

Microsatellites: The Bridge from the Field to the Lab

Cheryl A. Schmidt. Department of Biology, Central Missouri State University,
Warrensburg, MO 64093 CAS8785@cmsu2.cmsu.edu

Many undergraduate students perceive a chasm between the field biologist and the laboratory. Recent developments in molecular biology have provided "bridges" between these areas, resulting in unprecedented ability to elucidate the effects of ecological parameters on the genetic structure of populations and species. The use of molecular techniques such as PCR (Polymerase Chain Reaction) combined with recently developed methods to locate hypervariable DNA such as microsatellites, has yielded sets of genetic markers which allow us to address issues ranging from population genetics to behavioral ecology. Microsatellites are hypervariable DNA sequences comprising tandem repeats of simple motifs (1-6 nucleotides) with overall lengths usually of 100 base pairs or less. The hypervariable nature of microsatellite loci makes them particularly suited to addressing questions of genetic mating systems, patterns of relatedness at intra- and interpopulational levels, and overall assessment of genetic diversity within a species. Primers for PCR amplification of microsatellite loci are available for many vertebrate species including laboratory, domestic and wild (native) species. The proposed module describes the use of microsatellite primers developed for the native white-footed mouse (*Peromyscus leucopus*) to amplify hypervariable markers for assessment of geographic variation within that species. The techniques and principles established in the introductory module could easily be applied to a diversity of extended independent undergraduate research questions (e.g. determination of genetic relatedness and pedigrees; parentage determination; and elucidation of genetic mating systems), as well as to different taxa for which microsatellite primers are available. A vast array of undergraduate research questions could be developed with such a system.

Microsatellites: The Bridge from the Field to the Lab

Cheryl A. Schmidt. Department of Biology, Central Missouri State University, Warrensburg, MO 64093

Abstract. Many undergraduate students perceive a chasm between the field biologist and the laboratory. Recent developments in molecular biology have provided "bridges" between these areas, resulting in unprecedented ability to elucidate the effects of ecological parameters on the genetic structure of populations and species. The use of molecular techniques such as PCR (Polymerase Chain Reaction) combined with recently developed methods to locate hypervariable DNA such as microsatellites, has yielded sets of genetic markers which allow us to address issues ranging from population genetics to behavioral ecology. Microsatellites are hypervariable DNA sequences comprising tandem repeats of simple motifs (1-6 nucleotides) with overall lengths usually of 100 base pairs or less. The hypervariable nature of microsatellite loci makes them particularly suited to addressing questions of genetic mating systems, patterns of relatedness at intra- and interpopulational levels, and overall assessment of genetic diversity within a species. Primers for PCR amplification of microsatellite loci are available for many vertebrate species including laboratory, domestic and wild (native) species. This paper provides a brief description of the process of using microsatellite primers to address questions about genetic variation among individuals or taxa. The techniques and principles described could easily be applied to a diversity of extended independent undergraduate research questions (e.g. determination of genetic relatedness and pedigrees; elucidation of mating systems, etc.), as well as to different taxa for which microsatellite primers are available. A vast array of undergraduate research questions could be developed with such a system.

Microsatellites are tandem repeats of simple (1-6 nucleotides) motifs with overall lengths usually of 100 base pairs or less (Litt and Luty 1989; Epplen *et al.* 1991). This class of simple sequence DNA has been implicated in a wide variety of functions including gene regulation (Hamada *et al.* 1984) and signals for gene conversion and recombination (Jeffreys *et al.* 1985). Microsatellites have been shown to be a major source of genetic variation (Tautz *et al.* 1986) such that microsatellite polymorphisms have been found between individuals (Litt and Luty 1989; Tautz 1989; Weber and May 1989), among inbred strains of rats (Serikawa *et al.* 1992), and among populations of invertebrates (Choudhary *et al.* 1993) and vertebrates (Ellegren *et al.* 1992).

Microsatellites are common in eukaryotes, with some motifs being more common than others (Tautz *et al.* 1986; Moran 1993). In particular, the GT/CA tandem repeat has been shown to be very abundant in a variety of mammal species examined (Hamada *et al.* 1982; Tautz and Renz 1984; Van Den Bussche *et al.* 1995). Janecek *et al.* (1993) compared the abundance of all possible dinucleotide motifs in the native North American white-footed mouse, Peromyscus

leucopus, and concluded that GT/CA was the most common dinucleotide repeat (87% of clones hybridized producing an estimate of 75,000 copies in the haploid genome), followed by CT (59% of clones hybridized, producing an estimate of 50,000 copies in the haploid genome. Screening of the same genomic library of P. leucopus used by Janecek *et al.* (1993) with (GATA)_n produced 16% of the clones hybridizing with the (GATA)_n probe, yielding an estimate of over 13,000 copies in the haploid genome (Baker 1994). Epplen's (1988) review of the ubiquitous occurrence and high levels of polymorphism of this microsatellite in animal genomes, indicates that (GATA)_n may be a fruitful microsatellite even though it has much lower copy number than the two dinucleotide repeats. (CT)_n, (GT)_n, and (GATA)_n microsatellite loci have great potential for yielding markers which can distinguish among individuals and populations of animals. Microsatellite sequences (from which primers can be designed) are available for many economically important species (e.g. chickens, Khatib and Soller 1995; cotton – Baker *et al.*, 1995; cows – Burns *et al.* 1995; horses – Ewen and Matthews 1995) as well as a growing number of native species (e.g. North American wolflike canids – Roy *et al.* 1994; marine turtles – FitzSimmons *et al.* 1995; and European wild rabbits – Rico *et al.*, 1994). Sequences for microsatellites for many taxa can also be easily acquired through searches of GenBank (<http://www.ncbi.nlm.nih.gov>).

A detailed description of methods and analyses employed in the use of microsatellite variation to address questions of population structure in the white-footed mouse (P. leucopus), as an example of one type of research program in which microsatellites are effective tools, is provided in Schmidt (1995). The current paper provides general methods for primer pair development and use which may be applied to sequences from a variety of organisms.

Methods & Materials

Primer Development and Optimization

When searching GenBank or other sources for potential microsatellite loci, keep in mind that alleles of tetranucleotide repeats will be easier to distinguish using nonradioactive visualization (e.g. separation using MetaPhorTM agarose), than are alleles of dinucleotide repeats. Primers for amplification of microsatellite loci are developed complementary to the sequence of bases upstream and downstream of the selected microsatellite region. Three criteria should be remembered when selecting regions for development of complementary primers. The first is a relatively even distribution of the four nucleotides to avoid developing primers complementary to base sequences which are inherently variable due to tandem repeats. The second desirable condition is that the sequence to which the complementary primer is to be developed should contain at least 40% Cytosine/Guanine. Cytosine and guanine, with their triple hydrogen bonds, enhance primer annealing and stability on the priming site during amplification. The last criterion is to make the distance between primers of a pair range between 100 and 300 bp. Under most circumstances, primer pairs should be developed only for microsatellite clusters containing 10 or more tandem repeats. Whenever possible, primers should be 20 bp in length, however, shorter (e.g. 19-mer and 15-mer) primers have been effectively used. Primers are synthesized on the 0.2 uM scale, cleaved, deprotected, desalted, and nonphosphorylated.

Each primer pair is initially put through a survey of PCR conditions to determine optimal [MgCl₂], [genomic DNA], and annealing temperature. Seventy-five nanograms of genomic DNA from the model organism serves as the trial genomic DNA in the initial surveys. Synthesized DNA based on the microsatellite sequence used to develop the primer pair serves as the positive control, and the negative control uses 2.5 ul of ddHOH in place of DNA. Concentrations of MgCl₂ range from 0-2.5 ul/reaction in 0.5 ul/reaction steps. All trials are initially conducted with a thermal profile of 35 cycles as follows: 95°C for 1 minute (strand separation), 55°C for 30 seconds

(primer annealing), and 72°C for 30 seconds (elongation). If sufficient product is achieved but spurious banding is visible when the products are separated on a 2% agarose gel, annealing temperature may be increased (a suggested temperature step is to 58°C) and the reactions tried again at the optimal [MgCl₂] determined previously. The above PCR conditions are suggested starting points only, and some time should be invested in determining the most efficient conditions under which adequate product can be acquired. Once optimal conditions for amplification with the primer pair are determined, the primers can be used to address a variety of questions based on variation among individuals at that specific microsatellite locus.

Application of Microsatellite Primers

If radioisotope visualization is to be used, one primer of each primer pair is endlabeled by a standard g³⁵S-ATP, T₄ Polynucleotide Kinase reaction (Sambrook *et al.* 1989) prior to PCR amplification of homologous loci from genomic DNA of the individuals among which variation is to be assessed. Amplification is carried out using the optimal parameters established for each locus. endlabeled amplification products are denatured at 80°C for a minimum of 3 minutes prior to loading onto a denaturing polyacrylamide gel. The percentage polyacrylamide should be based on the predicted size range of the amplification products (i.e. products close to 250 bp in length might be effectively separated on a 4% polyacrylamide gel, while products only 150 bp long might require a 10% gel for adequate separation). Sequenced M13 is electrophoresed on each side of the population samples as a size standard. The gel rig used should have a continuous buffer well from top to bottom which provides a constant temperature across the gel, minimizing "smiles" and allowing lines to be drawn between identical base positions of the two M13 sequences. This provides an efficient means for determining the size of alleles in the population samples. Cold (non-endlabeled) amplification products may be separated using a high-resolution agarose such as MetaPhor™ Agarose from FMC.

Data Analysis

There are a plethora of analysis packages available for both Macintosh and IBM type personal computers which can be used to calculate measures of genetic variability. One such widely available package is Biosys-1 (Swofford and Selander, 1981). Measures of genetic variability can then be further analyzed using SAS (SAS Institute, Inc., 1985) or SYSTAT™ (Wilkinson *et al.* 1992).

Literature Cited

- Baker, R.J. (1994). Some thoughts on conservation, biodiversity, museums, molecular characters, systematics, and basic research. *Journal of Mammalogy* 75:277-297.
- Baker, R.J., J.L. Longmire, and R.A. Van Den Bussche. (1995). Organization of repetitive elements in the upland cotton genome (*Gossypium hirsutum*). *Journal of Heredity*, 86:178-185.
- Burns, B.M., J.F. Taylor, K.L. Herring, A.D. Herring, M.T. Holder, J.S. Collins, T.M. Guerra, J.O. Sanders, S.K. Davis. (1995). Bovine microsatellite dinucleotide repeat polymorphisms at the TEXAN11, TEXAN12, TEXAN13, TEXAN14, and TEXAN15 loci. *Animal Genetics*, 26:201-202.
- Choudhary, M., J.E. Strassman, C.R. Solis, and D.C. Queller. (1993). Microsatellite variation in a social insect. *Biochemical Genetics* 31:87-96.
- Ellegren, H., M. Johansson, K. Sandberg, and L. Andersson. (1992). Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* 23:13-142.
- Epplen, J.T. (1988). On simple repeated GT/CA sequences in animal genomes: a critical reappraisal. *Journal of Heredity* 79:409-417.
- Epplen, J.T., H. Ammer, C. Epplen, C. Kammerbauer, R. Mitreiter, L. Roewer, W. Schwaiger, V. Steimle, H. Zischler, E. Albert, A. Andreas, B. Beyermann, W. Meyer, J. Buitkamp, I. Nanda, M. Schmid, P. Nurnberg, S.D.J. Pena, H. Poche, W. Sprecher, M. Scharl, K. Weising, and A. Yassouridis. (1991). Oligonucleotide fingerprinting using simple repeat motifs: A convenient, ubiquitously applicable method to detect hypervariability for multiple purposes. Pages 50-69 in *DNA fingerprinting: Approaches and applications* (T. Butke, G. Dolf, A. Jeffreys, and R. Wolff, Eds). Birkhauser Press, Basil.
- Ewen, K.R., and M.E. Matthews. (1995). An equine microsatellite repeat at the VIAS-H64 locus. *Animal Genetics*, 26:277.
- FitzSimmons, N., C. Moritz, and S. Moore. (1995). Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* 12(3):432-440.

- Hamada, H., M. Petrino, and T. Kakunaga. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Science* 79:6465-6469.
- Hamada, H. M. Seidman, B. Howard, and C. Gorman. (1984). Enhanced gene expression by the poly (dT-dG)poly(dC-dA) sequence. *Molecular and Cellular Biology* 4:2622-2630.
- Janecek, L.L., J.L. Longmire, H.A. Wichman, and R.J. Baker. (1993). Genome organization of repetitive elements in the rodent, Peromyscus leucopus. *Mammalian Genome* 4:374-381.
- Jeffreys, A.J., V. Wilson, and S. Thein. (1985). Hypervariable 'minisatellite' regions in human DNA. *Nature* 314:67-72.
- Litt, M., and J.A. Luty. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44:397-401.
- Moran, C. (1993). Microsatellite repeats in pig (Sus domesticus) and chicken (Gallus domesticus) genomes. *Journal of Heredity* 84:274-280.
- Rico, C., I. Rico, N. Webb, S. Smith, D. Bell, and G. Hewitt. (1994). Four polymorphic microsatellite loci for the European wild rabbit, Oryctolagus cuniculus. *Animal Genetics*, 25:367.
- Roy, M.S., E. Geffen, D. Smith, E.A. Ostrander, and R.K. Wayne. (1994). Patterns of differentiation and hybridization in North American wolf-like canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution* 11(4):553-570.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. (1989). *Molecular cloning: a laboratory manual*. 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- SAS Institute, Inc. (1985). *System for elementary statistical analysis*. SAS Institute, Inc., Cary, North Carolina, 1042 pp.
- Schmidt, C.A. (1995). *Geographic variation of Peromyscus leucopus: microsatellite variation at regional and local levels*. Unpublished Ph.D. Dissertation, Texas Tech University, Lubbock, Texas.
- Serikawa, T., T. Kuramoto, P. Hilbert, M. Mori, J. Yamada, C. Dubay, K. Lindpainter, D. Ganten, J. Guenet, F. Lathrop, and J. Beckmann. (1992). Rate gene mapping using PCR-analyzed microsatellites. *Genetics* 131:701-721.
- Swofford, D.L., and R.B. Selander. (1981). BIOSYS-I: a FORTRAN program for the comprehensive analysis of electrophoretic data in population genetics and systematics. *Journal of Heredity*, 72:281-283.
- Tautz, D. (1989). Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Research* 17:6463-6471.

Tautz, D. and M. Renz. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research* 12:4127-4138.

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

Van Den Bussche, R., J.L. Longmire, and R.J. Baker. (1995). How bats achieve a small C value: frequency of repetitive DNA in Macrotus. *Mammalian Genome* 6:521-525.

Weber, J.L. and P.E. May. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44:388-396.

Wilkinson, L., M. Hill, and E. Vang. (1992). SYSTAT for the Macintosh, version 5.2 SYSTAT Inc., Evanston, Illinois.

Materials List

Thermal Cycler: Midwest Scientific #P-5200

Polyacrylamide Gel Rig: CMS gel-Pro #283-961

SequaGel Sequencing System: Midwest Scientific #EC-833

Taq DNA Polymerase: Promega #M1865

g35S-ATP: ICN Biomedicals, Inc. #55000H

Polynucleotide Kinase: New England Biolabs #201S

Primer and Oligonucleotide Construction: check back of Science for nearest and cheapest source

Autoradiography Film: Midwest Scientific #SXF-1417

Waterbath: Midwest Scientific, Standard Aquabath #18005

Heat block: Midwest Scientific, Heater Base #2000LL and Heater Blok #2068

Powerpack: VWR Scientific #3000

Sequenase™ Version 2.0 DNA Sequencing Kit: USB Product #70770

Microcentrifuge: Midwest Scientific Mini Micro-Centrifuge #1200 and #1206

Gel Dryer: Midwest Scientific #D-2000

Vacuum Pump: Midwest Scientific #93010085

DNA Isolation Kit: GIBCO BRL DNAzol Reagent #10503-027

MetaPhor™ Agarose: Midwest Scientific #50181