	297 (60)	15 (36)	696 (62)
Unique, non-truncated (>25 bp flanking sequence) SSR clones (percentage of clones sequenced)	210 (36)	174 (35)	13 (31)	397 (35)

[a] For description of libraries see Table 1
[b] Number of clones with a readable DNA sequence
[c] SSRs were defined as sequences containing at least five uninterrupted dinucleotide repeats, four uninterrupted trinucleotide repeats or three uninterrupted tetranucleotide (or larger) repeats
[d] Duplicate SSRs with >95% sequence similarity

Table 3 Frequency and type of di-, tri- and ≥tetra-nucleotide repeats found across six *T. repens* enriched libraries based on a total of 696 unique SSRs

Item	Type			Total
	Perfect	Imperfect	Compound[a]	
Dinucleotide repeats[b]	57%	15%	21%	93%
Trinucleotide repeats[c]	4%	2%	0%	7%
≥Tetranucleotide repeats[d]	<1%	<1%	0%	<1%
Total	62%	17%	21%	

[a] Compound repeats include perfect as well as imperfect repeats
[b] SSRs with at least five uninterrupted dinucleotide repeats
[c] SSRs with at least four uninterrupted trinucleotide repeats
[d] SSRs with at least three uninterrupted tetra (–) (or larger) nucleotide repeats

Table 4 Frequency of dinucleotide and trinucleotide SSR motifs found across six *T. repens* enriched libraries

Type	Total	
	Number	Percentage
Dinucleotide SSRs[a]	645	
(AC), (CA), (TG), (GT)	485	75.2
(AG), (GA), (CT, (TC)	9	1.4
(AT), (TA)	3	0.5
(CG), GC)	0	0.0
Compounds[b]	148	22.9
Trinucleotide SSRs[c]	46	
(GAA), (TTC)	23	50.0
(AAG), (CTT)	18	39.1
(AGT), GAT), (TAG), (GTT)	5	10.9

[a] SSRs with at least five uninterrupted dinucleotide repeats
[b] All compund dinucleotide repeats contained (AC), (CA), (GT) or (TG) motifs
[c] SSRs with at least four uninterrupted trinucleotide repeats

compared to TRSSRB (14.0). Table 4 shows that more than 75% of all dinucleotides contained SSRs of the motif that was used to construct the enriched libraries [(AC),(CA)/(GT),(TG)]. Together with the compound dinucleotide repeats, which all also contained (AC),(CA)/(GT),(TG) repeats, these motifs accounted for

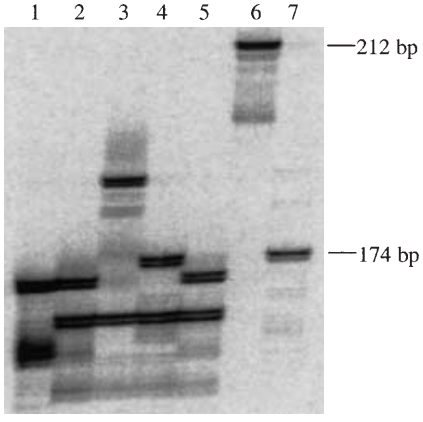

Fig. 1 Amplification across seven *T. repens* genotypes (*1*=Dusi 35, *2* =Haifa 27, *3*=Haifa 38, *4*=Prop16, *5*=Prop 39, *6*=I5J, *7*=I4R) for SSR locus TRSSRA01H11

91% of all SSRs. While some dinucleotide SSRs consisted of (AG),(GA)/(CT),(TC) and (AT)/(TA) repeats, the (CG)/(GC) motif was never found. The most frequent trinucleotide motif was (GAA)/(TTC) (50%), closely followed by the (AAG)/(CTT) motif.

Primer evaluation

Primer pairs were designed and evaluated for 117 SSR loci. Primers were 18–27-bp long (average 21 bp) and had a GC content of 27 to 58% (average 42%). The size of expected amplification products ranged from 84 to 312 bp with an average of 178 bp. Details of ten primer pairs are given in Table 5. Out of the 117 primer pairs, 101 amplified a product of the expected size (Table 6). There was no correlation between primer length, GC content or expected product size, and successful amplification. However, the percentage of successful amplification was highest for imperfect SSRs, dinucleotide repeats, and SSRs with 11–20 repeats. Eighty eight percent of all amplified SSRs were polymorphic across seven *T. repens* genotypes and detected an average of 4.8 alleles (Table 6.) An example of a typical amplification pattern is shown in Fig. 1. The average polymorphism informa-

Table 5 A selection of primer sequences designed for SSR loci that yielded amplification products of the expected size across seven *T. repens* genotypes

SSR	Primer sequence (5′–3′)	Repeat motif/repeat class	Expected	Polymorphic	No. of alleles	PIC[a]
TRSSRA01H11	F AGAAAGGTGAATGATGAAA R TCTAATTCTTCCAATAGGG	$(GAA)_{20}$ Perfect	212	Yes	6	0.78
TRSSRA02A05	F CAGTAAAGGAATCTGTTCAAACT R AAACACCAATCAGACCGAAA	$(GT(_5T(GT)_2C(GT)_3$ Imperfect	119	Yes	4	0.66
TRSSRA02B08	F TTTTGCTAATAAGTAATGCTGC R GGACATTATGCAATGGTGAG	$(TG)_{11}$ Perfect	121	Yes	3	0.53
TRSSRA02B09	F TTTTCGCATTGTTTCAGACC CCCTTTCTCAACCCACATC	$(CA)_5CC(CA)_3$ Imperfect	169	No	n.a.[b]	n.a.
TRSSRA02C02	F AAATAAAACCACAAGTAACTAG R TATAGGTGATTTGAAATGGC	$(CA)_{10}(GA)_6$ Compound	147	Yes	7	0.79
TRSSRA02C03	F TATGCTGGTAGATAAACTTAAA R TGCTCTGGAGATTGATGG	$(CA)_9$ Perfect	117	Yes	6	0.81
TRSSRA02C04	F TGGCTATTACAACTTGGAGA R CGAGGCATACTTGATGATGG	$(CA)_3CC(CA)_7$ Imperfect	84	No	n.a.	n.a.
TRSSRAXX31	F TCTGTTTTGTTGGCCATGC R TTGCAAAGTGTTTGGAAGGA	$(GT)_7$ Perfect	123	Yes	3	0.60
TRSSRAXX34	F TGACAGAAGACCTGATGTACCG R TTCCACTCTTAGCATCAACTGG	$(CA)_8(GA)_5(GG(GA)_{15}$ Compound	196	Yes	5	0.78
TRSSRDXX16	F AAGTGTTGGACAAGGAAACTAGG R TCTCTAGATCACCGGCATTG	$(TA)_7(CA)_{19}$ Compound	167	Yes	5	0.75

[a] Polymorphism information content as described in Materials and methods
[b] Not applicable

Table 6 Amplification and polymorphism across eight *T. repens* genotypes for different SSR types

Type	No. of primer pairs screened	Successful amplification (percentage of primer pairs screened)	Polymorphic across test genotypes (percentage of successful amplifications)	Average number of alleles detected	Average PIC[a]
All SSRs	117	101 (86)	89 (88)	4.8	0.69
By SSR class					
Perfect SSRs	69	59 (86)	53 (90)	4.4	0.68
Imperfect SSRs	25	23 (92)	19 (83)	5.2	0.66
Compound SSRs	23	19 (83)	17 (89)	6	0.78
By SSR type[b]					
Dinucleotide SSRs	108	94 (87)	84 (89)	4.8	0.69
Trinucleotide SSRs	8	6 (75)	5 (83)	5.5	0.77
By repeat number					
5–10 repeats	66	57 (86)	49 (86)	4.2	0.68
11–20 repeats	37	33 (89)	30 (91)	5.6	0.69
21–32 repeats	14	7 (79)	10 (91)	6	0.80

[a] Polymorphism information content as described in Materials and Methods
[b] Primers to one heptanucleotide SSR were tested but failed to amplify

tion content (PIC) for all polymorphic SSRs was 0.69. Compound SSRs had a PIC and detected more alleles than perfect or imperfect SSRs. The PIC and the number of alleles increased with increasing repeat length (Table 6). Four primer pairs were tested for Mendelian segregation in a segregating population. In all cases, primers detected a single SSR allele in the inbred parents, the two parental alleles in each F_1 and segregating parental alleles in the F_2. Segregation did not significantly deviate from Mendelian ratios ($P<0.05$) (Fig. 2).

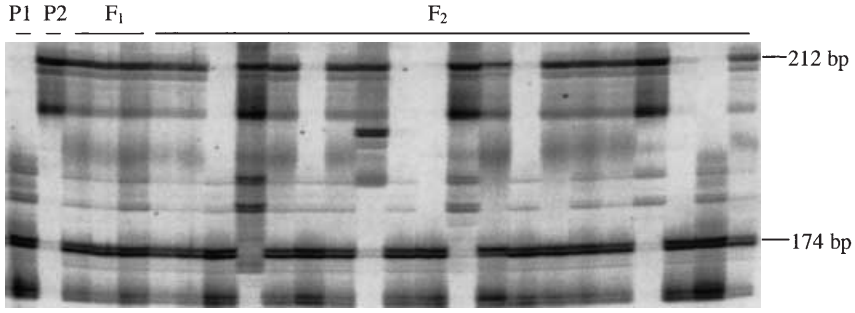

Fig. 2 Amplification pattern of SSR locus TRSSRA01H11 in a segregating *T. repens* population (P1, P2=inbred parents, F_1=three F_1 plants from P1×P2, F_2=20 F_2 plants obtained through selfing of a single F_1 plant). Size markers indicate segregating alleles

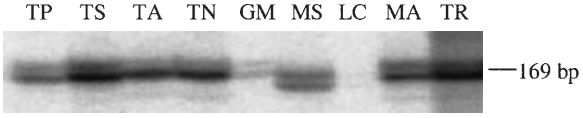

Fig. 3 Amplification across nine legume species [*T. pratense* (*TP*), *T. subterraneum* (*TS*), *T. ambiguum* (*TA*), *T. nigrescens* (*TN*), *G. max* (*GM*), *M. sativa* (*MS*), *L. corniculatus* (*LC*), *M. alba* (*MA*) and *T. repens* (*TR*) for SSR locus TRSSRA02B09

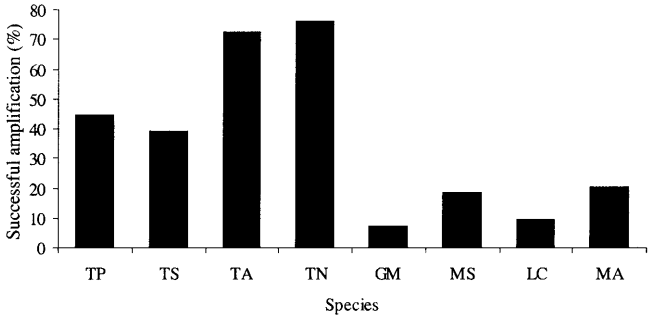

Fig. 4 Cross-species amplification of *T. repens* SSRs. Fifty four primer pairs were tested for amplification in *T. pratense* (*TP*), *T. subterraneum* (*TS*), *T. ambiguum* (*TA*), *T. nigrescens* (*TN*), *G. max* (*GM*), *M. sativa* (*MS*), *L. corniculatus* (*LC*) and *M. alba* (*MA*). Successful amplification in each species is shown as a percentage of the total number of primer pairs screened

Cross-species amplification

Of the 54 primer pairs screened for cross-species amplification, 46 amplified a product in at least one species comparable in size to the product detected from *T. repens* (Fig. 3). The rate of successful amplification ranged from 76% for *T. nigrescens* to 7% for *Glycine max* and was significantly higher for *Trifolium* species when compared to species of other genera (Fig. 4). Fifty percent of all primer pairs that amplified a product in at least one other species were shown to be polymorphic across the seven white clover genotypes (data not shown).

All of the 18 PCR products which were cloned from successful cross-species amplifications were shown to contain SSR loci upon sequencing. At least one of the three to six clones sequenced for each PCR product contained an SSR locus, but clones from the same PCR reaction were not always identical. Figure 5 shows DNA sequence alignment for SSR locus TRSSRA02B09 using one representative clone per PCR reaction. The sequences are highly conserved across species and similarity ranged from 82% (when comparing *M. sativa* to *T. nigrescens*) to 100% for the comparison of *T. repens*, *T. pratense* and *M. alba*. However, similarity between clones of the same PCR reaction

Fig. 5 Sequence comparison of SSR loci in *T. repens* (*TR*), *T. pratense* (*TP*), *T. subterraneum* (*TS*), *T. ambiguum* (*TA*), *T. nigrescens* (*TN*), *M. sativa* (*MS*), and *M. alba* (*MA*). Sequences were obtained by cloning PCR products amplified using primers for locus TRSSRA02B09 and aligned using the multiple alignment procedure of Clustal W (Thompson et al. 1994). *Blocked areas* indicate sequence identity. Primer sequences are indicated by *arrows* and the *line* indicates the SSR motif. TR (1) is the sequence in which the SSR was originally discovered, TR (2) and TR (3) are two *T. repens* inbred genotypes (I5J and I4R) which were used as a control in cross-species amplification

```
TR(1)  :  CCCTTTCTCAACCCACATCATCACATTCTCCACACACACACCCACACAAACCGGTCT  :  57
TR(2)  :  CCCTTTCTCAACCCACATCATCATATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57
TR(3)  :  CCCTTTCTCAACCCACATCATCATATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57
TP     :  CCCTTTCTCAACCCACATCATCATATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57
TS     :  CCCTTTCTCAACCCACATCATGACATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57
TA     :  CCCTTTCTCAACCCACATCATCACATTCTCCACACACACTCCCATGCAAACCGGTTT  :  57
TN     :  CCCTTTCTCAACCCACATCATCATATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57
MS     :  --CTTTCTCAACCCACATCATCACATTCTCCATACACACCCCCGAGCAAACTGGCTT  :  55
MA     :  CCCTTTCTCAACCCACATCATCATATTCTCCACACACACTCCCATGCAAACCGGCTT  :  57

TR(1)  :  ACACGGTCTACACTTAGTTCCACTTAATGGACCGGGACTCGACAATGCCGCAACCGG  :  114
TR(2)  :  ACACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCAACCAG  :  114
TR(3)  :  ACACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCAACCAG  :  114
TP     :  ACACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCAACCAG  :  114
TS     :  GCACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCGACAAG  :  114
TA     :  GCACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCGACTAG  :  114
TN     :  ACACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCGACCAG  :  114
MS     :  GCACGGTCTGCACTTAGCTCCGCTTAAAGGACCGGGACTAGACGATGCTGCAACCAG  :  112
MA     :  ACACGGTCTACACTTAGCTCCACTTAATGGACCAAGACTAGACAATGCCGCAACCAG  :  114

TR(1)  :  TACACTATACCGGGCTGGTGGAAAATCGTATTTTCGGTCTGAAACAATGCGAAAA-  :  169
TR(2)  :  TACACCAAATCGGGCTGGGGGAAAGATATATTTCCGGTCTGAAACAATGCGAAAA-  :  169
TR(3)  :  TACACCAAATCGGGCTGGGGGAAAGATATATTTCCGGTCTGAAACAATGCGAAAA-  :  169
TP     :  TACACCAAATCGGGCTGGGGGAAAGATATATTTCCGGTCTGAAACAATGCGAAAA-  :  169
TS     :  TACACCGAACCGGGCCGGAGGAAAAATATATTTTCGGTCTGAAACAATGCGAAAA-  :  169
TA     :  TACACCAAATCGGGTTGGAGGAAAAATATATTTTCGGTCTGAAACAATGCGAAAA-  :  169
TN     :  TACACCAAATCGGGCTGGGGGAAAGATATATTTCCGGTCTGAAACAATGCGAAAA-  :  169
MS     :  TACACCATACTGTGCCGGAGGAAAATCGTATTTTCGGTCTGAAACAATGCGAAAA-  :  167
MA     :  TACACCAAATCGGGCTGGGGGAAAGATATATTTCCGGTCTGAAACAATGCGAAAA-  :  169
```

also ranged on average from 89% to 99%, making comparisons between individual species difficult.

Discussion

The enrichment procedure used in this study proved to be highly effective and enabled the development of several *T. repens* DNA libraries highly enriched for SSR loci. Across all libraries, more than 70% of the clones that were sequenced contained SSR loci. This corresponds to a 200 to 700-fold enrichment when compared to the 0.1 to 0.3% of SSRs identified by screening non-enriched genomic libraries (Liu et al. 1995; Szewc-McFadden et al. 1996; Kubik et al. 1999; Saal and Wricke 1999). The percentage of clones containing no SSRs is comparable to the rate of false positives encountered by colony screening (Kubik et al. 1999). However, the enrichment procedure used in this paper does not require the screening of several thousands of colonies, and, once established, libraries provide an extensive source of SSR loci, resulting in a considerable reduction of time and labour per locus.

Due to the use of $(CA)_n$ oligonucleotides for library enrichment, the vast majority of SSR repeats were dinucleotides with a (AC),(CA)/(GT),(TG) motif. Although database surveys showed that these motifs are relatively rare in plants (Morgante and Olivieri 1993; Wang et al. 1994), the high SSR enrichment reported in this study suggests that (AC),(CA)/(GT),(TG) repeats occur at a significant frequency in the white clover genome. Substantial proportions of the (AC),(CA)/(GT),(TG) repeats have also been found by hybridisation screening of genomic libraries of seashore paspalum (*Paspalum vaginatum* Schwartz; Liu et al. 1995) and rye (*Secale cereale* L.; Saal and Wricke 1999). Among the SSR motifs that did not correspond to the oligonucleotide used for enrichment, the (AT)/(TA) motif, usually the most frequent repeat in plants (Morgante and Olivieri 1993; Wang et al. 1994; Powell et al. 1996), was found at a much lower frequency when compared to the (GAA)/(TTC) or (AAG)/(CTT) motifs. These trinucleotide repeats may be of particular interest since (AAG) repeats, together with (AAT), have been found to be particularly common in intron sequences (Smulders et al. 1997).

The restriction enzyme used for the construction of genomic libraries had a clear influence on the level of SSR enrichment. A similar effect was also observed by Hamilton and Fleischer (1999) between the restriction enzymes α*Taq*I and *Dpn*II. Although the number of clones sequenced per library in the preliminary screening was small, a difference in enrichment between the two libraries selected for large-scale SSR discovery (TRSSRA and TRSSRB) persisted after more than 1000 clones were sequenced. However, the TRSSRB library yielded significantly more clones with greater than 25 bp of flanking sequence, resulting in little difference between the two libraries in terms of potentially useful SSRs. *Alu*I, the enzyme used in the construction of the TRSSRA library, has a four base-pair recognition site and is therefore more likely to cut close to an SSR array than *Dra*I, the enzyme used in the construction of the TRSSRB library, which has a six base-pair recognition site. The use of more than one restriction enzyme in library construction not only maximises the yield of potentially useful SSRs, it may also support an even distribution of SSRs across the genome and may prevent the clustering of markers in future mapping applications (Hamilton et al. 1999).

For 86% of the 117 primer pairs that were designed, an amplification product of the expected size was amplified. Given a total number of 397 unique SSR clones suitable for primer design (i.e. with more than a 25-bp flanking region), 30% of all clones sequenced are expected to successfully amplify SSRs. This rate is significantly higher than the 12% reported for tea tree [*Melalueuca alternifolia* (Maiden and Betche) Cheel] using a similar enrichment procedure (Rossetto et al. 1999), but comparable to rates of successful amplification found in rye and perennial ryegrass (*Lolium perenne* L.) using library screening techniques (Kubik et al. 1999; Saal and Wricke 1999). Primers designed for compound SSRs, trinucleotide SSRs and SSRs with more than 21 repeats showed a lower rate of successful amplification when compared to other SSR types. Compound repeats and trinucleotide repeats were generally larger than dinucleotide repeats, averaging 20 bp and 29 bp, respectively (data not shown), and larger SSRs may be more difficult to amplify. Mutational decay of the SSR and its flanking region is unlikely to be responsible for the lower amplification success of these loci since imperfect SSRs, which are believed to be degraded perfect-SSRs (Taylor et al. 1999), showed the highest rate of amplification.

Some SSR primer pairs amplified more than one allele in either or both of the inbred genotypes I5J and I4R. Although both plants were inbred for four to five generations (Michaelson-Yeates et al. 1997), a certain amount of residual heterozygosity cannot be excluded particularly in the vicinity of highly heterotic loci. Additional alleles could also be due to the allotetraploid nature of the species, corresponding to homoeolocus detection (Williams 1987). Residual heterozygosity and allotetraploidy too could be the reasons for the different variants of the same SSR detected during the sequencing of clones from enriched libraries. However, we observed one redundant SSR with four variants of the SSR motif and, in addition, single base-pair mutations in the flanking sequences of six clones. Since no more than four alleles at a single locus would be expected in a tetraploid species, other factors such as sequencing errors, mutations in the cloning step or multiple copies of the SSR locus and its flanking regions in the genome must be responsible for this variation. Mapping allelic variants amplified with diagnostic primers designed to the different SSR variants may help to determine the source of these variants.

As expected for a highly heterogeneous, outbreeding species, the percentage of polymorphic SSR loci (88% of

successful amplifications) was very high. The polymorphism information content (PIC) was comparable to values found in other outbreeding species such as rye (Saal and Wricke 1999) or apple (Gianfranceschi et al. 1998). As previously observed (Weber 1990), there was an increase in the level of polymorphism with increasing SSR array length. The differences were small, possibly due to the small range of SSR lengths (5–32 bp) observed. However, even the shortest SSRs were highly polymorphic, detecting more than four alleles with an average PIC of 0.68. Polymorphism not only occurred among genotypes of different cultivars or accessions, but also between genotypes of the same cultivar. Therefore the SSR markers developed in this study provide a valuable tool for the separation of closely related individuals.

The application of molecular markers in trait mapping and marker-assisted selection is greatly facilitated by the availability of a genetic map for the species under investigation. At this time, no such map is available for *T. repens*. The four SSR loci we tested for segregation in an F_2 population all detected single alleles in the parents and did not significantly deviate from Mendelian segregation ratios. Thus, the SSR markers reported here provide an ideal tool for the construction of a linkage map of *T. repens*.

Extending the application range of SSRs can considerably enhance the cost-benefit ratio for the development of these markers. Although the cross-species transferability of most SSRs seems to be limited to species of the same genus (Peakall et al. 1998), some primer pairs developed in one genus may amplify SSR loci in allied taxa (Brown et al. 1996). Of the 54 primer pairs tested across eight legume species, 46 showed successful amplification in at least one species. More than 70% percent of the primers amplified products in *T. ambiguum* and *T. nigrescens*. These two species are very close relatives and may have played an important role in the origin of white clover (Williams 1987). As expected, successful amplification was more frequent for species of the same genus than for more-distantly related species. The relatively high success rate in red clover (*T. pratense*) is particularly noteworthy since this is another important pasture legume species in which there has been limited development of molecular marker technologies. A large proportion of the primers showing successful cross-species amplifications are expected to detect polymorphism in these species since 50% of these primers detected polymorphism in *T. repens*.

Cloning and sequencing confirmed the presence of SSR motifs in all species from which an amplification product was analysed. Large parts of the flanking sequences of these SSR loci were highly conserved across species. However, some sequence variation was observed between clones isolated from the same PCR amplification. This could be due to the fact that SSR loci were directly cloned from the PCR amplification product and not by isolation of single bands from the gel. Sequence variation between clones of the same PCR amplification was also observed by Brohede and Ellegren (1999) who assumed the variation to be due to allelic variants. The variation between clones may also be due to the same processes responsible for the often-observed stuttering of SSR bands or to the presence of several copies of the SSR and its flanking regions.

In conclusion, this paper reports the development of a large set of highly polymorphic SSR markers for *T. repens*. Their reproducibility will allow the exchange of information across laboratories and eventually result in a framework SSR-based linkage map for *T. repens*. The possible applications of these markers range from cultivar identification, the analysis of genetic diversity to trait mapping and marker-assisted selection. Although the direct costs involved in SSR development are relatively high, SSRs will almost certainly become the markers of choice for the molecular breeding of *T. repens*.

Acknowledgments The authors thank Dr. T. Michaelson-Yeates and Dr. M. Abberton of the Department of Legume Breeding, IGER, Aberyswyth, UK, for supplying genomic DNA from the parents and progeny of the I5J×I4R population and Prof. G. Spangenberg for scientific advice and support throughout the project. This research was funded by the Department of Natural Resources and Environment, Victoria, Australia, and the Swiss National Science Foundation (fellowship to R.K.). All experiments carried out in this study comply with current Australian laws.

References

Barcaccia G, Albertini E, Tavoletti S, Falcinelli M, Veronesi F (1999) AFLP fingerprinting in *Medicago* spp.: its development and application in linkage mapping. Plant Breed 118: 335–340

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

Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. Proc Roy Soc Lond B (Biol) 266:825–833

Brouwer DJ, Osborn TC (1999) A molecular marker linkage map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 99:1194–1200

Brown SM, Szewc-McFadden AK, Kresovich S (1996) Development and application of simple sequence repeat (SSR) loci for plant genome analysis. In: Jauhar PP (ed) Methods of genome analysis in plants. CRC Press Inc., Boca Raton, Florida, pp 147–159

Chen C, Sleper DA, Johal GS (1998) Comparative RFLP mapping of meadow and tall fescue. Theor Appl Genet 97:255–260

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758–759

Faville MJ, Adam KD, Wedderburn ME (1994) Allozyme markers suitable for pasture persistence studies in cocksfoot (*Dactylis glomerata*) cv Grasslands Wana. N Z J Agric Res 37: 135–141

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. Theor Appl Genet 96:1069–1076

Hamilton MB, Fleischer RC (1999) Cloned microsatellite repeats differ between 4-base restriction endonucleases. J Hered 90: 561–563

Hamilton MB, Pincus EL, DiFiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27:500–507

Hayward MD, Forster JW, Jones JG, Dolstra O, Evans C, McAdam NJ, Hossain KG, Stammers M, Will J, Humphreys MO, Evans GM (1998) Genetic analysis of *Lolium*. I. Identification of linkage groups and the establishment of a genetic map. Plant Breed 117:451–455

Hayward MD, McAdam NJ, Jones JG, Evans C, Evans GM, Forster JW, Ustin A, Hossain KG, Quader B, Stammers M, Will JK (1994) Genetic markers and the selection of quantitative traits in forage grasses. Euphytica 77:269–275

Hedrick PW (1985) Genetics of populations. Jones and Bartlett Publ. Inc., Boston, Massachusetts

Jenczewski E, Gherardi M, Bonnin I, Prosperi JM, Olivieri I, Huguet T (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (*Leguminosae*). Theor Appl Genet 94:682–691

Joyce TA, Abberton MT, Michaelson-Yeates TPT, Forster JW (1999) Relationships between genetic distance measured by RAPD-PCR and heterosis in inbred lines of white clover (*Trifolium repens* L.). Euphytica 107:159–165

Karagyozov L, Kalcheva ID, Chapman VM (1993) Construction of random small-insert genomic libraries highly enriched for simple sequence repeats. Nucleic Acids Res 21:3911–3912

Kidwell KK, Bingham ET, Woodfield DR, Osborn TC (1994) Relationships among genetic distance, forage yield and heterozygosity in isogenic diploid and tetraploid alfalfa populations. Theor Appl Genet 89:323–328

Kijas JMH, Fowler JCS, Garbett CA, Thomas MR (1994) Enrichment of microsatellites from the citrus genome using biotinylated oligonucleotide sequences bound to streptavidin-coated magnetic particles. Biotechniques 16:656–662

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

Kumar LS (1999) DNA markers in plant improvement: an overview. Biotechnol Adv 17:143–182

Liu ZW, Jarret RL, Kresovich S, Duncan RR (1995) Characterization and analysis of simple sequence repeat (SSR) loci in seashore paspalum (*Paspalum vaginatum* Swartz). Theor Appl Genet 91:47–52

Mather RDJ, Melhuish DT, Herlihy M (1996) Trends in the global marketing of white clover cultivars. In: Woodfield DR (ed) New Zealand's competitive edge. Proceedings of a symposium held November 21–22, Lincoln University, New Zealand, pp 7–14

Michaelson-Yeates TPT, Marshall A, Abberton MT, Rhodes I (1997) Self-compatibility and heterosis in white clover (*Trifolium repens* L.). Euphytica 94:341–348

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3:87–103

Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plants. Plant J 3:175–182

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera – implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238

Rafalski JA, Vogel JM, Morgante M, Powell W, Andre C, Tingey SV (1996) Generating and using DNA markers in plants. In: Birren B, Lai E (eds) Non-mammalian genomic analysis: a practical guide. Academic Press, Inc., San Diego, pp 75–135

Rossetto M, McLauchlan A, Harriss FCL, Henry RJ, Baverstock PR, Lee LS, Maguire TL, Edwards KJ (1999) Abundance and polymorphism of microsatellite markers in the tea tree (*Melaleuca alternifolia*, *Myrtaceae*). Theor Appl Genet 98:1091–1098

Saal B, Wricke G (1999) Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome 42:964–972

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphism among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor Appl Genet 97:264–272

Staub JE, Serquen FC, Gupta M (1996) Genetic markers, map construction, and their application in plant breeding. Hortscience 31:729–741

Sweeney PM, Danneberger TK (1995) RAPD characterization of *Poa annua* L. populations in golf course greens and fairways. Crop Sci 35:1676–1680

Szewc-McFadden AK, Kresovich S, Bliek SM, Mitchell SE, McFerson JR (1996) Identification of polymorphic, conserved simple sequence repeats (SSRs) in cultivated *Brassica* species. Theor Appl Genet 93:534–538

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Taylor JS, Durkin JMH, Breden F (1999) The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. Mol Biol Evol 16:567–572

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1–6

Weber JL (1990) Human DNA polymorphism based on length variations in simple-sequence tandem repeats. In: Davies KE, Tilghman SM (eds) Genetical and physical mapping. Volume 1 Genome analysis. Cold Spring Harbour Laboratory, Cold Spring Harbor, New York, pp. 159–181

Weising K, Winter P, Huttel B, Kahl G (1998) Microsatellite markers for molecular breeding. In: Basra AS (ed) Crop sciences: recent advances. The Haworth Press Inc., Binghamton, New York, pp 113–143

Williams WM (1987) Genetics and breeding. In: Baker MJ, Williams WM (eds) White clover. CAB International, Wallingford, Oxon, pp 343–419

Woodfield DR, Caradus JR (1994) Genetic improvement in white clover representing six decades of plant breeding. Crop Sci 34:1205–1213