charles river

Summary
It is now possible to resolve some of the most difficult organisms commonly observed in manufacturing industries by combining standard genotypic identification methods with multilocus sequence typing (MLST) or single-locus sequence typing (SLST).

MICROBIAL SOLUTIONS

Highly Accurate Strain Analysis Through Sequence Typing

The tracking and trending of microorganisms and investigations of excursions in manufacturing settings are important components of a comprehensive EM program. Likewise, accurate confirmation of identity of master cell banks, dietary ingredients, and supplements is critical. Standard genotypic sequence-based identification systems, utilizing regions of the ribosomal DNA (rDNA), increase the ability to accurately identify and track and trend microorganisms at the species level by using the sequence data rather than just the organism's name. However, some common contaminants isolated from both sterile and non-sterile manufacturing facilities cannot be resolved by this approach alone and some cell banks, ingredients, and supplements require a more in-depth characterization for verification of the strain. By combining standard genotypic identification methods with multilocus sequence typing (MLST) or single-locus sequence typing (SLST), it is now possible to resolve some of the most difficult organisms commonly observed in manufacturing industries for trending and tracking purposes.

MLST and SLST are well-established, highly accurate sequence-based methods that are used to distinguish closely-related microorganisms to the strain-level. Strain-level differentiation is made by analyzing protein coding or housekeeping genes that encode for proteins necessary for the normal cellular functions and which contain more sequence variability. Since more variation can be detected by analyzing the differences in gene sequences and not just the protein product, it results in a greater differentiation of subtypes within a population of organisms. The analysis of multiple loci provides additional variation to further differentiate between closely related strains.

Confirmation of species identity for all isolates is first determined by standard rDNA sequencing (16S or ITS2). Accurate identification is essential, as the target genes for sequence typing are different for each species. SLST methods involve sequencing one gene that is known to harbor variable DNA sequences, and with MLST, multiple genes of moderate variability are sequenced. The gene sequences from each isolate are then aligned and compared in a phylogenetic tree to show the amount of conservation and divergence in the DNA of that gene region. The goal is to determine a gene or gene combinations that can give a high level of variability to differentiate to the strain level.

This genotypic technique, AccuGENX-ST™, is designed to deliver the most informative data and, as a result, is highly discriminatory. Since the foundation of M/SLST is built upon DNA sequencing results, which can be easily cataloged and referenced, these techniques are highly reproducible and unambiguous. Additionally, databases can be created for historical trending and tracking. Given the reproducibility of M/SLST between experiments and over time, these methods can be used more reliably to determine if isolates recovered from one area are the same or different as another isolate – a trait that allows for high-resolution trending and tracking projects.

Case in Point

As an illustration of this technology, we present a case study involving *Micrococcus luteus*, which is a Gram-positive, spherical, saprotrophic bacterium usually isolated from human skin, water, dust, and soil. Due to its ubiquitous nature, *M. luteus* is the most frequently encountered contaminant species in sterile and nonsterile manufacturing environments. *M. luteus* is generally considered a harmless bacterium, though it can be an opportunistic pathogen, particularly for immunocompromised individuals. In addition to being an industrial contaminant, *M. luteus* is also exploited by the chemical and pharmaceutical industries for studying antibacterial activity and for its ability to convert biomass into isoprene, terpene, and long-chain alkenes that are used as raw materials for biofuel production. Its dual role as a production component and potential contaminant makes *M. luteus* an organism of interest for the biotech and pharmaceutical industries, where strain level identification is critical for both validation of stock cultures and as part of EM programs for identifying the source of contamination and developing possible remediation strategies.

Despite its biological significance and biotechnological potential, surprisingly very little is known about the diversity within this species. Furthermore, due to the inadequacy of robust strain typing methods for industrial strains, biotech and pharmaceutical industries still rely on expensive and low discriminatory methods like ribotyping for their strain characterization of *M. luteus*. This fragment-based strain typing method is very limited in regards to capability, resolution, and repeatability. Although molecular subtyping methods, especially MLST, are well-accepted for characterizing and monitoring clinical pathogens, application to industrial strains has been limited to a few yogurt culture strains. A novel MLST approach was developed to characterize *M. luteus* strains isolated from different industrial facilities. In this study, three protein coding genes (*gsp*, *sahH*, *prfA*) were used for MLST and 32 isolates were sequenced. MLST showed distinct patterns for the seven closely related 16S rDNA clusters, which were defined by 31 MLST-sequence types (ST). The technique was validated by comparison with ribotyping, the widely-used strain typing method for *M. luteus*. Sequence-based strain typing revealed that even a single protein coding gene can outperform ribotyping.

Methods

32 *M. luteus* isolates that have 16S rDNA sequence similarities between 99.7-100% were selected for analysis from the Accugenix® collection. Strain characterization for all 32 isolates was also carried out using the RiboPrinter® as per DuPont Qualicon's recommendations. For MLST, an alkaline lysis method was used for DNA extraction and was followed by PCR amplification of three target genes (*gsp*, *sahH*, *prfA*) using gene-specific primers. Sequencing of the three target genes, sequence assembly, and data analysis was performed using the Charles River Accugenix® group's proprietary methods and software. A neighbor-joining (NJ) tree was constructed to discern the relationship among the isolates.

Results

The 16S rDNA region was sequenced and analyzed for 32 isolates. All of the isolates clustered into seven groups showing very limited utility of 16S rDNA for strain-level differentiation (Figure 1). The genetic distance, or percent nucleotide difference, of the isolates from the type strain (T) ranged from 0.0%-0.3%. A concise alignment which shows the exact nucleotide differences at each of four positions in the 16S gene (read vertically, e.g., 156, 179, etc.) between each of the groups of organisms in the NJ tree is presented

in the box. Since the 16S rDNA is a multicopy gene, mixed-base or polymorphic positions are common and are listed with the International Union of Pure and Applied Chemistry, IUPAC, nucleotide code (Y = C or T; R = A or G; M = A or C). Our Accugenix® group's standard genotypic identification method using the first 500 bp of the 16S rDNA was useful for species-level identification; however, it failed to give enough resolution at the strain level. The isolates in this figure are color-coded based on their RiboGroup, as shown in Figure 2.

Ribotyping analysis was done for all 32 isolates, including the type strain (T) studied using DuPont Qualicon's RiboPrinter® (Figure 2). Restriction enzymes *Eco*RI and *Pvu*II were used to restrict DNA, as per protocols from DuPont Qualicon. However, the *Eco*RI restriction enzyme failed to cut *M. luteus* genomic DNA within the ribosomal region detected in this analysis (data not shown) and one isolate failed to cut with *Pvu*II (arrow). Results from the *Pvu*II digests were assigned to three distinct RiboGroups by the commercial software (listed in the middle panel and color coded). For *M. luteus*, ribotyping was unable to adequately discriminate between the strains.

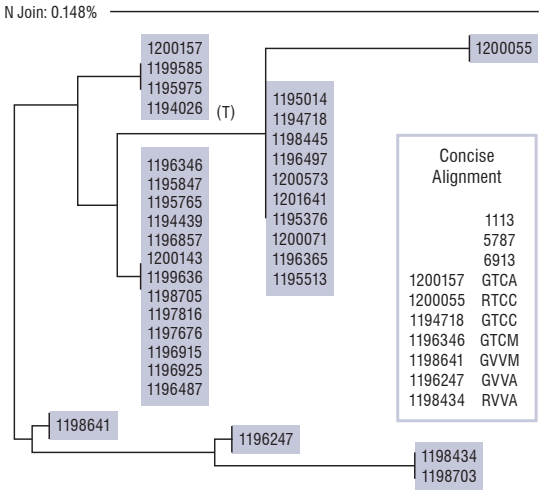

Figure 1. Construction of a neighbor-joining tree of 16S rDNA showing the phylogenetic relatedness among the isolates of M. luteus

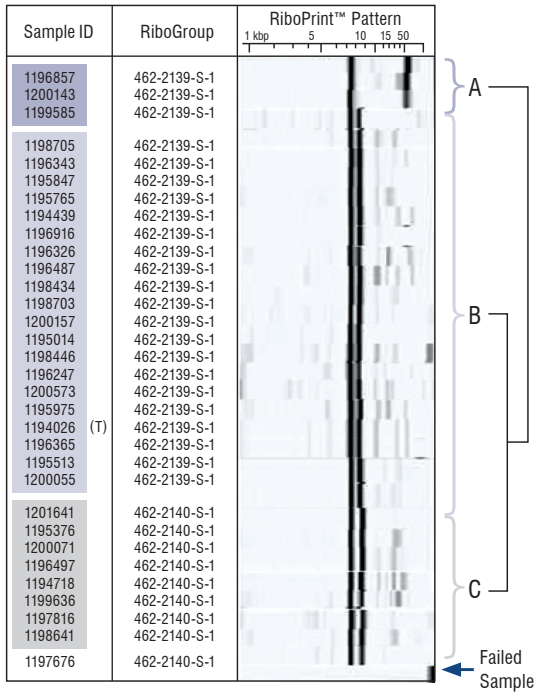

Figure 2. RiboPrinter® analysis of M. luteus *indicates three RiboGroups*

Three target genes were identified by sequence analysis as providing significant information regarding strain-level relatedness. The three loci (*sahH*, *prfA*, and *gsp*) were sequenced and the data analyzed. The number of nucleotides included in the analysis from each of the gene targets ranged from 523 to 575 bp. The highest percentage of polymorphic sites, a base position in the DNA sequence that has different nucleotides, was found with the *sahH* gene (7.13%), followed by *prfA* (4.97%) and *gsp* (3.17%). Likewise, *sahH* had the largest genetic or nucleotide distance, followed by *prfA* and *gsp*. These genetic loci yielded a total of 24, 14, and 19 individual sequence types, respectively, with each gene individually performing better than the RiboPrinter®. The *lysP* and *gyrA* genes were also tested, but were not used in the typing scheme as the *gyrA* gene primers showed secondary structural interference during PCR amplification and the *lysP* gene did not provide any additional information for subtyping.

Gene	No. of Polymorphic Sites	% of Nucleotide Distance	No. of Sequence Variants
sahH	41	3.64%	24
prfA	17	0.89%	14
gsp	26	0.84%	19
77.4 (70.7-87.3)	68.5*† (61.4-75.3)	78.0 (68.9-85.9)	77.0 (70.9-85.4)

All three MLST gene targets were amplified for all isolates. The amplified sequence lengths for MLST target genes, *sahH*, *prfA*, and *gsp* were 575 bp, 536 bp, and 523 bp, respectively. The individual sequences of the *sahH*, *gsp,* and *prfA* genes were concatenated, i.e., placed end to end, the genetic distance was calculated, and an NJ tree assembled. Concatenation of the sequences identified 31 different sequence types (Figure 3), indicating a high level of discrimination with these loci. We were able to conclude with high confidence that all isolates, except 1194439 and 1195765 (shaded green), are unique ST.

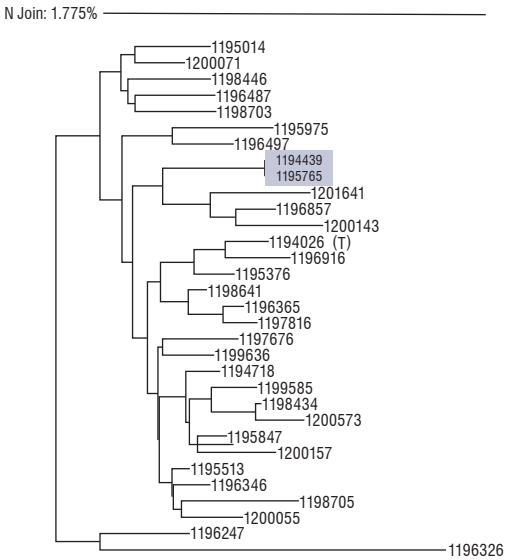

Figure 3. Concatenated neighbor-joining tree of sahH, prfA *and* gsp *gene sequences for* M. luteus *isolates*

Conclusions

The Charles River Accugenix® group has always been committed to providing the most accurate and reliable species level identifications to our clients. Now, we are proud to include sequence-based strain-level typing to our service list with AccuGENX-ST™. In order to perform molecular subtyping, it is first necessary to confirm the species-level identification of an organism with 16S rDNA gene sequencing to correctly choose the target genes for analysis. Additionally, 16S sequencing prior to typing can eliminate many subspecies from further consideration if there are differences in the 16S sequences.

Strain characterization by our cGMP RiboPrinter method is performed by two enzymes, *Eco*RI and *Pvu*II. However, strain typing of *M. luteus* isolates showed very poor resolution with the RiboPrinter®, since only *Pvu*II was capable of restricting *M. luteus* genomic DNA in the ribosomal region and *Eco*RI failed to cut with consistency. Frequent failure of enzymes to cut genomic DNA drastically affects the reliability and resolution and increases the time investment needed to try to obtain an answer, all of which can be critical, especially during investigations. Molecular subtyping of *M. luteus* strains using genes *sahH*, *gsp,* and *prfA* individually all gave better resolution than RiboPrinter®.

Subtyping of *M. luteus* was very promising with all the genes tested, and *sahH* showed the highest nucleotide distance (3.67%), followed by *prfA* (0.89%), and *gsp* (0.83%). Combined information from the three genes showed better subtyping utility, placing all the strains except two, into different sequence types. Thus, generating an MLST method for strain-level analysis for a critical and frequently-observed organism like *M. luteus* provides a powerful and accurate technology to aid manufacturers in demonstrating compliance.