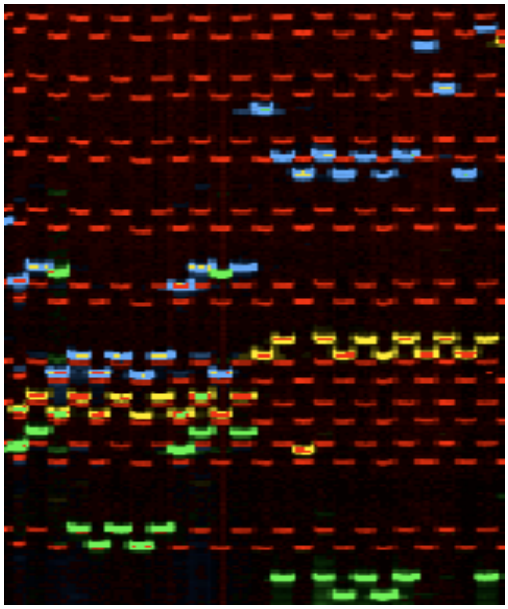
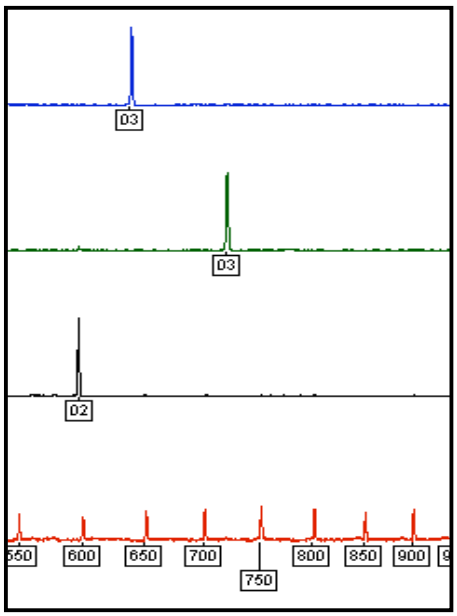




Multilocus Variable Number Tandem Repeat

Genotyping of

Mycobacterium tuberculosis

Technical Guide

86	2	2	5	3	2	5	1	5	3	3	2	3	2	3	2	3	6	8	3	4	4	2	3
87	2	2	3	3	2	5	1	4	3	3	2	3	2	3	2	3	6	2	3	4	4	2	3
88	2	2	7	2	2	5	1	1	3	2	2	3	3	5	4	3	9	2	4	4	4	2	3
90	2	2	3	3	2	5	1	7	3	2	3	3	2	4	4	2	12	6	4	4	4	2	3
92	2	3	4	4	2	4	2	2	4	2	2	1	3	4	3	2	5	8	2	3	4	2	2
93	2	5	4	3	2	11	2	2	3	6	3	3	3	4	4	8	6	3	5	3	1	4	3
95	2	2	4	2	2	6	1	5	3	3	2	1	5	4	4	3	2	4	2	4	1	1	5
96	2	2	3	1	2	5	1	5	3	3	2	2	2	4	2	2	5	3	2	4	2	2	3
97	2	1	3	3	1	5	1	2	3	4	2	3	3	4	1	3	5	2	4	4	2	1	3
98	2	2	4	2	2	6	1	6	3	3	2	1	4	2	4	4	2	4	2	4	1	2	3

Philip Supply

INSERM U629

Institut de Biologie/Institut Pasteur de Lille

May 2005

Acknowledgements

I would like to thank first Camille Locht for his vision, continuous support and commitment for the development of mycobacterial Variable-Number Tandem Repeat typing.

I would like also to thank especially Edith Mazars, Sarah Lesjean, Eve Willery, Caroline Allix, Mara Cardoso-Oeleman and Evgueni Savine for their valuable technical contribution, their work and personal dedication, as well as Vincent Vatin for providing us access to sequencers and for his expert advises in sequencer utilization. Marie-José Truong is thanked for her efficient organizational help.

I gratefully acknowledge Maria Cristina Gutierrez for her support and for helpful discussions about the manual typing method, and Mickaël Vanrumbeke (Applera/Applied Biosystems France) for useful advises.

Finally, I would like to thank the numerous collaborators involved in our work, and in particular, Kristin Kremer for the gift of her valuable reference strain collection.

Without the persons mentioned above, the realization of this guide would not have been possible.

To C., E., C., A. and A.

Table of contents

Preface

What you should know before beginning:

This manual is written for principle investigators and laboratory staff who are planning to perform Variable Number Tandem Repeat genotyping of *Mycobacterium tuberculosis* complex isolates, using agarose gel electrophoresis or using a 3100 Genetic Analyzer. Although the general principles are the same, some adaptations may be needed for 3130, 3100 Avant, and 3700 and/or other software versions.

Before attempting the procedures in this manual, you should be familiar with the following topics:

1. General techniques and safety procedures for obtaining and handling *Mycobacterium tuberculosis* complex isolates.
2. General techniques and safety procedures for handling DNA samples, performing PCR, preparing PCR products for electrophoresis, and performing electrophoresis.
3. Windows operating systems
4. The basics for installing, operating and maintaining a 3100 Genetic Analyzer. Detailed information for these procedures are given in Applied Biosystems' manuals.

1 Introduction

Variable Number Tandem Repeat (VNTR) sequences have emerged as valuable markers for genotyping of several bacterial species, especially for genetically homogeneous pathogens such as *Bacillus anthracis* (1, 2), *Yersinia pestis* (2, 3) and the *M. tuberculosis* complex members (see below). VNTR genotyping basically rely on PCR amplification using primers specific for the flanking regions of the VNTRs and on the determination of the sizes of the amplicons, after electrophoretic migration. As the length of the repeat units is known, these sizes reflect the numbers of the amplified VNTR copies. The final result is a numerical code, corresponding to the repeat number in each VNTR locus. Such numerical genotypes are intrinsically portable and are thus particularly convenient for both intra- and inter-laboratory comparative studies (4, 5). In addition, compared to IS*6110*-RFLP, MIRU-VNTR typing has the advantages of being faster, and appropriate for virtually all *M. tuberculosis* isolates, including strains that have a few IS*6110* copies (4, 5).

Initial VNTR typing systems for *M. tuberculosis* complex strains made use of very limited sets of loci (4, 6-9), which turned out to not be sufficiently discriminatory (10). More extensive sets of VNTR loci have been described subsequently (11-15), including a system based on 12 loci (14), which has been shown to be applicable for reliable genotyping and molecular epidemiology studies of *M. tuberculosis* (5, 16). These loci contain VNTR of genetic elements named Mycobacterial Interspersed Repetitive Units (MIRUs) that are located mainly in intergenic regions dispersed throughout the *M. tuberculosis* genome (14, 16, 17). As the other VNTRs sequences mentioned above, the lengths of MIRU repeat units are in the range of 50-100 bp, and belong therefore to the "minisatellite" VNTR category (16). **All above loci are collectively designated as MIRU-VNTR loci in this guide.**

A MIRU-VNTR-based high-speed genotyping system has been developed, which combines the analysis of multiplex PCRs for the target loci on a fluorescence-based DNA analyzer with computerized automation of the genotyping (16). Both this system and the simpler system using electrophoresis with agarose gels are highly reproducible at intra- and inter-laboratory levels (14,

18). A recent population-based study indicated that the use of the 12 locus-based MIRU-VNTR typing as a first-line method, together with spoligotyping, provides adequate discrimination in most cases for large-scale, prospective genotyping of *Mycobacterium tuberculosis* in the United States. IS*6110* fingerprinting can subsequently be used as a secondary typing method to type the clustered isolates, when additional discrimination is needed (19). Other studies have shown the interest of this typing method for clinical mycobacteriology (*e.g.* 20) or for local outbreak investigation (21).

In order to further reduce the numbers of isolates to be analyzed by IS*6110* fingerprinting, we have recently evaluated the additional information provided by a set enlarged to 29 loci, using a total of 824 tubercle bacillus isolates, including representatives of the main lineages identified worldwide so far (Supply *et al*., in preparation). Five loci (i.e. QUB-3232, -3336, -1895, -18 (alias VNTR 1982), -11a (alias VNTR 2163a)) were excluded for lack of robustness and/or stability in serial isolates or isolates from epidemiologically-linked patients. The use of the 24 remaining loci increased the number of types by 40% - and by 23% in combination with spoligotyping - among isolates from cosmopolitan origins, as compared to those obtained with the original set of 12 loci. Consequently, the clustering rate was decreased by four-fold – by three-fold in combination with spoligotyping - under the same conditions. A discriminatory subset of 15 loci with the highest evolutionary rates was then defined, that concentrated 96 % of the total resolution obtained with the full 24-loci set. Its predictive value for evaluating *M. tuberculosis* transmission was found to be equal to that of IS*6110* RFLP typing, as shown in a companion population-based study. This 15-loci system is therefore proposed as the new standard for routine epidemiological discrimination of *M. tuberculosis* isolates, and the 24-loci system as a high-resolution tool for phylogenetical studies (Supply *et al*., *J. Clin. Microbiol*., in press).

A table giving the correspondence between different nomenclatures for the 24 loci retained is given in Appendix 4.1.

2 Methods

Note: Aerosol resistant pipet tips are used at all experimental steps before PCR amplification, to avoid potential contamination problems.

2.1 *DNA extraction and dilution*

2.1.1 Principle

As it is based on amplification by PCR, MIRU-VNTR typing can be performed on heat inactivated mycobacterial colonies or mycobacterial pellets from liquid cultures without extensive DNA purification. It can be applied to various biological materials, including non-viable material, permitting for instance retrospective analyses of stocks of non-viable cells.

Note: alternatively, purified DNA obtained by the internationally standardized method as described by van Soolingen *et al*. (22) or by other standard methods for IS*6110*-RFLP analysis can also be used.

2.1.2 Procedure

1. Resuspend mycobacterial colonies grown on solid media (*e.g.* Löwenstein-Jensen) medium or mycobacterial pellets obtained from liquid cultures (*e.g.* MGIT) into 200 µl 10 mM Tris-HCl, 1 mM EDTA (pH 7.0), in a screwed cap tube or a safe lock tube.
2. For colonies recovered from solid media, go to step 3. Re-centrifuge mycobacterial pellets obtained from liquid cultures at 15,000 g x 5 min, discard the supernatant and resuspend the bacterial pellet into 200 µl 10 mM Tris-HCl, 1 mM EDTA (pH 7.0).
3. Incubate at 95°C for 45 min, using an oven or a PCR cycler with a hot lid, if available in the microbiological security facility.
4. Centrifuge the suspension at 15,000 g x 1 min, to pellet the cell debris.
5. Harvest the supernatant containing the DNA and transfer into a new tube.
6. Store concentrated stocks at –20 °C until further use, or dilute at 1:50 into sterile water in a new tube. Diluted solutions can be also stored at -20 °C.

2.2 PCR amplification

2.2.1 Principle

PCR amplification of different VNTR regions is performed using primers specific for the flanking regions of each VNTR region.

When analysis of the PCR products is done using a DNA sequencer, 8 different multiplex PCRs are performed in order to analyze eight groups of three loci simultaneously. For each multiplex, one primer per oligonucleotide pair is tagged with a specific fluorescent dye (appendix). In case of amplification failure of some loci in some multiplex reactions, these loci are then usually amplified separately by simplex PCR, in a second round.

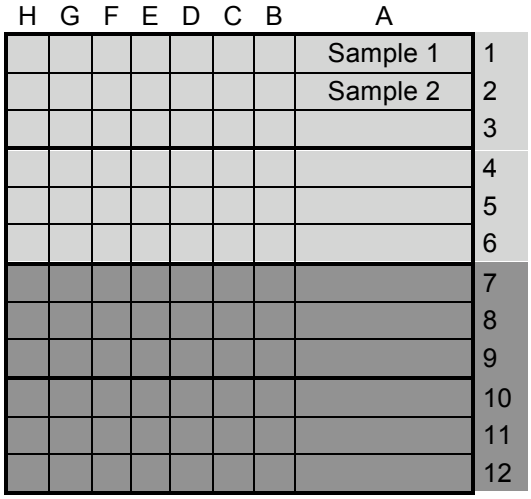
When analysis of the PCR products is done using only electrophoresis with agarose gels, separate amplification of each locus is performed, using unlabeled oligonucleotides.

2.2.2 Procedure

2.2.2.1 Preparation of a PCR spreadsheet

1. Prepare a PCR spreadsheet, indicating the position of each sample in the PCR 96-well microplate (Figure 1).
2. If different multiplex or simplex PCRs are performed in a same microplate, use specific extensions after sample names (e.g. sample 1-mix1) or use different colors to specify the different reactions, as shown in the example below. Samples analyzed with the same multiplex or the same locus should be grouped together for convenience.
3. Include positions for a negative control (sterile water) and a positive control (H37Rv or BCG Pasteur) for each multiplex or simplex set, to validate the analysis.

Figure 1 PCR spreadsheet model



2.2.2.2 Preparation of PCR premixes

PCR premixes can be prepared extemporaneously just before the addition of genomic DNA, or stock solutions can be prepared, aliquoted and stored at -20°C until further use. The final concentration of $MgCl_2$ varies from 1.5 mM (default concentration using the 10 X buffer) to 3 mM (by including additional $MgCl_2$), according to the multiplex or the MIRU locus. The final volume per reaction (after addition of DNA) is 20 µl.

The use of Qiagen Hotstart Taq Polymerase kit including Q solution is strongly recommended. If this kit is not used, PCR failures may be observed, especially for MIRU-VNTR locus 20, and more intense stutter peak ladders (see below) may be seen, leading to possible misinterpretation problems. **Note: the Qiagen PCR Multiplex PCR kit may be specifically used for mix 5**, to reduce pronounced stutter peaks seen with large alleles of locus 4052 (see Annex 5).

Using our conditions, Q solution is the limiting reactive in the Qiagen kit. However, the use of Betaine (identical to Q coumpond, according to Sigma) at a final concentration of 1M in replacement of Q solution is a good and cheap alternative.

1. In a DNA-free area, prepare the PCR reaction premixes for the different multiplex and simplex reactions, according to Table 1, Table 2, and Table 3. **Important:** to take into account void volumes, prepare a 5-10 % proportion in excess of the volume needed.

2. Label a 96-well PCR microplate with date and experiment numbers. Mark the middle of the microplate for better visualization of the positions. Optionally, indicate separations between zones with different multiplexes as shown in Figure 2.
3. Dispense 18 µl of the PCR premix into each well of the microplate.

Figure 2 Delineation of different multiplex zones on PCR microplate

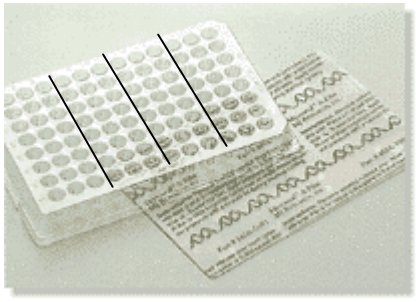


Table 1 Volumes (µl) for the discriminatory multiplex premixes

mix	1	2	3	4	5
Loci	4-26-40	10-16-31	0424- 0577- 2165	2401-3690- 4156	2163b- 1955- 4052
$MgCl_2$ final concentration	3mM	2mM	1,5 mM	3mM	1,5 mM
H2O	7,5	8,3	8,7	7,5	8,7
Buffer 10 X	2	2	2	2	2
Q Solution 5x	4	4	4	4	4
$MgCl_2$ 25 mM	1,2	0,4	0	1,2	0
DNTP 5mM	0,8	0,8	0,8	0,8	0,8
Primers EACH[a]	0,4	0,4	0,4	0,4	0,4
Hotstart DNA pol	0,08	0,08	0,08	0,08	0,08
Total premix	18	18	18	18	18

[a] Six in total, *i.e.* one forward and one reverse primer for each of the 3 pairs. Initial concentration for all unlabeled primers: 20 pmol/µl. Initial concentration for the labeled oligonucleotides: 2 pmol/µl for locus 0577, 3690 and 1955, 8 pmol/µl for locus 4052, 20 pmol/µl for locus 4156 and 4 pmol/µl for the other loci. See appendix for the primer sequence and labeling.
Note: the Qiagen PCR Multiplex PCR kit may be specifically used for mix 5, to reduce pronounced stutter peaks seen with large alleles of locus 4052 (see Annex 5).

Table 2 Volumes (µl) for the other multiplex premixes

mix	6	7	8
Loci	2-23-39	20-24-27	2347-2461-3171
$MgCl_2$ final concentration	2.5 mM	1,5 mM	2mM
H2O	7,9	8,7	8,3
Buffer 10 X	2	2	2
Q Solution 5x	4	4	4
$MgCl_2$ 25 mM	0.8	0	0,4
DNTP 5mM	0,8	0,8	0,8
Primers **EACH**[a]	0,4	0,4	0,4
Hotstart DNA pol	0,08	0,08	0,08
Total premix	18	18	18

[a] Six in total, *i.e.* one forward and one reverse primer for each of the 3 pairs. Initial concentration for all unlabeled primers: 20 pmol/µl. Initial concentration for all labeled oligonucleotides: 4 pmol/µl. See appendix for the primer sequence and labeling.

Table 3 Volumes (µl) for the simplex PCR premixes

Loci	4-26-40-2401-3690-4156	10-16-31-2347-2461-3171	20-24-27-0424- 0577-2165-2163b-1955- 4052	2-23-39
$MgCl_2$ final concentration	3mM	2mM	1,5 mM	2.5 mM
H2O	9,1	9,9	10,3	9,5
Buffer 10 X	2	2	2	2
Q Solution 5x	4	4	4	4
$MgCl_2$ 25 mM	1,2	0,4	0	0.8
DNTP 5mM	0,8	0,8	0,8	0,8
Reverse primer[a]	0,4	0,4	0,4	0,4
Forward primer[a]	0,4	0,4	0,4	0,4
Hotstart DNA pol	0,08	0,08	0,08	0,08
Total premix	18	18	18	18

[a] See tables above for the initial concentration. When analysis of the PCR products is performed using only agarose gel electrophoresis (manual typing), all primers are unlabeled.

2.2.2.3 Addition of genomic DNA

1. In another PCR product-free area, dispense 2 µl of extracted DNA previously diluted 1:50 into each well. If purified DNA is used, dispense 2 µl of solution at 1 ng/µl. Include H37Rv DNA and sterile water as positive and negative controls, respectively.
2. **Tightly** seal the PCR microplate using an adhesive PCR film, to avoid evaporation during amplification.

2.2.2.4 PCR amplification

1. Amplify the target loci using the PCR conditions indicated in Table 4. These conditions have been successfully tested on Hybaid PCR express and Perkin Elmer cyclers.
2. For analysis on automated sequencers, notice that the number of cycles may need to be adjusted depending on the DNA concentrations routinely obtained in the laboratory, i.e. amplification from weakly or highly concentrated mycobacterial cultures may require 40 cycles or only 25 (Mix 1, 2, 4, 6, 7, 8) to 30 (Mix 3, 5) cycles, respectively.
3. Store the PCR products at 4°C or -20°C until further use.

Table 4 PCR cycling conditions

15 min	95°C	
1 min	94°C	Cycle numbers: - Automated typing: 30 for Mix 1, 2, 4, 6, 7, 8 35 for Mix 3 and 5 - Manual typing: 40 cycles for all mixes
1 min	59°C	
1 min 30s	72°C	
10 min	72°C	
∞	4°C	

2.3 PCR product analysis using agarose gel electrophoresis

When each locus is amplified separately (i.e. by simplex PCR), the amplified fragments can be analyzed by electrophoresis using agarose gels. This method is inexpensive and easy, as it only requires a size resolution of about 50 bp (except for locus 4, see below). It is accurate, provided that adequate electrophoresis conditions and controls are used. It is suitable for laboratories with relatively low turnovers of isolates to be analyzed. The use of multi-channel pipettes compatible with gel combs is useful for both the speed and the reliability of the genotyping process.

This method is also used to size PCR fragments from reference strains, (in addition to H37Rv), selected to sample the allelic range of each MIRU-VNTR locus. When labeled with fluorescent dyes, the corresponding reference PCR products can then be used to calibrate the sizing using electrophoresis with your DNA sequencer.

The quality of the DNA fragment resolution and the use of a control for possible migration smiling effects are critical for the sizing accuracy. Therefore, long gels and sufficient electrophoresis time should be used. The use of Nu-Sieve agarose gels offering high resolution for small DNA fragments is recommended, especially for the detection of variant alleles in locus 4. However, electrophoresis grade agaroses such as Ultra Pure Electrophoresis Grade Agarose from Gibco-BRL can yield satisfactory results.

2.3.1 Gel preparation and electrophoresis

1. Prepare a 3 % suspension of Nu-Sieve agarose gel in 1 x TBE solution previously chilled at 4°C, to facilitate clump collapse.
2. Melt the agarose using a microwave oven, and agitate periodically until complete dissolution.
3. When the temperature is endurable to the touch, cast a 25-cm gel, using a shark tooth comb.
4. After solidification, place the gel into an electrophoresis tank containing 1 X TBE.

5. Load 10 µl of a 100-bp ladder size standard marker **in both external wells and in the central well** of the gel. The standard in central position can be used to detect possible migration smiling effects. A 50-bp ladder or a 20-bp ladder (however less easy to read sometimes) can be used in addition.
6. For each reaction, load, per well, a mixture of 2 ml of PCR product with 2 ml of loading buffer. **Migrate amplicons from a given locus together on a same gel**, rather than migrating different loci from a same isolate on a same gel. By this way, allelic assignation is facilitated by visualization of the band ladders generated by different repeat numbers (see 2.3.2).
7. Run at 120 V for 5 hours.
8. Stain the gel in sterile water containing 0.7 µg/µl of ethidium bromide for 15-30 min.
9. Expose the gel to UV light and take a photo.

2.3.2 Sizing and allele assignation

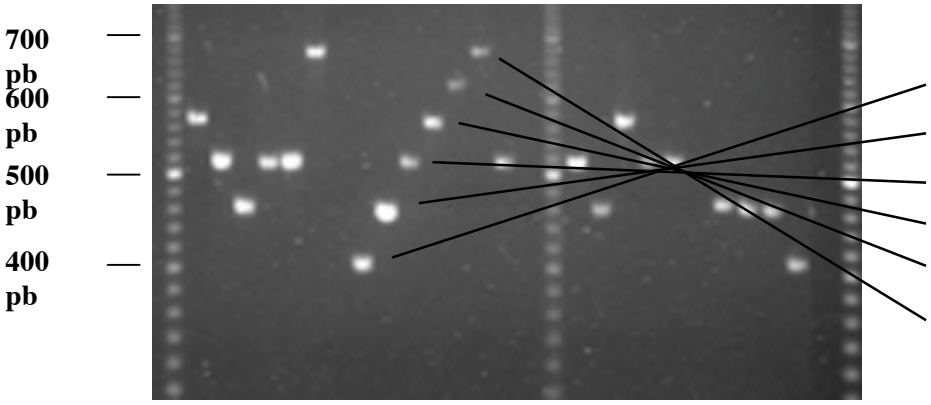
Most often, amplification results in single sharp PCR products (Figure 3). Depending on the locus, this sharp amplicon can be accompanied by a ladder of much lower intensity bands, called stutter peaks (see 2.3.3 and Figure 4). However, some problematic amplification of loci with large repeat numbers can result in a ladder of bands, with no clearly sharpest band or with a "bell-shaped" distribution of band intensities (see 2.3.3). **In this case, do not assign any result and re-amplify.**

1. Determine the size of the sharp PCR product by comparison with the position of the size standard marker.
2. Determine the corresponding repeat number, using the provided table containing the allele calling for each VNTR locus, as shown in Figure 3. **Verify the consistence with the usual allelic range.**
3. Verify that the allele assignation of the H37Rv control is correct.
4. Verify the consistence of the results by judging the incremental spacing between PCR products from different isolates (*i.e.* co-migrating fragments =

same alleles, fragments smaller by one repeat increment = allele-1, etc). **This control is best done starting from the smallest amplicon**, as sizing is usually more accurate for small products.

5. If classical stutter peak ladders (see 2.3.3 and Figure 4) are present, they can also be used as internal sizing controls to verify the consistency of the allele assignation. Similarly, this control is best done starting from the smallest stutter peak.

Figure 3 Example of allele identification for MIRU-VNTR locus 40



Allele	MIRU 40
0	354
1	408
2	462
3	516
4	570
5	624
6	678
7	732

2.3.3 Detection and interpretation of stutter peaks

Stutter peaks are common during genotyping of tandem repeat sequences, and mostly reflect artifactual strand slippage of the polymerase during PCR. Such stutter peaks are also quite frequently observed for PCRs of various MIRU-VNTR loci, more often when containing large repeat numbers.

In most cases, they can be easily diagnosed, as they appear as a ladder of much lower intensity peaks, corresponding to sizes of PCR fragments that lack one or more repeats, or more rarely that contain one or more additional repeat. The positions of stutter peaks can be used to confirm the allelic assignation of the principle fragment (see point 5 in 2.3.2).

As indicated above, sub-optimal amplification of loci with large repeat numbers (such as 4052, alias QUB 26) can sometimes result in a band ladder with no clearly sharpest band or with a "bell-shaped" distribution of band intensities (see next section). This can be typically seen when the

Q buffer is not used. **In this case, do not assign any result and re-amplify. Optimal amplification will often result in a single sharp band, with a highest size than could perhaps have been anticipated based on the initial band ladder pattern.**

Figure 4 Example of stutter peak detection in a MIRU-VNTR

The example shown corresponds to amplicons from MIRU-VNTR locus 27. Dotted arrows show positions of stutter peaks. The size increments between the stutter peaks correspond to the size of one repeat unit (53 bp in this case). The size standard (M) is a 20-bp ladder.

Allele	MIRU 27
0	498
1	551
2	604
3	657
4	710
5	763

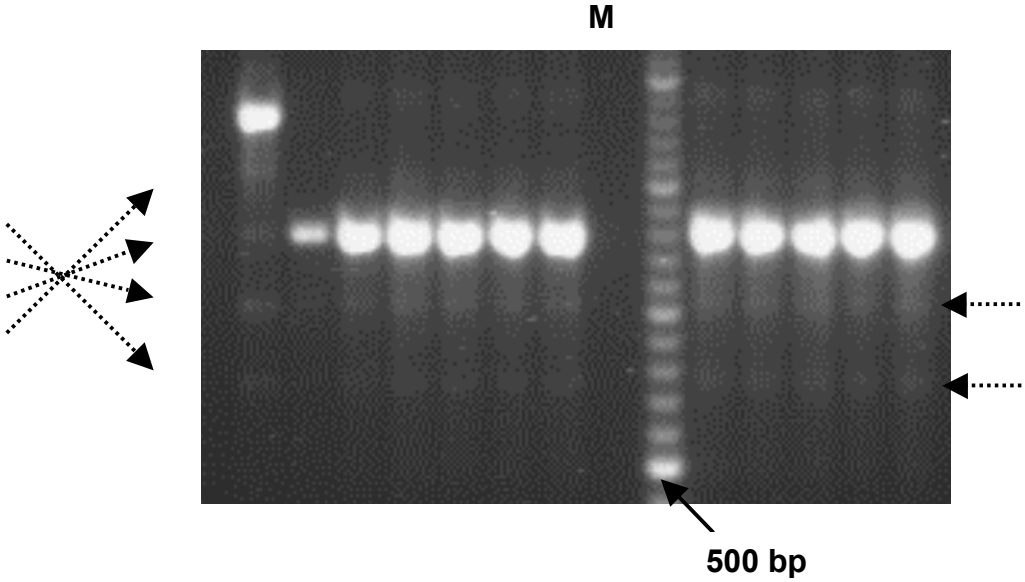
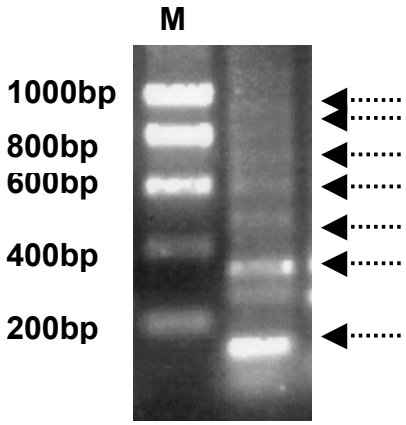


Figure 5 Example of "bell-shaped" ladder

The example shown corresponds to fragments amplified from MIRU-VNTR locus 4052 (alias QUB-26). Arrows show positions of stutter peaks. The size increments between the stutter peaks correspond to the size of one repeat unit (111 bp in this case). M, size standard (M).



2.4 PCR product analysis using capillary electrophoresis

2.4.1 Principle

The PCR fragments labeled with the three different fluorescent dyes from each multiplex are combined with an internal size standard, labeled by a fourth dye and then analyzed in individual capillaries on a DNA sequencer for size determination.

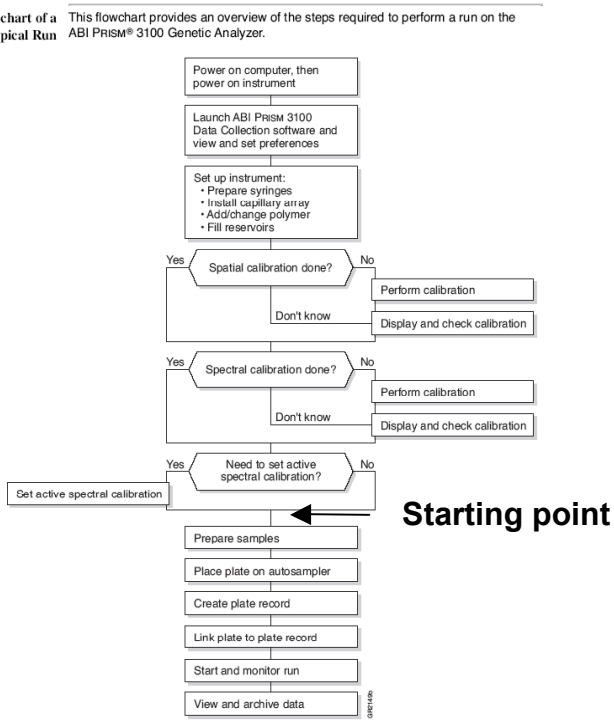
2.4.2 Capillary electrophoresis

The procedure is described for ABI 3100 sequencers, using the Data Collection software version 1.1. It is applicable to 3100 Avant and/or to software versions 2.x with minimal adaptations. Some additional adaptations may be needed for 3130, and 3700 and/or other software versions.

A summary flowchart of a typical sequencer run is displayed in Figure 6. Operations specific to MIRU-VNTR typing are described starting from sample preparation (see arrow in Figure 6). Information for spectral calibration for the current dye set used in MIRU-VNTR typing is given in 2.4.2.1.

Figure 6 Summary flowchart of a typical sequencer run

(From ABI PRISM® 3100 Genetic Analyzer User Guide)



2.4.2.1 Dye set and spectral calibration

If not done yet for other types of fragment analysis, initial spectral calibration is required for the appropriate detection of the current dye set used in MIRU-VNTR analysis. This combination of dyes (6FAM™, VIC®, NED™, for the amplicons, and ROX™ for the size standard) corresponds to dye set D on ABI sequencers.

This spectral calibration is performed using ABI DS-31 Matrix Standard Kit, containing 4 oligonucleotides respectively labeled with these 4 dyes, pre-pooled in a single tube. Spectral calibration procedure is described in the instructions for use and in the sequencer user guide.

2.4.2.2 Preparing the samples

Information on how samples are scheduled for injection based on plate configuration is provided in Figure 7.

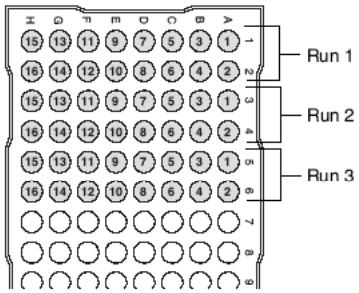
Figure 7 96-well plate mapping

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

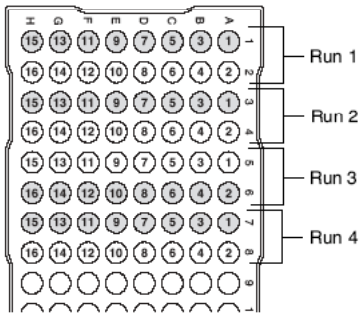
96-Well Plate Mapping

For a 96-well plate, injections are made from every well in two consecutive rows, starting with an odd row. A full 96-well plate requires six runs to inject all samples.

Below is an example of a 96-well plate. The gray circles represent samples, and the number in the well indicates capillary number. It takes three runs to inject 48 samples.

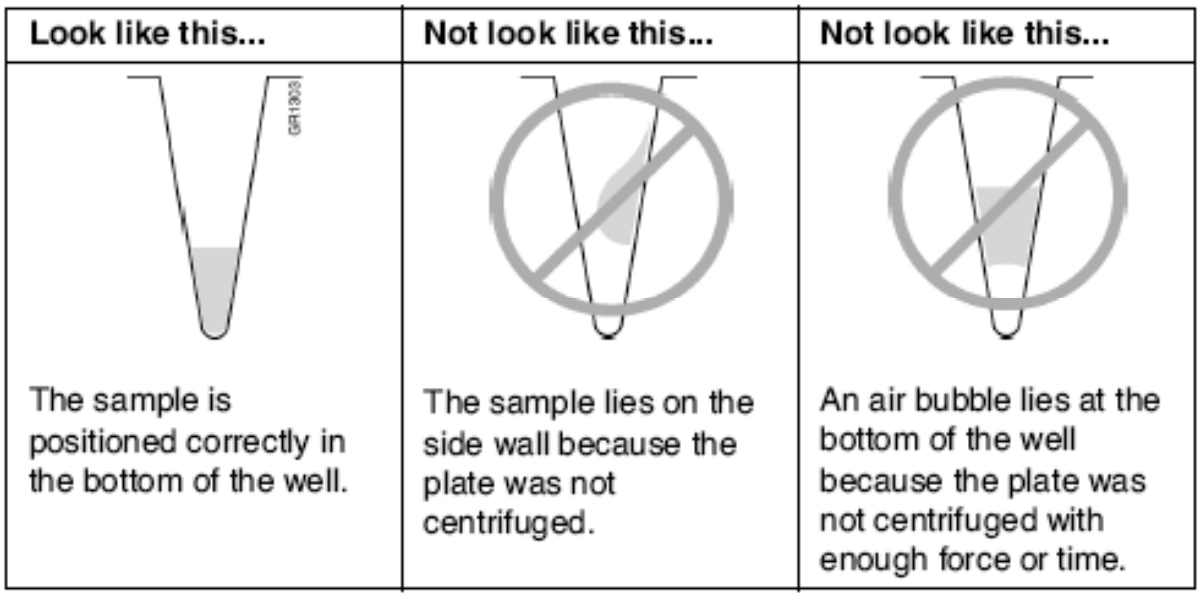


Below is an example of incorrect sample placement. To inject 24 samples requires four runs.



The sample are prepared as follows:

1. Dispense 1 µl of PCR product into the wells of a plate adapted for the ABI sequencer (e.g. Semi-skirted PCR plates, Sorenson, BioScience). Verify that each well contains solution.
2. Prepare extemporaneously a mix of 10 µl of formamide (Applied Biosystems) and 0.2 µl of MapMarker 1000 ROX™ size standard (BioVentures), per sample.
3. Dispense 10 µl of this mix into the wells of the plate. Verify that each well contains solution.
4. **Tap down or centrifuge the plate so that each sample is positioned at the bottom of its well.** Air injection should be avoided. Therefore your sample should:



(From ABI PRISM® 3100 Genetic Analyzer User Guide)

5. Keep the plate at 4°C until you are ready to prepare the plate assembly and place the assembly on the autosampler. (*note*: such DNA/formamide mixtures can be kept for a few days, but no longer than one week at 4°C).

2.4.2.3 Editing or creating a specific run module

Create a specific run module, called GS36_POP4_3000sec, as decribed in Figure 8, according to parameters in Table 5.

Figure 8 Creating a specific module

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

Step	Action
1	Click the Module Editor button on the toolbar to open the **Module Editor** dialog box. 
2	Select a run module to use as a template. → **Select GeneScan36_POP4DefaultModule**
3	Edit the parameter values that you want to change. **IMPORTANT** Only whole numbers are accepted. **IMPORTANT** Be sure that all values are red. Values in black are not saved.
4	Click **Save As** to create a new run module. Enter a unique descriptive name and click **OK**.  → **Enter GS36_POP4_3000sec** **Note** **Save** cannot be applied to default run modules.
5	When you are finished, click the Close button ( ) to exit the Module Editor.

Table 5 Run module parameters for ABI 3100, 3100 Avant

Run Temperature:	60 degrees
Cap Fill Volume:	184 steps
Current Tolerance:	100 micro Amps
Run Current:	100 micro Amps
Voltage Tolerance:	0.6 kVolts
Pre Run Voltage:	15 kVolts
Pre Run Time:	180 sec
Injection Voltage:	1 kVolts
Injection Time:	22 sec
Run Voltage:	15 kVolts
Number of Steps:	10 nk
Voltage Step Interval:	60 sec
Data Delay Time:	1 sec
Run Time:	3000 sec

2.4.2.4 Creating a plate record template

Plate records (or sample sheets) are data tables that store information about the plates and the samples they contain. Specifically for fragment analysis, a plate record contains the following information:

♦ Plate name, type, and owner

♦ Position of the sample on the plate (well number)

♦ Sample name

♦ Dye color of size standard

♦ Comments about the plate and about individual samples

♦ Dye set information

♦ Project name (**this entry is mandatory, even when Sequence Collector is not used**)

♦ Name of the run module (run modules specify information about how samples are run)

♦ Name of the analysis module (analysis modules specify how raw data is auto-analyzed at the end of the run). Auto-analysis is usually not performed.

Plate records can be generated using the Data Collection software, by following the steps in Figure 9. Specific notes for MIRU-VNTR analysis are indicated in bold. **Do not forget to fill in color info with sample name for each color.** This information is needed to create tables with final results in Genotyper templates.

Alternatively, plate records can be conveniently created using Excel and then subsequently imported as tab files (see next section).

Figure 9 Creating a plate record for fragment analysis

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

To enter plate record information:

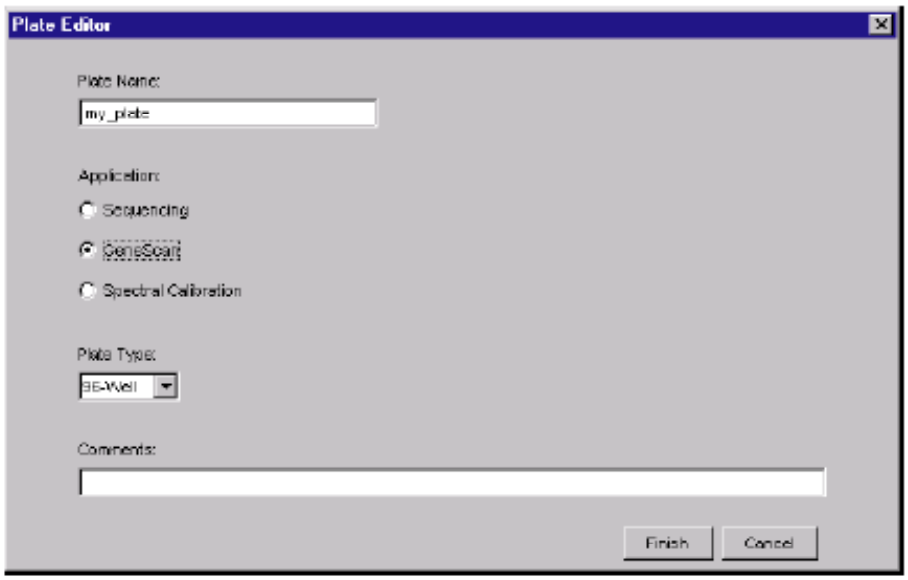
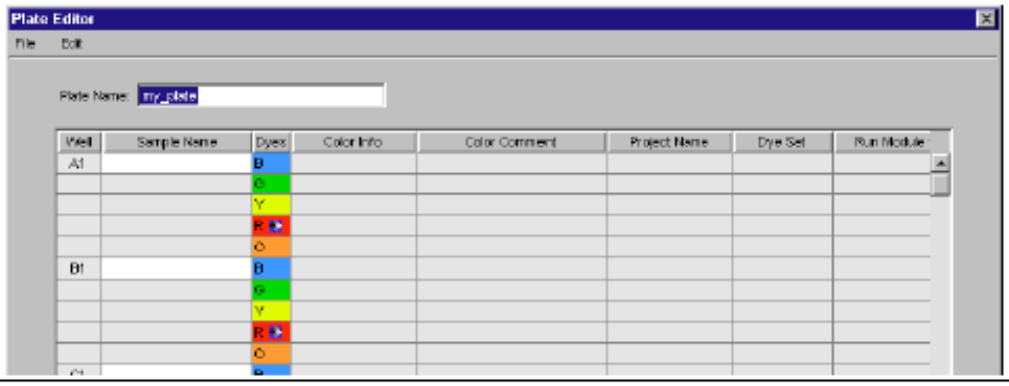
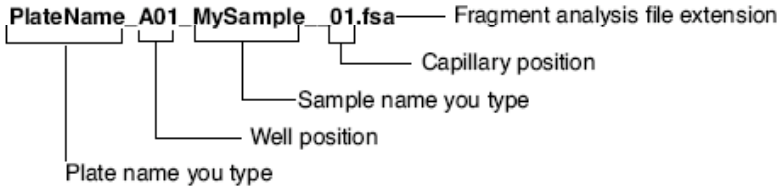
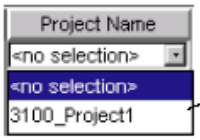
Step	Action
1	Click the **Plate View** tab in the data collection software window to go to the Plate View page. 
2	In the Plate View page, click **New**. The Plate Editor dialog box opens.
3	Enter your plate name and select the application and plate type. Comments are optional.  **Note** When naming the plate, you can use letters, numbers, and the following punctuation only: -_(){}#.+. DO NOT USE SPACES.
4	When done, click **Finish**. The plate editor spreadsheet opens. 

Figure 9 (continued)

Entering Sample Information

To enter sample information and save the plate record:

Step	Action
1	In the **Plate Editor** spreadsheet, type the names of all the samples in the Sample Name column. ♦ Sample names are limited to a maximum of 32 characters. ♦ When naming the samples, you can use letters, numbers, and the following punctuation only: -_(){}#.+. DO NOT USE SPACES. ♦ In the default naming convention, the sample name you type is incorporated into the sample file name. For example:  **Note** The sample file naming convention used can be changed in the **Preferences** dialog box. See page 2-12 for details.
2	Change the size standard dye color, if necessary. The default is red. Use red for 4-dye applications and orange for 5-dye applications. a. Click on the dye color you want to use. b. Select all the samples. c. Select **Edit > Fill down**.
3	Type in Color info and Color Comment, if needed. **Note** Color Info and Color Comment information is the same as Sample Info and Sample Comment in the ABI PRISM® Genotyper® software. Refer to the *ABI PRISM® Genotyper® Software User's Manual* for more information.
4	Enter a Project name. **Note** A project name is required for every sample, even if a Sequence Collector database is not used. a. Click in the **Project Name** cell for Well A1. b. Select a project name in the drop-down list.  **Note** You must select a project. c. To assign the same project name to each sample in the plate record: – Click the column header to select the whole column. – Press **Ctrl+D**. **Note** Press Ctrl+D to fill down whenever a field is the same for all samples in the plate record.

Important !

Not applicable
Size standard color is red (ROX)

Important !

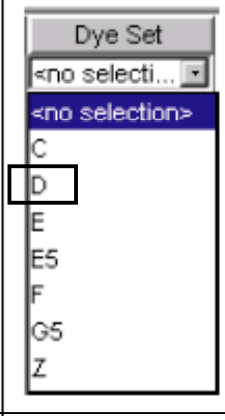
Figure 9 (continued)

To enter sample information and save the plate record: *(continued)*

Important ! Use Ctr+D To fill down

Step	Action
5	For each sample, select the appropriate **Dye Set** in the drop-down list.

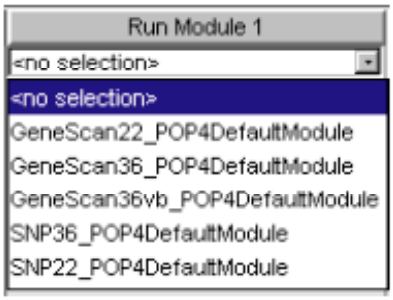
Select D



To enter sample information and save the plate record: *(continued)*

Select GS36_POP4_ 3000sec

Step	Action
6	For each sample, select the appropriate **Run Module** in the drop-down list.



To enter sample information and save the plate record: *(continued)*

Choose no selection

Step	Action
7	For each sample, select the appropriate **Analysis Module** from the drop-down list. **IMPORTANT** The Enable AutoAnalysis preference must be selected if analysis is to take place automatically after the run (see page 2-12).

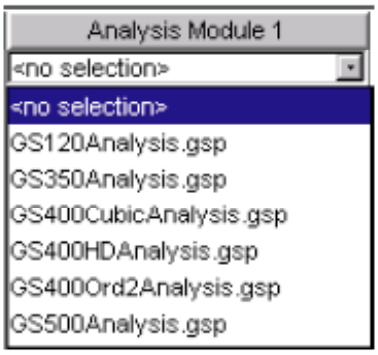


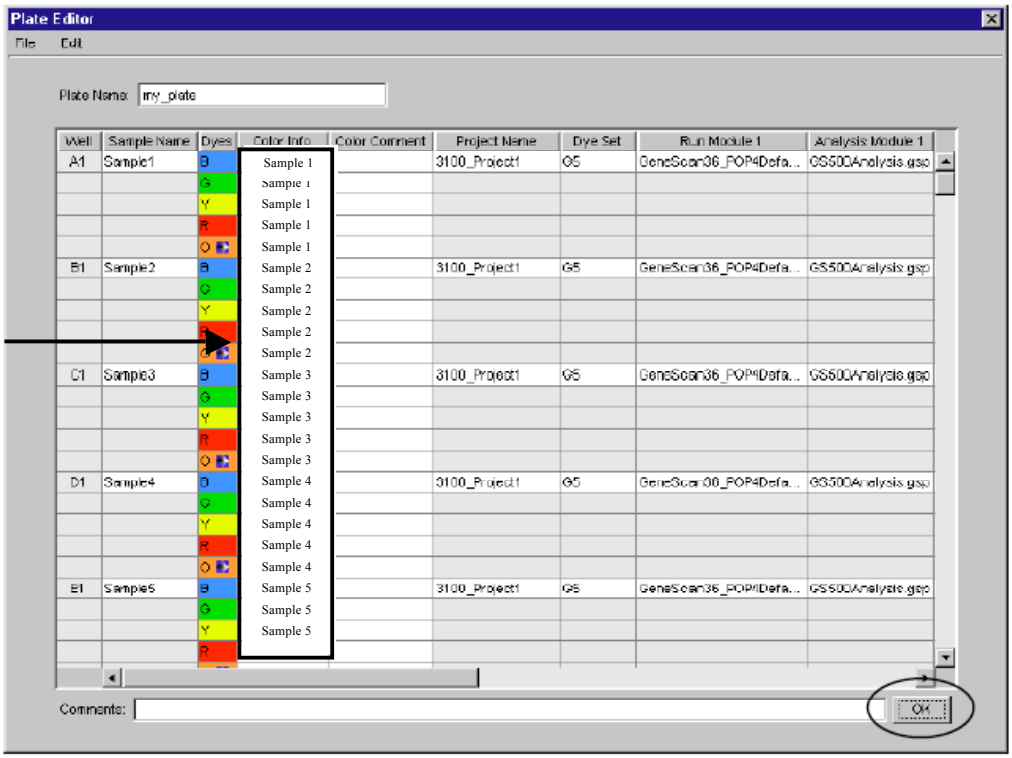
Figure 9 (continued)

8	If you want to run the same sample again, select a second run module and a second analysis module. You can run a sample in a linked plate up to five times. Run Module 2 | Analysis Module 2 **Note** Samples will be automatically grouped so that all samples with the same run module are run sequentially.

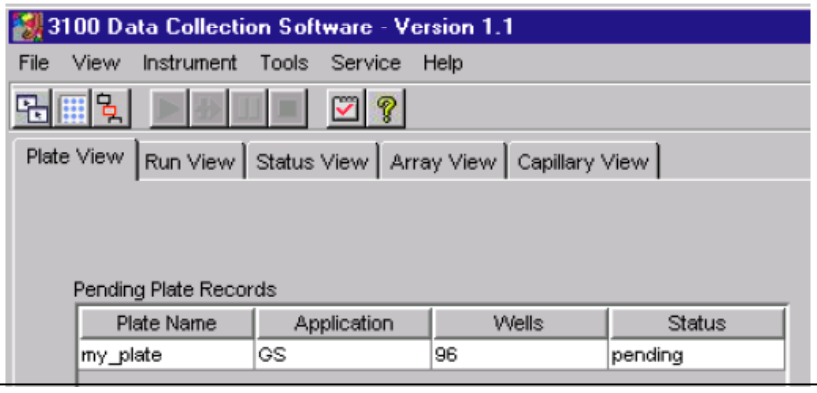
9	a. Verify that the plate record is correct and complete (contains the following required information: sample name, dye set, size standard color, project, and run module). b. Click **OK**. An example of a completed plate record is shown below.

Fill color info with sample name for each color

But not applicable for Data Collection version 3.x Must be entered in Genotyper See 2.6.2.4a



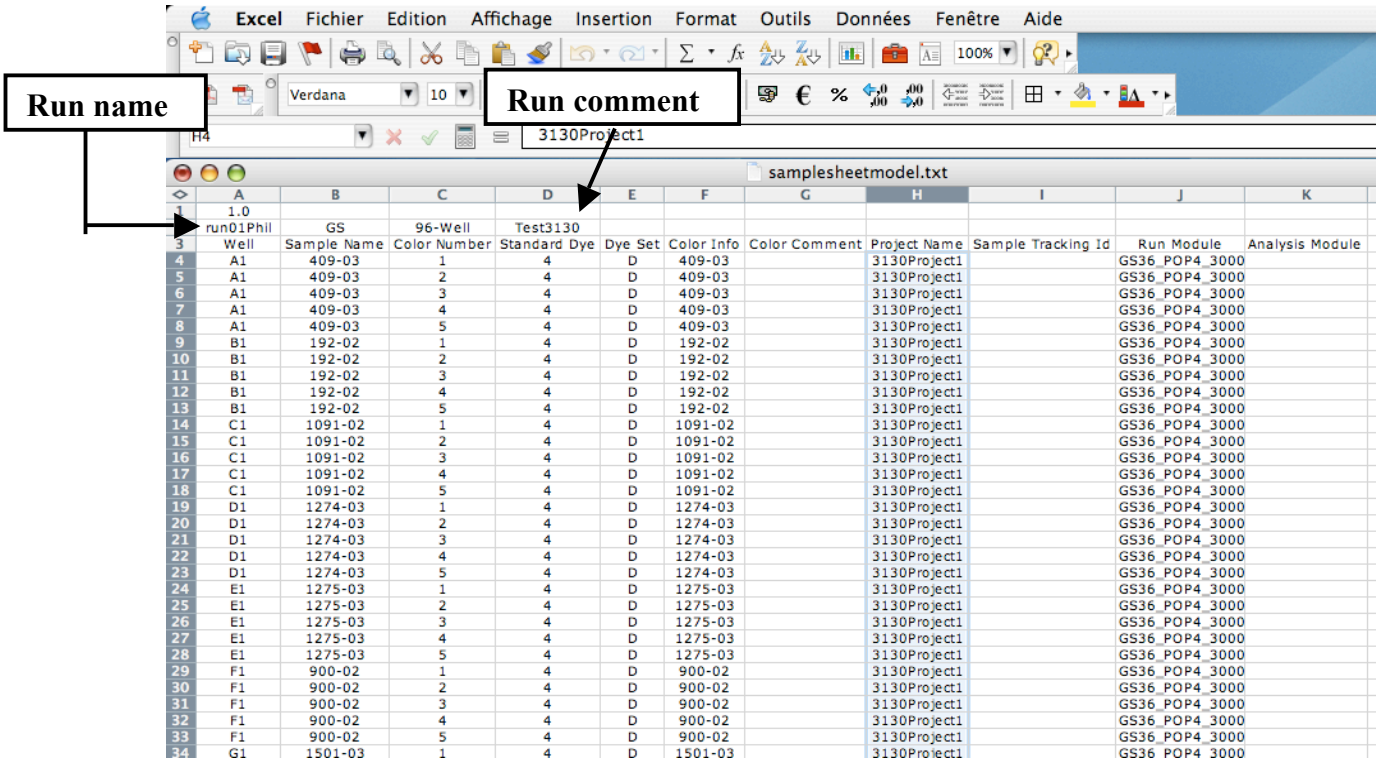
Note It may take a few minutes for the new plate record to be saved to the database and added to the Pending Plate Records table as shown below.



2.4.2.5 Importing plate record tab files generated with Excel

1. Create an Excel table following the model shown in Figure 10. Type sample names in cells B4 to B483. **Add extensions to sample names, if different multiplex or loci are used in a same plate (*e.g.* -M1 for Mix 1, -04 for locus 4). Do not use spaces, or punctuation incompatible with PC file name formats.**
2. Select the sample name cells, Copy (Ctrl+C) and Paste (Ctrl+V) these cells into cells F4 to F483. This will be useful to retrieve samples in Genotyper.
3. For positions containing no samples, type a 0 into all corresponding sample name cells for the five colors.

Figure 10 Plate record model using Excel



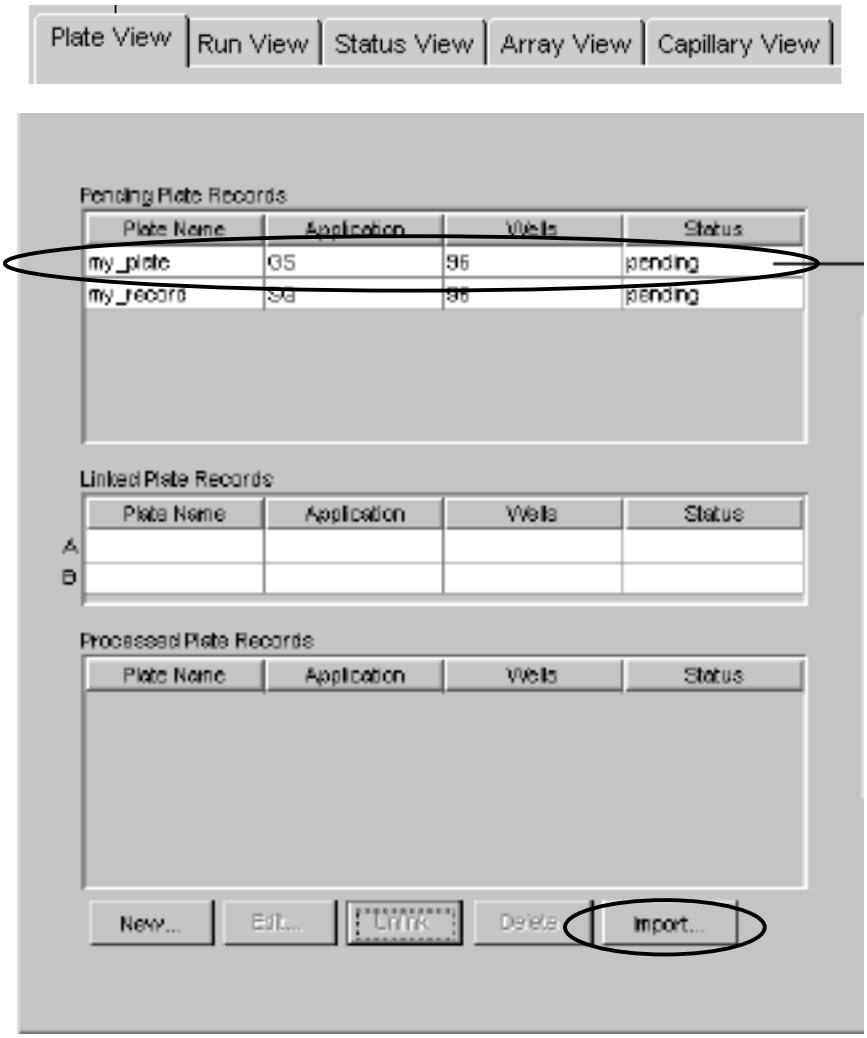
	A	B	C	D	E	F	G	H	I	J	K
1	1.0										
2	run01Phil	GS	96-Well	Test3130							
3	Well	Sample Name	Color Number	Standard Dye	Dye Set	Color Info	Color Comment	Project Name	Sample Tracking Id	Run Module	Analysis Module
4	A1	409-03	1	4	D	409-03		3130Project1		GS36_POP4_3000	
5	A1	409-03	2	4	D	409-03		3130Project1		GS36_POP4_3000	
6	A1	409-03	3	4	D	409-03		3130Project1		GS36_POP4_3000	
7	A1	409-03	4	4	D	409-03		3130Project1		GS36_POP4_3000	
8	A1	409-03	5	4	D	409-03		3130Project1		GS36_POP4_3000	
9	B1	192-02	1	4	D	192-02		3130Project1		GS36_POP4_3000	
10	B1	192-02	2	4	D	192-02		3130Project1		GS36_POP4_3000	
11	B1	192-02	3	4	D	192-02		3130Project1		GS36_POP4_3000	
12	B1	192-02	4	4	D	192-02		3130Project1		GS36_POP4_3000	
13	B1	192-02	5	4	D	192-02		3130Project1		GS36_POP4_3000	
14	C1	1091-02	1	4	D	1091-02		3130Project1		GS36_POP4_3000	
15	C1	1091-02	2	4	D	1091-02		3130Project1		GS36_POP4_3000	
16	C1	1091-02	3	4	D	1091-02		3130Project1		GS36_POP4_3000	
17	C1	1091-02	4	4	D	1091-02		3130Project1		GS36_POP4_3000	
18	C1	1091-02	5	4	D	1091-02		3130Project1		GS36_POP4_3000	
19	D1	1274-03	1	4	D	1274-03		3130Project1		GS36_POP4_3000	
20	D1	1274-03	2	4	D	1274-03		3130Project1		GS36_POP4_3000	
21	D1	1274-03	3	4	D	1274-03		3130Project1		GS36_POP4_3000	
22	D1	1274-03	4	4	D	1274-03		3130Project1		GS36_POP4_3000	
23	D1	1274-03	5	4	D	1274-03		3130Project1		GS36_POP4_3000	
24	E1	1275-03	1	4	D	1275-03		3130Project1		GS36_POP4_3000	
25	E1	1275-03	2	4	D	1275-03		3130Project1		GS36_POP4_3000	
26	E1	1275-03	3	4	D	1275-03		3130Project1		GS36_POP4_3000	
27	E1	1275-03	4	4	D	1275-03		3130Project1		GS36_POP4_3000	
28	E1	1275-03	5	4	D	1275-03		3130Project1		GS36_POP4_3000	
29	F1	900-02	1	4	D	900-02		3130Project1		GS36_POP4_3000	
30	F1	900-02	2	4	D	900-02		3130Project1		GS36_POP4_3000	
31	F1	900-02	3	4	D	900-02		3130Project1		GS36_POP4_3000	
32	F1	900-02	4	4	D	900-02		3130Project1		GS36_POP4_3000	
33	F1	900-02	5	4	D	900-02		3130Project1		GS36_POP4_3000	
34	G1	1501-03	1	4	D	1501-03		3130Project1		GS36_POP4_3000	

4. Save it as a tab-delineated file (.txt).
5. In Data Collection Software, click **Plate View** tab, then click **Import** (see Figure 11).
6. Select **all types** (*.*) from **files of type**.
7. Select your plate and click **OK**.

8. Double click on the pending plate and verify that the record is correct and complete (see example in step 9 of Figure 9).
9. If positions contain no sample (now indicated by a 0 after plate record importation), click on the **Run Module** cell of the first empty well and select **No Run Module**. Select the first and last empty positions by pressing **Shift** key**,** and then press **Ctrl+D** to fill down.

Figure 11 Plate record importation and checking

(From ABI PRISM® 3100 Genetic Analyzer User Guide)



2.4.2.6 Preparing and placing a plate assembly

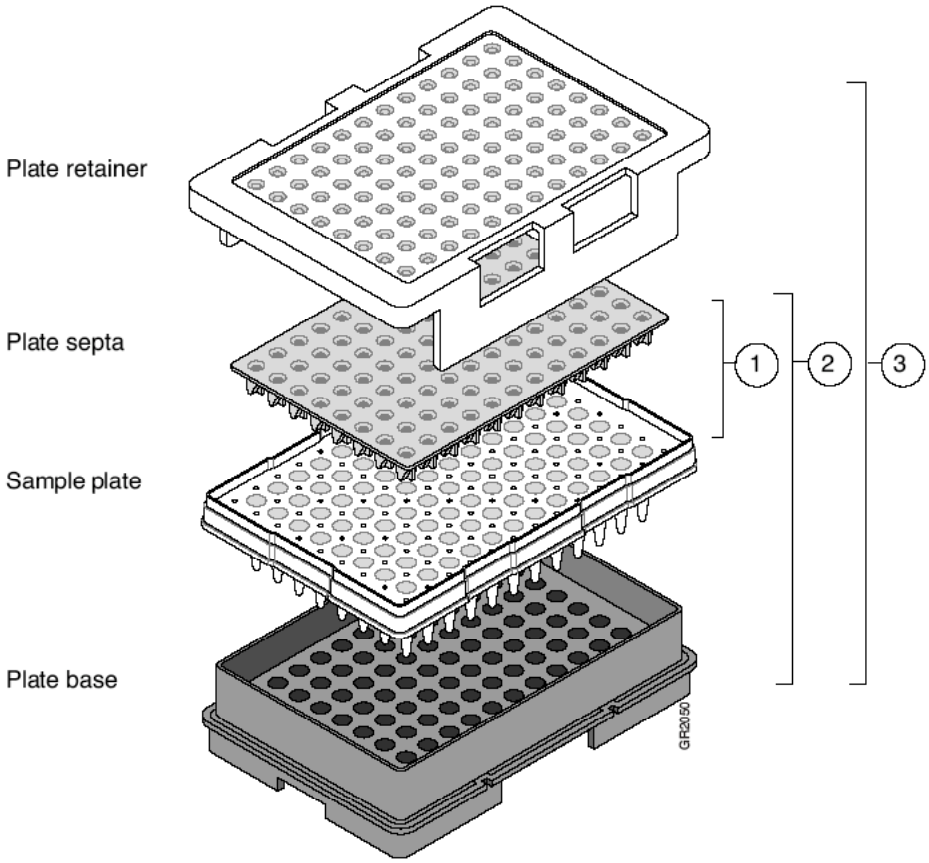
Verify the absence of any air bubble in plate wells. If needed, centrifuge the plate to remove them. Prepare a plate assembly and place it on the Autosampler by carefully following the steps described in Figure 12.

Figure 12 Plate assembly and placing onto the Autosampler

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

Plate Assembly Components

The plate assembly components are assembled as follows:

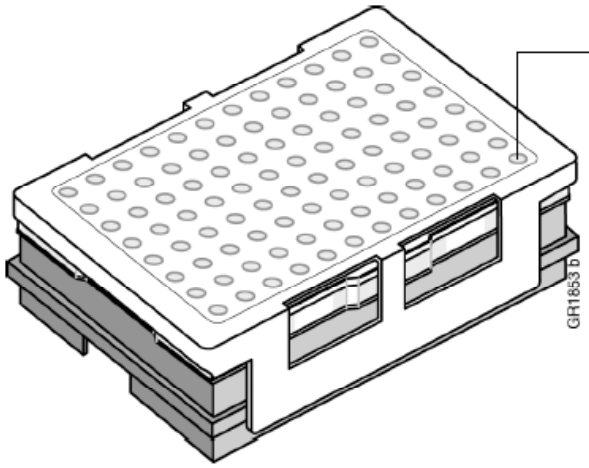


Preparing a Plate Assembly

To prepare a plate assembly:

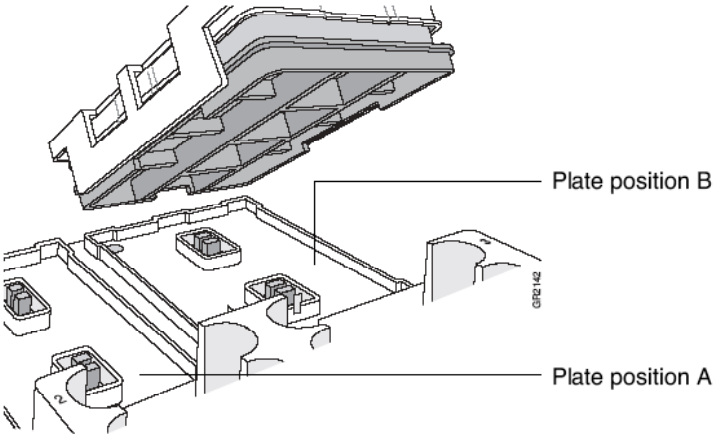
Step	Action
1	Secure a clean and dry plate septa on the sample plate. **IMPORTANT** Never use warped plates. **IMPORTANT** Ensure that the plate septa lies flat on the plate.
2	Place the sample plate into the plate base.

To prepare a plate assembly: *(continued)*

Step	Action
3	Snap the plate retainer onto the plate and plate base.
4	Ensure that the plate retainer holes are aligned with the holes in the septa strip. **IMPORTANT** Damage to the array tips will occur if the plate retainer and septa strip holes do not align correctly. 

Placing the Plate onto the Autosampler

To place the plate onto the autosampler:

Step	Action
1	Place the plate assembly on the autosampler in position A or B. **Note** There is only one orientation for the plate, with the notched end of the plate base away from you.  **IMPORTANT** Ensure the plate assembly fits flat in the autosampler. Failure to do so may allow the capillary tips to lift the plate assembly off of the autosampler.
2	When the plate is correctly positioned, the plate position indicator on the Plate View page changes from gray to yellow. Verify that this has happened.
3	Close the instrument doors. **Note** Closing the doors returns the autosampler to the home position, placing the tips of the capillaries in buffer.

2.4.2.7 Linking a plate

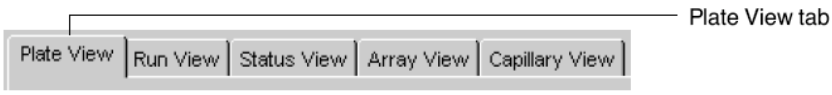
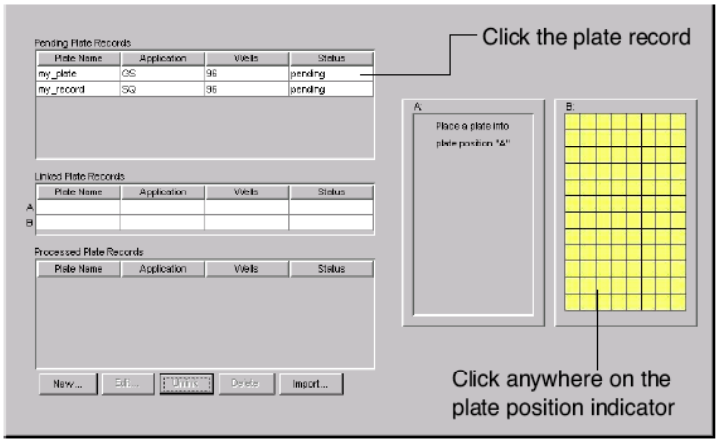
A link must be created between the plate on the auto-sampler and the corresponding plate record, before a run can be performed. This procedure and the verification of the run schedule are described in Figure 13.

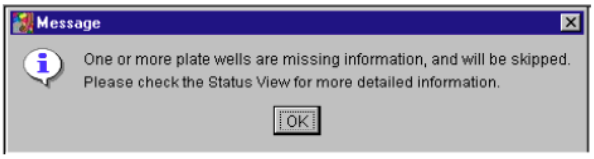
Figure 13 Linking a plate and viewing the run schedule

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

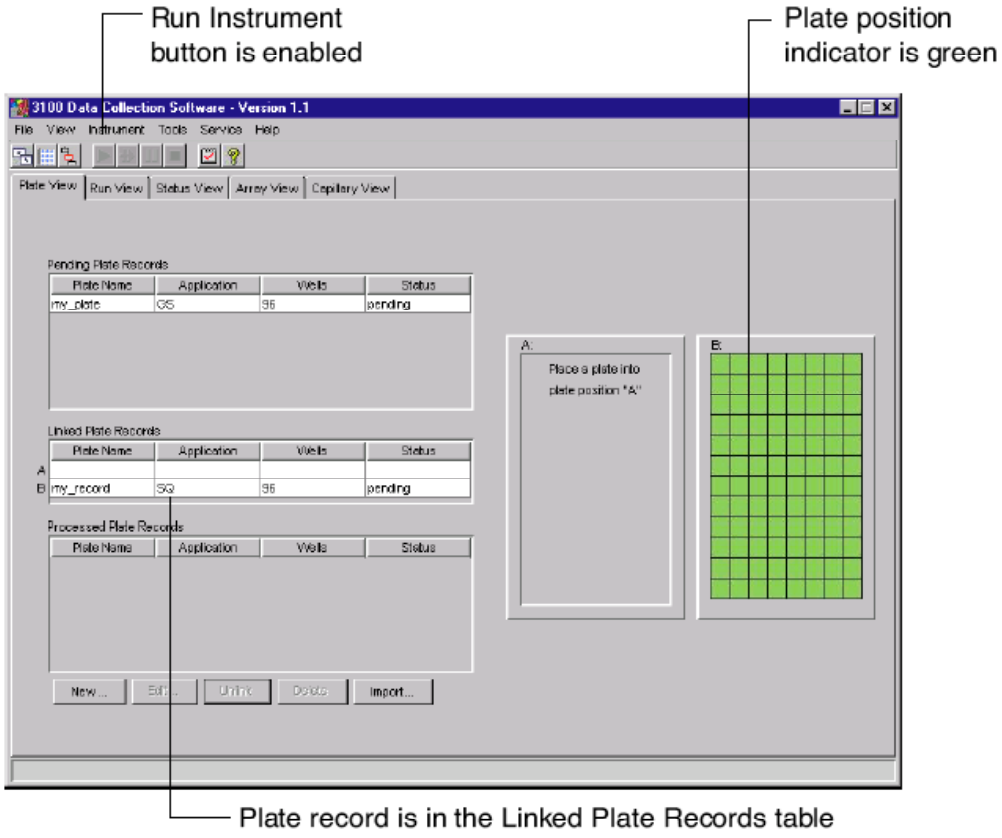
Linking a Plate to a Plate Record

To link a plate to a plate record:

Step	Action
1	Click the **Plate View** tab. 
2	In the Plate View page: a. In the Pending Plate Records table, click the plate record for the plate you are linking. b. Click the plate position indicator that corresponds to the plate you are linking. 
3	Add the required information to the plate record: a. Unlink the plate record, if necessary. (The plate record returns to the Pending Plate Records table.) b. Double-click the plate record name to open it. c. Correct the plate record and click **OK**. d. Link the plate record to the plate again.

Does the following message display?	Then proceed to step…
	
Yes	3
No	4

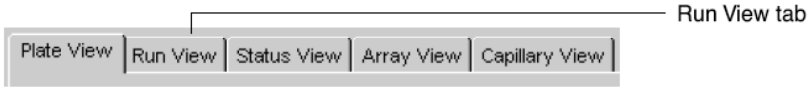
To link a plate to a plate record: *(continued)*

Step	Action
4	Verify that the plate has been linked. Once the plate has been linked, the: ♦ **Run Instrument** button in the toolbar is enabled, meaning that the instrument is ready to run. ♦ Plate position indicator for the linked plate becomes green. ♦ Plate record moves from the Pending Plate Records table to the Linked Plate Records table. 

Viewing the Run Schedule

After a plate is linked, use the Run View page to verify that runs are scheduled correctly.

To view the scheduled runs:

Step	Action
1	Click the **Run View** tab. 
2	Select a row for any run. The corresponding wells to be injected for that run becomes highlighted in the plate diagram on the left. For more information about the Run View page, see page 2-62.

Note Although you can delete individual runs, you cannot alter the order in which the runs are scheduled.

2.4.2.8 Launching and controlling the run

The procedure for starting and controlling a run is displayed in Figure 14.

Figure 14 Launching and controlling a run

(From ABI PRISM® 3100 Genetic Analyzer User Guide)

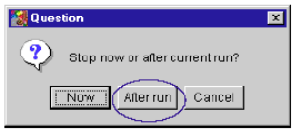
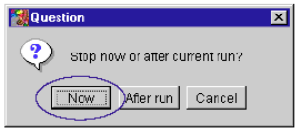
Starting the Run

To start a run:

Step	Action
1	If you want to review the run schedule before beginning the run, click the **Run View** tab.
2	Click the green **Run Instrument** button in the toolbar. 
3	The software automatically checks the available space in the database and drive D. **If the database or drive D are…** / **Then…** full: a. proceed to "Automatic Checking of Available Space on Drives D and E" on page 2-55. b. Make more space. c. Click **Run Instrument**. not full: the run starts.

Controlling the Run Using the Toolbar

You can also use the toolbar at the top of the data collection software window to control the run.

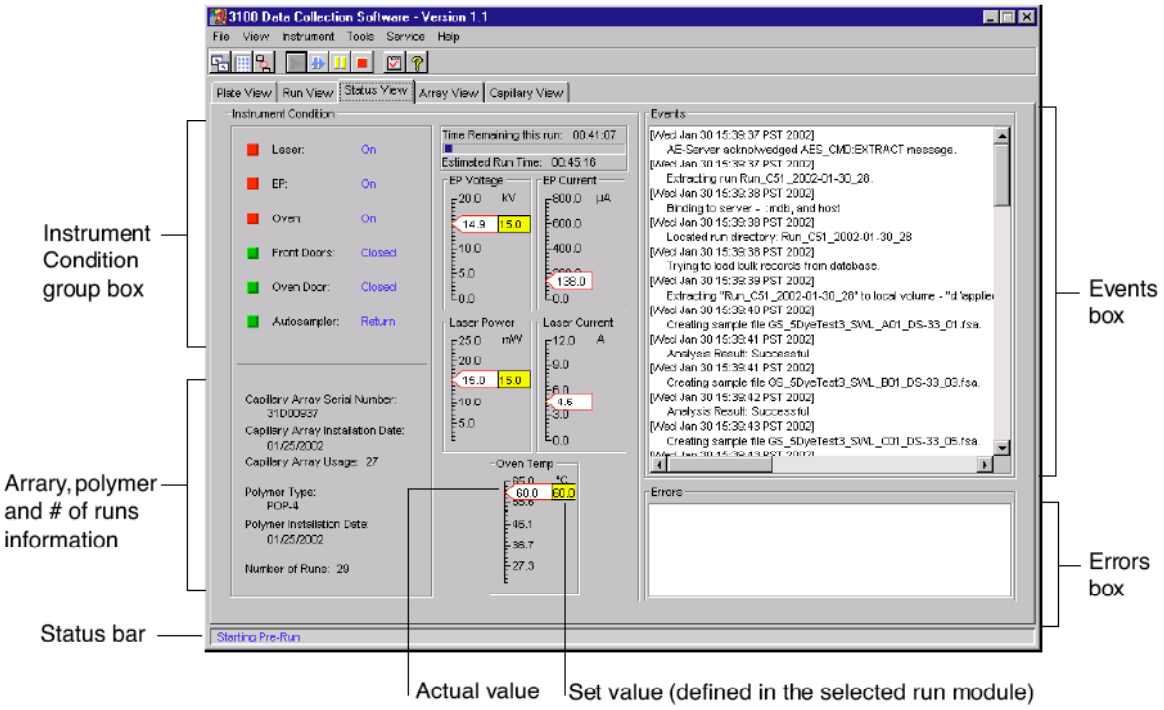
To...	Click...	Comment
Start the run	**Run Instrument** 	♦ This begins all scheduled runs. ♦ The run starts only when set temperature is reached.
Pause the run	**Pause** 	Pausing the instrument for too long, especially after sample injection, will affect data quality. The best time to pause is before sample injection
♦ Complete the current run, and ♦ Stop the other scheduled runs	a. **Stop**  b. **After run** in the **Question** dialog box 	
♦ Stop the current run, and ♦ Stop the other scheduled runs	a. **Stop**  b. **Now** in the **Question** dialog box 	When you click **Now**, the run files extract automatically. The files will be automatically analyzed if the AutoAnalysis preference is enabled. To recover data from a stopped run, see "Recovering Data If Autoextraction Fails" on page 2-70.
♦ Stop the current run, and ♦ Continue the other scheduled runs	**Skip to Next Run** 	To recover data from a stopped run, see "Recovering Data If Autoextraction Fails" on page 2-70.

2.4.2.9 Monitoring the run status

The run status can be monitored by clicking the Status View tab. An example of Status view page is provided in Figure 15.

Figure 15 Example of Status View page

(From ABI PRISM® 3100 Genetic Analyzer User Guide)



2.5 PCR product sizing using Genescan

After electrophoresis, the sizes of the PCR fragments are estimated using the GeneScan Analysis software (Applied Biosystem). **Size data editing and automated conversion of these data into MIRU-VNTR alleles is only done subsequently**, using customized templates of the Genotyper software (see next section). The procedure described below is based on GeneScan version 3.7.1.

IMPORTANT: this version must be initially updated with GeneScan version 3.7.1 Updater and switched to Large Fragment Analysis using Large Fragment Enabler, as indicated in the ReadMe instructions included the Updater CD.

2.5.1 Principle

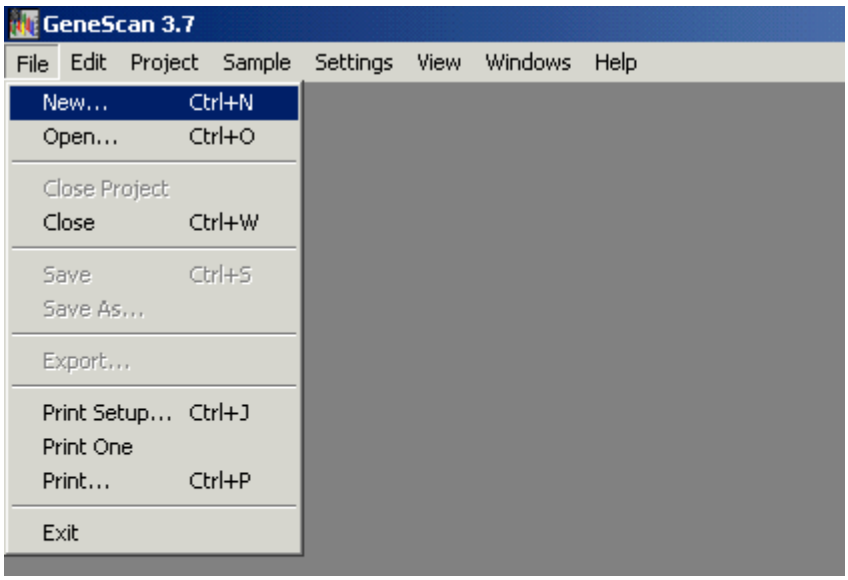
The GeneScan Analysis software is used to calculate the apparent sizes of the PCR products, by correlating their migration data (in terms of laser scan numbers needed for detection in the CCD window) to those of size standard bands.

2.5.2 Procedure

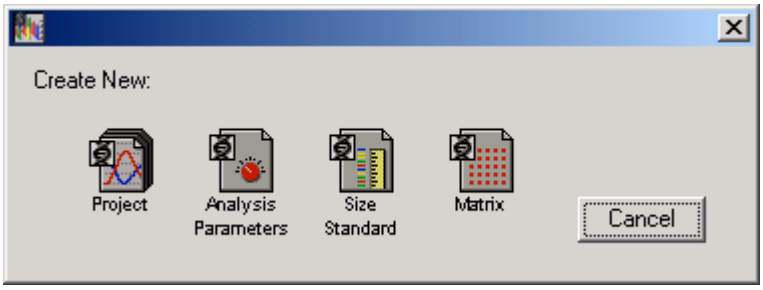
2.5.2.1 Creating a GeneScan project

1. Open Genescan 

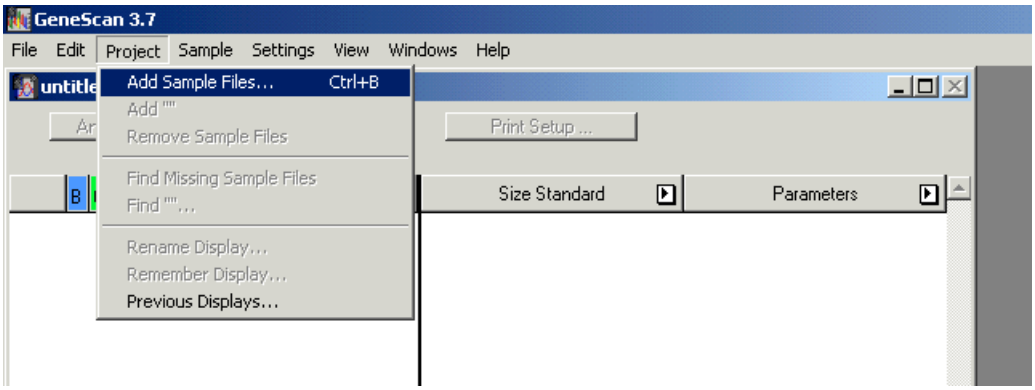
2. From the main menu, click **File**, then **New**.



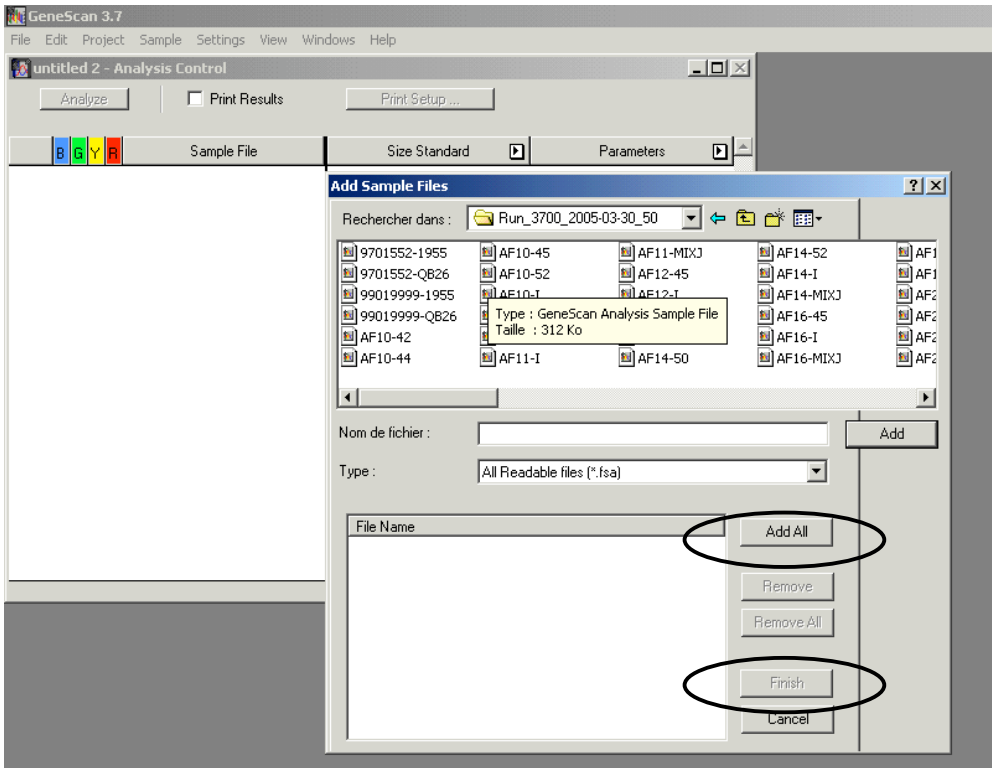
3. Click Project on the window as displayed below, to open the Analysis Control Window.



4. Click **Project** from the main menu, and then **Add Sample Files**.



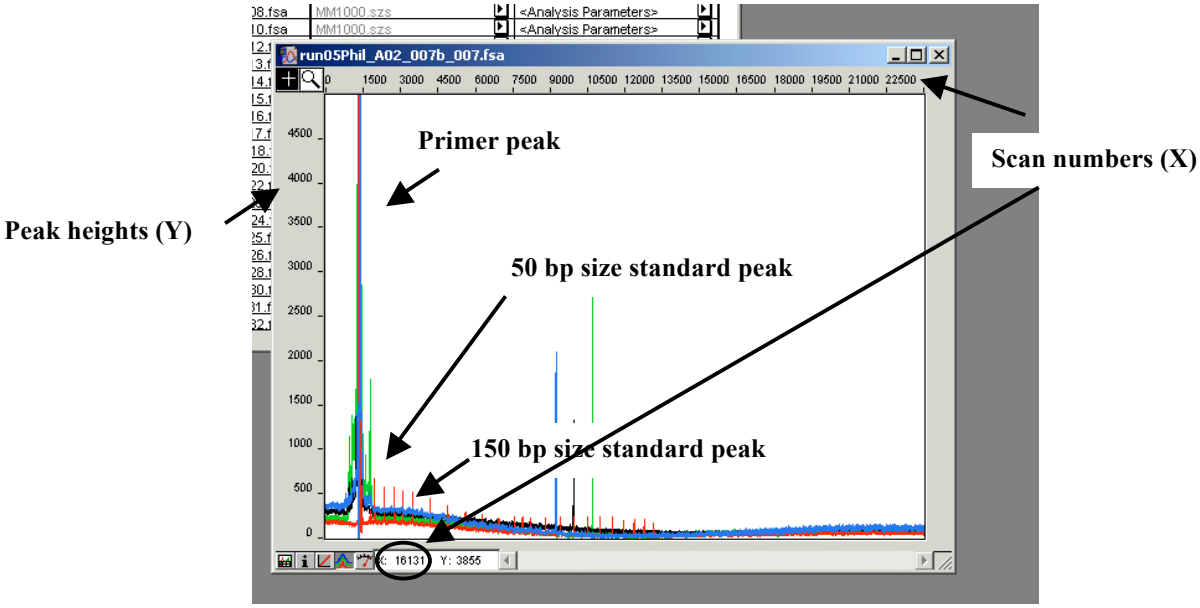
5. Select the folder containing the sample files, and then click **Add All**, then **Finish**.



2.5.2.2 Setting analysis parameters and size standard reference

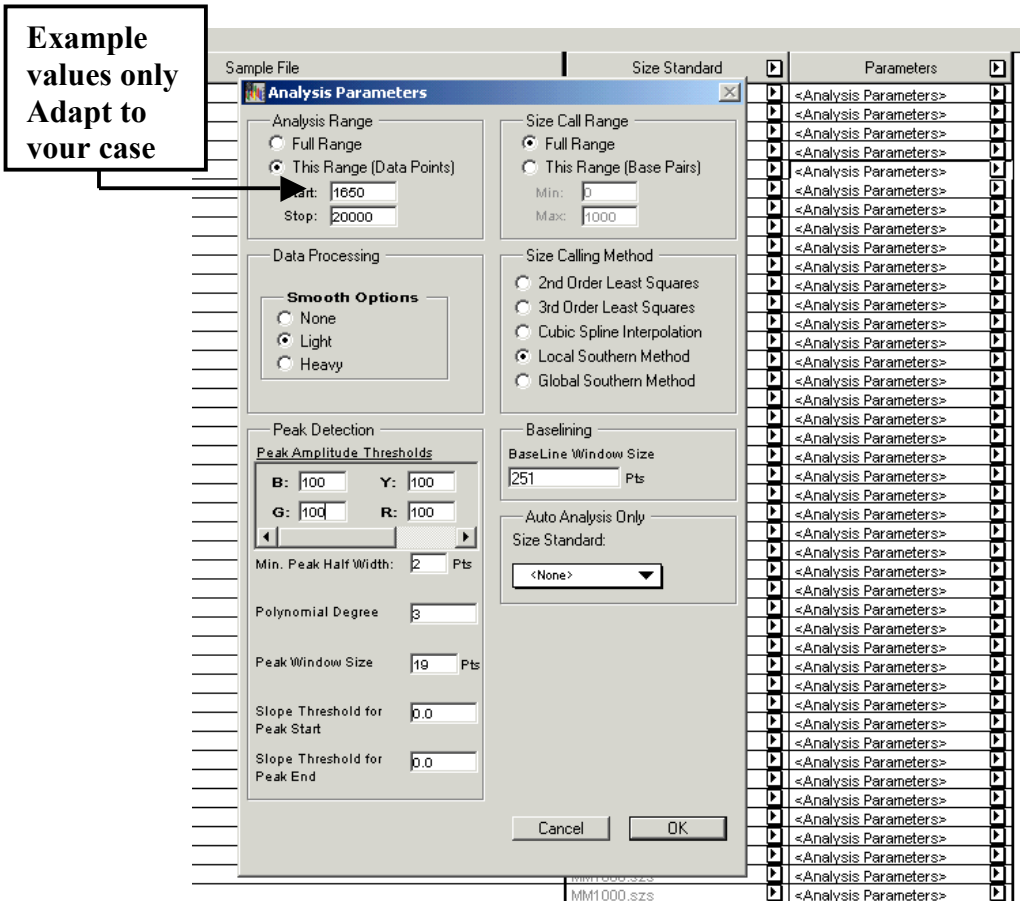
6. Double click on a Sample File and then click  at the bottom of the pop-up window, to visualize an electropherogram which should look like the example below.

7. Using the zoom and the pointer  , visualize and notice the scan numbers (X) corresponding to the zone immediately after the primer peak, and to the end of the run, respectively. This range should include all size standard peaks (in red) starting from 50 bp (see below for the size standard definition).

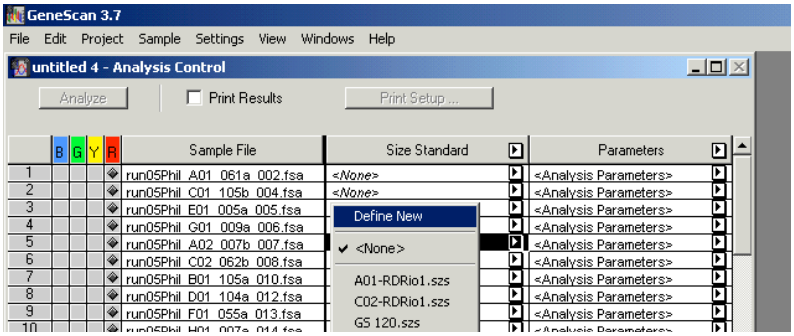


8. Close the electropherogram window. Double-click any Analysis Parameters cell, to open the Analysis Parameters window.

9. Define the start and stop of the analysis range by introducing the scan numbers corresponding to the zone comprised **between the primer peak and the 50 bp peak**, and **just before to the end of the run**, respectively. Define the other parameters as shown below.
10. Save as Analysis Parameters. This file will then be used as default.
11. *Note:* Peak amplitude thresholds, defining the minimum peak height detection, can be subsequently increased if higher backgrounds are noticed for some samples. In this case, save the modified analysis parameters under a new name.

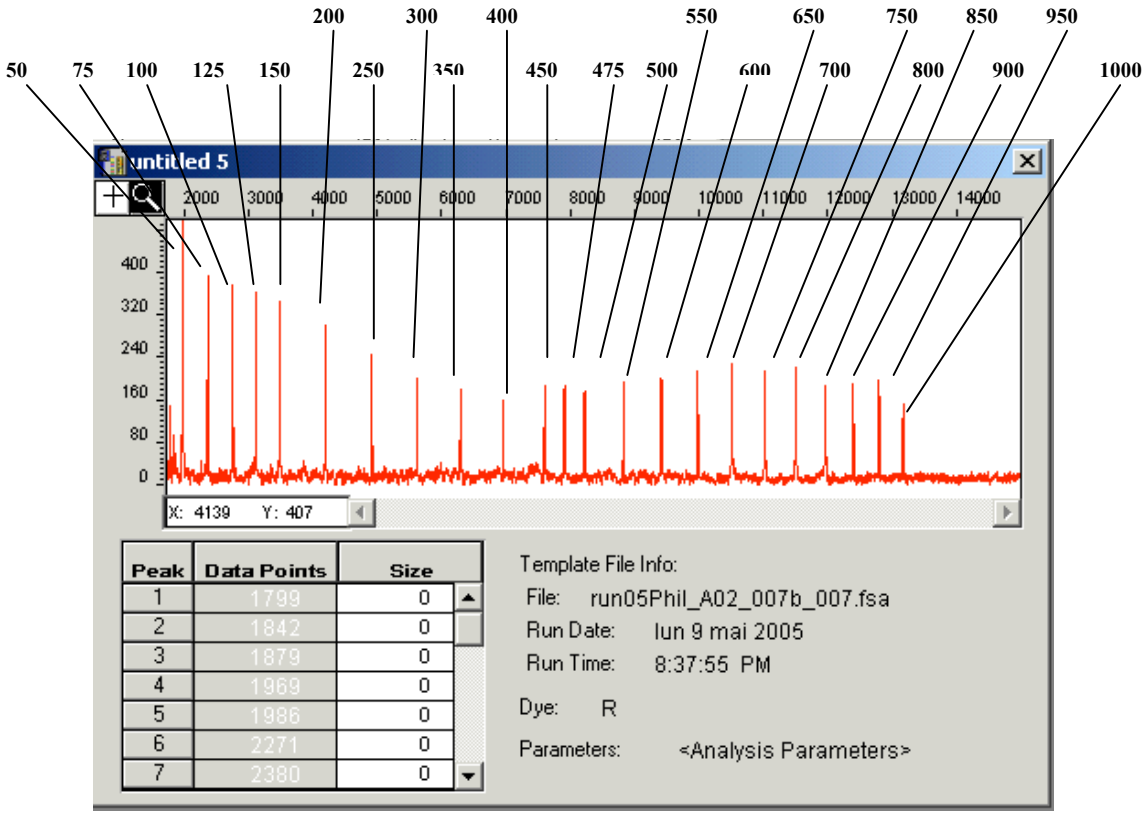


12. Define the size standard reference, by clicking <*None*> as shown below, better from the capillary containing the PCR negative control. This reference can be used for subsequent analysis using the same conditions and run module. **If conditions such as voltage or the polymer are changed, this size standard will need to be redefined.**

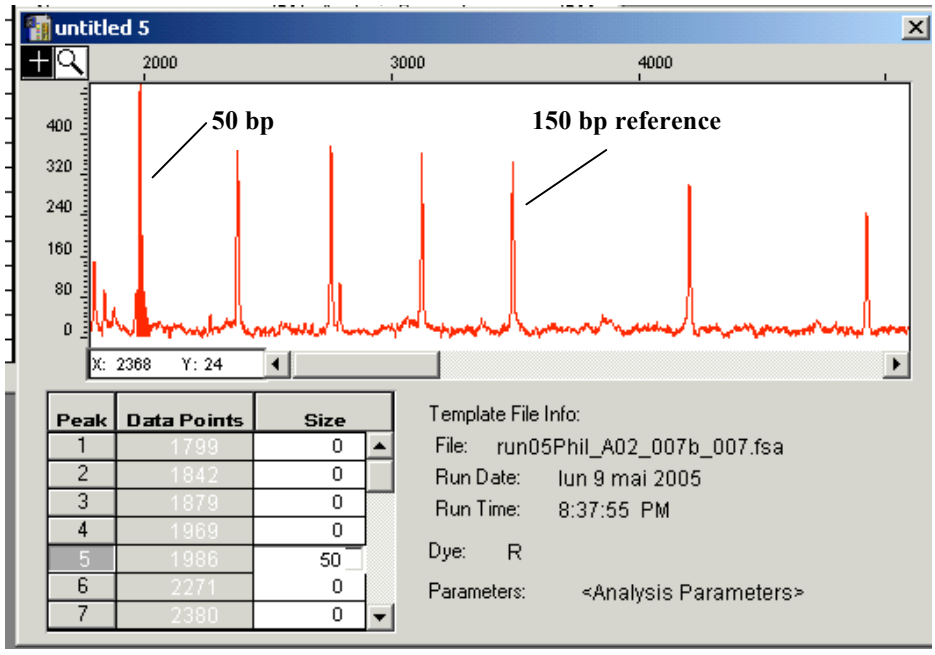


13. The size standard electropherogram corresponding to the selected capillary appears, as indicated below. The sizes of the size standard peaks of the MapMarker 1000 (except the 25 bp peak) are displayed.

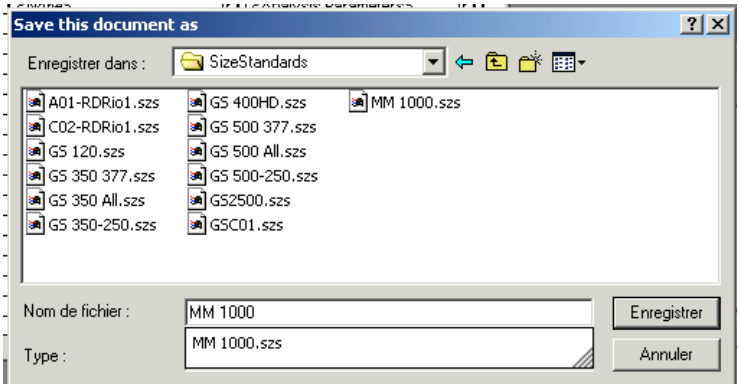
14. Using the zoom and the pointer  , zoom in to better visualize and select peaks.



15. Enter the sizes corresponding to the data points of each size standard peak, **using the 150 bp peak as a reference**. Carefully verify the correspondence with the general size standard pattern shown above, especially the presence of the typical 450-475-500 bp pattern.

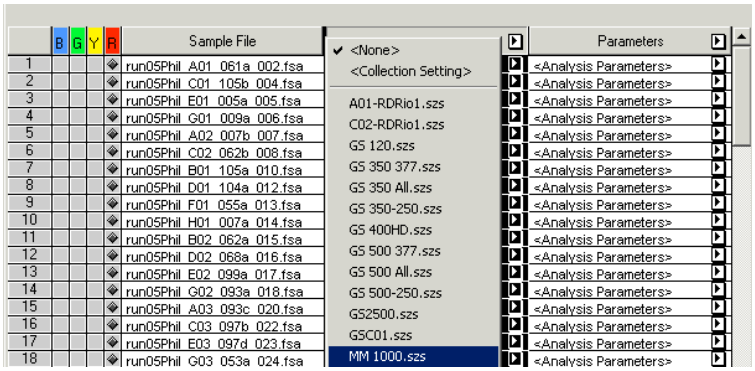


16. Save as MM 1000 in the **Applied Bio>Shared>Analysis>Size Caller>Size Standard Folder**.



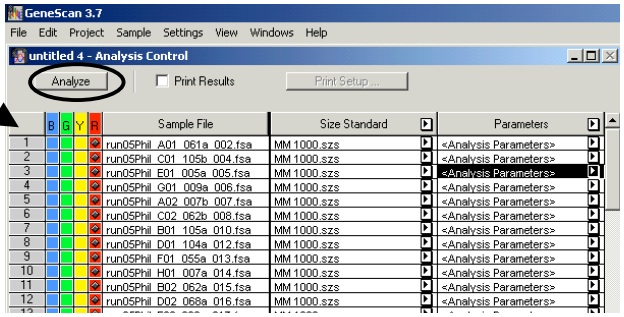
2.5.2.3 Analyzing sample files

17. Apply this size standard to all samples using the scroll-down menu as shown below.

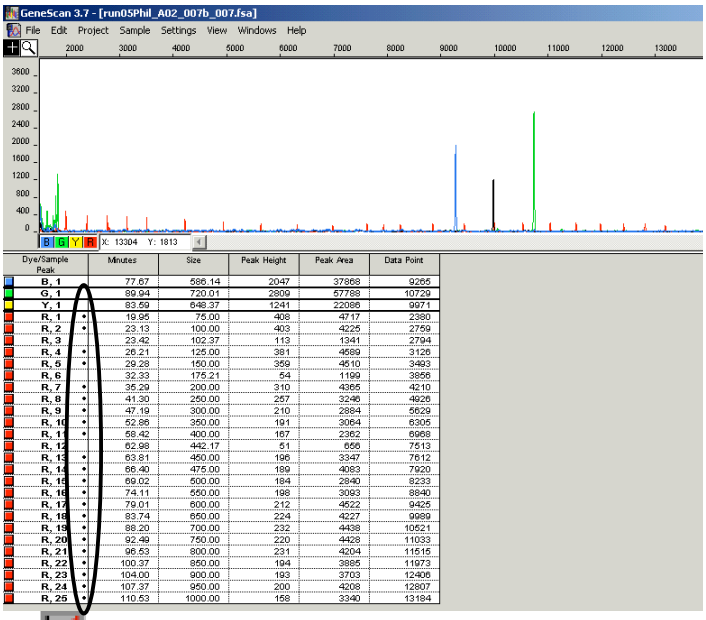


18. Click on the zone indicated below to select all sample files for analysis of all fluorochromes, and then click **Analyze**.

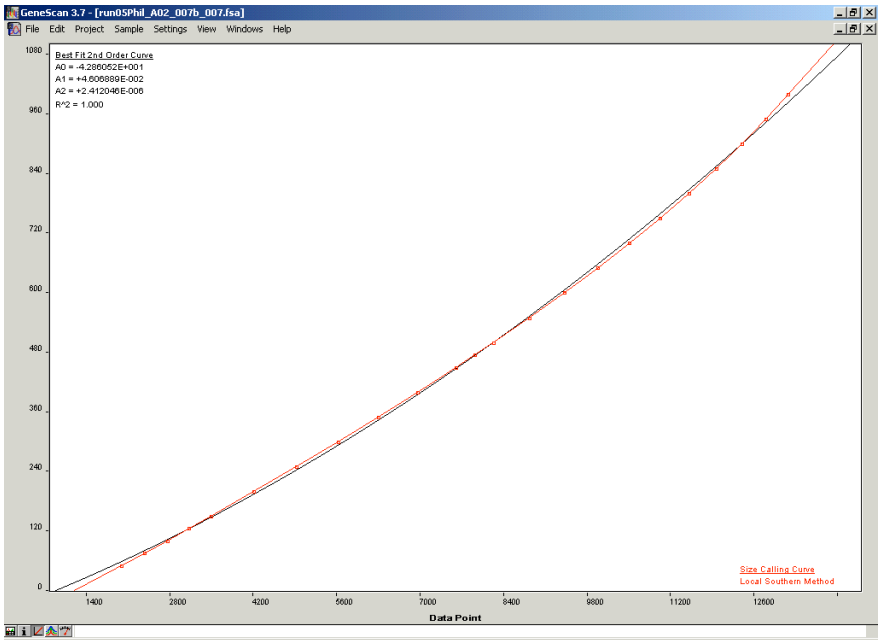
Click to select color analysis



19. Double click on a few sample files to visualize results, as shown on the example below. Dots in Dye/Sample Peak cells indicate peaks recognized as size standard peaks.



20. Click on the  button at the bottom of the window to control the size calling curve, which should look as in the example shown below.



21. Close the window, then select **File** from the main menu, and **Save as** My project.

2.6 Editing sizing data and assigning alleles using Genotyper

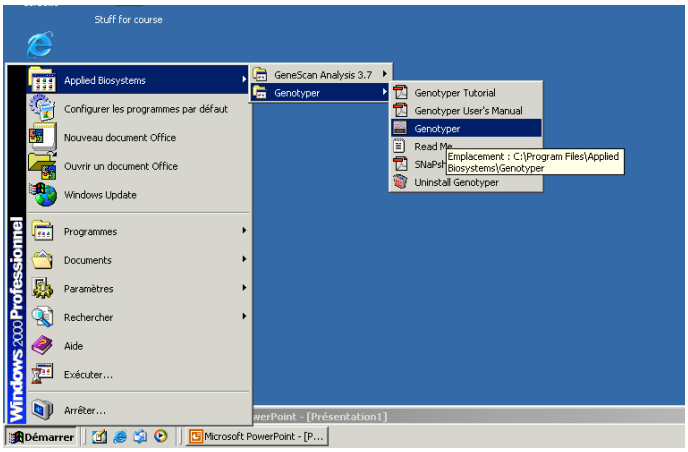
2.6.1 Principle

Size data obtained using GeneScan are edited and converted into MIRU-VNTR alleles, by using the Genotyper software. Therefore, customized files, called templates, are used for each multiplex (or for any of the 3 loci composing it). These templates contain allele-calling information for the corresponding loci, as well as macros that are used to automate the genotyping process.

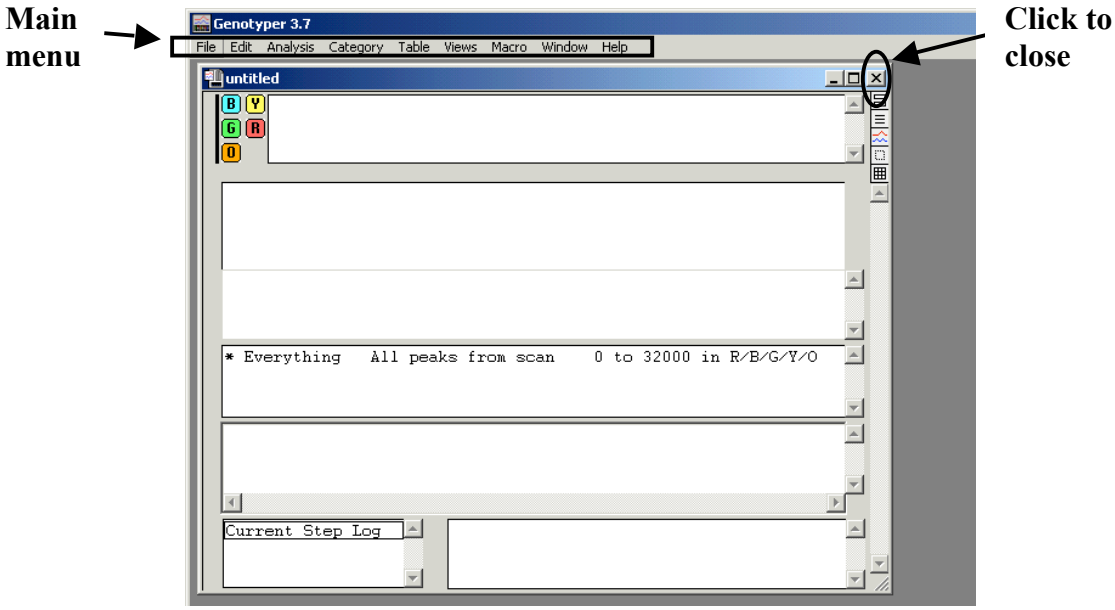
2.6.2 Procedure

2.6.2.1 Opening templates and checking import preferences

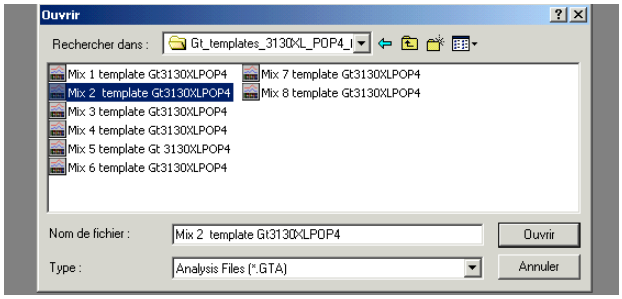
1. Open the application from the desktop or the Windows menu by clicking 



2. The Genotyper main menu is boxed below. Close the untitled file in the pop-up window.

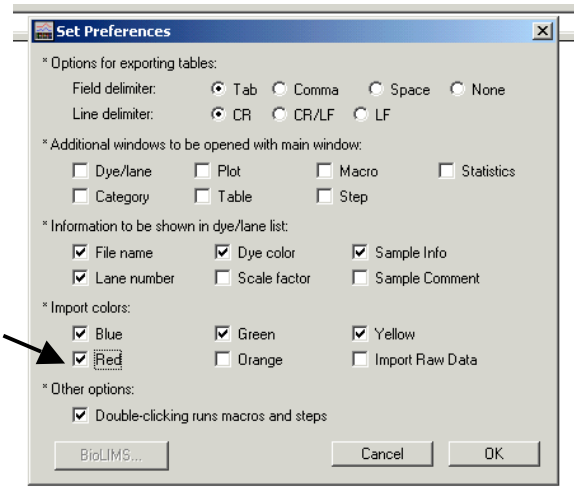


3. From the Genotyper main menu, select **File** and then **Open**. Select the Genotyper template corresponding to the multiplex (or individual loci) to be analyzed (see Table 2).



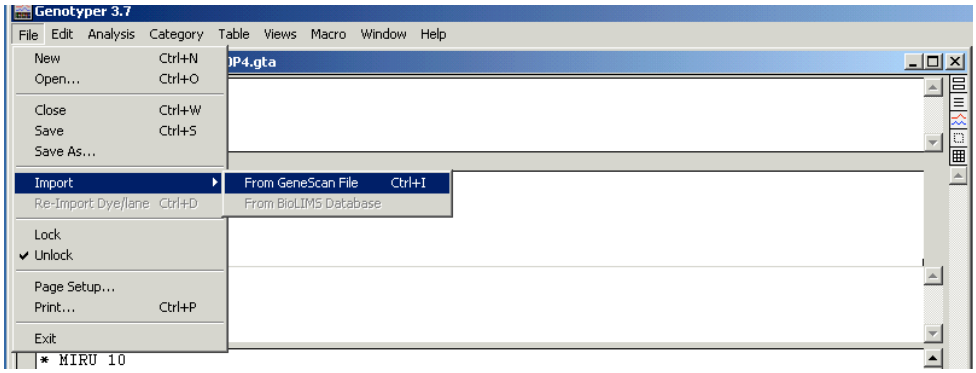
4. Before importation of GeneScan files, importation preferences **must be set as to import red color information** (corresponding to the size standard), in addition blue, green and yellow information for the PCR products. Therefore, from Genotyper main menu, select **Edit** and then **Set Preferences**. Make sure that Red is selected under Import colors. If not, tick the selection box as indicated, then click **OK**, then select **Save** from the main menu.

Tick to select

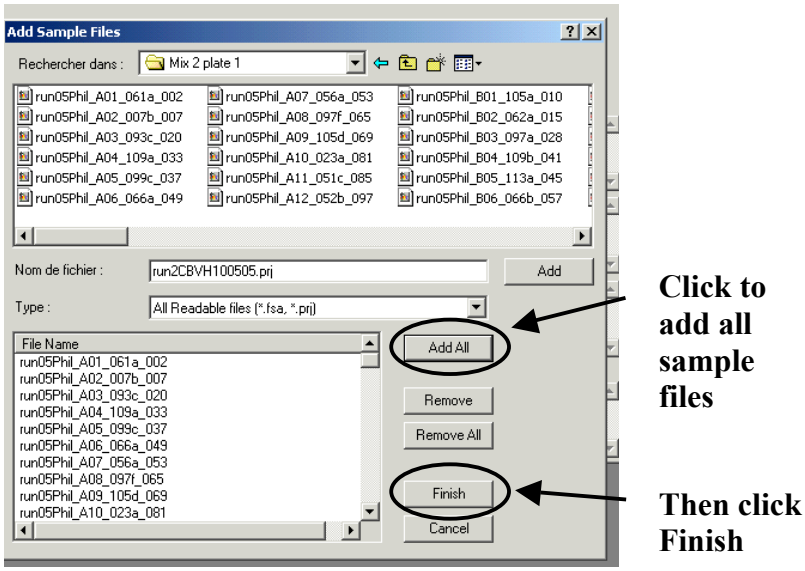


2.6.2.2 Importing GeneScan files

5. From the main menu, select **File**, then **Import From GeneScan File**.



6. Select the folder containing GeneScan sample files corresponding to the multiplex to be analyzed. If the sample folder contains files analyzed for a single multiplex, then click **Add All**, and **Finish**. If the folder contains files from different multiplexes, select files corresponding to the Genotyper template by **clicking+Ctrl** the sample names, then click **Add** and **Finish**.

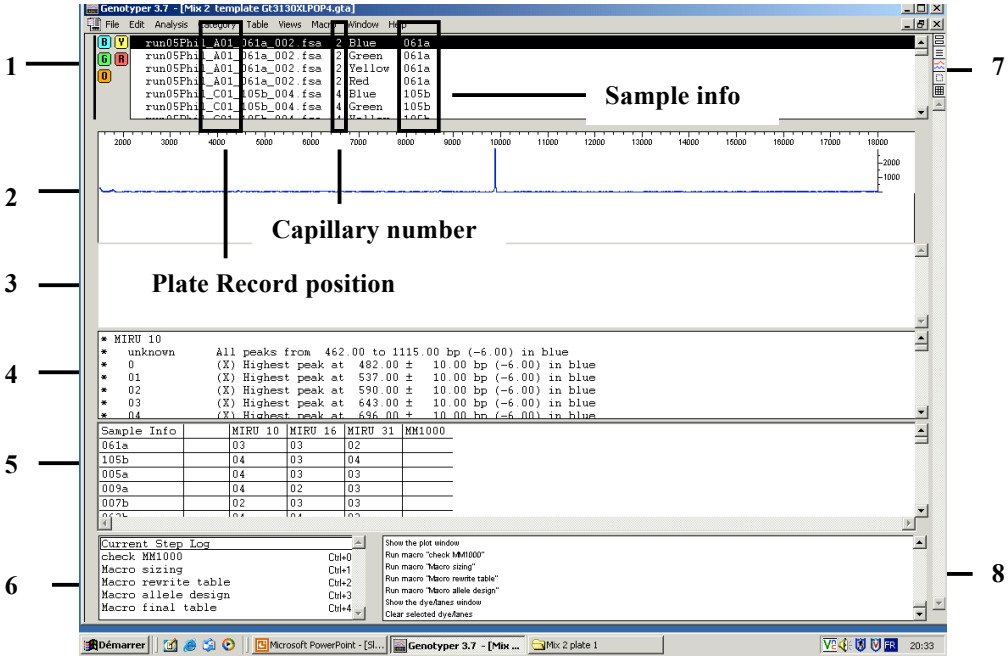


Click to add all sample files

Then click Finish

2.6.2.3 Main window and other windows of interest

7. The Genotyper Main Window should open as shown in the example below.



8. Parts of the Main Window are described below (from Genotyper User's Manual).

Parts of the Main Window

The following table describes the part of the Main Window in the above figure.

Parts of the Main Window:

Item	Name	Description
1	Dye/lane list	Shows specific dye/lanes available for analysis.
2	Upper Graphical Area	Shows electropherogram plots.
3	Lower Graphical Area	Shows peak labels.
4	Category list	Shows criteria for a group of peaks selected on the basis of parameters you define using the category features. For example, dye color, size, or height.
5	Table Area	Shows tabular data for created tables.
6	Macro list	Lists the names of the macros that you have created and can run.
7	Window selection buttons	Open windows for a particular Genotyper Document.
8	Step list	Contains the list of steps for the current Step Log or the macro selected in the Macro list.

Button	Description
	Main window
	Dye/lane List window
	Plot window
	Category window
	Table window

9. The category Window, opened by clicking  contains the allele calling information specific for each multiplex. The example below corresponds to multiplex 1.

```
* MIRU 04
*   unknown      All peaks from  102.00 to 1115.00 bp (-12.00) in blue
*   0          (X) Highest peak at  175.00 ±   10.00 bp (-5.00) in blue
*   0'         (X) Highest peak at  122.00 ±   10.00 bp (-3.00) in blue
*   1          (X) Highest peak at  252.00 ±   10.00 bp (-7.00) in blue
*   1'         (X) Highest peak at  199.00 ±   10.00 bp (-6.00) in blue
*   10         (X) Highest peak at  945.00 ±   10.00 bp (-27.00) in blue
*   11         (X) Highest peak at 1022.00 ±   10.00 bp (-29.00) in blue
*   12         (X) Highest peak at 1099.00 ±   10.00 bp (-31.00) in blue
*   13         (X) Highest peak at 1176.00 ±   10.00 bp (-33.00) in blue
*   14         (X) Highest peak at 1253.00 ±   10.00 bp (-35.00) in blue
*   15         (X) Highest peak at 1330.00 ±   10.00 bp (-37.00) in blue
*   2          (X) Highest peak at  329.00 ±   10.00 bp (-11.00) in blue
*   2'         (X) Highest peak at  276.00 ±   10.00 bp (-7.00) in blue
*   3          (X) Highest peak at  406.00 ±   10.00 bp (-14.00) in blue
*   3'         (X) Highest peak at  353.00 ±   10.00 bp (-12.00) in blue
*   4          (X) Highest peak at  483.00 ±   10.00 bp (-15.00) in blue
*   5          (X) Highest peak at  560.00 ±   10.00 bp (-17.00) in blue
*   6          (X) Highest peak at  637.00 ±   10.00 bp (-19.00) in blue
*   7          (X) Highest peak at  714.00 ±   10.00 bp (-21.00) in blue
*   8          (X) Highest peak at  791.00 ±   10.00 bp (-23.00) in blue
*   9          (X) Highest peak at  868.00 ±   10.00 bp (-25.00) in blue
* MIRU 26
*   unknown      All peaks from  265.00 to 1115.00 bp (-5.00) in green
*   00         (X) Highest peak at  285.00 ±   10.00 bp (-5.00) in green
*   01         (X) Highest peak at  336.00 ±   10.00 bp (-5.00) in green
*   02         (X) Highest peak at  387.00 ±   10.00 bp (-5.00) in green
*   03         (X) Highest peak at  438.00 ±   10.00 bp (-5.00) in green
*   04         (X) Highest peak at  489.00 ±   10.00 bp (-5.00) in green
*   05         (X) Highest peak at  540.00 ±   10.00 bp (-5.00) in green
*   06         (X) Highest peak at  591.00 ±   10.00 bp (-5.00) in green
*   07         (X) Highest peak at  642.00 ±   10.00 bp (-5.00) in green
*   08         (X) Highest peak at  693.00 ±   10.00 bp (-5.00) in green
*   09         (X) Highest peak at  744.00 ±   10.00 bp (-5.00) in green
*   10         (X) Highest peak at  795.00 ±   10.00 bp (-5.00) in green
*   11         (X) Highest peak at  846.00 ±   10.00 bp (-5.00) in green
*   12         (X) Highest peak at  897.00 ±   10.00 bp (-5.00) in green
*   13         (X) Highest peak at  948.00 ±   10.00 bp (-5.00) in green
*   14         (X) Highest peak at  999.00 ±   10.00 bp (-5.00) in green
*   15         (X) Highest peak at 1050.00 ±   10.00 bp (-5.00) in green
* MIRU 40
*   unknown      All peaks from  334.00 to 1115.00 bp (-1.00) in yellow
*   0          (X) Highest peak at  354.00 ±   10.00 bp (-4.00) in yellow
*   1          (X) Highest peak at  408.00 ±   10.00 bp (-8.00) in yellow
*   10         (X) Highest peak at  894.00 ±   10.00 bp (-49.00) in yellow
*   11         (X) Highest peak at  948.00 ±   10.00 bp (-54.00) in yellow
*   12         (X) Highest peak at 1002.00 ±   10.00 bp (-59.00) in yellow
*   13         (X) Highest peak at 1056.00 ±   10.00 bp (-64.00) in yellow
*   14         (X) Highest peak at 1110.00 ±   10.00 bp (-69.00) in yellow
*   15         (X) Highest peak at 1164.00 ±   10.00 bp (-74.00) in yellow
*   2          (X) Highest peak at  462.00 ±   10.00 bp (-12.00) in yellow
*   3          (X) Highest peak at  516.00 ±   10.00 bp (-16.00) in yellow
*   4          (X) Highest peak at  570.00 ±   10.00 bp (-20.00) in yellow
*   5          (X) Highest peak at  624.00 ±   10.00 bp (-25.00) in yellow
*   6          (X) Highest peak at  678.00 ±   10.00 bp (-30.00) in yellow
*   7          (X) Highest peak at  732.00 ±   10.00 bp (-35.00) in yellow
*   8          (X) Highest peak at  786.00 ±   10.00 bp (-39.00) in yellow
*   9          (X) Highest peak at  840.00 ±   10.00 bp (-44.00) in yellow
```

Note: ± x bp and (- y bp) define size tolerance and size offsets (see **2.7**) for allele identification, respectively; blue, green and yellow define the dye labeling for the respective loci. The Unknown category finds and labels any peaks that do not belong to any of the allelic bins, but are within the size range for that locus. Unknown labels are useful hints for detection of potential problematic peaks.

10. The Category Window also contains the information for size standard peak labeling, common to all multiplexes.

```
* MM1000
*   025       (X) Highest peak at   25.00 ±    0.50 bp in red
*   050       (X) Highest peak at   50.00 ±    0.50 bp in red
*   075       (X) Highest peak at   75.00 ±    0.50 bp in red
*   100       (X) Highest peak at  100.00 ±    0.50 bp in red
*   1000      (X) Highest peak at 1000.00 ±    0.50 bp in red
*   125       (X) Highest peak at  125.00 ±    0.50 bp in red
*   150       (X) Highest peak at  150.00 ±    0.50 bp in red
*   200       (X) Highest peak at  200.00 ±    0.50 bp in red
*   250       (X) Highest peak at  250.00 ±    0.50 bp in red
*   300       (X) Highest peak at  300.00 ±    0.50 bp in red
*   350       (X) Highest peak at  350.00 ±    0.50 bp in red
*   400       (X) Highest peak at  400.00 ±    0.50 bp in red
*   450       (X) Highest peak at  450.00 ±    0.50 bp in red
*   475       (X) Highest peak at  475.00 ±    0.50 bp in red
*   500       (X) Highest peak at  500.00 ±    0.50 bp in red
*   550       (X) Highest peak at  550.00 ±    0.50 bp in red
*   600       (X) Highest peak at  600.00 ±    0.50 bp in red
*   650       (X) Highest peak at  650.00 ±    0.50 bp in red
*   700       (X) Highest peak at  700.00 ±    0.50 bp in red
*   750       (X) Highest peak at  750.00 ±    0.50 bp in red
*   800       (X) Highest peak at  800.00 ±    0.50 bp in red
*   850       (X) Highest peak at  850.00 ±    0.50 bp in red
*   900       (X) Highest peak at  900.00 ±    0.50 bp in red
*   950       (X) Highest peak at  950.00 ±    0.50 bp in red
```

11. The Macro Window contains information about shortcut commands (Ctrl+0, etc) to run macros, and corresponding operation steps. These operations can be viewed by clicking macros in this Window, and are summarized in the Table below.

Macro	Command	Action
Check MM 1000	Ctrl+0	Labels and visualizes size standard peaks with size in bp
Sizing	Ctrl+1	Labels, filters and visualizes locus peaks with size in bp (see below), and writes intermediate result table
Rewrite table	Ctrl+2	Updates intermediate result table after data editing
Allele design	Ctrl+3	Converts sizes into alleles
Final table	Ctrl+4	Writes a final table

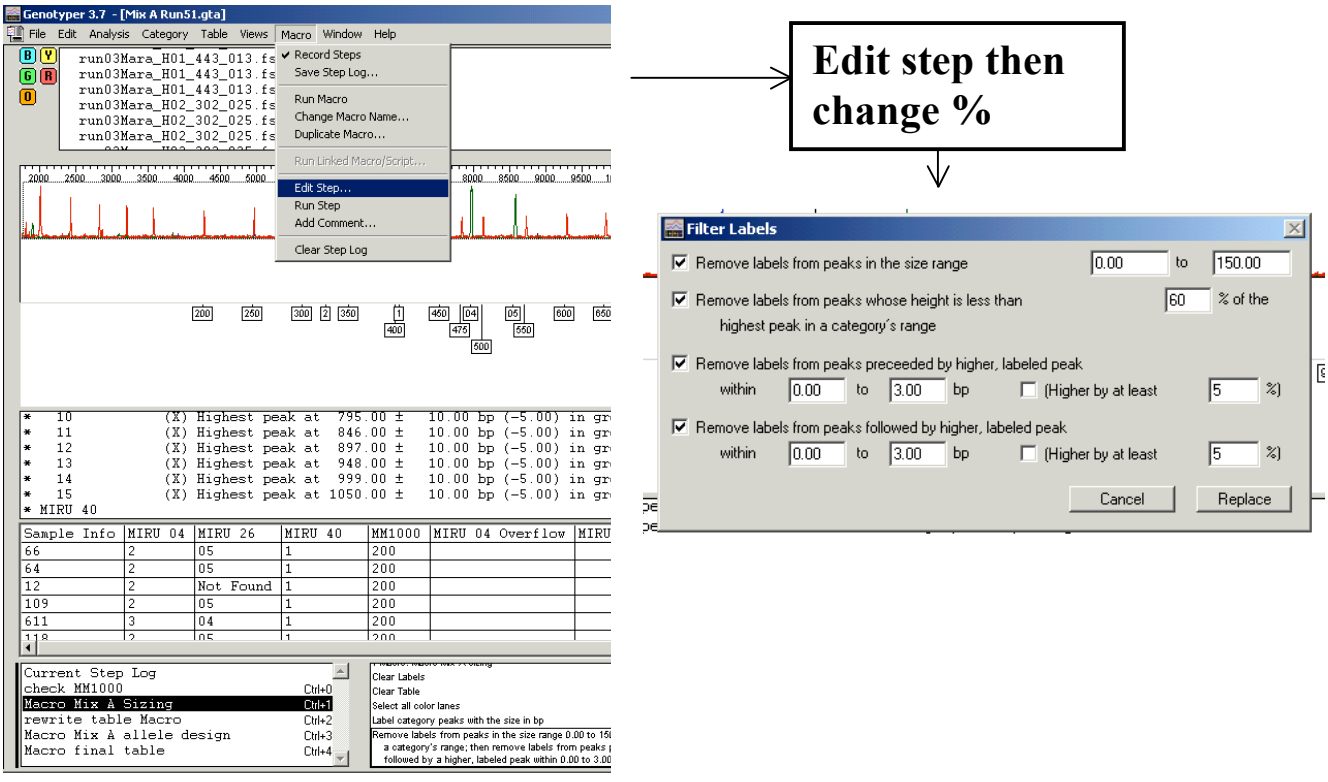
12. Filtering steps in the sizing macro (operated by pressing Ctrl+1) are described below. The “Remove labels from peaks in the size range xx to yy bp” step is the only macro step that differs between multiplex templates.

Steps	Effect
Remove labels from peaks in the size range 0.00 to 90.00 bp	Removes labels in size range below size of the smallest amplicon in any locus in multiplex
then remove labels from peaks whose height is less than 32% of the highest peak in a category’s range	Removes labels from **SUSPECTED** noise and stutter peaks
then remove labels from peaks preceeded by a higher, labeled peak within 0.00 to 10.00 bp	Removes labels from **SUSPECTED** shoulder peaks, if peak height is lower than allele peak
then remove labels from peaks followed by a higher, labeled peak within 0.00 to 10.00 bp	Removes labels from **SUSPECTED** shoulder peaks, if peak height is lower than allele peak

Note: when PCR signals are strong and cause frequent pull-up peaks, peak filtering level can be set to 60 % instead of 32 %, as follows:

a) Select Macro sizing from the main window, and from menu, select **Macro>Edit Step**.

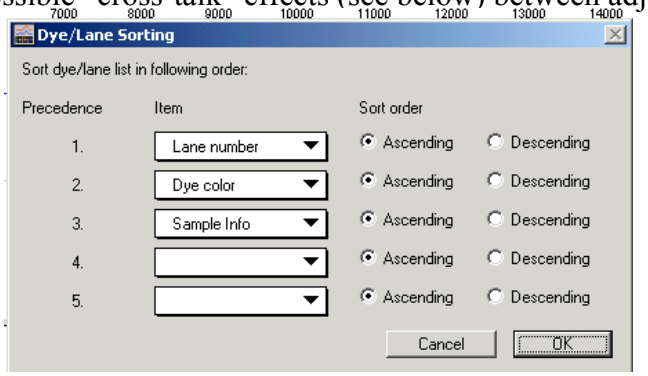
b) Replace 32 % by 60 % in the % of the highest peak in a category’s name box, and click **Replace**.



2.6.2.4 View, dye/lane sorting, and peak labeling options

13. From Genotyper main menu, select **Views**, then **Display by Scan**. *Note:* **Important.** Display problems may appear if results are displayed by sizes, due to possible inconsistencies among sizing data obtained from some capillaries.

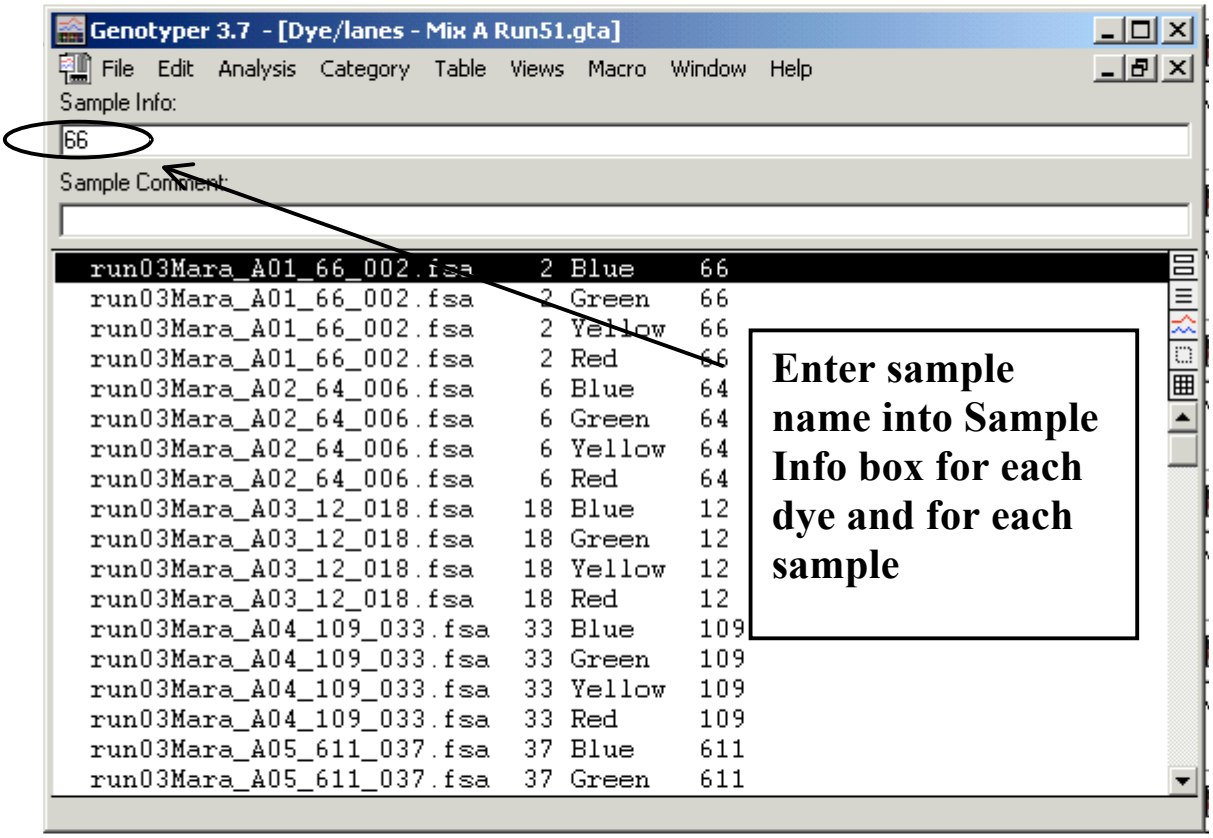
14. Select **Views** from Genotyper main menu, then **Dye/Lane Sorting**, then select display **Sample name**, **Dye color**, and **Sample Info** as criterions 1, 2, and 3, respectively, from the pull-down menus shown below, and press **OK**. *Note:* **for 3700 or 3730 systems,** sorting by Lane number/**Dye color**/**Sample Info** is required for detection of possible "cross-talk" effects (see below) between adjacent capillaries.



2.6.2.4a. Entering Sample Info (only when using Data Collection version 3.x)

Entering sample info is mandatory for construction final result table in Genotyper. Unfortunately, it is not possible to enter this info in the plate record (see page 25), when using Data Collection software version 3.x. To by-pass this problem, this info must be entered for each sample file at the level of Genotyper. Therefore,

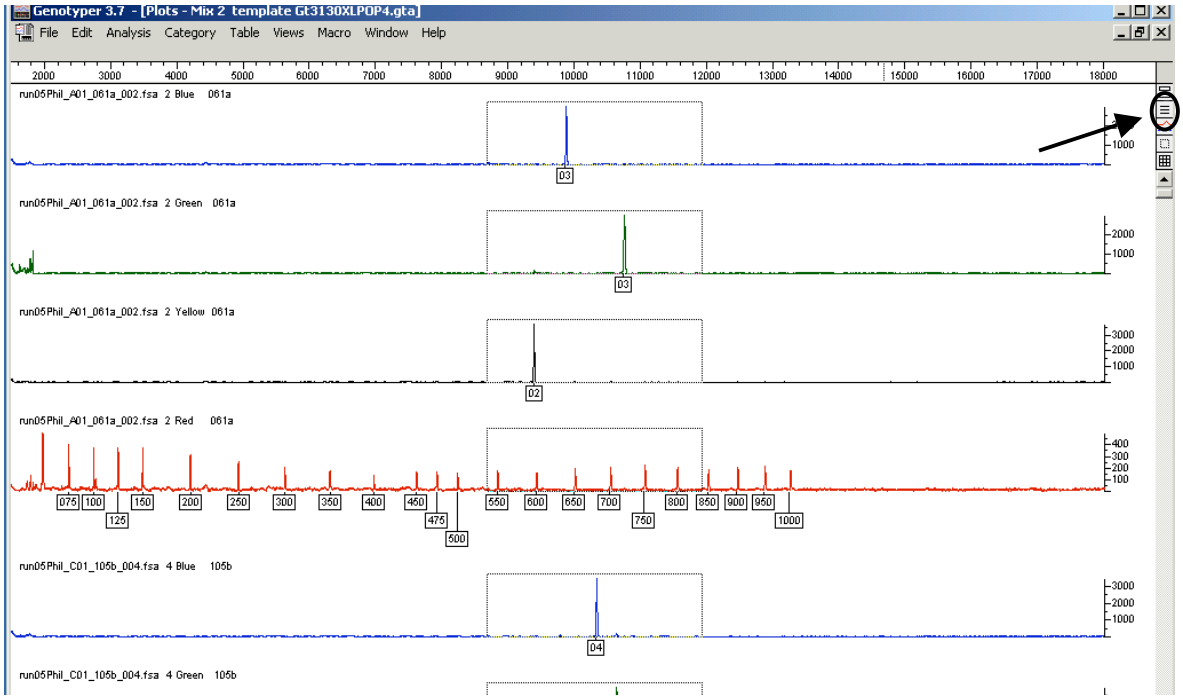
15. Open the Dye/Lane List window.
16. In the list, select successively **each Lane for each Dye** and enter the corresponding sample name into Sample Info box in the top of the Dye/Lane window, as shown below. **Beware of clerical errors !**
17. When finished, select **Save As** from main **Menu>File**, and save as your experiment multiplex x in the appropriate folder.



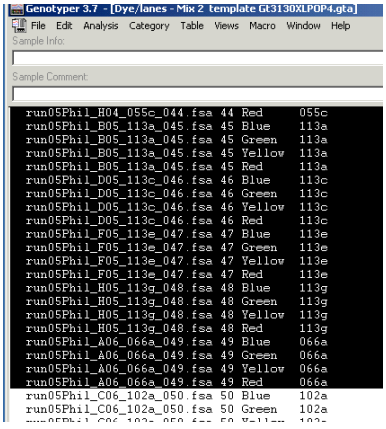
2.6.2.5 Running Macros and viewing results

18. Press **Ctrl+0**, immediately followed by **Ctrl+1**, **Ctrl+2**, and **Ctrl+3**. Wait a while until the process is finished. *Note:* These Macros can be run and their action viewed separately.

19. An example of result is displayed below.



20. **<u>When using a PC Genotyper version, a bug in plot display appears when analyzing many sample files in a single template</u>**. To overcome it, dye/lanes must be analyzed by halves. Therefore, click the Dye/Lane Window button, then select one first half of the dye lanes as shown below, by clicking the first Dye/Lane and then **Shift+clicking** the last Dye/Lane. **Notice the last sample of this half.**

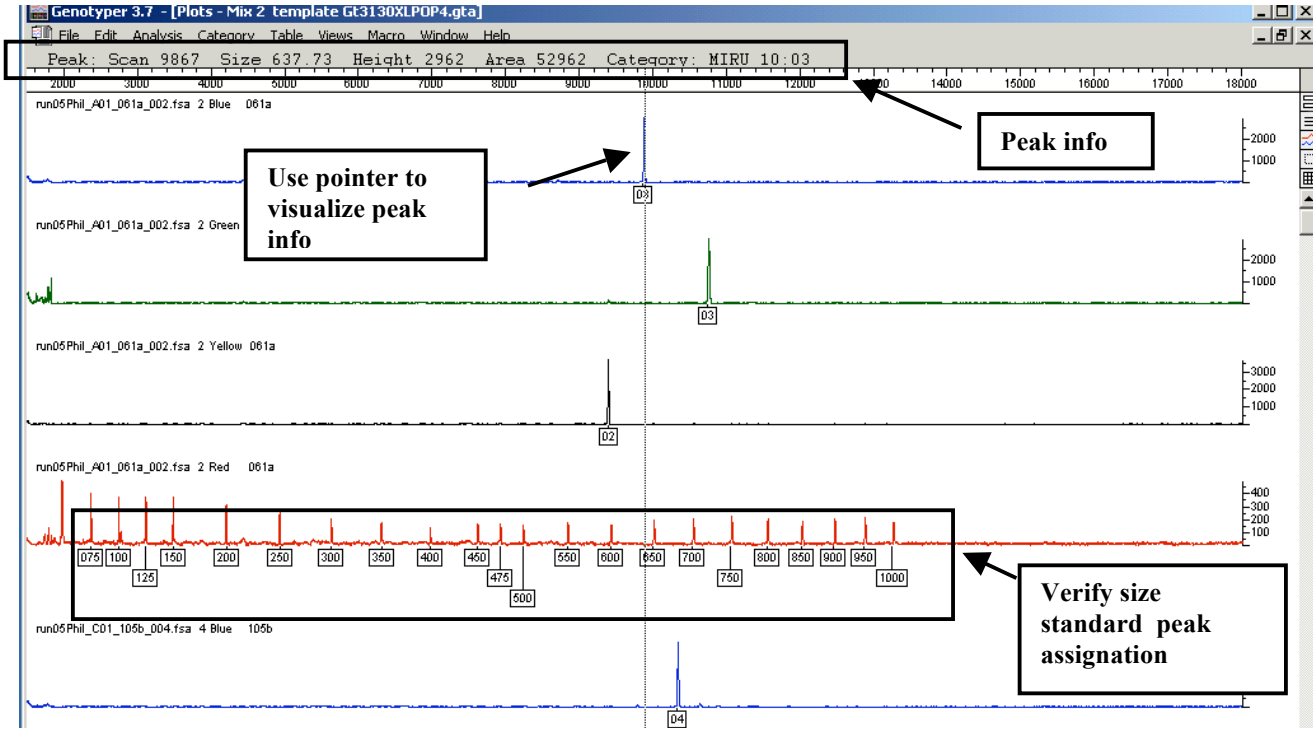


2.6.2.6 Verifying size standard assignation and peak information

21. For each lane/capillary, verify that the size standard assignation is correct, *i.e.* all peaks within the analysis range (defined by data points in GeneScan analysis, see 2.5.2.2) must be labeled correctly as indicated in the example below.

22. Verify the presence of allele label for each colour. Unless some clonal variant subpopulation or mixed DNA population are present (see below), **<u>one label is expected per color. If pull-up peaks (see 2.6.2.7.2) are too frequent, change the peak filtering level in the sizing macro (see 2.6.2.7.2) and rerun macros as above.</u>**

23. Peak information, including the positions in scan numbers, the size in bp, the peak height and area, and the corresponding allele (category), can be visualized for each color detected a capillary, by placing the cross-hair cursor on the peak of interest as shown below.



2.6.2.7 Detecting and removing Genotyper labels from problematic peaks

Occasionally, size standard assignation may be incorrect. Consequently, sizes/alleles are not assigned correctly for PCR products.

In addition, despite filtering by using the sizing macro, various kinds of artifactual peaks may be labeled. Such peaks include:

1) "Pull-up" peaks, resulting from spectral overlap from another dye. Such peaks may result from the presence of large amounts of a particular dye-labeled PCR product, or from sub-optimal spectral calibration, or from both causes.
2) Peak shoulders or "mirror" peaks in a same dye. Such peaks result from the presence of large amounts of a particular dye-labeled PCR product, and are therefore often associated with pull-up peaks in other dyes.
3) Spurious peaks. Such peaks are often thin peaks, which can be observed at a same position across several dyes.
4) Peaks resulting from fluorescence cross-talk between flanking capillaries (**on 3700 or 3730 sequencers)**. Such peaks result from the presence of very large amounts of PCR products in a proximal capillary. They should not observed on 4-, or 16-capillaries 3100/3130 sequencers.
5) Stutter peaks, classically caused by slippage of the polymerase enzyme during PCR.

Importantly, peaks of types 1) to 4) are frequently tagged with "unknown" labels, as their size do not belong to any of the allelic bins. **Thus, the presence of "unknown" labels particularly prompts for attention**.

However, this rule is not systematic, as some false peaks can fall into an allelic bin by chance, and can therefore be tagged with allelic labels.

Labels from all types mentioned above should be removed, according to algorithms explained below.

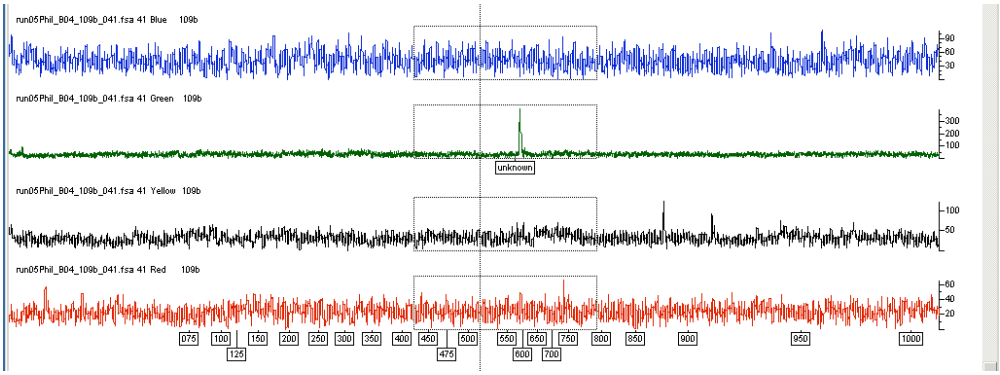
The procedure for label removal is the following:

1. Place the cross-hair cursor on the peak of interest. The vertical peak locator will be aligned with the midpoint of the peak.
2. Click **once** to remove the label from any incorrectly sized, artifactual, or stutter peak.

2.6.2.7.1 Incorrectly sized peaks

If size standard assignation is incorrect for results from a capillary, i.e. if peaks within the analysis range (defined by data points/scan numbers defined in GeneScan analysis parameters, see 2.5.2.2) are not labeled correctly, sizes/alleles will not be assigned correctly for the corresponding PCR products. Such a problem may for instance happen consecutively to a sample injection problem, resulting in low or absence of size standard peaks and incorrect size standard assignation to background signals, as shown in the example below (in this case the peak height detection threshold was set to 50).

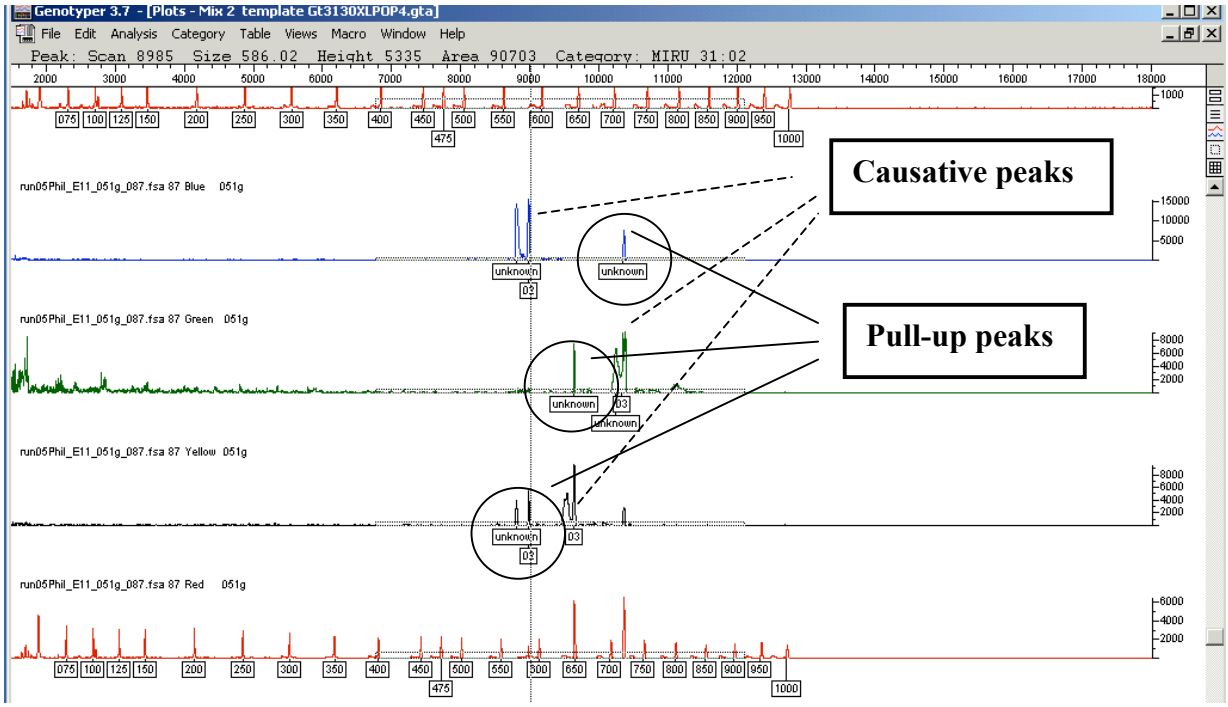
1. Inspect the size standard label pattern.
2. In case of size assignation problems, remove the labels**, corresponding to either unknown or apparently recognized alleles,** from **all** PCR product peaks from the same capillary.



2.6.2.7.2 Pull-up peaks

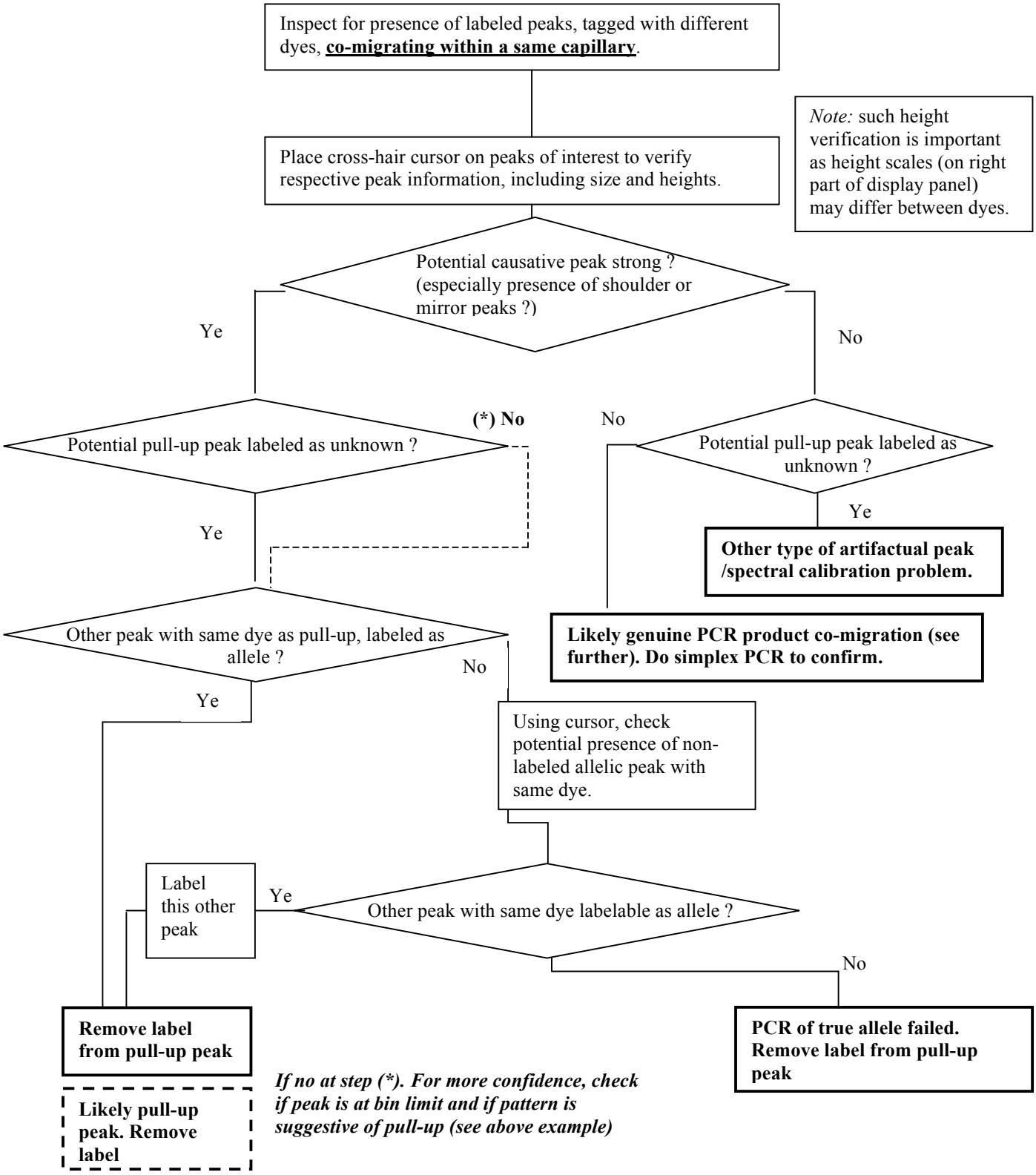
"Pull-up" peaks result from spectral overlap from another dye within a same capillary. They can be diagnosed, as they usually appear as peaks exactly co-migrating with a strong peak labeled with another dye.

Diagnostics is especially facilitated when 1) pull-up peaks are labeled as "unknown", **and** 2) another peak tagged with the same dye is labeled as a recognized allele, **and** 3) the causative peak labeled with a different dye displays itself shoulder or mirror peaks indicative of signal saturation, as in the example shown below.



1. **Systematical observation of pull-up peaks caused by strong genuine PCR products, is indicative of excess of PCR products**. In this case, reduce genomic DNA concentration, PCR cycle numbers, or dilute PCR products before addition to the loading mix.
2. **Systematical observation of pull-up peaks, even when the heights of causative true peaks are small, is indicative of sub-optimal spectral calibration**. Re-perform spectral calibration.
3. More generally, labels from co-migrating, potential pull-up peaks can be detected and interpreted according to the algorithm shown in Figure 16.

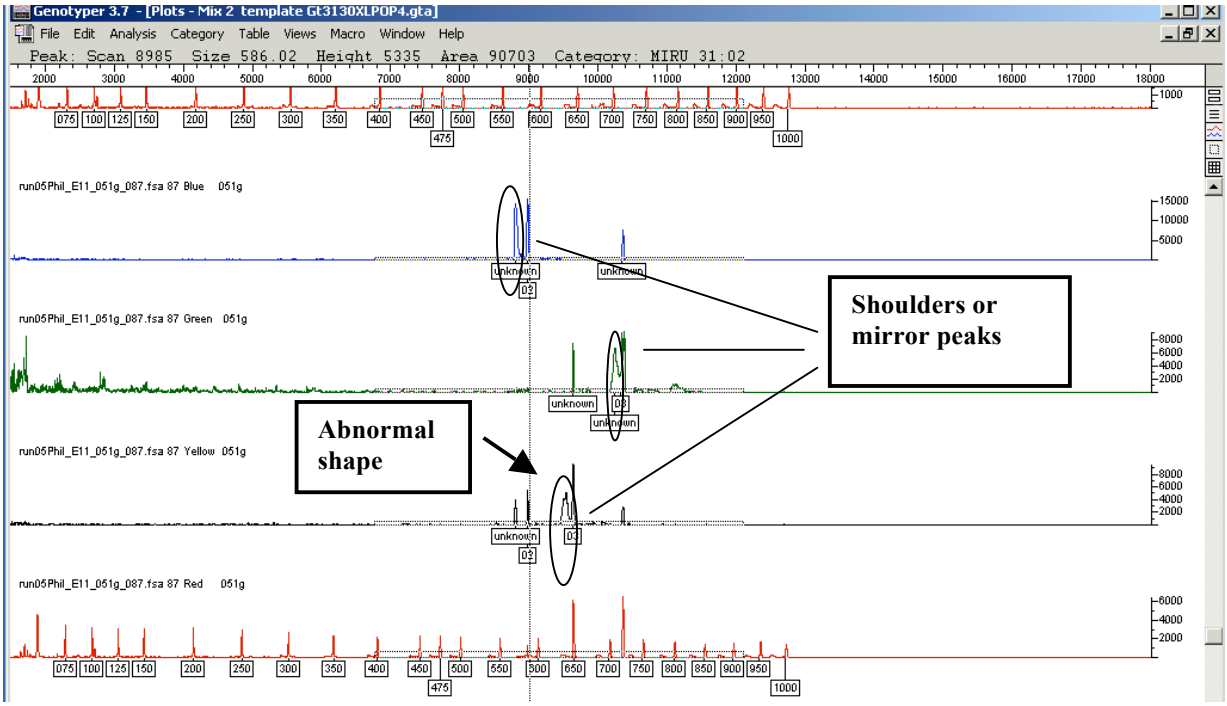
Figure 16 Algorithm for interpretation of potential pull-up peaks



2.6.2.7.3 Peak shoulders or mirror peaks

The presence of large amounts of a particular dye-labeled PCR product may cause the appearance of shoulders besides the genuine allelic peak. Furthermore, so-called mirror effects may occur, resulting in apparent and typical digging of the central part of large peaks, and even apparent reduction of the peak boundaries (see example below).

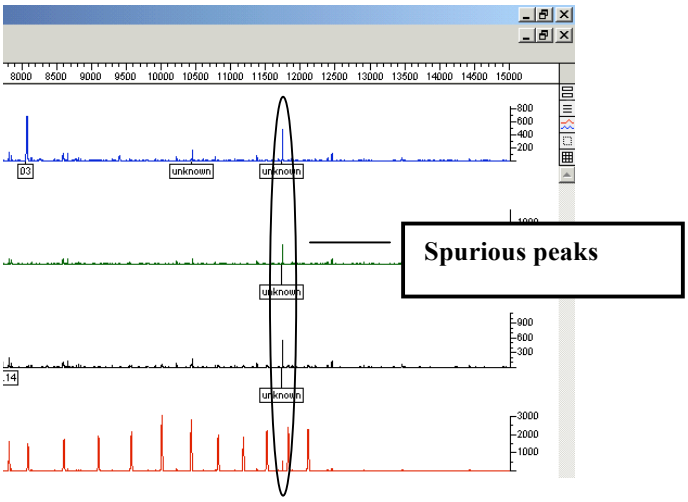
In most cases, the boundary with the lowest apparent size in bp is labeled as unknown, while the boundary with the largest size in bp is recognized as a true allele. **However, in extreme cases, the two boundaries may be separated by apparent sizes as large as one repeat length, and may thus appear as two independent allelic peaks at first glance.** Therefore, the diagnosis of such peaks is also based on their shape, and with frequent association with the presence of pull-up peaks in other dyes.



1. Remove labels from obvious peak shoulder or mirror peaks, initially assigned as unknown.
2. Check the presence of mirror peaks with abnormal shapes as seen above, initially assigned as alleles. Remove their labels.

2.6.2.7.4 Spurious peaks

Spurious peaks appear as thin peaks, that may be observed at a same position across several dyes within a capillary. It may be caused by the presence of a bubble during migration.

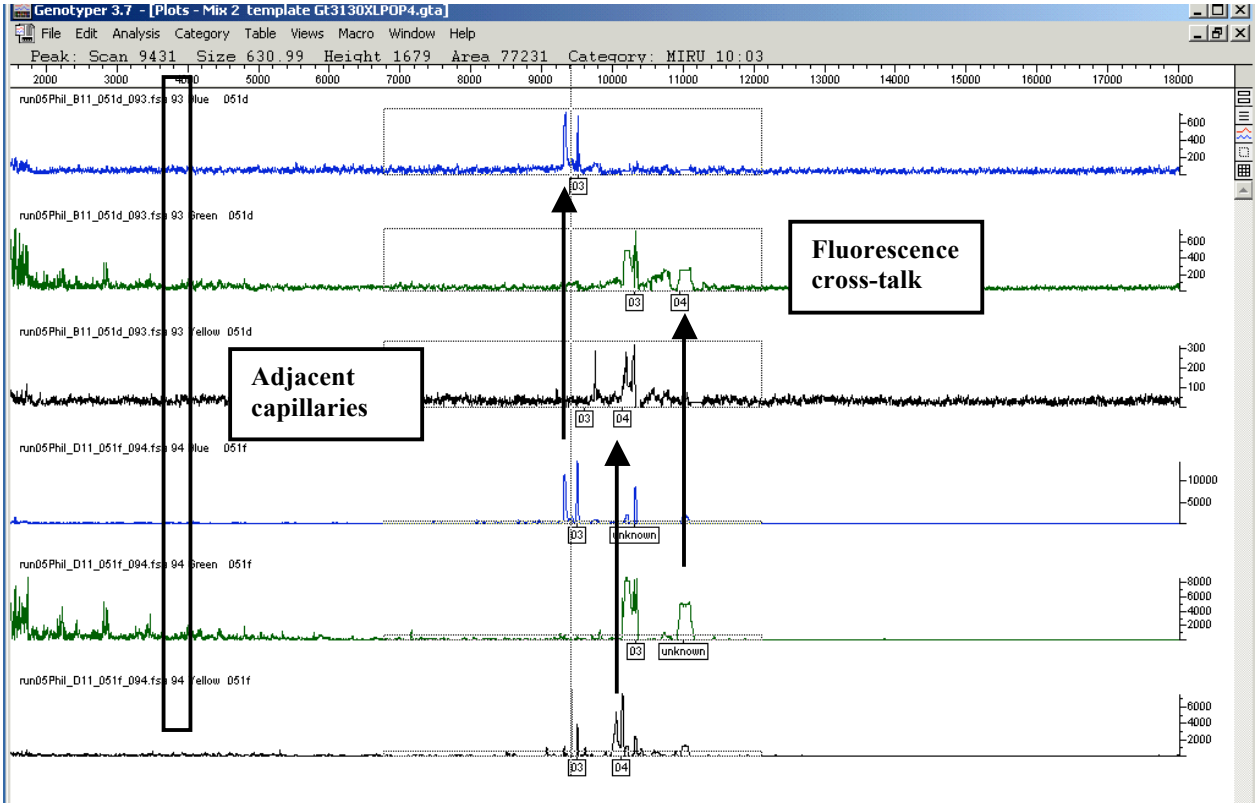


1. Check for the presence of any labeled thin peaks, observed at identical positions across all dyes.
2. Remove their labels.

2.6.2.7.5 Cross-talks between flanking capillaries (for 3700 and 3730)

Artifactual peaks may occur in a capillary, when very large amounts of PCR products are present in the adjacent capillary (see example below). This phenomenon is described as fluorescence cross-talk. It is similar to "contamination" effects, which may be seen between adjacent gel lanes when using a 377 sequencer. **Cross-talk effects may be observed on 96-capillaries 3700 or 3730 sequencers, but should not observed on 4-, or 16-capillaries 3100 sequencers.**

1. Compare patterns from immediately adjacent lanes/capillaries
2. Check for the possible occurrence of exactly co-migrating peak patterns, between a lane containing labeled weak peaks and the immediately proximal lane (see below).
3. Remove the corresponding labels, if any.



2.6.2.7.6 Stutter peaks

As indicated in 2.3.3, stutter peaks can be easily diagnosed, as they appear most often as a regular ladder of much lower intensity peaks (see example A below), corresponding to sizes of PCR fragments that lack one or more repeats, or more rarely that contain one or more additional repeat. Therefore, as for electrophoretic analysis using agarose gels, the positions of stutter peaks can be used to confirm the allelic assignation of the principle fragment (see point 5 in 2.3.2).

However, sub-optimal amplification of some loci, especially with larger repeat numbers, may sometimes result in stronger intensities of stutter peaks (example B). Exceptionally, band ladder with no clearly sharpest band can occur, preventing any allelic assignation (example C).

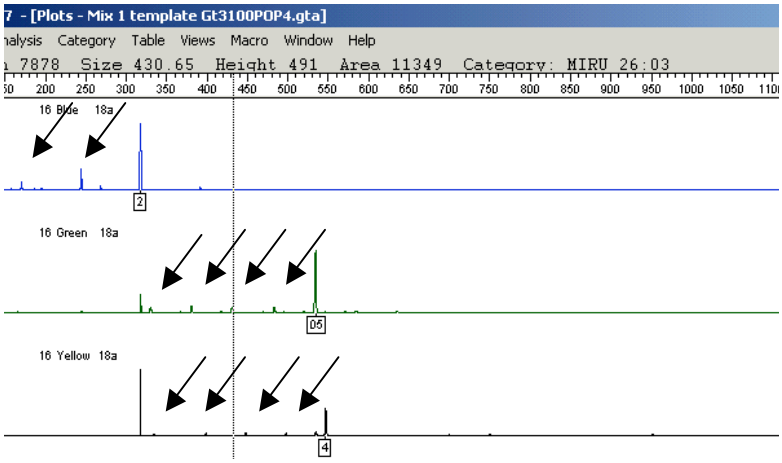
Based on analysis of appropriate controls, we have determined a rule for correctly distinguishing true alleles from stutter peaks as follows:

1. The true allele is assigned as the peak having the highest repeat number, **beyond which a sharp decrease is observed in heights of further stutter peaks (if any)**. **If further stutter peaks are present, they must be uniformly small.**
2. Remove labels from other peaks, if any.

3. If a “bell-shaped” distribution of band intensities is observed without any clear sharp band at the extremity of the ladder**, remove all labels and re-amplify.**

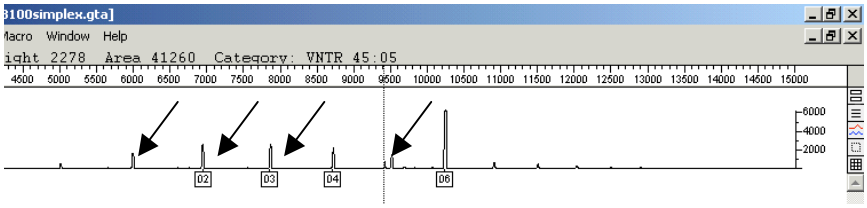
Example A: usual stutter pattern

Stutters are indicated by arrows



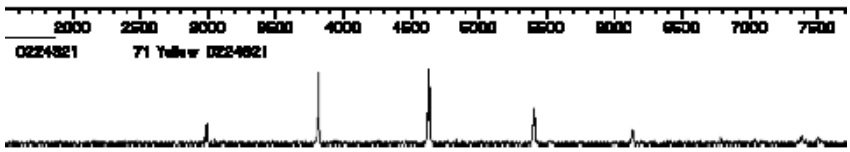
Example B: more complex stutter pattern

Stutters are indicated by arrows. The corresponding locus (VNTR 1982) is not included in our panel.



Example C: non-interpretable stutter pattern

The corresponding locus (VNTR 1982) is not included in our panel.



2.6.2.8 Adding labels to true alleles

Due to the label filtering process, true alleles may occasionally not be labeled. This can typically occur when some intense pull-up peaks are observed. In this case, because of the sizing macro parameters, true alleles with a peak height less than 32% of these pull-up peaks will not be labeled. The corresponding peaks must then be labeled manually.

In such conditions, because of disproportionate impact of pull-up peaks on the plot scale, these true alleles might remain unnoticed, especially when they correspond to large repeat numbers. Therefore, particular attention must be paid to detect them.

1. When intense pull-up or other artifactual peaks are present, inspect the plot patterns according to the algorithm in Figure 16, by moving the cursor along the whole analysis range to facilitate detection of potential true alleles.
2. Taking into account Place the cross-hair cursor on the peak of interest. The vertical peak locator will be aligned with the midpoint of the peak.
3. Click **once** to add a label to the true peak.
4. Press **Ctrl+3** to change the size label into an allele label.

2.6.2.9 Genuine co-migration between PCR products tagged with different dyes

Depending on the polymer used, nearly exact co-migration of two genuine PCR products labeled with different dyes can occur within some multiplex (*e.g.* allele 2 for loci 47 and 52 in our conditions). This can be predicted, after size calibration and application of appropriate size offsets, from comparison of the allele-calling table between the respective loci within the different multiplexes. However, it is sometimes to distinguish genuine co-migration from pull-up effects

1. To distinguish between these two possibilities, use the algorithm in Figure 16. Allelic labels can be assigned only if both respective peaks have moderate heights.
2. In case of doubts, re-amplify the two loci by simplex PCR, and analyze them separately.

2.6.2.10 Genuine double alleles

After application of the filtering process described above, genuine double alleles can be observed in certain cases. If 2 alleles are concordantly observed in several independent MIRU-VNTR loci for a given sample, this indicates the presence of a mixed DNA population in this sample. This mixed population can result from a true mixed infection, or from culture or DNA contamination.

In contrast, the occurrence of 2 alleles is observed in a single locus rather suggests the presence of a given allelic variant within a clonal isolate.

2.6.2.11 Viewing and analyzing the second half of the capillary lanes

If applicable, the second half of the capillary lanes can be analyzed according to the same procedure as described above.

To select this half, click the **Dye/Lane Window** button, then select the appropriate dyes, by clicking the dyes from the first lane and then **Shift+clicking** the dyes from the last lane. *Note:* when using a 96-capillary system, include also the last sample from the first half, to control the occurrence of potential fluorescence cross-talks.

2.6.2.12 Final sorting by file names

After analysis of all capillaries, all sample files should be sorted by file names, as indicated below. This mode of final sorting choice is more reliable and convenient to compare result tables from different analysis runs, among which sample plate mappings are not necessarily identical.

1. To select all samples, click the **Dye/Lane Window** button, then press Ctrl+A to select all lanes.
2. From Genotyper main menu, select **Views**, then **Dye/Lane Sorting**, then select **Sample Name** and **Dye Color** from the first and second pull-down menus, respectively, both in ascending order, and then press **OK.** *Note:* this sorting can also be performed after the creation of the final results table, by selecting **Table** from the main menu, then **Sort Table,** and selecting **Sample Info** from the first pull-down menu.

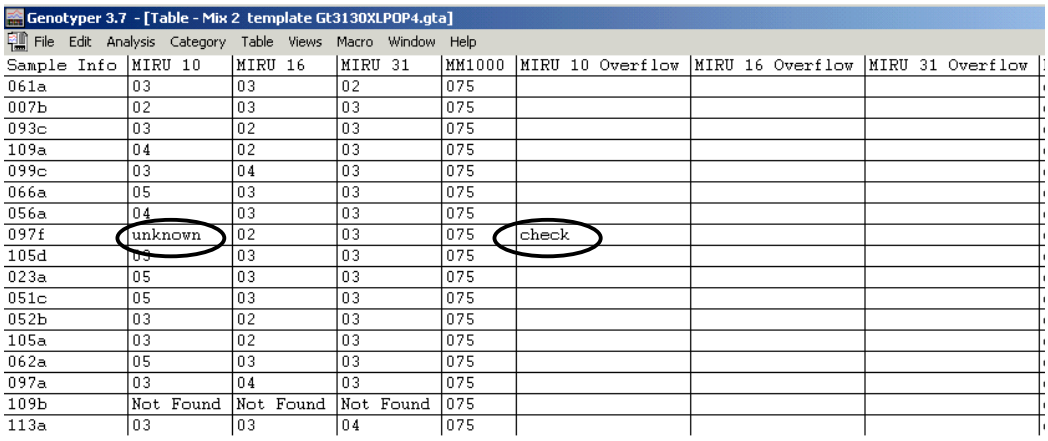
2.6.2.13 Creating final results tables and verifying results

Final results tables (see example below) are created using a macro called “final table” in the Genotyper templates. These tables are used to check, and subsequently export the results.

Columns of interest in this final table include the following information:

- Sample information

- Peak label information for the 3 loci for each sample. The expected information corresponds to alleles. In contrast, **"Unknown" tags are unexpected and prompt for attention**. "Not Found" tags are displayed when no label/result could not be assigned for some loci.
- "Overflow" status for each locus. **"Check" tags (see below) indicate the presence of two or more alleles in a given locus, which is normally unexpected and prompts for attention**.

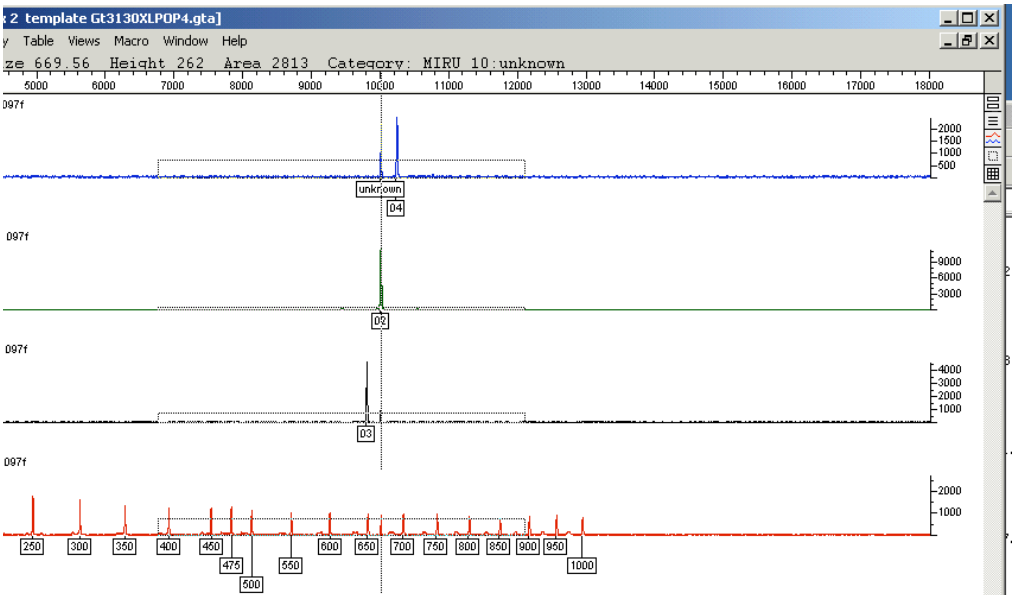


Sample Info	MIRU 10	MIRU 16	MIRU 31	MM1000	MIRU 10 Overflow	MIRU 16 Overflow	MIRU 31 Overflow
061a	03	03	02	075			
007b	02	03	03	075			
093c	03	02	03	075			
109a	04	02	03	075			
099c	03	04	03	075			
066a	05	03	03	075			
056a	04	03	03	075			
097f	unknown	02	03	075	check		
105d	03	03	03	075			
023a	05	03	03	075			
051c	05	03	03	075			
052b	03	02	03	075			
105a	03	02	03	075			
062a	05	03	03	075			
097a	03	04	03	075			
109b	Not Found	Not Found	Not Found	075			
113a	03	03	04	075			

To create a final table and verify the results, proceed as follows:

1. Make sure that all lanes are selected in the Dye/Lane Windows.
2. Press Ctrl+4.
3. Click the Table window button.
4. If any "unknown" or "check" tags are present in locus or locus overflow columns, respectively, click the corresponding "sample info" cell to visualize the peak patterns.
5. If appropriate, remove labels as described in 2.6.2.7.
6. Update the table by pressing **Ctrl+4**.
7. When all results are checked, select **File** from the main menu, then **Save as** Gt_Results_mixz_runx in an appropriate folder.

The following image shows verification of the sample with “unknown” and “check” labels in the example above. In this case, the presence of these tags were indicative of a pull-up peak with the blue dye, clearly caused by a strong peak with the green dye.



2.6.2.14 Exporting final table to Excel

1. From the main menu, select **Table**, then **Export to File**, and **Save** as Results_mixz_runx.
2. From Excel main menu, select **Open**, then **All Types of Files**.
3. Import the table, by clicking twice **Next**, and then **Finish**.

2.7 Size offset calibration

Differences in relative migration can occur between the size standard and the amplicons on the polymer depending on the polymer used for electrophoresis, and on the locus. In some loci, these differences can be fixed, or while in others it increases linearly upon the number of repeats. Also depending on the locus, offsets may be negative or positive. Therefore, **it is essential to take this effect into account, otherwise allelic assignation in terms of actual repeat numbers will be wrong**.

Application of size offsets in the Genotyper allele calling information correct these differences. These offsets are determined experimentally, by using reference PCR products with

alleles and sizes pre-determined using genomic data (*e.g.* from H37Rv), sequence analysis, or by using agarose gel.

Using identical electrophoretic conditions (*i.e.* essentially the same polymer and run temperature), we have observed that differences in size offsets are minimal for most loci between different sequencers of a same or comparable type (*i.e.* between 3100s, or between a 3100 and a 3130 XL). However, more important differences are seen with a few loci. **Therefore, size offsets must be verified experimentally**.

1. Select reference PCR products to cover the allelic range in the different loci, in order to determine if offsets increase upon repeat numbers. Alternatively, using stutter peaks from large alleles can to reduce the number of analyses.
2. Using the allele calling information in the Genotyper templates, compare the sizes observed by capillary electrophoresis with those predicted from sequence analysis, or from alleles carefully pre-determined by electrophoresis using agarose gels.
3. In Genotyper, click the Category Window button.
4. If the size offset is fixed upon repeat numbers, select all alleles/categories in the locus of interest, then select **Category** from the main menu, and **Offset categories**.
5. If the size offsets differ upon repeat numbers, select allele/categories separately.
6. Introduce the corresponding offsets.
7. When all offsets have been introduced, save the new template with a different name. These new templates can now be used for MIRU-VNTR analysis on your sequencer.

3 Bibliographic References

1. **Keim, P., A. M. Klevytska, L. B. Price, J. M. Schupp, G. Zinser, K. L. Smith, M. E. Hugh-Jones, R. Okinaka, K. K. Hill, and P. J. Jackson.** 1999. Molecular diversity in *Bacillus anthracis*. J. Appl. Microbiol. **87:**215-217.

2. **Le Fleche, P., Y. Hauck, L. Onteniente, A. Prieur, F. Denoeud, V. Ramisse, P. Sylvestre, G. Benson, F. Ramisse, and G. Vergnaud.** 2001. A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. BMC Microbiol. **1:**2.

3. **Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim.** 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. J. Clin. Microbiol. **39:**3179-3185.

4. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in *the Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144:**1189-1196.

5. **Mazars, E., S. Lesjean, A. L. Banuls, M. Gilbert, V. Vincent, B. Gicquel, M. Tibayrenc, C. Locht, and P. Supply.** 2001. High-resolution minisatellite-based typing as a portable approach to global analysis of *Mycobacterium tuberculosis* molecular epidemiology. Proc. Natl. Acad. Sci. U S A **98:**1901-1906.

6. **Goyal, M., D. Young, Y. Zhang, P. A. Jenkins, and R. J. Shaw** 1994. PCR amplification of variable sequence upstream of *katG* gene to subdivide strains of *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **32:**3070-3071.

7. **Magdalena, J., P. Supply, and C. Locht** 1998a. Specific differentiation between *Mycobacterium bovis* BCG and virulent strains of the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. **36:**2471-2476.

8. **Magdalena, J., A. Vachee, P. Supply, and C. Locht** 1998b. Identification of a new DNA region specific for members of Mycobacterium tuberculosis complex. J. Clin. Microbiol. **36:**937-943.

9. **Namwat, W., P. Luangsuk, and P. Palittapongarnpim** 1998. The genetic diversity of *Mycobacterium tuberculosis* strains in Thailand studied by amplification of DNA segments containing direct repetitive sequences. Int. J. Tuberc. Lung Dis. **2:**153-159.

10. **Kremer, K., D. van Soolingen, R. Frothingham, W. H. Haas, P. W. Hermans, C. Martin, P. Palittapongarnpim, B. B. Plikaytis, L. W. Riley, M. A. Yakrus, J. M. Musser, and J. D. van Embden.** 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. **37:**2607-2618.

11. **Roring, S., A. Scott, D. Brittain, I. Walker, G. Hewinson, S. Neill, and R. Skuce.** 2002. Development of variable-number tandem repeat typing of *Mycobacterium bovis*: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J. Clin. Microbiol. **40:**2126-2133.

12. **Skuce, R. A., T. P. McCorry, J. F. McCarroll, S. M. Roring, A. N. Scott, D. Brittain, S. L. Hughes, R. G. Hewinson, and S. D. Neill.** 2002. Discrimination of *Mycobacterium tuberculosis* complex bacteria using novel VNTR-PCR targets. Microbiology. **148:**519-528.

13. **Smittipat, N., and P. Palittapongarnpim.** 2000. Identification of possible loci of variable number of tandem repeats in *Mycobacterium tuberculosis*. Tuber. Lung Dis. **80:**69-74.

14. **Supply, P., E. Mazars, S. Lesjean, V. Vincent, B. Gicquel, and C. Locht.** 2000. Variable human minisatellite-like regions in the *Mycobacterium tuberculosis* genome. Mol. Microbiol. **36:**762-771.

15. **Le Fleche, P., M. Fabre, F. Denoeud, J. L. Koeck and G. Vergnaud.** 2002. High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. BMC Microbiol. **2**:37.

16. **Supply, P., S. Lesjean, E. Savine, K. Kremer, D. van Soolingen, and C. Locht.** 2001. Automated High-Throughput Genotyping for Study of Global Epidemiology of *Mycobacterium tuberculosis* Based on Mycobacterial Interspersed Repetitive Units. J. Clin. Microbiol. **39:**3563-3571.

17. **Supply, P., J. Magdalena, S. Himpens, and C. Locht.** 1997. Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol. Microbiol. **26:**991-1003.

18. **Steinlein Cowan, L., L. Mosher, L. Diem, J.P. Massey, and J.T. Crawford.** 2002. Variable-number tandem repeat typing of *Mycobacterium tuberculosis* isolates with low-copy numbers of IS*6110* by using mycobacterial interspersed repetitive units. J. Clin. Microbiol. **40:**1592-1602.

19. **Cowan L.S., L. Diem, T. Monson, P. Wand, D. Temporado, T.V. Oemig, and J.T. Crawford.** 2005. Evaluation of a two-step approach for large-scale, prospective genotyping of *Mycobacterium tuberculosis* isolates in the United States. J. Clin. Microbiol. **43:**688-695.

20. **Allix, C., P. Supply, and M. Fauville-Dufaux.** 2004. Usefulness of fast mycobacterial interspersed repetitive unit-variable-number tandem repeat typing in clinical mycobacteriology: A study of three cases. Clin. Infect. Dis. **39:**783-789

21. **Hawkey P.M., E.G. Smith, J.T. Evans, *et al.*** Mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* compared to IS*6110*-based restriction fragment length polymorphism analysis for investigation of apparently clustered cases of tuberculosis. 2003. J. Clin. Microbiol. **41:**3514-3520

22. **van Soolingen, D., P. E. de Haas, P. W. Hermans, and J. D. van Embden.** 1994. DNA fingerprinting of *Mycobacterium tuberculosis*. Methods Enzymol. **235:**196-205.

4 Appendixes

4.1 Conventional and alias designations of MIRU-VNTR loci

MIRU-VNTR convention[a]	Alias 1	Alias 2
154	MIRU 02	
424	VNTR 42	
577	VNTR 43	ETRC
580	MIRU 04	ETRD
802	MIRU 40	
960	MIRU 10	
1644	MIRU 16	
1955		
2059	MIRU 20	
2163b		QUB-11b
2165	ETRA	
2347	VNTR 46	
2401	VNTR 47	
2461	VNTR 48	ETRB
2531	MIRU 23	
2687	MIRU 24	
2996	MIRU 26	
3007	MIRU 27	QUB-5
3171	VNTR 49	
3192	MIRU 31	ETRE
3690	VNTR 52	
4052		QUB-26
4156	VNTR 53	QUB-4156c
4348	MIRU 39	

[a]MIRU-VNTR loci are listed according to their position in kbp on the H37Rv genome

4.2 PCR primer sequences

Multiplex	Locus	Alias	Repeat unit length, bp[1]	PCR primer pairs (5' to 3', with labeling indicated[c])
Mix 1	580	MIRU 4	77	GCGCGAGAGCCCGAACTGC (FAM) GCGCAGCAGAAACGCCAGC
	2996	MIRU 26	51	TAGGTCTACCGTCGAAATCTGTGAC CATAGGCGACCAGGCGAATAG (VIC)
	802	MIRU 40	54	GGGTTGCTGGATGACAACGTGT (NED) GGGTGATCTCGGCGAAATCAGATA
Mix 2	960	MIRU 10	53	GTTCTTGACCAACTGCAGTCGTCC GCCACCTTGGTGATCAGCTACCT (FAM)
	1644	MIRU 16	53	TCGGTGATCGGGTCCAGTCCAAGTA CCCGTCGTGCAGCCCTGGTAC (VIC)
	3192	MIRU 31	53	ACTGATTGGCTTCATACGGCTTTA GTGCCGACGTGGTCTTGAT (NED)
Mix 3	424	42	51	CTTGGCCGGCATCAAGCGCATTATT GGCAGCAGAGCCCGGGATTCTTC (FAM)
	577	43	58	CGAGAGTGGCAGTGGCGGTTATCT (VIC) AATGACTTGAACGCGCAAATTGTGA
	2165	ETR A	75	AAATCGGTCCCATCACCTTCTTAT (NED) CGAAGCCTGGGGTGCCCGCGATTT
Mix 4	2401	47	58	CTTGAAGCCCCGGTCTCATCTGT (FAM) ACTTGAACCCCCACGCCCATTAGTA
	3690	52	58	CGGTGGAGGCGATGAACGTCTTC (VIC) TAGAGCGGCACGGGGGAAAGCTTAG
	4156	53	59	TGACCACGGATTGCTCTAGT GCCGGCGTCCATGTT (NED)
Mix 5	2163b	QUB-11b	69	CGTAAGGGGGATGCGGGAAATAGG CGAAGTGAATGGTGGCAT (FAM)
	1955		57	AGATCCCAGTTGTCGTCGTC (VIC) CAACATCGCCTGGTTCTGTA
	4052	QUB-26	111	AACGCTCAGCTGTCGGAT (NED) CGGCCGTGCCGGCCAGGTCCTTCCCGAT
Mix 6	154	MIRU 2	53	TGGACTTGCAGCAATGGACCAACT TACTCGGACGCCGGCTCAAAAT (FAM)
	2531	MIRU 23	53	CTGTCGATGGCCGCAACAAAACG (VIC) AGCTCAACGGGTTCGCCCTTTTGTC
	4348	MIRU 39	53	CGCATCGACAAACTGGAGCCAAAC CGGAAACGTCTACGCCCCACACAT (NED)
Mix 7	2059	MIRU 20	77	TCGGAGAGATGCCCTTCGAGTTAG (FAM) GGAGACCGCGACCAGGTACTTGTA
	2687	MIRU 24	54	CGACCAAGATGTGCAGGAATACAT GGGCGAGTTGAGCTCACAGAA (VIC)
	3007	MIRU 27	53	TCGAAAGCCTCTGCGTGCCAGTAA GCGATGTGAGCGTGCCACTCAA (NED)
Mix 8	2347	46	57	GCCAGCCGCCGTGCATAAACCT (FAM) AGCCACCCGGTGTGCCTTGTATGAC
	2461	48	57	ATGGCCACCCGATACCGCTTCAGT (VIC) CGACGGGCCATCTTGGATCAGCTAC
	3171	49	54	GGTGCGCACCTGCTCCAGATAA (NED) GGCTCTCATTGCTGGAGGGTTGTAC

[a] Based on the allelic diversity of the MIRU loci observed in *M. tuberculosis*

[b] locus 4 in clinical isolates contains an additional invariable MIRU of 53 bp in terminal position of the repeat array. Locus 4052 (QUB-26) may display limited variations in repeat unit length

[c] Compared to Supply et al. 2001, *J. Clin. Microbiol.*, 39, 3563-3571, HEX labeling has been replaced by VIC labeling

4.3 Reagent and material references

Reagent/Material	Manufacturer	Reference	Packaging
ART 10 Pipet Tips	ART	2139	10 trays
ART 200 Pipet Tips	ART	2069	10 trays
ART 1000 PipetTips	ART	2079E	8 trays
Thermo-Fast 96, non-skirted plates	ABGENE	AB-0600	25 plates
Adhesive PCR Film	ABGENE	AB-0558	100 sheets
Hot Star Taq Polymerase	QIAGEN	203205	1000U
DNTP set	ROCHE	1969064	4x25 µMol
Unlabeled Oligonucleotides	Eurogentec	OliGold	20 nmol
FAM-, VIC-, NED-labeled oligonucleotides (lyophilized)	ABI	450007	10 nmol
Mapmarker1000-Rox	EUROGENTEC	MW-0195-80Rox	400 µl
PCR 20 and 100 bp low ladder set	SIGMA	D-7808	25 µg each
FG,Hi-Di Formamide	ABI	4311320	25 ml
3100 POP-4	ABI	4316355	7 ml
Semi-Skirted PCR Plates	SORENSON	35800	25 plates/ pack

4.4 Allele Calling table (top) and H37Rv/Ra genotype (using the 24 standard MIRU-VNTR loci.

Note: predicted allelic sizes below are those observed using agarose gel electrophoresis. **Offsets MUST be applied to these sizes when using capillary electrophoresis (see text).**

Locus	MIRU 02	MIRU 04	MIRU 10	MIRU 16	MIRU 20	MIRU 23	MIRU 24	MIRU 26	MIRU 27	MIRU 31	MIRU 39	MIRU 40	VNTR 42	VNTR 43	VNTR 1955	VNTR 2163b	ETRA	VNTR 2347	VNTR 2401	VNTR 2461	VNTR 3171	VNTR 3690	QUB-26	QUB-4052	VNTR 4156
Convention	154	580	960	1644	2059	2531	2687	2996	3007	3192	4348	802	424	577	1955	2163	2165	2347	2401	2461	3171	3690	4052	4156	
H37Rv-Ra genotype	2	3'	3	2	2	6	1	3	3	3	2	1	2	4	2	5	3	4	2	3	3	5	5	2	

Allele	MIRU 04
1'	112
1	199
2	274
3'	333

4.5 Alternative PCR conditions for Mix 5

The table below includes alternative PCR conditions that can reduce stutter peaks observed with large alleles of locus 4052, especially in multiplex PCR. In this case, the Multiplex PCR kit (Qiagen, Hilden, Germany) is used with DMSO, and NOT the Hotstart Qiagen kit with Q solution.

Volumes (µl) for the multiplex premix of Mix 5

mix	5
Loci	2163b- 1955-4052
H2O	4,6
Multiplex PCR Master Mix 2 X	10
DMSO	1
Primers EACH[a]	0,4
Total premix	18

[a] Six in total, *i.e.* one forward and one reverse primer for each of the 3 pairs. Initial concentration of 14 pmol/µl for unlabeled and labeled primers for locus 2163b; 4 pmol/µl for unlabled and labeled primers of locus 1955; 50 pmol/µl for unlabled and labeled primers of locus 4052. See appendix 2 for primer sequence and labeling.