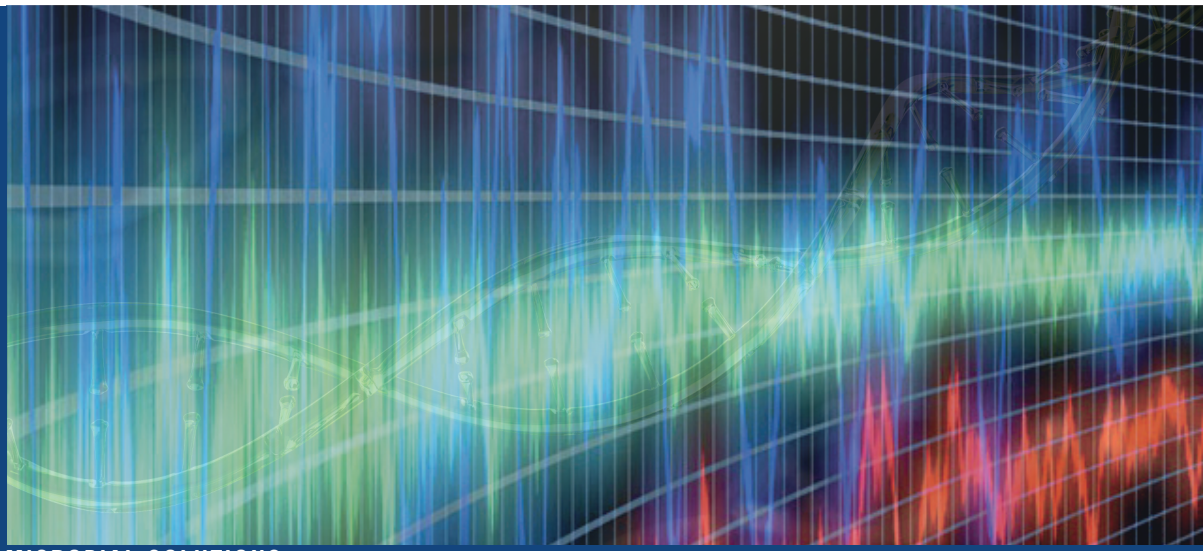
charles river

Summary

Having access to the most relevant, accurate, and compliant microbial libraries is crucial to obtaining correct species-level identifications for pharmaceutical, biotechnology, medical device, and other regulated manufacturing operations.



MICROBIAL SOLUTIONS

Creating and Maintaining Validated Microbial Identification Libraries

The Accugenix® business at Charles River recognized early on that the libraries associated with commercial systems were limited or clinically focused, and thus not sufficient. To minimize errors and allow for expansion of the reference libraries, we undertook the construction of our own validated, proprietary sequence identification libraries in 2004. The bacterial and fungal entries used to build these sequence-based libraries consisted primarily of type strains, when available, as these are the organisms that define each species. Further, with the introduction of our identification service using MALDI-TOF mass spectrometry, we began constructing our own validated, proprietary MALDI library entries in 2010. These MALDI entries consist of type and non-type strains that represent the normal variation and diversity seen within species. Since the initial development, we have been committed to performing maintenance of our libraries while adhering to validated processes. We give utmost priority to accurately reviewing and updating our sequencing and MALDI reference libraries by employing highly-qualified and knowledgeable PhD scientists with expertise in microbial taxonomy, phylogenetics, mass spectrometry, and bioinformatics to manage this process.

The results of this combined expertise are genotypic and proteotypic identification systems that show a high speciation rate, with superior accuracy, for samples obtained from regulated manufacturing industries.

Sources of Bacterial and Fungal Strains and Information

Charles River depends on superior scientific sources to obtain microorganism cultures and reliable taxonomic information for the species that are being identified using our validated Accugenix® bacterial and fungal reference libraries. New organisms are described daily in peer-reviewed scientific journals such as the *International Journal of Systematics and Evolutionary Microbiology* (IJSEM), *Systematic and Applied Microbiology, Journal of Clinical Microbiology, Applied and Environmental Microbiology, Mycologia, Mycology*, etc. When new species are published, the very first strain isolated and characterized is defined as the type (although this designation is not always the case with fungal classification). The “type strain” is the original reference specimen for the species name; it is the permanent example of the species.

Each species is composed of the type strain plus all other strains considered to be sufficiently similar to the type to warrant inclusion in the species. The type strain, and its accompanying information, should be deposited by the original author into at least two global culture collections such as the American Type Culture Collection (ATCC), the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), the English National Collection of Type Cultures (NCTC), or the Korean Collection for Type Cultures (KCTC), for example. Along with the type strain, many other non-type strains are deposited in the collections to be used as reference specimens. These cultures can be purchased from these collections. Additionally, since the 1990s, authors publishing new species are required to deposit type strain sequences in databases such as GenBank®. GenBank is an open-access, annotated collection of all publicly available nucleotide sequences that is maintained by the National Center for Biotechnology Information (NCBI) and a part of the International Nucleotide Sequence Database Collaboration, which also includes the DNA DataBank of Japan (DDBJ) and the European Nucleotide Archive (ENA). These three organizations exchange data on a daily basis. The GenBank accession numbers are routinely specified in the publications that describe novel species. The author is authorized to make changes to the content of the accession file, and a revision code or history is maintained. The GenBank database is revision controlled, but not curated or validated from a scientific or quality assurance perspective. Thus, sequence comparisons made against GenBank may match the originally described type strain for a species, but can also match submissions with incorrect nomenclature, with poor quality sequence data, from cloned DNA, or uncultured organisms, etc. A prime source for published bacterial type strain sequence information is the EzBioCloud™/BIOiPLUG™ genetic sequences database. The EzBioCloud/BIOiPLUG server contains a manually annotated and curated database of 16S ribosomal RNA gene sequences for bacterial type strains with validly described species names and other strains that have been described in the literature. The sequence information is not only from entries present in GenBank, but also from other relevant publications. Because of the time involved in researching the accuracy of sequences, release of updates to this database is a few months behind GenBank. Caution needs to be taken, however, as errors in sequence information that were present in GenBank can be propagated into these curated databases.

The most highly-regarded source for information about new taxa and changes in bacterial and yeast nomenclature is *International Journal of Systematics and Evolutionary Microbiology* (IJSEM). IJSEM is the "official journal of record for novel prokaryotic taxa. It is the official publication of the International Committee on Systematics of Prokaryotes" (ICSP, http://ijs.microbiologyresearch.org/content/journal/ijsem/about). The ICSP is the body that oversees the nomenclature of prokaryotes, determines the rules by which prokaryotes are named, and whose Judicial Commission issues Opinions concerning taxonomic matters and revisions to the Bacteriological Code. As the international standard for nomenclature and publication of new species, IJSEM publishes articles pertaining to all phases of the systematics of bacteria and yeast, and publishes the lists of valid species names twice a month. The classification and nomenclature of fungi is governed by the International Code of Botanical Nomenclature, which was renamed to the International Code of Nomenclature (ICN) for algae, fungi, and plants. Historical descriptions were based on observable morphological features of the multicellular structures. However, the same organism can look different when grown on different media, and the connection between the different forms or morphologies may not have been made. Thus, many fungi have more than one name; there are synonyms. Additionally, the anamorph and the teleomorph forms (sexual stages) of a species were traditionally described or classified with alternate state names. Fortunately, the practice of assigning separate names for anamorphs and teleomorphs was discontinued on January 1, 2013 as dictated in the ICN Congress Melbourne Code of 2011; however, the multitude of names still persists for many fungi. A further complication to fungal nomenclature is that a type strain for many species has not been defined. Recently, there has been a collaborative effort to improve fungal nomenclature, all of which has been published in multiple sources. Because the information is not centrally controlled, the accepted approach is to base

fungal nomenclature on information currently appearing in sites maintained by the International Mycological Association (MycoBank) and the Index Fungorum, for example, and culture collections including Centraalbureau voor Schimmelcultures (CBS).

Library Validation, Generation, and Maintenance

The Charles River validation approach for the Accugenix® services is based on our expertise in the arena of microbial identification. The validation program is designed to be in accordance with the current guidelines for equipment, processes, and computerized systems. We use an integrated team approach to validation that includes expertise from multiple departments. Our quality system includes a validation plan that specifies the studies or tests to use, the criteria appropriate to assess outcomes, the timing of qualification activities, the responsibilities of the relevant departments, and the procedures for documenting and approving the qualification. All validation activities are documented and summarized in a report with conclusions that address criteria in the plan. Quality management reviews and approves the plan and report. Validation of the libraries is one component that assures the quality of the services we offer. In order to confirm that the libraries are maintained in a state of control over the life of the maintenance process, verification activities are also conducted for each release and more extensive validation testing is performed annually. All library generation and maintenance activities are conducted in accordance with Quality Assurance-approved Standard Operating Procedures. A risk-based approach is utilized during the library entry generation process and when building and testing the completed libraries. It is also needed during the maintenance of the libraries. Maintenance updates to identification libraries are needed to keep current with newly described organisms, published taxonomic nomenclature changes, and for correcting errors in the names or supporting sequence or spectral data for entries. The specific activities for validation, entry generation, and maintenance vary by service as the data supporting our sequence and phylogenetic-based identification and MALDI-TOF identification are fundamentally different.

Libraries in Support of Our AccuGENX-ID® Service Offering

Since the original generation, validation, and implementation of our identification libraries, we have had an established maintenance program that enables the incorporation of new entries, or the modification of existing entries, in our databases to reflect the discovery and publication of novel species, and the reclassification of current type strains and species. For the AccuGENX-ID® sequence-based identification service, maintenance can also be initiated by customer samples that have a genus-level, or higher, confidence, but that are closely related to a validly named organism not currently in the library (Figure 1).

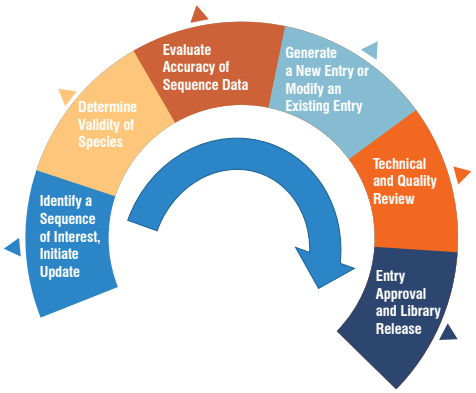


Figure 1. Overview of the Accugenix Sequencing Library Maintenance Process. A library entry is initiated when a need is identified. After determination and evaluation of validity and accuracy, an entry is created or modified. The entry is then reviewed and approved by a scientist and QA. New valid for use entries are incorporated in the next library release.

Sequence Entry Creation and Modification

If it is determined that a new entry is needed either due to a recent taxonomic publication or after reviewing the sequence data for customer samples that have failed to produce species-level identifications due to a lack of matches in the reference database, the maintenance process is initiated. For newly described species, the published and deposited type strain, when available, and the type strain sequence for the organism are part of the official description for the species and are, by definition, correct. For sequences from customer samples that fail to align closely with an existing library entry the consensus sequences are searched against EzBioCloud, or another

public curated database, for a closer match. If a published type strain or reference strain is found, the organism's scientific validity is verified by confirming that the type or reference strain has been published in a peer-reviewed journal and that the name is recognized as scientifically valid. In both situations, the quality of the sequence data associated with the original publication is evaluated. If the sequence quality is low, the organism will be purchased and sequenced in-house. If the quality is high, an entry can be made using the data. However, for organisms that may be frequently isolated from manufacturing environments, the library entry is built by purchasing the type strain from a culture collection and sequencing the organism in-house. For organisms that may not be encountered often, the library entry is built by acquiring the highest quality type strain sequence that can be found in public databases. Once a sequence is associated with a reference strain or type strain, the sequence goes through a comprehensive evaluation. It is tested for validity against at least one other sequence from an independent source, if available. This is most often the original sequence deposited in GenBank, a higher quality sequence from a more recent taxonomic study, or from a genome sequence database. Phylogenetic justification requires that the new species entry clusters with other species in the same genus, if appropriate, at a genetic distance that is consistent with the average genetic distance that separates other species in that genus (Figure 2).

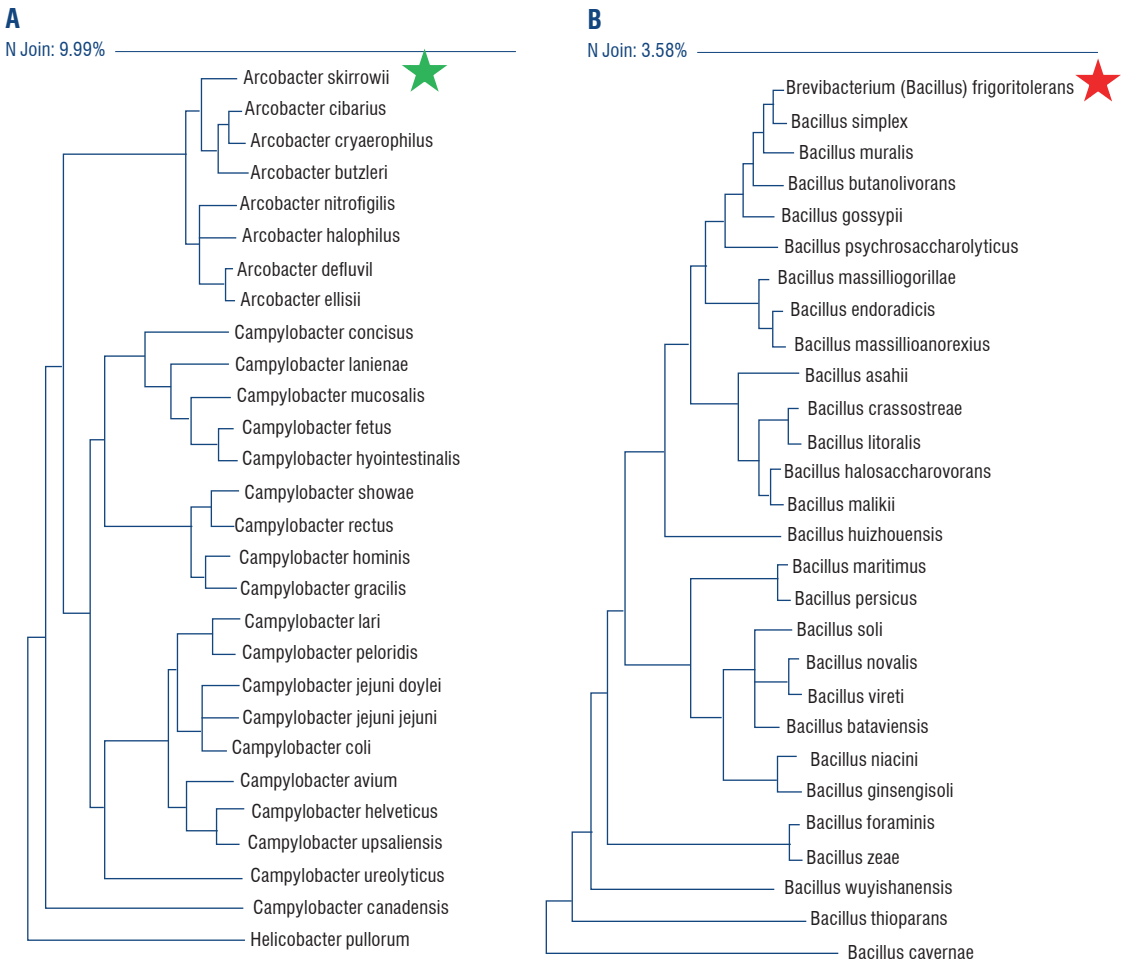
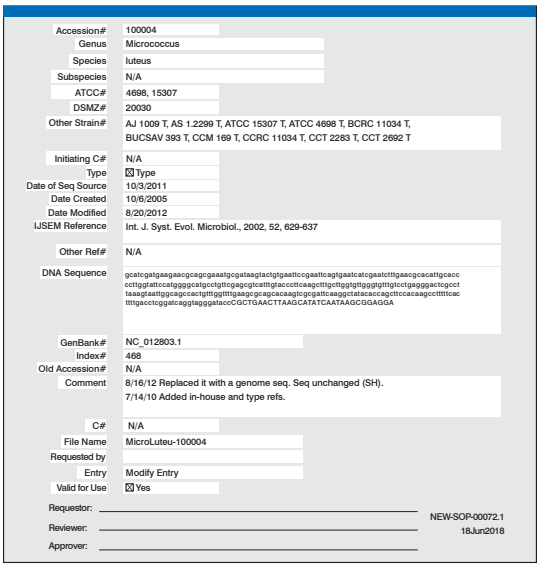


Figure 2. Example Neighbor Joining phylogenetic trees demonstrating analysis of genetic distance acceptance criteria for a library entry. (A) Species *A. skirrowii* (green star) segregates with other species within the genus *Arcobacter* with a similar genetic distance. (B) Analysis of *Brevibacterium frigoritolerans* (red star) indicates that it is most closely related to the genus *Bacillus*, and may have been incorrectly classified, which we denote by the more appropriate genus name in parentheses.

Accugenix® can modify an existing entry due to a reclassification or change in taxonomic relationship when it is published in a peer-reviewed journal, and thus has supporting documentation. A taxonomic reclassification that affects our bacterial or fungal libraries occurs approximately 10 times per month with the current species count of our bacterial and fungal sequencing libraries being > 11,000. We can also modify or inactivate an existing entry due to the detection of misinformation in the entry, changes to sequence information, or due to a retraction published in a peer-reviewed journal. Phylogenetic justification requires any changes to a species entry not affect how it clusters with other species in the same genus, if appropriate, and with a consistent distance separating the species from one another.

Sequence Library Entry Review and Approval

For all entries, new or modified, all the pertinent information is recorded in controlled database entry forms (Figure 3). These forms, and all the supporting scientific, phylogenetic, and publication records for each species entry are compiled. For all purchased type strains sequenced by Accugenix®, library entries are considered acceptable if their sequences are consistent with the taxonomic publication that defines the species. For species defined before sequence technology was used for this application, the sequence generated in-house must be phylogenetically consistent with other related species in the same genus. For all sequences acquired from curated databases, library entries are considered acceptable if the sequence is referenced in the taxonomic publication that defines the species or the sequence is derived from the published type strain or reference strain. In either case, the sequence generated must be phylogenetically consistent with other related species in the same genus. All proposed entries are independently reviewed by trained scientists for technical accuracy. All documentation associated with each proposed entry undergoes final review and approval by Quality Management. The new or modified Quality-approved library entry is then "valid for use" and available for generating identifications via our AccuGENX-ID® service.



Accession#	100004
Genus	Micrococcus
Species	luteus
Subspecies	N/A
ATCC#	4698, 15307
DSMZ#	20030
Other Strain#	AJ 1009 T, AS 1.2299 T, ATCC 15307 T, ATCC 4698 T, BCRC 11034 T, BUCSAV 393 T, CCM 169 T, CCRC 11034 T, CCT 2283 T, CCT 2692 T
Initiating C#	N/A
Type	☒ Type
Date of Seq Source	10/3/2011
Date Created	10/6/2005
Date Modified	8/20/2012
IJSEM Reference	Int. J. Syst. Evol. Microbiol., 2002, 52, 629-637
Other Ref#	N/A
DNA Sequence	gcatcgatgaagaacgcagcgaaatgcgataagtactgtgaattccgaattcagtgaatcatcgaatctttgaacgcacattgcacc cctggtattccatgggcgatgcctgttcgagcgtcatttgtacctttcaagctttgctttggtgtggtttgtcctgagggactcgcct taaagtaattggcagccactgtttggtttgaagccgcagcacagtccgcgattccaaggctatacaccagcttccacaagcctttttcac tttgacctcggatcaggtagggataccCGCTGAACTTAAGCATATCAATAAGCGGAGGA
GenBank#	NC_012803.1
Index#	468
Old Accession#	N/A
Comment	8/16/12 Replaced it with a genome seq. Seq unchanged (SH). 7/14/10 Added in-house and type refs.
C#	N/A
File Name	MicroLuteu-100004
Requested by	
Entry	Modify Entry
Valid for Use	☒ Yes

Requestor: _____

Reviewer: _____ NEW-SOP-00072.1 18Jun2018

Approver: _____

Figure 3. Sequencing Library Entry Form. Electronic database file documenting all pertinent information for the type or reference species that are used for AccuGENX-ID, our sequence-based phylogenetic identification service. Information includes genus and species names, culture collection number(s), type strain status, peer-reviewed citation(s) such as IJSEM or other references, 16S or ITS2 rDNA sequence, GenBank accession number, and any additional comments on the library entry.

Sequence Library Update and Quarterly Maintenance

Updates to the library are routinely generated and include all the approved, valid entries compiled since the last library version was issued, and adhere to strict quality documentation. The completed library is tested for performance to verify incorporation of the approved entries in the latest update and to ensure consistency in the name and sequence information. Quality Management approves all documentation associated with library updates prior to being used to generate identifications for customer samples. In addition to the review performed during updates, we perform quarterly maintenance on the libraries. This test requires that all new or modified sequencing library entries for that quarter are searched against databases, such as EzBioCloud, to verify that each entry is valid and correct. Any differences between the Accugenix® library and the public databases are investigated. We consistently

find a small percentage of results that are different due to errors in GenBank®, from which EzBioCloud obtained their sequences, errors our in-house sequencing has clarified (such as undefined nucleotides designated as N), or due to variances in nomenclature.

Validation and Performance Qualifications of the Sequence Libraries

Sequence library validation consists of testing activities that are designed to ensure that the entries in the libraries can reproducibly produce accurate identifications. The tests are designed to check the capacity of the libraries to produce identifications that are consistent with the identities of the type or reference strains. The sequencing libraries are requalified at least once per year if additions or modifications have been made. The test requires that a proportion of the most frequently identified organisms are tested to verify that the correct species-level identifications are obtained. Consensus sequences from at least 50% of the top 80% most frequently identified organisms are randomly chosen, and identification reports are generated to see if the identity associated with the sequence is as expected. Additionally, a subset of new and modified entries is selected, the strains obtained from recognized culture collections, and processed. Consensus sequences and identification reports are generated to verify that they also produce correct species-level identifications. In addition, as part of annual library maintenance, we challenge a set of library entries to make sure that they are consistent with current nomenclature and taxonomy. For all qualifications, the protocols, raw data, identification reports, and final report are all maintained as part of the official validation package. The validations are approved by Quality Management and are maintained on file for review by our customers during on-site auditing activities.

Libraries in Support of Our AccuPRO-ID® Service Offering and Axcess® System

Since the original generation, validation, and implementation of our MALDI identification library, we have had an established maintenance program that enables the incorporation of new entries, or the modification of existing entries, in our databases. Development of a library in support of our AccuPRO-ID service and Axcess System is fundamentally different than the development of sequencing libraries, and the methods we use are specific to the Bruker Biotyper MALDI-TOF mass spectrometry system. Identifications utilizing the MALDI are made by comparing mass spectral protein patterns. Species show intraspecific variation with respect to mass spectral patterns, thus species may need to be represented by reference spectra from multiple strains to cover the natural diversity within the species. The Bruker Biotyper system uses one strain to make each library entry and multiple entries cover that normal diversity. The number of entries that represent each species varies in the library based on the diversity we encounter. Currently, the Charles River MALDI library contains ~10,000 entries with a proportion of the entries created by Bruker and others by Charles River. There are, on average, three entries per bacterial and yeast species in the library. However, some species that are encountered often and that have a large amount of intraspecific variation have many entries (5-20). There are 36 entries for *Candida albicans*, 40 for *Neisseria gonorrhoeae*, and 39 for *Streptococcus mitis*, for example. Further, many species are represented in the library by only one entry (e.g. *Bacillus novalis, Malikia spinosa, Xenorhabdus szentirmaii*).

MALDI Entry Creation and Modification

New MALDI library entries are made from type and reference strains that we have purchased when we need to increase the strain and species diversity in commonly encountered genera. MALDI library entries are also made from novel and unique strains when they are encountered and have resulted in a “no identification” result as part of our testing services (Figure 4). The Accugenix AccuPRO-ID® service offering includes the identification of samples using our MALDI-TOF system and the Charles River MALDI library. If a sample fails to identify to the species level, that sample will be identified using our AccuGENX-ID® service. In this way, we know the identity of the samples that have not been classified on the MALDI. By evaluating these data on a periodic basis, we can determine what diversity

is needed in our MALDI library and target our efforts on these relevant organisms. Once a MALDI library candidate is recognized, its identification is confirmed by 16S or ITS rDNA sequencing and a phylogenetic analysis. Organisms that have a species-level taxonomic confidence enter the development pipeline. Spectra are generated for the library candidates as per our SOPs. These spectra are evaluated for acceptable quality and are used to build a MALDI library entry, known as an MSP, which stands for main spectrum. Each MSP comprises at least 20 spectra from the candidate isolate, and also has to pass additional quality criteria for robustness and accuracy. Maintenance on existing entries also occurs. If species are reclassified, entries can be modified to reflect the name changes. Entries can also be inactivated or deleted, as needed, if we see evidence that data quality is affecting the accuracy of an identification, or the taxonomic confidence of the isolate used to build the entry is no longer species-level because of a new species publication.

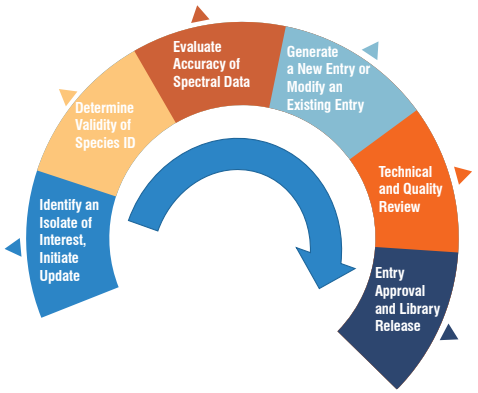
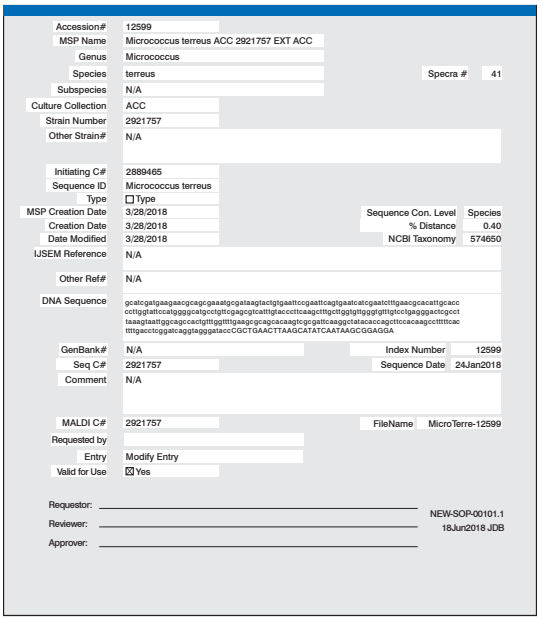


Figure 4. Overview of the Accugenix MALDI Library Maintenance Process. A library entry is initiated when a need is identified. After verification of the identity of the isolate and taxonomic confidence, spectral data are collected and compiled into an MSP. The MSP undergoes quality testing to determine its robustness and accuracy. The entry is then reviewed and approved by a scientist and QA. New valid for use entries are incorporated in the next library release.

MALDI Library Entry Review and Approval

Each new MSP must be tested against the existing library. The new MSP should not match to other MSPs already in the library indicating there are no duplicates. One spectrum acquired from the candidate isolate, which was not used to create the entry, is tested against the library containing the new entry. The test spectrum should match the new MSP. Finally, if there is a spectrum from an original unknown, which resulted in the "no identification" during processing, it is tested against the library containing the new entry. The original spectrum should match the new MSP. For all entries, all the pertinent information is recorded in controlled database entry forms (Figure 5). These forms, and all the supporting scientific, phylogenetic, and test data for each isolate entry are compiled. All proposed entries are independently reviewed by trained scientists for technical accuracy. All documentation associated with each proposed entry undergoes final review for completeness and approval by Quality Management.



Accession#	12599		
MSP Name	Micrococcus terreus ACC 2921757 EXT ACC		
Genus	Micrococcus		
Species	terreus	Specra #	41
Subspecies	N/A		
Culture Collection	ACC		
Strain Number	2921757		
Other Strain#	N/A		
Initiating C#	2889465		
Sequence ID	Micrococcus terreus		
Type	☐ Type		
MSP Creation Date	3/28/2018	Sequence Con. Level	Species
Creation Date	3/28/2018	% Distance	0.40
Date Modified	3/28/2018	NCBI Taxonomy	574650
IJSEM Reference	N/A		
Other Ref#	N/A		
DNA Sequence	gcatcgatgaagaacgcagcgaaatgcgataagtacttgtgaattccgaattcagtgaatcatcgaatctttgaacgcacattgcacc cctggtattccatgggcgatgcctgttcgagcgtcatttgtaccttcaagcttttgcttgggtgtttgtgcctgagggactcgcct taaagtaattggcagccactgtttggtttgagcgcagcacaagtcgcgatccaaggctatacaccagcttccacaagcctttttcac tttgacctcggatcaggtagggatacccGCTGAACTTAAGCATATCAATAAGCGGAGGA		
GenBank#	N/A	Index Number	12599
Seq C#	2921757	Sequence Date	24Jan2018
Comment	N/A		
MALDI C#	2921757	FileName	MicroTerre-12599
Requested by			
Entry	Modify Entry		
Valid for Use	☒ Yes		

Requestor: _____ NEW-SOP-00101.1

Reviewer: _____ 16Jun2018 JDB

Approver: _____

Figure 5. MALDI Library Entry Form. Electronic database file documenting all pertinent information for the strains that are used for AccuPRO-ID and Axcess, our MALDI-based identification services. Information includes genus and species names, number of spectra in the MSP, culture collection number(s), strain numbers, taxonomic confidence level, percent distance, 16S or ITS2 rDNA reference sequence, and any additional comments on the library entry.

MALDI Library Update

For the AccuPRO-ID® service and Axcess® System, the new and modified entries are incorporated into the Charles River MALDI library during a maintenance update

to enhance the diversity of the representation for relevant species in the library and to increase the accuracy of the current taxonomy. Updates occur as frequently as needed to incorporate new and modified entries for bacteria and yeast. Thus, organisms that are encountered more frequently in clinical testing or in manufacturing environments tend to have a greater diversity of representation. Those that are not encountered often generally have limited representation.

Validation and Performance Qualifications of the MALDI Library

MALDI library validation consists of testing activities that are designed to ensure that the entries in the libraries can reproducibly produce accurate identifications. The tests are designed to check the capacity of the libraries to produce identifications that are consistent with the identities of known strains. The MALDI library requalification is performed after Bruker library releases, or at least once per year. Spectra acquired from at least 10% of the strains added by Charles River since the last requalification are tested. Additional species' spectra, which represent the diversity of organisms we receive, and the most frequently observed organisms are used to generate reports to see if the samples identify as expected. For all qualifications, the protocols, raw data, identification reports, and final report are all maintained as part of the official validation package. The validations are approved by Quality Management and are maintained on file for review by our customers during on-site auditing activities.

Conclusions

Charles River understands the importance of having a compliant quality system in place to oversee all operations. Our quality system is purposely designed to ensure that laboratory processes, computerized systems, and equipment, as well as the reference libraries that are used to generate test results, are all adequately maintained in a state of control. For reference libraries specifically, we recognize that it is not acceptable to rely on non-curated public databases, clinically focused databases, or the outdated libraries that are distributed with some commercial identification systems. These commercial systems are typically not designed for identifying environmental isolates found in manufacturing facilities. Having direct knowledge of the organisms that are routinely isolated from pharmaceutical, biotechnology, medical device, and other manufacturing facilities worldwide has given us an advantage in determining what organisms are necessary in our reference libraries. We are committed to maintaining the most up-to-date, relevant, and compliant reference libraries available for our AccuGENX-ID®, AccuPRO-ID®, and Axcess® identification services.