Theor Appl Genet (1998) 97: 591–604

K. Loridon · B. Cournoyer · C. Goubely
A. Depeiges · G. Picard

Length polymorphism and allele structure of trinucleotide microsatellites in natural accessions of *Arabidopsis thaliana*

Received: 4 November 1997 / Accepted: 3 March 1998

Abstract The objective of this work was to assess the degree of trinucleotide microsatellite length polymorphism in the selfing species *Arabidopsis thaliana*. PCR amplifications of 12 microsatellite loci among 49 natural populations revealed between one to eight length variants (alleles) for each locus. The average number of alleles per locus was four and the average genetic diversity index was 0.43. Divergence between length variants was investigated at the nucleotide level. Several observations emerge from the sequence data: (1) for most loci, length polymorphism results only from variations in the number of trinucleotide repeats; (2) for a few others, some variability was noted in the flanking sequences; (3) for compound and interrupted loci containing two arrays of trinucleotide repeats, length variations preferentially affect the longest one. Five of the *Arabidopsis thaliana* accessions were clearly composed of two sublines. In 2 other accessions, some heterozygous individual plants, probably resulting from recent outcrosses, were found. A phylogenetic tree constructed on the basis of trinucleotide microsatellite allelic diversity shows that genetic relationships among the accessions are not correlated with their geographic origin.

Key words Microsatellite · Trinucleotide repeats · Simple sequence length polymorphism · *Arabidopsis thaliana* · Genetic diversity

Communicated by J. Beckmann

K. Loridon · C. Goubely · A. Depeiges (✉) · G. Picard
Laboratoire Biomove, UMR-CNRS 6547, Université Blaise Pascal Clermont-Ferrand II, 24, avenue des Landais,
F-63177 Aubière Cedex, France
E-mail: depeiges@cicsun.univ-bpclermont.fr

B. Cournoyer
Laboratoire d'Ecologie Microbienne du Sol, UMR-CNRS 5557, Université Claude Bernard Lyon I, F-69622 Villeurbanne Cedex, France

Introduction

Microsatellites are composed of tandemly repeated short (1–5 bp) DNA sequences (Litt and Luty 1989; Tautz 1989). They are abundant and widely distributed in eukaryotic genomes (Weber 1990; Stallings et al. 1991; Lagercrantz et al. 1993). Microsatellites have been identified in a wide variety of species, including plants (Morgante and Olivieri 1993; Wang et al. 1994), mammals (Moore et al. 1991; Beckmann and Weber 1992), birds (Moran 1993; Burt et al. 1995), fish (Estoup et al. 1993; Rico et al. 1996) and invertebrates (Goldstein and Clark 1995; Viard et al. 1997).

Microsatellite loci can be detected by the polymerase chain reaction (PCR) technique using primers complementary to the non-repetitive sequences flanking the repeat region (Beckmann and Soller 1990). In most cases, they show high levels of allelic diversity due to variations in the number of tandem repeats between different genotypes, possibly caused by slippage events during replication (Levinson and Gutman 1987; Schlötterer and Tautz 1992; Strand et al. 1993). Their great abundance, dispersion throughout the entire genome, high levels of polymorphism and rapid detection make microsatellites valuable genetic markers for genetic mapping (Wu and Tanksley 1993; Dib et al. 1996; Dietrich et al. 1996), multilocus DNA fingerprinting (Zietkiewicz et al. 1994; Sharma et al. 1995) and population genetic studies (Bowcock et al. 1994; Jarne and Lagoda 1996).

In the plant kingdom, microsatellites have been reported mostly in crop species, e.g. soybean (Rongwen et al. 1995), corn (Taramino and Tingey 1996), rice (Panaud et al. 1995), barley (Saghai Maroof et al. 1994) and tomato (Broun and Tanksley 1996). Though *Arabidopsis thaliana* has become a model organism for plant molecular biology (Meyerowitz 1989), little data are available about microsatellites in this self-fertilizing crucifer. Degrees of polymorphism of microsatellites

containing mono- or dinucleotide repeats have been assessed in 6 common laboratory accessions (Bell and Ecker 1994), in 12 Japanese natural populations (Todokoro et al. 1995) and in 42 ecotypes (Innan et al. 1997). In a previous work (Depeiges et al. 1995), we showed that trinucleotide microsatellites are very abundant in *Arabidopsis thaliana*, with the most represented repeated motifs being AAG/CTT, ATG/CAT and GTG/CAC, in a decreasing order of frequencies. However, no data regarding their length polymorphism have been reported yet.

In this paper, we report the degree of polymorphism of 12 trinucleotide microsatellites among 49 *Arabidopsis thaliana* ecotypes made available from the Nottingham Stock Centre and examine some of the allelic variations at the nucleotide level. Trinucleotide microsatellites were also used to determine genetic relationships between natural populations of *Arabidopsis thaliana* collected in various countries.

Materials and methods

Plant materials and DNA extraction

Seeds of 49 *A. thaliana* ecotypes coming from various areas were obtained from the Nottingham Arabidopsis Stock Centre (NASC). The common name and geographical origin of these ecotypes are given in Table 1.

Seeds were sown on sterile ATG germination medium according to Valvekens et al. (1988) and grown at 23°C in a 16-h light/8-h dark cycle. Three-week-old plants (of each ecotype) were removed from the germination medium, pooled (30–50 individuals) and ground in liquid nitrogen. DNA was extracted according to the CTAB method (Doyle and Doyle 1987). Alternatively, 2-week-old individual plants were ground separately at room temperature in a 10 m*M*, TRIS-HCl, 1 m*M* EDTA pH 8.0 buffer. DNA was purified by adding Gene Releaser (BioVentures) following the supplier's instructions.

Polymerase chain reaction and polymorphism detection

Table 2a shows the 12 microsatellite loci examined in this study, with their primer sequences and pattern of repeats. These loci were isolated from cDNA and genomic libraries or came from a search in the GenBank database (Depeiges et al. 1995). PCR primer sets were synthesized by Eurogentec.

PCR reactions were run with 50–100 ng genomic DNA and 0.25 units of Vent DNA polymerase (New England Biolabs) in a total volume of 25 µl, with 1 × DNA polymerase buffer, 1 µ*M* of each primer and 0.2 m*M* of each dNTP including 10 µ*M* of α-[^{33}P]-dATP (Amersham). After an initial denaturation step at 94°C for 2 min, the reaction went through 25 cycles of 30 s/94°C, 30 s/primer-specific annealing temperature and 30 s/72°C, followed by a final extension of 5 min at 72°C in a Techne PHC-3 thermocycler. The optimized annealing temperatures ranged between 55°C and 65°C. The PCR products were electrophoresed on 6% denaturing polyacrylamide gels, a technique that enables the separation of fragments differing by 1 nucleotide. To assess the size of PCR fragments, we loaded a M13mp18 DNA sequencing ladder near the amplification products. Gels were dried and exposed 24 h to X-ray films.

Table 1 Name and geographical origin of 49 *Arabidopsis thaliana* ecotypes according to the NASC Seed List (1993)

Ecotype	Geographical origin
C24	Derivative of ecotype Columbia
Ler-0	Koornneef group (The Netherlands)
Col-4	Dean Group (UK)
Be-0	Bensheim/Bergstr. (Germany)
RLD1	Koornneef group (The Netherlands)
S96	Koornneef group (The Netherlands)
Condara	Khurmatov (Tadjikistan)
Enkheim-T	Usmanov laboratory (Tadjikistan)
Rubezhnoe-1	Rubezhnoe (Ukraine)
Shahdara	Khurmatov (Tadjikistan)
Aa-0	Aua/Rhön (Germany)
Ag-0	Argentat (France)
An-1	Antwerpen (Belgium)
Ang-0	Angleur (Belgium)
Bay-0	Bayreuth (Germany)
Bl-1	Bologne (Italy)
Bla-12	Blanes (Spain)
Bs-1	Basel (Switzerland)
Bsch-0	Buchschlag/FFM (Germany)
Bur-0	Burren (Eire)
Ca-0	Camberg/Taunus (Germany)
Cal-0	Calver (UK)
Cen-0	Caen (France)
Chi-1	Chisdra (former URSS)
Co-1	Coimbra (Portugal)
Ct-1	Catania (Italy)
Cvi-0	Cape Verdi Islands
Di-0	Dijon (France)
Dr-0	Dresden (former GDR)
Dra-0	Drahonin (former Czechoslovakia)
Edi-0	Edinburgh (UK)
Eil-0	Eilenburg (former GDR)
Es-0	Espoo (Finland)
Ge-1	Genève (Switzerland)
Gr-2	Graz (Austria)
Gy-0	La Minière (France)
Je-0	Jena (former GDR)
Jl-1	Vranov u Brno (former Czechoslovakia)
Kas-1	Kashmir (India)
Kä-0	Kärnten (Austria)
Kn-0	Kaunas (Sweden)
Lc-0	Loch Ness (UK)
Lip-0	Lipowiec/Chrzanow (Poland)
Ll-0	Llagostera (Spain)
Lm-2	Le Mans (France)
Lz-0	Lezoux (France)
Mh-0	Mühlen (Poland)
Mt-0	Martuba/Cyrenaika (Lybia)
Nc-1	Ville-en-Vermois (France)

Cloning and sequencing of PCR products

PCR-amplified products (400 ng initial genomic DNA – 40 cycles) were ligated to pGEM-T vector (Promega) following the supplier's protocol or to pBluescript digested with *Sma*I (New England Biolabs). The nucleotide sequences of the inserts were determined using the dideoxychain termination method (Sanger et al. 1977). Radioactive sequencing with α-[^{33}P]-dATP (3.7×10^8 Bq/ml, 9.2×10^{13} Bq/mmol, Amersham) was performed using the T7 sequencing kit (Pharmacia). Sequence reactions were electrophoresed on 8% denaturing polyacrylamide gels.

Table 2 Characteristics of the 12 microsatellite loci used in this study. **a** Pattern of repeats in the reference ecotype, primer sequences for PCR amplification and gene or cDNA in which they were found, **b** physical and genetical mapping, chromosomal position and flanking markers

a

Locus[a]	Pattern of repeats	Ecotype	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	Gene/cDNA
ATTIF4A2	$(GAA)_9$	Unknown	TGTGGTTGTTGAGGAGC	GATGTATCGATGATTCAG	Translation initiation factor 4A-2 (3′UTR cDNA)
ATTS0191	$(TGA)_7$	Col-4	CTGATGTTGATGGAGATGGTC	GCATATCTCCACCAATCATGC	Clone similar to calmodulin-1 (3′UTR cDNA)
ATHGAPAB	$(CTT)_7CT(CTT)_2$	C24	GGTCACCATGGCTTCGG	AGCAGAGGAGCTTCGC	Glyceraldehyde-3-phosphate dehydrogenase A subunit (GapA) gene (intron 1)
ATPOSF21	$(CAA)_8CAGCAACACCAA(CAG)_4$	Ler-0	GCAAACAATCTTAGCTGC	GCTGTTGAAGCCGCTGC	pos F21 gene (exon 4)
ATCSR12	$(CAA)_6$	Col-4	AACCCTTAAACTTCAAGTCTC	GGAGGGAGAATCTGGAG	Acetolactate synthase gene (exon transit peptide)
ATTOP1	$(ATG)_7$	Col-4	TGTTGATGGAGATGGTC	ATCTCCACCAATCATGC	Topoisomerase 1 (cds cDNA)
ATAAG1	(CTT)7(CCT)2	C24	GAAGGTCTTAAATCTGGC	GAAGCGTTAGGCGGATGC	Unknown
ATTS0392	$(GAA)_7$	Col-4	GGAGTTAGACACGGATC	GATCGCAGCTTGATAAGC	Unknown (EST)
TO4351	$(CTC)_6$	Col-4	GAGTTCGTCTCTACCTCTG	GCGAATAGGATCTGCATCTC	Unknown (EST)
ATHHANKA	$(AGC)AG(AGC)_5$	Ler-0	ATGAGATCCGGCCGAGCG	GGCCTGACTTCCCGTTCG	Ankyrin repeat-containing protein (5'UTR cDNA)
ATTS0966	$(CTT)_2(CAT)_4(CTT)_3$	Col-4	CTGGAATCTGAATCGACG	CATCAGATTCAATCGCTG	Unknown (EST)
ATTS0262	$(GAA)_7CAACTTG(T)_6G(T)_4$	Col-4	GAATAGTATCGATGTTGC	GATACCTACACAAATCCTC	Clone similar to tomato histone H2A (3′UTR cDNA)

b

Locus[a]	CIC YAC	Chromosome	Flanking markers[c] (position on the RI map)	Chromosome (position on the RI map)
ATTIF4A2	IIIC9	Unknown	Unknown	nd[d]
ATTS0191	VE10	Unknown	Unknown	5 (61.1 cM)
ATHGAPAB[b]	VH4, VIB11, XID3	3	mi178 (40.1 cM)	3 (40.1 cM)
ATPOSF21	XIE1	2	ve096 (61.2 cM)	nd
ATCSR12	XIG6	3	I18 (61.8 cM)	nd
ATTOP1	VIC5, VIIIH1, XH7	5	m233 (85.7 cM)	nd
ATAAG1	IIIB10, IIID2, VIIA9, XID12, XIG6	3	ASN1 (57.8 cM), I18 (61.8 cM)	nd
ATTS0392[b]	IIIH12, VIIH4, XIIF3, XIIA10, XIIA11, XIIB11, XIIE10	1	UFO (47.5 cM)	1 (44.6 cM)
TO4351	VIA8, XF2	Unknown	Unknown	nd
ATHHANKA[b]	IIIB1, VB3, VC3, IXG4	5	m555 (100.4 cM)	5 (100.8 cM)
ATTS966	IID1, IVF1	1	pPC8.94 (114.4 cM)	nd
ATTS0262[b]	IIIA4, XIIF8	5	mi125 (48.6 cM)	5 (48.3 cM)

[a] All microsatellite loci were mapped on the CIC YAC library (Creusot et al. 1995). The genetic position of 5 loci was determined using 100 recombinant inbred (RI) lines derived from a cross between Col-4 and Ler-0 (Lister and Dean 1993)

[b] For loci ATHGAPAB, ATTS0392, ATHHANKA and ATTS0262, both physical and genetical mapping could be done and gave similar results

[c] Positions of markers on the RI map are given after Lister and Dean (1997)

[d] nd = not determined (lack of polymorphism between Col-4 and Ler-0)

Southern analysis

Five micrograms of genomic DNA were digested 6 h with either *Dra*I, *Eco*RV or *Hin*dIII (these enzymes, supplied by New England Biolabs, are insensitive to methylation) and hybridized, after alkali Southern blotting, with PCR fragments corresponding to 12 mapped restriction fragment length polymorphism (RFLP) markers (Liu et al. 1996) kindly provided by Dr. R. F. Whittier. Each of these markers shows polymorphism between the Landsberg *erecta* (Ler-0) and Columbia (Col-4) accessions. Each PCR product was labeled with α-[^{32}P]-dCTP (3.7×10^8 Bq/ml, 1.1×10^{14} Bq/mmol, Amersham) using the Megaprime Kit (Pharmacia) and hybridized to one of the membranes described above.

Genetic diversity index and phylogenetic analyses

Allele frequencies were determined for each microsatellite marker without taking heterogenous patterns (Table 3) into account. As an indicator of polymorphism in a group of homozygous individuals from a self-fertilizing species, a genetic diversity index (GDI) was defined as $GDI = 1 - \Sigma (Pi)^2$ where Pi is the frequency of the i^{th} allele in the population studied for each microsatellite locus (Saghai Maroof et al. 1994).

The microsatellite distance program "microsat" (E. Minch, Department of Genetics, Stanford University, Calif.) was used to estimate phylogenetic distances between microsatellites of *A. thaliana* ecotypes. The following distance methods were used and compared: (1) the delta mu genetic distance (Goldstein et al. 1995a) and (2) the average square distance (Goldstein et al. 1995b). The neighbor-joining (NJ) tree-building method of the PHYLIP package (J. Felsenstein, 1991–1993, PHYLIP package, Version 3.5c, University of Washington, Wash.) was used to infer phylogenetic trees which were visualized by the njplot program (M. Gouy, UMR CNRS 5558, Université Claude-Bernard-Lyon 1, France). The "consense" program of the PHYLIP package was used to compare tree topologies.

Results

Polymorphism of trinucleotide microsatellites in *Arabidopsis thaliana*

A set of 49 *A. thaliana* ecotypes (Table 1) was used to measure the level of polymorphism of 12 trinucleotide microsatellite loci (Table 2a) which were mapped on the CIC YAC library and on the RI lines (Table 2b). PCR amplification was carried out on genomic DNA extracted from pooled plants. Of the 588 locus/accession combinations, 524 gave a single fragment, 44 produced 2–3 fragments and 20 did not give any product. PCR product sizes, assessed by comparison with a sequencing ladder, are given in Table 3.

The 12 trinucleotide microsatellites showed various degrees of polymorphism, as illustrated in Fig. 1. The number of alleles varied from 1 (ATTS966) to 8 (ATTS0191, ATTS0392 and ATTS0262), with an average of 4 alleles per locus. Except for a few alleles at loci ATHHANKA and ATTS0262 (Fig. 2), alleles at a same locus differed in length by multiples of 3 base pairs, which can be explained by variations in the number of trinucleotide repeats. The maximum difference in allele size detected at 1 locus was 27 bp (ATTS0191, ATTS0392). For each microsatellite, the allele frequencies and genetic diversity index were calculated (Fig. 1). It appears that the allele frequencies were not evenly distributed at a same locus (1 or 2 major alleles). Genetic diversity index values ranged from 0.00 (ATTS966) to 0.80 (ATTS0392, ATTS0262), with an average of 0.43.

Sequence analyses

To test whether alleles at a microsatellite locus diverge by the number of repeats or by a modification of the flanking regions, we amplified, cloned and sequenced several alleles at 6 microsatellite loci (3–10 alleles per locus, 1–4 clones per allele). The sequence comparisons are reported in Fig. 3.

Two microsatellite loci (ATTS0191, ATTS0392) were found to consist of a stretch of perfect tandem repetitions. The alleles just differed by the number of trinucleotide repeats and by a few nucleotide substitutions in adjacent sequences. In 3 other loci, the microsatellite consisted of two stretches of perfect trinucleotide repeats which differed in length. In ATHGAPAB, the two stretches of (CTT) were interrupted by a CT motif which probably resulted from a T deletion in an original perfect block of repeats. In ATAAG1, the two arrays displayed different trinucleotide motifs, (CTT) and (CCT). In ATPOSF21, the two stretches also carried different motifs, (CAA) and (CAG), and were separated by four trinucleotide-related repeats. In each case, the longer stretch appears to be more variable than the shorter. It is also noteworthy that a base substitution inside the block of repeats allows 2 alleles of the same size in the Ag-0 accession to be distinguished.

Locus ATTS0262 was more complex, with the presence of one stretch of $(GAA)_n$ and two blocks of $(T)_n$ repeats which varied simultaneously in sequenced alleles. Some deletions and additions of one nucleotide as well as base substitutions have occurred, preferentially between the repeated motifs. This explains why alleles did not always differ in size by multiples of 3 bp at ATTS0262 (Table 3 and Fig. 2).

For all of the microsatellite loci, the sequence data were in good agreement with the assessment of the size of PCR products (Table 3). For Bur-0 at ATPOSF21, 2 alleles were detected, as expected. All sequenced clones from a PCR product were identical, except in two cases: (1) for Ag-0 at ATPOSF21, a base substitution appeared in the core motif; (2) for Lc-0 at ATTS0262, 3 alleles were found, which differred by the number of T repeats in a poly(T) block. These size differences may be due either to some polymerase stuttering during PCR or to natural variations not detected in the PCR product because of its complex pattern.

Table 3 Allele sizes (in bp) of PCR products from 49 *Arabidopsis thaliana* accessions for 12 trinucleotide microsatellites (na = no amplification)

Locus / Accession	ATTIF4A2	ATTSO191	ATHGAPAB	ATPOSF21	ATCSR12	ATTOP1	ATAAG1	ATTSO392	TO4351	ATHHANKA	ATTS0966	ATTSO262
C24	142	155	163	170	201	145	125	136	194	118	111	111
Ler-0	142	164	163	173	207	142	125	151	197	122	111	121
Col-4	142	155	154	173	207	142	125	136	197	118	111	103
Be-0	142	161	154	173	207	145	128	151	197	118	111	103
RLD1	142	161	154	173	na	142	125	136	na	118	111	106
S96	145	152	154	176	207	142	128	145	197	118	111	103
Conaara	na	na	166	173	207	142	128	na	197	118	111	115
Enkheim-T	142/145[a,*]	155	163	173/176*	207	na	na	142	197	na	111	103/111/118*
Rubezhnoe-1	na	155	166	173	207	142	125	145	na	118	111	113*
Shahdara	142	161	166	173	207	142	128	145	197	118	111	106
Aa-0	142	155	163	173	207	142	na	142	197	118	111	103/118*
Ag-0	142	158	154	179	207	142	128	136	197	118	111	103
An-1	142	161	154	173	201	142	128	127	197	127	111	118
Ang-0	142	155/158*	154	173	207	148	128	145	197	118	111	103/113*
Bay-0	na	na	154	173	207	145	125	145	197	118	111	106
Bl-1	142	164	163	170	207	145	125	142	197	127	111	106
Bla-12	142	155	154	173	207	142	125	136	197	118	111	103
Bs-1	142	155	154	173	207	142	125	127	197	118	111	111
Bsch-0	142	155	154	173	207	145	125	145	na	118	111	103
Bur-0	142	155	154	170/173*	201/207*	142	128	136/142*	194/197*	118	111	103/111*
Ca-0	142	161	154	173	207	142	125	148	197	118	111	103
Cal-0	142	155	154	170/173*	201/207*	142	128	142	194/197*	118	111	103
Cen-0	142	158	154	176	207	142	125	154	197	118	111	113
Chi-1	142	na	163	173/176	207	142	128	145	197	127	111	113
Co-1	142	na	163	170	201	145	125	136	na	118	111	111
Ct-1	142	na	154	173	207	142	128	136	197	127	111	111*
Cvi-0	145	155	154	176	207	na	125	136	197	118	111	118
Di-0	145	155	163	173	207	142	125	151	197	122	111	103
Dr-0	142	152	163	173	207	145	128	145	197	127	111	111*
Dra-0	142	155	154	170/173*	201	142	128	136/145*	194/197*	118	111	111*
Edi-0	142	161	154	173	201	142	125/128	151	197	118	111	103*
Eil-0	142	155/164*	154/163*	170/173*	201	142/145*	128	136/151*	194/197*	118/127*	111	118*
Es-0	142	161	154	173	207	142/145	125	145	197	127	111	111
Ge-1	142	158	154	173	201	142	125	136	197	118/127	111	111*
Gr-2	142	164	163	170	207	145	125	142	197	127	111	106
Gy-0	142	158	154	173	207	142	128	151	197	118	111	99
Je-0	145	155	154	173	201	142	125	148	197	118	111	113
Jl-1	142	155/158*	154/163*	173/176*	201/207*	142/145*	125	145	197	118/127*	111	103*
Kas-1	142	170	166	173	207	142	125	139	197	118	111	106*
Kä-0	145	155	154	182	207	142	128	154	197	118	111	111*
Kn-0	145	167	154	173	207	142	128	136	197	127	111	111
Lc-0	142	143	154	173	207	142	125	142	197	118	111	118
Lip-0	142	143/164*	154/163*	173	207	145	128	136	197	118/127*	111	108/115*
Ll-0	142	155/164*	154	173	201/207*	142	131	136	194/197*	118/127*	111	111/115/121*
Lm-2	145	155	154	173	201	142	128	151	197	118	111	103
Lz-0	142	161	154	176	207	142	125	139	197	127	111	103
Mh-0	142	158	154	173	201	142	125	145	197	118	111	115*
Mt-0	142	155	154	na	207	142	125	136	197	118	111	106
Nc-1	142	155	154	173	207	142	128	151	197	118	111	111/115*

[a] Locus/accession combinations with an asterisk have been further examined by PCR amplifications on at least 15 individual plants

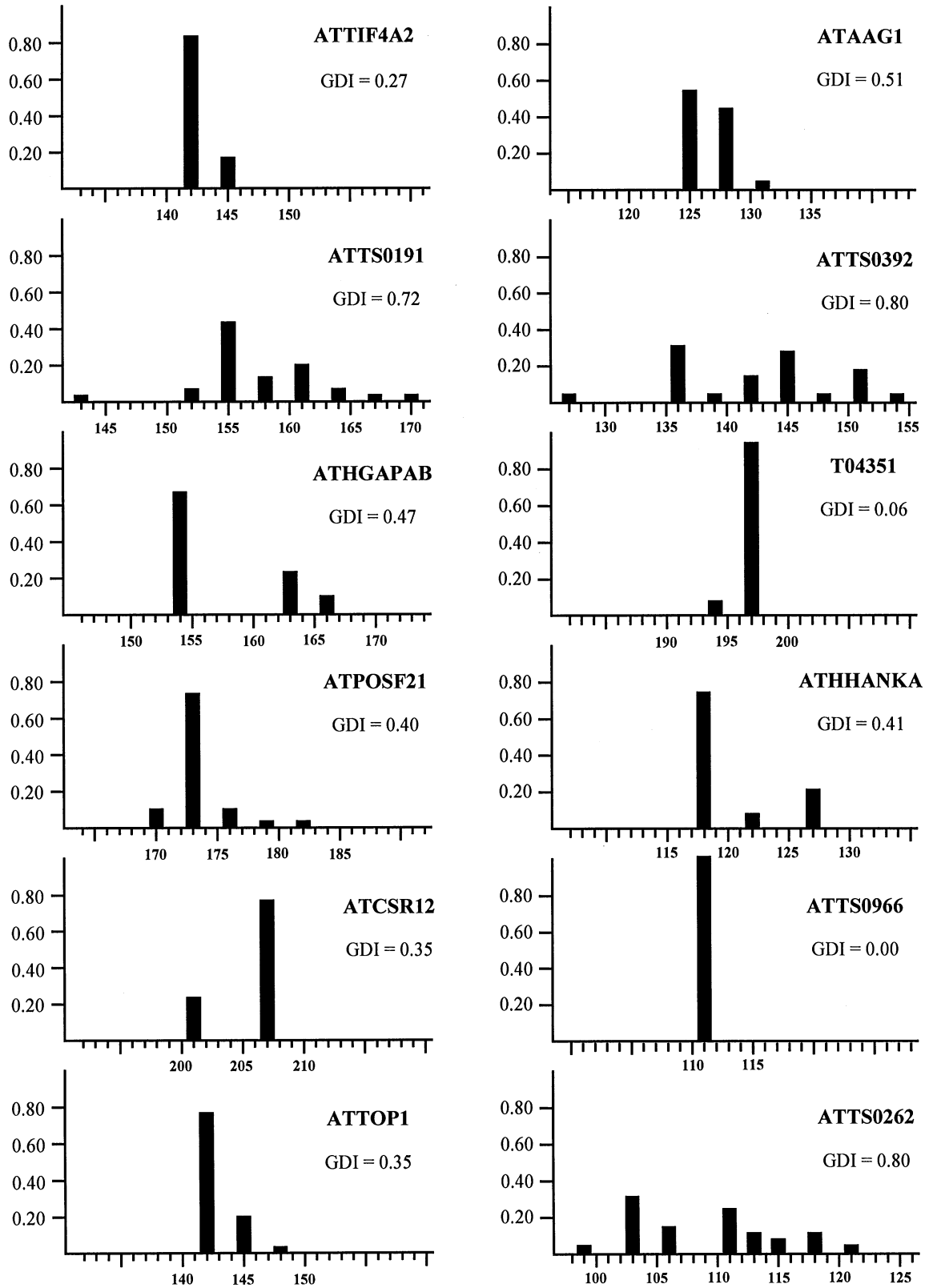


Fig. 1 Allele frequencies and genetic diversity index (GDI) of 12 trinucleotide microsatellite loci among 49 *Arabidopsis thaliana* accessions. The *X axis* represents allele lengths (in basepairs); the *Y axis* represents allele frequencies

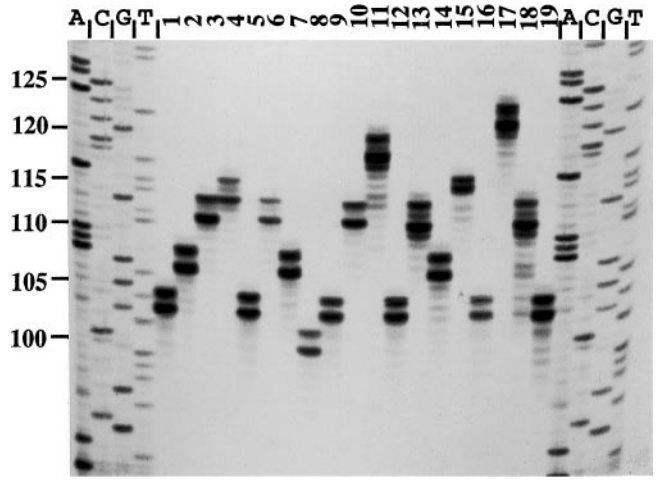


Fig. 2 Autoradiogram of PCR amplifications at locus ATTS0262 on 19 *Arabidopsis thaliana* accessions. *Lanes: 1* Ag-0, *2* Bl-1, *3* Ge-1, *4* Cen-0, *5* Di-0, *6* Dr-0, *7* Gr-2, *8* Gy-0, *9* Jl-1, *10* Kä-0, *11* Lc-0, *12* Lm-2, *13* Kn-0, *14* Mt-0, *15* Mh-0, *16* Col-4, *17* Ler-0, *18* C24, *19* Be-0. This microsatellite locus is characterized by a specific pattern with two major bands; the *upper one* was chosen to define the allele size. The lanes labeled *A*, *C*, *G* and *T* correspond to the sequence ladders of M13 phage as size references

Analysis of "multiple bands" patterns

Multiple bands corresponding to 2 or 3 different alleles were found in about 7% of the PCR reactions (Table 3). For at least 1 microsatellite locus 15 ecotypes are concerned. To investigate this phenomenon, we performed amplifications from DNA of at least 15 individual plants from 8 of the 9 ecotypes showing these types of patterns for at least 2 loci. In most cases (6 accessions of 8), a single fragment was found in each individual plant that corresponded to one of the bands observed in the pooled plants. For example, at locus ATTS0392, 5 Bur-0 individuals showed a 142-bp fragment and 10 Bur-0 individuals a 136-bp fragment (Table 4a); both alleles were found in pooled plants (Table 3).

Moreover, the data showed that only two haplotypes are present in Bur-0, 5 plants being homozygous for the allelic association 173-207-142-197-103 and 10 for the association 170-201-136-194-111. The same observation was made for ecotypes Cal-0, Dra-0, Eil-0 and Jl-1, where only two haplotypes were found, and for Enkheim-T, which displayed three haplotypes (Table 4b). Lip-0 and Ll-0 (Tables 4c and 4d) were the only accessions in which individual plants showing a two-band pattern could be found. The hypothesis of a somatic mosaicism is very unlikely because a 1 : 1 stoichiometry between the two bands was always observed. In Ll-0, only 1 plant out of the 15 tested was heterozygous for a single locus, T04351 (Table 4c), while in Lip-0, 20 plants out of 50 were heterozygous for 1–3 loci (Table 4d). In these 2 ecotypes, at least four different haplotypes can be observed. In Lip-0, however, one of them, 164-154-118-115, was clearly more frequent than the others. There was a close association between alleles present at loci ATTS0191 and ATTS0262, which can probably be explained by their genetic linkage since they are separated by only 12.8 cM, as determined in our mapping study. It can also be noted that in this ecotype, only homozygous plants were found at locus ATHHANKA while the 3 other loci analyzed showed a significant proportion of heterozygous plants.

To determine whether the observed intra-ecotype polymorphism is specific to microsatellite loci or extends to other parts of the genome, 12 RFLP markers were hybridized with genomic DNA extracted from pooled plants belonging to 23 ecotypes (Table 5): 8 accessions showed a single PCR band at each microsatellite locus (used as controls) and 15 accessions showed multiple bands for at least 1 microsatellite locus. The number of RFLP and microsatellite markers displaying multiple bands is given for each ecotype. The results show a clear correlation between the variability of the microsatellite loci and the RFLP markers, demonstrating that the heterogeneity observed in some ecotypes was not restricted to microsatellite sequences.

Microsatellite phylogenetic distances

Phylogenetic analyses were performed using the variations in microsatellite length polymorphisms with or without considering the allele frequencies observed. The data set was mainly derived from Table 3 and the allele frequencies from Tables 3 and 4. Two main stepwise mutation model (SMM)-based distances were used to estimate microsatellite-derived phylogenetic relationships among *A. thaliana* ecotypes: (1) the delta mu genetic distance and (2) the average square distance. Identical neighbor-joining (NJ) trees were recovered with both distance measures. However, differences in the NJ tree topologies were observed for each method whether the allele frequencies were considered or not. These differences in topology were restricted to the poorly resolved clusters/nodes of the trees. The NJ tree derived from the delta mu genetic distance and having taken the allele frequencies into consideration is shown in Fig. 4a. The main lineages, which were always recovered independent of allele frequencies and distance method used, are as follows: (A) a lineage grouping ecotypes Aa-0, Cvi-0 and Enkheim-T, (B) one grouping Co-1, Rubezhnoe-1, Bsch-0, and RLD1, (C) one grouping Chi-1, Ct-1, Bay-0 and Condara, and two well-separated individual lineages corresponding to ecotypes Lc-0 and Mt-0. Other much less well-resolved clusters but ones still recovered with or without using the allele frequencies are as follows: (D) Mt-0 groups with Kas-1, Shahdara, Gr-2, Bl-1 and C24; (E) ecotypes of group A are clustered with Bs-1, Col-4, Bla-12,

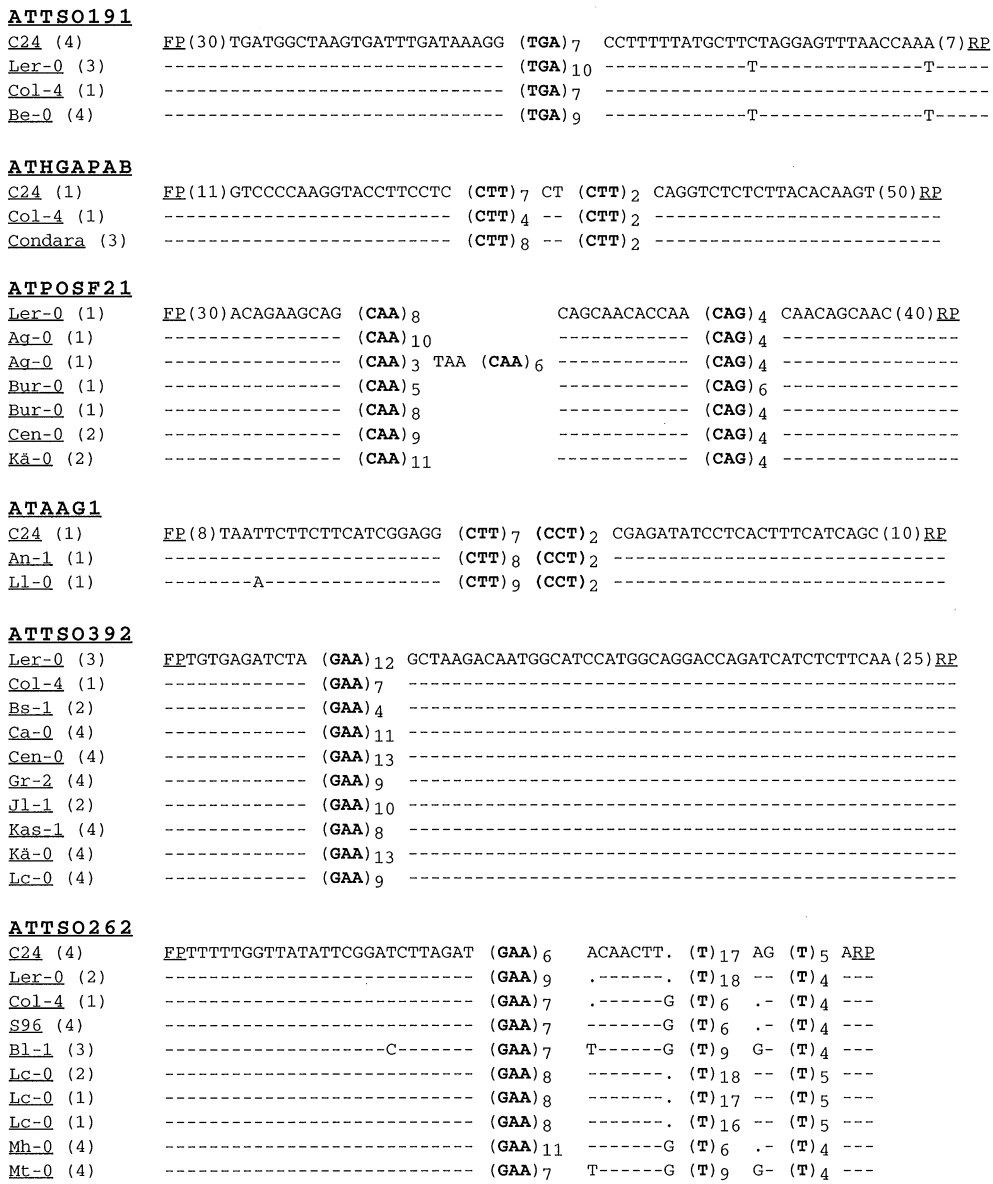


Fig. 3 Sequence comparisons between trinucleotide microsatellite alleles. The number of sequenced clones is noted in *brackets* near the *underlined* ecotypes. Microsatellite motifs are written in *bold*. *FP* = Forward primer, *RP* = reverse primer. The *number in brackets* next to *FP* or *RP* indicates the number of basepairs between a primer and the following sequence. *Dashes* denote identities; *dots* denote insertions/deletions. These sequences are available from GenBank/EMBL/DDBJ under accession numbers AF027613 to AF027649

Ag-0 and Lz-0; (F) a cluster grouping Kn-0, Ll-0, and An-1; (G) one grouping Eil-0 and Es-0; (H) one grouping Dr-0, Lc-0, and clusters (B) and (C); (I) one grouping Gy-0, Be-0, Ca-0, Edi-0 and Lm-2; and (J) one grouping Di-0 and Jl-1.

Discussion

One of the main attributes of microsatellite loci is their hypervariability, which makes them generally more informative than other molecular markers, as RFLP (Wu and Tanksley 1993; Rongwen et al. 1995; Taramino and Tingey 1996). The average heterozygosity reported for microsatellites has generally been between 0.5 and 0.8 in mammals (Takezaki and Nei 1996) and between 0.5 and 0.9 in plants (Wu and Tanksley 1993; Saghai Maroof et al. 1994; Rongwen et al. 1995; Echt et al. 1996; Taramino and Tingey 1996).

Allelic diversity appears to depend on the pattern of repeats and on the species being studied. Broun and Tanksley (1996) found a very weak variability at (GA)n and (GT)n loci in tomato, whereas high levels of polymorphism were reported for various microsatellites in rice (Wu and Tanksley 1993), soybean (Morgante and Olivieri 1993), barley (Saghai Maroof et al. 1994) and maize (Taramino and Tingey 1996).

In *A. thaliana*, dinucleotide loci have been shown to be very polymorphic, with an average allele number

of 4.1 among 6 accessions (Bell and Ecker 1994), of 5.3 among 12 Japanese natural populations (Todokoro et al. 1995) and of 10.6 among 42 natural populations (Innan et al. 1997). The average gene diversity values were, respectively, 0.67 and 0.79 (Todokoro et al. 1995; Innan et al. 1997). In this report, the extent of polymorphism was studied at 12 trinucleotide microsatellites among 49 *A. thaliana* accessions. They displayed variable levels of polymorphism, with 1–8 alleles (average number of alleles per locus = 4) and genetic diversity indices varying from 0.00 to 0.80 (average genetic diversity index = 0.43). These values are consistent with the lower polymorphism levels reported in other plant species for trinucleotide microsatellites when compared to dinucleotide microsatellites (Saghai Maroof et al. 1994; Rongwen et al. 1995; Echt et al. 1996).

To investigate the molecular basis of difference between length variants at the nucleotide level, we cloned and sequenced several alleles from 6 microsatellite loci. As expected, length polymorphism resulted only from variations in the number of trinucleotide repeats (except at locus ATTS0262), since no addition or deletion events were observed in the flanking sequences. It is interesting to note that when a locus is made of two distinct blocks of trinucleotide repeats, length variations preferentially affect the longest one.

Many authors have found a positive correlation between polymorphism and structure and/or length of microsatellites (Weber 1990; Thomas and Scott 1993; Wu and Tanksley 1993; Saghai Maroof et al. 1994) though others have not come to the same conclusions (Valdes et al. 1993; Bell and Ecker 1994). In their analysis of Arabidopsis dinucleotide microsatellites, Innan et al. (1997) showed a clear correlation between the estimated number of repeats in the alleles present at a locus and the number of alleles. Similarly, we estimated the number of repeats present at an allele from the length of the corresponding PCR product, except for loci ATHHANKA and ATTS0262, where alleles did not always differ by multiples of 3 bp and sequencing results revealed variations in the flanking sequences. From our data, no significant relationship ($r = 0.307$; $P > 0.1$) was found between the average number of repeats and the standard deviation of the repeat number, but a positive and significant correlation ($r = 0.592$; $P < 0.05$) was found between the average number of repeats and the number of alleles present at a microsatellite locus. This last observation is in agreement with our findings that length variations preferentially affect the longest stretch at loci composed of two distinct arrays of trinucleotide repeats and that no polymorphism was detected at locus ATTS966, which consists of three short blocks of repeats (Tables 2 and 3). These conclusions suggest that the degree of polymorphism is positively correlated with the number of contiguous repeats, which could be explained by slipped-strand mispairing during replication.

Table 4 PCR amplifications of microsatellite loci from individual plants from *Arabidopsis thaliana* ecotypes Bur-0, (**a**), Enkheim-T (**b**), Ll-0 (**c**), Lip-0 (**d**). Allele lengths are noted in basepairs; haplotypes are reported at the bottom of each table, (na = no amplification)

a

Bur-0	ATPOSF21	ATCSR12	ATTS0392	T04351	ATTS0262
1	173	207	142	197	103
2	170	201	136	194	111
3	170	201	136	194	111
4	170	201	136	194	111
5	173	207	142	197	103
6	173	207	142	197	103
7	170	201	136	194	111
8	173	207	142	197	103
9	173	207	142	197	103
10	170	201	136	194	111
11	170	201	136	194	111
12	170	201	136	194	111
13	170	201	136	194	111
14	170	201	136	194	111
15	170	201	136	194	111
	173	207	142	197	103
	170	201	136	194	111

b

Enkheim-T	ATTIF4A2	ATPOSF21	ATTS0262
1	145	176	103
2	142	170	111
3	145	na	103
4	142	173	118
5	145	176	103
6	142	173	118
7	142	173	118
8	145	176	103
9	145	176	103
10	142	170	111
11	145	176	103
12	145	na	103
13	142	173	118
14	142	na	118
15	145	176	103
	145	176	103
	142	170	111
	142	173	118

c

Ll-0	ATTS0191	ATCSR12	T04351	ATTS0262
1	na	201	na	121
2	na	201	197	121
3	na	207	197	na
4	155	207	197	115
5	155	207	197	111
6	164	201	197	121
7	164	201	194	121
8	155	207	197	115
9	na	201	197	121
10	na	201	194/197	121
11	164	201	194	121
12	na	201	na	121
13	155	207	197	111
14	155	207	na	na
15	155	207	na	115
	155	207	197	115
	164	201	197	121
	164	201	194	121
	155	207	197	111

Table 4 Continued
d

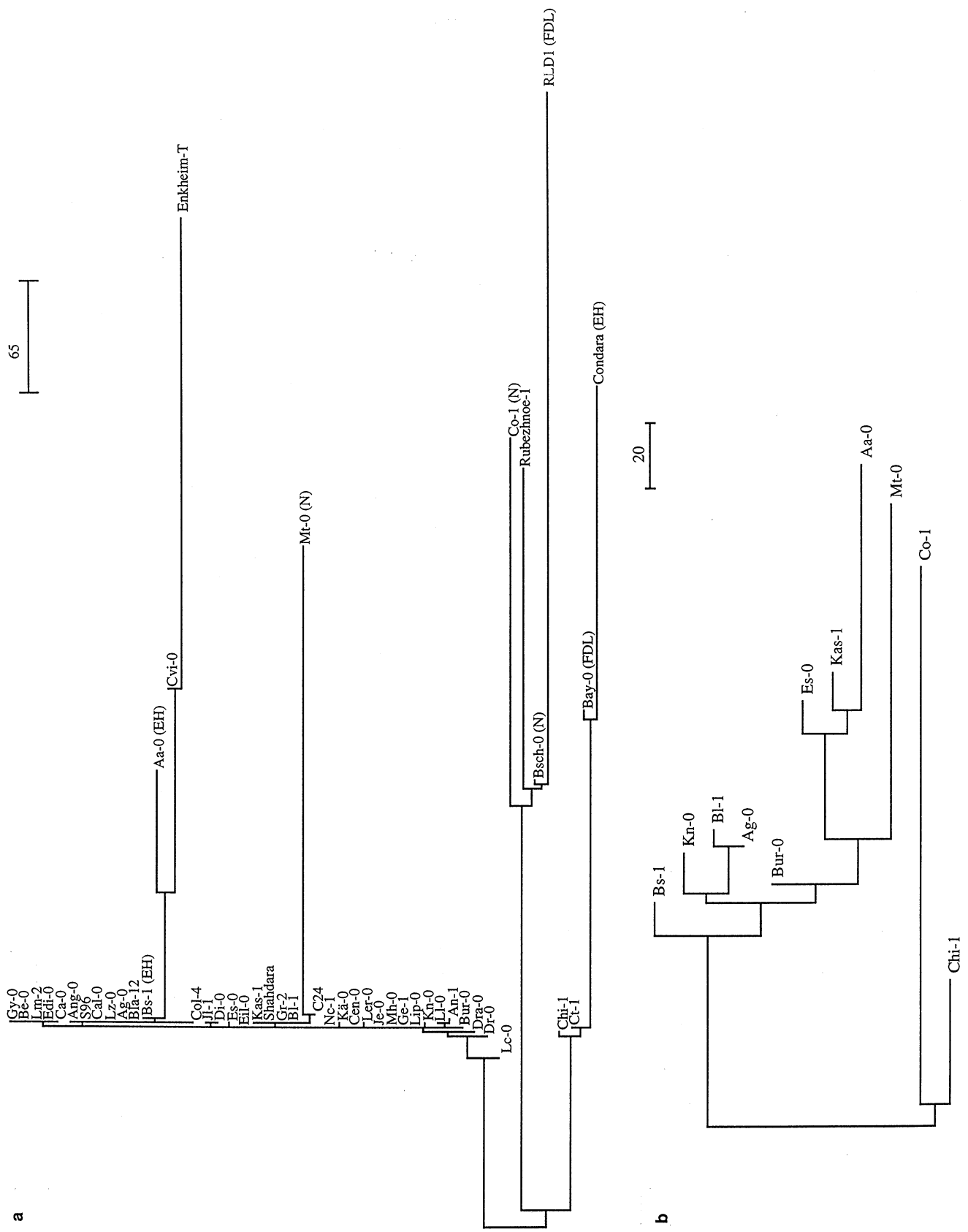
Lip-0	ATTS 0191	ATHG APAB	ATHH ANKA	ATTS 0262	Lip-0	ATTS 0191	ATHG APAB	ATHH ANKA	ATTS 0262
1	143/164	154/163	127	108/115	26	164	154	118	115
2	164	154	118	115	27	143/164	163	127	108/115
3	164	154	118	115	28	143/164	154	127	108/115
4	164	154	118	115	29	164	154	118	115
5	143/164	154/163	127	108/115	30	164	154	118	115
6	164	154	118	115	31	164	154	118	115
7	164	154	118	115	32	164	154	118	115
8	164	154	127	108/115	33	143/164	163	127	108/115
9	143/164	154/163	127	108/115	34	164	154	127	115
10	164	154	118	115	35	143/164	154	127	108/115
11	164	154	127	115	36	164	154	118	115
12	143	163	127	108	37	164	154	118	115
13	143/164	163	127	108/115	38	143/164	154/163	127	108/115
14	143	154/163	127	108	39	143/164	163	127	108/115
15	164	154	118	115	40	164	154/163	127	115
16	143/164	154/163	127	108/115	41	164	154	118	115
17	164	163	127	115	42	164	154	118	115
18	164	163	127	115	43	164	154	118	115
19	143/164	163	127	108/115	44	164	154/163	127	na
20	143	154	127	108	45	na	154	na	na
21	164	154	118	115	46	164	163	127	115
22	164	154	118	115	47	143/164	154	127	108/115
23	164	154	118	115	48	143	154/163	127	108
24	143/164	154/163	127	108/115	49	143	154/163	127	108
25	164	154	118	115	50	164	154	118	115
						164	154	118	115
						164	154	127	115
						143	163	127	108
						164	163	127	115
						143	154	127	108

Table 5 Comparisons between RFLP and microsatellite markers

Ecotype[a]	Number of mb[b]/ Total number of RFLP	Number of mb/ Total number of microsatellites	Ecotype	Number of mb/ Total number of RFLP	Number of mb/ Total number of microsatellites
*Ler-0	0/12	0/12	- Cal-0	4/12	2/12
*Col-4	0/12	0/12	- Chi-1	1/12	1/12
*C24	0/12	0/12	- Dra-0	4/12	3/12
*Be-0	0/11	0/12	- Edi-0	0/12	1/12
*S96	1/12	0/12	- Eil-0	5/12	8/12
*Bl-1	0/12	0/12	- Es-0	1/12	1/12
*Kä-0	0/12	0/12	- Ge-1	0/12	1/12
*Lm-2	1/12	0/12	- Jl-1	6/12	6/12
- Enkheim-T	7/12	3/12	- Lip-0	2/11	4/12
- Aa-0	0/12	1/12	- Ll-0	9/12	4/12
- Ang-0	2/12	2/12	- Nc-1	1/12	1/12
- Bur-0	6/12	5/12			

[a] * = accessions showing a single PCR fragment for all microsatellite loci (controls), - = accessions showing multiple bands for at least 1 microsatellite locus
[b] Number of mb = number of markers displaying multiple bands

Fig. 4a, b Neighbor-joining (NJ) phylogenetic trees of *Arabidopsis thaliana* ecotypes using microsatellite loci. Genetic distances were computed according to the delta mu square method. *Horizontal distances* represent the squared mean difference between alleles of two ecotypes. *Vertical distances* are for clarity only. The trees were arbitrarily separated into two parts at their midpoint. **a** NJ tree of 49 Arabidopsis ecotypes using trinucleotide repeats microsatellite loci. Reaction phenotypes of Arabidopsis challenged with *Peronospora parasitica Noco*2 isolate (as reported by Reignault et al. 1996) are given in brackets: N = necrotic phenotype (resistant host), FDL = necrotic flecks with delayed and light sporulation, EH = early and heavy sporulation (susceptible host). **b** NJ tree of 11 Arabidopsis ecotypes using di- and trinucleotide repeats microsatellite loci. The data matrix was derived from Table 3 and from Innan et al. (1997), but frequencies were not considered


a
65
Gy-0
Be-0
Lm-2
Edi-0
Ca-0
Ang-0
S96
Cal-0
Lz-0
Ag-0
Bla-12
Bs-1 (EH)
Aa-0 (EH)
Cvi-0
Enkheim-T
Col-4
Jl-1
Di-0
Es-0
Eil-0
Kas-1
Shahdara
Gr-2
Bl-1
Mt-0 (N)
C24
Nc-1
Kä-0
Cen-0
Ler-0
Je-0
Mh-0
Ge-1
Lip-0
Kn-0
Ll-0
An-1
Bur-0
Dra-0
Dr-0
Lc-0
Chi-1
Ct-1
Bay-0 (FDL)
Bsch-0 (N)
Rubezhnoe-1
Co-1 (N)
RLD1 (FDL)
Condara (EH)
b
20
Bs-1
Kn-0
Bl-1
Ag-0
Bur-0
Es-0
Kas-1
Aa-0
Mt-0
Co-1
Chi-1

In their analyses on Arabidopsis dinucleotide microsatellites, Bell and Ecker (1994), Todokoro et al. (1995) and Innan et al. (1997) always found only 1 allele per accession, except for 1 locus in 1 accession (Bell and Ecker 1994), for which the heterogeneity was not further investigated. However, 2 alleles per accession have been observed in a few cases in soybean, a selfing species (Morgante and Olivieri 1993) and in maize inbred lines (Taramino and Tingey 1996) at 2 microsatellite loci (1 di- and 1 trinucleotide) and 1 dinucleotide locus, respectively. In the present report, double-banded patterns were found in about 7% of the PCR amplifications from pooled DNA in Arabidopsis accessions and involved 15 ecotypes out of the 49 studied. Eight of these, showing polymorphism at 2 loci or more, were analyzed. In most cases (6 ecotypes), only homozygous plants were detected. Moreover, only a small number of haplotypes, two (Bur-0, Cal-0, Dra-0, Eil-0 and Jl-1) or three (Enkheim-T) were detected. This means that these accessions consist of two sublines (three for Enkheim-T) which were propagated together without having recently exchanged genetic information, since no heterozygous plant was found. This situation probably reflects the original diversity of the natural populations when these accessions were collected.

The other 2 ecotypes showed a more complex situation. In Lip-0, a significant proportion of heterozygous plants for 3 of the 4 polymorphic loci were found: 11/50 for ATHPAGAB, 14/50 for ATTS0191, 15/50 for ATTS0262. Moreover, various haplotypes could be distinguished, indicating that this accession very likely consists of a mix of plants resulting from segregations of recent crosses. Since no heterozygous plant was found for locus ATHHANKA, the most simple hypothesis is to assume that at least three original sublines coexisted in this ecotype. One is characterized by the haplotype 164-154-118-115, which was found in 22 out of the 29 homozygous plants at the 4 loci and was not involved in these recent crosses. Two other sublines shared the 127 allele of ATHHANKA and differed by the alleles present at the 3 other loci. Crosses between these two lines will generate, in the F_2 and the following generations, all the allelic associations observed in the 27 plants not belonging to the first subline. Since the history of this ecotype is not known precisely (M. Anderson, personal communication), it is not possible to know whether the crosses occurred in the natural population or after its collection. In the ecotype Ll-0, some crosses probably also occurred recently since 1 plant was found to be heterozygous at locus T04351, and four different haplotypes were identified in 15 individual plants. These results suggest that the outcrossing rate might be higher than expected in *A. thaliana*, and the observation of plants heterozygous for isozymes (M. Lefranc and B. Godelle, personal communication) strengthens this hypothesis.

Phylogenetic distances between *A. thaliana* ecotypes derived from microsatellites length polymorphisms and their frequencies were estimated according to stepwise mutation model (SMM)-based distances. These distances imply changes in allelic sizes by one or very few number of repeats (Ohta and Kimura 1973). They increase linearly with time, making them a preferable method for estimating phylogenetic relationships and the relative separation times (for review, see Goldstein et al. 1995b). The phylogenetic relationships among *A. thaliana* ecotypes using SMM-based distances were compared with a recently published *A. thaliana* phylogenetic tree based on nuclear and chloroplast DNA sequences (Hardtke et al. 1996). In this DNA-derived phylogenetic tree, the Mt-0 lineage is clearly apart from the Kas-1, RLD1, Di-0 and Ll-0 lineages. In our analyses, Mt-0 is also clearly apart from the other ecotypes, suggesting that the right tree topology is likely to have been recovered (at least for the distant ecotypes).

Among the main lineages recovered using our estimated microsatellite phylogenetic distances, lineages B and C are grouped together and closely related to the Lc-0 lineage but apart from group A and Mt-0 lineages which are closely related. The geographic origin of the ecotypes was compared with the SMM NJ trees main lineages, but no obvious concordance between the country of origin and the phylogenetic clusters could be drawn. This does not seem atypical of our data set since the same observation was made with other data sets (King et al. 1993; Innan et al. 1996, 1997). The various reaction phenotypes of pathogens on *A. thaliana* (e.g. Reignault et al. 1996; Holub et al. 1995; Lummerzheim et al. 1993) and the above phylogenetic clusters/lineages were also compared. In one particular case, i.e. the reaction phenotypes of *A. thaliana* against *Peronospora parasitica*, some phenotypes seemed to be associated with specific clusters (see Fig. 4A): the Mt-0 cluster was associated strictly with the necrotic phenotype (resistant hosts); the Aa-0 lineage and closely related Bs-1 lineage with the early and heavy sporulation phenotype (susceptible hosts); and the necrotic phenotype was not observed among a same main phylogenetic cluster of the NJ tree together with the early and heavy sporulation phenotype (susceptible hosts) but sometimes an intermediate phenotype (necrotic flecks with delayed and light sporulation) was observed among each group. These observations suggest that *A. thaliana* interactions with pathogens might have affected its evolution, leading to the emergence of disease-specific phylogenetic clusters.

The above well-resolved clusters of the NJ trees were also compared to the recently reported microsatellite NJ tree based on the diversity of dinucleotide repeats loci (Innan et al. 1997). These phylogenetic trees show some similarities that we have reevaluated by the delta mu square method (without frequencies) while restricting the analyses to ecotypes used in both studies. The

NJ trees derived from these analyses are similar. The main lineages are conserved i.e. Co-1 groups with Chi-1 apart from Mt-0 and Aa-0 which are grouped together. Changes in the tree topology are confined to nodes that are related to the previous poorly resolved nodes/clusters observed among the global SMM-based distances NJ tree (Fig. 4a). The data sets were pooled to improve the resolution and robustness of the analysis. The inferred NJ tree derived from the pooled data sets is likely to be a better reconstruction of the phylogenetic relationships among *A. thaliana* ecotypes (Fig. 4b).

Acknowledgements We would like to thank the Nottingham Arabidopsis Stock Centre for providing the ecotype seed collection, R. Whittier for providing RFLP markers, J. Lafleuriel for genetic mapping analyses and S. Genestier for technical assistance. We are grateful to C. Camilleri and D. Bouchez (INRA, Versailles) for their help in the physical mapping of microsatellite loci using the CIC YAC library. This work was supported by CNRS, Université Blaise Pascal, and GREG (Groupement de Recherches et d'Etudes sur les Génomes).

References

Beckmann JS, Soller M (1990) Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. Biotechnology 8:930–932

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

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455–457

Broun P, Tanksley SD (1996) Characterization and genetic mapping of simple repeat sequences in the tomato genome. Mol Gen Genet 250:39–49

Burt DW, Bumstead N, Bitgood JJ, Ponce de Leon FA, Crittenden LB (1995) Chicken genome mapping: a new era in avian genetics. Trends Genet 11:190–194

Creusot F, Fouilloux E, Dron M, Lafleuriel J, Picard G, Billault A, Paslier DL, Cohen D, Chabouté M-E, Durr A, Fleck J, Gigot C, Camilleri C, Bellini C, Caboche M and Bouchez D (1995) The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. Plant J 8:763–770

Depeiges A, Goubely C, Lenoir A, Cocherel S, Picard G, Raynal M, Grellet F, Delseny M (1995) Identification of the most represented repeated motifs in *Arabidopsis thaliana* microsatellite loci. Theor Appl Genet 91:160–168

Dib C, Fauré S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5264 microsatellites. Nature 380:152–154

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalis KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L, Goodman N, Copeland NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES (1996) A comprehensive genetic map of the mouse genome. Nature 380:149–152

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

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterization of microsatellite markers in eastern white pine. Genome 39:1102–1108

Estoup A, Presa P, Krieg F, Vaiman D, Guyomard R (1993) (CT)-n and (GT)-n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity 71:488–496

Goldstein DB, Clark AG (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. Nucleic Acids Res 23:3882–3886

Goldstein DB, Ruiz-Linares A, Feldman M, Cavalli-Sforza LL (1995a) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6720–6727

Goldstein DB, Ruiz-Linares A, Feldman M, Cavalli-Sforza LL (1995b) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Hardtke CS, Müller J, Berleth T (1996) Genetic similarity among *Arabidopsis thaliana* ecotypes estimated by DNA sequence comparison. Plant Mol Biol 32:915–922

Holub EB, Brose E, Tör M, Clay C, Crute IR, Beynon JL (1995) Phenotypic and genotypic variation in the interaction between *Arabidopsis thaliana* and *Albugo candida*. Mol Plant Microbe Interact 8:916–928

Innan H, Tajima F, Terauchi R, Miyashita NT (1996) Intragenic recombination in the *Adh* locus of a wild plant *Arabidopsis thaliana*. Genetics 143:1761–1770

Innan H, Terauchi R, Miyashita NT (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146:1441–1452

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. Trends Ecol Evol 11:424–429

King G, Nienhuis J, Hussey C (1993) Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms. Theor Appl Genet 86:1028–1032

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

Lister C, Dean C (1993) Recombinant inbred for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750

Lister C, Dean C (1997) The latest RI map using Lister and Dean RI lines. Weeds World 4:1–10

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

Liu Y-G, Mitsukawa N, Lister C, Dean C, Whittier RF (1996) Isolation and mapping of a new set of 129 RFLP markers in *Arabidopsis thaliana* using recombinant inbred lines. Plant J 10:733–736

Lummerzheim M, de Oliveira D, Castresana C, Miguens FC, Louzada E, Roby D, van Montagu M, Timmerman B (1993) Identification of compatible and incompatible interactions between *Arabidopsis thaliana* and *Xanthomonas campestris* pv. *campestris* and characterization of the hypersensitive response. Mol Plant Microbe Interact 6:532–544

Meyerowitz EM (1989) Arabidopsis, a useful weed. Cell 56: 263–269

Moore SS, Sargeant LL, King TJ, Mattick JS, Georges M, Hetzel DJS (1991) The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10: 654–660

Moran C (1993) Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. J Hered 84:274–280

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

Ohta T, Kimura M (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22 : 201–204

Panaud O, Chen X, McCouch SR (1995) Frequency of microsatellite sequences in rice (*Oryza sativa* L.). Genome 38 : 1170–1176

Reignault P, Frost LN, Richardson H, Daniels MJ, Jones JDG, Parker JE (1996) Four *Arabidopsis RPP* loci controlling resistance to the *Noco*2 isolate of *Peronospora parasitica* map to regions known to contain other *RPP* recognition specificities. Mol Plant Microbe Interact 9 : 464–473

Rico C, Rico I, Hewitt G (1996) 470 million years of conservation of microsatellite loci among fish species. Proc R Soc London (Ser) B Biol Sci 263 : 549–557

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

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74 : 5463–5467

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

Sharma PC, Winter P, Bünger B, Weigand F, Weising K, Kahl G (1995) Abundance and polymorphism of di-, tri- and tetranucleotide tandem repeats in chickpea (*Cicer arietinum* L.). Theor Appl Genet 90 : 90–96

Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE, Moyzis RK (1991) Evolution and distribution of (GT)n repetitive sequences in mammalian genomes. Genomics 10 : 807–815

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365 : 274–276

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144 : 389–399

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39 : 277–287

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

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

Todokoro S, Terauchi R, Kawano S (1995) Microsatellite polymorphisms in natural populations of *Arabidopsis thaliana* in Japan. Jpn J Genet 70 : 543–554

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133 : 733–749

Valvekens D, van Montagu M, van Lijsebettens M (1988) Agrobacterium tumefaciens-mediated transformation of Arabidopsis thaliana root explants by using kanamycin selection. Proc Natl Acad Sci USA 85 : 5536–5540

Viard F, Justy F, Jarne P (1997) Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. Genetics 146 : 973–982

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 polymorphisms based on length variations in simple-sequence tandem repeats. In: Tilgman S, Davies K (eds) Genome analysis vol. 1: genetic and physical mapping. Cold Spring Harbor Laboratory Press, New York, pp 159–181

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

Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20 : 176–183