Theor Appl Genet (1994) 88:1–6

Z. Wang · J. L. Weber · G. Zhong · S. D. Tanksley

Survey of plant short tandem DNA repeats

Received: 27 April 1993 / Accepted: 26 July 1993

Abstract Length variations in simple sequence tandem repeats are being given increased attention in plant genetics. Some short tandem repeats (STRs) from a few plant species, mainly those at the dinucleotide level, have been demonstrated to show polymorphisms and Mendelian inheritance. In the study reported here a search for all of the possible STRs ranging from mononucleotide up to tetranucleotide repeats was carried out on EMBL and GenBank DNA sequence databases of 3026 kb nuclear DNA and 1268 kb organelle DNA in 54 and 28 plant species (plus algae), respectively. An extreme rareness of STRs (4 STRs in 1268 kb DNA) was detected in organelle compared with nuclear DNA sequences. In nuclear DNA sequences, $(AT)_n$ sequences were the most abundant followed by $(A)_n \cdot (T)_n$, $(AG)_n \cdot (CT)_n$, $(AAT)_n \cdot (ATT)_n$, $(AAC)_n \cdot (GTT)$, $(AGC)_n \cdot (GCT)_n$, $(AAG)_n \cdot (CTT)_n$, $(AATT)_n \cdot (TTAA)_n$, $(AAAT)_n \cdot (ATTT)_n$ and $(AC)_n \cdot (GT)_n$ sequences. A total of 130 STRs were found, including 49 $(AT)_n$ sequences in 31 species, giving an average of 1 STR every 23.3 kb and 1 $(AT)_n$ STR every 62 kb. An abundance comparable to that for the dinucleotide repeat was observed for the tri- and tetranucleotide repeats together. On average, there was 1 STR every 64.6 kb DNA in monocotyledons versus 1 every 21.2 kb DNA in dicotyledons. The fraction of STRs that contained G-C basepairs increased as the G + C contents went up from dicotyledons, monocotyledons to algae. While STRs of mono-, di- and tetranucleotide repeats were all located in non coding regions, 57% of the trinucleotide STRs containing G-C basepairs resided in coding regions.

Communicated by F. Salamini

Z. Wang (✉) · J. L. Weber
Marshfield Medical Research Foundation, 1000 N. Oak Avenue, Marshfield, WI 54449, USA

G. Zhong
Department of Agronomy and Range Sciences, University of California, Davis, CA 95616, USA

S. D. Tanksley
Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA

Key words Short tandem repeat (STR) · DNA sequences · Molecular marker

Introduction

Short tandem repeats (STRs or microsatellites) can be defined as relatively short runs of tandemly repeated (STR) DNA with repeat lengths of 6 bp or less. STRs have been proven to be highly informative and abundant and evenly distributed in human and other mammalian genomes (Tautz and Renz 1984, Weber and May 1989). They can also be efficiently analyzed using the polymerase chain reaction (PCR). Thousands of mammalian STRs DNA polymorphisms have been developed and mapped to specific chromosomes (NIH/CEPH Collaborative Mapping Group 1992; Weissenbach et al. 1992; McAleer et al. 1992, Serikawa et al. 1992).

Several initial investigations of plant STRs in chickenpea (Weising et al. 1989), barley (Weising et al. 1989), tropical trees (Condit and Hubbell 1991), soybean (Akkaya et al. 1992; Morgante and Olivieri 1993) and rice (Zhao and Kochert 1992; Wu and Tanksley 1993) have demonstrated the informativeness, random distribution and Mendelian inheritance of STRs in a few plant species. Morgante and Olivieri (1993) searched the sequences of EMBL and GenBank databases up to the level of trinucleotide repeats and reported that STRs are also frequently and widely distributed in most plant species with one STR every 50 kb (taking di- and trinucleotide repeats together) (Morgante and Olivieri 1993). It seems that the application of STRs may be as important in plant genetics as in mammalian genetics.

However, the above-mentioned searches of genomic libraries were all conducted with limited number of repeat motifs of di- and trinucleotides, and almost all of the STRPs developed have been only dinucleotide repeats for a few plant species. Morgante and Olivieri's findings that $(AT)_n$ sequences were the most frequently observed class of STRs with $(ATA)_n$ prevailing among trinuclneotides in plant species should serve as a general guide in the development of plant STRs, but detailed information on individual species is needed especially when types of STRs and their frequencies are concerned. In

addition, information about tetranucleotide repeats in plants needs to be collected since STRs with a repeat length of 3 bp or longer may be of particular importance as their PCR artifacts are reduced compared to those of dinucleotides. Finally, a separate search or analysis of organelle DNA sequences from nuclear DNA sequences needs to be done in the interest of developing genetic markers in linkage mapping since organelle DNA does not obey Mendel's laws of inheritance.

A computer sequence search of 3026 kb nuclear DNA and 1268 kb organelle DNA in 54 plant species (plus algae) was conducted to detect all of the STRs through to tetranucleotide repeats within EMBL and GenBank databases. In this article, relative abundances, lengths and locations relative to the coding regions of plant STRs are presented. More importantly, we report here information on tetranucleotide STRs in plant species, dramatic differences in the relative abundances of STRs between organelle and nuclear DNA sequences and differences in both frequencies and types of STRs among monocotyledons, dicotyledons and algae.

Materials and methods

Sequence search

Searches were performed using PC/GENE software (Intergenetics) release 6.7. Sequences of both nuclear and organelle DNA of 54 plant species (plus algae) found within EMBL database version 31 and UGenBank 72-31 (containing sequences which were not present in EMBL) were first extracted using the program SELECT. In 28 of the 54 plant species (plus algae) 1268 kb of organelle DNA sequences were excluded from the nuclear DNA sequences and carried out in a separate search. The QSEARCH program was then used to identify all possible mononucleotide through to tetranucleotide STRs with a minimum length of 20 bp (for example ten dinucleotide or five tetranucleotide repeats) without mismatch. An independent search was carried out on EMBL database version 29 as a check of the primary results.

Data processing

To avoid bias in the results, all STRs were counted for each individual sequence entry. Duplications of specific STRs at one locus within a single species were removed. Ten $(AAC)_n$ STRs in wheat for example, all from sequences associated with gliadin genes, were reduced to a single hit. Species having 39 or more entries were arbitrarily termed 'major' species and were separated in analysis from "other" species that contain 29 entries or less.

Results and Discussion

Low frequency of short tandem repeats in organelle DNA

An extremely low frequency of STRs was found in organelle DNA sequences compared with nuclear DNA sequences. Only 4 STRs were detected from 1268 kb of organelle DNA which contains 961 kb chloroplast and 307 kb mitochondrial DNA in contrast with the 130 STRs found in 3026 kb nuclear DNA. This was unexpected considering that organelle DNA had roughly the same G + C content as the nuclear DNA of dicotyledons which, in turn, was rich in STRs. All 4 STRs were found in chloroplast (cp) DNA sequences. The sequences are $(AT)_n$ in CHPSGPA1 (pea GPA1 gene for subunit A of chloroplast glyceraldehyde-3-phosphate dehydrogenase), $(ATC)_n$ in CCPATPC2 (*Arabidopsis thaliana* chloroplast ATP synthase Gamma subunit gene), $(AAG)_n$ in ATGAPAD (chloroplast *Arabidopsis thaliana* glyceraldehyde-3-phosphate dehydogenase A subunit gene) and $(AATT)_n$ in CHCPPSBA (*Cyanophora paradoxa* cyanelle PSBA gene). $(AT)_n$ in CHPSGPAI was found to be located in the coding region (exon 5) and the other 3 were in noncoding regions. Below are results from only nuclear DNA sequences.

Pooled results over all plant species

Numbers of STRs classified by repeat motif and species are displayed in Table 1. Note that each entry in the table includes all of the permuted and complementary repeats. The AAC entry for example includes all STRs with AAC, ACA, CAA, GTT, TGT or TTG repeats. A total of 49 possible different repeat motifs through to the level of tetranucleotides exist: 2 mononucleotides, 4 dinucleotides, 10 trinucleotides and 33 tetranucleotides. Twenty-three motifs with a minimum length of 20 bp were found in the plant species on the whole; 26 motifs were absent from this version of the database.

$(AT)_n$ sequences were the most abundant, with 1 $(AT)_n$ every 62 kb, followed by the $(A)_n \cdot (T)_n$, $(AG)_n \cdot (CT)_n$, $(AAT)_n \cdot (ATT)_n$, $(AAC)_n \cdot (GTT)$, $(AGC)_n \cdot (GCT)_n$, $(AAG)_n \cdot (CTT)_n$, $(AATT)_n \cdot (TTAA)_n$, $(AAAT)_n \cdot (ATTT)_n$ and $(AC)_n \cdot (GT)_n$ sequences (Table 1). The $(AT)_n$ sequence was the most abundant not only in total number but also individually in almost all of the major species (Table 1). In comparison, other types of STRs occurred in only 6 or fewer major species. Out of a total of 3026 kb DNA, 130 STRs (> 20 bp in length) were found, including 49 with $(AT)_n$ repeats.

Table 2 summarizes the richness of STRs up to the level of tetranucleotide repeats. Dinucleotide STRs occurred the most frequently (0.477, as a fraction of the total) while mono- and tetranucleotide STRs were the least common (0.162 and 0.138, respectively). There existed 1 STR every 144 kb for mononucleotide repeat motifs, every 49 kb for di-, 104 kb for tri-, every 168 kb for tetranucleotide repeat motifs and every 23.3 kb if all were combined (Table 2). There was still 1 STR per 27.8 kb DNA even when mononucleotide STRs were excluded from the calculations.

Of the plant STRs detected in this survey 64% had total uninterrupted lengths of 20–29 bp and 82% were 20–39 bp long (Table 3). The frequencies decreased with the increased length of STRs of no interruption (Table 3). A similar tendency has been reported for humans (Beckmann and Weber 1992).

The locations of STRs relative to the coding regions are also shown in Table 3. STRs of mono-, di- and tetranucleotide repeats are all located in noncoding regions; STRs of trinucleotide repeats can be categorized into two types on the basis of whether or not they contained G-C basepairs: 12 out of 21 (57%) of those containing G-C basepairs were found in coding regions. In contrast, all of the 8 $(AAT)_n \cdot (TTA)_n$ repeats were in noncoding regions.

Table 1 Number of short tandem repeats of 23 motifs in different plant species (genus)

Species[a]	Motifs																										
	Mbp/C[b]	%G + C	AT	A	AG	ATA	AAC	AGC	AAG	AATT	AAAT	AC	ACTC	CGC	AAAG	AAAC	ATC	AGAT	ACAT	ACCC	ACGC	ACGG	GAG	GGGC	C	Total
Algae		59.2						1				2								1	1			1		6
Monocotyledonous plants																										
Hordeum vulgare	4 873	48.6	1					1	1	1	2				1								1			9
Oryza sativa	431	45.9	2				2	3						1												7
Triticum aestivum	15 966	48.0	2					1																		7
Zea mays	2 504	48.3	4		1																					
		$\bar{X} = 47.7$																								
Dicotyledonous plants																										
Arabidopsis thaliana	145	59.2	5	4			1																			10
Brassica	759	40.0		2	1																					3
Glycine max	1 115	35.2	8	2	1	1					1															13
Lycopersicon esculentum	954	35.2	5	2		1																				8
Nicotiana	3 960	37.1	9	1	1	1																			1	13
Petunia	1 274	37.0	2	1		2						1														6
Pisum sativum	4 172	35.5	2			2			1	1																6
Solanum tuberosum	1 730	34.0	3		1				1	1			2													8
		$\bar{X} = 39.2$																								
Other species			6	8	5	1	3		2	1	1			1		1	1	1	1			1				33
Total			49	20	10	8	6	6	5	4	4	3	2	2	1	1	1	1	1	1	1	1	1	1	1	130

[a] Some of species groups (genus) contain more than one species: algae – *Chlamydomonas reinhardtii, Euglena gracilis*; *Brassica-Brassica napus, Brassica nigra, Brassica oleracea*; *Nicotiana-Nicotiana plumbaginifolia, Nicotiana rustca, Nicotiana sylvestris, Nicotiana tabacum*; *Petunia-Petunia hybrida, Petunia* sp.; other-*Antirrhinum majus, Armoracia rusticana, Avena sativa, Bertholletia excelsa, Chlorella ellipsoidea, Cucumis sativus, Cucurbita pepo, Cyanophora paradoxa, Daucus carota, Dianthus caryophyllus, Fragaria ananassa, Gossypium hirsutum, Ipomoea batatas, Lemna gibba, Lupinus luteus, Medicago sativa, Mesembryanthemum crystallinum, Oryza rufipogon, Petroselium crispum, Phaseolus vulgaris, Raphanus sativus, Ricinus communis, Scilla siberica, Secale cereale, Sesbania rostrata, Sinapis alba, Sorghum bicolor, Sorghum vulgare, Spinacia oleracea, Symbiodinium* sp., *Trifolium repens, Tripsacum dactyloides, Triticum durum, Triticum* sp., *Vicia faba, Vigna radiata, Volvox carteri*

[b] Number of nucleotides (megabase) per haploid genome

Table 2 Numbers, relative abundance and distances between adjacent short tandem repeats (STRs) for mono-, di-, tri- and tetranucleotide STRs

Motifs		Number of STRs	Fraction of total	Distances (kb) between adjacent STRs
Mono		21	0.162	144.1
Di		62	0.477	48.8
	AT	49	0.377	61.8
	AG	10	0.077	302.6
	AC	3	0.023	1008.7
Tri		29	0.223	104.3
Tetra		18	0.138	168.1
Total		130		23.3(27.8)[a]

[a] When mono-STRs are excluded from the calculation

Differences among plant species

The data were also analyzed separately for 13 major plant species (Tables 1 and 4). For 10 out of 13 major species there was 1 STR from every 11kb (*Petunia*) to every 25.4 kb (*Brassica*). Less frequent rates of STRs were found in *Arbidopsis thaliana*, *Zea mays* and *Hordeum vulgare*, with 1 STR every 42.4, 58.6 and 156 kb DNA, respectively. Notice that when only di-, tri- and tetranucleotide repeats were considered the average distance between STRs for most of the dicotyledonous plants rose since mononucleotide STRs were found only in dicotyledonous plants. Rice was most the abundant in types of STRs (7 motifs detected) among the 31 species in which STRs were uncovered, while 3–5 STRs were found for most of the major species, with only the $(AT)_n$ type of STRs being found in *Hordeum vulgare*.

When species were broken down into algae, monocotyledons and dicotyledons, differences could be seen in both numbers and types of STRs. STRs occurred roughly 3 times more frequently in dicotyledons than in monocotyledons. On average, there was 1 STR every 64.6 kb DNA in monocotyledons versus 1 every 21.2 kb DNA in dicotyledons (Table 4). The frequency (distance between adjacent STRs) of 19.4 kb in algae was comparable to that of dicotyledon (Table 4). Of the STRs found in dicotyledons 15% contained G-C basepairs while as many as 50% STRs with G-C basepairs could be found in monocotyledons (Table 1). STRs in algae all contained G-C basepairs (Table 1).

The nuclear DNA contents (Armuganathan and Earle 1991) and G + C contents are also listed in Tables 1 and 4. G + C contents were calculated from sequences in this study and found to be in good agreement ($r = 0.87$) with previous estimations (Messeguer et al. 1991). Nuclear DNA contents and G + C contents were investigated for associations with both frequencies and types of STRs in monocotyledons, dicotyledons and algae. Between monocotyledons and dicotyledons it seemed, on average, that G + C contents were positively correlated with the fraction of STRs containing G + C basepairs (Table 1) and inversely correlated with the frequencies of STRs (Table 4). The observation that algae had the highest G + C content (59.2%) could account for the finding that all of the STRs found in algae contained G + C basepairs. As to the nuclear DNA contents there was no associations of genome size with either frequencies or types of STRs in the plant species.

The above results were confirmed by an independent search of EMBL database version 29. The only differences were that 5 rare STRs were missed in that version, GAG, ACAT, AAAG, AAAC and GCCC (data not shown here), which can be explained by the fact that 366 entries of version

Table 3 Size (bp), fraction of total and locations relative to coding region for 23 repeat motifs in plant species

Motif	Number	Size (bp) 20–29	30–39	40–49	≥ 50	Fraction	Location
AT	49	25	12	9	3	0.3551	Non-coding
A	20	17	2	1		0.1449	Non-coding
AG	10	4	4	1	1	0.0725	Non-coding
ATA	8	2	2	1	3	0.0580	Non-coding
AAC	6	3	2		1	0.0435	4/6 coding
AGC	6	4	1	1		0.0435	3/5 coding
AAG	5	4	1			0.0362	2/5 coding
AATT	4	4				0.0290	Non-coding
AAAT	4	4				0.0290	Non-coding
AC	3	1		1	1	0.0217	Non-coding
ACTC	2	2				0.0145	No information
CGC	2	2				0.0145	Coding
AAAG	1	1				0.0072	No information
AAAC	1	1				0.0072	Non-coding
ATC	1	1				0.0072	Non-coding
AGAT	1	1				0.0072	Non-coding
ACAT	1	1				0.0072	Non-coding
ACCC	1				1	0.0072	Non-coding
ACGC	1	1				0.0072	Non-coding
ACGG	1	1				0.0072	Non-coding
AGG	1	1				0.0072	Coding
GGGC	1	1				0.0072	Non-coding
C	1	1				0.0072	Non-coding
Total	130	82	24	14	10		

Table 4 Number of STRs and distance between adjacent STRs for some plant species (Mbp/C, number of nucleotides (megabase) per haploid genome)

Species	Nuclear DNA content		DNA Surveyed		STRs[a]	
	Mbp/C	%G + C	Entries	Length(kb)	Number	Distance (Kb)
Algae		59.2	48	116.7	6	19.4
Monocotyledonus plants						
Hordeum vulgare	4 873	48.6	59	156.4	1	156
Oryza sativa	431	45.9	168	208	9	23
Triticum aestivum	15 966	48.0	97	144	7	20.6
Zea mays	2 504	48.3	205	410.5	7	58.6
		$\bar{X}$(mono) = 47.7				$\bar{X}$ = 64.6
Dicotyledonous plants						
Arabidopsis thaliana	145	59.2	214	424.2	10(6)	42.4(70.7)
Brassica	759	40.0	39	76.1	3(1)	25.4(76.1)
Glycine max	1 115	35.2	101	199.7	13(11)	15.4(18.2)
Lycopersicon esculentum	954	35.2	94	195.3	8(6)	24.4(32.6)
Nicotiana	3 960	37.1	96	203.8	13(12)	15.7(17.0)
Petunia	1 274	37.0	46	66	6(5)	11(13.2)
Pisum sativum	4 172	35.5	51	105.9	6	17.1
Solanum tuberosum	1 730	34.0	85	145.8	8	18.2
		$\bar{X}$(di) = 39.2				$\bar{X}$ = 21.2
Other species			309	1287.1	33(25)	39.0(51.5)
Total			1612	3026.2	130(109)	23.3(27.8)

[a] Number in brackets were obtained when mono-STRs, if any, were removed

31 were absent from version 29. However, caution should be exercised when using results obtained on different major species with respect to the relative frequencies and number of types of STRs: the lengths of DNA sequences surveyed in each species showed a six-fold variation, from 66 to 424.2 kb, and this could bring some potential bias in comparisons among different species.

Our results support Morgante and Olivieri's findings that $(AT)_n$ sequences are by far the most abundant STRs in plants and that the dinucleotide STRs detected are all located in noncoding regions while only some trinucleotide STRs are found to be located in coding regions. We extend their findings by showing that all of the STRs of mono-, di- and tetranucleotide repeats are located in noncoding regions; it is only the trinucleotide STRs containing G-C basepairs that were often located in coding regions in contrast to noncoding regions of all $(AAT)_n$ · $(TTA)_n$ sequences. One STR every 50 kb (including di- and trinucleotide repeats only) was reported in Morgante and Olivieri's study in which organelle and nuclear DNA sequences were searched and analyzed together. A separate search of organelle DNA sequences in this study made it possible to detect the dramatic difference in abundance of STRs between organelle and nuclear DNA sequences. Taking di- and trinucleotide repeats together we found there was 1 STR every 33.2 kb of nuclear DNA sequence versus 1 every 422.7 kb of organelle DNA sequence. If organelle and nuclear DNA are pooled together there will be one STR every 45.7 kb of DNA sequence, which approximates the 50 kb reported by Morgante and Olivieri (1993).

The physical distance of 1 cM in plants ranges roughly from 150 kb (Arabidopsis), 300 kb (rice), 500 (tomato), 1000 kb (potato) to 1500 kb (maize). Given the frequency of $(AT)_n$ every 62 kb, it would be expected to find, on average, at least 2.5 $(AT)_n$ STRs within 1 cM for as small a genome as that of *Arabidopsis*. For a medium-sized genome like in tomato, 8 $(AT)_n$ STRs would be expected per cM. STRs of di-, tri- and tetranucleotide repeats should themselves be in sufficient numbers as to be a reservoir of genetic markers in plants for linkage mapping and even for STS physical mapping.

An important result of this survey is the finding that STRs with tri- and tetranucleotide repeats are relatively abundant in plant genomes. Seven motifs of trinucleotide repeats and 11 motifs of tetranucleotide repeats were uncovered in this survey. Together, tri- and tetranucleotide STRs were as abundant as $(AT)_n$ sequences (Table 2), with 1 STR of a tri- or tetranucleotide level every 62 kb DNA. The apparent strand-slippage which occurs during PCR amplification of dinucleotide repeat STRs (Smeets et al. 1989; Luty et al. 1990) is significantly diminished as the repeat length increases (Economou et al. 1990; Zuliani and Hobbs 1990). The development of STR markers with repeat lengths $\geq$ 3 bp which match the informativeness of dinucleotide repeat markers would represent a significant improvement in the utility of these polymorphisms.

Another interesting finding is that 57% of the trinucleotide repeat STRs that contain G-C basepairs are within coding regions. The distribution of interspersed repeats close to and even within genes has brought the trinucleotide repeat polymorphisms into the spotlight of human molecular genetics. The instability of the trinucleotide repeat has been found to be responsible for several important hereditary human diseases. In the case of the fragile X syndrome, amplification of the $(CCG)_n$ repeat is involved (Kremer et al. 1991), while for myotonic dystrophy and Kennedy's disease the amplified repeat is the $(AGC)_n$ (Brook et al. 1992; Fu et al. 1992). The instability of the trinucleotide repeat sequence is likely to provide a new mechanism of mutation.

In summary, the relative abundance of STRs in plants should be encouraging news for those who are anxious to

develop highly informative markers in linkage mapping. Because of their high informativeness and ease of analysis, STRs in plants, as in humans, could be of great benefit in linkage mapping. STRs can also provide a new tool for various studies in plant genetics and breeding, systematics, strain identification and other germ-plasm-related studies.

References

Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–218

Beckmann JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627–631

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, Sohn R, Zemelman B, Snell RG, Rundle SA, Crow S, Davies J, Shelbourne P, Buxton J, Jones C, Juvonen V, Johnson K, Harper PS, Shaw DJ, Housman DE (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3'end of a transcript encoding a protein kinase family member. Cell 68: 799–808

Condit R, Hubbell SP (1991) Abundance and DNA sequence of 2-base repeat regions in tropical tree genomes. Genome 34:66–71

Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc Natl Acad Sci USA 87:2951–2954

Fu Y-H, Pizzuti A, Fenwick RG, King J Jr, Rajnarayan S, Dunne PW, Dubel J, Nasser GA, Ashizawa T, de Jong P, Wieringa B, Korneluk R, Perryman MB, Epstein HF, Caskey CT (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. Science 255:1256–1258

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Waren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CCG)_n$. Science 252:1711–1714

Luty JA, Guo Z, Willard HF, Ledbetter DH, Ledbetter S, Litt M (1990) Five polymorphic microsatellite VNTRs on the human X chromosome. Am J Hum Genet 46:776–783

McAleer MA, Aitman TJ, Cornall RJ, Ghosh S, Hall JRS, Hearne CM, Love JM, Prins JB, Ramachandran S, Rodrigues N, DeLarato NH, Peterson LB, Wicker LS, Todd JA (1992) Linkage analysis of 84 microsatellite markers in intra and interspecific backcrosses. Mamm Genome 3:457–460

Messeguer R, Ganal MW, Steffens JC, Tanksley SD (1991) Characterization of the level, target sites and inheritance of cytosin methylation in tomato nuclear DNA. Plant Mol Biol 16:753–770

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

NIH/CEPH Collaborative Mapping Group (1992) A comprehensive genetic linkage map of the human genome. Science 258:67–86

Serikawa T, Kuramoto T, Hilbert P, Mori M, Yamada J, Dubay CJ, Lindpainter K, Ganten D, Genenet J-L, Lathrop M, Beckmann JS (1992) Rat gene mapping using PCR-analyzed microsatellites. Genetics 131:701–721

Smeets HJM, Brunner HG, Ropers HH, Wieringa B (1989) Use of variable simple sequence motifs as genetic markers: Application to study of myotonic dystrophy. Hum Genet 83:245–251

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of aukaryotic genomes. Nucleic Acids Res 12:4127–4138

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weising K, Weigand F, Driesel AJ, Kahl G, Zischler H, Epplen JT (1989) Polymorphic simple GATA/GACA repeats in plant genomes. Nucleic Acids Res 17:10128

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Wu K, Tanksley SD (1993) Abundances, polymorphisms and genetic mapping of microsatellites in rice. Mol Gen Genet (in press)

Zhao X, Kochert G (1992) Characterization and genetic mapping of a short, highly repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol Gen Genet 231:353–359

Zuliani G, Hobbs HH (1990) A high frequency of length polymorphisms in repeated sequences adjacent to Alu sequences. Am J Hum Genet 46:963–969