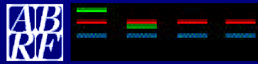
Practical Considerations for PCR and Electrophoresis in a Genotyping Core Facility

Lynn Petukhova

Starr Center for Human Genetics

Rockefeller University



Fragment Analysis Research Group

PCR ➾ Electrophoresis ➾ Analysis

This is the workflow in a typical Genotyping Core Facility. My talk will focus on the first two steps, PCR and electrophoresis.

PCR

- Selecting Markers
- Setting Up Panels
- Choosing a Type of STR
- Ramping Up Throughput

Fragment Analysis Research Group

For PCR, I will discuss these four topics.

Choosing Markers

Commercially Available Markers and Marker Sets

- Convenient
  - Sets are configured into panels
  - Individual markers can be ordered without sequence information

Custom Synthesized Primers

- Cost Effective
- Allow Flexibility
  - Design your own panels
  - Redesign primers
- Time Consuming
  - Requires sequence information of primers

Fragment Analysis Research Group

When we began genotyping, we used commercially available markers. These come prepackaged with specific labels that allow the markers to be organized into particular panels. It is very convenient ordering markers and getting started.

A few years ago, I realized that if we ordered the same markers as a custom synthesis, the cost savings would be huge, assuming that the order is consumed. In other words, for a very small project, it may not save money. Primer sequences are publicly available. It is somewhat time consuming finding and organizing this information for a custom synthesis order.

Primer Information

Marker Sets	Primer Sequences	CEPH Genotypes
Marshfield http://research.marshfieldclinic.org/genetics/sets/combo.html	**Sequences for tri/tetra Markers** http://research.marshfieldclinic.org/genetics/sets/Primer10Frames.htm http://lpg.nci.nih.gov/html-chlc/GCT.seqtable.html	**Marshfield** http://research.marshfieldclinic.org/genetics/indexmark.htm
Research Genetics http://www.resgen.com/products/HuSCREENSET.php3	**Sequences for di Markers** http://lpg.nci.nih.gov/data/Markers/all.MFD.primers	**CEPH** http://www.cephb.fr/cgi-bin/wdb/ceph/systeme/form
Applied Biosystems http://www.appliedbiosystems.com/ga/gaww0002.htm#genescan	**Primer Query** http://gdbwww.gdb.org/gdb-bin/genera/generaSF/hgd/Amplimer?!action=queryform	

http://www.abrf.org

Fragment Analysis Research Group

These are the web sites I use most frequently in compiling information about markers. The list is organized by topic. First, I find out what markers I need for a project. Information about screening sets are in the first column, including marker names and allele ranges. Then I find the sequence of the primers, using web sites in the center column. Finally, genotypes for my control samples are downloaded from the sites listed in the last column.

Custom Synthesized Primers are Screened on CEPH DNA

- Control Values are as expected
- Label is correct
- PCR product peak heights
- Determine Quality of the Data

Fragment Analysis Research Group

Whenever we receive new primers in the laboratory, we set up reactions on a few control DNAs to assess their quality before using them on "real" samples. At times, we have found that control values do not match up. This was found to be due to errors in the public databases. On rare occasions, our primer vendor has used the wrong label on a primer. PCR peak heights are used to determine post-amplification pooling ratios. Data quality is assessed.

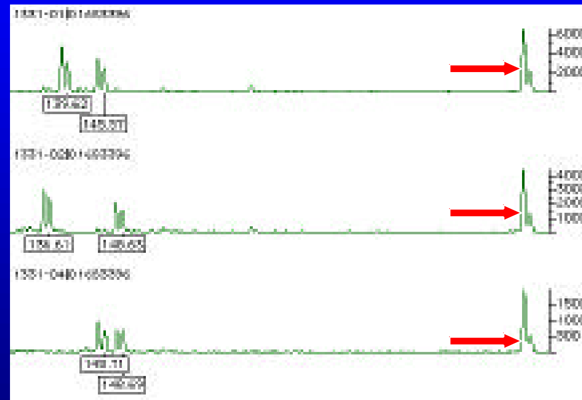
Problems with Data

- No Data or Low Signal
- Nonspecific Amplification
- Preferential Amplification

Fragment Analysis Research Group

In general, there are three types of problems we see with poor primer sets. No data or low signal can be due to one of two things: first, the primer set does not work well under our conditions. As a core lab, we use one PCR buffer and one thermal cycling program. If a marker doesn't work under our conditions, we replace it with one that will. Second, the lack of amplification could be due to synthesis problems. I have not examined this possibility. I mention it simply as something to consider.

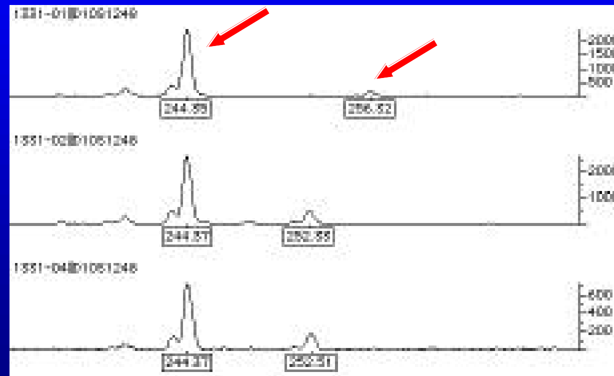
Nonspecific Amplification



Fragment Analysis Research Group

Nonspecific amplification is evidenced by the presence of one or more extra peaks in a sample. Here you can see the microsatellite amplification product on the left. The peak on the right is present in all samples. We find these extra peaks in about 10% of our markers. In a case such as this, where there is only one extra peak, we will generally use the marker in a panel that does not have a similarly labeled marker in the size range of the artifact peak. If a marker has many artifact peaks, we will replace it.

Preferential Amplification



Fragment Analysis Research Group

Preferential amplification occurs in a heterozygote when the smaller allele is amplified preferentially to the larger allele. This affect is marker specific. We have found that lowering the reaction stringency will decrease this effect. We add 5-10% glycerol to the reaction.

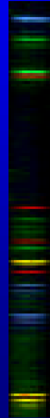
PCR

- Selecting Markers
- Setting Up Panels
- Choosing a Type of STR
- Ramping Up Throughput

Fragment Analysis Research Group

The next part of my talk will focus on a few issues related to setting up panels.

Configuring Panels



A set of markers that can be loaded together in one lane

- Allele size ranges do not overlap
- Different labels

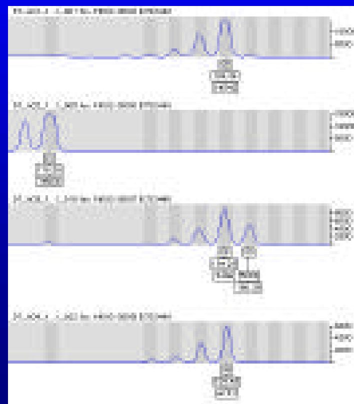
Fragment Analysis Research Group

Here is the definition of a panel. Here is how you can make a panel:

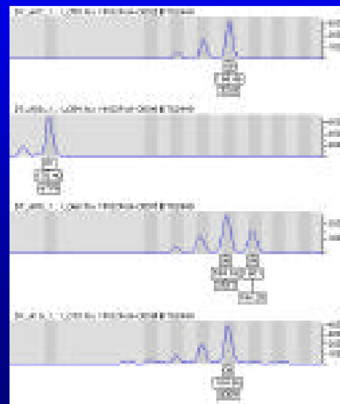
1. Compile markers with different size ranges
2. Distinguish those with overlapping size ranges by using different labels

Signal Quenching

Diluted 1:8 in water

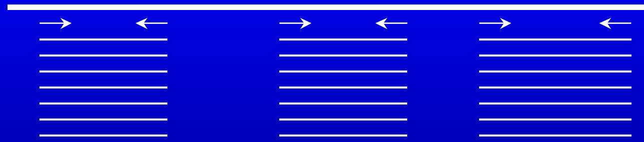


Pooled with 7 other PCRs



There are several dye sets available for genescan. We switched to the Ned set when we began using the 3700. There is one artifact that we noticed at that time. When a PCR product is diluted in water (run segregated), it will have much higher signal strengths than when it is diluted with other PCR products (run in a panel).

Multiplex PCR

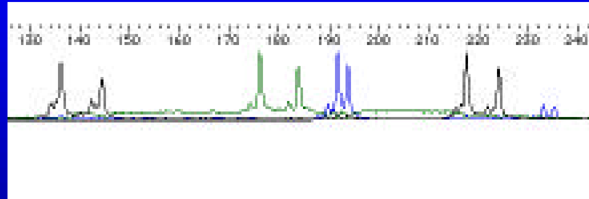


Markers are pooled before PCR for simultaneous amplification.

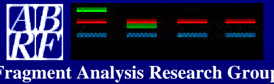
Fragment Analysis Research Group

When primer sets can be pooled together before PCR in a process known as multiplex PCR.

Multiplex PCR Data

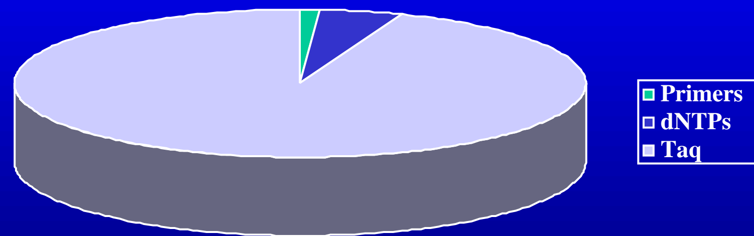


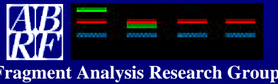
Doug Bintzler, University of Cincinnati



Here is an example of the result of a multiplex PCR reaction, courtesy of Doug. These five primer sets were amplified in a single reaction.

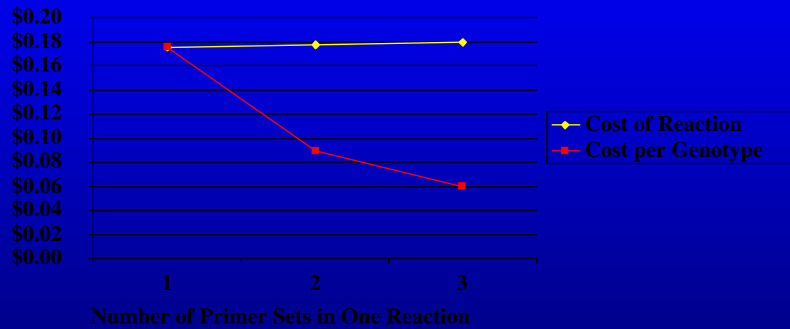
Cost Breakdown of PCR reagents

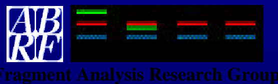




Multiplexing cut the cost of PCR reagents. In the reaction, you increase the cheapest reagent (primers) modestly, but you use the same amount of Taq enzyme, which is the most expensive PCR reagent.

Multiplex PCR Reduces Cost of PCR reagents





This dramatically drives down the per genotype cost of your PCR, as illustrated in this graph. Per genotype cost of the PCR reagents is on the y-axis. Number of primer sets in the reaction is on the x-axis. As you increase the number of markers, there is a modest increase in reaction cost (yellow line), but a huge decease in the per genotype cost (in red).

PCR

Selecting Markers

Setting Up Panels

Choosing a Type of STR

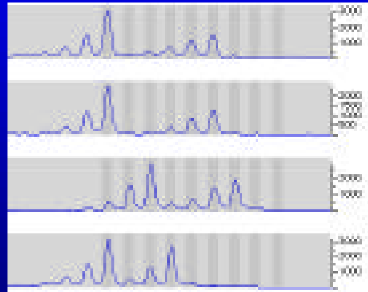
Ramping Up Throughput

Fragment Analysis Research Group

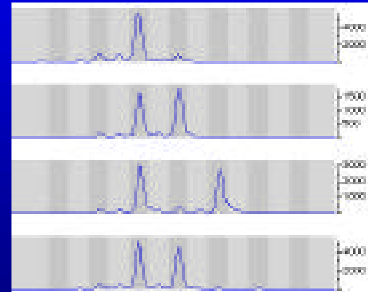
I will discuss issues that we considered when choosing a primary type of short tandem repeat marker for our laboratory.

Type of STR

Dinucleotide

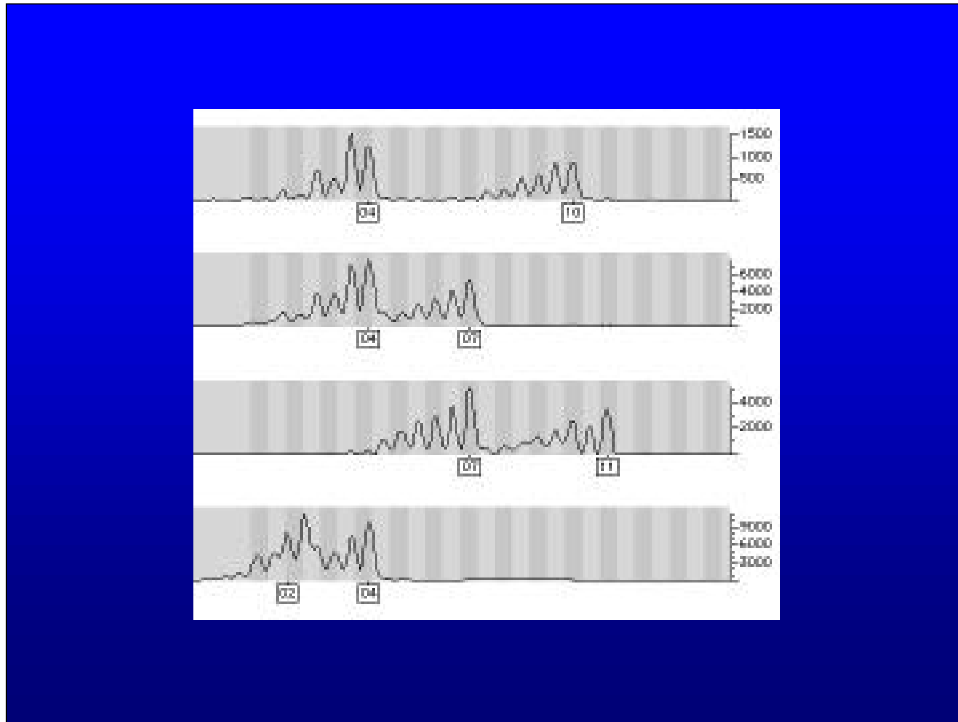


Trinucleotide and Tetranucleotide



Fragment Analysis Research Group

There are three types of repeat units in microsatellites commonly used in genetic studies. Two 10cM screening sets that are frequently used differ in the primary type of STR. While the screening set from Applied Biosystems consists entirely of dinucleotide repeats, the set developed at Marshfield contains mostly tri and tetranucleotide markers (78%).

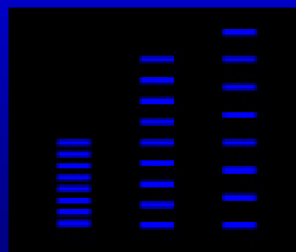


This is an example of some ugly data that we encountered when we used dinucleotide markers. In addition to the stutter bands you see every two base pairs preceding the true product, there is one base pair stutter which is due to Taq enzyme nontemplated nucleotide addition. This type of data inhibits automated scoring and increases genotyping errors.

Dinucleotide Markers Increase Throughput

Size of Repeat	2 bp	3 bp	4 bp
Number of Alleles	8	8	8
Size Range of Alleles	16 bp	24 bp	32 bp

Simulated Gel Image



The advantage to using dinucleotide markers rather than tri or tetranucleotide markers is that they allow for a higher degree of multiplexing. Because the repeat unit is shorter, allele ranges are smaller, and this allows you to run more markers in a single lane. This is illustrated above. A dinucleotide marker with eight alleles will have an allele size range of 16 base pairs. A tetranucleotide marker with eight alleles will have a size range of 32 base pairs.

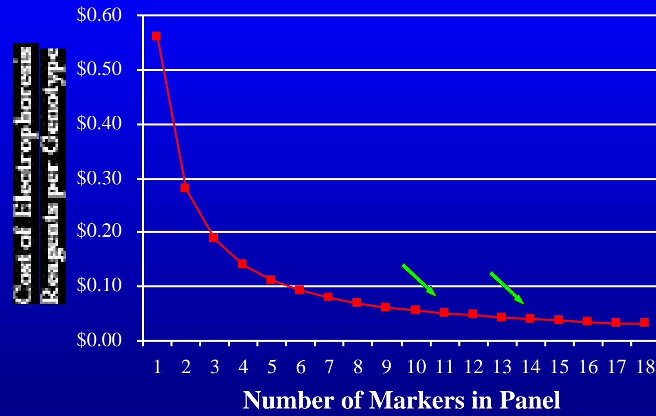
Create Larger Panels

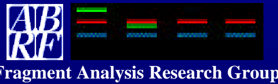
	Markers	Panels	Markers per Panel
ABI (100% dinucs)	400	28	14
Marshfield (22% dinucs)	405	36	11

Fragment Analysis Research Group

The effect on throughput that the repeat type has is illustrated in this table, comparing two different 10cM screening sets. Both sets have approximately the same number of markers. The screening set that consists entirely of dinucleotide marker contains 28 panels for an average of 14 markers per panel whereby the screening set with mostly tri and tetranucleotide markers contains an average of 11 markers per panel.

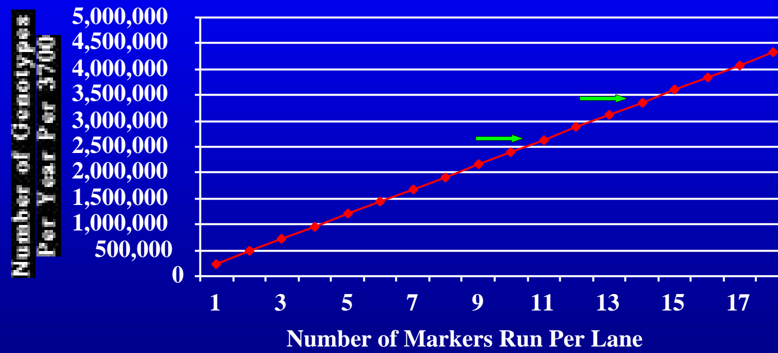
Larger Panels are More Cost Efficient

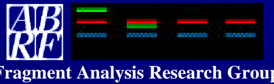




Here you can see the effect this has on cost. Plotted on the y-axis if the cost of electrophoresis reagents per genotype. In general, the more markers that are run in a lane together, the lower the cost. The difference between and 11 marker panel and a 14 marker panel comes to about a penny a genotype.

Larger Panels Increase Throughput





When you look at the effect on throughput, you can see here that utilizing 14 marker panels rather than 11 marker panels will result in an increase of 960,000 genotypes per year. The cost savings on electrophoresis reagents, $0.01 per genotype, is now more significant: $9600 per year.

PCR

- Selecting Markers
- Setting Up Panels
- Choosing a Type of STR
- Ramping Up Throughput

Fragment Analysis Research Group

I will touch briefly on the topic of automation.

PCR needed for one 3700

Application	Sequencing	Genescan
Runs per Day	8	10
Lanes per Day	768	960
Reactions per Day	768	8640-14,400

Fragment Analysis Research Group

Many Genotyping Core Facilities have their roots in a DNA Sequencing Facility. Here you can see the workflow difference between the applications. With sequencing, one PCR is needed per lane, so the bench work bottleneck is with the electrophoresis instrument. With genescan, multiple PCRs are electrophoresed together, shifting the bottleneck upstream to reaction set up and increasing the necessity of automating pipetting procedures.

There are many options available for automating PCR set up. I will show you the solution that I have found.

Automating the PCR Workflow

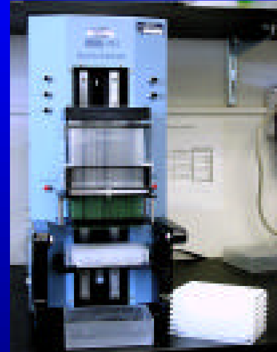
- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer

Fragment Analysis Research Group

The PCR workflow can be broken down into these five steps.

PCR Workflow

- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer



Fragment Analysis Research Group

We use a Hydra 384 from Robbins Scientific for aliquoting DNA. DNA samples are submitted in 384 well titer plates. We stamp out 1uL aliquots for the PCR.

PCR Workflow

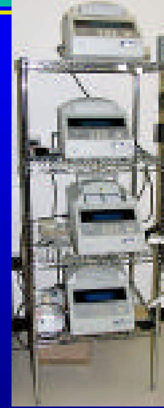
- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer

Fragment Analysis Research Group

We use a Tecan Genesis100 to distribute cocktail mixes, which contain DNA polymerase, primers, dNTPs and PCR buffer. The instrument has eight pipetting channels and has a deck that has been customized for pipetting into 384 well plates. It takes about 2.5 minutes per plate. It sets up 5uL reactions.

PCR Workflow

- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer

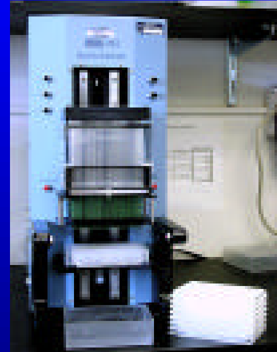


Fragment Analysis Research Group

We have 4 Perkin Elmer 9700s. This setup is a great space saver!

PCR Workflow

- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer

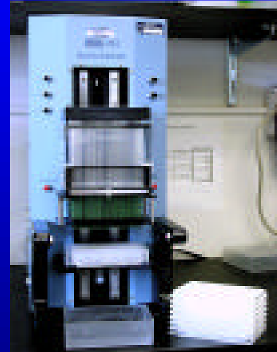


Fragment Analysis Research Group

We use our Hydra 384 for pooling PCR products.

PCR Workflow

- Aliquot DNA
- Aliquot PCR cocktail mix
- Thermal cycler
- Pool PCR products
- Aliquot pooled product to loading buffer



Fragment Analysis Research Group

And for aliquoting pooled products to loading buffer.

PCR ⇒ Electrophoresis ⇒ Analysis

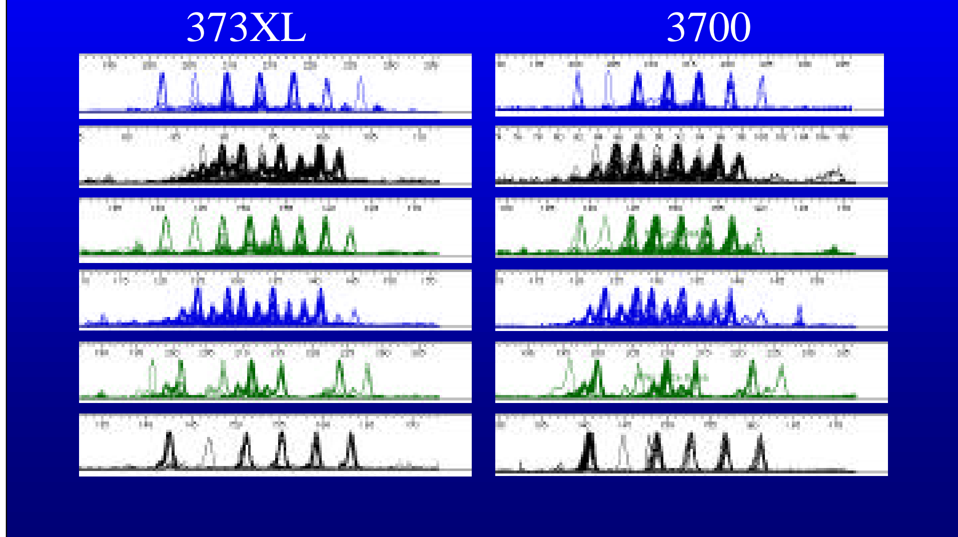
About 2 years ago we switched from slab gel instruments, 373XLs, to a capillary instrument, a 3700. In order to understand how a change in technologies would affect our data, we conducted a study to directly compare the two platforms.

Comparing 373XL to 3700

- 1152 people genotyped with 6 markers
  - 6912 PCRs set up
  - 72 negative controls
  - 218 failed reactions
- 6622 genotypes

Fragment Analysis Research Group

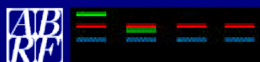
We set up a six marker panel on 1152 samples giving us a set of 6622 genotypes that were electrophoresed on each instrument. The PCR was set up only once.



Here are the six markers as they appeared on each instrument. As you can see, there is a mobility shift between the instruments that varies by marker and ranges from one base pair up to as many as four base pairs.

Failure Rates

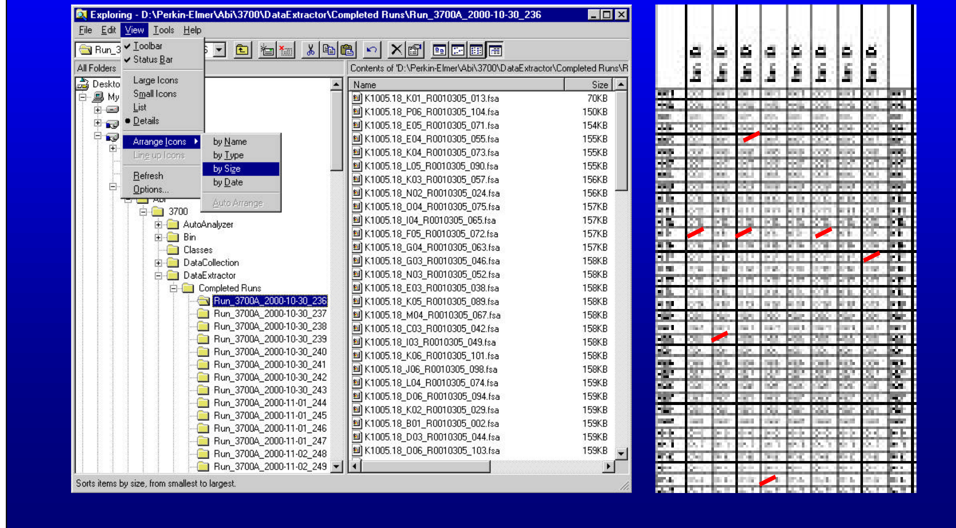
Total Number of Reactions	**6622**	
Samples with no data 373XL	**89**	**1.34%**
Samples with no data 3700	**290**	**4.38%**



Fragment Analysis Research Group

There is a dramatic increase in failure rates from the 373 to the 3700. There were 89 samples that did not yield any data on the 373s, but produced data on the 3700. These samples included those with low signals (3700 is more sensitive), as well as samples that were lost to problems loading the gels or whose data was obscured by gel artifacts (lane overflow, bubbles, etc.).

Reducing Failure Rate on 3700



We choose to monitor the failures in order to detect any patterns. As a fast way to achieve this, at the end of a run, we use NT explorer to view the contents of the run folder, sorted by file size. Samples that could not be analyzed are much smaller in size. The capillary number is in the file name. We track these failures in a log sheet, shown on the right. By doing so, we have noticed that certain capillaries are associated with failures. By turning off these capillaries once they are identified, we have reduced the failure from 4% to 1%.

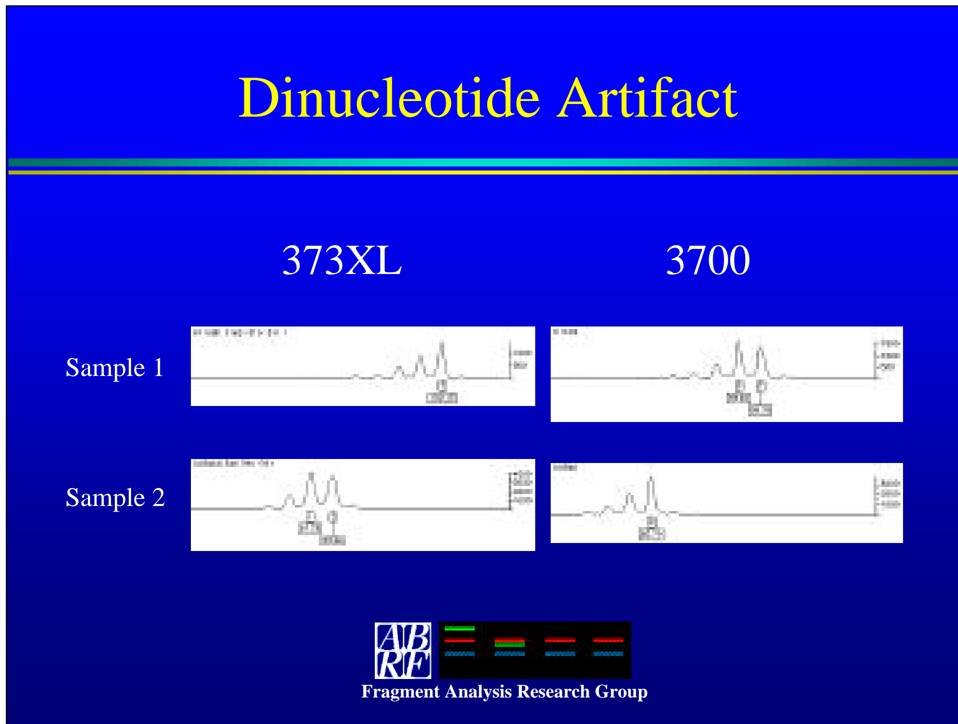
Platform Dependant Errors

20 Samples with Discrepant Genotypes

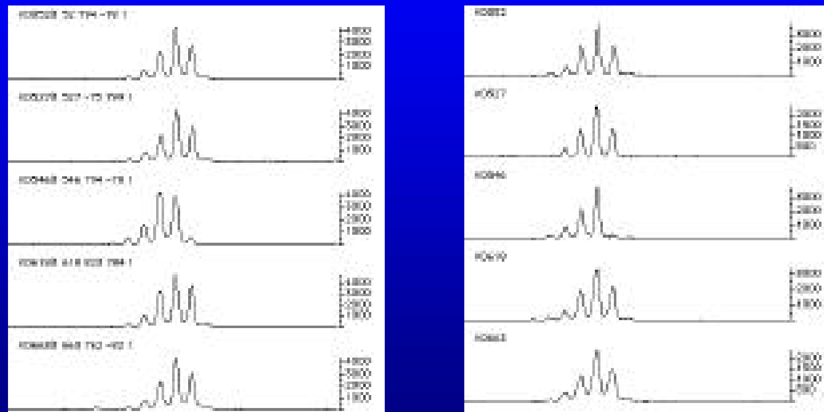
	373XL	3700
pull-up	6	1
overflow	9	0
dinucleotide	1	1
lane standard	0	1
spike	0	1

Fragment Analysis Research Group

There were 20 samples that produced different genotypes on each platform. These data for these samples were closely examined to determine the source of the problem. The problems fell into five categories listed in this slide. Two are unique to slab gels, two are unique to capillary platform and one appears to be platform independent. Pull-up is attributed to a bad matrix file and can be caused by a bad file, a corrupted file, or overloaded product. It is evidenced by the presence of one PCR product peak in more than one color (for example, a fam product produces a peak in blue and green and yellow). Overflow is due to gel loading: either leaky lanes, or incorrect loading, or can be due to bleedthrough when an adjacent lane has very high peak heights. The following slides will depict the last three types of problems.

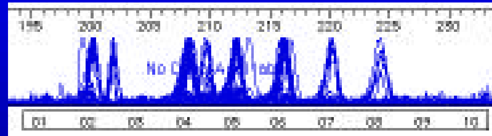


The platform independent problem appeared with a dinucleotide marker and was found in two different samples. Both samples appeared to be homozygous on one instrument and heterozygous on the other instrument. Because the PCR was only performed once, this artifact is thought to result from the electrophoresis. Gel files were checked to rule out lane tracking as the source of error. The true genotype of these samples is not known.



Sample 2 from the previous slide is shown here in a group of samples that were electrophoresed together.

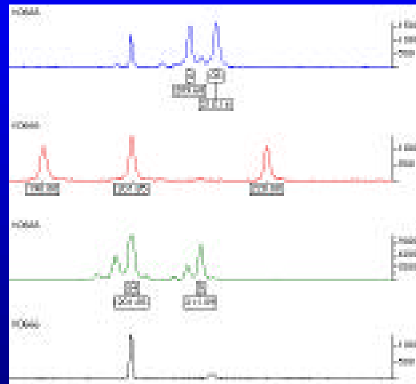
Sizing Errors



Fragment Analysis Research Group

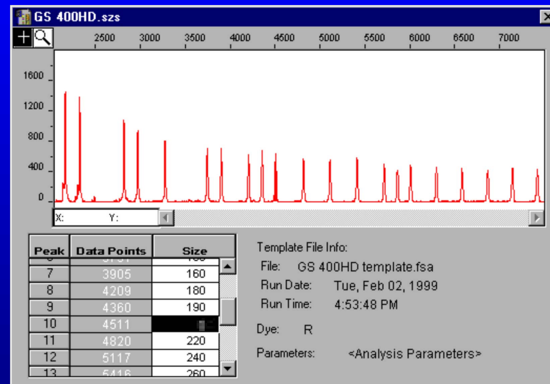
Two of the genotype discrepancies found were due to sizing errors on the 3700. Shown here is an overlay of 96 samples for one marker. You can see that some of these PCR products are falling outside of the allele bins.

Pull-up Peak



Here is an example of one such sample. When the lane standard for this sample was examined, it was found that the fragment at 200bp is incorrectly sized. One can also see that a hex-labeled marker with a fragment at 200bp is overloaded and causing bleedthrough in the yellow and blue. These bleedthrough is thought to be responsible for the incorrect sizing of the lane standard peak.

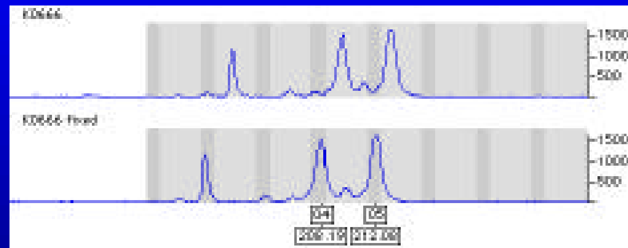
Correcting the Error



Fragment Analysis Research Group

This problem can be corrected by analyzing the file with a lane standard definition that does not include the 200bp fragment.

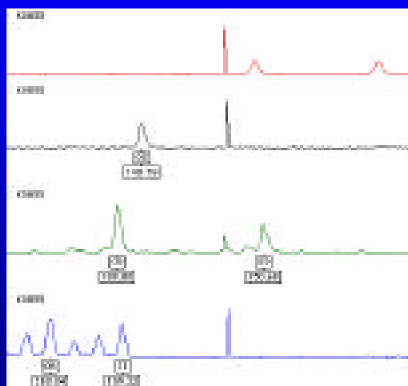
Corrected



Fragment Analysis Research Group

The top illustration depicts sizing when the 200bp fragment is defined in the lane standard. Below, the product is correctly sized when the 200bp fragment is not defined.

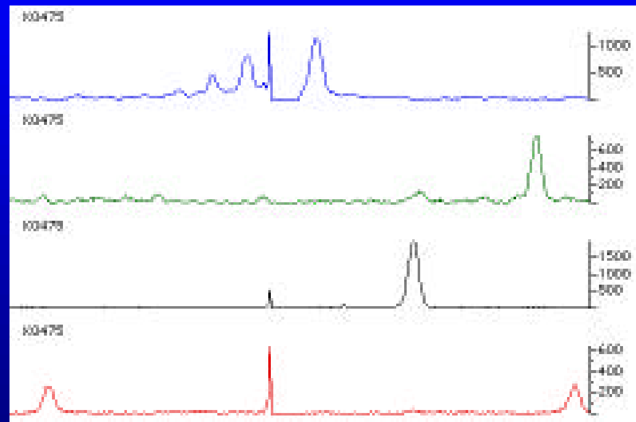
3700 spikes



Fragment Analysis Research Group

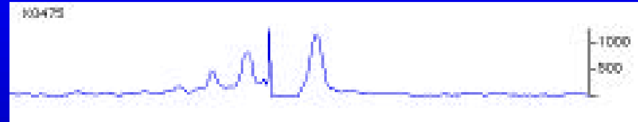
This spike is an artifact that is seen from time to time in data generated on our 3700. It is thought to be due to an air bubble. It is a narrow peak that appears in all four colors. Most of the time it does not interfere with data interpretation.

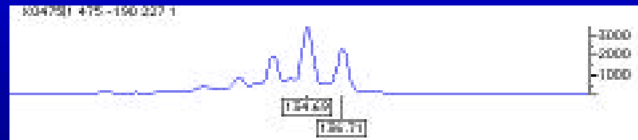
Spike Artifact



In one sample in this study, it did interfere.

Spike Artifact





Fragment Analysis Research Group

The spike obscured the view of the smaller allele.

Acknowledgements

Sandra Oplanich

Dvora Shmulewitz
Stephanie Boger
Zhihua Hans

Peter C. Verlander

Markus Stoffel
Jeffrey Friedman

Here I acknowledge and thank all of the people that contributed to the study comparing electrophoresis platforms.