Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers

Geoffrey C. Waldbieser, Sylvie M.A. Quiniou, and Attila Karsi
United States Department of Agriculture, Agricultural Research Service,
Stoneville, MS, USA

BioTechniques 35:976-979 (November 2003)

We developed a technique to improve the efficiency of producing TAA repeat microsatellite markers linked to interspecific conserved genes. Template DNA was prepared from cultures derived from single bacterial artificial chromosome (BAC) colonies using a simple alkaline lysis miniprep. The presence of conserved genes in each BAC clone was verified by sequencing with gene-specific primers. The BAC templates were directly sequenced using short tandem repeat-anchored primers (STRAPs), consisting of TAA repeats with one or two unique 3' terminal bases. At least one STRAP provided sufficient 3' flanking sequence from each clone for the design of a BAC-specific primer. The BAC-specific primer was used to sequence back through the tandem repeat and obtain 5' flanking sequence, and a second BAC-specific primer was designed for microsatellite genotype analysis. This technique quickly provided microsatellite markers with an average of 15 tandem repeats for the BAC clones tested. The identification of polymorphic microsatellite loci in these clones permits the identification of alleles linked to candidate genes, placement of conserved genes on genetic linkage maps, and integration of linkage and physical maps.

INTRODUCTION

Type I genetic markers are polymorphic loci within or near genes conserved between species. Genetic maps containing type I markers allow researchers to identify regions of chromosomal synteny between species, integrate physical and genetic linkage maps, and identify potential candidate genes for marker-assisted selection. Two classes of polymorphic loci commonly used are single nucleotide polymorphisms (SNPs) and short tandem repeats (microsatellites). SNPs can be identified within intronic or untranslated regions of genes. However, several SNPs are necessary at one locus to identify multiple allelic haplotypes in a population. Alternately, one can identify multiple alleles in a population at one microsatellite locus due to the variation in the number of tandem repeats. Microsatellites containing tandem TAA repeats have been useful markers in several species (1–3) but may be underutilized in other species

because the low annealing temperature of TAA probes hinders the efficient identification of TAA-containing clones by hybridization screening.

The presence of large sizes of genomic DNA (70–300 kb) in bacterial artificial chromosome (BAC) clones increases the potential for the identification of microsatellite loci relatively close to genes of interest. However, microsatellite identification through the primer walking or sequencing of subcloned DNA can be labor-intensive. Recently, microsatellite sequences were identified by the direct sequencing of BAC clones with tandem repeat primers containing 3' degenerate bases (4). The method works by annealing the anchored tandem repeat primer to the 3' end of a microsatellite locus, and sequence is obtained from the 3' flanking region. A gene-specific primer is then designed to sequence the complementary strand back across the tandem repeat and into the 5' flanking DNA (Figure 1). However, this method was untested for TAA repeats and required a miniprep

from large BAC cultures to provide 1 µg template DNA per sequencing reaction. Here we report a simple BAC miniprep protocol and the production of microsatellite markers containing TAA repeats by the direct sequencing of BAC clones with short tandem repeat-anchored primers (STRAPs).

MATERIALS AND METHODS

Channel catfish BAC clones in pBeloBAC11 vector containing conserved genes were obtained by PCR screening of pooled clones from an arrayed library (5). Individual clones isolated from streaked agar plates were grown in a shaking incubator for 22–24 h at 37°C at 250 rpm in 17 × 100 mm Falcon® round-bottom polypropylene tubes (BD Biosciences, San Jose, CA, USA) containing 5 mL LB medium and 12.5 µg/mL chloramphenicol. All miniprep solutions (Qiagen, Valencia, CA, USA) were kept at room temperature. After centrifugation at 1900× g for 10 min and removal of the supernatant, the pellet was resuspended in 250 µL solution P1 containing 100 µg/mL RNase A (Qiagen) and transferred to a 1.5-mL microcentrifuge tube. Solution P2 (250 µL) was added and mixed slowly by inversion. Within 5 min, 350 µL solution P3 were added and mixed by inversion, and the tube was centrifuged at 11,500× g for 10 min. Eight hundred microliters of supernatant were transferred to another microcentrifuge tube containing 560 µL isopropanol and vortex mixed. Precipitated BAC DNA was pelleted by a 30-min centrifugation at 11,500× g. After a 200-µL 70% ethanol wash, the pellet was dried for 10–15 min and resuspended in 50 µL deionized water. The yield was estimated by the comparison of ethidium-stained BAC DNA with λ DNA standards on agarose gels. The typical BAC DNA yield was 1–2 µg per 5 mL culture.

The STRAPs were designed with 10 tandem repeats followed by 1–3 bases at the 3' end that interrupted the tandem repeat (Table 1). Reverse complement BAC-specific primers were designed from the flanking DNA of STRAP-primed sequences using GeneRunner (Hastings Software, Hastings-on-Hudson, NY, USA). All primers were obtained from Integrated DNA

Table 1. Read length of Sequences Immediately Downstream of TAA Repeats Using STRAP-Based Sequencing of Catfish MHC Class II Bacterial Artificial Chromosome (BAC) Clones

STRAP Name	Primer Sequence and 3' Anchor	Templates ^a	Repeats ^b (n)	Sequence Read Length ^c (bases)
TAA _n A	(TAA) ₁₀ A	3	14, 14, 17	0
TAA _n C	(TAA) ₁₀ C	0	—	0
TAA _n G	(TAA) ₁₀ G	0	—	0
TAA _n TAC	(ATA) ₁₀ C	0	—	0
TAA _n TAG	(ATA) ₁₀ G	1	15	544
TAA _n TAT	(ATA) ₁₀ T	0	—	0
TAA _n TC	(AAT) ₁₀ C	0	—	0
TAA _n TG	(AAT) ₁₀ G	1	14	487
TAA _n TT	(AAT) ₁₀ T	4	10, 15, 17, 22	0

STRAP, short tandem repeat-anchored primer; MHC, major histocompatibility complex.

^aNumber of potential primer binding sites in the MHC class II BAC clone.

^bLength of the tandem repeat motif at each template.

^cPhred40 quality bases.

Figure 1. Model for the determination of DNA sequence on both flanks of short tandem repeats by sequencing with short tandem repeat-anchored primers (STRAPs). (1) A bacterial artificial chromosome (BAC) clone containing TAA repeats (in bold) is denatured. (2) The A (bold and italic) on the 3' end of the (TAA)₁₀A STRAP forces specific annealing at the 3' end of this TAA repeat, and primer extension provides downstream flanking sequence. (3) A BAC-specific primer (BSP1) designed from the downstream sequence is used for sequencing through the repeat into the upstream flanking region. (4) A second BAC-specific primer (BSP2) is designed from the upstream flanking sequence, and the microsatellite locus is PCR-amplified using the BSP1 and BSP2 primers.

Technologies (Coralville, IA, USA) and resuspended in deionized water. The sequencing reactions contained 1.5 µL ABI PRISM® BigDye™ 3.0 or 3.1 Terminator premixture (Applied Biosystems, Foster City, CA, USA), 2.25 µL sequencing buffer (250 mM Tris-Cl, pH 9.0, 10 mM MgCl₂), 5 pmol primer, and 100–200 ng BAC miniprep. Cycle sequencing was performed in a PTC-200™ DNA Engine thermal cycler (MJ Research, Waltham, MA, USA) with an initial denaturation of 3 min at 95°C, followed by 99 cycles of 95° for 30 s, 50° for 30 s, and 60°C for 4 min. Sequencing products were precipitated using sodium acetate/ethanol (Big Dye 3.0) or EDTA/ethanol (Big Dye 3.1), according to manufacturer's instructions. The sequences were resolved on an ABI PRISM 3100 Avant Genetic Analyzer (Applied Biosystems) with 1.0-kV injection voltage, 10-s injection time (if the peaks were too low, then the sample was reinjected for 22 s), POP4 polymer, and 50-cm capillary array. Chromatograms were visualized with Trev version 7 (6). We determined the sequence quality with Phred version 0.000925.c (7,8) using a quality score cut-off of 40.

One gene-specific primer flanking the microsatellite repeat was 5' tailed with the sequence 5'-GAGTTTTTC-CCAGTCACGAC-3' (long) to permit product labeling, and the other gene-specific primer was 5' tailed with the sequence 5'-GTTT-3' to promote non-template adenylation (9). Amplification reactions also included the primer 5'-GAGTTTTTCCAGTCACGAC-3', which was 5' labeled with 6-FAM, VIC, NED, or PET (Applied Biosystems). Amplification reactions included 0.5 µL HotStarTaq® DNA polymerase and 1.5 µL 10× buffer (Qiagen), 67 µM each dNTP, 1 pmol fluorescence-labeled primer, 2 pmol long primer, 6 pmol short primer, and 50 ng genomic DNA. The products were amplified in a PTC-200 thermal cycler at 95° for 3 min, and then 2 cycles of 95° for 1 min and 60° for 1 min, 27 cycles of 95° for 30 s and 60° for 30 s, and a final extension of 72°C for 4 min. Genotypes were resolved by capillary electrophoresis on an ABI PRISM 3100 Genetic Analyzer, and fragment sizes were determined with GeneMapper™ software (Applied Biosystems).

Table 2. Polymorphic Microsatellite Markers Linked to Interspecific Conserved Genes via STRAP Sequencing of Bacterial Artificial Chromosome (BAC) Clones

Gene	BAC Clone	STRAP	Repeat Motif in Clone	Other Repeats in Flanking Sequence	Polymorphic
<i>Adora</i>	9d14	(TAA) ₁₀ A	(TAA) ₁₃		+
<i>Fst</i>	33d19	(TAA) ₁₀ C	(TAA/TAC) ^a		n.d.
	33d19	(TAA) ₁₀ TG	(TAA) ₁₅	(TAA) ₁₈	+
<i>Hsp70</i>	13j21	(TAA) ₁₀ TAG	(TAA) ₁₇	(TA) ₃₂ , (TA) ₁₆	+
	13j21	(TAA) ₁₀ G	(TAA) ₁₅	(TA) ₃₆	n.d.
	13j21	(TAA) ₁₀ AA	(TAA) ₂₃		n.d.
	13j21	(TAA) ₁₀ AT	(TAA) ₂₂		n.d.
<i>Mx</i>	4p07	(TAA) ₁₀ A	(TAA) ₁₇		+
	4p07	(TAA) ₁₀ G	(TAA) ₁₂	(AC) ₁₂	+
	118a11	(TAA) ₁₀ A	(TAA) ₁₁		n.d.
<i>NK-Lysin</i>	9p14	(TAA) ₁₀ C	(TAA) ₁₃	(TAA) ₆	+
<i>TCRα</i>	Tcra	(TAA) ₁₀ G	(TAC) ₈ (TAA) ₁₄		+
<i>TNFα</i>	3h20	(TAA) ₁₀ C	(TAA) ₇	(TG) ₂₂ , (ATTT) ₁₀	+

STRAP, short tandem repeat-anchored primer; n.d., not determined.
^aCompound repeat > 400 bp.

dinucleotide and trinucleotide microsatellites in the catfish BAC clones prohibited the use of degenerate STRAPs. The MHC class II BAC clone sequence revealed from 4 to 17 potential templates for each degenerate dinucleotide STRAP that has been previously described (4). Although this may not be typical of catfish BAC clones, there were no unique templates for nondegenerate dinucleotide STRAPs with one-base extensions. Furthermore, sequencing reactions that contained equimolar amounts of (TAA)₁₀A, (TAA)₁₀C, and (TAA)₁₀G rarely produced useful sequence (data not shown). Other catfish BAC clones containing conserved genes were screened with TAA repeat STRAPs. In each case, at least one polymorphic TAA repeat was identified (Table 2). A (TAA)₅ template in the MHC class II

Table 3. Length of STRAP-Primed Sequencing Products for Two Short Tandem Repeat Templates in Bacterial Artificial Chromosome (BAC) Clone 13j21

Template	(ATT) ₂₃ TT			(ATT) ₂₂ TA
STRAP	(TAA) ₁₀ AA	(TAA) ₁₀ AC	(TAA) ₁₀ AG	(TAA) ₁₀ AT
T _{anneal}				
45.0	0	102	449	280
45.5	21	21	555	604
46.5	25	174	419	22
48.2	97	0	505	484
50.5	333	118	354	324
53.4	428	0	0	465
56.7	37	0	0	0
59.6	0	0	0	0

Lengths are shown in base pairs.

RESULTS AND DISCUSSION

A search of the complete sequence of a channel catfish BAC clone containing the major histocompatibility complex (MHC) class II α and β genes (unpublished data) revealed 9 loci with 10 or more tandem TAA repeats (Table 1). Only two loci contained unique anchor sequences, TAG and TG, and these loci were correctly identified in the flanking sequence obtained from sequencing reactions primed by the correspond-

ing STRAP. Three loci contained an A anchor, and four loci contained a TT anchor, and the sequences obtained with the corresponding STRAP primers were ambiguous. While the Phred 40 sequence was as long as 544 bases, some runs were manually stopped because a large amount of flanking DNA was unnecessary for the design of the first BAC-specific primer.

Previous research described the use of degenerate bases in the anchored portion of the STRAPs (4), but the frequency of

BAC clone was not utilized by any TAA STRAP; however, a (TAA)₇ template was utilized in the tumor necrosis factor α (TNFα) clone. One repeat in clone 33d19 was longer than 400 bp, but on average, the TAA templates contained 15 repeats. The selection for longer loci likely enhanced the probability of marker polymorphism.

Sequencing experiments revealed that primer extension from multiple templates in a clone could be resolved by the addition of two anchored bases to the primer. Multiple overlapped sequences were obtained using the (TAA)₁₀A STRAP in clone 13j21 because there were two ATTNt templates (Table 3), but the addition of a second anchored base allowed for the discrimination of the two unique 3'-(ATT)₂₃TT-5' and 3'-(ATT)₂₂TA-5' templates in this clone. The relatively low annealing temperatures of cycle sequencing permitted mispriming at the 3' end of the TAA STRAPs. The 3'-(ATT)₂₃TT-5' template was utilized by TAA STRAPs containing AA, AC, and AG anchors (Table 3) at 50°C. Mispriming was inhibited at 53.4°C, and no efficient priming was observed above that temperature. However, mispriming was not a problem

in this case because each of the three primers provided the same high-quality 3' flanking sequence.

STRAPs should be designed according to the abundance of repeat sequence motifs in each genome. Loci with TAA repeats were abundant in channel catfish (2), but TAA STRAPs were not successful in BAC clones of rainbow trout and swine. Dinucleotide STRAPs were required to produce polymorphic markers from BAC clones of these species (data not shown). Dinucleotide STRAPs were not analyzed in catfish clones because of the large number of potential dinucleotide STRAPs required for screening a BAC, our prior success with TAA STRAPs, and the low level of band stutter in genotyping data obtained from trinucleotide repeat markers.

Here we describe the rapid, efficient production of polymorphic microsatellite markers from BAC clones by direct sequencing from a simple miniprep. This is an efficient method for developing polymorphic markers linked to conserved genes, and sometimes other microsatellite sequences were discovered in the initial 3' flanking sequence (Table 2). STRAP sequencing a BAC template produced from a sequence-verified single colony is crucial to provide the linkage of the microsatellite marker with the conserved gene. This method overcomes potential difficulties in screening subcloned libraries by colony hybridization with TAA repeat primers at low annealing temperatures (1). The production of markers based on other trinucleotide or tetranucleotide repeats should be possible depending on their abundance in each species and may require higher annealing temperatures in the sequencing reaction to overcome false priming. These markers can be used to investigate allelic variation at candidate genes, integrate physical and genetic maps by identifying BAC clone locations on genetic linkage maps, and support comparative mapping by the addition of conserved genes to genetic linkage maps.

ACKNOWLEDGMENTS

The authors thank Debra Harris for technical assistance, Xiaofan Liu and Mary Duke of the USDA, ARS, Mid-South Area Genomics Laboratory for

assistance with BAC sequencing, Dan Nonneman for swine BAC clones, and Scott Gahr and Caird Rexroad III for trout BAC clones. We also thank Lanie Bilodeau and Jeff Ray for their critical review of the manuscript. Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the U.S. Department of Agriculture and does not imply approval to the exclusion of other products that may be suitable.

REFERENCES

1. Cregan, P.B., J. Mudge, E.W. Fickus, L.F. Marek, D. Danesh, R. Denny, R.C. Shoemaker, B.F. Matthews, et al. 1999. Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theor. Appl. Genet.* 98:919-928.
2. Waldbieser, G.C., B.G. Bosworth, D.J. Nonneman, and W.R. Wolters. 2001. A microsatellite-based genetic linkage map of the channel catfish, *Ictalurus punctatus*, genome. *Genetics* 158:727-734.
3. Song, Q.J., E.W. Fickus, and P.B. Cregan. 2002. Characterization of trinucleotide SSR motifs in wheat. *Theor. Appl. Genet.* 104:286-293.
4. Fujishima-Kanaya, N., D. Toki, K. Suzuki, T. Sawazaki, H. Hiraiwa, M. Iida, T. Hayashi, H. Uenishi, et al. 2003. Development of 50 gene-associated microsatellite markers using BAC clones and the construction of a linkage map of swine chromosome 4. *Anim. Genet.* 34: 135-141.
5. Quiniou, S.M.A., T. Katagiri, N.W. Miller, M. Wilson, W.R. Wolters, and G.C. Waldbieser. 2003. Construction and characterization of a BAC library from a gynogenetic channel catfish, *Ictalurus punctatus*. *Genet. Sel. Evol.* 35:1-11.
6. Bonfield, J.K., K.F. Beal, M.J. Betts, and R. Staden. 2002. TREV: a DNA trace editor and viewer. *Bioinformatics* 18:194-195.
7. Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175-185.
8. Ewing, B. and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8:186-194.
9. Brownstein, M.J., J.D. Carpten, and J.R. Smith. 1996. Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* 20:1004-1010.

Received 9 July 2003; accepted 26 August 2003.

Address correspondence to Geoffrey C. Waldbieser, USDA, ARS, Catfish Genetics Research Unit, P.O. Box 38, Stoneville, MS 38776, USA. e-mail: gwaldbieser@ars.usda.gov