Fragment analysis applications guide

Go beyond sequencing

applied biosystems

Contents

Chapter 1: Introduction to DNA fragment analysis

Fragment analysis—what does it mean?

In the context of Applied Biosystems™ genetic analyzers, fragment analysis is (1) the preparation, electrophoretic separation, and detection of fluorescently labeled biomolecules (typically DNA), followed by (2) software-assisted analysis and interpretation of the resulting electrophoretic peak data.

The predominant use and application of fragment analysis data is to obtain information from the visualization of fluorescently labeled DNA molecules in an annotated electropherogram or digital data in table format.

For most fragment analysis applications, DNA fragments are prepared as fluorescently labeled PCR products (amplicons). Fragment analysis data are colored peaks that line up on an electropherogram: the x-axis represents size measured in nucleotides, and the y-axis represents the peak signal intensity measured in relative fluorescent units (RFU).

So what kind of information can you get from fragment analysis data?

Examples of widely used fragment analysis applications include the determination of:

- The identity of a single species or multiple species (e.g., viral, bacterial, fungal, plant, animal)
- The identity of an individual (e.g., human identity for forensic, paternity, or maternity purposes) or a different organism
- The allele status of a region or locus of interest (e.g., insertion or deletion, copy number variation, SNP, mutation, variant)
- Discrete genomic differences between multiple specimens (e.g., biological species diversity, molecular fingerprinting, or CRISPR-Cas9–mediated mutagenesis screening)
- The “state of health” or decay of a genome (e.g., apoptosis, microsatellite instability, methylation status, and expansion or shrinkage of genomic loci)
- The secondary structure of a biomolecule
- The activity of DNA-interacting enzymes
- The precise size of restriction fragments or genetic-engineering constructs
- And many more applications

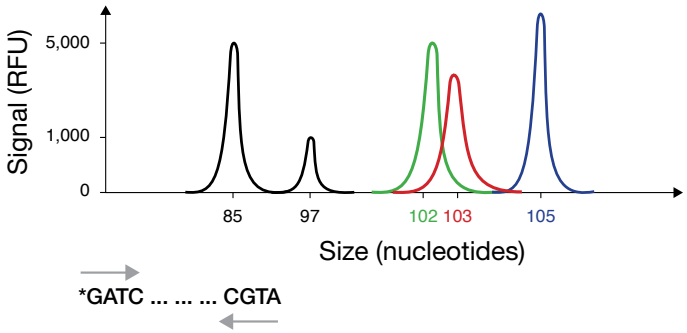


Figure 1.1. A simplified electropherogram to illustrate the concept of four "dimensions of information" in fragment analysis data. Peaks are in different colors according to the fluorescent label of the DNA fragment. The DNA fragments (visualized as peaks) are separated by capillary electrophoresis (CE) according to their size as a function of migration time through the polymer gel matrix. Peaks have different peak heights reflecting different quantities of DNA fragments present in the CE run. For this example, the fragments (peaks) were generated by PCR with defined primers (*); therefore, we know the identity of the fragment and, very likely, also the DNA sequence of the insert. A deviation of the observed size in CE and the size predicted by bioinformatics might indicate a change—a deletion if the fragment is shorter than expected or an insertion if the fragment is longer than expected. Note that overlapping or same-sized peaks can be distinguished due to multicolor detection. (Axes are not drawn to scale.)

What kind of information is contained in the data?

There are essentially four "dimensions of information" embedded in fragment analysis data (Figure 1.1):

- The size of the DNA fragment(s) measured in nucleotides, which gives information on the length of the fragment.
- The peak heights or peak areas of the fluorescent signal detected from the DNA fragment(s) give information on the relative abundance and can be used for quantitation under certain experimental conditions.
- The peak color (i.e., the fluorescent label detected), which can give customized encoded information (e.g., a specific primer (pair), a G, A, T, or C nucleotide, a specific genus or species, a particular panel, or any other encoded information). Currently, five dye channels are available for labeling PCR primers with fluorescent dyes.
- The primer sequences that are used for generating the DNA fragment(s); they provide information on the genomic location and the expected identity (including DNA sequence) of the amplified fragment(s).

Researchers have a toolbox to obtain answers to both simple and more complex biological questions with relatively low effort using these four parameters.

In the following chapters, we provide thorough coverage of the principles of fragment analysis technologies, the basics of multi-color capillary electrophoresis, the steps in the various workflows, required materials, experimental design strategies, and predeveloped solutions for some commonly used applications.

An overview of fragment applications can be found at thermofisher.com/fragmentanalysis.

If you are entirely new to DNA fragment genetic analysis, we recommend the following videos:

- Fragment analysis—the other half of your Applied Biosystems genetic analyzer
- How does fragment analysis work?

What is the difference between fragment analysis and sequencing?

Fragment analysis involves:

- Labeling DNA fragments with fluorescent dyes (Figure 1.2). Multiple fluorescent dyes can be detected from one sample. One of the dyes is used as a size standard in each sample. The size standard is used to extrapolate sizes (bp) of the peaks from each sample.
- Amplifying labeled fragments using polymerase chain reaction (PCR) on a thermal cycler.
- Separating the fragments by size using capillary electrophoresis (CE) on an Applied Biosystems™ genetic analyzer.

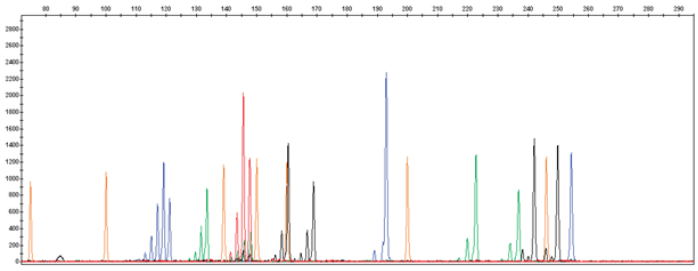


Figure 1.2. Fragment analysis is the process of separating and sizing fluorescently labeled fragments.

- Analyzing the data using software to determine:
  - **Size:** The analysis software uses the size standard in each sample to create a standard curve for each sample. The software then determines the relative size of each dye-labeled fragment in the sample by comparing fragments with the standard curve for that specific sample.
  - **Genotype:** The analysis software assigns allele calls based on user-defined markers (loci).
  - **Relative quantity:** Peak height is measured in relative fluorescent units (RFU) or as integrated numerical values for the peak area.

Sequencing is the determination of the base-pair sequence of a DNA fragment by the formation of extension products of various lengths amplified by PCR (Figure 1.3). For more information, refer to the "Chemistry guide: DNA sequencing by capillary electrophoresis" (Pub. No. 4305080). Sequencing is a special format of fragment analysis, namely a high-density fragment multiplex end-labeled with four fluorescent dyes. In contrast to DNA fragment analysis, no internal size standard is added to the sample before electrophoretic separation. However, a sequencing reaction sample can be treated as a fragment analysis sample by combining it with a size standard (e.g., Applied Biosystems™ GeneScan™ 600 LIZ™ dye standard) and processing the CE run in fragment analysis mode.

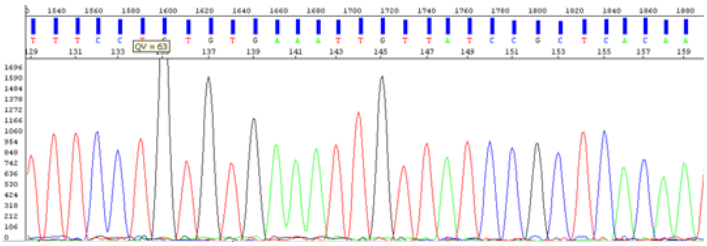


Figure 1.3. Sequencing is the process of identifying the nucleotides of a DNA sequence

What can I do with fragment analysis?

- Microsatellite analysis (see Chapter 4, "Microsatellite analysis")
  - Microsatellite markers (loci), also known as short tandem repeats (STRs), are polymorphic DNA loci consisting of a repeated nucleotide sequence. In a typical microsatellite analysis, microsatellite loci are amplified by PCR using fluorescently labeled forward primers and unlabeled reverse primers. The PCR amplicons are separated by size using electrophoresis. Applications include:
    - Linkage mapping
    - Animal breeding
    - Human, animal, and plant typing
    - Pathogen subtyping
    - Genetic diversity
    - Microsatellite instability
    - Loss of heterozygosity (LOH)
    - Multiple locus variant analysis (MLVA)
- SNP genotyping (see Chapter 5, "Single-nucleotide polymorphism (SNP) genotyping")
  - A single-nucleotide polymorphism (SNP) consists of a single base pair that varies from the known DNA sequence, thereby creating up to four alleles or variations of the marker. Applications include:
    - Applied Biosystems™ SNaPshot™ Multiplex Kit for single-base extension (SBE) labeling
- Allele-specific PCR or amplification-refractory mutation system (ARMS)
- Oligonucleotide ligation assay (OLA)
- Insertion or deletion (indel) detection: a fluorescently labeled PCR primer pair flanking a suspected indel can reveal size information of an indel event

- Fingerprinting (see Chapter 6, “Fingerprinting”)
  - Several AFLP™ (amplified fragment length polymorphism) technologies use restriction enzyme length polymorphism and PCR to generate a fingerprint for a given sample, allowing differentiation between samples of genomic DNA based on the fingerprint. Applications include:
    - Microbial genome typing
  - Inter-simple sequence repeat (ISSR)
- Animal or plant genome typing
- Creation of genetic maps of new species
- Genetic diversity and molecular phylogeny studies
- Establishment of linkage groups among crosses
- Relative fluorescence (see Chapter 7, “Relative fluorescence quantitation (RFQ)”)
  - Relative fluorescence applications compare peak height or area between two samples. Common techniques include:
    - Qualitative fluorescence (QF) PCR
    - Quantitative multiplex PCR of short fluorescent fragments (QMPSF)
    - MLPA™ (multiplex ligation-dependent probe amplification) analysis
  - Applications include:
    - Loss of heterozygosity (LOH) in tumor samples
    - Copy number variation (CNV)
    - Aneuploidy detection
- Analysis of double-stranded DNA (see Chapter 8, “Application note—Analysis of double-stranded DNA”)
  - Biochemical characterization of nucleic acid metabolic enzymes
  - Fluorescent DNA template as substrate and reporter of biochemical activity

Note: Not all applications are covered in this document. Please visit **thermofisher.com/fragmentanalysis** for more information.

What is capillary electrophoresis?

Capillary electrophoresis (CE) is a process used to separate ionic fragments by size. Applied Biosystems™ CE systems use an electrokinetic injection to inject DNA fragments from solution into each capillary.

During CE, the amplification products of the PCR reaction (and any other negatively charged molecules such as salt or unincorporated primers and nucleotides) enter the capillary as a result of electrokinetic injection. A short high-voltage charge applied to the sample forces the negatively charged fragments into the capillaries. The electrokinetically injected molecules are separated by size based on their total charge.

Categories of applications in this guide.

Category	Description
Supported by Thermo Fisher	Thermo Fisher Scientific has tested and validated this protocol on the instrument system specified. The technical support and field application specialists have been trained to support this protocol.
Demonstrated by Thermo Fisher	Thermo Fisher has tested this protocol but has not validated for the instrument system specified. Certain components of the protocol workflow such as reagent kits and other protocols for preparation of reagents may not be available through Thermo Fisher. Supporting documentation such as application notes may be available from Thermo Fisher and/or third parties. Limited support is available from Thermo Fisher.
Demonstrated by customer	Thermo Fisher has not tested this protocol. However, at least one customer or third party has reported successfully performing this protocol on the instrument system specified. Thermo Fisher cannot guarantee instrument and reagent performance specifications with the use of customer-demonstrated protocols. However, supporting documentation from Thermo Fisher and/or third parties may be available and Thermo Fisher may provide basic guidelines in connection with this protocol.

The electrophoretic mobility of the sample can be affected by the run conditions: the type, concentration, and pH of the buffer; the run temperature; the amount of voltage applied; and the type of polymer used.

Shortly before reaching the positive electrode, the fluorescently labeled DNA fragments, which are separated by size, move across the path of a laser beam. The laser beam causes the dyes attached to the fragments to fluoresce. The dye signals are separated by a diffraction system, and a CCD camera detects the fluorescence (Figure 1.4).

Because each dye emits light at a different wavelength when excited by the laser, all colors, and therefore loci, can be detected and distinguished in one capillary injection.

The fluorescence signal is converted into digital data, then the data is stored in a file format compatible with an analysis software application.

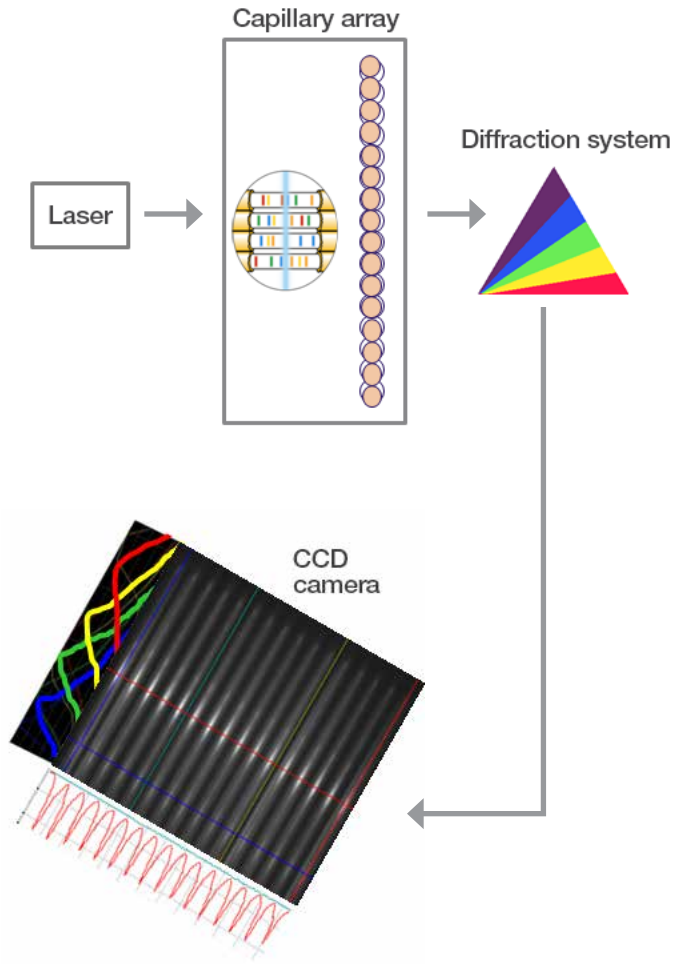


Figure 1.4. Overview of the physics of fragment analysis.

Hardware required for fragment analysis

To perform fluorescent DNA fragment analysis by capillary electrophoresis, two types of instruments are required:

- PCR thermal cycler
- Capillary electrophoresis (CE) genetic analyzer

Thermo Fisher offers a variety of thermal cyclers and genetic analyzers to match your budget and throughput needs.

A brief overview of the steps involved in the fragment analysis workflow.

Phase	Technology	Products
1. DNA isolation	Depends on sample source and application	DNA isolation method depends on the starting DNA source. For guidelines on isolating DNA, visit **here**.
2. DNA purification	Depends on sample source and application	For information on appropriate products, visit **here**.
3. DNA quantification	Dye-labeling and fluorometric detection	Invitrogen™ Qubit™ fluorometer and quantitation kits For more information, visit **thermofisher.com/qubit**.
4. PCR amplification	Dye-labeling and amplification of fragments using a thermal cycler	Applied Biosystems™ ProFlex™ PCR System Applied Biosystems™ VeritiPro™ Thermal Cycler Applied Biosystems™ SimpliAmp™ Thermal Cycler Applied Biosystems™ MiniAmp™ Thermal Cycler
5. Capillary electrophoresis	Separation of fragments based on size using a genetic analyzer	Applied Biosystems™ 3500 or 3500xL Genetic Analyzer Applied Biosystems™ 3730 or 3730*x*l Genetic Analyzer Applied Biosystems™ SeqStudio™ Genetic Analyzer Applied Biosystems™ SeqStudio™ 8 Flex Genetic Analyzer Applied Biosystems™ SeqStudio™ 24 Flex Genetic Analyzer
6. Data analysis	Sizing and optional genotyping	Applied Biosystems™ GeneMapper™ Software Applied Biosystems™ Microsatellite Analysis Software
	Sizing	Applied Biosystems™ Peak Scanner™ Software

Chapter 2: Fragment analysis applications in common research fields

Fragment analysis is a highly flexible genetic analysis method that can be applied to a wide variety of research fields. The flexibility afforded with the choice of PCR primers (PCR being a necessary step in any fragment analysis experiment) means that a specifically sized fragment corresponding to a PCR target sequence is straightforward to generate. The ability to precisely select target fragment sizes and label fragments with up to four different fluorophores gives researchers great flexibility in the experimental design for fragment analysis.

Here we describe some examples of how fragment analysis can be used in common research fields. However, this is not a complete or comprehensive list. The potential applications of fragment analysis depend only on the investigators' imagination and experimental creativity.

Oncology research

Mutation detection

Oncology researchers frequently need to investigate or verify a limited number of variations in one or more genes of interest. The Applied Biosystems™ SNaPshot™ Multiplex System is a versatile and economical method of SNP genotyping. Up to ten SNP markers (on different genes) can be investigated simultaneously by using PCR amplification, followed by dideoxy single-base extension (SBE) with an unlabeled primer, and then capillary electrophoresis (CE) of the resulting fragments. After electrophoresis and fluorescence detection, the alleles of a single marker appear as different colored peaks at roughly the same size in the electropherogram plot. This technology has been used for simultaneous detection of 4 hotspot mutations in *PIK3CA*, a gene involved in the phosphatidylinositol 3-kinase pathway that plays an important role in cancer pathogenesis [1].

Epigenetic modification

Changes in DNA methylation patterns are also characteristic of oncogenesis. Genome-wide hypomethylation or hypermethylation has been shown to be associated with cancer development [2]. This is particularly true of cytosines that are 5′ to guanosines (CpG) within the promoter regions of tumor repressor genes. The SNaPshot assay can also be used to investigate DNA methylation patterns following bisulfite treatment of genomic DNA. In this process, sodium bisulfite converts unmethylated cytosines into uracils, which are subsequently replaced as thymines in PCR; in contrast, methylcytosines remain unchanged. This feature can be used to distinguish base differences by PCR amplification and T/C genotyping using SNaPshot SBE followed by fragment analysis.

Microsatellite instability assay

Microsatellite instability (MSI) is a hallmark of several cancers and is characterized by changes to the length of microsatellites due to defects in mismatch repair mechanisms (MMR). Microsatellites are genetic motifs consisting of 1–6 base pair repeats. These sequences are susceptible to replication errors that can result in deletions and insertions. Normally, these errors are corrected by the DNA MMR mechanism; however, when deficiencies in the DNA MMR system are present, microsatellite replication errors accumulate in the genome [3]. In addition to its prognostic value for certain types of cancer, including colorectal [4] and endometrial, MSI can also serve as a predictive biomarker for immune checkpoint therapy response [5].

The Applied Biosystems™ TrueMark™ MSI Assay contains 13 mononucleotide MSI markers. Eight of these markers are derived from the literature and the guidelines from the National Cancer Institute; the other five markers were internally identified for monomorphism and high sensitivity in multiple cancer types. The assay also contains two sample identification markers to detect sample mix-up or contamination.

Loss of heterozygosity

MLPA™ (multiplex ligation-dependent probe amplification) technology, developed and commercialized by MRC Holland, is a flexible technique that is commonly used to detect aberrations in gene copy number such as loss of heterozygosity. It is based on the ligation and PCR amplification of up to 50 multiplexed pairs of probe oligonucleotides, which hybridize to the loci of interest. Each oligonucleotide pair is designed to give an amplification product of a specific length; by using sequence-tagged ends, all ligated probes can be amplified with a single primer pair in a PCR reaction. The forward PCR primer carries a fluorescent label, allowing for the detection and quantification of size-separated probes on an automated capillary electrophoresis system.

Several cancers are characterized by germline point mutations in specific genes, such as *BRCA1* mutation in breast cancer, *BRCA2* mutation in ovarian cancer, etc. However, even for such well-characterized cancers, mutations involving genomic rearrangements have been found in a subset of cases [6]. The MLPA assay can be used to interrogate these genomic loci in cases where the initial screen is negative even though there is a high probability of disease inheritance. For instance, the MLPA assay has been used to show deletions and duplications in the *BRCA1* gene in people who had a history of breast cancer but tested negative for *BRCA* point mutations [7].

MLPA analysis can also be used to detect epigenetic modifications involving DNA methylation that contribute to cancer formation [8]. The methylation specific–MLPA (MS-MLPA) assay has probes that bind and ligate over a GCGC sequence, which is also a cleavage site for the methylation-sensitive restriction enzyme HhaI. The enzyme cleaves the probes that are ligated and hybridized to unmethylated DNA, while the probes bound to methylated DNA remain intact and are subsequently amplified by PCR. This assay has been used to determine promoter methylation status of tumor suppressor genes *p15* and *p73* in acute myeloid leukemia cell lines [8].

Cell line authentication and human sample matching

Cancer cell lines are heavily relied upon for biomedical research including cancer-related studies.

To ensure reproducibility of scientific research, cell line authentication (CLA) is of paramount importance for confirming the cell's origin as well as checking for contamination and genomic instability. CLA is performed by generating a profile of highly variable short tandem repeat (STR) markers from microsatellite loci with varying number of repeats for a particular cell line or type. This profile is then compared with the allelic profiles present at these loci in known standards. A study aimed at authenticating 278 human tumor cell lines used in China found that nearly 46% of the samples were either cross-contaminated or misidentified [9]. Needless to say, such findings can have massive implications on the conclusion of studies that utilize those cell lines. The Applied Biosystems™ CLA GlobalFiler™ PCR Amplification Kit generates a molecular fingerprint for 24 different STR loci, while the Applied Biosystems™ CLA IdentiFiler™ Plus PCR Amplification Kit analyzes 16 STR loci. These fingerprints can then be used for CLA.

In addition to CLA, STR typing can be used for human sample matching. CAR T cell therapy is a potential cellular immunotherapy that has seen great success in the treatment of hematological malignancies. CAR T cell preparation involves the isolation, activation, engineering, and expansion of patient-derived T cells. Every step of this process can affect the efficacy and consistency of a CAR T cell product and requires thorough characterization to ensure that the product meets predesigned specifications. STR typing can be used to confirm that the CAR T cells match the starting T cell population, ensuring the identity of the final CAR T product [10].

Neuroscience research

Spinal muscular atrophy

Spinal muscular atrophy (SMA) is an autosomal recessive genetic disease caused by mutations in the survival motor neuron 1 (*SMN1*) gene located on chromosome 5q. It is characterized by progressive loss of motor neurons leading to loss of muscle control and muscular atrophy. The majority of SMA cases are caused by homozygous loss of *SMN1* due to gene deletion or point mutation in exon 7 of the gene or conversion into a less functional *SMN2* gene [11].

Detection of *SMN1* mutations can be simplified using the Applied Biosystems™ CarrierMax™ SMN1/SMN2 Reagent Kit that can accurately detect exon 7 copy number states for *SMN1* and *SMN2*. The kit can also detect gene conversion and haplotyping that can affect the copy number outcome by a fragment analysis–based approach. The first reaction is to detect copy number of *SMN* genes. *SMNP*, a retropseudogene with no known function located on 9p21, is employed as the endogenous control. *SMN1* and *SMNP*, but not *SMN2*, are amplified by one pair of primers. *SMN2* and *SMNP*, but not *SMN1*, are amplified by another pair of primers. These primers are incorporated in a single multiplex reaction. For more information, see **this poster**.

Epigenetic modification

DNA methylation is intricately involved in regulating human brain development, learning, and memory function. Dysregulation in the DNA methylation patterns of neural-specific genes has been shown to lead to disease development. Alzheimer's disease, for example, has been shown to be associated with differential DNA methylation patterns in genes involved in neurogenesis and neurodevelopment [12]. In another study investigating the role of methylation in behavior, using rats as the animal model, it was found that DNA methylation in the amygdala was significantly increased in animals that displayed behavior associated with anxiety and depression [13]. Interrogating the methylation status of genes of interest using fragment analysis can be used to identify and/or confirm the epigenetic regulatory mechanism involved in the neurological condition being studied. The SNaPshot fragment analysis assay or MLPA assays described above are suitable methods to assess the methylation status of discrete CpG loci involved in the development or regulation of the nervous system.

Cell line authentication and human sample matching

In addition to brain-derived cell lines, induced pluripotent stem cell (iPSC)-derived neurons are used for disease modeling and screening compounds for neurological diseases. One of the main concerns around iPSCs is the genetic drift that occurs following prolonged continuous culture such as *p53* mutations and/or spontaneous differentiation. STR profiling is essential to set a baseline for cells in culture and regularly monitor their genetic stability over time to ensure validity and reproducibility of data.

Triplet repeat analysis for fragile X mental retardation (FXMR) syndrome

Fragile X syndrome is one of the leading causes of developmental delay and one of the key targets for carrier screening in individuals planning for conception. This syndrome arises from silencing or genetic disruption of the fragile X mental retardation 1 (*FMR1*) gene by an increase of CGG repeats or point or structural mutations. The majority (99%) of detected aberrations comprise the CGG repeat expansion that increases the risk of gene silencing.

The Applied Biosystems™ CarrierMax™ FMR1 Reagent Kit is used to detect the number of CGG repeats in the *FMR1* gene. This kit uses a dual PCR system combining full-length and triplet-primed PCR (TP-PCR) amplification, followed by fragment analysis to accurately determine up to 200 CGG repeats and detect alleles that have more than 200 CGG repeats.

Infectious disease research

Pathogen detection and strain typing

Genotyping pathogenic organisms is important for tracing the source of the pathogen and preventing its spread. STR typing is a cost-effective method to identify the pathogenic isolate and determine its source. This method has been used to distinguish several clades of *Candida auris*, a pathogenic yeast known to cause outbreaks in health care facilities, by examining 12 STRs for identification [14]. It was also shown that STR typing was comparable to whole-genome sequencing for genotyping the pathogen [14]. The US Centers for Disease Control and Prevention (CDC) has developed validated protocols for fingerprinting food-borne pathogens using the multiple-locus variable number tandem repeat analysis (MLVA) technique for analyzing fragments.

Amplification of DNA sequences by PCR is a straightforward method for detecting the presence of specific strains. However, there is an increasing need to analyze many targets in a single sample. One method that can address this is fragment analysis by CE in which multiple fluorescent dyes can be used. Here, a very large number of targets can be analyzed from a single sample, since unique and specific amplicons can be separated based on both size and fluorescence in a single capillary.

To demonstrate how multiplex PCR by fragment analysis can be used for pathogen analysis, a custom panel of PCR primers able to detect 12 different respiratory RNA viruses was designed. We showed that this multiplex PCR panel, coupled with fragment analysis, serves as a powerful tool to detect several RNA viruses in a single sample with high sensitivity and specificity. Moreover, this approach can be scaled up to accommodate a large number of targets for a given sample, and it can be broadly applied to other research areas in which simultaneous detection of several targets, at the DNA or RNA level, is desired. For more information, see **this application note**.

Epigenetic modification

Certain pathogens are known to induce epigenetic modifications including DNA methylation. For instance, HIV infection has been shown to increase DNA methylation in antiretroviral treatment–naïve subjects compared to a control group of uninfected subjects, which suggests accelerated epigenetic aging [15]. DNA methylation signatures may also be useful in determining the infection status of an individual as well as predicting disease severity, as has been observed for SARS-CoV-2 infections [16]. Fragment analysis following bisulfite treatment can be used to investigate infectious disease pathogenesis caused by methylation of specific host genes as well as to confirm predictive biomarkers.

Inherited disease research

Multiplex ligation-dependent probe amplification (MLPA) analysis

MLPA analysis is a widely used molecular biology technique for copy number determination of multiple DNA sequences in the study of human genetic diseases. MLPA assays specifically designed for Duchenne and Becker muscular dystrophy have provided insights into genomic structural changes that lead to the pathology [17]. Similarly, MLPA assays have been used to analyze rearrangements that may lead to cystic fibrosis phenotypes [18]. The *BRCA1* and *BRCA2* genes are frequently mutated in hereditary breast and ovarian cancers. A deletion, when present in coding sequences of a gene, often leads to a disruption in protein function. The SALSA™ MLPA™ Probemix P002 BRCA1 Assay (MRC Holland) is widely used by clinical researchers to screen for deletions and/or duplications in the human *BRCA1* gene.

Epigenetic modification

Epigenetic inheritance, or inheritance of traits without a change to the DNA sequence, occurs through transfer of germline factors such as DNA methylation to the gametes through detrimental exposures or experiences of the individual [19]. Some examples of diseases that may be caused by epigenetic inheritance involving DNA methylation include metabolic diseases (such as diabetes) and cardiovascular diseases [20]. An epidemiological study examining the role of stress induced in Holocaust survivors showed that the DNA methylation levels of the gene encoding the FK506 binding protein 5 (*FKBP5*) was present in the genomes of both the Holocaust survivors as well as their offspring [21]. The *FKBP5* gene is involved in the body's stress response by modulating the glucocorticoid receptor [22]; genetic polymorphisms of the gene have been shown to be associated with stress-related psychiatric disorders [23]. Investigation of the role of DNA methylation in inherited diseases can be performed by examining base differences in the target gene through bisulfite treatment followed by PCR and fragment analysis.

Agricultural and crop science

SNP genotyping

The SNaPshot assay can be used to investigate several SNPs for genetics and breeding applications in animal husbandry and agriculture. For instance, the SNaPshot assay based on 4 SNP markers was developed for identification of male-sterility markers in Japanese cedar trees [24]. These trees are important for Japanese forestry but are also associated with allergy-triggering pollen that affects a sizable portion of the Japanese population. Early identification of the male-sterile trees by SNP assessment can be used for breeding trees that have desired traits. The SNaPshot assay has also been used for animal breeding purposes as a tool for identifying SNPs linked to certain traits of the animal, such as the coat color of horses [25].

STR genotyping

Microsatellite analysis has been widely used in breeding programs, parentage identification, and studying biodiversity of plants and animals of interest for agriculture or conservation purposes. For example, this method was used to investigate the impact of human and environmental factors on genetic biodiversity of the alpine meadow-grass *Poa alpina*, an important fodder grass, by assessing five microsatellite markers of the grass originating in natural and agriculturally used land [26]. It was found that the diversity of the grass was negatively impacted by human land use.

References

1. Hurst CD, Zuiverloon TC, Hafner C, Zwarthoff EC, Knowles MA. A SNaPshot assay for the rapid and simple detection of four common hotspot codon mutations in the PIK3CA gene. *BMC Res Notes*. 2009;2:66. Published 2009 Apr 29. doi:10.1186/1756-0500-2-66
2. Yamashita K, Hosoda K, Nishizawa N, Katoh H, Watanabe M. Epigenetic biomarkers of promoter DNA methylation in the new era of cancer treatment. *Cancer Sci*. 2018;109(12):3695–3706. doi:10.1111/cas.13812
3. Baretti M, Le DT. DNA mismatch repair in cancer. *Pharmacol Ther*. 2018;189:45–62. doi:10.1016/j.pharmthera.2018.04.004
4. Kang S, Na Y, Joung SY, Lee SI, Oh SC, Min BW. The significance of microsatellite instability in colorectal cancer after controlling for clinicopathological factors. *Medicine* (Baltimore). 2018;97(9):e0019. doi:10.1097/MD.0000000000010019
5. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II keynote-158 study. *J Clin Oncol*. 2020;38(1):1–10. doi:10.1200/JCO.19.02105
6. Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes. *Hum Mutat*. 2005;25(5):415–422. doi:10.1002/humu.20169
7. Hogervorst FB, Nederlof PM, Gille JJ, et al. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Res*. 2003;63(7):1449–1453.
8. Nygren AO, Ameziane N, Duarte HM, et al. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res*. 2005;33(14):e128. Published 2005 Aug 16. doi:10.1093/nar/gni127
9. Huang Y, Liu Y, Zheng C, Shen C. Investigation of cross-contamination and misidentification of 278 widely used tumor cell lines. *PLoS One*. 2017;12(1):e0170384. Published 2017 Jan 20. doi:10.1371/journal.pone.0170384
10. MacArthur CC, Pradhan S, Wetton N, et al. Generation and comprehensive characterization of induced pluripotent stem cells for translational research. *Regen Med*. 2019;14(6):505–524. doi:10.2217/rme-2018-0148
11. Singh RN, Singh NN. Mechanism of splicing regulation of spinal muscular atrophy genes. *Adv Neurobiol*. 2018;20:31–61. doi:10.1007/978-3-319-89689-2_2
12. Altuna M, Urdánoz-Casado A, Sánchez-Ruiz de Gordoa J, et al. DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clin Epigenetics*. 2019;11(1):91. Published 2019 Jun 19. doi:10.1186/s13148-019-0672-7
13. McCoy CR, Glover ME, Flynn LT, et al. Altered DNA methylation in the developing brains of rats genetically prone to high versus low anxiety. *J Neurosci*. 2019;39(16):3144–3158. doi:10.1523/JNEUROSCI.1157-15.2019
14. de Groot T, Puts Y, Berrio I, Chowdhary A, Meis JF. Development of Candida auris short tandem repeat typing and its application to a global collection of isolates. *mBio*. 2020;11(1):e02971-19. Published 2020 Jan 7. doi:10.1128/mBio.02971-19
15. Nelson KN, Hui Q, Rimland D, et al. Identification of HIV infection-related DNA methylation sites and advanced epigenetic aging in HIV-positive, treatment-naive U.S. veterans. *AIDS*. 2017;31(4):571–575. doi:10.1097/QAD.0000000000001360
16. Konigsberg IR, Barnes B, Campbell M, et al. Host methylation predicts SARS-CoV-2 infection and clinical outcome. *Commun Med* (Lond). 2021;1(1):42. doi:10.1038/s43856-021-00042-y
17. Sheikh O, Yokota T. Advances in genetic characterization and genotype-phenotype correlation of Duchenne and Becker muscular dystrophy in the personalized medicine era. *J Pers Med*. 2020;10(3):111. Published 2020 Sep 3. doi:10.3390/jpm10030111
18. Duz MB, Ozyavuz Cubuk P. Analysis of rearrangements of the CFTR gene in patients from Turkey with CFTR-related disorders: frequent exon 2 deletion. *J Hum Genet*. 2021;66(3):315–320. doi:10.1038/s10038-020-00859-w
19. Ghai M, Kader F. A review on epigenetic inheritance of experiences in humans [published online ahead of print, 2021 Nov 18]. *Biochem Genet*. 2021. doi:10.1007/s10528-021-10155-7
20. Salameh Y, Bejaoui Y, El Hajj N. DNA methylation biomarkers in aging and age-related diseases. *Front Genet*. 2020;11:171. Published 2020 Mar 10. doi:10.3389/fgene.2020.00171
21. Yehuda R, Daskalakis NP, Bierer LM, et al. Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biol Psychiatry*. 2016;80(5):372–380. doi:10.1016/j.biopsych.2015.08.005
22. Zannas AS, Wiechmann T, Gassen NC, Binder EB. Gene-stress-epigenetic regulation of FKBP5: clinical and translational implications. *Neuropsychopharmacology*. 2016;41(1):261–274. doi:10.1038/npp.2015.235
23. Häusl AS, Brix LM, Hartmann J, et al. The co-chaperone Fkbp5 shapes the acute stress response in the paraventricular nucleus of the hypothalamus of male mice. *Mol Psychiatry*. 2021;26(7):3060–3076. doi:10.1038/s41380-021-01044-x
24. Mishima K, Hirao T, Tsubomura M, et al. Identification of novel putative causative genes and genetic marker for male sterility in Japanese cedar (Cryptomeria japonica D.Don). *BMC Genomics*. 2018;19(1):277. Published 2018 Apr 23. doi:10.1186/s12864-018-4581-5
25. Rendo F, Iriondo M, Manzano C, Estonba A. Identification of horse chestnut coat color genotype using SNaPshot. *BMC Res Notes*. 2009;2:255. Published 2009 Dec 16. doi:10.1186/1756-0500-2-255
26. Rudmann-Maurer K, Weyand A, Fischer M, Stöcklin J. Microsatellite diversity of the agriculturally important alpine grass Poa alpina in relation to land use and natural environment. *Ann Bot*. 2007;100(6):1249–1258. doi:10.1093/aob/mcm203

Chapter 3: Multiplex PCR analysis

Polymerase chain reaction (PCR) is the workhorse for modern molecular biology research. It involves the amplification of a targeted segment of DNA—typically 200–1,000 bp long for standard PCR and 75–150 bp long for quantitative PCR (qPCR)—using a set of primers . Primers are oligonucleotides that hybridize with their complementary sequences on the sense and antisense strands of the DNA. They not only serve to define the region that needs to be amplified, but also provide the 3′-OH group to which the DNA polymerase adds nucleotides.

For certain applications, simultaneous detection of multiple genomic sequences is preferred. Multiplex PCR is an approach wherein a pool of primers is used to amplify several genomic targets in one reaction. For qPCR however, amplification can be technically limited by the number of distinct fluorophores that are conjugated to the detection probes (in the case of Applied Biosystems™ TaqMan™ chemistry) or the number of amplicons with different melting temperatures (in the case of Applied Biosystems™ SYBR Green™ chemistry). Methods that can circumvent these challenges and analyze multiple sequences using a simple workflow are highly desirable.

Fragment analysis is a genetic analysis method that can separate fluorescently labeled genomic products from multiplex PCR based on their color and size, with a resolution of 2 bp differences. This enables a theoretical multiplexing capability of a large number of targets from a single sample. In addition, each sample can be amplified by PCR with a different set of fluorophore-labeled dyes that allow multiple samples to be analyzed in a single well, provided that the marker ranges do not overlap. This approach can be very powerful to detect and distinguish genomic targets in a highly complex sample when optimized for PCR primers and thermocycling conditions.

Fragment analysis–based detection of multiplex PCR products is a two-step process:

- **Step 1:** Standard PCR using a pool of fluorophore-conjugated primer pairs. Here, a pool of dye-labeled primers (only one of the primers from the primer pair is conjugated to the fluorophore) is mixed with the PCR master mix and the sample, and amplified using standard or modified (depending on the primers and/or template) thermocycling conditions. The master mix including the DNA polymerase selected for this application must be suitable for multiplexing applications (see “PCR enzymes for fragment analysis” section in the Appendix).
- **Step 2:** Fragment analysis by capillary electrophoresis. The amplicons generated by PCR (in step 1) are resolved on a genetic analyzer based on their size and fluorescence. A typical reaction may include 1 µL of each PCR product, 0.5 µL of the Applied Biosystems™ GeneScan™ size standard in 9 µL of Applied Biosystems™ Hi-Di™ Formamide (for denaturing applications) or distilled, deionized water (for non-denaturing applications). Data can be analyzed on Applied Biosystems™ GeneMapper™ Software or Peak Scanner™ Software.

Depending on the primers and target template, optimization of the protocol and panel design may be required. The following guidelines are recommended:

1. **Sample input:** Any DNA source, such as genomic DNA (gDNA), complementary DNA (cDNA), or plasmid DNA can be used as an input for PCR. The sample DNA concentration can impact the robustness of the PCR and/or the peak heights observed in capillary electrophoresis. The optimal input range for PCR is 5–50 ng for gDNA and 1–20 ng for cDNA or plasmid DNA.
2. **Panel design:** Primers in the panel should be chosen for optimal discrimination of the target sequences in a given amplicon size. The size of the amplicons should be between 100 and 300 nucleotides; larger amplicons are possible but might have more variable PCR results. For best results, we recommend separating the different peaks by at least 5 nucleotides. Primer sets can be distributed across the four different fluorescent channels (FAM, VIC, NED, and PET) to increase the number of targets detectable and target resolution, if needed.
3. **Including an internal control:** An internal control can and should be included. This can be an exogenous target that is spiked into the PCR reaction before amplification. For panels detecting DNA targets, simply choose a target that will produce a unique size fragment in the same size range as targets that will be queried. For panels detecting RNA targets, an exogenous RNA, such as the Applied Biosystems™ TaqMan™ Xeno™ RNA controls, can be used as a control for reverse transcription and PCR efficiency for normalization. For these controls, it is best to design an amplicon that will produce a fragment slightly larger than the queried targets—this ensures the cDNA synthesis and PCR can cover the desired lengths.

4. **Primer cross-reactivity determination:** Primer pairs should be tested individually and against other primer pairs to ensure that there is no cross-reactivity during PCR against the target and nontarget sequences.
5. **Limits of quantitation and detection (LOQ and LOD):** Determine the LOQ and LOD for an assay by titrating the amount of input DNA or RNA in both directions where the signal starts to plateau or becomes undetectable, respectively. This establishes the range for semi-quantitative estimation of target input.
6. **Primer concentration optimization for quantitative input measurements or quantitative fluorescent PCR (qfPCR):** If the intended application is target quantitation, a standard curve using known DNA amount is required to determine the peak intensity to its corresponding input. Primer concentrations should be adjusted such that the peak intensities for the same input of any target sequence are similar for all primer/target pairs. If primer concentration adjustment does not yield the desired results, design new primers against the target and test again.

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data.

Thermo Fisher offers a complete set of tools for multiplex PCR and fragment analysis. See the Appendix for more details.

Category	Recommended products
Thermal cycler	Applied Biosystems™ 2720, Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Polymer and capillary array	See the Appendix for combination of polymer and capillary array length supported on each instrument
Size standard	Applied Biosystems™ GeneScan™ 600 LIZ™ size standards
Matrix standard	Applied Biosystems™ DS-33 Matrix Standard Kit (Dye Set G5)
Fragment analysis	Applied Biosystems™ AmpliTaq Gold™ DNA Polymerase or other reagents (see Appendix)

Applications of multiplex PCR and fragment analysis

Simultaneous detection of multiple respiratory pathogens

Identification of the etiological agent of infection in a clinical research sample is an important first step to guide the treatment approach. In complex samples containing genomic DNA from potentially several different pathogens, a multiplex PCR approach coupled to fragment analysis–based detection is very suitable. Fluorophore-conjugated PCR primers against various targets that generate amplicons distinguished by size can be designed for the detection of multiple pathogens in a given sample. To demonstrate this, we designed a set of PCR primers that could detect as many as 10 different respiratory RNA viruses including SARS-CoV-2, influenza, and respiratory syncytial virus. All forward primers were synthesized with a 6-FAM fluorophore conjugation at the 5´ end. The targets were distinguished by the different sizes of the amplicons, so the set of primers could be pooled and used to detect the pathogens in a single step. The resulting amplicons could be separated and examined using any of the genetic analyzers from the previous table. For more information see **this application note**.

In addition to pathogen detection, this approach can be broadly applied to other research areas where simultaneous detection of several targets at the DNA or RNA level is desirable. For instance, fragment analysis can be used to determine genetic aberrations such as translocation for multiple rearrangements in cancer [2]; various mutations can be screened across a particular gene by designing primers against the suspect region to determine the etiological cause of the disease [3]; and genetic polymorphisms can be identified across several variants simultaneously to predict pharmacological response to drugs (for an example, see [4]). These are just a few examples that illustrate the versatility and applications of the approach. **Learn more about fragment analysis and how you can use it to power your area of research**.

References

1. Algeciras-Schimnich A, et al. Evaluation of the PAX8/PPARG translocation in follicular thyroid cancer with a 4-color reverse-transcription PCR assay and automated high-resolution fragment analysis. *Clin Chem.* 2010;56(3): 391–398. doi:10.1373/clinchem.2009.134015
2. Furtado LV, et al. A multiplexed fragment analysis-based assay for detection of JAK2 exon 12 mutations. *J Mol Diagn.* 2013;15(5): 592–599. doi:10.1016/j.jmoldx.2013.04.006
3. Bouvet R, et al. PharmFrag: An easy and fast multiplex pharmacogenetics assay to simultaneously analyze 9 genetic polymorphisms involved in response variability of anticancer drugs. *Int J Mol Sci.* 2020;21(24). doi:10.3390/ijms21249650.
4. Taylor C, Crosby I, Yip V, Maguire P, Pirmohamed M, Turner RM. A review of the important role of CYP2D6 in pharmacogenomics. *Genes* (Basel). 2020;11(11):1295. Published 2020 Oct 30. doi:10.3390/genes11111295

Chapter 4: Microsatellite analysis

Microsatellite markers are polymorphic DNA loci that contain a repeated nucleotide sequence. Each repeat unit can be from 1 to 100s of nucleotides in length, and alleles differ by the number of repeats (Figure 4.1). The number of nucleotides per repeat unit is the same for a majority of repeats within a microsatellite locus—in other words, if a single unit is three nucleotides long, then all the units in the region will be three nucleotides long.

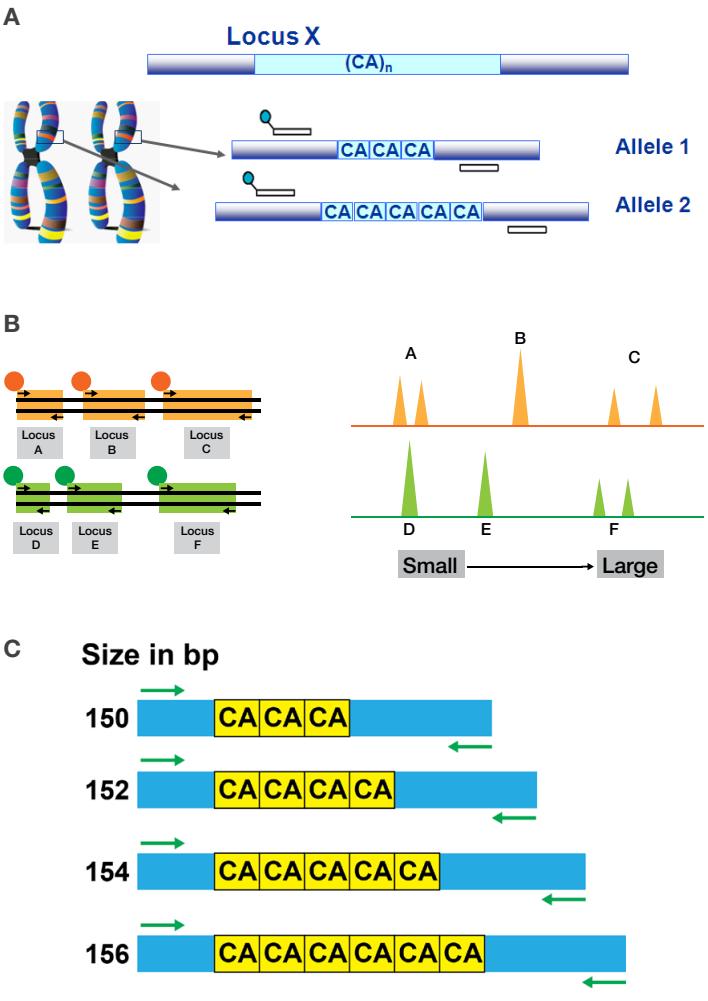


Figure 4.1. Basics of microsatellite analysis. (A) Microsatellite sequences. **(B)** Analysis of microsatellite sequences by multiplex PCR. **(C)** Different repeats lead to PCR fragments of different lengths (arrows indicate forward and reverse primers).

Microsatellite markers are also known as:

- Short tandem repeats (STRs)—each repeated unit is typically 2–7 nucleotides long
- Simple sequence repeats (SSRs)— each repeated unit is typically 1–3 nucleotides long
- Variable number tandem repeats (VNTRs)—each repeated unit is typically 10–100 nucleotides long

Note that the size associated with the different definitions is not exact. Often, the length of the repeating unit correlates with its frequency within a genome. For example, in the human genome, mononucleotide repeats are the most common form of microsatellites found, and pentanucleotide and hexanucleotide repeats are the least common [1].

In any microsatellite region, the number of repeat units can vary due to mispriming and polymerase slippage during replication. Thus, alleles of many different lengths are possible at each locus. Along with other features described below, the variability of microsatellites makes them ideal for certain types of genotyping analyses:

- Microsatellites are present in large numbers.
- Microsatellites are relatively evenly spaced throughout the genome and often physically situated near or within a gene.
- Microsatellite alleles are inherited in a Mendelian manner and can be stable over multiple generations.
- Microsatellites show a varying, but relatively high, mutation rate relative to non-microsatellite loci: the mispriming and slippage results in a 100- to 1,000-times faster mutation rate than for single-copy nuclear DNA. The mutation rate of microsatellite loci is 10^{-2} to 10^{-6} events per locus per generation [2]. The rate is believed to be dependent on the number of nucleotides in the repeated unit [3].
- Microsatellite alleles can be unique to specific populations.
- Detailed data on allelic variation, number of repeats, and allelic frequencies are widely available for a large number of microsatellite markers.

- The small size of microsatellite loci improves the chance of obtaining a result, particularly for samples containing very low amounts of DNA and/or degraded DNA.
- The small size range of microsatellite loci makes them ideal candidates for coamplification while keeping all amplified alleles smaller than 350 base pairs. Many microsatellite loci can therefore be typed from a single PCR.
- Microsatellite alleles have discrete sizes, allowing for simplified interpretation of results.

Many applications make use of microsatellite genotyping. For example, forensic analysis of molecular fingerprints based on microsatellite length variability is commonly used for human identification. Microsatellite analysis is also used for detecting certain types of enzymatic deficiencies in tumors. It is also useful for detecting pathogenic expansion of repetitive sequences in some neurodegenerative diseases.

Capillary electrophoresis is a simple and trusted method for examining microsatellite loci (Figure 4.2). PCR primer pairs are designed such that the forward and reverse primers hybridize to sequences that flank the microsatellite locus of interest. These PCR primers usually bind to the nonrepetitive, unique sequences and are typically 15–30 nucleotides long. To be detectable in a fragment analysis application, one of the primers needs to be labeled at the 5′ end with a fluorophore. By choosing primers such that different sizes and different fluorophores correspond to different loci, multiple targets can be analyzed in one reaction.

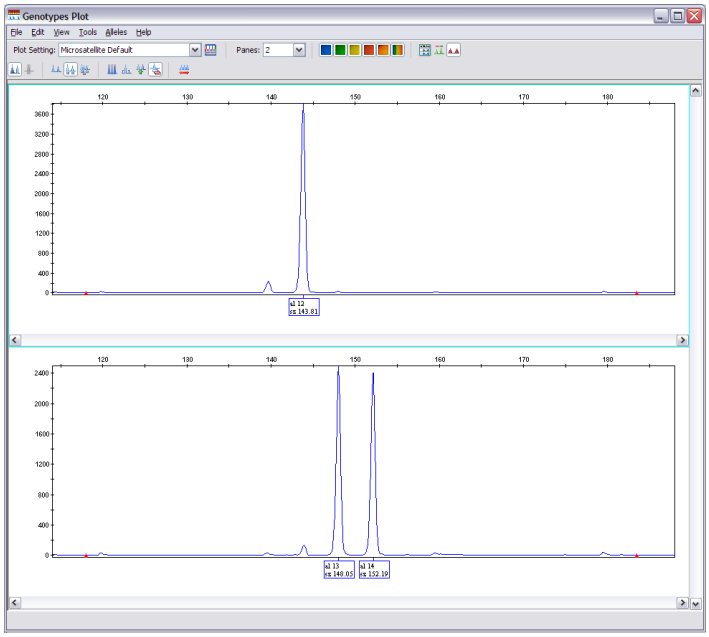


Figure 4.2. Example of microsatellite analysis of two samples by capillary electrophoresis. Samples have different genotypes as shown by the different peaks for the same marker.

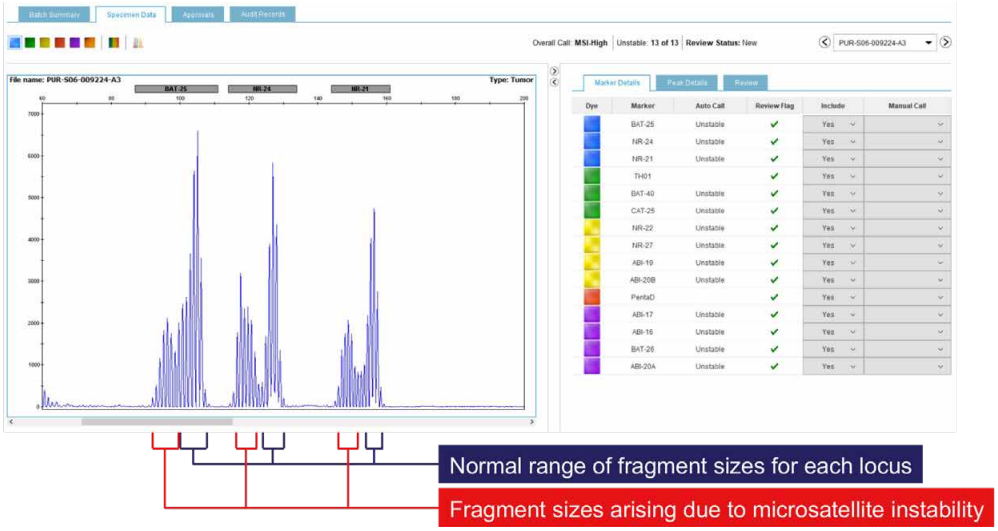


Figure 4.3. Example of results produced by TrueMark MSI Analysis Software. The TrueMark MSI Assay analyzes 13 microsatellite loci for instability, including the widely used Bethesda standards. Loci that are determined to be unstable can be autocalled; the software will then interpret the totality of the calls to make an overall call for the sample. The assay includes 2 highly variable STR sequences (THO1 and PentaD) that can be used to confirm sample identity. The proprietary algorithms used by the software do not require side-by-side analysis of normal, nontumor tissue in order to make stable and unstable calls.

Thermo Fisher offers a complete set of tools needed for microsatellite analysis. See the Appendix for more details.

Category	Recommended products
Thermal cycler for generating amplicons	Applied Biosystems™ 2720, Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Capillaries for genetic analyzer	36 cm or 50 cm for 3500 Series, 3730 Series, and SeqStudio Flex genetic analyzers; Capillary, polymer, and buffer cartridge for SeqStudio genetic analyzer
Polymers and electrode buffers	See Appendix
Size standard	Applied Biosystems™ GeneScan™ LIZ™ or ROX™ size standards
Spectral calibration matrices	See Appendix
Fragment analysis	Applied Biosystems™ AmpliTaq™ and AmpliTaq Gold™ DNA polymerases, and other PCR reagents as necessary
Software	Applied Biosystems™ GeneMapper™ Software, Peak Scanner™, and MicroSatellite Analysis Software

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data. Signal intensity may vary from instrument to instrument, capillary to capillary, etc. Fragment analysis applications on CE genetic analyzers are not designed for absolute quantitation, although relative quantitation may be assessed within a project and a specific system (see chapter 7 to learn more about loss of heterozygosity (LOH)).

Applications of microsatellite analysis

Cell line authentication and human sample matching

The research of human development and diseases relies on analysis of samples obtained and manipulated *ex vivo*. These include human cell lines, induced pluripotent stem cells (iPSCs), and chimeric antigen receptor T (CAR T) cells. Mix-ups and contamination can occur and impact research results or put lives in danger. For example, the International Cell Line Authentication Committee (ICLAC) has found that there are at least 451 cell lines being used by researchers that are misidentified with no known authentic stock [4]. Furthermore, a study from 2019 found that 32,755 articles reported results obtained with misidentified cells—these were in turn found to be cited in an estimated half a million other papers [5]. Finally, as research progresses into *ex vivo* cell therapies—that is, cells that are removed from a donor, manipulated in a lab, and returned to the host—it is critical to confirm that cells have the expected genotype and match the intended recipient. It is therefore extremely important to know the provenance of human cells and confirm that they have the expected identity.

Thermo Fisher has been the leader in providing solutions for identifying human samples by capillary electrophoresis–based fragment analysis of STRs. These are microsatellite sequences that are highly variable and provide a unique molecular fingerprint for a human sample. The Applied Biosystems™ CLA GlobalFiler™ kit generates a molecular fingerprint for 24 different STR loci, while the Applied Biosystems™ CLA IdentiFiler™ Plus kit analyzes 16 STR loci. Both easy-to-use kits are optimized for use on Applied Biosystems™ genetic analyzers. For more information, please see **this application note**.

TrueMark MSI Assay

Many types of cancer display deficiencies in mismatch repair (MMR) [6], producing an overall higher mutation rate across the genome. A higher mutation rate often means a higher rate of neoantigen production, providing opportunities for immune therapy treatments [7]. There have been at least 11 different loci implicated in MMR [8]. Looking for an inactivating event in the sequence of all these loci can be complicated, time consuming, and expensive. Instead, analyzing the functional outcome of perturbations in the MMR pathway by microsatellite instability (MSI) analysis is a practical alternative. MMR deficiencies are often detectable as changes in lengths of microsatellite sequences. However, MSI analysis is often difficult on next-generation sequencing (NGS) systems due to the highly repetitive nature of the microsatellite sequences. Fragment analysis of microsatellite loci length is therefore a widely used method for detecting MSI. The Applied Biosystems™ TrueMark™ MSI Assay was developed for microsatellite analysis. The assay analyzes a panel of 13 microsatellite loci, including the Bethesda set of markers [9]. The assay also provides STR information at two highly variable loci, allowing for sample identity confirmation.

To ease the analysis of MSI raw data, the Applied Biosystems™ TrueMark™ MSI Analysis Software was developed for analyzing results that simplifies the calls at each locus (Figure 4.3). Furthermore, we incorporated analysis algorithms that do not require side-by-side analysis of normal, nontumor tissue, reducing the number of samples needed and the expense of MSI analysis. For more information, please see this **application note**.

Note: It is recommended to use the free TrueMark MSI Analysis Software. If GeneMapper Software is used for analysis, tumor-normal pairs are required.

Microsatellite repeat expansion diseases

More than 40 different neurological diseases, including fragile X syndrome, amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), myotonic dystrophy, and Huntington's disease, are linked to microsatellite repeat expansions that are found within translated or untranslated gene regions. These regions can include 5′ and 3′ untranslated regions (UTRs), introns, and protein-coding regions [10]. The pathology of these sequences often arises when the repeat sequence, which in normal alleles can range from 15 to 40 repeats, expands to exceed a disease-specific threshold, usually greater than 45 repeats. The disease severity may worsen from generation to generation due to *de novo* germline expansion of repeats. It is therefore critical to monitor the repeat length of these microsatellite sequences, especially in individuals who are at risk.

PCR followed by fragment analysis has been used to quickly and easily identify changes in repeat length of these regions [11-13]. Applied Biosystems™ genetic analyzers have been an integral part of these analyses. In addition to providing the platform and consumables that facilitate these analyses, we have partnered with other research entities to provide solutions for investigators. For example, AmplideX™ PCR/CE reagents (Asuragen) leverage repeat-primed PCR and CE to query the microsatellite region of the *FMR1* gene for fragile X syndrome. For more details, see the white paper "**The SeqStudio Genetic Analyzer simplifies the analysis of triplet repeat expansions with AmplideX PCR/CE reagents**".

Designing a custom microsatellite analysis panel

For some studies, it may be necessary to design an assay that queries loci that are not covered by a preexisting kit. For example, a custom STR panel was used to identify family relationships among beluga whales [14]. In these cases, it is necessary to identify the relevant sequences by examining the existing scientific literature for a specific marker or cross-species marker, or by following a microsatellite development protocol [15-17]. Keep in mind that the discovery and naming of new microsatellite markers across different organisms at multiple institutions has led to inconsistent nomenclatures for microsatellite loci. For more information on nomenclature standards for specific genomes, visit **data.ccmb.res.in/msdb**.

Although the design and implementation of a custom panel is highly idiosyncratic, some general best practices should be followed:

- Design primers so the range of amplicon lengths for markers in the study is within the size-standard fragment range, with two size-standard peaks preceding the smallest fragment of interest and two size-standard peaks following the largest fragment of interest.
- Use 5′-end labeled primers. The success of microsatellite analysis depends on the ability to detect small mobility differences. The reproducible sizing and sharp peaks obtained when using the 5′-end labeling method are crucial to the success of this application.
- If you plan on performing a multiplex analysis, design primers with similar annealing temperatures (~60°C).
- Use reverse-primer tailing on one primer in each set of primers to help differentiate between peaks made by the forward and reverse DNA strands and to promote single 3′-adenine addition. For more information, see [17].
- Based on sample DNA concentration, robustness of the PCR, and/or peak heights observed in capillary electrophoresis, determine whether you need to dilute the PCR products. Dilutions can range from undiluted to 1:20 in water. You can pool the diluted PCR products if desired.
- Dye-labeled PCR products must be mixed in different ratios because each dye has a slightly different fluorescence signal strength.
- To avoid inaccuracies associated with pipetting small volumes, prepare a master mix of reagents. Prepare sufficient master mix for at least 10% overage.
- Store the master mix in the dark at 2–6°C for up to 1 month or at –15 to –25°C for longer storage time.

- After generating amplicons, prepare for electrophoresis by combining 1 µL of each PCR product and 0.5 µL of the GeneScan LIZ size standard in 9 µL of Applied Biosystems™ Hi-Di™ Formamide.
- Analyze the results using GeneMapper Software. GeneMapper Software includes a Microsatellite Default Analysis method that you can use as a starting point for analysis (Figure 4.4).

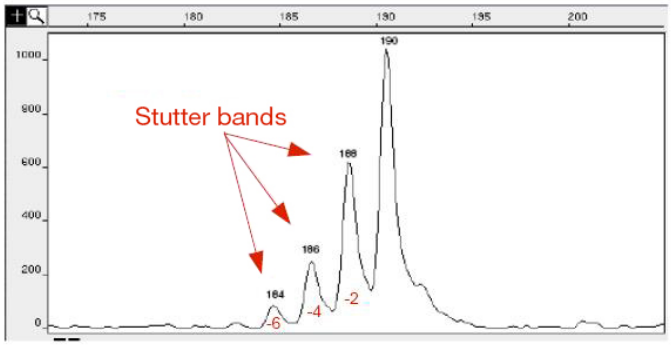


Figure 4.4. Stutter peaks in a dinucleotide repeat electropherogram. Electropherogram of a dinucleotide repeat marker from a homozygous individual (190 bp, 190 bp). The peaks at 188 bp, 186 bp, and 184 bp show the typical 2-bp stutter pattern seen with dinucleotide repeats. They represent the –2 bp, –4 bp, and –6 bp stutter peaks from the 190-bp true allele peak.

Common problems encountered with microsatellite analysis

The most commonly encountered problems during microsatellite analysis are:

- Poor or nonspecific amplification.
- Size differences of a single nucleotide. *Taq* polymerase adds a non-templated single adenosine to the 3´-blunt end of PCR products. This increases the length of the amplicon by one base pair. However, incomplete addition of these—for example, if the reaction is stopped prematurely—can introduce single-nucleotide size heterogeneity. A final, single 72°C extension step for 10 minutes up to one hour may reduce this. In addition, the single, 3'-adenine overhang can eventually get lost over time. It is therefore best to run the reactions in a timely manner. For more information, see [18].
- Stutter. During PCR amplification of di-, tri-, and tetra-nucleotide microsatellite loci, minor products that are 1 to 4 repeat units shorter than the main allele can be produced. The minor product peaks are referred to as "stutter peaks". Stutter peaks may be caused by polymerase slippage during elongation [19,20]. Stutter is expected, particularly with mono-, di-, and tri-nucleotide repeats. GeneMapper Software can handle this type of data. If two different alleles are present and heterozygous, such as A1 and A2, it may further complicate analysis with respect to proper allele calling. In addition, the use of HPLC-purified primers may assist in preventing similar "messy" data.

Understanding stutter

Stutter peaks appear as multiple smaller peaks that precede the true allele peak (Figure 4.4). These stutter peaks differ in size from the true allele peak by multiples of the length of the repeat unit. The number of peaks and their intensities are proportional to the length of the repeat and the number of repeats in the PCR product [21]. Shorter repeat units (e.g., di- or tri-nucleotide) generate more stutter, and dinucleotide repeats tend to generate more stutter peaks than trinucleotide repeats. Stutter peaks can also be caused by off-scale data.

The percent stutter can be estimated by calculating the ratio of the combined heights of the stutter peaks with the height of the true allele peak. Note the following:

- The longer the repeat unit, the less stutter product produced. For microsatellite loci with the same number of repeat units, the percent stutter is greater for dinucleotide microsatellite loci than it is for trinucleotide microsatellite loci, and so on [22]. Figure 4.5 illustrates the greater stutter in dinucleotide (left) as compared to tetranucleotide (right) repeat loci. Each locus is homozygous, with the largest peak in each figure representing the "true" allele.

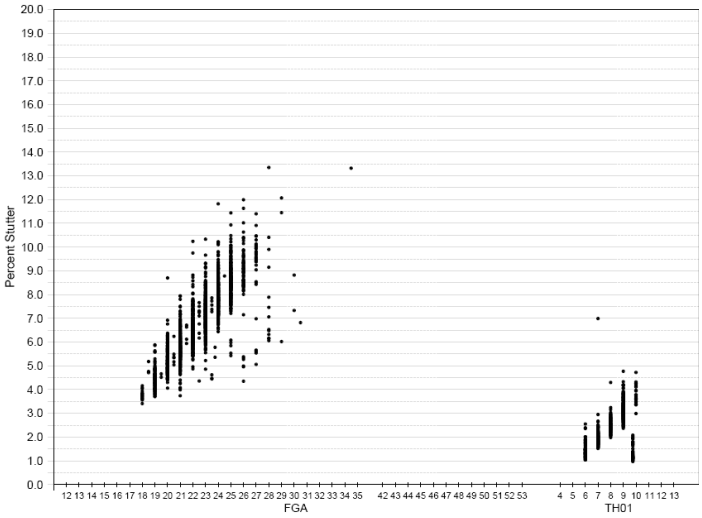


Figure 4.5. Stutter percentages for the FGA and TH01 loci. Data points indicate loci labeled with NED dye.

- The percent stutter increases with increasing allele length (i.e., increasing number of repeat units) (Figure 4.6). However, if some of the repeats are partial repeats, you may not see the proportionate increase in percent stutter.
- GeneMapper Software is optimized to attempt to filter out stutter peaks.

Evaluating data with stutter

The multipeak pattern seen for stutter can complicate analysis, particularly for samples with two or more alleles that are close in size. For example, small peaks in a position that is one-repeat unit smaller than the true allele can be interpreted either as a stutter peak or as an allele in a minor component of a mixed sample. The possible presence of stutter peaks makes precise quantitation especially important, to allow the filtering algorithm of GeneMapper Software to interpret the peak pattern accurately.

The percent stutter for a given allele is reproducible and does not depend on the quantity of input DNA or the number of loci amplified during multiplex PCR. The relative reproducibility of percent stutter is important for a few reasons:

- In many cases, the Peak Amplitude Threshold in the analysis method of GeneMapper Software can be adjusted to filter out stutter peaks and detect only true allele peaks. For more information, refer to the "GeneMapper Software Getting Started Guide: Microsatellite Analysis" (Pub. No. 4403672).
- Amplifications with abnormally high percent stutter can indicate mixed samples or some other problems with PCR amplification or electrophoresis.
- Dinucleotide repeats are particularly sensitive to stutter issues. Successful amplification of dinucleotide repeat markers yields allele peaks and associated stutter peaks within a maximum range of eight base pairs from the allele peak. In addition, the number of allele peaks depends on whether the individual tested is a heterozygote or homozygote. For examples, see Figures 4.6–4.9.

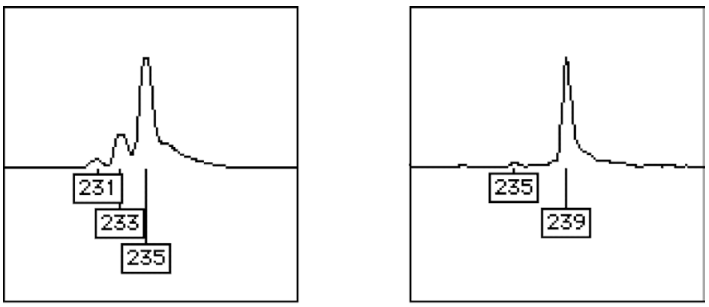


Figure 4.6. Example of greater stutter in dinucleotide (left) compared to tetranucleotide (right) repeat loci. Each locus is homozygous, with the largest peak in each figure representing the true allele.

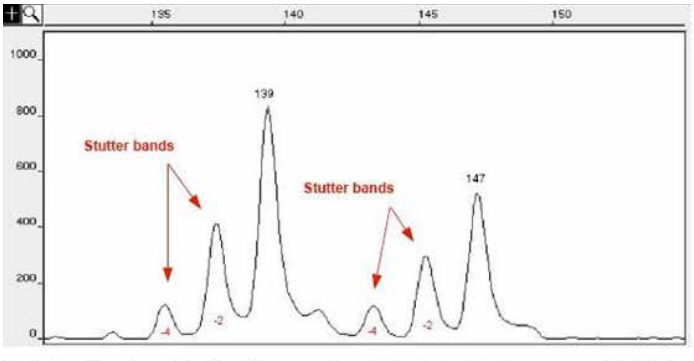


Figure 4.7. Electropherogram of a dinucleotide repeat marker from a heterozygous individual (139 bp, 147 bp). Allele sizes differ by 8 bp. The 2-bp stutter peak to the left of each allele peak is always of lower intensity than the allele peak itself. The larger 147-bp allele peak is of lower intensity than the smaller 139-bp allele. In heterozygotes, the higher molecular weight allele (i.e., the allele peak further to the right in electropherograms) often produces a fluorescence signal of lower intensity than the lower molecular weight allele, suggesting less efficient amplification of the larger fragment. This phenomenon could also be caused by preferential injection of the smaller fragments.

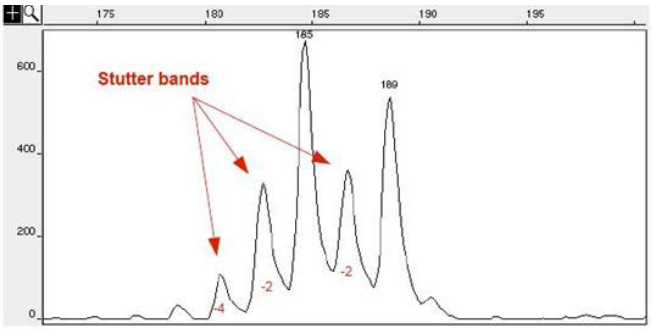


Figure 4.8. Stutter peaks in a dinucleotide repeat electropherogram from a heterozygous individual (185 bp, 189 bp). Allele sizes differ by 4 bp. When the difference between the allele sizes is ≤4 bp, a shift occurs in the height ratio between the two allele peaks (compare the Figures 4.6 and 4.7). The fluorescence signal from the –4 bp stutter of the 189-bp allele is added to the signal from the 185-bp allele.

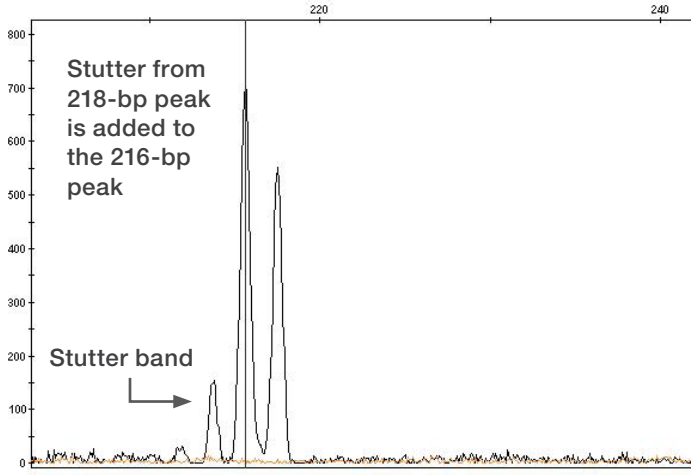


Figure 4.9. Electropherogram from a dinucleotide repeat marker of a heterozygous individual (216 bp, 218 bp). Allele sizes differ by >2 bp. The stutter peak from the 218-bp peak is added to the peak height of the 216-bp peak.

Once understood, stutter does not pose a major problem for microsatellite analysis and can aid in allele calling by:

- Distinguishing true allele peaks from nonspecific PCR products. This is because nonspecific PCR products are not associated with stutter peaks.
- Identifying alleles that fall far outside the reported allele range. The percent stutter is often specific to a particular locus. You can sometimes identify alleles that fall far outside the previously reported range based on percent stutter.

References

1. Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet.* 2004;5(6):435–445. doi:10.1038/nrg1348
2. Wan QH, Wu H, Fujihara T, Fang SG. Which genetic marker for which conservation genetics issue? *Electrophoresis.* 2004;25(14):2165–2176. doi:10.1002/elps.200305922
3. Eckert KA, Hile SE. Every microsatellite is different: intrinsic DNA features dictate mutagenesis of common microsatellites present in the human genome. *Mol Carcinog.* 2009;48(4):379–388. doi:10.1002/mc.20499
4. International Cell Line Authentication Committee (2020). Register of Misidentified Cell Lines. https://iclac.org/databases/cross-contaminations
5. Horbach SPJM, Halffman W. The ghosts of HeLa: how cell line misidentification contaminates the scientific literature. *PLoS One.* 2017;12(10):e0186281. Published 2017 Oct 12. doi:10.1371/journal.pone.0186281
6. Gologan A, Sepulveda AR. Microsatellite instability and DNA mismatch repair deficiency testing in hereditary and sporadic gastrointestinal cancers. *Clin Lab Med.* 2005;25(1):179–196. doi:10.1016/j.cll.2004.12.001
7. Zhao P, Li L, Jiang X, Li Q. Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. *J Hematol Oncol.* 2019;12(1):54. Published 2019 May 31. doi:10.1186/s13045-019-0738-1
8. Reyes GX, Schmidt TT, Kolodner RD, Hombauer H. New insights into the mechanism of DNA mismatch repair. *Chromosoma.* 2015;124(4):443–462. doi:10.1007/s00412-015-0514-0
9. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998;58(22):5248–5257.
10. Nguyen L, Cleary JD, Ranum LPW. Repeat-associated non-ATG translation: molecular mechanisms and contribution to neurological disease. *Annu Rev Neurosci.* 2019;42:227–247. doi:10.1146/annurev-neuro-070918-050405
11. Nethisinghe S, Pigazzini ML, Pemble S, et al. PolyQ tract toxicity in SCA1 is length dependent in the absence of CAG repeat interruption. *Front Cell Neurosci.* 2018;12:200. Published 2018 Jul 31. doi:10.3389/fncel.2018.00200
12. Kim SI, Kim H, Park JW, et al. Coexistence of dentatorubral-pallidoluysian atrophy and Parkinson's disease: An autopsy case report. *Neuropathology.* 2021;41(3):196–205. doi:10.1111/neup.12720
13. Gu H, Kim MJ, Yang D, et al. Accuracy and performance evaluation of triplet repeat primed PCR as an alternative to conventional diagnostic methods for fragile X syndrome. *Ann Lab Med.* 2021;41(4):394–400. doi:10.3343/alm.2021.41.4.394
14. O'Corry-Crowe G, Suydam R, Quakenbush L, et al. Migratory culture, population structure and stock identity in North Pacific beluga whales (Delphinapterus leucas). *PLoS One.* 2018;13(3):e0194201. Published 2018 Mar 22. doi:10.1371/journal.pone.0194201
15. Fleischer, RC, Lowe S (1995). Construction and screening of microsatellite-enriched genomic libraries. In J. a. P. Ferraris, S. Wiley-Liss (Ed.), Molecular Zoology: Advances, Strategies and Protocols (pp. 459–468). New York.
16. Kandpal RP, Kandpal G, Weissman SM. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proc Natl Acad Sci U S A.* 1994;91(1):88–92. doi:10.1073/pnas.91.1.88
17. Witmer PD, Doheny KF, Adams MK, et al. The development of a highly informative mouse simple sequence length polymorphism (SSLP) marker set and construction of a mouse family tree using parsimony analysis. *Genome Res.* 2003;13(3):485–491. doi:10.1101/gr.717903
18. Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques.* 1996;20(6):1004–1010. doi:10.2144/96206st01
19. Magnuson VL, Ally DS, Nylund SJ, et al. Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *Biotechniques.* 1996;21(4):700–709. doi:10.2144/96214rr03
20. Murray V, Monchawin C, England PR. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucleic Acids Res.* 1993;21(10):2395–2398. doi:10.1093/nar/21.10.2395
21. Shinde D, Lai Y, Sun F, Arnheim N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* 2003;31(3):974–980. doi:10.1093/nar/gkg178
22. Walsh PS, Fildes NJ, Reynolds R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* 1996;24(14):2807–2812. doi:10.1093/nar/24.14.2807

Chapter 5: Single-nucleotide polymorphism (SNP) genotyping

Overview of SNP genotyping

A single-nucleotide polymorphism (SNP) consists of a single base pair that varies from the known DNA sequence, thereby creating up to four alleles or variations of the marker (Figure 5.1).

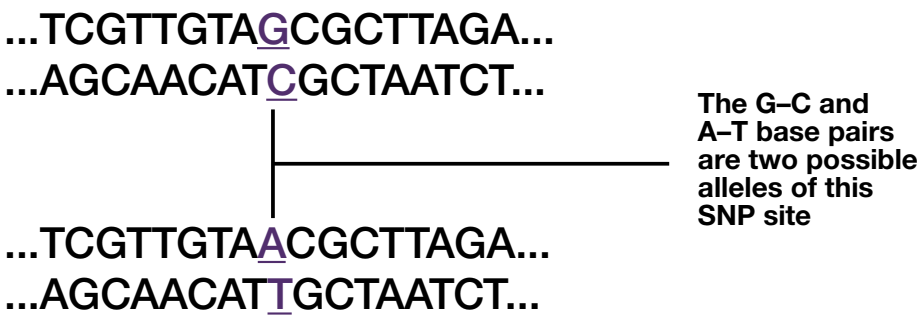


Figure 5.1. Example of a SNP.

SNPs occur in the human genome at a frequency of about 1 in every 1,000 bp, with a total number of over 10 million SNPs distributed evenly over the 3 billion bp of the human genome. They have been shown to be responsible for differences in genetic traits, susceptibility to disease, and response to drug therapies. SNPs are excellent genetic markers to construct high-resolution genetic maps.

SNPs can be genotyped by a variety of methods. Thermo Fisher products support the following fragment analysis method:

- Single-base extension (SBE)

Applications of SNP genotyping include:

- Study of mutations implicated in various cancers
- Genetic disease research
- Mitochondrial DNA investigations
- Scrapie susceptibility in sheep
- Loss of heterozygosity (LOH)
- Performance assessment in food animal production
- Differentiation of drug and non-drug forms of cannabis
- Age determination by SNP-typing certain methylated loci (forensics research)

SNaPshot Multiplex System

The Applied Biosystems™ SNaPshot™ Multiplex System investigates up to ten SNP markers simultaneously by using PCR amplification, then dideoxy single-base extension of an unlabeled primer flanking the SNP of interest, followed by capillary electrophoresis (CE). After CE, the alleles of a single marker appear as different colored peaks at roughly the same size in the electropherogram plot. The sizes of the different allele peaks will vary slightly due to differences in the molecular weight of the dyes (Figure 5.2).

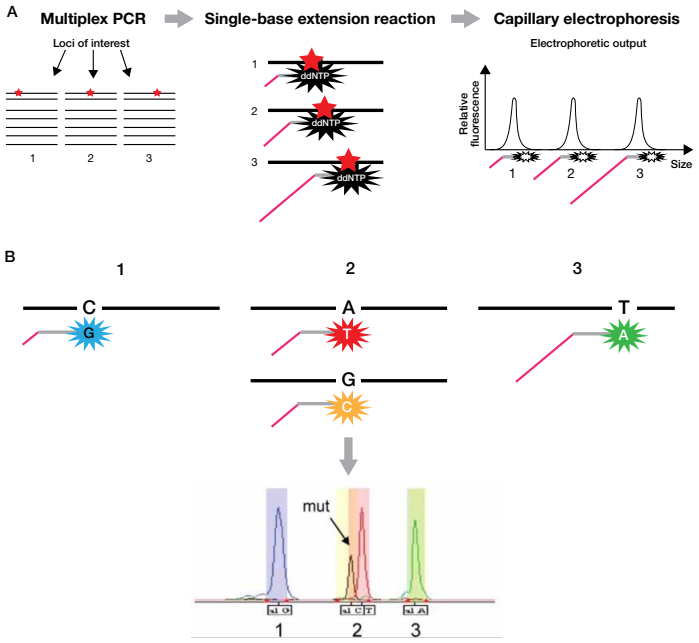


Figure 5.2. SNaPshot assay. (A) Overview of the workflow. **(B)** Single-base extension at loci of interest; each base has a distinctive color. Homozygous alleles are present as double peaks of different colors.

The components of the SNaPshot Multiplex System are:

- **SNaPshot™ Multiplex Kit**—includes SNaPshot™ Multiplex Ready Reaction Mix, control primer mix, and control template.
- **Applied Biosystems™ GeneScan™ 120 LIZ™ Size Standard**—five-dye size standard that is designed for reproducible sizing of small fragment analysis data generated with the SNaPshot Multiplex System. It accurately sizes samples ranging from 20 to 120 nucleotides (nt). When used with GeneMapper Software, the GeneScan 120 LIZ Size Standard eliminates the need for manual genotyping.
- **Applied Biosystems™ DS-02 Matrix Standard Set**—used for spectral calibration.
- **Applied Biosystems™ GeneMapper™ Software**—genotype analysis for data generated with the SNaPshot Multiplex System.

Principle of the analysis

For the single-base extension technique, an unlabeled primer is designed to anneal to the sequence adjacent to the SNP site. After the primer anneals, the single-base extension occurs by the addition of the complementary dye-labeled ddNTP (dye terminator) to the annealed primer. Each of the four ddNTPs is fluorescently labeled with a different color dye (Figure 5.3).

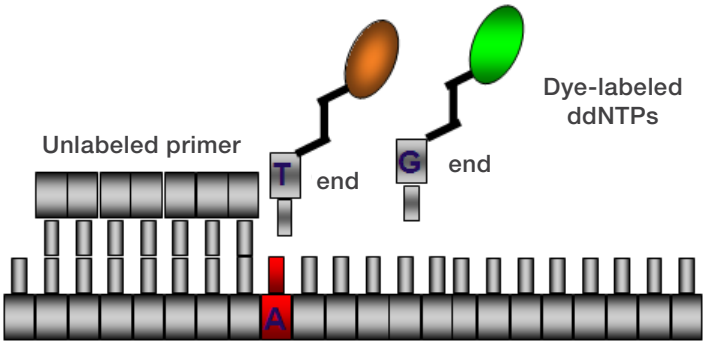


Figure 5.3. Single-base extension with dye-labeled ddNTPs.

The addition of ddNTPs generates marker fragments for the different SNP alleles that are all the same length but vary by color.

After electrophoresis and fluorescence detection, the alleles of a single marker appear as different colored peaks at roughly the same size in the electropherogram plot. The sizes of the different allele peaks will vary slightly due to differences in molecular weight of the dyes.

Advantages of the SNaPshot Multiplex System

- Cost effective—uses unlabeled (i.e., regular) oligonucleotide primers that are customized for your target
- Offers multiplexing capability—up to 10-plex, regardless of SNP positions on the chromosome or the amount of separation from neighboring SNP loci
- Sensitive allele frequency detection (5%)
- Compatible with all Applied Biosystems™ genetic analyzers
- Automated analysis using specific GeneMapper Software data analysis module

Applications of the SNaPshot Multiplex System

- Fluorescent labeling of any restriction fragment (population) with a 3´ overhang
- BAC fingerprinting (see Chapter 6, “Fingerprinting”)
- DNA methylation
- Low- to medium-throughput linkage and association studies
- Single-locus fragment analysis
- Screen and confirm SNPs
  - The SNaPshot Multiplex System includes a variety of SNaPshot Multiplex Kits used for SNP screening and validation. Each kit offers a one-tube single-base extension/termination reagent to label DNA fragments.
- Screen for a prion-encoding gene or any other mutation of interest in clinical research
  - The single–base pair sensitivity of the SNaPshot Multiplex System enables you to accurately screen samples for codon differences in prion-encoding genes. Prion diseases are caused by abnormally folded isoforms of host-encoded proteins. The SNaPshot Multiplex System can screen for SNPs in the genes that code for these proteins. For instance, polymorphisms at codons 136, 154, and 171 of the *PrP* gene in sheep and goats can lead to abnormally folded isoforms of the protein product that results in scrapie.

Thermo Fisher offers a complete set of instruments and consumables for SNP genotyping. See the Appendix for more details.

Category	Recommended products
Thermal cycler	Applied Biosystems™ 2720, Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Polymer and capillary array	Capillary, polymer, and buffer cartridge for SeqStudio genetic analyzer. See the Appendix for the polymer and capillary array length combinations supported on each instrument.
Size standard	Applied Biosystems™ GeneScan™ LIZ™ size standards
Matrix standards	Applied Biosystems™ DS-02 Matrix Standard Kit

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data.

Experiment and primer design recommendations

This is a protocol supported by Thermo Fisher.

- Minimum primer length is 23 nt; however, it is strongly recommended that primers shorter than 36 nt be tested before multiplexing.
- HPLC or PAGE purification of primers longer than 30 nt is recommended. Heterogeneous primer populations will lead to major analysis issues.
- Each primer should have 23 nt complimentary to the genomic DNA (gDNA) sequence.
- Use 5´ tails to create primers of different lengths for spatial separation of query assays of interest.
- Add a poly(dGACT) tail to generate a size difference of at least 4–6 nt.
- Primers can be complimentary to the antisense strand of the DNA if the sense strand is difficult to assay.
- Always run a negative control (no template DNA) when evaluating a new primer.

Workflow

1. Design primers
2. Design template by PCR of target, then clean up
3. Prepare SNaPshot assays
4. Post-extension by PCR, then clean up
5. Capillary electrophoresis
6. Analyze data

Data analysis

GeneMapper Software includes a default analysis method for SNaPshot assays that can be used as a starting point for analysis.

To order the SNaPshot Multiplex System, visit the **product page**.

Chapter 6: Fingerprinting

Overview

DNA fingerprinting is a technique that is used to identify patterns that occur in genetic markers. These fingerprints are specific to particular organisms. Several techniques are available for fingerprinting.

Amplified fragment length polymorphism (AFLP) analysis

Amplified fragment length polymorphism (AFLP™) analysis is a mapping technique used to visualize polymorphisms in genomic DNA (gDNA) [1]. AFLP analysis combines the restriction fragment length polymorphism (RFLP) technique and PCR to generate a large number of amplified restriction fragments from prepared gDNA. When separated by electrophoresis, the samples yield unique band patterns; when visualized by PAGE, agarose gel electrophoresis, or fluorescence-based fragment analysis, these unique band patterns can be used for high-resolution genotyping, polymorphism detection, or cladistics [2].

Principle of the analysis

AFLP analysis involves digestion of gDNA to produce a population of restriction fragments, ligation of priming sites, then amplification by PCR (Figure 6.1) [1].

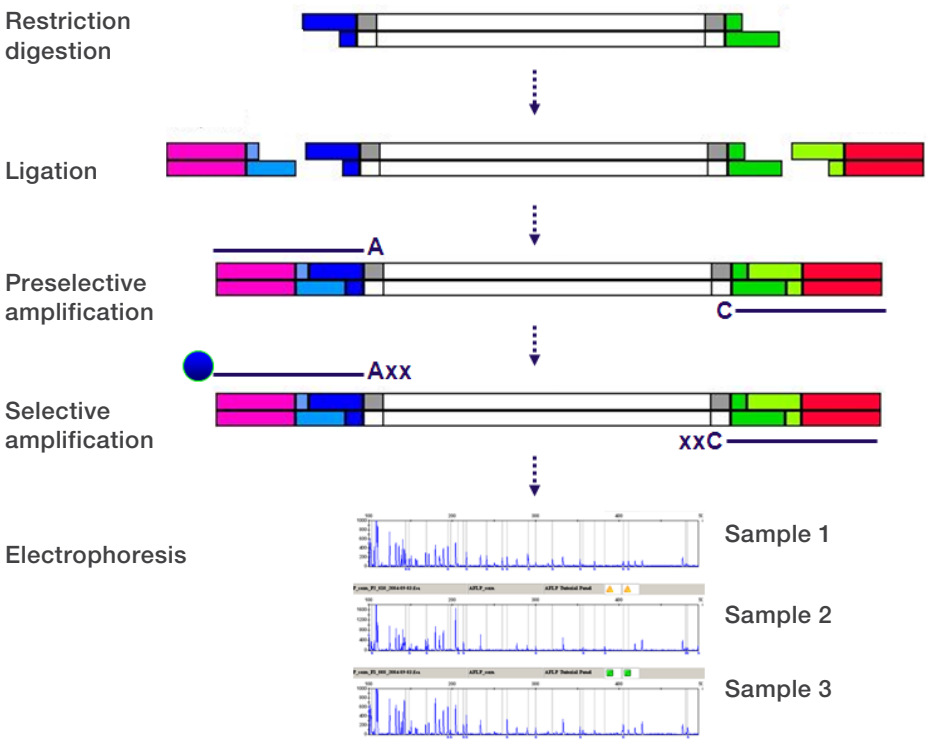


Figure 6.1. Overview of AFLP analysis.

AFLP analysis is possible because the abundant complexity in eukaryotic gDNA means that it is statistically likely that enough restriction fragments will be short enough to successfully produce PCR amplicons that yield a unique “fingerprint” profile.

Advantages of AFLP analysis

The power of AFLP analysis derives from its ability to quickly generate large numbers of marker fragments for any organism, without prior knowledge of the genomic sequence. In addition, AFLP analysis requires only small amounts of starting template and can be used for a variety of gDNA samples.

Applications of AFLP analysis

Fingerprints, or AFLP band patterns, can be used for many purposes. For example, AFLP analysis is often used in plant research where fingerprints can be compared to determine the plant variety or to compare the similarities between different plant varieties. AFLP analysis has also been used as an informative and economical research method to detect the association of variable regions of the human genome with diseases [3].

Some additional applications for AFLP analysis include:

- Molecular diversity studies [4,5]
- Phylogeny studies [6]
- Breeding [5]
- Backcross studies [5,6]
- Mapping of cloned fragments in bacterial and yeast artificial chromosomes (BACs and YACs) [7,8]
- Identifying new species or subspecies [2,5]

Thermo Fisher offered kits for performing AFLP analysis on microbes and plants as well as other organisms until 2017. While reagents for AFLP analysis are no longer available in a comprehensive kit format, it is still possible to perform AFLP analysis by ordering the required components separately from Thermo Fisher. Please note that Thermo Fisher does not provide technical support for customer-developed AFLP reagents and assays. The information is provided only to make researchers aware of this application.

Required components for the AFLP workflow.

Product	Cat. No.	Notes
Anza 11 EcoRI restriction enzyme	IVGN0116	–
Tru1I (MseI) restriction enzyme	ER0981	Tru1I is same as MseI.
T4 DNA Ligase	EL0014	–
AmpliTaq Gold 360 Master Mix	4398876 (smallest volume; larger volumes are available)	Use this reagent for convenience (2X master mix) and consistent reagent formulation.
Adaptor primers	Custom oligo synthesis (see “Fluorescent DNA labeling methods” in the Appendix)	5´ end of upper strand oligo of adaptor needs to be modified with phosphate group (for ligation). For sequences, see Vos et al.
Unlabeled selective primers (MseI+N..)	Custom oligo synthesis (see “Fluorescent DNA labeling methods” in the Appendix)	For sequences, see Vos et al. and AFLP user guide.
5´ fluorescent-labeled selective primers (EcoRI adaptor+N..)	Custom oligo synthesis (see “Fluorescent DNA labeling methods” in the Appendix)	Substitute 6-FAM for 5-FAM dye (original protocol). Substitute VIC for JOE dye (original protocol). For sequences, see Vos et al. and AFLP user guide

The protocols can be downloaded from the link below. The primer sequences and molar concentrations are described in the original publication by Vos et al. [1]. For a comparative microbiological AFLP application, see [9].

- **AFLP microbial fingerprinting protocol** (associated kit is no longer available)
- **AFLP plant mapping protocol** (associated kit is no longer available)

Note the suggested substitutions of 5-FAM with 6-FAM dye and JOE with VIC dye (see table above).

Thermo Fisher offers a complete set of instruments and consumables for customer-developed protocols for AFLP analysis. See the Appendix for more details.

Category	Recommended products
Thermal cycler	Applied Biosystems™ 2720, Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Polymer	See the Appendix for the polymer and capillary array length combinations supported on each instrument
Capillary array	50 cm for 3500 Series and SeqStudio Flex systems; 36 cm for 3730 Series systems
Size standard	Applied Biosystems™ GeneScan™ ROX™ size standards
Matrix standards	Applied Biosystems™ DS-31 Matrix Standard Kit

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data.

Experiment and primer design recommendations

DNA extraction and purification

Because AFLP analysis requires only a small amount of DNA (10–100 ng), DNA purification is critical.

We recommend the following kits for extracting DNA for AFLP analysis:

- **Plant analysis:** Invitrogen™ DNAzol™ Reagent or Invitrogen™ PureLink™ Genomic Plant DNA Purification Kit
- **Microbial analysis:** Invitrogen™ PureLink™ Genomic DNA Mini Kit

Restriction

For AFLP analysis on genomes of unknown content, determine whether your gDNA is digested properly with EcoRI and MseI enzymes.

Empirical guidelines suggest that if the GC-content of the genome is >65%, restriction digest with MseI will not give a significant number of fragments. Optimal results are obtained with MseI when the GC-content is <50%. EcoRI also tends to produce more fragments in GC-poor genomes. In cases where an organism's GC-content is unknown, the effectiveness of the restriction enzymes must be determined empirically.

Primers

For the selective amplification step, the primers that target the EcoRI cut site are fluorescently labeled at the 5′ end. The primers that target the MseI cut site are unlabeled.

You may need to optimize to identify primer combinations that generate sufficient unique marker fragments for a study.

The appendix of the AFLP plant mapping protocol (**Pub. No. 4303146**) shows primer combinations that have been successfully used for a variety of plant species, and the AFLP microbial fingerprinting protocol (**Pub. No. 402977**) shows primer combinations that have been successfully used for a variety of microbial organisms. Note that if your organism of interest does not appear in the list, you can still conduct experiments by choosing primers from the most closely related species that is available.

In general, the strategy for AFLP analysis is to generate informative fragments or enough fragments so that individuals are distinguishable. However, too many fragments complicate the analysis, so you must empirically determine the optimal number of fragments needed for adequate discrimination. As a general rule, it is best to have 50–200 peaks as the "fingerprint" after amplification.

Workflow

1. Restriction digestion
2. Ligation
3. Preselective amplification
4. Selective amplification
5. Capillary electrophoresis
6. Data analysis

Data analysis

The GeneMapper Software includes a default method for AFLP analysis that you can use as a starting point for analysis.

This method contains analysis parameters for pattern recognition of fragments across samples to generate a fingerprint for every sample. This method can be used for analysis of any type of data from fragment length polymorphism assays such as AFLP analysis or T-RFLP analysis. Features useful for AFLP analysis include:

- Ability to generate a panel (the collection of markers) from sample files that have been added to a project.
- Sizing quality and genotyping quality values flag poor quality samples, enabling easy identification and decreasing manual review.
- Automatic generation of final marker genotypes in a standard binary format where “1” represents the presence of a given fragment while “0” represents the absence of the corresponding fragment.

Up to four profiles are expected for each sample because:

- Both the forward and reverse PCR primers may be fluorescently labeled.
- Two restriction enzymes are used.

Using the default method for AFLP analysis will generate panels and binsets. You can then routinely analyze data using this panel.

A change in the fragment profile is indicated by the absence of a peak as well as a reduction in the height of a peak when comparing different samples.

The following two figures are examples of AFLP analysis on typical (Figure 6.2) and polymorphic (Figure 6.3) reactions.

These peak patterns are automatically converted to a table of binary marker genotypes (Figure 6.4), which can be exported and analyzed for similarity and generation of dendrograms using a statistical software package or other downstream analysis software for this type of clustering analysis.

For file conversion of genotype tables to other downstream analysis formats, see [10].

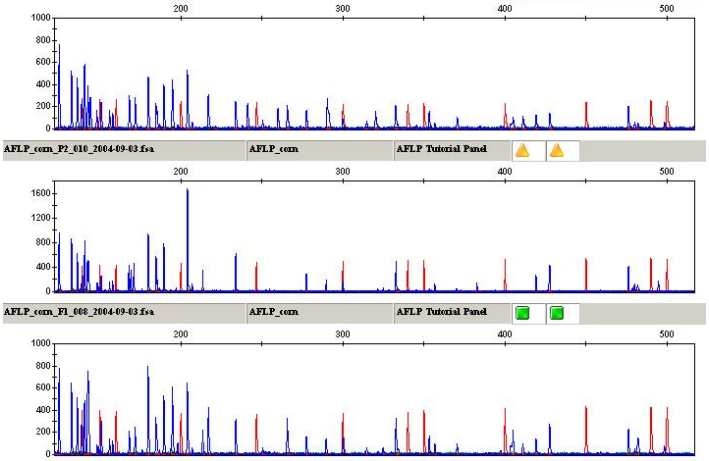


Figure 6.2. Electropherogram of a typical AFLP reaction.

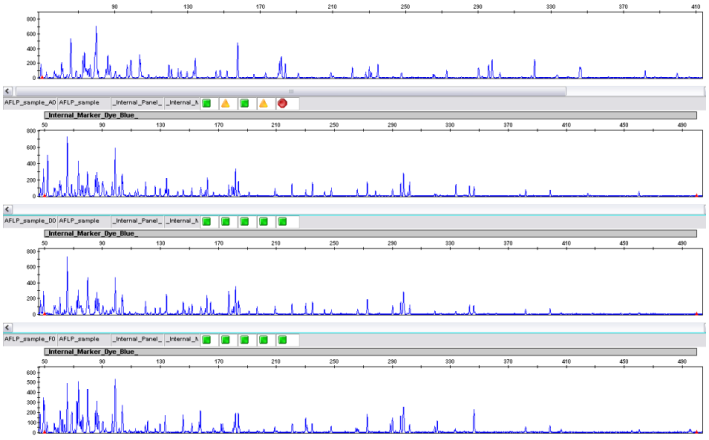


Figure 6.3. Electropherogram of a polymorphic AFLP reaction.

Samples | Genotypes

	Sample File	Panel	Dye	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Allele 9	Allele 10	Allele 11	Allele 12	Allele 13	Allele 14	Allele 15	Allele 16	Allele 17	Allele 18
1	AFLP_corn_F1_	AFLP Tutorial Pan	B	0	1	0	1	0	0	0	0	0	0	1	1	0	1	1	0	0	1
2	AFLP_corn_P1_	AFLP Tutorial Pan	B	1	1	0	0	1	0	0	1	1	0	1	0	0	0	1	1	1	1
3	AFLP_corn_P2_	AFLP Tutorial Pan	B	1	0	1	1	1	1	1	0	1	1	0	0	1	1	0	0	0	0

Figure 6.4. AFLP genotypes in GeneMapper Software.

Terminal restriction fragment length polymorphism (T-RFLP) analysis

Overview

Terminal restriction fragment length polymorphism (T-RFLP) analysis is a mapping technique used to study complex microbial communities based on variation in the 16S rRNA gene [11] or other species-relevant targets and markers of interest. T-RFLP analysis is culture-independent, rapid, sensitive, reproducible, and does not require genomic sequence information other than the primer sequences of the target region(s).

Principle of the analysis

For T-RFLP analysis, fluorescently labeled DNA (i.e., a pool of multiplexed PCR amplicons or a single amplicon) is digested with restriction enzymes. This step generates terminal restriction fragments that are fluorescently labeled. The fragments in the digest are then separated by capillary electrophoresis. Profiles can then be compared between samples or matched to a database of known species (Figure 6.5).

Applications of T-RFLP analysis

- Examination of microbial community structure and community dynamics in response to changes in different environmental parameters or to study bacterial populations in natural habitats
- Study of complex microbial communities in diverse environments such as soil [12], activated sludge [13], and marine [14] systems
- Characterization of oral bacterial flora in saliva from healthy and periodontitis samples [15]
- Food and beverage testing: multiplex T-RFLP test to detect and identify *Salmonella enterica* and all six species of *Listeria* inoculated into milk at minimal levels [16]
- Sensitive and high-throughput assessment of bacterial diversity patterns in human anterior nares [17]
- Preliminary screening of microorganisms before analysis using Applied Biosystems™ MicroSEQ™ microbial identification kits

Thermo Fisher offers a complete set of instruments and consumables for T-RFLP analysis. See the Appendix for more details.

Category	Recommended products
Thermal cycler	Applied Biosystems™ 2720, Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Polymer and capillary array	See the Appendix for the polymer and capillary array length combinations supported on each instrument
Size standard	Applied Biosystems™ GeneScan™ LIZ™ size standards
Matrix standards	Applied Biosystems™ DS-33 Matrix Standard Kit

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data.

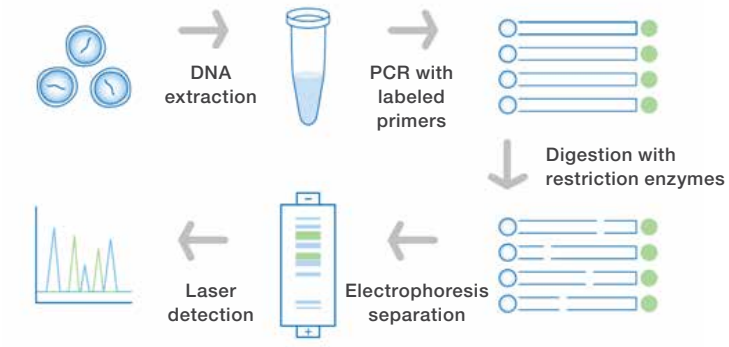


Figure 6.5. Principle of T-RFLP analysis.

Workflow

1. DNA isolation and purification
2. PCR amplification and restriction enzyme digestion
3. Separation and detection of the digested products via electrophoresis
4. Analysis of data to generate the fragment profile for each sample
5. Clustering analysis based on the profile of samples from step 4

Data analysis

T-RFLP analysis uses the same data analysis technique as AFLP analysis.

For more information

An application on T-RFLP analysis can be seen **here**.

Labeling of restriction fragments and bacterial artificial chromosome (BAC) fingerprinting

Overview

Any DNA restriction fragment (RF) of interest or population of RFs can essentially be fluorescently labeled and mapped based on the first nucleotide incorporated on a 3´ DNA overhang. For example, bacterial artificial chromosome (BAC) fingerprinting provides an efficient and cost-effective method of characterizing large genomic fragment libraries for genome sequencing, positional cloning, and physical mapping efforts. Restriction enzyme digestion of BAC clones followed by fluorescent-dye labeling can be used to generate a profile or fingerprint. Overlap between fingerprints are subsequently used to assemble contiguous sequences (contigs) in the construction of whole-genome physical maps. Physical maps are important resources for genome sequencing efforts, positional cloning, comparative genomics, and determining the size and structure of genomes.

The Applied Biosystems™ SNaPshot™ Multiplex Kit provides an easy and cost-effective method for high-throughput BAC fingerprinting of any RF (or population of RFs) of interest [18].

Principle of the analysis

During BAC fingerprinting analysis using the SNaPshot Multiplex Kit, BAC clones are subjected to restriction-enzyme treatment to generate fragments of various lengths that end in A, C, G, or T (Table 6.1). The fragments are then labeled with the corresponding bases by single-base extension to create a distinct DNA fragment pattern, or "fingerprint", for each clone (Figure 6.6). The clones are then mapped based on the order of the overlapping parts of "fingerprints" with other clones of the same genome [19,20].

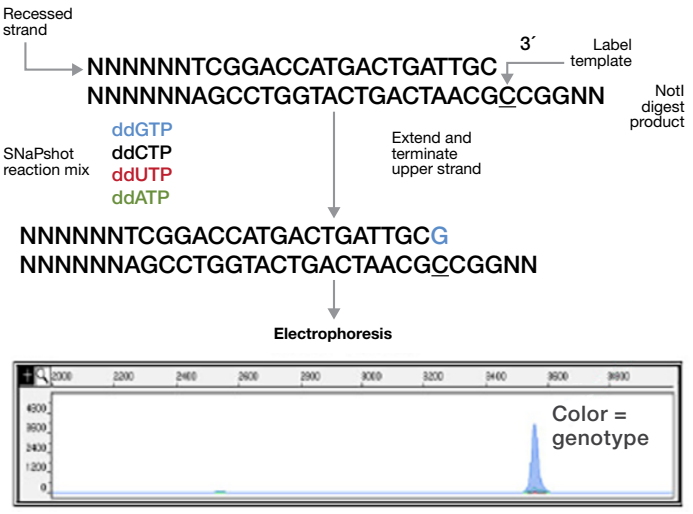


Figure 6.6. Fragment labeling using the SNaPshot Multiplex Kit.

Table 6.1. Examples of possible restriction enzymes (5-base cutters) and dyes used in the SNaPshot Multiplex Kit. Note that blunt ends, such as those created by HaeIII, cannot be labeled since there is no 3´ overhang.

Restriction enzyme	Restriction site	ddNTP	Fluorescent dye	Restriction fragment color
EcoRI	G–AATTC	A	dR6G	Green
BamHI	G–GATCC	G	dR110	Blue
XbaI	T–CTAGA	C	dTAMRA	Yellow
XhoI	C–TCGAG	T	dROX	Red
HaeIII	GG–CC	None	–	–

Applications

With BAC fingerprinting, you can create whole-genome physical maps that are important resources for:

- Genome sequencing
- Positional cloning
- Comparative genomics

Note: BAC fingerprinting is based upon pattern recognition; therefore, data analysis is focused on relative size and distribution. We recommend using a dedicated instrument platform to minimize low random error caused by sizing imprecision.

Thermo Fisher offers a complete set of instruments and consumables for BAC fingerprinting. See the Appendix for more details.

Category	Recommended products
Thermal cycler	Applied Biosystems™ Veriti™, ProFlex™ thermal cyclers
Genetic analyzer	Applied Biosystems™ 3500 Series, 3730 Series, SeqStudio™, SeqStudio™ Flex genetic analyzers
Polymer	See the Appendix for the polymer and capillary array length combinations supported on each instrument
Capillary array	50 cm for 3500 Series systems; 36 or 50 cm for 3730 Series systems
Size standard	Applied Biosystems™ GeneScan™ LIZ™ size standards
Matrix standards	Applied Biosystems™ DS-02 Matrix Standard Kit

Note: Throughout a set of experiments, use the same equipment, run parameters, polymers, dyes, etc. Consistent conditions are required to avoid mobility shifts that interfere with accurate interpretation of data.

Experiment and primer design recommendations

This is a protocol demonstrated by Thermo Fisher.

- Protocols may differ based on the kind of restriction enzymes and the BAC DNA purification kits that are used.
- Enzymatic digestion and labeling using the SNaPshot kit can be performed in one tube or in separate reactions.

Workflow

Figure 6.7. Workflow for BAC fingerprinting.

A detailed application note of the BAC fingerprinting workflow using the SNaPshot kit can be seen **here**.

A detailed application note on sizing BAC clone fragments can be seen **here**.

High-coverage expression profiling (HiCEP)

Overview

The high-coverage expression profiling (HiCEP) method of fragment analysis was developed to address the shortcomings in gene expression profiling and to provide a sensitive method for detecting a large proportion of transcripts in both known and unknown genes, with a low false-positive rate.

As an AFLP analysis–based gene expression profiling method, HiCEP does not require sequence information and has a reduced rate of false positives with a high degree of detection of both coding and noncoding transcripts. After HiCEP analysis, fragments of interest can be purified and cloned from agarose gels and sequenced to identify the transcripts. If whole-genome sequence information for the organism under study is known, the fragments of interest can be identified by bioinformatic prediction using the sequence information available from public databases and the restriction enzyme recognition sites used in the HiCEP workflow.

Principles of the analysis

HiCEP is an AFLP analysis–based method. The analysis involves digesting cDNA to produce a population of restriction fragments. Priming sites are then ligated onto the ends of the restriction fragments so that they can be amplified by PCR [6].

Application

Fingerprinting

Recommendations

This is a protocol demonstrated by a customer.

Workflow

A detailed workflow of HiCEP analysis can be seen in Figure 6.8.

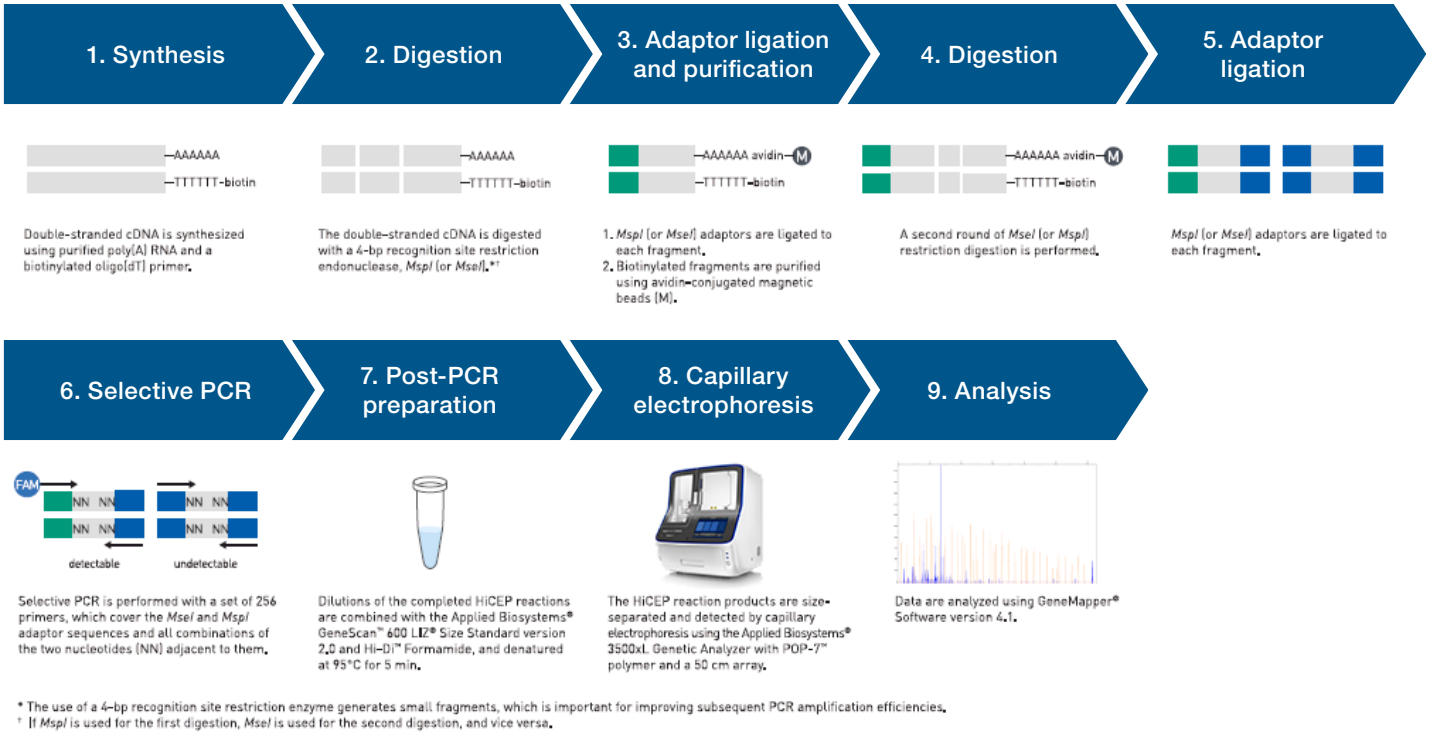


Figure 6.8. Workflow for HiCEP analysis.

Inter-simple sequence repeat (ISSR) PCR

Overview

Inter-simple sequence repeat (ISSR) PCR is a fast and inexpensive genotyping technique with a wide range of uses, including the characterization of genetic relatedness among populations. ISSR PCR is a genotyping technique based on variation found in the regions between microsatellites.

In addition to the use of long fragments for accurate analysis, this technique provides additional benefits over agarose gels. The higher sensitivity of Applied Biosystems™ genetic analyzers over traditional analysis methods routinely allows the detection of more peaks, and this increased resolution results in better discrimination between individuals being compared in populations.

However, the primers that are designed to anneal to the di- or tri-nucleotide repeats can lack specificity in PCR and are a major contributor to a lack of reproducibility. Also, the lack of complexity of ISSR primers can lead to nonspecific amplification, particularly if coupled with poor-quality gDNA extraction methods and suboptimal PCR amplification conditions.

Principle of ISSR PCR

ISSR PCR uses a single fluorescently labeled primer to target the region between identical microsatellites (Figure 6.9). An ISSR PCR primer comprises three parts:

- A fluorescent tag
- Eight dinucleotide repeat units (or 6 trinucleotide repeat units)
- One or more anchor nucleotides designed with a dual purpose: to target the end of a microsatellite region and to prevent primer dimerization

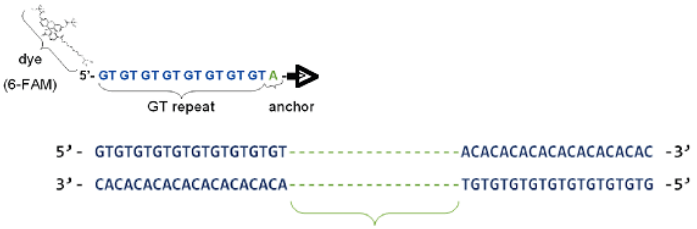


Figure 6.9. Example of an inter-simple sequence repeat (ISSR) region.

Because ISSRs are dominant markers, the amplified regions in an ISSR PCR are scored as diallelic. Between individuals within a population, changes in the amplified products can arise through structural changes to the region (insertions or deletions) or the loss of primer binding sites.

Advantages

- Faster and requires a lower start-up investment than other genotyping methodologies such as AFLP and RFLP analysis
- Several studies have compared results from AFLP and ISSR analysis and have found ISSR PCR to be preferable because of the reduced number of protocol steps required and less DNA consumed
- Less expensive and less time-consuming than microsatellite-based genotyping
- No need to clone and characterize microsatellites
- Capillary electrophoresis delivers significantly higher resolution than traditional agarose gel electrophoresis, thus increasing the amount of information obtained from each experiment

Applications

ISSR PCR has been used to investigate many plant and animal species using the following techniques:

- Genetic fingerprinting [21]
- Gene tagging [22]
- Detection of clonal variation [23]
- Cultivar identification [24]
- Phylogenetic analysis [25]
- Detection of genomic instability [26]
- Assessment of hybridization [27]

The versatility of this genotyping technique makes ISSR PCR useful for researchers interested in diverse fields such as conservation biology and cancer research.

Considerations for experiment and primer design

- Cetyltrimethyl ammonium bromide (CTAB) gDNA isolation delivers high and consistent amplification [28].
- For DNA amplification, primers and PCR master mixes should be tested for robustness and consistency when amplifying ISSR targets in both types of specimens. Subsequently thermal cycling conditions can be refined, with particular focus on primer annealing temperature and primer annealing time.

Workflow

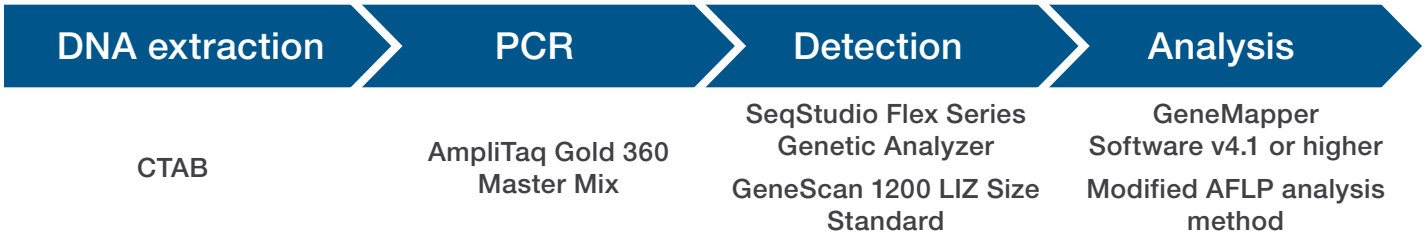


Figure 6.10. Workflow for ISSR analysis.

Data analysis

Using GeneMapper Software:

- In the panel manager, create a panel for each dye color (primer) with bins, centered at whole base pairs, one base pair wide covering the entire range of 80–1,200 bp (Figure 6.11).
- In the panel manager, modify the AFLP analysis method (Figure 6.12).
- This method detects peaks above a minimum peak height as an allele and applies a binary label of either 1 or 0 for the presence of a peak in a particular bin.

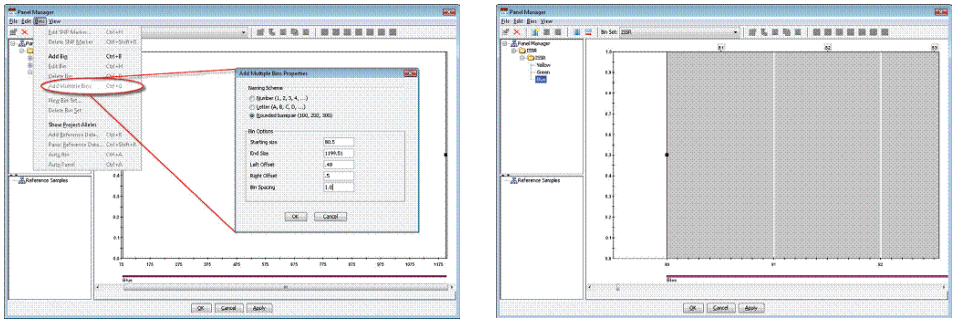


Figure 6.11. Creating multiple ISSR bins and example of multiple bins centered at whole base pairs for the blue marker in an ISSR panel.

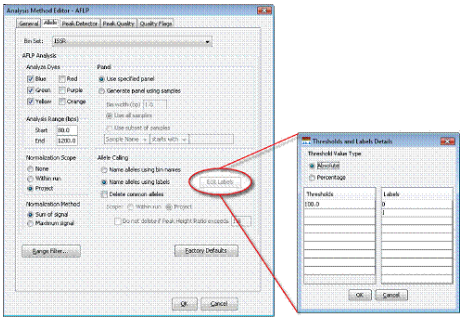


Figure 6.12. Example of the allele tab settings for an ISSR analysis method.

Example analysis

ISSR analysis was used to compare two samples: *Agave shawii shawii* (also known as Shaw's agave) from Rosarito, Mexico, and *Agave shawii shawii* from the US/Mexico border. Figure 6.13 shows the distinct peak patterns of these two samples.

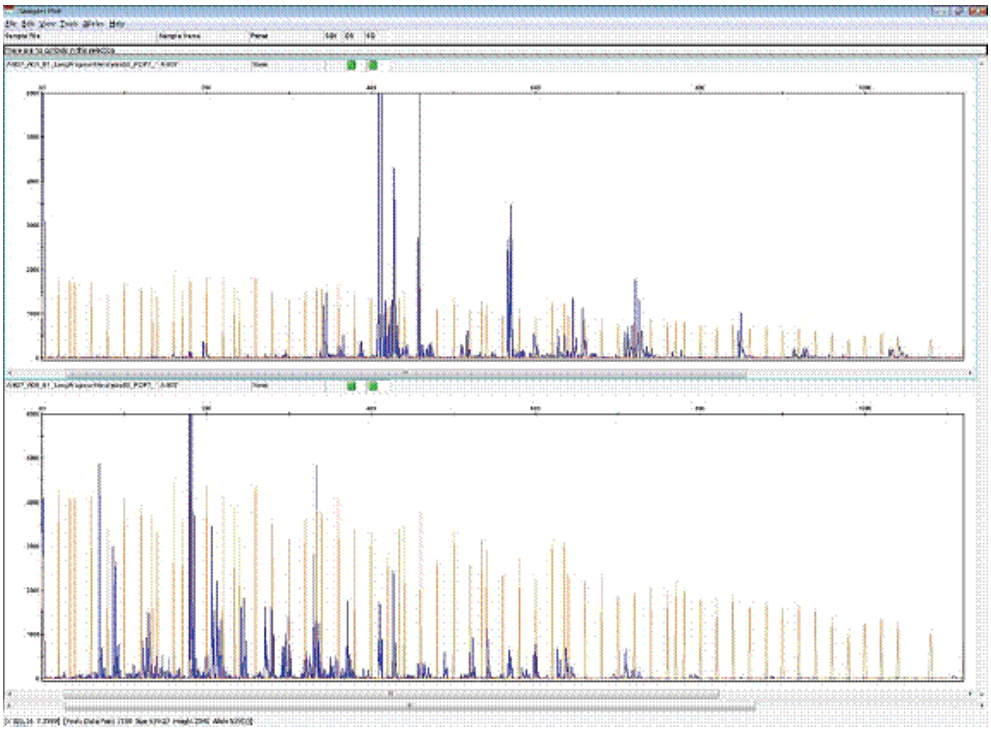


Figure 6.13. ISSR analysis of two *Agave shawii shawii* samples, from Rosarito, Mexico (top) and the US/Mexico border (bottom).

After analyzing the data in GeneMapper Software, genotypes were exported and evaluated using a spreadsheet program to:

- Assess the consistency of genotyping for four replicate ISSR polymerase chain reactions for each primer analyzed.
- Calculate the alleles shared between the replicates.

Only alleles with 100% concordance were scored as true alleles and used in subsequent phylogenetic analyses.

True allele data for each individual for each primer were concatenated into a single list of binary states. The binary data were then analyzed using the phylogenetic software, MrBayes program [29,30]. A phylogram was generated using the MrBayes program (Figure 6.14). It indicates with high confidence that three distinct populations of *Agave shawii shawii* existed.

For more information, view the application note **here**.

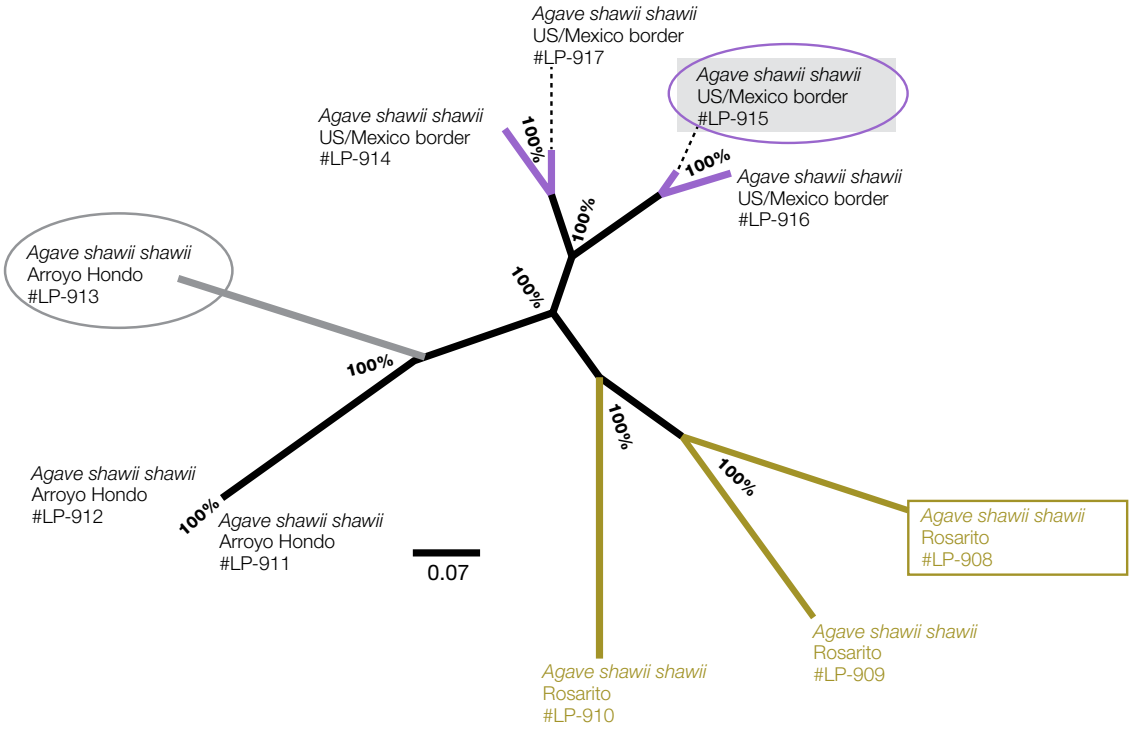


Figure 6.14. Phylogram generated using MrBayes software shows three distinct populations of agave. Individuals collected from Rosarito, Mexico; Arroyo Hondo, New Mexico, US; and the US/Mexico border are shown in gold, gray, and purple, respectively. Highlighted individuals correspond to the data presented in Figure 6.13. Nodes with posterior probability values above 95% are considered to be informative in Monte Carlo Markov Chain (MCMC) and Bayesian analysis.

References

1. Vos P, Hogers R, Bleeker M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 1995;23(21):4407–4414. doi:10.1093/nar/23.21.4407
2. Savelkoul PH, Aarts HJ, de Haas J, et al. Amplified-fragment length polymorphism analysis: the state of an art. *J Clin Microbiol.* 1999;37(10):3083–3091. doi:10.1128/JCM.37.10.3083-3091.1999
3. Koblihova J, Srutova K, Krutska M, Klamova H, Machova Polakova K. AFLP-AFLP in silico-NGS approach reveals polymorphisms in repetitive elements in the malignant genome. *PLoS One.* 2018;13(11):e0206620. Published 2018 Nov 8. doi:10.1371/journal.pone.0206620
4. Zhao H, Bughrara SS, Oliveira JA. Genetic diversity in colonial bentgrass (Agrostis capillaris L.) revealed by EcoRI-MseI and PstI-MseI AFLP markers. *Genome.* 2006;49(4):328–335. doi:10.1139/g05-113
5. Kandpal RP, Kandpal G, Weissman SM. Construction of libraries enriched for sequence repeats and jumping clones, and hybridization selection for region-specific markers. *Proc Natl Acad Sci U S A.* 1994;91(1):88–92. doi:10.1073/pnas.91.1.88
6. Goel S, Chen Z, Akiyama Y, et al. Comparative physical mapping of the apospory-specific genomic region in two apomictic grasses: Pennisetum squamulatum and Cenchrus ciliaris. *Genetics.* 2006;173(1):389–400. doi:10.1534/genetics.105.054429
7. Serra R, Cabañes FJ, Perrone G, et al. Aspergillus ibericus: a new species of section Nigri isolated from grapes. *Mycologia.* 2006;98(2):295–306. doi:10.1080/15572536.2006.11832702
8. Naimuddin M, Nishigaki K. Genome analysis technologies: towards species identification by genotype. *Brief Funct Genomic Proteomic.* 2003;1(4):356–371. doi:10.1093/bfgp/1.4.356
9. Bertani G, Savo Sardaro ML, Neviani E, Lazzi C. AFLP protocol comparison for microbial diversity fingerprinting. *J Appl Genet.* 2019;60(2):217–223. doi:10.1007/s13353-019-00492-0
10. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics.* 2012;28(2):298–299. doi:10.1093/bioinformatics/btr642
11. Osborn AM, Moore ER, Timmis KN. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol.* 2000;2(1):39–50. doi:10.1046/j.1462-2920.2000.00081.x
12. Derakshani M, Lukow T, Liesack W. Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl Environ Microbiol.* 2001;67(2):623–631. doi:10.1128/AEM.67.2.623-631.2001
13. Eschenhagen M, Schuppler M, Röske I. Molecular characterization of the microbial community structure in two activated sludge systems for the advanced treatment of domestic effluents. *Water Res.* 2003;37(13):3224–3232. doi:10.1016/S0043-1354(03)00136-2
14. Danovaro R, Luna GM, Dell'anno A, Pietrangeli B. Comparison of two fingerprinting techniques, terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analysis, for determination of bacterial diversity in aquatic environments. *Appl Environ Microbiol.* 2006;72(9):5982–5989. doi:10.1128/AEM.01361-06
15. Sakamoto M, Takeuchi Y, Umeda M, Ishikawa I, Benno Y. Application of terminal RFLP analysis to characterize oral bacterial flora in saliva of healthy subjects and patients with periodontitis. *J Med Microbiol.* 2003;52(Pt 1):79–89. doi:10.1099/jmm.0.04991-0
16. Elliott GN, Thomas N, Macrae M, Campbell CD, Ogden ID, Singh BK. Multiplex T-RFLP allows for increased target number and specificity: detection of Salmonella enterica and six species of Listeria in a single test. *PLoS One.* 2012;7(8):e43672. doi:10.1371/journal.pone.0043672
17. Camarinha-Silva A, Wos-Oxley ML, Jáuregui R, Becker K, Pieper DH. Validating T-RFLP as a sensitive and high-throughput approach to assess bacterial diversity patterns in human anterior nares. *FEMS Microbiol Ecol.* 2012;79(1):98–108. doi:10.1111/j.1574-6941.2011.01197.x
18. Luo MC, Thomas C, You FM, et al. High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics.* 2003;82(3):378–389. doi:10.1016/s0888-7543(03)00128-9
19. Muñoz-Amatriaín M, Lonardi S, Luo M, et al. Sequencing of 15 622 gene-bearing BACs clarifies the gene-dense regions of the barley genome. *Plant J.* 2015;84(1):216–227. doi:10.1111/tpj.12959
20. Quiniou SM, Waldbieser GC, Duke MV. A first generation BAC-based physical map of the channel catfish genome. *BMC Genomics.* 2007;8:40. Published 2007 Feb 6. doi:10.1186/1471-2164-8-40
21. Blair M, Panaud O, McCouch S. Inter-simple sequence repeat (ISSR) amplification for analysis of microsatellite motif frequency and fingerprinting in rice (Oryza sativa L.). *Theor Appl Genet.* 1999;98:780–792. doi:10.1007/s001220051135
22. Ammiraju J, Dholakia B, Santra D, et al. Identification of inter simple sequence repeat (ISSR) markers associated with seed size in wheat. *Theor Appl Genet.* 2001;102:726–732. doi:10.1007/s001220051703
23. Leroy XJ, Leon K. A rapid method for detection of plant genomic instability using unanchored-microsatellite primers. *Plant Mol Biol Rep.* 2000;18:283. doi:10.1007/BF02824000
24. Wang HZ, Wu ZX, Lu JJ, et al. Molecular diversity and relationships among Cymbidium goeringii cultivars based on inter-simple sequence repeat (ISSR) markers. *Genetica.* 2009;136(3):391–399. doi:10.1007/s10709-008-9340-0
25. Gupta SK, Souframanien J, Gopalakrishna T. Construction of a genetic linkage map of black gram, Vigna mungo (L.) Hepper, based on molecular markers and comparative studies. *Genome.* 2008;51(8):628–637. doi:10.1139/G08-050
26. Anderson GR, Brenner BM, Swede H, et al. Intrachromosomal genomic instability in human sporadic colorectal cancer measured by genome-wide allelotyping and inter-(simple sequence repeat) *PCR. Cancer Res.* 2001;61(22):8274–8283.
27. Wolfe AD, Xiang QY, Kephart SR. Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. *Mol Ecol.* 1998;7(9):1107–1125. doi:10.1046/j.1365-294x.1998.00425.x
28. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 1987;19:11–15.
29. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics.* 2001;17(8):754–755. doi:10.1093/bioinformatics/17.8.754
30. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19(12):1572–1574. doi:10.1093/bioinformatics/btg180

Chapter 7: Relative fluorescence quantitation (RFQ)

Overview

Relative fluorescence quantitation (RFQ) is a technique used in a variety of fragment analysis applications to compare peak heights across samples.

Relative fluorescence applications compare peak height or area between two samples. Common techniques include:

- Qualitative fluorescence (QF) PCR
- Quantitative multiplex PCR of short fluorescent fragments (QMPSF)
- Multiplex ligation-dependent probe amplification (MLPA)

Principle of the analysis

The data for an RFQ experiment can be analyzed with the microsatellite module or AFLP™ analysis module on the Applied Biosystems™ GeneMapper Software, or other application-specific software (e.g., the Applied Biosystems™ TrueMark™ MSI Analysis Software or Coffalyser.Net™ software for MLPA analysis).

Peak height or peak area can be used to compare differences in the same marker across multiple samples. However, you may see a difference in results depending on whether peak height or peak area is used.

Note: Variations in signal intensity adversely affects results in RFQ experiments.

Figure 7.1 is an electropherogram from an RFQ experiment of a microsatellite marker in DNA from a healthy sample and a tumor sample. The reduced peak height in the tumor sample indicates potential loss of heterozygosity (LOH) in the sample.

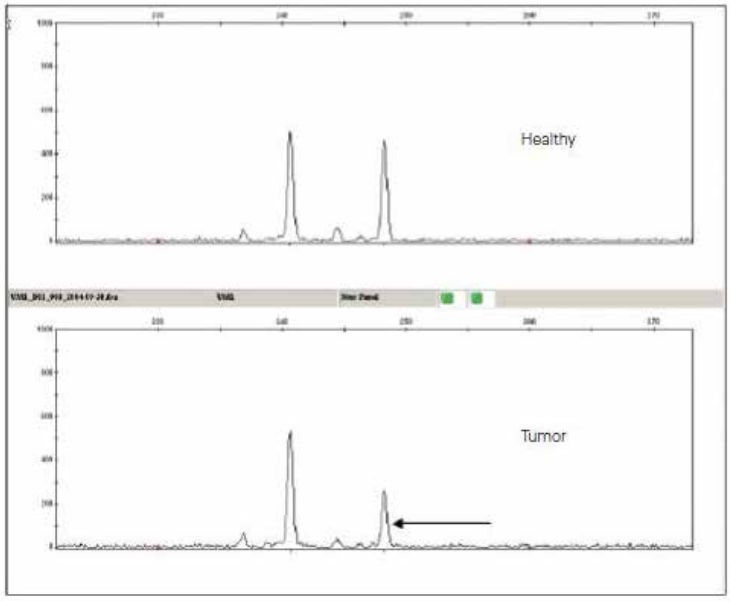


Figure 7.1. Electropherograms from a healthy sample (top) and a tumor sample (bottom).

Applications

- Screening for LOH using microsatellites or single-nucleotide polymorphisms (SNPs)
- Aneuploidy assays
- Detection of large chromosomal deletions
- Multiplex ligation-dependent probe amplification (MLPA)

Experiment and primer design recommendations

Recommendations

- Do not use internally labeled (fluorescent dNTP) fragments in quantitative experiments. Variations in the number of labeled nucleotides per fragment and the increased peak spreading with this method make relative quantitation unreliable.
- For more information, see "Microsatellite analysis" on page 14, and "Amplified fragment length polymorphism (AFLP) analysis" on page 24.

Minimizing signal intensity variation

To minimize variations, consider the ionic strength of samples and consumables. The amount of DNA injected is inversely proportional to the ionic strength of the solution:

- Samples high in salt result in poor injections. Polymerase chain reactions vary in efficiency; therefore, some reactions may result in higher ionic concentration post-amplification. A pre-CE sample cleanup, e.g., by PCR cleanup methods (column- or bead-based), is recommended.
- Conductivity of the solvent used for injection will affect the sample injection and can cause variation in peak height. A constant ratio of sample to mixture of Applied Biosystems™ HiDi™ Formamide and size standard should be obtained.

Note: Preferential amplification can decrease the accuracy of relative quantitation measurements.

LOH workflow

1. Design the primers and select the primer dye set
2. PCR:
   - Run two DNA samples from each individual, for example:
     - One from healthy tissue (H)
     - One from tumor tissue (T)
   - **Note:** Some normal tissue contaminating the tumor tissue sample is typical.
   - Run 3 to 4 independent injections for each sample (H and T) to obtain sufficiently accurate quantitative estimates for subsequent data analysis
   - Run control DNA:
     - Amplify at least one control DNA sample for every round of PCR amplification
     - Run at least one injection of amplified control DNA for every set of microsatellite markers used
     - Run at least one injection of amplified control DNA whenever you change the capillary or electrophoresis conditions
3. Capillary electrophoresis
4. Data analysis

Data analysis

Precise peak detection

Optimize peak detection parameters to ensure precise peak detection.

If noise peaks are detected, increase the minimum peak half-width or use a stronger smoothing option when analyzing noisy data.

Determining relative quantities

You can determine the relative quantities of two 5´-end labeled fragments by comparing the corresponding peak areas or peak heights on electropherograms generated from GeneMapper Software or Applied Biosystems™ Peak Scanner™ Software.

Assess the reproducibility of peak height and area for each new analysis. Note the following:

- Use area for slow migrating or wide peaks at high concentration.
- Use height for sharp peaks at low concentration.
- There is a linear relationship between migration time and reproducibility. As the migration time increases, the peak width and area increase. Therefore, fast migrating peaks result in higher reproducibility as measured by the peak area. However, improved reproducibility calculated using peak height has been observed as the migration time increases [1].

Determining relative number of molecules

To determine the relative number of molecules of two different-size fragments, calculate the ratio of respective peak areas or heights. Make sure to compare peak area to peak area, or peak height to peak height:

- If two fragments are similar in size, compare peak heights, especially if the peaks overlap slightly. If the peaks are well defined, peak area and peak height will give similar results. If the peaks are irregularly shaped or have shoulders, peak heights will often give better results than peak areas.
- If two fragments differ greatly in size, compare peak areas because large peaks tend to spread considerably more than small peaks.
- A practical approach has been developed for applying microsatellite multiplex PCR analysis (MMPA) for determining allele copy-number status and percentage of normal cells within tumors [2].

An application note on relative fluorescent quantitation is available **here**.

Microsatellite instability (MSI)

Microsatellite instability is the genetic mutation of microsatellites that can be caused by errors in DNA mismatch repair (MMR). The presence of high levels of MSI (MSI-H) may potentially be predictive of Lynch syndrome, a hereditary condition associated with increased cancer risk across a variety of tumor types including colorectal, gastric, and endometrial.

Defects in MMR also result in novel somatic mutations in unrelated loci throughout the genome. These mutations can produce novel immunogenic proteins that can "prime" an immune response to the foreign neoantigens in the tumor. MSI-H tumors are potentially more sensitive to immune checkpoint inhibitor treatment than microsatellite stable (MSS) tumors. There is therefore substantial interest in studying MSI in diverse tumors to identify those in which growth might be slowed by immune checkpoint inhibitor treatments.

The Applied Biosystems™ TrueMark™ MSI Assay is a low-input, affordable assay with an expanded panel of 13 MSI markers (Table 7.1) and includes Applied Biosystems™ TrueMark™ MSI Analysis Software that features automated calling to potentially accelerate accurate decision-making. Using only 2 ng of DNA derived from formalin-fixed, paraffin-embedded (FFPE) samples, the entire workflow takes as little as 4 hours to determine accurate sample status (Figure 7.2). The TrueMark MSI Assay leverages a 6-dye set to incorporate additional markers and does not require the use of a tumor-normal match when using the TrueMark MSI Analysis Software, cutting the number of wells required per sample in half and reducing sample-handling burden. The TrueMark MSI Assay is designed to run on Applied Biosystems™ SeqStudio™ and 3500 Series genetic analyzers.

Table 7.1. The TrueMark MSI Assay includes 13 MSI markers and 2 sample identification markers.

Markers	Description
BAT-25, BAT-26, BAT-40, CAT-25, NR-21, NR-22, NR-24, NR-27	Publicly known MSI markers
ABI-16, ABI-17, ABI-19, ABI-20A, ABI-20B	MSI markers proprietary to Thermo Fisher Scientific
PentaD, TH-01	Sample identification markers

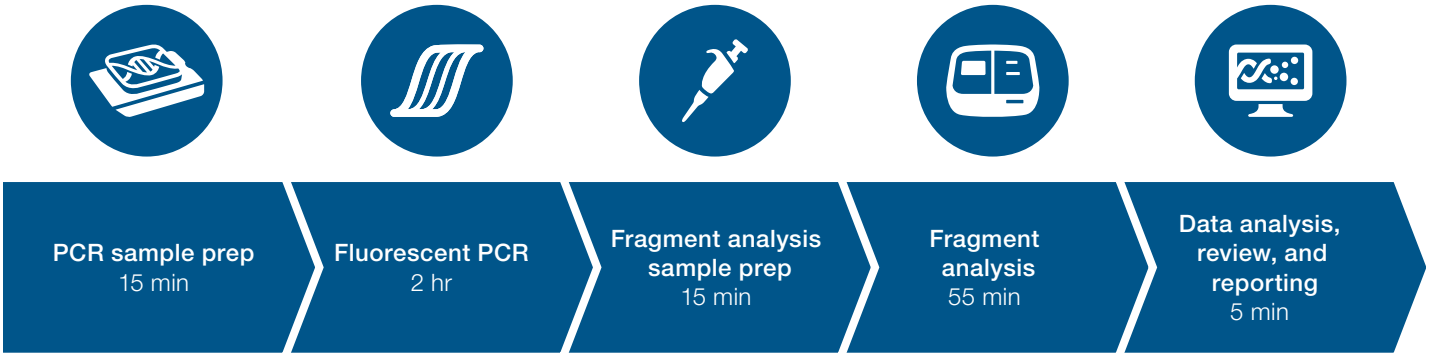


Figure 7.2. Workflow of the TrueMark MSI Assay. Go from extracted DNA to results in as little as 4 hours with less than 1 hour of hands-on time.

The TrueMark MSI Assay offers:

- **An expanded panel**—13 microsatellite markers to test for MSI associated with cancers related to Lynch syndrome and to study immunotherapy
- **Optional normal control**—no requirement to use a tumor-normal match for interpretation with the TrueMark MSI Analysis Software
- **Low input**—use as little as 2 ng of FFPE-derived DNA
- **Rapid turnaround**—go from DNA to results in as little as 4 hours with less than 1 hour of hands-on time
- **Automated calling**—simple analysis and reporting software includes automated calling of markers and samples

Included at no additional cost, Applied Biosystems™ TrueMark™ MSI Analysis Software is designed specifically for the TrueMark MSI Assay to enable greater ease of analysis and reporting. TrueMark MSI Analysis Software offers:

- Automated calling of markers and samples (Figure 7.3).
- Straightforward QC and review steps.
- Simple review of individual specimens and batches.
- An easy workflow for selecting markers to include in analysis and making manual calls when needed.
- Fast export of results into multiple file formats (PDF, CSV, VCF).
- Compatibility with Ion Torrent™ Oncomine™ Reporter to produce clear and concise reports that summarize relevant biomarkers with information from targeted therapies, guidelines, and global clinical trials (Figure 7.4).

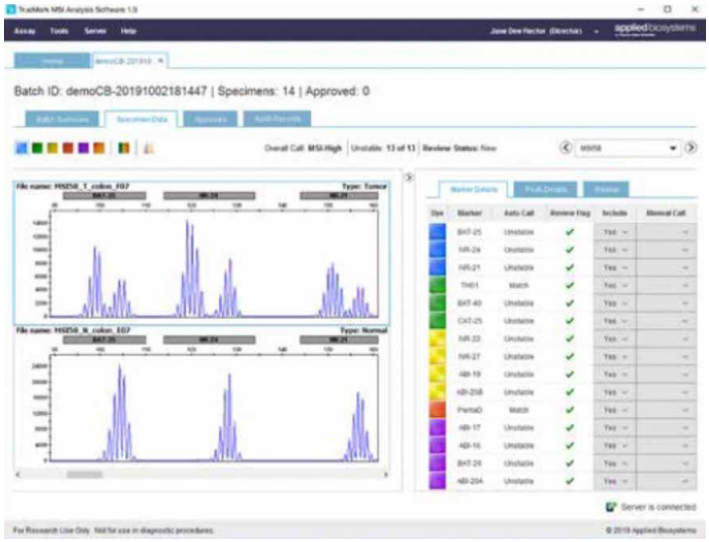


Figure 7.3. TrueMark MSI Analysis Software offers automated calling. Custom algorithms analyze the MSI status and return an overall call that considers all loci. The status at each locus can be examined and modified manually, if needed.

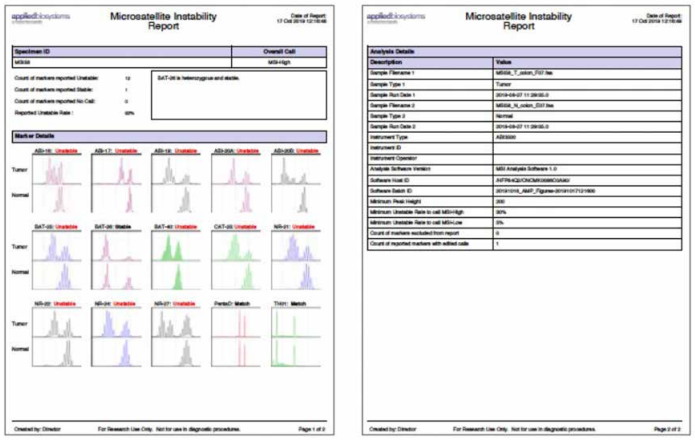


Figure 7.4. Example report from TrueMark MSI Analysis Software. Results are summarized in an easy-to-interpret manner.

References

1. Shihabi ZK, Hinsdale ME. Some variables affecting reproducibility in capillary electrophoresis. *Electrophoresis*. 1995;16(11):2159–2163. doi:10.1002/elps.11501601348
2. Garcia-Linares C, Mercadé J, Gel B, et al. Applying microsatellite multiplex PCR analysis (MMPA) for determining allele copy-number status and percentage of normal cells within tumors. *PLoS One*. 2012;7(8):e42682. doi:10.1371/journal.pone.0042682

Chapter 8: Application note—Analysis of double-stranded DNA

Analysis of double-stranded DNA fragments using capillary electrophoresis

In this study, we show:

- Double-stranded DNA (dsDNA) can be analyzed on capillary electrophoresis (CE)-based genetic analyzers with a simple workflow requiring no additional equipment
- Sizes of DNA can be calculated using well-defined dsDNA samples to recalibrate Applied Biosystems™ GeneScan™ LIZ™ standards
- The quality of next-generation sequencing (NGS) libraries can be analyzed using this method prior to sequencing

Introduction

Many steps in genomic analyses begin with a pool of DNA. It is often necessary to determine the size and amount of DNA in that pool before initiating a study or continuing to the next steps. For example, DNA extracted from formalin-fixed, paraffin-embedded (FFPE) tissue may be fragmented and, thus, too small to provide useful information; at the nucleic acid isolation stage, it would help to know the overall size of the DNA fragments before moving experiments forward. In NGS experiments, libraries of pooled fragments are often generated before being loaded onto the sequence-reading chip; knowing the size and abundance of fragments in that library is important for predicting the success of the NGS result. Finally, it can be useful to know whether a specific amplicon was successfully generated from a PCR reaction. Agarose gels are often used for this application, but while they are inexpensive, they are not very precise or high-throughput. Methods that can accurately and precisely characterize DNA quality (i.e., fragment size and abundance) in a sample are therefore highly advantageous.

Here we demonstrate how the Applied Biosystems™ genetic analyzer family of CE instruments can be used for evaluating the quality of dsDNA samples. The simple fragment analysis method relies on incubating DNA with a dsDNA-intercalating dye, followed by CE using existing polymers and buffers at low temperature. The resulting electropherogram shows the relative size of a sample and gives a semiquantitative measurement of the amount of DNA in the sample. Since this method does not require purchase of new reagents or instrumentation, it provides additional flexibility and utility for existing CE instruments.

Materials and methods

The size and relative abundance of dsDNA fragments can be analyzed on Applied Biosystems genetic analyzers (Figure 8.1). For this analysis, Invitrogen™ TOTO™-1 Iodide (514/533), a fluorescent DNA intercalating dye, is incubated with a nucleic acid sample. The sample is then analyzed by CE using Applied Biosystems™ POP-7™ Polymer on Applied Biosystems™ 3500 Series or SeqStudio™ Flex Series Genetic Analyzers, or by using POP-1™ on the Applied Biosystems™ SeqStudio™ Genetic Analyzer with the capillary heater turned off or minimized. This helps maintain DNA in a mostly double-stranded, undenatured state. By comparing the resulting migration of a test fragment with the migration of known dsDNA standards (e.g., Thermo Scientific™ phage lambda DNA digested with HindIII restriction enzyme and Thermo Scientific™ phage phiX174 digested with BsuRI or HaeIII restriction enzyme), the size of the test fragment(s) can be determined. A detailed protocol and run conditions are in the “Detailed protocol and run conditions” section.

Figure 8.1. Workflow for dsDNA analysis. This simple workflow can check the quality of dsDNA using Applied Biosystems genetic analyzers, such as SeqStudio Flex Series Genetic Analyzers, without requiring additional instrumentation or complex chemistries.

Results

dsDNA fragments with phage markers and other standards

To illustrate the results obtained with the CE method, we analyzed DNA digested with HindIII (for phage lambda DNA) and BsuRI (for phage phiX174 DNA). When phiX174–BsuRI enzyme standards were analyzed, the characteristic 10 fragments ranging from 1,353 bp to 72 bp were generated (Figure 8.2A). Similarly, a lambda–HindIII enzyme digest resulted in the well-known pattern of eight fragments that range from 23,130 bp to 125 bp (Figure 8.2B). A mixture of phiX174–BsuRI and lambda–HindIII digests were used to visualize a wide range of sizes covered by the two digests (Figure 8.2C). Finally, an engineered artificial ladder, such as the Invitrogen™ 100 bp DNA Ladder, produced similar results (Figure 8.2D).

Redefining GeneScan LIZ size markers for precision size determination

Although the size of a DNA population can be estimated using side-by-side comparison of peak traces, a more accurate and precise determination can be made by making use of the GeneScan LIZ size standards that are run in the same capillary. The GeneScan LIZ standards are end-labeled molecules that are single-stranded, and thus do not migrate the same as double-stranded DNA fragments that are intercalated with a dye. Nevertheless, the sizes of the GeneScan LIZ standards can be redefined, using a known set of dsDNA fragments. For example, Applied Biosystems™ GeneMapper™ Software can use the mixture of phiX174–BsuRI and lambda–HindIII digests described previously (Figure 8.2C) as the size standard. A 100 bp DNA ladder is also shown (Figure 8.2D). Using these size definitions for the GeneScan LIZ markers, a new size-standard set can be defined and subsequently used to analyze the sizes of unknowns labeled with TOTO-1 dye. For details on this method, see the "Recalibrating GeneScan LIZ standards for dsDNA sizing" section.

To illustrate how this method works, we used the phiX174–BsuRI and lambda–HindIII markers to define a set of GeneScan LIZ markers for dsDNA size determination. Note that the variability in size estimation increases as DNA fragment sizes increase—small variations in migration can produce large differences in calculated size for large fragments. The accuracy of this sizing was determined by measuring the sizes in a 100 bp ladder, using the recalibrated GeneScan LIZ standards (Figure 8.3A). Overall, there was excellent agreement between the measured and the predicted sizes for these fragments (Figure 8.3B).

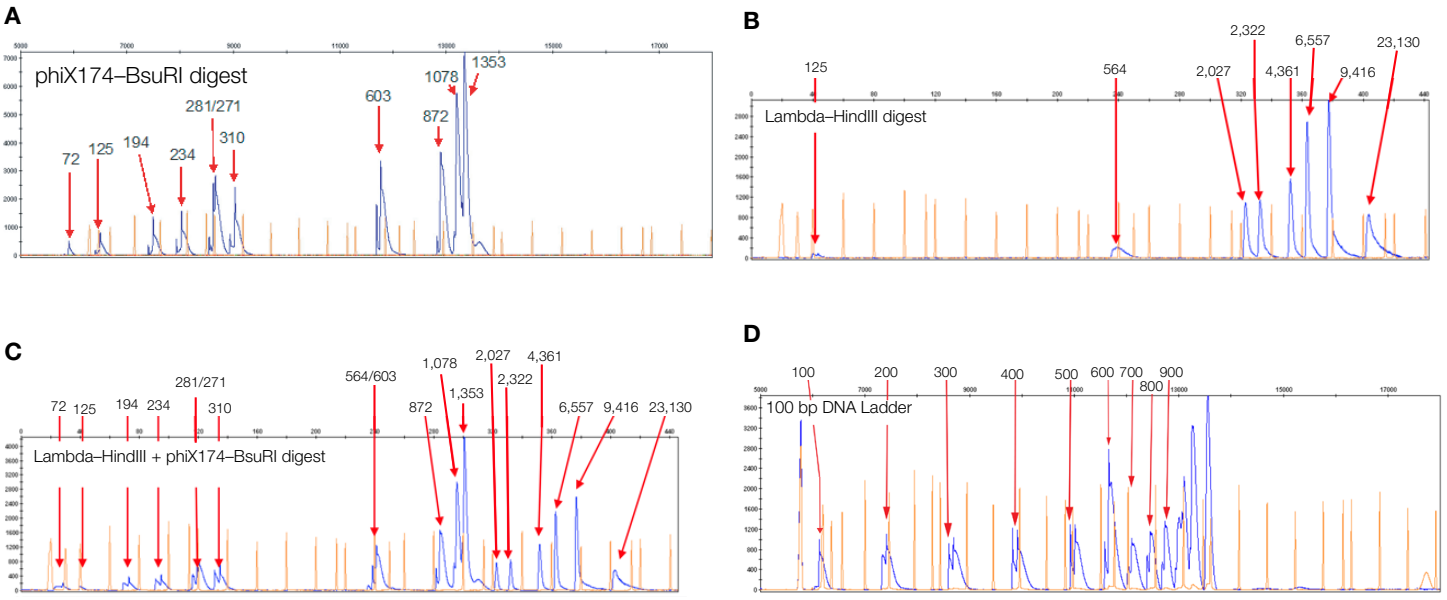


Figure 8.2. Migration of well-defined dsDNA fragments in capillaries. The typical fragment patterns seen on agarose gels were recapitulated by CE. GeneScan LIZ size standards displayed as orange peaks. **(A)** Phage phiX174 digested with BsuRI (HaeIII). **(B)** Phage lambda digested with HindIII. **(C)** Mixture of phiX174–BsuRI and lambda–HindIII digests. **(D)** 100 bp DNA ladder.

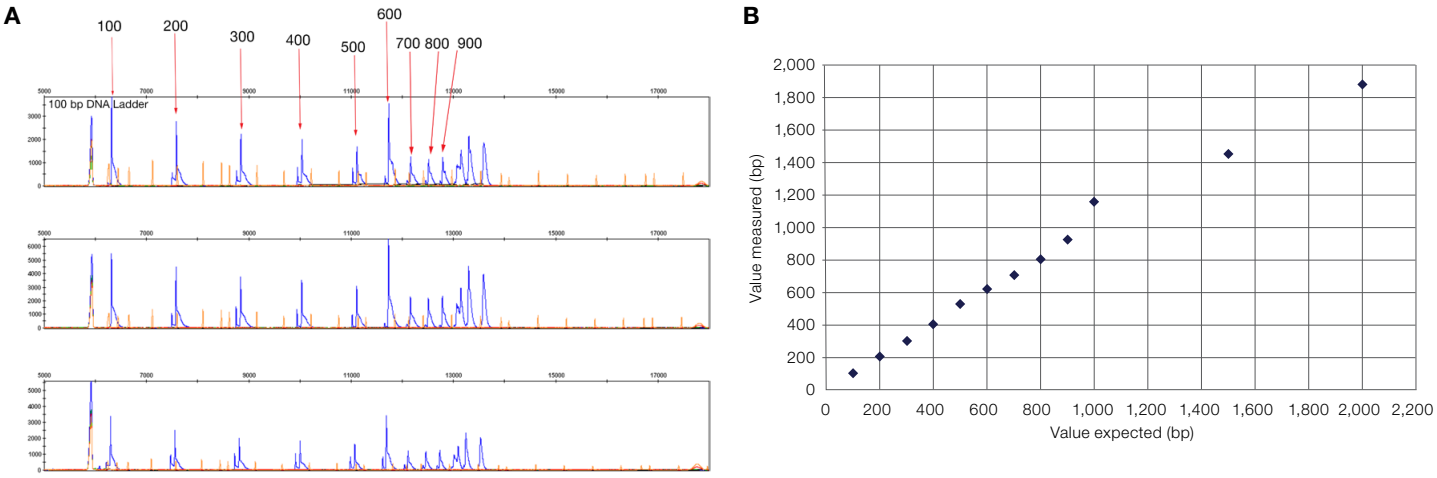


Figure 8.3. Calibration of LIZ size standards to dsDNA fragments. A mixture of phiX174–BsuRI and lambda–HindIII digests (not shown) were used to calibrate GeneScan LIZ markers for dsDNA size determination following the protocol on page 46. **(A)** Three replicates of a 100 bp ladder were used to characterize the precision of the measurements with the redefined GeneScan LIZ standards. **(B)** Correlation of expected sizes versus measured sizes of the 100 bp fragments. For all sizes, the correlation was $R^2 = 0.9887$. The standard deviations at each size for the replicates were calculated but are too small to be seen on the graph. Note that above 1,000 bp, the correlation drops off; if the fragments greater than 1,000 bp are removed, the correlation between expected and measured values rises to $R^2 = 0.9987$.

dsDNA sizing using genomic DNA as sample input

One of the reasons for dsDNA sizing is to determine the amount of fragmentation of a DNA sample. We isolated genomic DNA from HEK293 cultured cells and ran 1 µL of the product on a 2% agarose gel and on a CE genetic analyzer using the described protocol. As seen on the agarose gel, the majority of the DNA is large, essentially comigrating with the 23 kb lambda marker, but there is a smear of smaller fragments below that (Figure 8.4A, lane 5). When analyzed by CE, a similar pattern is clear: the majority of the DNA is large, with smaller, degraded fragments detectable (Figure 8.4B). Using the sizing method described above, the smear of large DNA was estimated to be between about 11 kb and 18 kb, with a smaller population also detectable.

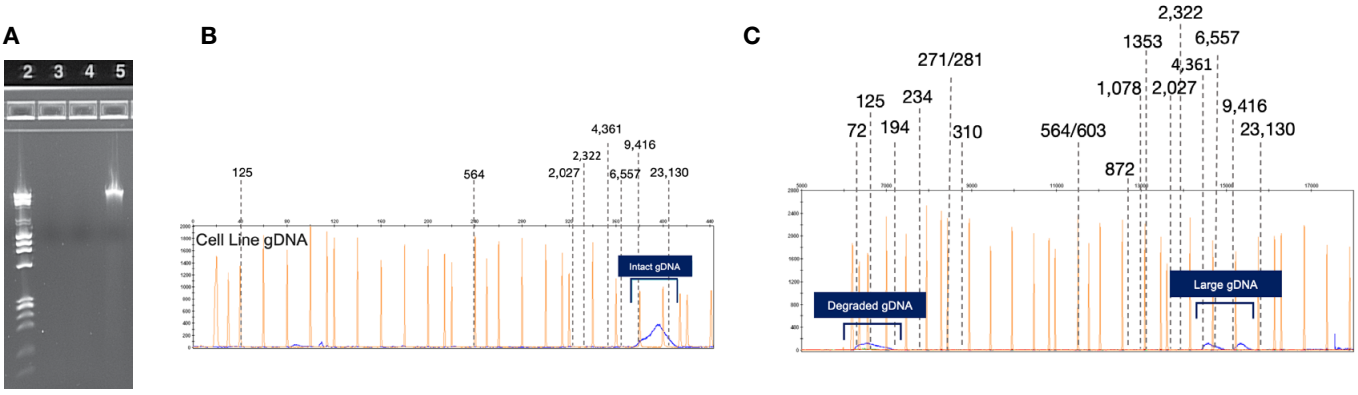


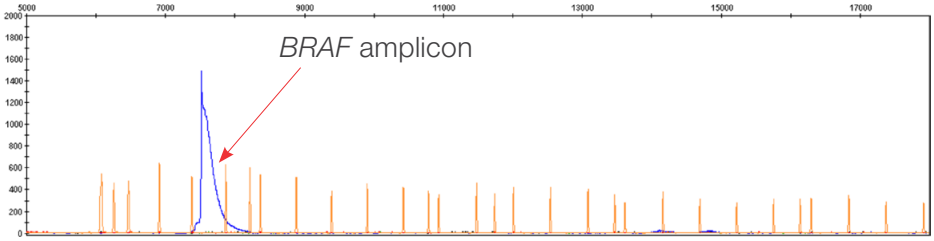
Figure 8.4. Genomic DNA analyzed by capillary electrophoresis. (A) Agarose gel showing migration of phiX174–BsuRI and lambda–HindIII mixture (lane 2); FFPE-extracted DNA (lanes 3 and 4); and genomic DNA from HEK293 cells (lane 5). **(B)** HEK293 genomic DNA with the major peak around 9,000–23,000 bp. **(C)** Genomic DNA extracted from an FFPE tumor sample. Note that there are distinct populations—a population of smaller fragments between about 85 bp and 185 bp and several populations of larger fragments at greater than 4,500 bp.

We also obtained samples of genomic DNA extracted from FFPE slides. When 1 µL each from two different samples was analyzed on an agarose gel, a very faint smear could be seen for each sample (Figure 8.4A, lanes 3 and 4). However, when analyzed by CE, two sets of DNA fragments could be seen: a small set, ranging in size from about 85 bp to 185 bp, and another set of fragments at greater than 4,500 bp (Figure 8.4C). These latter peaks could be intact genomic DNA, or they could be complexes of fixed nucleic acids, protein, and other molecules that resulted from the FFPE-fixation method.

dsDNA sizing using PCR amplicons as sample input

Some investigations depend on generating a PCR amplicon before performing more complex analyses. For example, Sanger sequencing from small amounts of template is facilitated by amplifying the region to be sequenced. Many researchers will want to verify that the PCR reaction was successful before proceeding with subsequent steps. To show how the CE method could be used to analyze PCR amplicons, we amplified the region surrounding *BRAF* V600 from FFPE samples. Using the sizing method described above, the peak in the positive control migrates at 209–218 bp (Figure 8.5A), and the FFPE-derived amplicon is similarly sized (Figure 8.5B). The expected size of this amplicon is 208 bp, which matches well with the observed size. For both amplicons, the subsequent cycle sequencing was successful (data not shown).

A



B

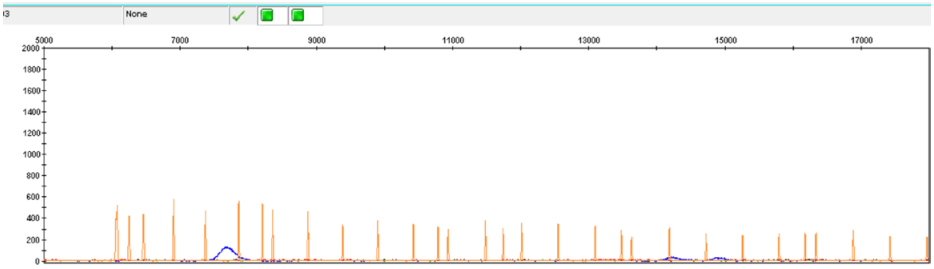


Figure 8.5. Quality analysis of a PCR amplicon to be used for sequencing. Two different PCR reactions from **(A)** HEK293 genomic DNA and **(B)** DNA extracted from an FFPE tumor section were run with GeneScan LIZ standards, and size was calculated as described in the "Recalibrating GeneScan LIZ standards for dsDNA sizing" section. The major peak is between 209 and 218 bp; the expected amplicon size is 208 bp.

dsDNA sizing using NGS libraries as sample input

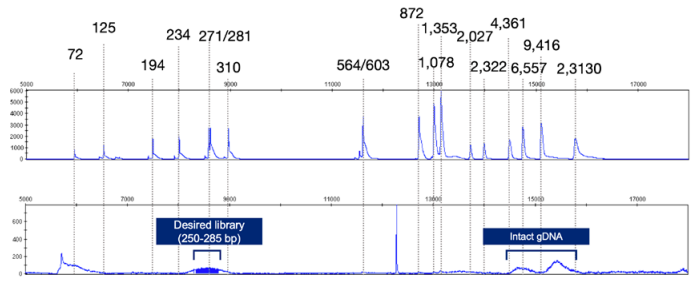
Another common instance in which prior information on amplicon size is desired is library generation for next-generation sequencing (NGS). In this scenario, researchers need to know whether their libraries meet the size expectations before loading them onto expensive NGS systems. As an example of an NGS library quality control (QC) experiment, we analyzed 2 µL of a library constructed using the Ion Torrent™ CarrierSeq™ workflow. The amplicons in this library migrated as a heterogeneous set of peaks, reflecting the different sizes of amplicons in the library. As measured by CE, the fragments ranged from 250 to 285 bp (Figure 8.6). This agreed with the expected value of around 200–300 bp.

Summary and conclusions

The CE method described here provides a path to fast and simple QC of dsDNA samples. We show that CE-based genetic analyzers provide information similar to an agarose gel, but with the potential capability of precise measurement of fragments. We illustrate how size determination by CE can be used to check the extent of fragmentation present in a genomic DNA sample and how this method can be used to analyze PCR amplicons before proceeding with more complex analyses. One workflow where this is important is QC of NGS libraries; we show that library amplicons can be detected and sized by CE.

The CE detection and sizing method described here allows users to get extra flexibility from their existing instruments and provides a semiquantitative estimate of fragment-specific DNA quantity. However, there are some things to consider. First, the molecules remain partially denatured when run in the capillary. That can lead to some broadening of the peaks, since the population is not expected to be in a uniform configuration. Thus, users should expect the double-stranded peaks to be wider than single-stranded peaks. Second, the incorporation of the intercalating dye can affect the migration of the molecules. This, in turn, may affect the measurement of the absolute size. Finally, there are many variables that affect the amount of DNA injected into the capillary. Because of this, quantification requires additional calibration and normalization. Other instruments should be used for precise quantification, such as the Invitrogen™ Qubit™ 4 Fluorometer or the Applied Biosystems™ QuantStudio™ Absolute Q™ Digital PCR System.

A



B

	Dye/Sample Peak	Sample File Name	Size	Height	Area	Data Point
	B,14	H02_H02_Sample	249.28	51	964	8309
	B,15	H02_H02_Sample	252.72	54	1966	8368
	B,16	H02_H02_Sample	254.37	57	1464	8394
	B,17	H02_H02_Sample	256.8	56	1773	8430
	B,18	H02_H02_Sample	258.55	68	1766	8454
	B,19	H02_H02_Sample	261.53	61	1049	8476
	B,20	H02_H02_Sample	265.32	62	1804	8502
	B,21	H02_H02_Sample	267.82	61	1000	8519
	B,22	H02_H02_Sample	270.97	67	1964	8540
	B,23	H02_H02_Sample	275.83	70	2064	8571
	B,24	H02_H02_Sample	282.31	76	3226	8609
	B,25	H02_H02_Sample	285.65	62	663	8640

Figure 8.6. Size analysis of an NGS library. (A) A phiX174–BsuRI and lambda–HindIII mixture (upper frame) in 2 µL of an Ion Torrent™ library (lower frame). Note there are three populations—a population of small fragments less than 75 nt, likely consisting of primer dimers and other small PCR artifacts; a population of large fragments between 4,000 and 23,000 nt, likely due to leftover genomic DNA; and the library at about 250–285 bp (blue area under the curve). A very narrow peak between 600 and 800 bp was also seen; the 1–2 pixel width suggests it is an artifact. **(B)** Table of GeneMapper data showing the breakdown of the fragment sizes under the curve highlighted in blue in panel A. The library was successfully run on Ion Torrent™ chips, generating over 95M reads.

Detailed protocol and run conditions

1. Create a 10 µM (1:100 dilution) intermediate stock solution of TOTO-1 dye in DMSO by adding 5 µL of 1 mM TOTO-1 dye to 495 µL of DMSO. Pipet 100 µL into microcentrifuge tubes. Store at –20°C, and thaw one tube as needed.
2. Create a 15 nM working solution by diluting the 10 µM stock solution in nuclease-free water. For a 96-well plate, add 1.5 µL of the 10 µM stock solution to 998 µL of water (1:667 dilution). Discard unused portions.

Note: The final working concentration of dye is 15 nM. For each use, prepare the 15 nM solution fresh as needed.

Labeling:

1. Add Applied Biosystems™ GeneScan™ 1200 LIZ™ Size Standard to 15 nM diluted TOTO-1 dye solution, and mix well. Use 0.5 µL/measurement.
2. Pipet 9 µL of 15 nM TOTO-1 dye–GeneScan LIZ solution into wells of a 96-well microtiter plate.
3. Add 1 µL of dsDNA sample into designated wells.
4. Add 1 µL of dsDNA standards (e.g., phage phiX174 digested with BsuRI or phage lambda digested with HindIII, or a mixture of the two, 10–100 ng) to at least one well.
5. Seal with adhesive film for the incubation period to decrease evaporation.
6. Incubate the sample according to the table:

Step		Temperature	Time
1x	Incubate	37°C	10 min
1x	Hold	4°C	Hold

7. Remove the adhesive film from the plate, and add the septum in preparation for the instrument run.

Load the instrument and run using the following running conditions:

Genetic analyzer	Injection time	Injection voltage	Run voltage	Run time	Oven temperature	Run module*	Dye set	Analysis protocol
SeqStudio Flex	15 sec	1,600	8,500	5,000 sec	40°C	DSDNA_QC	E5	(Default)
3500 Series	15 sec	1,600	8,500	8,000 sec	40°C	DSDNA_QC	E5	Fragment_ Analysis_PA_ Protocol
SeqStudio	7 sec	1,200	9,000	1,000 sec	40°C	FragAnalysis_1	E5	(Default)

* Guidance for setting up run modules on various genetic analyzers is in Figures 8.7–8.9.

Figure 8.7. Run conditions for SeqStudio Flex Genetic Analyzers.

SeqStudio Flex Series Genetic Analyzers

1. From the Maintenance screen, choose Library, then Run Module.
2. Select a fragment analysis run module, then choose Copy.
3. Edit conditions to match those in Figure 8.7 by touching the particular field, then entering the value.
4. After all changes are made, save the run module under a new name.

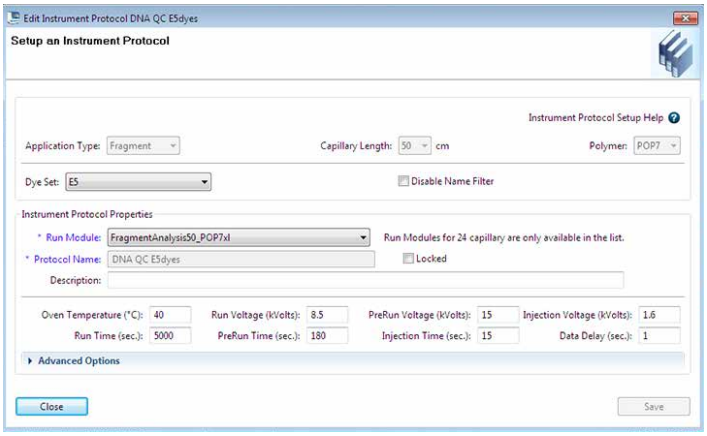
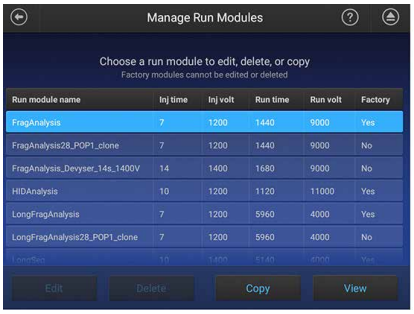


Figure 8.8. Run conditions for 3500 Series Genetic Analyzers.

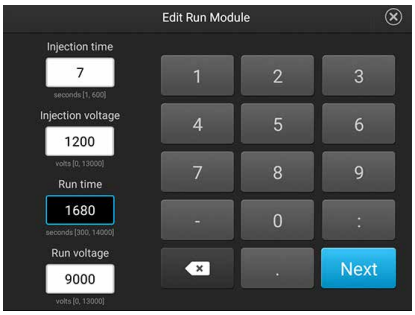
3500 Series Genetic Analyzers

1. From the Libraries tab, choose Instrument Protocols.
2. Select the Fragment Analysis run module to match the instrument configuration (3500 vs. 3500xL, 36 vs. 50 cm), and choose Copy.
3. Open the copied file.
4. Enter conditions to match those in Figure 8.8.
5. Save to the Library under a new name.

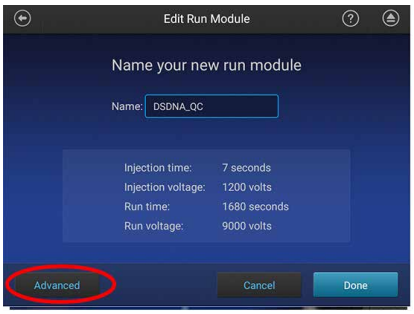
A



B



C



D

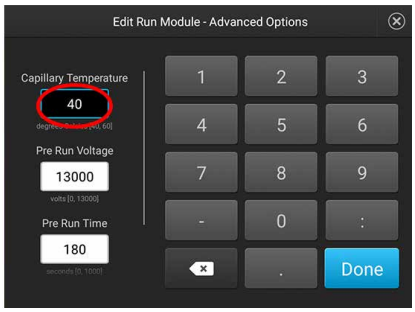
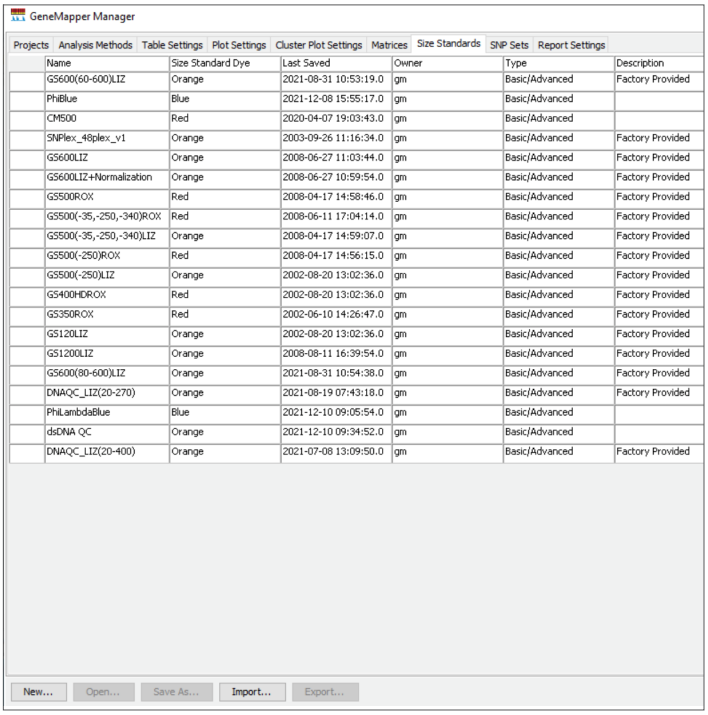


Figure 8.9. Run conditions for the SeqStudio Genetic Analyzer.

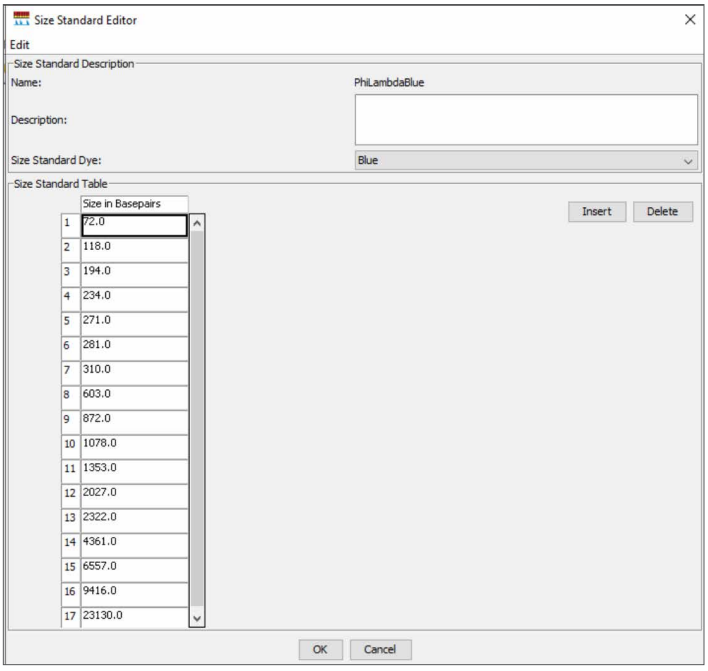
SeqStudio Genetic Analyzer

1. From the Maintenance screen, choose Manage Run Modules (Figure 8.9A).
2. Select FragAnalysis, followed by Copy.
3. Modify conditions on the Edit Run Module screen (Figure 8.9B), and select Next.
4. Select Advanced (Figure 8.9C).
5. Change Capillary Temperature from 60 to 40 (Figure 8.9D), and select Done.
6. Name the new run module.

Recalibrating GeneScan LIZ standards for dsDNA sizing

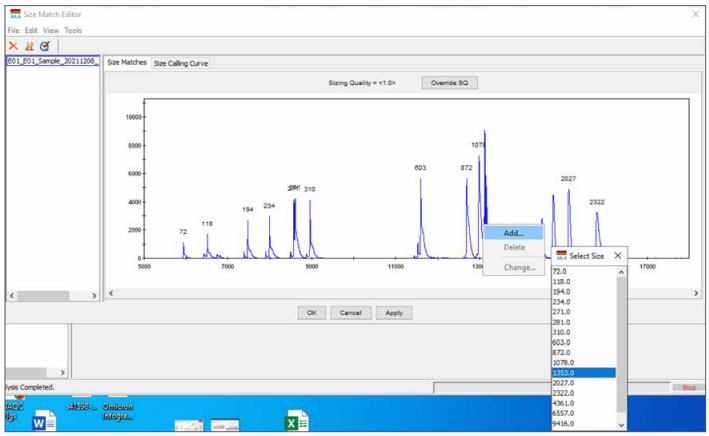


Step 1. In the GeneMapper Manager menu, open the Size Standards tab. Select the New button at the bottom left of the menu.

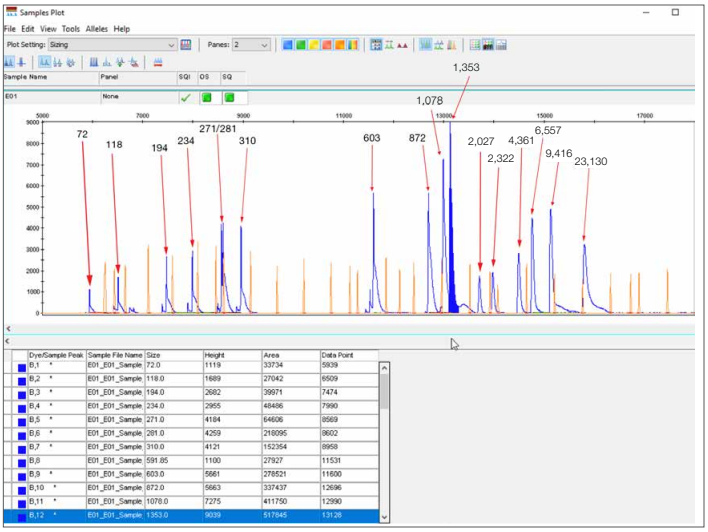


Step 2. Provide a name for the new size standards. For this example, since the sizes will be defined by phiX–lambda fragments, the set is called PhiLambdaBlue. In the Size Standard Dye selection, choose blue. For the Size Standard Table, enter the size in bp for each of the restriction fragments. Use the Insert button to add a new fragment. When finished, select OK and then Done.

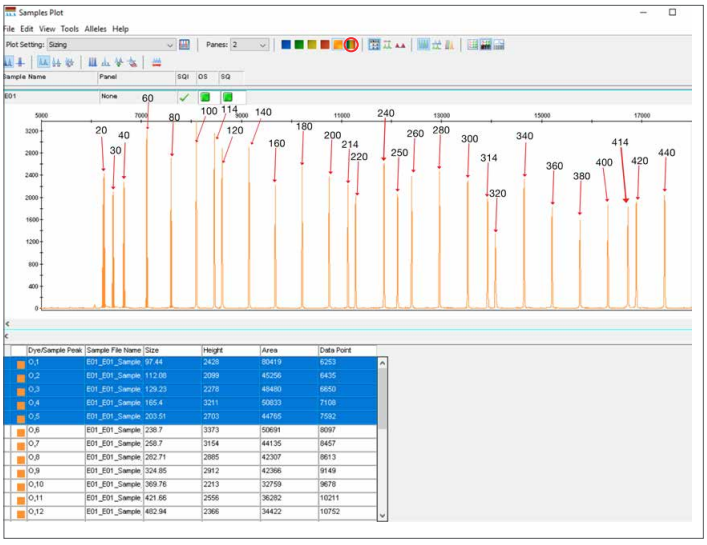
Step 3. Open the project with the .fsa files. Select the sample that has the fragments that were defined in step 2, DNA Library QC as the analysis method, and the newly defined double-stranded size standard (here, PhiLambdaBlue). Select the green triangle (arrow) to analyze the files.



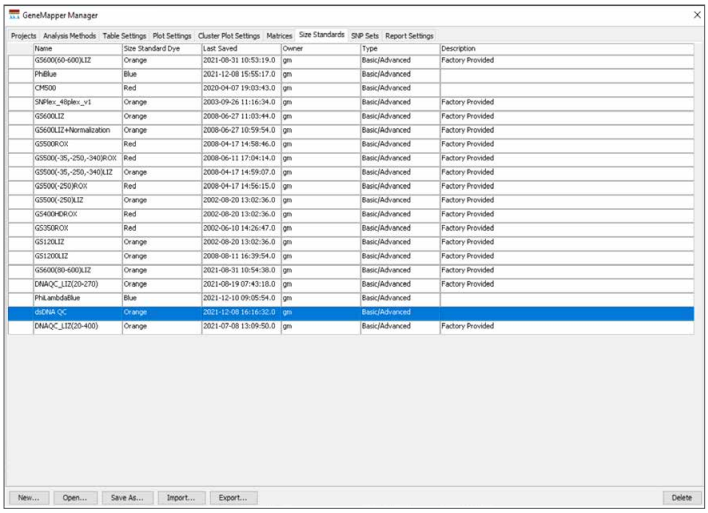
Step 4. Open the Size Match Editor for that sample. GeneMapper software will attempt to assign the fragments with the lengths defined. In some cases, it will be necessary to manually assign the fragment sizes. To do that, double-click on a peak—it will turn blue, and a menu will pop up. If there is an incorrect size, select Change. If there is no assignment, select Add. From the next pop-up that appears, choose the correct fragment size. Continue in this manner until all the fragments are assigned to the appropriate size. Finally, select the Override SQ button at the top, then Apply, and OK. This will take you back to the project view screen shown in step 3.



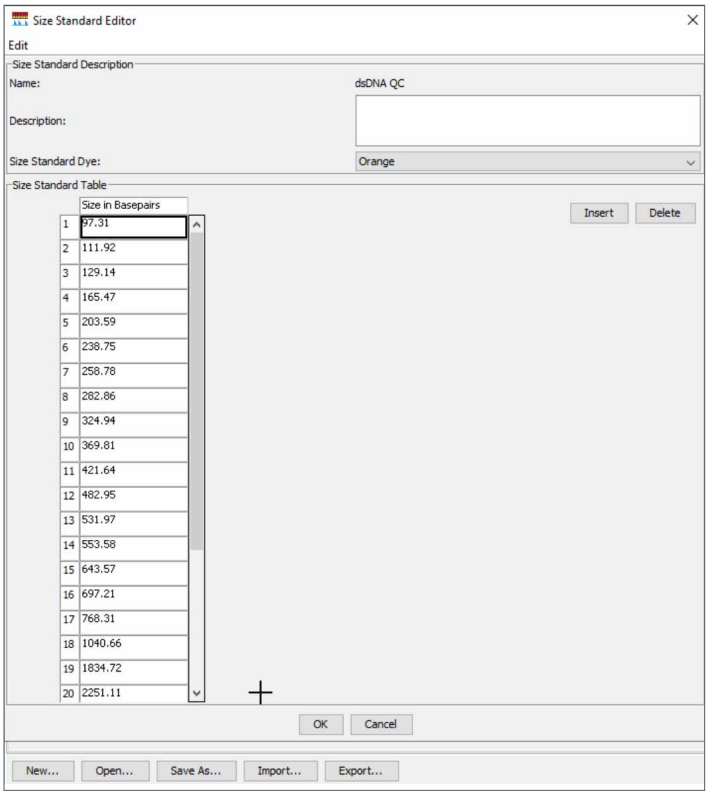
Step 5. In the project view, select the green arrow again (see step 3) to reanalyze the sample with the newly defined standards. When the analysis is finished, display the plot for the file. The peaks should now be assigned the correct sizes. In the table, the peaks that were defined will be indicated with an asterisk; peaks that are not part of the size-standard set will lack the asterisk. Clicking on a peak will highlight the peak in blue on the electropherogram and highlight the corresponding peak in blue in the table, as shown. Notice that in this example, the highlighted peak is defined in the size standards as 1,353 bp.



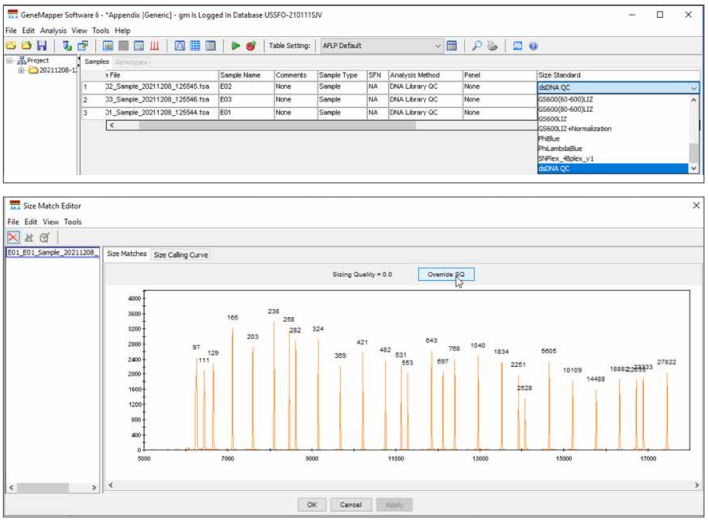
Step 6. Next, in the same window, highlight the orange peaks only by selecting the rainbow button, which is circled in red (turns off all colors), followed by the orange button to its left (turns on orange only). The orange GeneScan LIZ peaks will be shown, along with their calculated sizes. Make note of the calculated sizes of all the GeneScan LIZ peaks. For convenience, the entire table can be copied into a Microsoft™ Excel™ file or a text document.



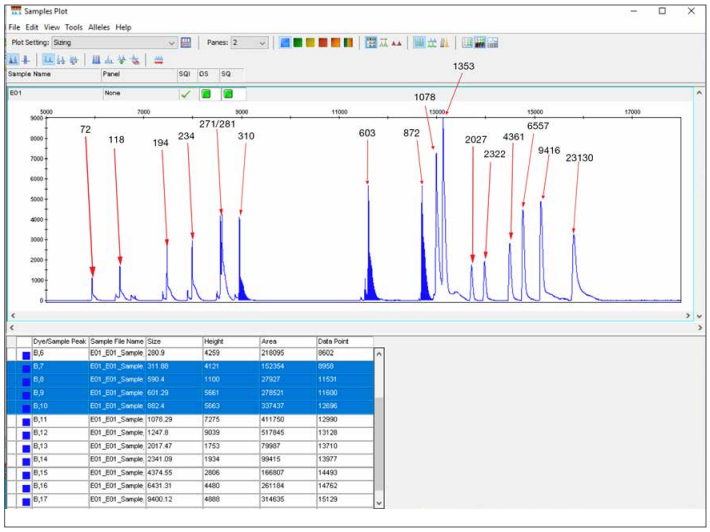
Step 7. Open the GeneMapper Manager, choose the Size Standards tab, and select New to create a new size standard.



Step 8. Similar to step 2, define a new set with a name (here, dsDNA QC) and size standard dye (orange this time, since we will be defining the GeneScan LIZ peaks). Enter the calculated values for the GeneScan LIZ peaks from step 6. Use the Insert button to create more entries as needed. When all are entered, select OK, then Done.



Step 9. In the project view, change the size standards column to the newly named set (in this example, dsDNA QC) and analyze by selecting the green arrow (top). Open the Size Match Editor, and make sure all the peaks have a size assigned and that the sizes make sense. If they are all assigned and they all make sense, select Override SQ at the top of the window, select Apply if needed, and then select OK.



Step 10. Select the green arrow to analyze the sample with the new standards. Here, the same phiX–lambda set was sized using the reassigned LIZ peaks. The highlighted large peaks are consistent with expected values (measured 311.88 bp, expected 310 bp; measured 601.29 bp, expected 603 bp; measured 882.4 bp, expected 872 bp). One smaller “pre-peak” was also measured—it precedes the 601 bp peak and was measured as 590.4 bp.

Appendix

Choose the thermal cycler that is right for you

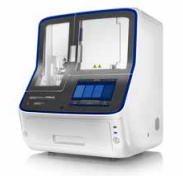
Applied Biosystems™ thermal cyclers enable precise, consistent PCR results for every challenge, application, and budget. Be confident in your results with instruments designed for reliability, accuracy, and user-friendly interfaces. Reduce time spent on optimization with the Applied Biosystems™ VeriFlex™ temperature control technology. Manage your instruments remotely with secure access.

Compare instruments

	ProFlex PCR System	VeritiPro Thermal Cycler	SimpliAmp Thermal Cycler	MiniAmp Thermal Cycler	Automated Thermal Cycler
					
Block formats (temperature optimization)	3 x 32-well, 0.2 mL (2-zone VeriFlex Block) 96-well, 0.2 mL (6-zone VeriFlex Block) 2 x 96-well, 0.2 mL 2 x flat block 2 x 384-well, 0.02 mL	96-well, 0.2 mL (6-zone VeriFlex Block) 384-well, 0.02 mL	96-well, 0.2 mL (3-zone VeriFlex Block)	96-well, 0.2 mL	96-well, 0.2 mL, compatible with full- or semi-skirted plates 384-well, 0.02 mL
Temperature optimization	6-zone VeriFlex Block on 96-well system 2-zone VeriFlex Block on 32-well system	6-zone VeriFlex Block on 96-well system	3-zone VeriFlex Block on 96-well system	3-zone VeriFlex Block on MiniAmp Plus Thermal Cycler	None
User interface	8.4 in. touchscreen	8.0 in. touchscreen	8.0 in. touchscreen	5.0 in. touchscreen	Via robotic platform or desktop software
Max. sample throughput	480,000 reactions	384 reactions	96 reactions	96 reactions	384 reactions
Max. block ramp rate	6.0°C/sec	6.0°C/sec	4.0°C/sec	3.0°C/sec	3.5°C/sec

Choose the genetic analyzer that is right for you

Applied Biosystems™ genetic analyzers are a trusted standard for Sanger sequencing and fragment analysis by capillary electrophoresis (CE). Versatile and innovative, our genetic analyzers have been designed with you in mind to meet the demands of your genetic testing needs. Whether you are analyzing a challenging sample or a routine sample, our genetic analyzers offer simplicity, scalability, and speed to your workflow, enabling you to analyze the most difficult cases with complete confidence.

	SeqStudio™ Genetic Analyzer	SeqStudio™ Flex Series Genetic Analyzers	3500 Series Genetic Analyzers	3730*x*/ DNA Analyzer
	Easy-to-use, flexible system	**Easy-to-use, flexible, connected system**	**Meets the needs of verified and process-controlled environments**	**Maximum throughput, scalability, and flexibility**
				UPGRADED 
Number of capillaries	4	8 or 24	8 or 24	48 or 96
Capillary array length (cm)	28	36, 50	36, 50	36, 50
Sample capacity	1 plate; 96-well plate; 8-strip tube (compatible)	4 plates; 96- or 384-well plates; 8-strip tube (compatible)	2 plates; 96- or 384-well plates; 8-strip tube (compatible)	16 plates; 96- or 384-well plates
Continual plate loading	No	Yes	No	Yes
Sample reprioritization	No	Yes	Yes	No
Polymer type	POP-1, integrated into click-in cartridge	POP-6, POP-7, and POP-4	POP-6, POP-7, and POP-4	POP-6, POP-7, and conformational analysis polymer
Radio-frequency ID	Yes	Yes	Yes	No
Configuration	Integrated computer with touchscreen; optional desktop	Integrated computer with touchscreen; optional desktop	External desktop required	External desktop required
Amazon™ Alexa™ voice command	No	Yes	No	No
Remote monitoring and data sharing	Yes	Yes	No	No
Integrated remote troubleshooting tools	No	Yes	No	No
Connectivity	USB, Ethernet ports, and Wi-Fi dongle	USB, Ethernet ports, and Wi-Fi dongle	Ethernet port	Ethernet port
Connectivity with Thermo Fisher™ Connect Platform	Yes	Yes	No	Yes

Fluorescent DNA labeling methods

Detection and high-fidelity sizing of DNA by fragment analysis on the capillary electrophoresis platform requires the presence of at least one fluorescent label attached to the molecule of interest. In most cases, the labeling can be performed by a variety of methods that are described below.

Method I. Generate a 5′-labeled DNA molecule by PCR with a primer that carries a fluorescent label modification at its 5′ end (Figure A.1).

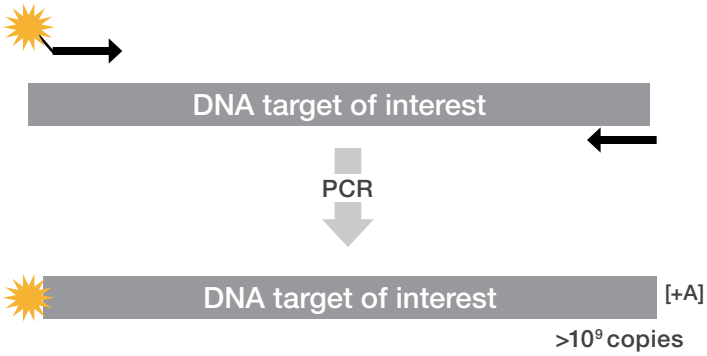


Figure A.1. Generation of a 5′-labeled DNA molecule using a fluorescently labeled PCR primer.

The benefits of direct and stoichiometric PCR-incorporated fluorescent labeling are:

- **Higher precision**—Different fluorophores have different mobilities. DNA fragments with the same 5′-end primer and fluorophore have comparable electrophoretic mobility and yield sharper fragment peaks because 5′-end primer labeling yields 1:1 incorporation (i.e., one fluorophore to one DNA fragment).
- **More consistent quantitation**—Every peak in an electropherogram is made up of multiple DNA fragments with an equal number of base pairs. When using 5′-end primer labeling, every DNA fragment contributes a single fluorophore to the total signal of a peak, and thus the peak area is proportional to the number of DNA molecules.
- **Distinct strands**—By attaching different fluorophores to the forward and reverse primers, it is possible to distinguish between the peaks that correspond to each strand and between residual double-stranded products.
- **Multiplexing**—This is the method of choice for more complex (e.g., multiplex) project designs.

Applied Biosystems™ 5′-labeled primers that are fully compatible with the standard dye matrix settings of the Applied Biosystems™ capillary electrophoresis platform can be ordered **here**.

Oligonucleotides are custom manufactured to your specifications. Choose from Applied Biosystems™ 5′-Fluorescent Labeled Oligo primers and primer pairs for use in fragment analysis.

5′-labeled single primers

Custom 5′-labeled primers are fluorescently labeled oligos with a choice of dye on the 5′ end. These products can be used for fragment analysis, gene expression studies, PCR, and genotyping research.

Features include:

- Choice of delivered scales
- All oligos are desalted; larger scales of labeled primers have an option for HPLC purification; HPLC purification reduces the yield by half

Cat. No.	Size	HPLC yield	5′ reporter dye available	Quantity
450007	10,000 pmol	NA	6-FAM, TET, VIC, HEX, NED, PET	1 tube
450006	80,000 pmol	40,000 pmol minimum	6-FAM, TET, VIC, HEX, NED, PET	1 tube
450017	300,000 pmol	150,000 pmol minimum	6-FAM, TET, VIC, HEX, NED, PET	1 tube

5′-labeled and unlabeled primer pairs

Custom 5′-labeled primer pairs include a fluorescently labeled oligo with a choice of dye on the 5′ end, along with an unlabeled oligo, in each pair. These products can be used in a variety of research applications.

Features include:

- Choice of delivered scales
- For use in fragment analysis, gene expression studies, genotyping, and PCR
- All oligos are desalted
- Available as “tailed” for the unlabeled primer to complete a single A addition during PCR
- Available as “di-repeat” for microsatellite typing of di-repeat loci

Cat. No.	Size	Product description	5′ reporter dye available	Quantity
450056	10,000 pmol	5′-labeled/ unlabeled primer pairs	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
450059	80,000 pmol	5′-labeled/ unlabeled primer pairs	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
450062	300,000 pmol	5′-labeled/ unlabeled primer pairs	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304976	10,000 pmol	5′-labeled/ unlabeled di-repeats	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304977	80,000 pmol	5′-labeled/ unlabeled di-repeats	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304978	300,000 pmol	5′-labeled/ unlabeled di-repeats	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304979	10,000 pmol	5′-labeled/ unlabeled di-repeats + tail	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304981	80,000 pmol	5′-labeled/ unlabeled di-repeats + tail	6-FAM, TET, VIC, HEX, NED, PET	2 tubes
4304982	300,000 pmol	5′-labeled/ unlabeled di-repeats + tail	6-FAM, TET, VIC, HEX, NED, PET	2 tubes

Method II. Label the 5′ end using a secondary primer with a fluorescent label.

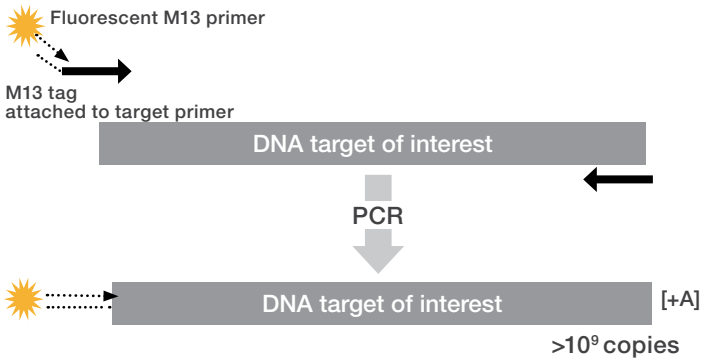


Figure A.2. Generation of a 5′-labeled DNA molecule using a secondary primer that is fluorescently labeled.

This is an economical method because the primer carrying the fluorescent label can be a "universal" M13 primer (or any other standardized sequence). The universal primer can therefore be used for coamplification with any other target-specific primer carrying the same M13 label (Figure A.2). Applications and variations of this labeling method have been described by various authors [1-4]. A fluorescently labeled primer can be ordered **here**.

The use of this method is not necessarily recommended for more complex multiplex amplifications because of the possibility of unwanted primer interactions. However, the method can be used for the detection of a small insertion or deletion (indel) in genomic DNA with a singular PCR amplification.

For investigations of indels in human exon DNA, we recommend the over 600,000 predesigned primer pairs offered in the "Primer Designer Tool for PCR and Sanger Sequencing" portal (**thermofisher.com/primerdesigner**). Suitable and economically priced primer pairs (including an M13 tail) are available for over 95% of the human exome.

Method III. Generate a 3′-labeled DNA molecule by addition of a fluorescent dideoxynucleotide (Applied Biosystems™ SNaPshot™ System).

The benefit of this method is that any double-stranded DNA molecule with an overhang of one or more nucleotides can be conveniently and economically labeled without any need for additional primers (Figure A.3). Nucleotide overhangs can be generated by:

i) PCR amplification with a polymerase that adds an A at the 3′ end

ii) A restriction enzyme digest that leaves a 5′ or 3′ overhang

iii) Tailing (e.g., single A addition) of blunt-end DNA

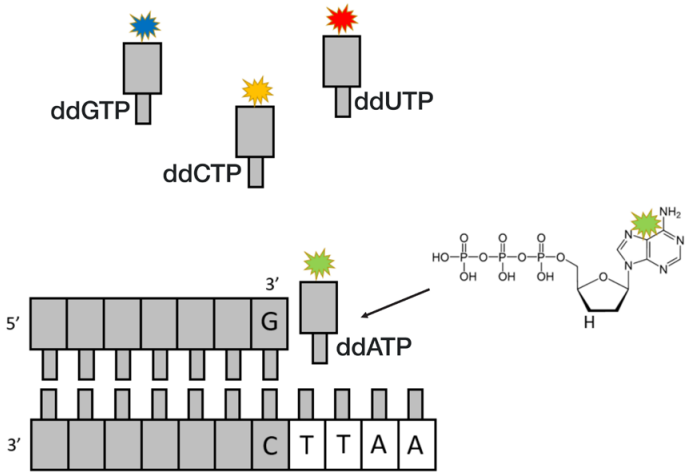


Figure A.3. Generation of a 3′-labeled DNA molecule by addition of a fluorescent dideoxynucleotide. Example of labeling an EcoRI-generated 5′ DNA overhang (G|AATTC) using a rhodamine dye–labeled (green fluorescent) dideoxyATP nucleotide that is incorporated by the SNaPshot™ DNA polymerase (see page 22). The dideoxy modification prevents further addition (polymerization) of other nucleotides and thus generates a single 3′ fluorescent end-label.

Method IV. Globally stain the DNA using nucleic acid stains.

An indiscriminate way of "labeling" DNA fragments for sizing and a general quantity assessment is to stain fragments using a dye with high-affinity dye for nucleic acids such as the Invitrogen™ TOTO™-1 Iodide (514/533) dye. Staining is then followed by capillary electrophoresis under non-denaturing conditions. An application of this method is described in chapter 8.

References

1. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol.* 2000;18(2):233–234. doi:10.1038/72708
2. Culley TM, Stamper TI, Stokes RL, et al. An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. *Appl Plant Sci.* 2013;1(10):apps.1300027. Published 2013 Oct 1. doi:10.3732/apps.1300027
3. Varshney GK, Carrington B, Pei W, et al. A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish. *Nat Protoc.* 2016;11(12):2357–2375. doi:10.1038/nprot.2016.141
4. Petree C, Varshney GK. MultiFRAGing: rapid and simultaneous genotyping of multiple alleles in a single reaction. *Sci Rep.* 2020;10(1):3172. Published 2020 Feb 21. doi:10.1038/s41598-020-59986-1

PCR enzymes for fragment analysis

Thermo Fisher has a variety of Applied Biosystems™ and Invitrogen™ PCR enzymes and master mix reagents available for your applications. The range of enzymes and reagents available gives you the flexibility needed to perform your experiments.

For most fragment analysis applications, Applied Biosystems™ AmpliTaq™ 360 DNA Polymerase and Platinum™ SuperFi II DNA Polymerase are recommended as reliable PCR enzymes. Master mix packs are also available for each of these polymerases.

Most of the PCR enzymes are available in "stand-alone" format or as "2X master mix" format (i.e., conveniently formulated as a 2X ready-to-use reagent containing optimized buffer, nucleotides, and enzyme). For fragment analysis applications, a colorless version is preferred to avoid any spectral interference during capillary electrophoresis (CE) on a genetic analyzer.

PCR types

Hot-start PCR

Goal: Increase specificity and yield.

	Invitrogen™ Platinum™ II Taq Hot-Start DNA Polymerase	AmpliTaq Gold 360 DNA Polymerase
Hot-start technology	Antibody	Chemical
Enzyme activation time	2 min	10 min
Universal annealing temperature of 60°C	Yes	No
DNA synthesis speed	15 sec/kb	60 sec/kb
Inhibitor tolerance	Yes	No
Amplification length	Up to 5 kb	Up to 5 kb
GC-rich format	Yes	Yes (add GC enhancer for optimal performance)

High-fidelity PCR

Goal: Obtain high sequence accuracy.

	Invitrogen™ Platinum™ SuperFi II DNA Polymerase	Invitrogen™ Platinum™ Taq DNA Polymerase High Fidelity
Fidelity vs. *Taq* enzyme	>300x	6x
Hot start for enhanced specificity	Yes	Yes
Universal annealing temperature of 60°C	Yes	No
Amplification length	Up to 20 kb*	Up to 20 kb
Amplicon overhangs	Blunt	3′ A and blunt
GC-rich amplification	Yes	No

* Amplification of >20 kb fragment sizes is possible (up to 40 kb) but may require additional optimization of reaction conditions and primer design.

Long-range PCR

Goal: Produce robust yields of templates longer than 10 kb.

	Platinum SuperFi II DNA Polymerase
Amplification length	Up to 20 kb*
Fidelity vs. *Taq* enzyme	>300x
Hot start for enhanced specificity	Yes
Universal annealing temperature of 60°C	**Yes**
GC-rich amplification	Yes
DNA synthesis speed	15–30 sec/kb

* Amplification of >20 kb fragment sizes is possible (up to 40 kb) but may require additional optimization of reaction conditions and primer design.

Multiplex PCR

Goal: Amplify multiple targets in one reaction. This is an important fragment analysis application.

	Platinum SuperFi II DNA Polymerase	Applied Biosystems™ Platinum™ Multiplex PCR Master Mix
No. of amplicons in single reaction	Up to 15-plex	Up to 20-plex
Amplification length	Up to 2.5 kb	Up to 2.5 kb
Hot start for enhanced specificity	Yes	Yes
Fidelity vs. *Taq* enzyme	>300x	1x
Universal annealing temperature of 60°C	**Yes**	No
GC-rich amplification	Yes	No

GC-rich PCR

Goal: To amplify DNA sequences with >65% GC content.

	Platinum SuperFi II DNA Polymerase	Platinum II Taq Hot-Start DNA Polymerase
Fidelity vs. *Taq* enzyme	>300x	1x
Hot start for enhanced specificity	Yes	Yes
Efficient amplification of >65% GC sequences	Yes	Yes
Universal annealing temperature of 60°C	**Yes**	**Yes**
Speed	15–30 sec/kb	15 sec/kb
Amplification length	Up to 20 kb	Up to 5 kb
Amplicon overhangs	Blunt	3´ A

Direct PCR

Goal: To amplify target DNA directly from samples without DNA purification.

	Invitrogen™ Platinum™ Direct PCR Universal Master Mix
Works across samples of various origins	Yes
Universal annealing temperature of 60°C	**Yes**
Hot start for enhanced specificity	Yes
Fidelity vs. *Taq* enzyme	1x
GC-rich amplification	Yes
Amplification length	Up to 8 kb**
Speed	20 sec/kb

** Using the lysis protocol.

Note: It is advisable to purify a direct PCR reaction (using column, bead, or precipitation methods) before CE to avoid contamination of capillaries with debris or fluorescent byproducts.

Broad-range PCR

Goal: To accurately detect and identify microbiome composition by 16S rRNA genes.

	Platinum II Taq Hot-Start DNA Polymerase	Invitrogen™ Platinum™ Taq DNA Polymerase, DNA-free	Platinum SuperFi II DNA Polymerase
Bacterial gDNA copy per 50 µL reaction	≤1 copy	≤0.01 copy	≤1 copy
Human gDNA copy per 50 µL reaction	≤0.2 copy	≤0.001 copy	≤0.3 copy
DNA synthesis speed	15 sec/kb	60 sec/kb	15–30 sec/kb
Inhibitor tolerance	Yes	No	Yes
Amplification length	Up to 5 kb	Up to 5 kb	Up to 20 kb
Fidelity vs. *Taq* enzyme	1x	1x	>300x
GC-rich amplification	Yes	No	Yes
Hot start for enhanced specificity	Yes	Yes	Yes
Universal primer annealing	Yes	No	Yes

Standard PCR

Goal: For everyday PCR amplification.

	Thermo Scientific™ DreamTaq™ DNA Polymerase	Invitrogen™ Taq DNA Polymerase	Applied Biosystems™ AmpliTaq™ DNA Polymerase
Hot-start modification	No	No	No
Amplification length	Up to 6 kb	Up to 5 kb	Up to 5 kb
5′→ 3′ exonuclease activity	Yes	Yes	Yes
Amplicon overhangs	3′ A	3′ A	3′ A
dUTP incorporation	No	Yes	Yes
DNA synthesis speed	1 min/kb	1 min/kb	1 min/kb
Format available with the native enzyme	No; recombinant expression in *E. coli*	Yes; purified from host *Thermus aquaticus*	No; recombinant expression in *E. coli*
Format for low bacterial DNA	–	–	Available

An overview of Thermo Scientific™ PCR enzymes and 2X master mixes can be found **here**.

High-fidelity PCR enzymes are used for applications requiring high accuracy during DNA amplification such as cloning, sequencing, or mutagenesis. Thermo Scientific™ Phusion™ high-fidelity DNA polymerases were created by fusing a dsDNA-binding domain to a polymerase-like proofreading enzyme from *Pyrococcus*. Due to this unique protein-fusion engineering, Phusion DNA polymerases have very low error rates, fast extension rates, and high tolerance for inhibitors.

ThermoFisher SCIENTIFIC

Select a Phusion DNA polymerase best suited for your application

Phusion high-fidelity DNA polymerases are available in a variety of formats. Please reference the table below to find an appropriate format for your application.

	Thermo Scientific™ Phusion™ Plus DNA Polymerase	Thermo Scientific™ Phusion™ High-Fidelity DNA Polymerase	Thermo Scientific™ Phusion™ Hot Start II DNA Polymerase	Thermo Scientific™ Phusion™ Flash High-Fidelity PCR Master Mix	Thermo Scientific™ Phusion™ U Hot Start DNA Polymerase	Thermo Scientific™ Phusion™ U Multiplex PCR Master Mix
Fidelity (vs. *Taq* enzyme)	>100x	52x	52x	25x	25x	25x
T_m calculation required	No	Yes	Yes	Yes	Yes	Yes
Enhanced specificity (hot-start version)	Yes	No	Yes	Yes	Yes	Yes
Amplification length	Up to 20 kb	Up to 20 kb	Up to 20 kb	Up to 20 kb	Up to 20 kb	Up to 2.5 kb
Speed	15–30 sec/kb	15–30 sec/kb	15–30 sec/kb	<15 sec/kb	15–30 sec/kb	15–30 sec/kb
GC-rich format	Yes	Yes	Yes	No	Yes	No
dUTP incorporation	No	No	No	No	Yes	Yes
Designed for multiplex PCR	No	No	No	No	No	Yes, up to 20 targets

Learn more at **thermofisher.com/fragmentanalysis**

applied biosystems