APPLICATION OF MOLECULAR SELECTION TO DNA FRAGMENTS THAT CONTAIN MICROSATELLITES

R. M. Thallman, J. F. Taylor, C. R. Young and S. K. Davis
Department of Animal Science
Texas A&M University
College Station, TX 77843, USA

SUMMARY

The theoretical framework for a different approach to the detection of genetic polymorphisms is presented. A model system utilizing microsatellites is described and several of its fundamental operations are demonstrated.

INTRODUCTION

The use of DNA markers for genetic improvement in livestock is dependent on the ability to detect polymorphisms either at or closely linked to genes that affect economically important traits. Two main approaches to the identification of markers related to quantitative trait loci have been proposed. One approach is to evaluate the effect of candidate genes of known physiological function on a trait of interest. The other approach is to saturate the genome with markers and evaluate the effect of each marker on the trait.

A fundamentally different approach, referred to as "Molecular Selection" was proposed by Thallman et al. (1993). Molecular Selection begins by cutting the entire genome into small fragments with a restriction enzyme. Those fragments containing polymorphisms that cosegregate with a particular phenotypic trait or DNA marker are selected, while the remaining fragments are progressively diluted out of the population. Consequently, it should be quite useful for the generation of new markers surrounding genes that affect economically important traits.

Molecular Selection is based on the concept of the Selectable Fragment Library (SFL), which is a collection of DNA fragments with a particular set of characteristics. The population of fragments in an SFL can be modified by selection. All of the DNA sequences that make up a particular SFL are located in the same tube and all laboratory operations are performed on all sequences simultaneously.

Selection would often be applied by comparing an SFL with the fragments in a second population. For example, an Angus-specific (relative to Brahman) library might be produced by selecting those fragments from an SFL constructed from an Angus animal that are different from (or not present in) a similar library constructed from a Brahman animal. Several methods for genomic subtraction have been described in the literature.

MATERIALS AND METHODS

Table 1. Oligonucleotides used in the experiment. Bases that are mismatched relative to their template are shown in bold. Nla III and Fok I recognition sites are in lower case.

Name	Sequence (5'-3')	Name	Sequence (5'-3')
P008	CCTAACCGCGACCAGCCGAGTGAAGCcatg	P010	CCCAACGTCAACTAACCAGACAGGACcatg
P108	CCTAACCGCGACCAGCCGAG	P110	CCCAACGTCAACTAACCAGA
P208	CCCGACCAGCCGAGTGAAG	P210	CTCAACTAACCAGACAGGA
R124	CTACACCCATACACCCATACACA	R127	TCGTGTTTGCGTGTTTGCGTGTG
R224	CTACACCCAAACACCCATggatgCACA	R227	TCGTGTTTGCGTGTTTGCggatgTGTG

Molecular Selection begins with an SFL of total genomic DNA. Genomic DNA was digested with the restriction enzyme Nla III which leaves a 4 base overhang on the 3' strand. Next, the 3' ends of the fragments were dideoxy terminated by the addition of a single nucleotide of ddGTP with terminal transferase, increasing the 3' overhang to 5 bases. Finally, single stranded oligonucleotide linkers were ligated to the ends of the fragments

using T4 DNA ligase. The extra dideoxy guanosine at the end of the genomic fragments prevented them from ligating to one another. SFLs were constructed from DNA of an Angus cow using linkers P008 and P010 (see Table 1) to produce libraries SFL8 and SFL10, respectively. Figure 1 illustrates the main features of fragments in a genomic SFL.

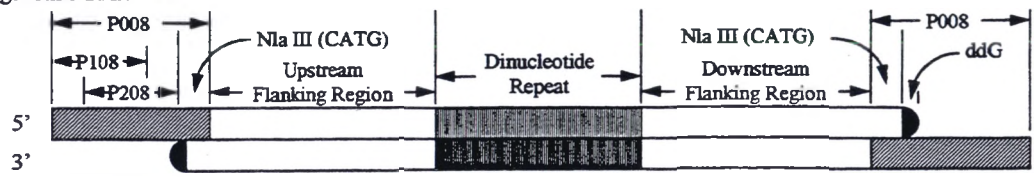

Figure 1. The organization of a microsatellite fragment in an SFL.

A model system utilizing fragments that contain $(GT)_n$ microsatellites consists of several components: they include selection of fragments that contain microsatellites from the total genomic SFL, isolation of individual microsatellite loci, and a system for assigning genotypes to individual animals once a locus has been isolated.

Selection of fragments that contain microsatellites from the total genomic SFL is a critical step. The polymerase chain reaction (PCR) is used to amplify the regions flanking the dinucleotide repeat. A fragment containing $(GT)_n$ in SFL8 will be used as an example. Its upstream flanking region (UFR) is amplified with P108 and R124. Figure 2 shows that the region upstream of the dinucleotide repeat will be amplified, but fragments without a $(GT)_n$ sequence are not amplified. This is the basis for selecting microsatellite fragments out of total genomic DNA. However, any mispriming event terminates in a sequence complementary to P108 and is readily amplified. Therefore, in the second stage of the PCR, nested primers P208 and R224 are used to select against these fragments. The products of this reaction are referred to as "left end fragments." "Right end fragments," containing the downstream flanking region are produced by amplifying with P108 and R127 followed by P208 and R227.

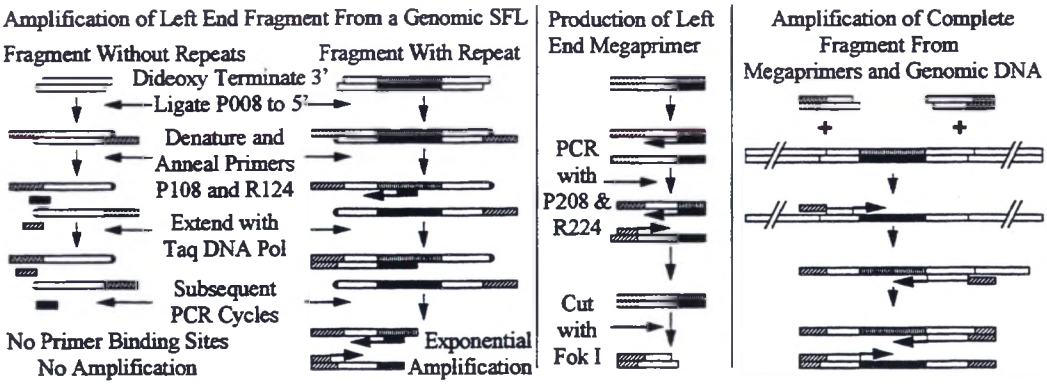

Figure 2. Selection of microsatellite fragments from a genomic SFL, production of megaprimers, and amplification of microsatellite fragments with megaprimers.

In the next step, the end fragments were used as megaprimers (Sarkar and Sommer, 1990). Consequently, it was important that the repeat primer ends of these fragments matched the genomic sequence exactly. Therefore, Fok I restriction sites were designed into R224 and R227. Fok I cuts 9 and 13 bases downstream of its recognition

sequence on the 5'-3' and 3'-5' strands, respectively, so it usually cuts within the flanking region rather than in the dinucleotide repeat sequence. The PCR product was digested by the direct addition of Fok I and was electrophoresed on a 4% Metaphor agarose gel and the megaprimer band was excised, melted, and diluted to 10% in water.

The megaprimers were used to regenerate the full length microsatellite-containing fragment, known as a "complete fragment." It consisted of the genomic fragment from one Nla III site to the other (dinucleotide repeat sequence and both flanking regions) with a double stranded linker on each end. This reaction began by amplifying whole genomic DNA with an upstream megaprimer and a downstream megaprimer using Deep Vent (exo$^-$) DNA polymerase (New England BioLabs) without any oligonucleotide primer. Figure 2 illustrates the megaprimer reaction.

The product of the initial amplification with megaprimers only was then diluted and reamplified with P108 and/or P110 using Taq polymerase. This was followed by dilution and reamplification with the nested primer set P208 and/or P210 to produce the complete fragment. This reaction could be performed with a complex mixture of megaprimers during the selection process or with a single set of megaprimers for assigning genotypes.

RESULTS

The amplification of the end fragments was a critical step. Primers that consisted entirely of dinucleotide repeats were expected to cause problems. $(GT)_7$ and $(GT)_{10}$ were tested and found to be unacceptable primers because they resulted in smears of DNA. Several modified primer designs were considered. Grist et al. (1993) used primers in which the 3' nucleotide broke the repeat motif. Taylor et al. (1992) used primers with modified 5' ends, $C_5(GT)_7G$ and $G_5(CA)_7C$ to successfully amplify flanking regions in cloned inserts.

In the preliminary stages of this project, $(GT)_9C$, $G_5(CA)_7C$, and $C_6(GT)_9C$ were characterized extensively. All of them could be made to produce sharp bands when amplifying cloned inserts with the appropriate repeat motif. However, they were very sensitive to annealing temperature, primer and Mg^{++} concentrations, and other reaction conditions. Under suboptimal conditions, smears of DNA resulted. These primers did not produce stable amplifications of complex mixtures of size selected end fragments from genomic SFLs. Instead, they produced smears both above and below the location of the expected band.

R124 (Table 1) was designed to overcome these problems. It is an AC-based primer with mismatches relative to $(AC)_n$ placed strategically throughout it in an eight base repeating motif. In the initial PCR cycle, it is equally likely to anneal anywhere within an $(AC)_n$ sequence. However, in subsequent cycles, any misalignment other than in multiples of eight bases generates numerous mismatches and would be highly unstable. Obviously, correct alignment results in 23 perfectly matched base pairs and is quite stable. Misalignments that would lengthen the fragment by eight or sixteen bases result in only fourteen or six matched base pairs, respectively, and, in the event of annealing, result in a 3' terminal A:G mismatch which is unlikely to be extended by Taq polymerase (Kwok et al., 1990). On the other hand, if a PCR fragment has eight or more bases of perfect CA repeat internal and adjacent to the R124 sequence, then a misalignment that shortens the fragment by eight bases results in only three mismatches, five and eight bases from the 3' end and at the 5' terminus of R124. This allows formation of a stable duplex that is readily extended by Taq polymerase.

The design of R124 makes it unlikely that daughter strands will be longer than their respective template strands, but they may be shorter. However, template strands that have less than eight bases of repeat beyond the R124 site are unlikely to produce daughter strands that are shorter than their template because an eight base misalignment would probably generate mismatches at or near the 3' end of R124 that would prevent primer extension. Theoretically, the PCR product should consist of approximately equal numbers of fragments with 0, 1, 2, or 3 CA repeats beyond the R124 site. R127 was designed with essentially the same features as R124.

R124 and R127 did in fact work as expected. They have been extensively characterized on a cloned bovine genomic fragment that contains $(GT)_{19}$. These primers have worked consistently at annealing temperatures ranging from 48 to 60°C.

Amplifications of size selected SFL8 and SFL10 with P108 or P110 and R124 or R127 followed by size selection of the resulting end fragments and reamplification with P208 or P210 and R224 or R227 have resulted in

stable amplification of fragments of the selected sizes. When these products are digested with Fok I, the expected shift of approximately 30 bp is visible on the gel.

Megaprimers produced from size selected fractions of SFL8 and SFL10 have been used in the megaprimer reaction on whole genomic DNA to regenerate fragments within the expected size range. An additional putative microsatellite fragment was generated by amplifying a size selected fraction of SFL8 with P108 and a megaprimer produced from a similar size fraction of SFL10. Left and right end fragments were produced from this complete fragment and were of the expected size. Amplification of whole genomic DNA with megaprimers produced from these end fragments regenerated a complete fragment of the same size. DNA sequencing verified that this fragment did indeed contain a $(GT)_7GC(GT)_5$ sequence in the predicted location.

DISCUSSION

This presentation has focused on several tools for manipulating DNA fragments associated with microsatellites. The objective is to apply these tools to select microsatellite loci that cosegregate with genes or markers of particular interest. The key to accomplishing this selection is in alternating between end and complete fragments to take advantage of the different information content in the two fragment types.

Figure 2 shows that the end fragments are locus-specific, but allele-independent. The sequence of the flanking region conveys locus information, but the number of repeats (allele) of the source has no effect on the end fragments. On the other hand, the complete fragments are both locus-specific and allele-specific because their length is a direct reflection of the length of the source.

The intersection of two sets of fragments will be referred to as "positive selection." The intersection of the loci present in two populations of fragments can be obtained by using left end megaprimer from one population and right end megaprimer from the other to amplify whole genomic DNA. A more stringent selection could be obtained by amplifying template from one of the populations with left and right end fragments from the other.

Positive selection could be used to select those microsatellite loci for which a particular individual is heterozygous for specified allele lengths. An SFL from the individual would be size selected into two fractions corresponding to the desired allele lengths. Left end megaprimer would be made from one and right end from the other. Then these two megaprimers would be used together to amplify whole genomic DNA from the individual.

An important application of Molecular Selection could be QTL mapping through pooled sampling. SFLs of microsatellite fragments would be constructed from pooled DNA samples of animals from the two tails of the distribution for some phenotypic trait. Fragments that had one allele in one library and a different allele in the other library would be selected. If desired, this procedure could be applied to several different families simultaneously and only those fragments that were selected in all families could be kept. The selected fragments could then be mapped to indicate genomic regions that deserved a more detailed QTL analysis.

Modification of the primers to amplify trinucleotide repeats should also provide a rapid approach to the isolation of genes in which disease is the result of microsatellite expansion. Fragments that are longer in diseased tissue than in normal tissue from the same individual would be amplified and isolated.

ACKNOWLEDGMENTS

The authors wish to express appreciation to Mark Holder, Doug Holder, and Albert Fliss for their assistance in the laboratory.

REFERENCES

GRIST, S.A., FIRGAIRA, F.A., and MORLEY, A.A. (1993) BioTechniques 15 : 304-309.

KWOK, S., KELLOGG, D.E., MCKINNEY, N., SPASIC, D., GODA, L., LEVENSON, C., and SNINSKY, J.J. (1990) Nucleic Acids Res. 18 : 999-1005.

SARKAR, G. and SOMMER, S.S. (1990) BioTechniques 8 : 404-407.

TAYLOR, G.R., HAWARD, S., NOBLE, J.S., and MURDAY, S. (1992) Anal. Biochem. 200 : 125-129.

THALLMAN, R.M., TAYLOR, J.F., and DAVIS, S.K. (1993) J. Anim. Sci. 71(Suppl. 1) : 94.