Theor Appl Genet (2004) 108:1010–1016
DOI 10.1007/s00122-003-1517-4

ORIGINAL PAPER

L. G. Fraser · C. F. Harvey · R. N. Crowhurst · H. N. De Silva

EST-derived microsatellites from *Actinidia* species and their potential for mapping

Received: 28 April 2003 / Accepted: 28 October 2003 / Published online: 20 November 2003

Abstract To increase the speed and reduce the cost of constructing a genetic map of *Actinidia* species (kiwifruit), for use in both breeding and functional genomics programmes, we sampled microsatellites from expressed sequence tags (ESTs) to evaluate their frequency of occurrence and level of polymorphism. Perfect dinucleotide repeats were the microsatellites selected, and these were found to be numerous in both the 5′ and 3′ ends of the genes represented. The microsatellites were of various lengths, the majority being repeats with the pattern $(CT)_n$/$(GA)_n$. One hundred and fifty microsatellites, each with more than 10 dinucleotide repeat units, were chosen as possible markers, and when these were amplified, 93.5% were found to be polymorphic and segregating in a mapping population, with 22.6% amplifying more than one locus. Four marker categories were identified. Fully informative markers made up 27% of the total, 36.2% were female informative, 25.8% were male informative and 10% partly informative. The mapping population was an intraspecific cross in the diploid species *Actinidia chinensis*, with parents chosen for their diversity in fruit and plant characteristics, and for their geographical separation. Linkage was tested using the software 'Joinmap' and a LOD value of 3. The distribution of the EST-based markers over the linkage groups obtained appeared to be random, taking into consideration the small sample size, that the number of linkage groups (31) exceeded the chromosome number of n=29, and that a number of markers were not assigned to any group. Some microsatellite markers which amplified more than one locus mapped to separate linkage groups. According to our study in *A. chinensis*, EST-derived microsatellites give large numbers of possible markers very quickly and at reasonable cost. The markers are highly polymorphic, segregate in the mapping population, and increase the value of the genomic map by providing some functional information.

Communicated by P. Langridge

L. G. Fraser (✉) · C. F. Harvey · R. N. Crowhurst · H. N. De Silva
The Horticulture and Food Research Institute of New Zealand,
120 Mt Albert Road, Auckland, New Zealand
e-mail: lfraser@hortresearch.co.nz
Tel.: +64-9-8154200
Fax: +64-9-8154201

Introduction

Microsatellites, or simple sequence repeats (SSRs), are ubiquitous in eukaryotic genomes and consist of repeats of 2–5 nucleotides in a continuous stretch. The number of core repeats is variable and has been attributed to strand slippage during DNA replication, or unequal exchange in meiosis. Experimental evidence suggests that the rate of slippage is dependent on the size of the repeat unit, being greatest for dinucleotides. The rate of slippage is also dependent on the unit sequence, GC repeats being the most stable in rice (Schlotterer and Tautz 1992; Valdes et al. 1993).

Polymorphism is revealed in the form of differences in length, which in turn depend on the number of repeat units. When comparing three marker systems (microsatellites, AFLPs and RAPDs) in barley and potato, Milbourne et al. (1998) found that microsatellites consistently showed the highest level of polymorphism, with 100% in barley and 90.8% in potato, while AFLPs exhibited the lowest level at 46.9% in barley and 41% in potato. RAPDs were intermediate at 66.3% in barley and 65.8% in potato. Microsatellites also have an advantage over AFLPs and RAPDs in that they are co-dominant markers and are able to distinguish all morphs at a locus, and are therefore more informative than the dominant markers.

Morgante et al. (2002) found that the overall frequency of microsatellites among species was inversely related to genome size and to the proportion of repetitive DNA, but remained constant in the transcribed regions, particularly the untranslated portions, of the genome. They reported a significantly higher frequency of microsatellites in ESTs than in the whole genomic DNA across species, with a higher frequency of CT/GA repeats and a lower frequency

of AT repeats in ESTs. AT repeats seemed to be typical of nontranscribed regions.

Previously, the cost of identifying and developing microsatellite markers has been high. This has, to a large extent, prevented their use in less important commercial crops. Enrichment protocols have been applied in an attempt to reduce costs, but as Scott et al. (2000) observed, one result of the direct bias created by these techniques is that these markers are less representative of all repeat motifs.

An alternative source of microsatellites is now being utilised. EST databases are being examined for their microsatellite-containing sequences, and these microsatellites are being used as markers in various studies (Scott et al. 2000; Eujayl et al. 2002). As the EST-derived microsatellites are found within transcribed regions of the genome these markers may be less polymorphic than those from untranscribed regions, but possible conservation of the primer sites could make them more transferable across species, an aspect which would increase their value in a breeding programme.

As EST-microsatellite markers directly sample variation in transcribed regions of the genome, they may provide an estimate of functional diversity. Instability in the number of repeats in microsatellites has been implicated in the inheritance and severity of some human diseases (Brook et al. 1992; The Huntington's Disease Collaborative Research Group 1993; Vincent et al. 2000). A similar variation in repeat numbers in plant systems could indicate variation in vegetative and fruit characters, which would increase their value in marker-assisted selection.

The genus *Actinidia* (kiwifruit) is native mainly to China and consists of over 60 species of long-lived, scrambling vines which bear edible fruit. The kiwifruit, while being a small crop on the fresh fruit markets of the world (about 1%), is nevertheless the most valuable horticultural crop grown in New Zealand, and the considerable flower and fruit variation in and between species offers potential both for breeding new cultivars, and for identifying valuable genes and gene products. *Actinidia* species, all dioecious, form a polyploid series from diploid to octaploid with n=29. The green-fleshed kiwifruit of commercial value, *A. deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson var. *deliciosa* 'Hayward', is hexaploid. A more recent introduction to the market, the yellow-fleshed kiwifruit (ZESPRI™ GOLD Kiwifruit) *A. chinensis* Planch. var. *chinensis* 'Hort16A', is diploid. Experimental evidence (McNeilage and Considine 1989; Huang et al. 1998) would suggest that while *A. chinensis* is a functional diploid, it is probably a cryptic (i.e. palaeo) polyploid.

An *Actinidia* map previously published by Testolin et al. (2001) was constructed with microsatellites from enriched genomic libraries together with AFLP markers. The population used to determine linkage was the result of an interspecific cross between two diploid species, *A. chinensis* and *A. callosa*. An interspecific cross was chosen with the aim of obtaining transportable markers, but proved unsatisfactory for segregation of plant characteristics.

Genetic markers are considered to be a useful tool both in breeding, where attributes only visible at maturity can be assessed at the seedling stage and the male contribution to fruit characters can be determined, and also for locating genes of interest in genomics programmes. In our laboratory a genetic map of the diploid species *A. chinensis* is under construction. We have used EST sequences from ten cDNA libraries, nine from *A. deliciosa* and one from *A. chinensis,* to assess the occurrence of microsatellites and to develop microsatellite primer pairs for amplification of polymorphic markers to evaluate their suitability for genetic mapping. It is thought that *A. deliciosa*, a species closely related to *A. chinensis*, is a polyploid derived from *A. chinensis* (Crowhurst and Gardner 1991).

Materials and methods

Microsatellite identification and primer design

Libraries of *A. chinensis* and *A. deliciosa* made from various tissues at various developmental stages were selected as the source of microsatellites. cDNA sequences from these libraries were analysed and annotated automatically using BioPipe, HortResearch's in-house bioinformatic sequence analysis platform (R. Crowhurst, M. Davy, C. Dong, unpublished). Microsatellites were mined using BioView (HortResearch's in-house viewer for BioPipe). Microsatellites with perfect dinucleotide repeats of more than 20 bp, and with sufficient sequence flanking the microsatellite to enable primer design, were chosen at random from the one *A. chinensis* and nine *A. deliciosa* libraries. Primers were designed from flanking regions using the software programme Primer3 (Rozen and Skaletsky 2000), and are given in Table 1. Primer pair sequences were selected to give a PCR product size between 200 and 300 bp, and with a preferred annealing temperature of approximately 60°C. Primers of 18–22 bases in length and with a GC content of approximately 50% were synthesised and fluorescently labelled (Dye Set DS-31) by Applied Biosystems Australia.

Plant materials

The mapping population used in this study is an intraspecific cross between two genotypes of diploid *A. chinensis* made in the spring of 1996. The female parent originated from seed from Henan province, Central China, while the male parent was selected from a seed accession from Guangxi province, South China. The parental genotypes were chosen for their geographic separation and the diversity in the fruit characteristics displayed by the progeny grown from each seed accession. Equal numbers of female and male seedlings from the cross were planted in a population of 300 plants in the HortResearch orchard in Te Puke, Bay of Plenty, New Zealand. Leaf tissue was taken from each plant at budbreak, held at 4°C for 24 h then stored at −80°C until required.

DNA extraction

DNA samples were extracted from leaf tissue from both parents and every individual in the mapping population. A small tissue sample was ground to powder in liquid nitrogen, then processed through a DNeasy Plant Mini Kit (Qiagen). The final eluate was 200 μl in volume.

Table 1 The sequences of the forward and reverse primers which amplify polymorphic and segregating microsatellites in an *Actinidia chinensis* mapping population. The library code (see Table 2 for tissue type), repeat motif and the alleles at each locus are given

Library code	SSR	Motif	Forward and reverse primers	Alleles (female×male)
KADA	Ke115	$(AG)_{21}$	5′ CGAAGAAACAGGAGAGAGAATCG GGCACGTTGTACATCAGGAG	aaxab, abxaa, 00xa0
KADA	Ke121	$(CT)_{17}$	5′ TCTGATTTCTCTCGCAGACG CATTTGTCCATCTGGCTGAA	abxcd
KABA	Ke129	$(CT)_{17}$	5′ CTGTGAAGATGGTGGGAAGC CAAATCAAGCCAATGACCAA	abxbb, a0xb0
KABA	Ke140	$(CT)_{18}$	5′ TCTTTCCCCTTCCCAAATCT TCTTGGGCTTGACAATCCAT	abxcc
A. chinensis	Ke/sp	$(CT)_{18}$	5′ ATGTGAATCGATACGTGCGTG CTTAAGTTCTCGATTTAATCAG	abxcd
KAIA	Ke150	$(CT)_{17}$	5′ CTCTCAGAGACAACTCTTCA GAGCTTGGATTTCTTGTCAG	abxcd
KAIA	Ke171	$(CT)_{19}$	5′ CGCAAAACAATACAGCAACTG GGCTTGATTGCTTGTGGAGT	abxcd
KABA	Ke183	$(CT)_{11}$	5′ TCAACCCATACCCAAGTGCT ATGTCGCGCTCAGATTCAC	aaxab
KABA	Ke184	$(CT)_{11}$	5′ CTCCATCTCTCTCCTAA GGTTCGAGCAATTGAGTCTC	abxaa
KALA	Ke199	$(CT)_{16}$	5′ CCGCAAGAACGGGTACATAG ATCACAGACGCAGACATGGA	abxcd
KABA	Ke200	$(AG)_{15}$	5′ GTCCCACATCCTTTTTGCAG ACGAGCTGGATTTGGATACG	abxcd
KABA	Ke207	$(CT)_{15}$	5′ ACCTGGTCTTTCCCCTTCC TCTTGGGCTTGACAATCCAT	abx00
KAIA	Ke209	$(AG)_{16}$	5′ GCTTGCCTCACCACTTATCC TGGTCTTGACCGGGACATA	abxcd
KAIA	Ke210	$(AG)_{16}$	5′ ATACAAAATCGCCGCTTGAA AGAACCCGAGCATCTGCTTT	a0xa0
KUBA	Ke214	$(CT)_{15}$	5′ CGTATTTAACCACGCACCAG TGGGGGTTTTAAGGGAGAAG	a0xbb, a0×00
KAIA	Ke217	$(AG)_{16}$	5′ AGCCCAGGGGAAACATCA GTGTGATCTGCACTCCCTGA	a0xb0, aaxab
KUFB	Ke221	$(AT)_{16}$	5′ TGAGTTGTGGGTATTGCAAGTT GCAGCAGTGCTAAACCTGTG	aaxab, a0×00
KAIA	Ke227	$(AG)_{16}$	5′ AGAGACTCGCCACTGCATTC GTGGCGACAGTGACGTAGAG	abxbb, a0×00, a0xb0
KABA	Ke239	$(AT)_{16}$	5′ ACTCGTGGACGTTCTGGTTC TCGAAGAGATCATACGCACAA	a0xab, a0×00
KADA	Ke244	$(GT)_{16}$	5′ CGCCTTCCTGCTG CCCACCACCCAAA	abxac, abxbb, a0xa0, 00xa0

Polymerase chain reaction

A total of 150 primer pairs were tested for PCR amplification and polymorphisms using the parental DNAs as template. PCR reactions were performed in a Techne Genius thermal cycler. A single cycle of 94°C for 3 min led into a programme of 35 cycles of denaturing at 94°C for 30 s, annealing at an optimal temperature for a particular primer pair at, or close to, 60°C for 30 s, and elongation at 72°C for 1 min. PCR products were assessed for size, concentration, and polymorphism using a DNA 500 LabChip Kit and an Agilent 2100 Bioanalyzer, supplied by Agilent Technologies.

Mapping

When a microsatellite was found to be polymorphic between the parents, linkage was established using DNA from the parents and a sub-set of 62 progeny. The allelic content of each genotype was determined by gel electrophoresis in the ABI PRISM 377 DNA Sequencer, and analysed with GeneScan Analysis and Genotyper software (Applied Biosystems). Segregation data were analysed using JoinMap (CPRO-DLO Wageningen), and a LOD threshhold of 3.0 for linkage.

Results

A batch of 50 possible microsatellite markers and the primers required for their amplification and testing for polymorphism and segregation could be prepared within 2–3 days of accessing the EST database. When compared with the weeks of work required to create an enriched genomic library, then to identify and sequence microsatel-

Table 2 A sample of 50 microsatellites (SSRs) which were assessed in 62 mapping population progeny. The allelic profile and information content of each marker is recorded. 0 represents a null allele. Formula for identification: female allelesx male alleles

SSR	Locus 1	Info	Locus 2	Info.	Locus 3	Info.	Locus 4	Info.
Ke199	abxcd	Fully						
Ke200	abxcd	Fully						
Ke201	abxc0	Fully	a0xb0	Fully				
Ke202	abxaa	Female						
Ke203	aaxab	Male						
Ke204	abxb0	Partly	a0×00	Female	a0×00	Female		
Ke205	abxbc	Fully						
Ke206	a0xa0	Partly						
Ke207	abx00	Female						
Ke208	abxbc	Fully						
Ke209	abxcd	Fully						
Ke210	a0xa0	Partly						
Ke211	abxcd	Fully						
Ke212	abxcc	Female						
Ke213	a0xb0	Fully	00xa0	Male	a0xa0	Partly		
Ke214	a0xbb	Female	a0×00	Female				
Ke215	Product too large							
Ke216	abxcd	Fully						
Ke217	a0xb0	Fully	aaxab	Male				
Ke218	a0×00	Female	a0×00	Female	a0xb0	Fully	a0xb0	Fully
Ke219	Non-polymorphic							
Ke220	abxcd	Fully						
Ke221	aaxab	Male	a0×00	Female				
Ke222	a0xb0	Fully						
Ke223	abxab	Partly						
Ke224	abxcd	Fully						
Ke225	aaxbc	Male						
Ke226	abxcd	Fully	a0×00	Female				
Ke227	abxbb	Female	a0×00	Female	a0xb0	Fully		
Ke228	00xa0	Male						
Ke229	No suitable PCR product							
Ke230	a0xa0	Partly						
Ke231	a0×00	Female	00xa0	Male				
Ke232	Too difficult to analyse							
Ke233	Product too large							
Ke234	abxaa	Female	aaxb0	Male	a0×00	Female	00xa0	Male
Ke235	Product too large							
Ke236	a0×00	Female	00xa0	Male				
Ke237	abxcd	Fully						
Ke238	Non-segregating							
Ke239	a0×00	Female	a0xab	Male				
Ke240	No suitable PCR product							
Ke241	Product too large							
Ke242	abx00	Female						
Ke243	Too difficult to analyse							
Ke244	abxac	Fully	a0xbb	Female	00xa0	Male	a0xa0	Partly
Ke245	a0×00	Female	00xa0	Male	00xa0	Male	a0×00	Female
Ke246	a0×00	Female						
Ke247	abxcd	Fully						
Ke248	a0×00	Female						

lite-containing clones and design primer pairs, the saving in time and cost was considerable.

Microsatellites composed of di-, tri- and tetranucleotide repeats were numerous in all the libraries used in this study. Considering repeat lengths of $\geq(XX)_5$, $(XXX)_4$ and $(XXXX)_3$, the relative proportions of each type were consistent. A comparison of two libraries, KAAA, made from vegetative tissue, and KABA, made from floral tissue, showed that of the total of 2,410 microsatellites assessed in the KAAA library, 1,551 or 64% were dinucleotide repeats, 692 or 28% were trinucleotide repeats and 167 or 6.9% were tetranucleotide repeats. In the KABA library 3,409 dinucleotide repeats represented 67.5% of the total of 5,053, while 1,308 or 25.8% were trinucleotide repeats, and 336 or 6.65% were tetranucleotide repeats.

The positions of the microsatellites in relation to the open reading frame (ORF) were also found to be constant. Dinucleotide repeats of $(XX)_5$ and longer were always 5′ and/or 3′ to the ORF, and no dinucleotide repeat longer than 8 bases was seen within the ORFs of this data set. In

Table 3 Summary of the data of the 150 EST-derived microsatellites used in this study for the evaluation of their potential for genomic mapping

Microsatellite-associated primer pairs screened	Number	Percentage
Markers amplifying more than one locus	34	22.6
Non-polymorphic loci	8	5.3
Non segregating loci	2	1.3
Too difficult to score	24	16
Segregating loci scored	155	103.3
Segregating loci mapped	138	89
Fully informative loci	42	27.1
Female informative loci	56	36.13
Male informative loci	40	25.81
Partly informative loci (not mapped)	17	10.1
Loci containing null alleles	81	51.6
CT repeats	68	45.3
GA repeats	38	25.3
AT repeats	38	25.3
GT, AC repeats	6	4

contrast, tri- and tetranucleotide repeats were found almost exclusively within the ORFs.

The data sets from the ten cDNA libraries contained different numbers of cloned ESTs, and showed variation in the percentage of ESTs containing microsatellites which fulfilled our criteria for mapping (Table 2). The summed (2.67%) and pooled (3.08%) percentages of microsatellites did not reflect the percentages recorded in the individual libraries, which showed quite wide variation, ranging from 0.55% to 4.9%. In the sample studied, microsatellites 5′ or 3′ relative to the ORF showed no appreciable difference in the level of polymorphism between the two positions.

The most frequently occurring dinucleotide repeat was CT, which made up 45.3% of the total, and together with GA (25.3%), gave a combined percentage of 70.5%. The AT motif was reasonably well represented at 25.3%. AC/GT repeats made up the remaining 4% of the repeat units. A small number (4 of the 50 primer pairs shown in Table 3) of the PCR products were considerably larger than expected from the sequence data of the ESTs. These could not be analysed as the increase in size put them beyond the range of the size standard employed. Genomic DNA was used for evaluation of the microsatellite markers, and hence was sometimes found to contain introns towards the 5′ end of the ORF. The inclusion of these introns affected the size of the PCR product.

The level of polymorphism in the EST-derived microsatellites was high (Tables 3, 4), and the quality of the markers was very good in that the allelic peaks were distinct, and generally relatively free of stutter. The amount of information derived from each microsatellite varied, since crosses between outcrossing diploid parents produce heterozygous full-sib progeny, which may have up to four different alleles at a single locus. The information content was recorded in four categories: fully informative from the cross ab×cd, segregating in the ratio of 1:1:1:1; female informative from the cross ab×cc or male informative from the cross cc×ab, both segregating in the ratio of 1:1; and partly informative from the cross ab×ab, segregating in the ratio of 1:2:1 (Table 3, Table 4). Of the 150 microsatellites sampled in this study, 116 amplified potential marker loci. From these, 89% of the polymorphic markers were suitable for genetic mapping as detailed in Table 4. A significant number of null alleles were recorded (Table 3). Of the 155 segregating loci scored, 81 (51.6%) were found to have one or more null alleles present.

The software ‘Joinmap’ gave 31 linkage groups when a LOD value of 3 was employed. The distribution of the markers appeared to be random over the linkage groups identified, but not all markers were assigned to a group. Of the 138 markers mapped, 107 were assigned to linkage groups, leaving 31 unassigned. The secondary loci which were amplified by 22.6% of the markers were often mapped to separate groups.

A large number (57%) of the genes giving rise to the ESTs were of unknown function when BLAST searches were made, but of those which did show homology to genes in the databases, many functions were represented, with no functional set of genes in our small sample being more likely to contain microsatellites than any other.

Discussion

The ESTs derived from *Actinidia* tissues, both vegetative and reproductive, gave us an excellent source of mi-

Table 4 The incidence of perfect dinucleotide microsatellites greater than 20 bp in length in the total number of ESTs sequenced from ten cDNA libraries of *Actinidia* species

Library	Species and tissue type	No. of EST seq.	No. with microsats	% containing microsats
1 KAAA	*A. deliciosa*, vegetative	9,555	194	2.03
2 KABA	*A. deliciosa*, floral	10,405	515	4.95
3 KADA	*A. deliciosa*, fruit	9,995	396	3.96
4 KSFA	*A. deliciosa*, fruit	3,442	19	0.55
5 KUBA	*A. deliciosa*, vegetative	5,325	56	1.05
6 KAFB	*A. deliciosa*, vegetative	4,598	94	2.04
7 KAIA	*A. chinensis*, fruit	7,064	289	4.1
8 KAEB	*A. deliciosa*, vegetative	1,032	14	1.36
9 KAKA	*A. deliciosa*, fruit	4,735	193	4.08
10 KALA	*A. deliciosa*, vegetative	9,310	244	2.62
				summed 2.67
Totals		65,461	2,014	pooled 3.08

crosatellites of suitable length to reveal polymorphisms between the parents and allow the construction of a useful map. The relative frequency of microsatellites in plants is considered to be lower than in animals, with the suggestion that the presence of repetitive DNA is fundamental to the development of microsatellites in both. Nadir et al. (1996) assigned a role to retrotransposons as "A-rich microsatellite generators", with the microsatellites possibly contributing through scaffold/matrix associated regions to the higher-order organisation of eukaryotic nuclei. Ramsay et al. (1999) also reported that 41% of dinucleotide microsatellites from enriched libraries showed either direct or indirect association with known repetitive elements, and considered this an underestimate due to the presence of unknown homologies. Nadir et al. (1996) acknowledged the presence of trinucleotide repeats associated with some human diseases in coding regions which have no association with retrotransposons, and gave these "unique status" within the family of microsatellite sequences.

The contrary results of Morgante et al. (2002), who reported data which showed a significant association between microsatellites and the low-copy fraction of plant genomes, could suggest an alternative origin for microsatellites. Genes and single copy regions are the most ancient portions of plant genomes since their presence pre-dates that of repetitive DNA, which is a feature of the recent genome expansion in plants. Our data from *Actinidia* species would support the view of Morgante et al. (2002) in that we recorded ranges of between 0.55 and 4.95% of microsatellite-containing clones in the cDNA libraries we sampled, with an average of 3.07% (Table 2). This figure contrasts with the 1% of microsatellite-containing clones reported in unenriched genomic libraries in *Actinidia* (Weising et al. 1996) and other plants. Microsatellites in this genus would appear to be more frequently associated with coding regions of the genome and possibly not derived from repetitive DNA. Our figures relate to dinucleotide repeats of 22 bp or longer.

It is interesting to note that the four libraries containing the highest percentage of dinucleotide microsatellites in the ESTs are those derived from reproductive tissues (floral and fruit) (Table 2). A similar trend was found in the incidence of trinucleotide microsatellites as shown in the comparison of the totals in the KAAA and KABA libraries. The trinucleotide microsatellites associated with human disease have been implicated in the onset and severity of the disease, and so have been accorded a regulatory function in gene expression. While a precise role for microsatellites in the transcribed regions of genes has not been shown, we wonder if the higher percentage of microsatellites in the reproductive tissues would give an adaptive advantage in adverse conditions, or during colonisation of new environments.

We found a high level of polymorphism between the parents of our intraspecific cross, with only 8 of the 150 microsatellites (5.3%), being non-polymorphic. This was in contrast to the results reported by Eujayl et al. (2002), who found that in wheat EST-microsatellite primers gave high quality markers devoid of stuttering, but had a low level of polymorphism (25%) when compared with genomic microsatellites (53%). Cho et al. (2000) studied the level of polymorphism in microsatellites between parental pairs of interspecific and intersubspecific crosses in rice, and also detected a higher level of polymorphism in those derived from genomic libraries (83.8%) than those from ESTs (54%). When analysing di- and trinucleotide microsatellites from grape ESTs Scott et al. (2000) found all ten functional primers tested were polymorphic across the accessions studied. They did report, however, that there were differences between microsatellites from the 3′ untranslated region (most polymorphic at the cultivar level), the 5′ untranslated region (most polymorphic between cultivars and species), and microsatellites within the coding sequence (most polymorphic between species and genera). Our microsatellites were all sourced from the 5′ and 3′ untranslated regions of the ESTs and were all dinucleotide repeats of at least 22 bp. Both wheat and rice have been subjected to selective breeding over a long period of time and this fact may have reduced the level of polymorphism in the coding regions due to selective pressure. *Actinidia* germplasm has not undergone a period of selective breeding and displays considerable phenotypic variation in its native state. A study of native American land races and wild populations of cultivated sunflower uncovered extraordinary allelic diversity in the land races and wild populations, and progressively less allelic diversity in germplasm produced by successive cycles of domestication and breeding (Tang and Knapp 2003).

We found that the most frequently occurring microsatellite was $(CT)_n$(45.2%), which together with $(GA)_n$ made up 70.54% of the microsatellites tested (Table 4). The frequency with which we found $(CT/GA)_n$ in *A. chinensis* EST libraries is in agreement with the results from similar studies in genomic libraries (Huang et al. 1998; Weising et al. 1996). These investigators conclude that in *Actinidia* (CT/GA) repeats are more common than (AC/GT) repeats. Huang et al. (1998) report they did not find any AT repeats, though AT was by far the most common type of repeat found in plants in a study carried out by Morgante and Olivieri (1993). We found AT repeats in equal numbers to GA repeats, and possibly the lack of AT repeats recorded by Huang et al. (1998) was the result of biased sampling as they suggest. In *Arabidopsis* AT repeats have been described as being typical of untranscribed regions, while CT/GA repeats are frequent in transcribed regions. It has been suggested that CT microsatellites in the 5′ UTRs of *Arabidopsis* genes are involved in their antisense transcription, and thus play a part in gene regulation (Martienssen and Colot 2001).

The presence of null alleles has been reported in many independent studies, and indicates their relative abundance (Weber et al. 1991; Callen et al. 1993; Bowcock et al. 1994). Null alleles most commonly arise from point mutations in the sequence flanking the repeat region (Lehmann et al. 1996), and are a complication in the

interpretation of microsatellite genotype data as their presence results in a reduced level of heterozygosity being observed (Callen et al. 1993). We found a high percentage (51.6%) of our loci contained null alleles, with a subsequent reduction in the information content of those markers. It is possible that had we sourced microsatellites from EST libraries derived only from *A. chinensis* tissues, the frequency of null alleles in our study may have been reduced. Some loci, while showing clear segregation of the alleles in the progeny, were unable to be scored as the allelic information did not conform to Mendelian ratios. We believe two loci were overlying each other and thus these alleles were unable to be paired.

The abundance of microsatellites in transcribed regions of the genome and the high level of polymorphism of these markers make EST libraries a valuable resource for the supply of markers for genetic mapping of *Actinidia*.

Acknowledgements This research was partially funded by the New Zealand Foundation for Research, Science and Technology (CO6X0213) and by the HortResearch Royalty Fund.

References

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

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. Cell 68:799–808

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of "null" alleles in the (AC)n microsatellite markers. Am J Hum Genet 52:922–927

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Parl WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza satica* L.). Theor Appl Genet 100:713–722

Crowhurst RN, Gardner RC (1991) A genomic-specific repeat sequence from kiwifruit *Actinidia deliciosa* var. *deliciosa*. Theor Appl Genet 81:71–78

Eujayl I, Sorrells ME, Baum M, Wolters P, Powell W (2002) Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor Appl Genet 104:399–407

Huang WG, Cipriani G, Morgante M, Testolin R (1998) Microsatellite DNA in *Actinidia chinensis*: isolation, characterisation, and homology in related species. Theor Appl Genet 97:1269–1278

Lehmann T, Hawley WA, Collins FH (1996) An evaluation of evolutionary constraints on microsatellite loci using null alleles. Genetics 144:1155–1163

Martienssen RA, Colot V (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293:1070–1074

McNeilage MA, Considine JA (1989) Chromosome studies in some *Actinidia* taxa and implications for breeding. NZ J Bot 27:71–81

Milbourne D, Russell J (1998) Comparison of molecular marker assays in inbreeding (barley) and outbreeding (potato) species. In: Molecular tools for screening biodiversity. Chapman and Hall, London

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

Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Nadir E, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellites spreading in the human genome: evolutionary mechanisms and structural implications Proc Natl Acad Sci 93:6470–6475

Ramsay L, Macaulay M, Cardle L, Morgante M, degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J 17:415–425

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana, Totowa, pp 365–386

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

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100:723–726

Tang S, Knapp SJ (2003) Microsatellites uncover extraordinary diversity in the native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003

Testolin R, Huang WG, Lain O, Messina R, Vecchione A, Cipriani G (2001) A kiwifruit (*Actinidia* spp.) linkage map based on microsatellites and integrated with AFLP markers. Theor Appl Genet 103:30-36

The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72:971–983

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genome 133:737–749

Vincent JB, Patterson AD, Strong E, Petronis A, Kennedy JL (2000) The unstable trinucleotide repeat story of major psychosis. Am J Med Genet 97:77–97

Weber JL, Polymeropoulos M, May P, Kwitek A, Xiao H, McPherson JD, Wasmuth JJ (1991) Mapping of human chromosome 5 microsatellite polymorphisms. Genomics 11:695–700

Weising K, Fung RWM, Keeling DJ, Atkinson RG, Gardner R (1996) Characterisation of microsatellites from *Actinidia chinensis*. Mol Breed 2:117–131