A Fast, Accurate, and Automated Workflow for Multi Locus Sequence Typing of Bacterial Isolates

Using Applied Biosystems 3130 and 3730 Series Capillary Electrophoresis Systems and SeqScape® Software

Multi Locus Sequence Typing (MLST) is a nucleotide sequence-based approach for the unambiguous characterization and sub-speciation of bacteria isolates and other organisms. The technique identifies alleles by direct DNA sequencing of fragments of housekeeping genes from known microorganisms. It is much more precise than indirect methods, which discriminate the electrophoretic mobility of large DNA fragments of gene products.

Introduction

Current methods of bacterial and fungal strain typing, such as pulsed-field gel electrophoresis and amplified fragment length polymorphism, are time-consuming and require extensive inter-laboratory comparison of results and coordination of protocols. For these reasons, DNA sequence-based approaches are becoming more popular for bacterial and fungal strain typing. Also, standardized sequencing methods yield accurate high-sensitivity sequence results as compared to current fragment based methods. Moreover, sequence data can be easily shared between laboratories via the Internet, and compared to well-established libraries. Among these DNA sequence-based techniques is Multi Locus Sequence Typing (MLST), which has been developed by a network of researchers and identifies alleles from the DNA sequences of several housekeeping genes.[1] The MLST technique has been used successfully for molecular epidemiology, examination of population structures, and evolution of various virulent bacterial species.[2] MLST is performed after identifying and classifying unidentified bacterial and fungal organisms at the species level.[3]

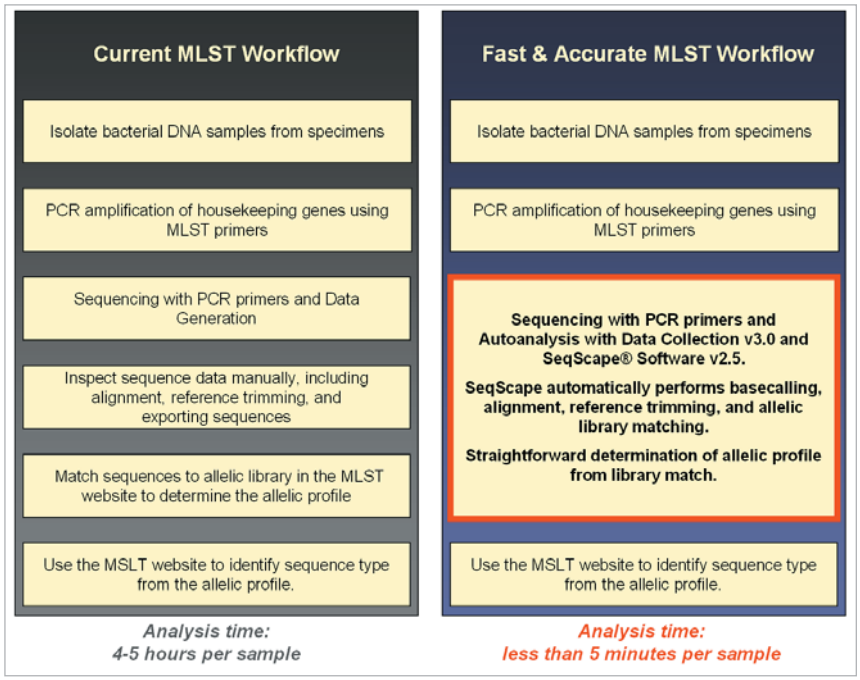

Figure 1. Comparison of current MLST workflow and proposed fast and automated workflow. Left panel shows the current MLST workflow. The data analysis requires a significant amount of post analysis data retrieval and processing which takes about 4–5 hours per sample. Right panel shows the proposed fast and accurate workflow using SeqScape Software to automatically perform basecalling, alignment, reference trimming, and allelic library matching. With minimal sequence checking, the ST of each sample can be obtained in five minutes or less (M. Langvik).

To demonstrate this technique, we typed several DNA samples from *Staphylococcus aureus* using Applied Biosystems BigDye® Terminator chemistries, Applied Biosystems Capillary Electrophoresis (CE)

sytems and SeqScape® Software v2.5 for automated data analysis. *Staphylococcus aureus,* a major pathogen, is associated with serious community- and hospital-acquired diseases. It causes skin and tissue infections, pneumonia, septicemia, and device-associated infections.[4]

MLST Workflow

The MLST technique characterizes isolates of bacterial species using internal fragment sequences from several housekeeping genes. The fragments, which are approximately 450–500 bp of each gene, can be accurately sequenced on both strands using BigDye® Terminators v1.1 or v3.1 chemistries on Applied Biosystems Capillary Electrophoresis systems. For each housekeeping gene, the various sequences present within a bacterial species are specified as distinct alleles. For each isolate, the alleles at all loci define the allelic profile or sequence type (ST).

Current MLST methods do not offer a fast, accurate, and automated workflow for determining the allelic profiles and STs of bacteria and other microorganisms (Figure 1). Instead, they often require 4–5 hours of analysis to determine the allelic profile of just one bacterial sample. This Application Note will provide an example of *S. aureus* typing with Applied Biosystems products that significantly reduces the time required for this procedure.

Sequence libraries of known STs can easily be imported into SeqScape software for automatic allele matching. Importing the libraries eliminates the need to manually upload individual sample sequences into the MLST database to determine specific isolates of a population. This integrated workflow significantly reduces the time required to analyze and type unknown microbial strains. Data analysis in the current MLST workflow requires 4–5 hours of post-analysis data retrieval and processing of each sample. By contrast, the fast and accurate workflow uses SeqScape software for automatic basecalling, alignment, reference trimming, and allelic library matching. With minimal sequence checking, the ST of each sample can be obtained in five minutes or less. This workflow enables researchers to use MLST for fast, accurate sequence typing of a wide variety of microorganisms.

Amplification of Seven Housekeeping Genes of *S. aureus*

DNA samples with unknown STs from *S. aureus* bacterial isolates were supplied by M. Langvik, Akershus University Hospital, Loerenskog, Norway. The following seven housekeeping genes were used in the MLST experiments: carbamate kinase *(arcC),* shikimate dehydrogenase *(aroE),* glycerol kinase *(glpF),* guanylate kinase *(gmk),* phosphate acetyltransferase *(pta),* triosephosphate isomerase *(tpi),* and acetyl coenzyme A acetyltransferase *(yqiL),* as specified by the MLST website.

The genes were amplified using two sets of primers (Figure 2). One set was derived from the MLST website; the second set includes the same MLST primers, tailed with -21 M13 forward or M13 reverse primers. The use of tailed primers further streamlines and simplifies the sequencing reactions, allowing for universal sequencing reactions.

Sequencing with Applied Biosystems Genetic Analyzer and BigDye Terminator v1.1

MLST PCR products can be sequenced with Applied Biosystems 3130 Series Genetic Analyzers or Applied Biosystems 3730 Series DNA Analyzers, together with either the BigDye Terminators Ready Reaction Cycle Sequencing Kits v1.1 or v3.1. Both kits offer improved peak-height uniformity for accurate heterozygous basecalling, and enhanced robustness for sequencing a wide range of template types and qualities.

To compare the two sets of primers (A and B in Figure 2), two sequencing reactions were performed with BigDye® Terminator v1.1, using standard cycle

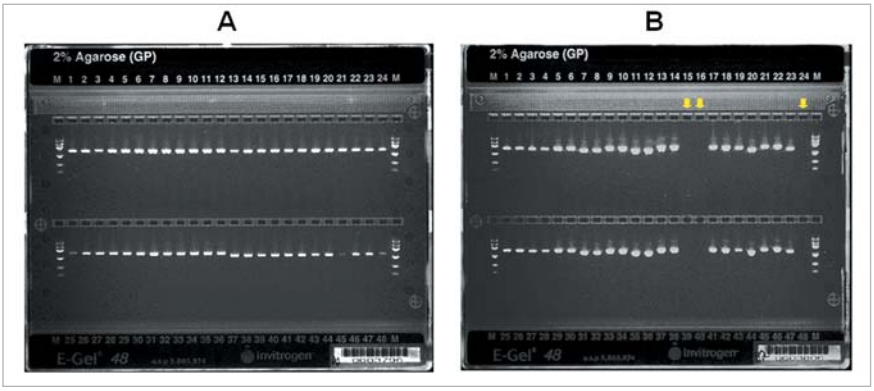

Figure 2. Quantitation of PCR Products using MLST Primers (A) and M13-tailed MLST Primers (B). PCR was conducted with 10-µL reaction volumes containing 22 ng of chromosomal DNA, 4 pmol of each primer, 0.4 µL of deionized water, 1.6 µL of 50% glycerol, and 5 µL of 2X AmpliTaq Gold® PCR Master Mix. The PCR was performed using an Applied Biosystems System 9700 thermocycler with an initial 5-min denaturation at 96°C, followed by 40 cycles of annealing at 60°C for 45 sec, extension at 72°C for 45 sec, and denaturation at 94°C for 30 sec, followed by a final extension step of 72°C for 10 min. Amplicons were purified using 2 µL of ExoSAP-IT® (USB), digested at 37°C for 30 min, and inactivated at 80°C for 15 min. The PCR products were run in a 2% agarose gel (E-gel-48 from Invitrogen) to determine the quantity of the amplicons. Negative controls are shown with yellow arrows.

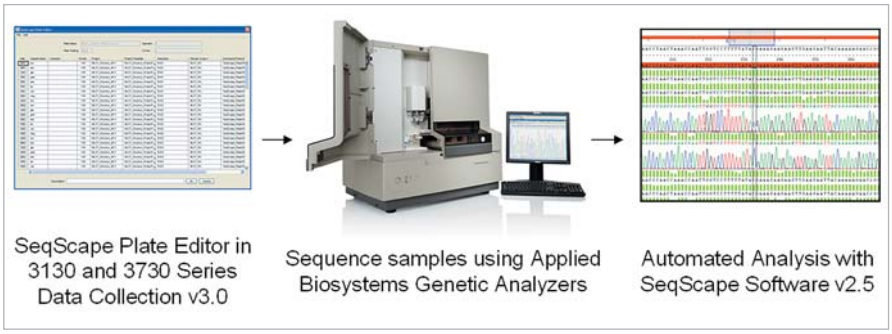

Figure 3. The SeqScape software Plate Editor in 3130*xl* Data Collection Software v3.0, shows sample integration with the SeqScape software project, project template, and specimen. This feature allows automated analysis for all samples after electrophoresis is complete. Analysis information can be viewed in SeqScape Software v2.5.

sequencing methods. Target-specific primers (MLST primers) were used for set A, while the universal -21 M13-forward and M13-reverse primers were used for set B. The M13-tailed MLST primers were used instead of MLST primers for the sequencing reaction, because they provide more uniform results and significantly streamline the sequencing workflow. Completed reactions were purified using ethanol/EDTA precipitation.

In the experiment described here, the sequences of both strands were determined by running the Applied Biosystems 3130*xl* Genetic Analyzer using a 36-cm capillary array, 3130 POP-7™ Polymer, and the RapidSeq36_POP7 run module. Automated data analysis was performed using 3130*xl* Data Collection Software v3.0 and SeqScape® Software v2.5.

Automated Data Analysis with SeqScape Software

Data Collection Software v3.0 for both the 3130 and 3730 Series Systems provides seamless integration between the instrument and analysis application SeqScape Software v2.5. It also automates sample loading, generates, sequencing data, and performs basecalling, trimming, assembly, and alignment of data to the reference. The convenient autoanalysis function in SeqScape software, which streamlines the whole process, can be configured in the following three easy steps:

1. A SeqScape software plate record is created in Data Collection software
2. The plate is then run on the instrument
3. After the plate is complete and all the data has been collected, the software automatically performs the analysis, generating immediate results (Figure 3)

This process, which is simple and straightforward, provides fast, automated, and accurate data analysis.

SeqScape Software v2.5

SeqScape software is a comprehensive tool that provides SNP detection and the identification of sequences that most closely match a target sequence. Analysis algorithms are designed and calibrated with Applied Biosystems instrument specifications and reagent characteristics to provide accurate basecalling with quality values for each mutation. SeqScape software allows researchers to compare sample data against a library of allelic sequences to identify the best matches.

The Reference Data Group (RDG), a feature in SeqScape software, offers powerful capabilities to configure a reference sequence. One reference sequence is constructed in the RDG with segments for each of the seven housekeeping genes. A sequence library is associated with each gene segment in the RDG. The reference sequence and the associated allelic libraries for each gene can be downloaded from the MLST website in a FASTA format and are easily imported into the Library Editor of SeqScape software. Each sequence in the library is labeled with the gene's allelic profile number for easy identification (Figure 4).

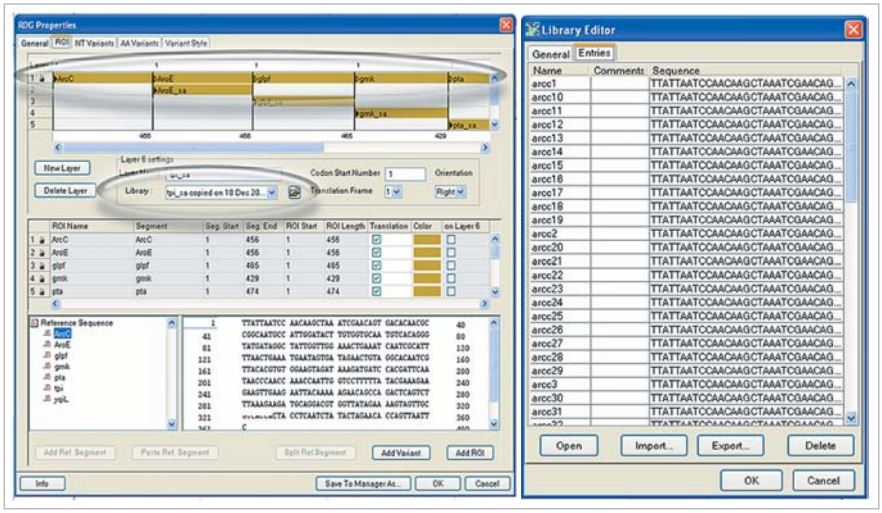

Figure 4. Left panel: Researchers can set up a reference sequence and associate an allelic library for each gene using the RDG feature in SeqScape software. Right panel: Importing a multi-aligned FASTA file of allelic sequences into the SeqScape Library Editor creates an allelic library for each gene.

Figure 5. DNA samples were analyzed in the Applied Biosystems 3130*xl* Genetic Analyzer. A sample sequence trace for the EG02 sample *arcC* in SeqScape Software v2.5 is shown. Both the forward and reverse sequences are used to derive a consensus sequence for EG02. During analysis, the sample sequences are compared to both the reference sequence and the allelic libraries in the RDG. The results for the library match are tabulated on the bottom portion of the Project View. For this example, the *arcC* amplicon for EG02 matches the number 1 allele, derived from the allelic library in the MLST website.

Analysis with SeqScape Software v2.5

The autoanalysis feature in SeqScape software generates and saves data as a project. This project contains numerous individual specimens based on the information provided in the SeqScape software Plate Editor within Data Collection software. Typically, each of these specimens contains sequence files in the forward and reverse sequence that are aligned to each housekeeping gene. The Project View option provides an easy way of identifying the allelic library match for each gene of every specimen consensus (Figure 5). Additionally, the Library Search Report tabulates the allele number that matches the sequence for each DNA sample.

Sequence Type Determination from the MLST Website

An allelic profile of each specimen or bacterial isolate can be determined by collecting the library match for each of the seven housekeeping genes. This set of numbers can be queried against the database in the *S. aureus* MLST website under the Allelic Query option. DNA samples that yield the same allelic profile will not need further determination. For our experiments, we were able to determine the STs for eight samples, as shown in Table 1.

Table 1. Sequence Types of Eight Unknown *S. aureus* DNA Samples

Sample	Housekeeping Genes Allele Number							Strain Type
	arcC	*aroE*	*glpF*	*gmk*	*pta*	*tpi*	*yqiL*	
EG02	1	3	1	14	11	51	10	**80**
EG05	1	3	1	14	11	51	10	**80**
EG06	2	3	1	1	4	4	3	**239**
EG07	1	3	1	14	11	51	10	**80**
EG08	1	3	1	14	11	51	10	**80**
EG09	1	4	1	4	12	1	10	**5**
EG10	3	3	1	1	4	4	3	**8**
EG11	22	1	14	23	12	4	31	**88**

Conclusion

MLST is an effective and simple method for bacterial and fungal strain typing. This technique can be performed faster, easier and better with our Applied Biosystems Genetic Analyzers, BigDye® Terminator Cycle Sequencing Kits v1.1 and 3.1, and SeqScape Software v2.5. The fast, accurate, and automated workflow allows researchers to determine the allelic profiles of unknown samples easily and confidently, and to discover new alleles for possible new sequence types.

Acknowledgements

We would like to thank *Cand. scient.* Magnus Langvik from the Epi-Gen Institute of Clinical Epidemiology and Molecular Biology, Akershus University Hospital, Loerenskog, Norway, for providing the *S. aureus* chromosomal DNA samples that were used in this experiment, and for his advice and assistance in the creation of this Application Note.

References

[1] http://www.mlst.net/. MLST was developed in the laboratories of Martin Maiden, Dominique Caugant, Ian Feavers, Mark Achtman, and Brian Spratt. MLST has been developed for *Staphylococcus aureus* by Mark Enright (currently at the University of Bath) in the laboratory of Professor Brian Spratt, Imperial College London, in collaboration with the laboratories of Drs. Nick Day and Sharon Peacock at the John Radcliffe Hospital, Oxford.

[2] Wieger L. Homan, David Tribe, Simone Poznanski, Mei Li, Geoff Hogg, Emile Spalburg, Jan D. A. van Embden, and Rob J. L. Willems. Multi Locus Sequence Typing Scheme for Enterococcus faecium. *J. Clin. Micro.* **40:** 1963–1971.

[3] Identification and classification of microorganisms can be achieved by comparing them to a validated microbial library. Applied Biosystems MicroSeq® ID Analysis Software v1.0 and associated kits can be used as a tool for identification of bacterial and fungal organisms. For more information, please visit www.microseq.com.

[4] Enright, M.C., Day, N.P.J., Davies, C.E., Sharon J.P, and Spratt, B.G. 2000. Multi Locus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin-Susceptible Clones of *Staphylococcus aureus. J. Clin. Micro.* **38:** 1008–1015.

Printed in the USA, 8/2005, LD Publication 106AP20-01

Applied Biosystems

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.345.5224
www.appliedbiosystems.com

International Sales
For our office locations, please call the division headquarters or refer to our Web site at
www.appliedbiosystems.com/about/offices.cfm