Theor Appl Genet (1996) 92:326–333
© Springer-Verlag 1996

M. Mörchen · J. Cuguen · G. Michaelis
C. Hänni · P. Saumitou-Laprade

Abundance and length polymorphism of microsatellite repeats in *Beta vulgaris* L.

Received: 27 April 1995 / Accepted: 8 September 1995

Abstract Simple sequence repeats (SSRs) are known to exhibit high degrees of variability even among closely related individuals. Their usage as nuclear genetic markers requires their conversion into sequence-tagged sites (STSs). In this paper we present the development of simple sequences as STSs for *Beta vulgaris*. This species comprises wild, cultivated, and weedy forms; the latter are thought to originate from accidental hybridisation between the other two. Two partial genomic libraries were screened with simple sequence motifs (AT, CA, CT, ATT, GTG, and CA, CT, respectively). Clones of 22 CA, nine CT, eight ATT, and one GTG sequence were obtained. AT microsatellites were present in compound motifs, not recognised by the probe. Sequence comparisons revealed that 20 CA clones containing short motifs (< 16 bp) were variants of a previously described approximately 320-bp satellite DNA (Schmidt et al. 1991), and hence did not correspond to unique loci. Polymorphism of one $(ATT)_{15}$ and three $(CT)_n$, with n = 15, 17 and 26, was detected by PCR on a sample of 64 plants from the different forms of *B. vulgaris*. 13 (ATT), 13 (CT), nine (CT) alleles and one (CT) allele were detected. One of the ATT alleles was much larger than the others (> 800 bp). Genetic variability was high among wild beets, lower among cultivated beets, and intermediate among weed beets. One allele of each locus was found at high frequencies in cultivated beets and, to a lower extent, in weed beets. The combination of three polymorphic loci allowed the individual identification of 17/17 wild and 15/15 weed beets, and 21/32, mostly homozygous, cultivated beets.

Key words Beet · VNTR · Simple sequence length polymorphism · Microsatellite · Sequence-tagged sites

Communicated by G. Wenzel

M. Mörchen · J. Cuguen · P. Saumitou-Laprade (✉)
Laboratoire de Génétique et Evolution des Populations Végétales, URA-CNRS 1185, Bât. SN2, Université de Lille 1, F-59655 Villeneuve d'Ascq Cedex, France

J. Cuguen
Université de Lille 1, Institut Agricole et Alimentaire de Lille, F-59655 Villeneuve d'Ascq Cedex, France

G. Michaelis
Botanisches Institut der Heinrich-Heine Universität Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Federal Republic of Germany

C. Hänni
Unité d'Oncologie Moléculaire, URA-CNRS 1160, Institut Pasteur de Lille, 1 rue Calmette, F-59019 Lille Cedex, France

Introduction

Concomitant with the development of novel techniques in molecular biology, population genetic methods have become more efficient at acquiring information on genetic polymorphism. Initially revolutionised by the RFLP (restriction fragment length polymorphism) method, the high variability of minisatellites or VNTRs (variable number of tandem repeats) (Jeffreys et al. 1985; Nakamura et al. 1987) has facilitated the distinction between closely related individuals and have also been used for paternity analysis.

In recent years another kind of VNTR has attracted much attention: namely, microsatellites or simple sequence repeats (SSRs) (Litt and Luty 1989; Tautz 1989; Weber and May 1989). These consist of tandem repeats of very short motifs (1–6 bp), which are dispersed throughout eukaryotic genomes (Tautz and Renz 1984). Most probably due to "slippage" events during replication (Levinson and Gutman 1987; Schlötterer and Tautz 1992) the number of repetitions is in general highly variable. This polymorphism can be detected by the PCR technique using primers complementary to unique flanking sequences. In distinction to PCR using arbitrary primers, microsatellites can therefore be converted into sequence-tagged sites (STSs) (Olson et al. 1989). Combining a high abundance with an even distribution in the genome, microsatellites are highly suitable as nuclear genetic markers.

Microsatellites have already served for the construction of genetic maps of the human (Weissenbach et al. 1992) and four other mammalian genomes (e.g. Serikawa et al. 1992). Applications in plant biology were initiated later, beginning with the screening of tropical tree genomes for CA and CT repeats (Condit and Hubbell 1991); these are also the most abundant dinucleotides in mammalian genomes. Subsequent reviews of the abundance and characteristics of SSRs in published plant DNA sequences (Lagercrantz et al. 1993; Morgante and Olivieri 1993; Wang et al. 1994) showed that the situation in plants is quite different. The average density of SSRs in plant genomes (one every 23.3 kb, Wang et al. 1994) is about four times lower than in the human genome. Nevertheless a comparison of the results for 12 angiosperm species of widely differing genome sizes reveals a large range of frequencies, varying from one repeat every 11 kb for *Petunia* to one every 156 kb for *Hordeum vulgare* (Wang et al. 1994). Concerning the character of the repeats, the five most abundant repeats in humans are A, CA, A_3T, CT, and AT in descending order, whereas in plants AT, A, CT, AAT, and AAC are the most abundant (Beckmann and Weber 1992; Wang et al. 1994). Thus screening for dinucleotides may be more difficult in plants, because the probe AT is self-complementary.

According to these results, the search for microsatellites in plants should employ the motifs AT, CT, and ATT. Screening for CA repeats should nevertheless be undertaken, since CA microsatellites showing polymorphism have been reported for grapevine (Thomas and Scott 1993) and rice (Wu and Tanksley 1993). CA repeats can, however, be rare and/or uninformative, as is the case in soybean (Akkaya et al. 1992) and *Arabidopsis* (Bell and Ecker 1994).

Due to their high levels of polymorphism, plant microsatellites have provided new markers for constructing genetic maps (Wu and Tanksley 1993) and for unambiguously distinguishing cultivars in different plant species (Yang et al. 1994; Rongwen et al. 1995). In addition to this high level of polymorphism, the application of microsatellites as sequence-tagged sites is more convenient than classical RFLPs, once the procedure for their isolation and definition has been overcome.

In the beet genome, the occurrence of short repetitive DNA sequences has been reported on the basis of oligonucleotide fingerprinting (Weising et al. 1991; Schmidt et al. 1993). The *Beta vulgaris* section comprises one of the major crop plants of Northern Europe, including the cultivated sugar beet (*Bv* ssp. *vulgaris*), and its wild relative (*Bv* ssp. *maritima*), which is largely distributed along the sea coasts, although is also present inland (South-Western France). Interestingly it also includes weed beets which have expanded into areas of European sugar beet production since the early 1970s, and which are thought to result from accidental hybridization between cultivated and wild beets in areas of commercial seed production (South-Western France) (Boudry et al., 1993). The characterisation of microsatellites as sequence-tagged sites in *B. vulgaris* is desirable for the further evaluation of several problems, such as the distinction between numerous cultivated variants, population genetic studies on wild plants, gene flow between the different forms of beets and the origin of the weed beets, as well as for risk assessments in relation to the use of transgenic plants.

In this paper we present the development of simple sequences as sequence-tagged sites for *B. vulgaris* and report on the screening results for two partial genomic libraries, the characterisation of positive clones, and the results of PCR for four loci in a sample from the three *Beta* forms.

Materials and methods

Plant material

The plants analysed comprised 64 *B. vulgaris* (L.) accessions:

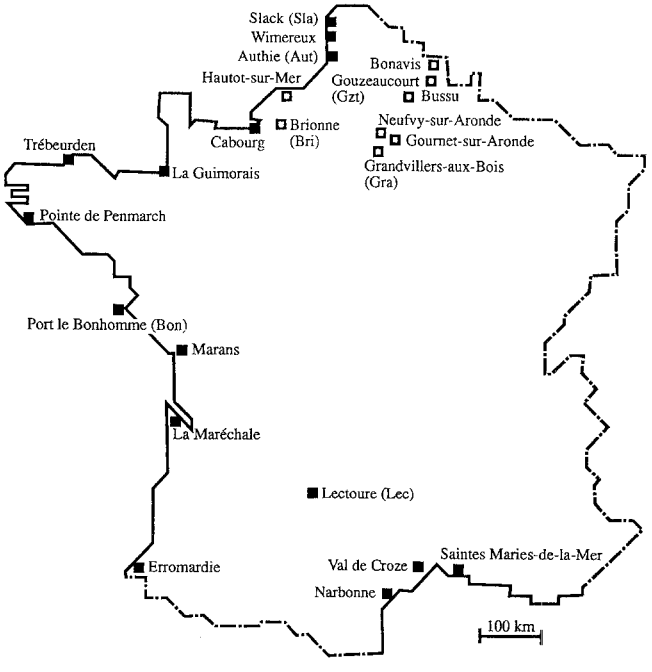
(1) *Bv.* ssp. *maritima* (wild beet). Seeds from 17 plants were collected in 13 coastal populations distributed along the French sea coast (Fig. 1), and in two inland populations located in Southern France (Val de Croze) and South-Western France (Lectoure).

(2) *Bv.* ssp. *vulgaris* (cultivated beet). Thirty two accessions were kindly supplied by two breeding companies, Florimond-Desprez, Cappelle-en-Pévèle (France) and Kleinwanzlebener Saatzucht AG, Einbeck (FRG), and by INRA, Dijon (France).

(3) Weed beets. Seeds collected on 15 plants from eight populations in three regions from the sugar production area in Northern France (Fig. 1) were kindly supplied by P. Boudry and H. van Dijk.

In order to test for Mendelian inheritance, parents and 20 plants of the F_1 generation from a reciprocal cross between two wild beets were used. Goodness of fit tests were performed following Sokal and Rohlf (1981), using BIOM package ver. 2.1 (Rohlf 1992).

Fig. 1 Location of the wild (*filled squares*) and weed (*open squares*) beet accessions



Cloning, selection and sequencing of microsatellite sequences

Nuclear DNA was extracted from leaves of a wild beet as described by Pillen et al. (1992). After digestion with *Alu*I, *Hae*III, and *Rsa*I, and separation on a 1% agarose gel, fragments of 200–500 bp were cut out and extracted using a dialysis tube. A Qiaex extraction (Diagen) was added in order to remove traces of agarose. Fragments were cloned into the *Sma*I site of pUC19 vector (Boehringer Mannheim) and transformed into *E. coli* DH5α cells (Clontech) by electroporation (Bio-Rad Gene Pulser). Transformants were plated on LB-ampicillin plates containing IPTG and X-Gal, allowing a selection of colonies carrying plasmids with inserts; 5200 colonies were selected (bank B) and transferred to Biodyne A nylon membranes (Pall). For preliminary experiments employing non-radioactive detection procedures a smaller bank (bank A) consisting of 3500 colonies was constructed in the same manner, using nuclear DNA from a cultivated beet and pBluescript SK(-) (Stratagene), digested with *Eco*RV, as the cloning vector.

Oligonucleotides representing microsatellite motifs were obtained from GenSet [$(CA)_8$, $(ATT)_7$, $(GTG)_5$] or kindly supplied by V. Laudet, Institut Pasteur Lille [$(AT)_{10}$, $(CT)_{10}$].

Radioactive labelling

The five oligonucleotides were 5′ end-labelled at 37 °C for 30 min, using *E. coli* polynucleotide kinase and the buffer supplied by Boehringer Mannheim. The reactions contained 50 pmoles of DNA and 5 µl of $\gamma[^{32}P]$ATP (5000 Ci/mmol) in a total volume of 20 µl. Non-incorporated nucleotides were removed by centrifugation through Quick Spin Sephadex G-25 columns (Boehringer Mannheim).

Non-radioactive labelling

$(CA)_8$ and $(CT)_{10}$ were 3′ end-labelled with terminal transferase, adding a polyadenosine tail, in which several bases are replaced by digoxigenin (DIG) dUMP. The protocol given by the supplier of the kit was followed. Free nucleotides were separated from the labelled fraction by ethanol precipitation. Labelling success was determined by a direct immunological detection of filter-bound probe, using the non-radioactive detection kit. Both kits were purchased from Boehringer Mannheim.

Radioactive and non-radioactive hybridizations were carried out at the same temperatures [$(AT)_{10}$ 35 °C, $(ATT)_7$ 37 °C, $(CA)_8$ 43 °C, $(GTG)_5$ 45 °C, $(CT)_{10}$ 55 °C].

Radioactive method

Filters were hybridised for 2–3 h as described by Ali et al. (1986) and washed in 5 × SSC, including a 5–10-min washing step at the hybridisation temperature. In cases where filters exhibited a high background activity, a further washing step in 2 × SSC/0.1%SDS was added. Filters were exposed over night at −70 °C with Kodak X-Omat AR films and intensifying screens (Dupont). Sequencing of both strands was performed using $\alpha[^{35}S]$ dATP (Amersham) and the T7 sequencing kit from Pharmacia. Reactions were separated on 6% denaturing polyacrylamide sequencing gels (LKB apparatus Pharmacia).

Non-radioactive method

Hybridisations were carried out for 1 h in a solution containing 5 × SSC, 1% dried milk, 0.1% lauryl-sarcosyl, and 0.02% sodium dodecylsulphate. Filters were washed in 5 × SSC for 2 × 30 min at room temperature and for 1 × 5 min at the hybridisation temperature. Non-radioactive detection was performed as described by Allefs et al. (1990), using CSPD instead of AMPPD as the chemiluminiscent dye (Tropix). Filters were exposed to Kodak X-Omat LS films for 1 h at 37 °C and 1–12 h at room temperature. Both strands of plasmids from positive clones were sequenced using the fluorescent dye-deoxy terminator kit and the 370 A DNA automated sequencer from Applied Biosystems.

DNA isolation, polymerase chain reaction, and sequence-tagged site polymorphisms detection

Total DNA was extracted from one individual per accession, according to a slightly modified procedure of Dellaporta et al. (1983).

Primers were selected with the aid of the programme *Primer Ver 0.5* (Daly et al. 1991) and ordered from GenSet. Amplifications were performed in a Perkin-Elmer thermal cycler 480; 25-µl samples overlaid with mineral oil (Sigma) contained about 25 ng of total DNA, 1.25 µmoles of each nucleotide (50 µM), 5 pmoles of each primer (0.2 µM), and one unit of *Taq* DNA polymerase, using the supplied buffer (Appligene). A 2-min denaturation at 97 °C, was followed by 25 cycles with 1 min at 94 °C, 40 s at the annealing temperature (between 48 and 60 °C), and 1 min at 70 °C. In the last cycle elongation was prolonged to 10 min. In two cases a 'touch down' procedure (Don et al. 1991) was included, beginning 5° above the annealing temperature. Then 10–15 µl of the reactions were separated on non-denaturing, 10% polyacrylamide/1 × TBE gels (Protean II, Bio-Rad) and stained with ethidium bromide.

Results

Frequencies of CA, AT, CT, ATT, and GTG repeats in the beet genomic libraries

The two partial libraries of nuclear DNA from beet were screened with two non-radioactively labelled, and five radioactively labelled, oligonucleotide probes, respectively; among them were repetitive sequences expected to be rare (CA) and abundant (AT, CT, ATT) in plants. The GTG motif is rare in plant DNA databases but was used here because it gave a smear when hybridised to digested total DNA (data not shown).

The highest number of signals was obtained with the CA probe. A total of 33 positive clones was found among 8700 colonies; four and seven clones gave clear, but considerably weaker, signals (Table 1). Considering only strong signals, the number of CT clones was considerably lower in bank A, whereas in bank B the frequencies of CA, CT, and ATT were nearly equal. Hybridisation with the AT probe gave no specific signals.

Table 1 Frequency of two- and three-base repeat regions detected by colony hybridisation in two different mini libraries containing genomic DNA inserts of *B. vulgaris*

A. Non-radioactive screening (3500 clones)

Probe	CA	CT
Number of signals	15(+4)[a]	3

B. Radioactive screening (5200 clones)

Probe	CA	CT	AT	ATT	GTG
Number of signals	7(+7)[a]	6	No specific signals	8	1

[a] 4 and 7 of the signals, respectively, were clearly weaker

Sequencing results

The average insert size of the clones was 510 bp (bank A) and 390 bp (bank B). All the positive clones contained simple sequences with different degrees of purity (Table 2). We found 31 perfect (27 di- and four tri-nucleotide) stretches, eight imperfect stretches with repetitive sequences interrupted by up to three nucleotides, and five compound stretches with at least two different adjacent repetitive sequences in the same clone. One of the compound repeats selected by the $(CT)_{10}$ probe contained a stretch of 14 AT motifs, confirming the inability of the $(AT)_{10}$ probe to detect positive clones under the hybridisation conditions employed. The very long $(CA)_{29}(TA)_{>20}$ stretch was difficult to sequence. Within the TA sequence, the readability declined after 20 motifs; its total number of motifs amounted to roughly 35. Most of the CT and ATT sequences contained more than ten and seven motifs, respectively, thus having a total length of more than 20 bp, which is known to be the lower limit of size polymorphism for human CA repeats (Weber 1990). Even if small sequences might correspond to small alleles of polymorphic microsatellites, a large repeat is thought to guarantee a higher probability of polymorphism. Apart from two stretches with 13 and 29 repetitions the CA sequences were below this limit. Four CA clones giving weak signals in non-radioactive hybridisation (Table 1, A) even had total lengths of less than ten nucleotides. Since the seven CA clones detected as strong signals by the radioactive hybridisation (Table 1, B) had total lengths of up to 16 bp, the seven CA clones exhibiting weak signals were not further characterised.

Moreover, a closer look at the sequences of the flanking regions revealed that 13 CA clones from bank A and all the seven CA clones from bank B represented variations of the same sequence. Sequence homologies were around 90%; differences consisted of single base changes and small deletions or insertions. The copy number of the CA repeat appeared in some cases to have been changed by point mutations, not by slippage, e.g. a CA changed to a TA. Sequence comparison of these 20 CA clones revealed about 90% homology with a satellite sequence previously described in *Beta* (Schmidt et al. 1991; Santoni and Bervillé 1992). The CA clones we isolated and the published sequences (pBV1 and BvBsat, respectively) are in fact variations of the same satellite DNA sequence. Comparing a 179-bp stretch from pBV1

Table 2 Di- and tri-nucleotide repeats in genomic DNA clones of *B. vulgaris* selected by colony hybridisation in libraries A and B

A. Clones selected by non-radioactive colony hybridisation in library A

Motif	Number of clones	Category[a]	Repeat sequence	
CA	4[b]	Perfect	$(CA)_n$	$n < 4$
	13	Perfect	$(CA)_n$	$3 < n < 9$
	1	Perfect	$(CA)_{13}$	
	1	Perfect compound	$(CA)_{29}(AT)_{>20}$	
CT	1	Perfect	$(CT)_6$	
	1	Perfect	$(CT)_{26}$	
	1	Imperfect	$(CT)_{17}$	

B. Clones selected by radioactive colony hybridisation in library B

Motif	Number of clones	Category[a]	Repeat sequence	
AT	None			
CA	7	Perfect	$(CA)_n$	$5 < n < 9$
CT	1	Imperfect	$(CT)_{10}$	
	1	Imperfect	$(CT)_{12}$	
	1	Imperfect	$(CT)_{22}$	
	1	5 imperfect	$(CT)_4N_{11}(TC)_7N_{15}(TC)_4N_{26}(TC)_8N_5(CT)_4$	
	1	Perfect compound	$(CT)_{15}CC(T)_{10}$	
	1	Perfect compound	$(CT)_{10}(AT)_{14}$	
ATT	3	Perfect	$(ATT)_n$	$n = 12, 30, 35$
	1	3 perfect	$(ATT)_8N_8(ATT)_{14}N_{25}(ATT)_{18}$	
	1	Imperfect	$(ATT)_4$	
	1	Imperfect	$(ATT)_{15}$	
	1	Imperfect compound	$(CTT)_{15}(ATT)_{37}$	
	1	Imperfect/perfect compound	$Imp[(ATT)_{35}A(TCA)_4(TTA)_5]N_{25}$ $Perf[(ATT)_6(GTT)_3]$	
GTG	1[c]	3 imperfect	$(GTG)_6N_{21}(GTG)_8N_{20}(GTG)_5$	

[a] Definition of categories is according to Weber (1990)

[b] One of the sequences was highly repetitive, showing 24 (TCAGT) motifs spread over 260 bp

[c] Clone contained in addition an imperfect $(TCAGT)_9$ repeat

to the corresponding regions in 13 of our clones, we found lengths between 165 and 184 bp. The included perfect $(CA)_n$ motif with n = 6 to 8 is in pBV1 an imperfect $(CA)_3$ repeat at position 64 to 71. We used the insert fragment of one CA clone as a probe against *Bam*HI-restricted total DNA. As expected the pattern observed was a ladder of fragments separated by approximately 330 bp. Thus a simple sequence appears to be part of a repetitive sequence of a higher repeat order.

Polymorphism of microsatellite loci in beets

One ATT and three CT clones having high motif numbers, which were sufficiently distant from the cloning site for primer construction, were selected for PCR primer design. Primers and repeat types are shown in Table 3. To facilitate the separation of fragments differing by 2 bp under the electrophoresis conditions used, we designed primers giving PCR products between 100 and 200 bp. Total DNAs of beets representing the three different genetic forms of the *B. vulgaris* species were used as templates. Amplification products were obtained for all primer pairs and consisted of one or two products for each individual. A single case of three fragments was found (Fig. 2, lane 9), most probably due to a triploid genome. Clear major PCR products were observed for the ATT locus and for small alleles of the CT repeats, with only a few weak secondary bands, whereas large alleles of the CT repeats were sometimes accompanied by multiple "shadow" bands.

Table 3 summarises the sizes and frequencies of alleles. Levels of polymorphism were conspicuously different for the four detected loci, ranging from complete monomorphism of Bvm1 to 13 alleles at the Bvm 2 and 3 loci.

The Bvm2 locus showed a very large allele size range. Some individuals (Fig. 2 A; lanes 2, 9 to 13) showed a PCR product of over 800bp (size verified on agarose

Table 3 *B. vulgaris* sp. microsatellite loci and allele frequencies in wild, cultivated, and weed beets

Locus	Primers 5′..... 3′	Repeat type[a]	Size classes [bp]	Allele frequencies			
				Wild beets $n=17$	Weed beets $n=15$	Cultivated beets $n=32$	Total 64 plants
Bvm1	CAGATGATTCACGAAGCAGG CCTAAGGACAACATAAGTTCTG	Imperfect $(CT)_{17}$	151[b]	1.00	1.00	1.00	1.00
Bvm2	AAGTAACCCAGGTAAAAGAC CAACATTCCAAGTAATCAACAT	Imperfect$(ATT)_{15}$	99	0.12	0.19	0.08	0.12
			102	0.03			<0.01
			111	0.24			0.06
			114	0.09		0.09	0.07
			123	0.06	0.13	0.27	0.18
			132	0.03			<0.01
			135[b]	0.23			0.06
			138	0.03			<0.01
			147	0.06			0.02
			150		0.16	0.03	0.05
			165			0.03	0.02
			171		0.03	0.02	0.02
			>800	0.12	0.49	0.48	0.39
Bvm3	ACCAAATGACTTCCCTCTTCTT ATGGTGGTCAACAATGGGAT	Compound $(CT)_{15}CC(T)_{10}$	96	0.03	0.07		0.02
			100	0.28	0.17	0.02	0.12
			108	0.12		0.03	0.05
			110	0.19	0.53	0.87	0.62
			112[b]	0.16		0.05	0.06
			114		0.13		0.03
			118	0.03	0.07		0.04
			120	0.13	0.03	0.03	0.04
			122	0.06			0.02
Bvm4	CATCCTACTTTCTCCGTT CAAAGTGTCGACATAGATT	Perfect$(CT)_{26}$	55	0.03	0.07	0.11	0.08
			63	0.22			0.06
			65	0.10	0.07	0.03	0.06
			69	0.03	0.03		0.015
			79		0.07		0.015
			81	0.03	0.10		0.03
			83	0.06	0.23	0.59	0.37
			85	0.06			0.015
			89	0.06			0.015
			93	0.03	0.07		0.02
			95	0.25	0.16	0.05	0.13
			97[b]	0.13	0.13	0.19	0.16
			103		0.07	0.03	0.03

[a] Repeat type present in the sequenced clone
[b] Expected size of the clone

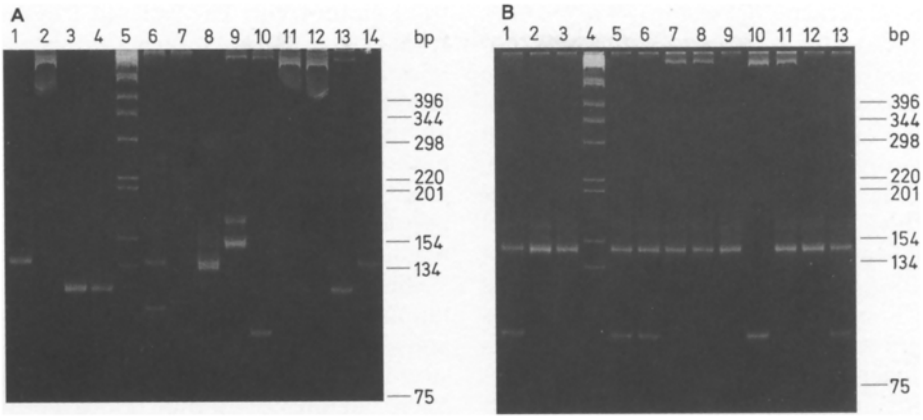


Fig. 2A Microsatellite genotypes of beet accessions at the Bvm2 locus. *Lane 1*, amplification product of the cloned fragment; *lane 2*, *B. vulgaris* Florimond Desprez maintainer of sterility TO5; *lane 3*, *B.v.* KWS maintainer of sterility TO7B; *lane 4*, *B.v.* INRA maintainer of sterility LO313; *lane 5*, molecular-weight marker *X* (Boehringer Mannheim). *Lane 6*, *B. maritima* Aut; *lane 7*, *B.m.* Sla; *lane 8*, *B.m.* 4113; *lane 9*, weed beet Bri3; *lane 10*, w.b. Cro2; *lane 11*, w.b. Gzt 1; *lane 12*, w.b. Gzt 3; *lane 13*, w.b. Gra 29; *lane 14*, *B.m.* Lec. For abbreviations of population names see Fig. 1. **B** Segregation of microsatellite alleles at the Bvm2 locus in progenies produced by reciprocal crosses of the P1 and P2 *B. maritima* plants. Results for 10 of 20 plants are shown: *lane 1*, P1; *lanes 2, 3, 5, 6, and 7* correspond to five progenies harvested on P1; *lane 8*, P2; *lanes 9 to 13* correspond to five progenies harvested on P2; *lane 4*, molecular-weight marker *X* (Boehringer Mannheim)

gels), which would correspond to approximately 250 repetitions of the trinucleotide motif. This large fragments (1f) was rare among wild beets, but represented nearly 50% of the alleles of cultivated and weed beets, which are supposed to originate from crosses between wild and cultivated beets (Boudry et al. 1993). Extensive secondary structures were not responsible for the slow migration, as confirmed by a denaturing gel electrophoresis.

Mendelian inheritance of alleles and genotype distribution

In order to verify Mendelian inheritance of the alleles at the three polymorphic loci, PCR amplification on two parents and 20 F_1 progenies produced by a controlled reciprocal cross were analysed. Amplification on parental DNAs gave two fragments each for all loci, suggesting four possible combinations of equal proportions for the progeny. All F_1 plants showed only parental fragments in proportions not significantly different from the expected ones (Table 4). Figure 2B shows the results from parental and 10 F_1 plants for the Bvm2 locus. Comparing the 17 wild to the 32 cultivated beets for the three polymorphic loci, genetic diversity was considerably higher for wild forms: ten alleles versus seven (10/7) for Bvm2, 8/5 for Bvm3, 11 vs 6 for Bvm4. Hence some alleles found in wild beets appear to be absent from cultivated plants. Genetic variability of the 15 weed beets was intermediate. For Bvm2, results for weed and cultivated beets were nearly identical, whereas for Bvm4 the genetic variability of the weed beets was comparable to those of wild ones. Nevertheless one allele of each locus was associated with cultivated and weed beets, showing frequencies of 0.59 and 0.23 (Bvm4), 0.48 and 0.49 (Bvm2), up to 0.87 and 0.53 (Bvm3). In Bvm2, this allele corresponded to the large fragment.

Combining data of the three polymorphic loci, it was possible to distinguish each of the 17 wild and 15 weed beet individuals, and 21 of the 32 cultivated plants. Data from single loci identified between 11 (Bvm3) and 14 (Bvm2) genotypes among the wild plants, but only five (Bvm3) to nine (Bvm2) genotypes among the cultivars. Concerning the weed beets, six (Bvm2) to 12 (Bvm4) genotypes could be identified.

Discussion

Abundance of simple sequence repeats in the *Beta* genome

Oligonucleotide fingerprinting of long restriction fragments had already shown that the beet genome contains

Table 4 Likelihood ratio tests (G-test) for goodness of fit for Mendelian inheritance of alleles at microsatellite loci of *B. vulgaris*. All the segregations were expected in the ratio 1:1:1:1 (3 degrees of freedom)

Locus	Parental genotypes	Observed segregation				G	*P*
Bvm2	99/147 × 147/1f	99/147	99/1f	147/147	147/1f		
	$n = 20$	4	3	6	7	2.048	0.56
Bvm3	96/110 × 112/120	96/112	96/120	110/112	110/120		
	$n = 19$	3	8	2	6	4.927	0.18
Bvm4	69/89 × 93/95	69/93	69/95	89/93	89/95		
	$n = 19$	4	6	2	7	3.397	0.33

simple repetitive sequences (Weising et al. 1991). Here, we converted some repetitive sequences into informative STSs for use as genetic markers. As the plasmid libraries represented only very small parts of the *Beta* genome, absolute frequencies could not be deduced. The proportions of the different motifs, however, were similar to those published for plants (Morgante and Olivieri 1993), except in the case of the CA repeats. CT and ATT repeats were the most abundant. GTG blocks were around six times less frequent than ATT blocks in our mini library; this ratio is in agreement with the value reported for plants (Morgante and Olivieri 1993).

The relative frequency of CA repeats was high for a plant species: $(CA)_n$ blocks were the commonest in the beet genomic libraries, the reverse of what is observed e.g. in rice (Wu and Tanksley 1993) and *Arabidopsis* (Bell and Ecker 1994). Nevertheless, apart from two long sequences (13 and 29 motifs), 24 CA stretches were below nine motifs. In addition, sequence comparisons showed that 20 of these clones were variations of a previously described satellite DNA (Schmidt et al. 1991; Santoni and Bervillé 1992). To be useful as PCR-detected genetic markers, microsatellites need to be embedded in unique sequences. Thus, the satellite-embedded CA repeats do not appear to be suitable for generating STSs. Simple repeats within classical satellites increase the frequency of positive clones in genomic banks when appropriate restriction-enzyme recognition sites are present. This leads to overestimates of microsatellite sequences in a given species. This phenomenon has been reported for *Pinus radiata*, where 25% of CA and/or CT clones were associated with other highly repetitive DNA, as revealed by hybridisation with total genomic DNA (Smith and Devey 1994). Therefore such a secondary hybridisation should routinely follow for plant genomes.

Hybridisation with the probe expected to be the most useful, $(AT)_{10}$, gave no positive signals, even in the case where a long AT stretch was part of a compound microsatellite. Similar results have been reported by Lagercrantz et al. (1993) and could be explained by the probe's self-complementarity. In fact, all reports of finding AT repeats in plants are based on screening DNA databases and not on colony hybridisation (e.g. Saghai Maroof et al. 1994). Nevertheless, in animals several clones have been isolated with the $(AT)_{10}$ probe (J.T. Epplen, personal communication).

Length polymorphism

The degree of polymorphism is one of the important features of a microsatellite, and is thought to depend on motif copy number; for CA repeats the rate of strand slippage was found to increase markedly when motif numbers exceed ten (Weber 1990).

Apart from the CA stretches, most of the microsatellites studied here were above this limit. Nevertheless PCR analysis of four Bvm loci showed a large range of polymorphism. Bvm1, an imperfect $(CT)_{17}$, appeared to be monomorphic. The lack of variability in such a large repeat suggests a location within sequences submitted to selection, such as coding regions. A comparison of the flanking sequences with gene data banks did not reveal any homologies. Dinucleotide microsatellites in plants have not yet been found within coding regions (Wang et al. 1994). Alternatively, the presence of variant motifs may depend on a low slippage rate at Bvm1, which would also be reflected by a low level of polymorphism.

The highly polymorphic loci possessed alleles containing numbers of repeats in the range of 5–29 (Bvm4) and 3–27 or possibly 250 (Bvm2). Therefore, short stretches can correspond to small alleles of polymorphic microsatellites and should not be ignored when developing PCR primers.

The occurrence of fragments much larger in comparison to other alleles of the same site was first reported for humans. In this case the phenomenon was associated with the fragile X syndrome; in carriers and affected individuals the copy number of a trinucleotide repeat was increased from 6 to 60 to more than 200 (for a review see Caskey et al. 1992).

The large allele range of the Bvm2 locus is to our knowledge the first case reported for plants. Its function and character have yet to be elucidated. We did not observe any influence of large fragments on the phenotype; the plants produced normal progenies with the large fragment segregating according to Mendelian rules. Moreover, this allele was the most frequent one in both cultivated beet accessions and weed beets. DNA sequence analysis of these large fragments will show whether they are in fact generated by a multiplication of the trinucleotide repeat, or whether the insertion of a different DNA sequence is responsible.

The numbers of alleles found at the three polymorphic loci in a sample of 64 wild, cultivated, and weed beets were nine and 13 observed twice, giving a mean value of 11.6. It is interesting to compare these numbers to those observed with four single-copy RFLP markers in different individuals of the same forms of beet (Boudry et al. 1995). The RFLP markers were found to be "highly" variable with up to 13 alleles (mean value 11) in the sample of 236 plants. Therefore, the advantage of the microsatellites defined in this study probably lies more in the technical advantages of PCR for analysing large numbers of individuals, than in their increased levels of polymorphism over RFLP markers.

Nevertheless the three microsatellites were capable of distinguishing 17/17 wild beets, 15/15 weed beets, and 21/32 (mostly homozygous) cultivated beets. Thus, their informativeness is sufficiently high to encourage further development: namely, an extension of the analysis to a larger number of markers with emphasis on trinucleotide repeats, which seem to be easily scored, as well as the localisation of these markers on existing linkage maps of *Beta* (Barzen et al. 1992; Pillen et al. 1992; Barzen et al. 1995). The development of sequence-tagged sites will allow the replacement of costly and time-consuming RFLP procedures for studies of genetic diversity in

cultivars, or of population genetic structures within and gene flow among the different forms of beets.

Acknowledgements We thank Florimond-Desprez, Cappelle-en-Pévèle (France), Kleinwanzlebener Saatzucht AG, Einbeck (FRG), and the INRA, Dijon (France) for providing the cultivated plant material, and Pierre Boudry and Henk van Dijk for providing the weed and wild beets. We are indebted to Vincent Laudet for synthesising oligonucleotides. We are very grateful to Jörg T. Epplen, John F. Dallas, Myriam Valero, and Philippe Vernet for thoughtful reading of the manuscript. This work was supported by the MRT grant 90.G.0519 and by the 'Contrat de plan Etat/Région Nord-Pas de Calais 1989–1993: Maîtrise de la qualité des produits agroalimentaires' (France).

References

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

Ali S, Müller CR, Epplen JT (1986) DNA finger printing by oligonucleotide probes specific for simple repeats. Hum Genet 74:239–243

Allefs JJHM, Salentijn EMJ, Krenz FA, Rouwendal GJA (1990) Optimization of non-radioactive Southern-blot hybridization, single-copy detection and reuse of blots. Nucleic Acids Res 18:3099–3100

Barzen E, Mechelke W, Ritter E, Seitzer JF, Salamini F (1992) RFLP markers for sugar beet breeding – chromosomal linkage maps and location of major genes for *Rhizomania* resistance, monogermy and hypocotyl color. The Plant Jour 2:601–611

Barzen E, Mechelke W, Ritter E, Schultekappert E, Salamini F (1995) An extended map of the sugar beet genome containing RFLP and RAPD loci. Theor Appl Genet 90:189–193

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

Bell CJ, Ecker JR (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144

Boudry P, Mörchen M, Saumitou-Laprade P, Vernet P, Van Dijk H (1993) The origin and evolution of weed beets: consequences for the breeding and release of herbicide-resistant transgenic sugar beets. Theor Appl Genet 87:471–478

Boudry P, Broomberg K, Saumitou-Laprade P, Mörchen M, Cuguen J, van Dijk H (1995) Gene escape in transgenic sugar beet: what can be learned from molecular studies of weed beet populations? In: Jones DD (ed) Proc 3rd Int Symp on the Biosafety Results of Field Tests of Genetically Modified Plants and Microorganisms. Montery, California (in press)

Caskey C, Pizzuti A, Fu Y-H, Fenwick RG, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–789

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

Daly MJ, Lincoln SE, Lander ES (1991) PRIMER, Ver. 0.5. Whitehead Institute, MIT Center for Genome Research, Cambridge, Massachusetts

Dellaporta SL, Wood VP, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) "Touchdown" PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res 19:4008

Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable "minisatellite" regions in human DNA. Nature 314:67–73

Lagercrantz U, Ellegren H, Andersson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucleic Acids Res 21:1111–1115

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

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

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E, White R (1987) Variable number of tandem repeat (VNTR) markers for human genome mapping. Science 235:1616–1622

Olson M, Hood L, Cantor D, Botstein D (1989) A common language for physical mapping of the human genome. Science 245:1434–1435

Pillen K, Steinrücken G, Wricke G, Herrmann RG, Jung C (1992) A linkage map of sugar beet (*Beta vulgaris*). Theor Appl Genet 84:129–135

Rohlf JF (1992) BIOM, a package of statistical programs to accompany the text Biometry. Applied Biostatistics Inc., Setauket, New York

Rongwen J, Akkaya MS, Bhagwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43–48

Saghai Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW (1994) Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466–5470

Santoni S, Bervillé A (1992) Two different satellite DNAs in *Beta vulgaris* L.: evolution, quantification and distribution in the genus. Theor Appl Genet 84:1009–1016

Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Schmidt T, Jung C, Metzlaff M (1991) Distribution and evolution of two satellite DNAs in the genus *Beta*. Theor Appl Genet 82:793–799

Schmidt T, Blobenz K, Metzlaff M, Kaemmer D, Weising K, Kahl G (1993) DNA fingerprinting in sugar beet (*Beta vulgaris*) – identification of double-haploid breeding lines. Theor Appl Genet 85:653–657

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

Smith DN, Devey ME (1994) Occurrence and inheritance of microsatellites in *Pinus radiata*. Genome 37:977–983

Sokal RR, Rohlf J (1981) Biometry (2nd edn.). Freeman and Co, San Francisco

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

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

Thomas MR, Scott NS (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor Appl Genet 86:985–990

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) Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

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, Ramser J, Kaemmer D, Kahl G, Epplen JT (1991) Oligonucleotide fingerprinting in plants and fungi. In: Burke T, Dolf G, Jeffreys AJ, Wolff R (eds) DNA fingerprinting: approaches and applications. Birkhäuser Verlag, Basel, pp 312–328

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 KS, Tanksley SD (1993) Abundance, polymorphism and genetic mapping of microsatellites in rice. Mol Gen Genet 241:225–235

Yang GP, Saghai Maroof GP, Xu CG, Zhang Q, Biyashev RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol Gen Genet 245:187–194