Theor Appl Genet (1995) 91:137–141

B. D. Dow · M. V. Ashley · H. F. Howe

Characterization of highly variable $(GA/CT)_n$ microsatellites in the bur oak, *Quercus macrocarpa*

Received: 29 August 1994 / Accepted: 22 September 1994

Abstract The objective of this study was to ascertain the usefulness of polymerase chain reaction (PCR)-based microsatellite analysis for studying pollination and parentage in a wind-pollinated temperate tree. A small insert genomic library of the bur oak (*Quercus macrocarpa*) was constructed and screened for the presence of $(CA/GT)_n$ and $(GA/CT)_n$ repeats. The proportion of positive clones yielded estimates of 3×10^5 such dinucleotide repeats per genome, roughly comparable to abundances reported in other eukaryotic genomes. Thirteen positive clones were sequenced. In contrast to mammalian genomes, the $(GA/CT)_n$ motif was more abundant than the $(CA/GT)_n$ motif in these clones. The $(GA/CT)_n$ repeats also showed longer average repeat length (mean $n = 16.2$ versus 7.3), suggesting that they are better candidates for yielding polymorphic genetic markers in oak genomes. Indeed, a survey of adult bur oaks and offspring in a small stand in northern Illinois at 3 of these $(GA/CT)_n$ microsatellite loci revealed Mendelian inheritance and extremely high levels of polymorphism, with the number of alleles at each locus ranging from 11–20 and heterozygosity ranging from 0.66 to 0.75. These results, indicating that $(GA/CT)_n$ microsatellites are both abundant and highly polymorphic in the bur oak genome, suggest that such genetic markers have tremendous potential for applications for studies of parentage, pollination and dispersal in temperate trees.

Key words Microsatellite · Oak · Polymorphism · Population biology · *Quercus*

Communicated by P. M. A. Tigerstedt

B. D. Dow (✉) · M. V. Ashley · H. F. Howe
University of Illinois at Chicago, Department of Biological Sciences, Room 3262 SES, M/C 066, 845 West Taylor, Chicago, IL 60607-7060, USA

Introduction

The usefulness of microsatellite analysis has been well-established in mammalian genomes, but the application of such analyses has been relatively rare in crop plants and almost nonexistent in natural plant populations. Database searches have revealed that microsatellites exist in many plant genomes (Lagercrantz et al. 1993; Wang et al. 1994) and have suggested that microsatellites may be as useful to plant genetics as they have been to animal genetics. Although microsatellites are often used in genome mapping, our purpose was to develop primers for variable microsatellite loci to address questions of parentage, gene flow and genetic diversity in a species of temperate deciduous forest trees.

Research on microsatellites in agricultural plant species such as soybean (Akkaya et al. 1992; Morgante and Olivieri 1993), sunflower (Brunel 1994) and rapeseed (Lagercrantz et al. 1993) suggests that the relative abundance of motifs may differ dramatically from that observed in animal genomes. In database searches of plants (Akkaya et al. 1992; Lagercrantz et al. 1993; Wang et al. 1994) the most common long dinucleotide repeat sequence was found to be $(AT/TA)_n$ followed by the relatively rare $(GA/CT)_n$. In contrast, $(CA/GT)_n$ sequences are the most common in mammals. Microsatellite surveys of natural populations of higher plants have not been reported. The only information on woody species comes from Condit and Hubbell (1991), in which DNA libraries of five tropical tree species were screened for $(CA/GT)_n$ and $(GA/CT)_n$ motifs. They found these repeats to be relatively abundant in higher plants and estimated that there were between 10^4 and 10^5 such sites in the genomes of the species examined. However, the only sequence information presented was for five relatively short $(CA/GT)_n$ microsatellites, unlikely to be highly polymorphic.

The objective of our study was to explore the potential of microsatellite markers to study parentage and pollen dispersal in a wind-pollinated temperate decidu-

ous tree, the bur oak (*Quercus macrocarpa*). Traditional indirect methods of evaluating pollen movement with pollen traps, pollen dyes and observation of pollinators (e.g. Levin and Kerster 1969; Mosquin 1971; Linhart et al. 1987; Handel 1983; Greenwood 1986) do not allow an assessment of the actual parentage of seeds and plants. Similarly, protein electrophoresis permits evaluation of similarities and differences of allozyme frequencies in populations (Mitton et al. 1979; Meagher and Thompson 1987; Wolf et al. 1990), but polymorphism is usually too limited to determine parentage with certainty (Chakraborty et al. 1988). Attempts to use standard multi-locus fingerprinting failed to provide the necessary resolution for definitively determining parentage at our study site, a small bur oak population in northern Illinois that is a disturbed fragment of pre-European settlement savanna. To assess the potential of microsatellite analysis for our study, we screened a library of bur oak genomic DNA for $(CA/GT)_n$ and $(GA/CT)_n$ repeats, sequenced a number of these repeats as well as their flanking regions and examined variability at three loci using the polymerase chain reaction (PCR). Our goals were to use these data to (1) estimate the relative abundance of microsatellites in the bur oak genome; (2) determine the sequences and lengths of several representatives of both types of repeats; and (3) assess their pattern of inheritance and level of variability and thus their relative potential for providing polymorphic genetic markers.

Materials and methods

Library preparation and screening

Genomic DNA was isolated from leaves of bur oak (*Quercus macrocarpa*) and purified on cesium chloride gradients following the protocol of Keim et al. (1989) except that 0.125 volume of 5 *M* potassium acetate was added to each tube and the mixture cooled for 30 min on ice before centrifugation to remove debris. For construction of a bur oak library, the general protocol of Rassmann et al. (1991) was followed. Genomic oak DNA was triple digested with *Alu*I, *Hae*III and *Rsa*I. Fragments of 300–500 bp were size-fractionated twice on 1% agarose gels and recovered using Spin-X^{TM} centrifuge filter units (Costar) and ethanol precipitation. Fragment ends were filled in by incubating with Klenow (Promega) and dNTP's, extracted with phenol/chloroform/isoamyl alcohol and ethanol-precipitated. M13mp19RF DNA was digested with *Sma*I and dephosphorylated with bacterial alkaline phosphotase (BAP, Gibco BRL Life Technologies). About 50 ng of insert DNA was ligated into 200 ng M13 for 1 h at room-temperature, transformed into competent cells (JM101) and plated with Bluo-gal and IPTG.

Plaque lifts were made using nylon membranes (Nytran^{TM} Nucleic Acid and Protein Transfer Media, Schleicher and Schuell). Two simple-sequence oligonucleotides, poly (CA/GT) and poly (GA/CT) (Pharmacia LKB Biotechnology), were labelled with α-[^{32}P]-dCTP by nick translation and simultaneously hybridized to the plaque-lift membranes overnight at 65 °C. The membranes were washed for 15 min in 2 × SSC and 1% SDS at room temperature and 15 min in 2 × SSC and 0.1% SDS at 45 °C. Membranes were exposed to Kodak XAR-5 film for 1–3 days at −80 °C with intensifying screens. Single-stranded DNA was extracted from selected positive plaques and used for sequencing (Sequenase 2.0, U.S.B.). Sequencing reaction products were separated on denaturing acrylamide gels containing 5% Long Ranger (J.T. Baker), 8 *M* urea, and 1.2 × TBE for 2–5 h. Gels were dried and exposed to Kodak XAR-5 film for approximately 16 h.

Leaf collection and quick DNA extraction

Young leaves (0.5–2.0 cm long) of 61 adult bur oaks were collected in the spring when tannin content was lowest (Feeny 1970), quick-frozen and stored at −70 °C. Acorns were collected from several trees, stratified (USDA 1974) and grown in a greenhouse. Fully expanded leaves were quick-frozen and stored as above. To avoid cross-contamination, single-use sterile "pestles" for grinding leaves were made by melting the points of blue pipet tips in a flame and autoclaving them. Leaves (4–5 from one emerged bud, about 0.15 g) were placed in a sterile 1.5-ml microcentrifuge tube and ground in liquid nitrogen. Aliquots of 800 µl of extraction buffer (Keim et al. 1989) at 65 °C and 10 µl of 20 mg/ml Proteinase K were added immediately to the ground tissue, and the tubes were incubated at 65 °C for 30 min. After incubation 100 µl of 5 *M* potassium acetate was added, and the tubes were placed on ice for 30 min. Debris was pelleted by centrifugation for 10 min. The supernatant was placed in a new tube, and 0.54 volume isopropanol at −20 °C was added. The tubes were chilled at −70 °C for 10 min and centrifuged again for 10 min to pellet the DNA. The pellet was washed in 70% ethanol and vacuum-dried.

The DNA was resuspended in TE (10 m*M* Tris pH 8.0, 1 m*M* EDTA) and the proteins removed by two extractions with phenol/chloroform/isoamyl alcohol. Following the last phenol extraction, the DNA was precipitated with ethanol and resuspended in water. This process yielded about 700 µg DNA from 0.15 g leaf tissue.

PCR amplification

Primer pairs were identified from flanking regions of microsatellites with the aid of Mac Vector 4.1 software (I.B.I.). Primers were made for two simple repeats, MSQ3 and MSQ4, and two compound repeats, MSQ11 and MSQ13. Amplification was performed in a total volume of 10 µl, which included 20 µ*M* dNTP's, 50 n*M* of each primer, 0.2 µl *Taq* polymerase, and 1 microcurie [^{35}S]-dATP. Amplification proceeded for 35 cycles of 92 °C for 1 min, 50 °C for 1 min and 72 °C for 30 s. PCR products were separated on 6% standard denaturing acrylamide gels. Gels were dried and exposed to Kodak XAR-5 film for 3–5 days.

Results

Abundance of microsatellites in the bur oak genome

A total of 1547 white (recombinant) plaques resulted from cloning short genomic inserts into M13. Of these, 39 (2.5%) hybridized to the simple-sequence probes. The number of microsatellites in the bur oak genome was roughly estimated following Condit and Hubbell (1991), using 400 bp as the average size of our inserted DNA and 5×10^9 as the approximate genome size. This calculation showed that the bur oak genome has about 3×10^5 $(CA/GT)_n$ and $(GA/CT)_n$ regions, or 1 every 16 kb if randomly distributed.

Characterization of bur oak repeats

Sequences of the genomic insert for 13 positive plaques were obtained, and each contained 1 or more microsatellite sequences. Table 1 shows the sequence and lengths of the microsatellites. Using the classification

Table 1 Length and type of microsatellites from 13 sequenced fragments of bur oak DNA. All microsatellites are perfect (no substitutions) unless otherwise indicated

	Number of repeat units (n) for each motif				
	$(GA/CT)_n$	$(CA/GT)_n$	$(TA/AT)_n$	$(AAT/TTA)_n$	$(A/T)_n$
Simple repeats					
MSQ1	22				
MSQ2	15				
MSQ3[a]	18				
MSQ4[a]	17				
MSQ5	14				
MSQ6	16				
Compound repeats					
MSQ7	23			8	9
MSQ8	12, 13, 15	7			
MSQ9	17	6			
MSQ10		9[b]	7		
MSQ11[a]	15		10		
MSQ12	12		10		
MSQ13[a]	14, 11				
Mean number of repeats	15.6	7.3	9	8	9
Total number of microsatellites	15	3	3	1	1

[a] These loci were used to make primers (see Table 2)
[b] This microsatellite had one substitution

scheme of Weber (1990), we determined that 6 of the inserts contained single "perfect" repeats (with no interruptions of the core motif) and 7 were "compound" repeats (with different motifs located in close proximity), yielding a total of 21 dinucleotide repeats. Three of these were $(TA/AT)_n$ repeats (for which we had not screened) that were part of compound repeats. Of the remaining 18, 15 were $(GA/CT)_n$ (83%). The average length of $(GA/CT)_n$ repeats was 16.2, substantially higher than the average number of either the $(AT/TA)_n$ repeats (mean $n = 9.0$) or $(GT/CA)_n$ repeats (mean $n = 7.3$).

The entire sequence for 8 genomic insets containing microsatellites was obtained (data not shown). The G+C content for these inserts was relatively low, approximately 35%. In a number of cases, the composition of the flanking sequences made it difficult to design PCR primers. PCR primers were found for 3 microsatellites, MSQ3, MSQ4 and MSQ13, which successfully amplified the target microsatellite region from genomic DNA of adult bur oaks in the stand. Information regarding these loci is given in Table 2. The fourth pair of primers, for MSQ11, produced many artifact bands and was therefore uninterpretable.

Figure 1 shows the results of PCR amplification of genomic DNA isolated from 13 adult bur oaks in our study stand using primers for MSQ4. A total of 11 alleles could be scored at this locus, with product lengths ranging from 203 to 227 bp (10–22 repeat units). High levels of variability were observed at the other microsatellite loci as well (Table 2), with the levels of heterozygosity at these 3 loci ranging from 0.66 to 0.75.

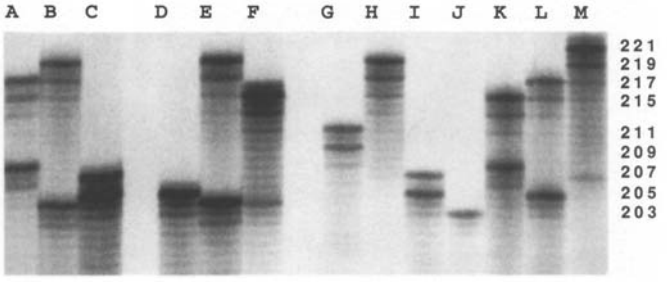

Fig. 1 Autoradiograph of PCR-amplified microsatellites of bur oak (*Quercus macrocarpa*). Microsatellites of 9 individuals were amplified with primer pair MSQ4 (Table 2). Interpretation of genotypes is as follows: *A* 217/207, *B* 219/203, *C* 207/205, *D* 205/205, *E* 219/203, *F* 215/215, *G* 211/209, *H* 219/217, *I* 207/205, *J* 203/203, *K* 215/207, *L* 217/205, *M* 221/221. PCR product length (bp) is indicated at the *right*. The lighter bands are the result of "stutter" thought to be due to PCR slippage in the microsatellite repeat region

Table 2 Summary of microsatellite primers sequences, number of alleles per locus, product sizes and heterozygosity among 61 adults in a small stand of bur oaks

Locus and microsatellite sequence	Forward primer	Backward primer	Number of alleles	Range of PCR product sizes (nt)	Heterozygosity
MSQ3 $(GA)_n$	CCCTTCTGACATTGCATATTCGA	CCAATTCGACAATTTCTTAGTGCA	20	191–231	0.66
MAQ4 $(AG)_n$	TCTCCTCTCCCCATAAACAGG	GTTCCTCTATCCAATCAGTAGTGAG	11	203–227	0.75
MSQ13 $(TC)_n$	TGGCTGCACCTATGGCTCTTAG	ACACTCAGACCCACCATTTTTCC	12	222–246	0.74

Preliminary screening of 96 offspring of 1 tree using primers for MSQ4 and MSQ13 showed that all offspring shared at least 1 allele with the maternal tree. Furthermore, the 2 maternal alleles were divided evenly among the offspring, indicating Mendelian inheritance.

Discussion

We report here the first application of PCR-based microsatellite analysis to a natural population of woody plants. Our preliminary results demonstrate that this approach holds excellent promise for studies of relatedness and genetic diversity that previously have been difficult in such systems. Furthermore, several considerations emerge with regards to the experimental design of microsatellite studies of trees.

Our estimate for the abundance of microsatellite repeats in bur oaks is similar to Condit and Hubbell's (1991) estimate of 10^4–10^5 simple sequence repeats in five species of tropical trees and comparable to dinucleotide repeat frequencies reported in other higher eukaryotes (Weber and May 1989; Tautz et al. 1986) and some plants (Wang et al. 1994). The number of microsatellites found in bur oaks contradicts the results of Lagercrantz et al. (1993), who report that microsatellites are 5 times less abundant in the genomes of plants than in mammals. The abundance reported here should provide ample microsatellite loci for population studies. An important result not available from previous work on trees is that the $(GA/CT)_n$ motif appears to be more abundant than the $(CA/GT)_n$ motif, and also to have longer average repeat length (mean $n = 16.2$ versus 7.3). Studies in humans suggest that loci with repeat lengths greater than 12–15 are much more likely to be variable and therefore useful as genetic markers (Weber 1990). Our results suggest that $(GA/CT)_n$ repeats hold much greater promise for useful genetic markers than the $(CA/GT)_n$ that have been the focus of mammalian studies.

The survey of adult bur oaks in this small stand provides strong evidence that higher plants show levels of microsatellite variability at least as high as that reported in other eukaryotes. The number of alleles at each of the 3 loci ranged from 11 to 20, and heterozygosities ranged from 0.66 to 0.75 (Table 2). With these levels of polymorphism, data from relatively few such loci will allow determination of parentage of seedlings and saplings at our study site. Furthermore, our results point to the general usefulness of $(GA/CT)_n$ microsatellite repeats for a variety of plant studies requiring polymorphic genetic markers.

In addition to the increased abundance of $(GA/CT)_n$ markers in the oak genome relative to mammalian genomes, two other findings also appear to reflect fundamental differences in the organization of mammalian and higher plant genomes. First, we observed low G + C content of the genomic inserts containing microsatellites, something which has not been noted in mammal studies. Although the G + C content of oak genomes has not been reported, if the low values seen in our inserts (35%) are reflective of the entire genome, bur oaks would fall at the low end of the range of G + C values reported for other higher plants (Smillie and Scott 1969; Wang et al. 1994). Secondly, the presence of 7 compound repeats in a sample of 13 positive clones (54%) suggest that compound repeats may be proportionally higher in plant genomes. In contrast, Weber (1990) found only 12 compound repeats in a survey of 112 human microsatellites (11%). Although some authors have suggested that compound repeats may be less variable than single repeats (Rassmann et al. 1991), the 1 compound locus assayed here (MSQ13) proved to be highly variable (Table 2). Primers for another compound repeat (MSQ11) produced so many artifact bands that the allele lengths could not be unambiguously scored. This difference may be due to the fact that the 2 microsatellite regions in MSQ13 were separated by 12 nonrepeating base pairs, while those in MSQ11 were adjacent, which may cause greater slippage in the PCR amplification. These results suggest that closely spaced compound repeats may prove useful for natural plant studies, while adjacent compound repeats should be avoided.

Acknowledgements We thank Lon Kaufman for the use of his laboratory for portions of this project. This research was supported by a Campus Research Board grant to Henry F. Howe from the University of Illinois at Chicago.

References

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

Brunel D (1994) A microsatellite marker in *Helianthus annuus* L. Plant Mol Biol 24:397–400

Chakraborty R, Meagher TR, Smouse PE (1988) Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118:527–536

Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:314–331

Feeny P (1970) Seasonal changes in oak leaf tannins and nutrients as a cause of spring feeding by winter moth caterpillars. Ecology 51:565–581

Greenwood MS (1986) Gene exchange in loblolly pine: The relation between pollination mechanism, female receptivity and pollen availability. Am J Bot 73:1443–1451

Handel SN (1983) Pollination ecology, population structure and gene flow. In: Real L (ed) Pollination biology. Academic Press, New York, pp 163–211

Keim P, Paige KN, Whitham TG, Lark KG (1989) Genetic analysis of an interspecific hybrid swarm of populus: occurrence of unidirectional introgression. Genetics 123:557–565

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

Levin DA, Kerster HW (1969) The dependence of bee-mediated pollen and gene dispersal upon plant diversity. Evolution 23:560–571

Linhart YB, Busby WH, Beach JH, Feinsinger P (1987) Forager behavior, pollen dispersal and inbreeding in two species of hummingbird-pollinated plants. Evolution 41:679–682

Meagher TR, Thompson E (1987) Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). Ecology 68:803–812

Mitton JB, Linhart YB, Sturgeon KB, Hamrick JL (1979) Allozyme polymorphisms detected in mature needle tissues of ponderosa pine. J Hered 70:86–89

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

Mosquin T (1971) Competition for pollinators as a stimulus for the evolution of flowering time. Oikos 22:398–402

Rassmann K, Schlötterer C, Tautz D (1991) Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. Electrophoresis 12:113–118

Smillie RM, Scott NS (1969) Organelle biosynthesis: The chloroplast. In: Hahn FE (ed) Progress in molecular and subcellular biology, vol. 1. Springer Berlin Heidelberg New York, pp 136–202

Tautz D, Trick M, Dover GA (1986) Cryptic simplicity in DNA is a major source of genetic variation. Nature 322:652–656

USDA Forest Service (1974) Seeds of woody plants in the United States. U.S. Department of Agriculture, Agricultural Handbook No. 450

Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem 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

Wolf PG, Soltis DE, Soltis PS (1990) Chloroplast DNA and allozymic variation in diploid and autotetraploid *Heuchera grossulariifolia* (Saxifragaceae). Am J Bot 77:232–244