Fragment Sizing

|  |
| --- |
| * [***Introduction***](https://drupal.unbc.ca/genetics/training/fragment-sizing#Intro) * [***Microsatellites***](https://drupal.unbc.ca/genetics/training/fragment-sizing#Micr) * [***LH-PCR***](https://drupal.unbc.ca/genetics/training/fragment-sizing#Lh) * [***AFLPs***](https://drupal.unbc.ca/genetics/training/fragment-sizing#Af) * [***T-RFLPs***](https://drupal.unbc.ca/genetics/training/fragment-sizing#TR) * [***Genotyping and Scoring***](https://drupal.unbc.ca/genetics/training/fragment-sizing#gen) |

**Introduction**

Labelled fragment analysis covers a broad category of genetic research, but each is based upon the same principle of analyzing varied length polymorphisms. The basic principle involves the labeling of a DNA fragment with dye markers (traditionally labeled by radioactive tags) during PCR followed byelectrophoresis, fluorescence reporting, and size estimation using labeled standards. The CEQ8000 has a laser that excites the dye label to fluoresce just as the varied length amplicons electrophoretically pass through a capillary detection window at rates proportional to their size.

**Microsatellites**

Microsatellite (or 'Short Tandem Repeat' [STR], 'Simple Sequence Repeats' [SSR], 'Sequence Tagged Sites' [STS]) loci are a powerful method for research of Mendelian markers. Microsatellites are found in all prokaryotic and eukaryotic genomes and  are defined as tandemly repeated motifs of  1-6 bases (usually di-, tri-, or tetranucleotide). In eukaryotes they are generally identified in large blocks of non-coding segments of the heterochrotic region, which is concentrated towards the centromere and differentiated from the euchromatic regions away from the centromere.

If you are starting out with an organism were microsatellite primers have yet to be developed, there are several approaches design and target effective primers. The simplest approach is to test primers that have been targeted and amplified in a related species. Genomes are often conserved even among very distant relatives, so it isn't surprising when microsatellite primers work for different species that separated 470 million years ago (Zane et al. 2002). More involved methods for microsatellite isolation involves fragmenting of the DNA, either by restriction enzymes or sonication, ligation of these fragments to a known sequence, vector, or an adaptor, and followed by the identification and recovery of microsatellite containing sequences by hybridizing labeled repeat sequence probes. I direct the reader to Zane et al. (2002) who provides a comprehensive review of the various methods and Hamilton et al. (1999) for a specific example.

The non-coding nature of microsatellites permits non-lethal mutations to occur at a higher rate than protein coding or signal regulatory sections of the genome. This does not mean, however, that microsatellites are selectively neutral. Robust statistical tests of the relative frequency of variants is required to reject neutrality and can be performed in such programs as ***[Arlequin](https://drupal.unbc.ca/genetics/links/index.html" \l "gene" \o "Genetic analysis software" \t "_self)***[***or MEGA3***](https://drupal.unbc.ca/genetics/links/index.html#gene). Depending on the type of analysis you are interested in there are many programs and methods to select from. Visit our links page where you can find links to [***phylogenetic software***](https://drupal.unbc.ca/genetics/genetic_links/index.html#analysis) and [***discussion forums***](https://drupal.unbc.ca/genetics/genetic_links/index.html#trouble).

The key objective of microsatellite analysis is to record the relative sizes of the duplicating segments. Each locus is categorized by electrophoretically separating labeled microsatellite PCR fragments and measuring the relative size (in base pairs).  Panels in fig 1 show microsatellite profiles for four individuals that were run on the CEQ8000. Notice that their are multiple peaks that look like a stegosaurus back. Depending on the type of marker, such profiles form the characteristic of the locus under investigation. The nature of the repeat pattern is thought to result from the polymerase enzyme 'slipping' along the target strand sometimes creating an amplicon that is slightly shorter of longer than the actual microsatellite, although the exact nature of what causes this artifact is still up for debate (Zane et al. 2002). It is generally assumed that the largest peak is the actual size. Individual one, in the top panel, is a heterozygote, because two large peaks are observed at 106.23 and 131.83 respectively. Since it isn't possible to have 0.23 or 0.83 of a bp the values are corrected during the binning stage (described below). These decimal values highlight that the electrophoretic sizing of fragments isn't solely due to the length of the amplicon, but the GC content also has an effect (albeit smaller) on DNA migration rates (Amos et al. 2007).

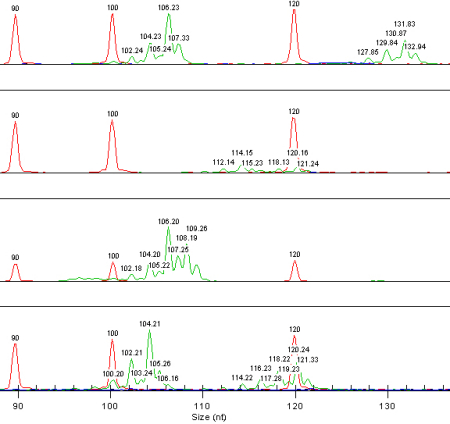


Fig. 1 Beckman-coulter profile with size standard and microsatellite sizes for four individuals shown in separate panels from top to bottom.

It takes time and repeated exercises for a researcher to become familiar with the patterns of their microsatellite so that they may effecitvely score alleles. For example, is the individual in the second panel a homo-or-heterozygote? By zooming in and changing the relative scale of the y-axis it is possible to more clearly see the pattern and determine the genotype for this individual. This process of sorting the raw size and pattern into allele categories is called binning (Dewoody et al. 2006; Amos et al. 2007) and Beckman-Coulter's CEQ8000 software is designed for comprehensive binning analysis. Unfortunately, the reality is that fully automated binning is still not feasible and the error rate can be astoundingly high. A program is available to assist in the search for the best fit of bin locations through[***Dr. Bill Amos***](http://www.zoo.cam.ac.uk/zoostaff/amos/) (Amos et al. 2007).

**LH-PCR**

Length-heterogeneity PCR was pioneered by Suzuki et al. (1998). The method discriminates the size and proportions of PCR amplications originating from different genes in community structures, such as soil. The technique focuses in on natural size variation that exists in the lengths of small-subunit rRNA genes or the ammonia monooxygenase (*amo*) operon (Mills et al. 2007). The method is being tested in our lab by [***Dr. Keith Egger***](http://web.unbc.ca/~egger/) as a means of quantifying biological community structure in soils. The relative area under each fluorescent peak in the electrophoretic profiles can be used to estimate the relative biomass of the species that corresponds to the DNA sequence generating the peak (Richie et al. 2000). There are, however, inherent limitations in the PCR amplification process for detecting certain members of the soil community with equal probability. The method is also imprecise because each peak at a particular length may or may not represent more than species. To address this limitation it is best to complement LH-PCR with clone libraries and sequencing of the length variants. Mills et al. (2007) gives a recent overview of the method.

**AFLPs**

Amplified fragment length polymorphisms combine elements of PCR and RFLP (Restriction Fragment Length Polymorphism)-based methods.  A subset of an entire genome is selectively amplified following restriction with a 4-bp and a 6-bp enzyme cutter (Usually EcoRI and MSeI). Adapters are incorporated into the sticky ends of the restricted sites. Adapter sequences, adjacent restriction sites, and one to three random nucleotides at the 3' end of the primer serve as binding sites. Vos et al. (1995) provide the initial desciption of the method. Polymorphisms are scored as either present or absent by each unique peak of varied length.

The technique suffers from genetic dominance in the markers, that is heterozygotes cannot be distinguished from a homozygotes. However, the approach has much discriminatory power for strains or populations because it generates a large number of fragments that are generally reproducible within an acceptable statistical degree of precision. Scoring of the AFLP electropherograms is the most crucial and time consuming step. The CEQ8000 Fragment Analysis software allows the user to score each AFLP profile using bin width, amplitude threshold, and range of fragment size, but a manual check of the data is necessary because the automated process is prone to errors. It is generally recommended that 5-10% of each profile be run in duplicate, preferably using separate DNA extractions, to reduce and quantify the potential error rates in your data sets. A comprehensive review of the method and scoring technique is provided by ***[Meudt and Clarke (2006)](http://www.softgenetics.com/TrendsPlantScience2007.pdf" \o "Meudt and Clarke (2006) AFLP review paper" \t "_blank)***. Hardy (2003)  discusses analytical approaches to use dominant markers for effective analysis of isolation-by-distance in populations.

**T-RFLPs**

Terminal restriction fragment length polymorphism is another variant of the RFLP technique and is used for community profiling. The method generally targets 16S rRNA and is being tested in our lab by[***Dr. Keith Egger***](http://web.unbc.ca/~egger/) .  A 5'-fluorescently labelled primer is incorporated into the targetted 16S rRNA amplicon and then digested with various restriction enzymes. The fragments are analyzed on the CEQ8000 and distinguished by length polymorphisms. The length polymorphisms are caused by variants in the DNA sequence of the bacteria analyzed plus mutations that alter the specificity the digestive enzymes.  Like all restriction fragment profiling methods there is much potential for non-specific or non-target species amplification, and limitations by restriction enzyme and PCR efficiency (Avis et al. 2006). Moeseneder et al. (1999, 2001) gives an overview of the method and demonstrates its application in microbiological screening applications.

**Genotyping and Scoring**

Genotyping refers to the categorization and interpretation, or scoring, of unique fluorescent peaks (or some other molecular marker) that corresponding to alleles.  This is perhaps the most difficult stage of a fragment study and experience in interpreting the data helps. It is important to execute a pilot study where replicates of samples are run to observe the nature of the fluorescent peaks. There are several artificacts that can hinder this analytical phase including differences in peak intensities, slight shift between two peaks from muliple runs, and shadow bands caused by slippage of *Taq* polymerase during PCR.  The critical assumption in each of the fragment methods is that each peak of similar length represents a single homolog. Improper scoring of the genotype is a common source of error and it is important that the user become familiar with the literature on this subject (*see* Ginot et al. 1996; Bonin et al. 2004; DeWoody et al. 2006; Meudt and Clarke 2006; Amos et al. 2007).

Microsatellite results differ from the rest of the fragment profiling methods described herein, because they reveal diploid Mendelian hetero- or homozygous loci. Hence, as a codominant marker the analytical treatment of the data can be more involved in terms of heterozygosity diversity indices. The remaining methods are classified as dominant markers (but see Piepho and Koch 2000).  There are two main approaches in the scoring of fragment data. The data can be treated as discrete (e.g. 01 binary) or continuous (e.g. 0 > 1 > 2 > 3 scale). Binary data sets are easily accepted into most[***phylogenetic software***](https://drupal.unbc.ca/genetics/genetic_links/index.html#analysis) packages, but for community profile analysis the different between a network and phylogenetic tree must be understood because the operational taxonomic units in a community profile violate the principle of common decent represented in a phylogeny. Continuous character statescan be obtained from relative peak heights in the electropherograms; standards, replicates and controls must be in effect for relative peak height analysis. The analysis of such data may involve[***principal component analysis***](http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf) (Maccaferri et al. 2007), but since most profile data are not normally distributed Mills et al. (2007) describes other forms of appropriate multivariate statistics.

**References**

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S and Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes*, **7**, 10-14.

Avis PJ, Dickie IA and Mueller GM (2006) A 'dirty' business: testing the limitations of terminal restriction fragment lengthy polymorphism (TRFLP) analysis of soil fungi. *Molecular Ecology*, **15**, 873-882.

Bonin A, Bellemain E, Eidesen B, Pompanon F, Brochmann C and Taberlet P (2004) How to track and assess genotyping errors in genetics studies. *Molecular Ecology*, **13**, 3261-3273

Dewoody J, Nason JD, and Hipkins VD (2006) Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes,* **6**, 951-957.

Ginot F, Bordelais I, Nguyen S, Gyapay G (1996) Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research*, **24**, 540-541

Hamilton MB, Pincus LP, Fiore AD and Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques*, **27,** 500-507.

Hardy (2003) Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Molecular Ecology*, **12**, 1577-1588.

Meudt HM, and Clarke AC (2007) Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science*, **12**, 106-117.

Mills DK, Entry JA , Gillevet PM , and Mathee K (2007) Assessing Microbial Community Diversity Using Amplicon Length Heterogeneity Polymerase Chain Reaction. *Soil Sci Soc Am J* **71**, 572-578.

Moeseneder MM, Arrieta JM, Muyzer G, Winter G Herndl GJ (1999) Optimization of Terminal-Restriction Fragment Length Polymorphism Analysis for Complex Marine Bacterioplankton Communities and Comparison with Denaturing Gradient Gel Electrophoresis. *Applied and Environmental Microbiology,***65,** 3518–3525.

Moeseneder  MM, Winter C, Arrieta JM, Herndl GJ (2001) Terminal-restriction fragment length polymorphism (T-RFLP) screening of a marine archaeal clone library to determine the  
different phylotypes. *Journal of Microbiological Methods,* **44**