Evaluating Library Databases for Microbial Identification: Critical Aspects and Recommendations

Abstract

A thorough evaluation process for microbial identification systems should consist of both a technical and financial review, regardless if you are performing internal testing and outsourcing. All evaluations should include a detailed list of specifications to meet your laboratory's objectives. Most laboratories perform evaluations by examining obvious system attributes and tangible features: work flow, automation, labor and technical requirements, cost of reagents and consumables, capital expenses and interfacing with laboratory information systems. A key component often neglected in the review process is the impact and effect of library databases on performance in a pharmaceutical environment, including comparing comprehensive depth of library entries, accuracy and relevancy to obtain a species level identification.

Introduction

Assessment of the library database used for microbial identifications is a critical component of evaluating a system or service in its ability to generate accurate identifications. Comprehensive depth of entries, accuracy and coverage of relevant species frequently found in aseptic and sterile manufacturing environments have a significant impact on both performance and cost. Databases that are not routinely updated and curated through a validated cGMP process can lead to taxonomic errors in species level identifications or generate inconclusive results. These errors cause higher repeat testing rates and time delays that affect operations and production, which effectively increases overall cost. As an independent contract laboratory service provider offering multiple technology solutions, Accugenix routinely compares differences in system performance for environmental isolates to establish product claims for our clients. This poster emphasizes the need to perform "Fitness for Use" studies beyond routine panels provided by manufacturers to understand both the performance and operational impact to your environmental monitoring system.

Methods

Accugenix sequenced 262,699 unknown bacterial isolates using 16S rDNA from June 2006 to March 2011. During that span, our proprietary library was updated 21 times enabling the detection of 4,158 unique, taxonomically distinct species. Utilizing the Accugenix Bacterial Library 04APR11 as the reference standard, we compared the databases of four commercially available identification systems for coverage, accuracy, relevancy and operational impact on cost of environmental isolate identifications. The bacterial library comparisons included the following four systems: bioMérieux Vitek® 2 Compact, Biolog GENIII®, Bruker MALDI-TOF BioTyper® v. 3.1.2 and ABI MicroSEQ® v. 2.2. The type strain is considered for each species, while redundant species entries were excluded because they would have no effect on projected performance. Taxonomically inaccurate or obsolete entries were considered errors. Projected performance expectations assume the assays are perfect and there are no laboratory errors or reagent quality issues. Case studies are also presented to demonstrate the relevancy that libraries have on accuracy and performance.

Results

The ability of any system to accurately identify bacteria to the species level is based on the coverage and the accuracy of the library entries. The Accugenix library contains substantially more unique library entries than any other test system, containing 4,158 unique, taxonomically distinct species. Specifically, in terms of relevancy, 1,979 unique species were identified by Accugenix, analyzing 262,699 isolates from aseptic and sterile manufacturing environments collected worldwide.

Table 1 is a comparative analysis demonstrating that all other identification systems are just a subset of the Accugenix library, comprising less than 50% of the coverage. Furthermore, differences in coverage of bacterial species relevant to the pharmaceutical and biotechnology were found to be significant. Such gaps and errors in entries prohibit speciation of many relevant bacteria. The lack of coverage observed is rather forgiving, as it does not take into account assay quality and laboratory operational errors. Moreover, the ability to determine an accurate species level identification can also be influenced by poor enzymatic expressions that organisms may yield in stressed growth conditions often found in manufacturing environments when tested using phenotypic systems.

Library	Total # of Unique Entries	# of Relevant Entries[1]	# of Missing Relevant Entries	Known Erroneous Entries
Accugenix	4158	1979	0	0
MicroSeq® v 2.2	1862	942	1037	25
BioTyper® v. 3.1.2	1801	893	1086	23
Biolog GENIII®	949	590	1389	17
Vitek® 2 Compact	432	325	1654	2

[1] = Observed at least once in testing 262,699 samples

Table 1. Library Comparison

Based on the data, one can conclude that a robust library curation process is required. The most basic cGMP compliant SOP for library generation must include a QC step to confirm the identity of strains received from culture collections. Culture collection labeling errors can be as high as 1%. Ultimately, both confidence and reliability of the platform being evaluated rests on coverage, relevancy and compliance of each new library entry and the overall library. Figure 1 illustrates the superior coverage of the Accugenix library based on missing relevant bacterial species. In comparison, the other systems contain only a fraction of the bacterial species observed over the last 5 years by Accugenix. The percentage of missing relevant entries varied from 52%, 55%, 70% and 84%respectively for MicroSeq, BioTyper, Biolog and Vitek 2. These deficiencies in the library limits the successful identification of many bacteria commonly observed in aseptic and sterile manufacturing industries. Since June 2010, 46,509 unknown isolates out of a total of 50,928 have been identified to the species level (91.3%).

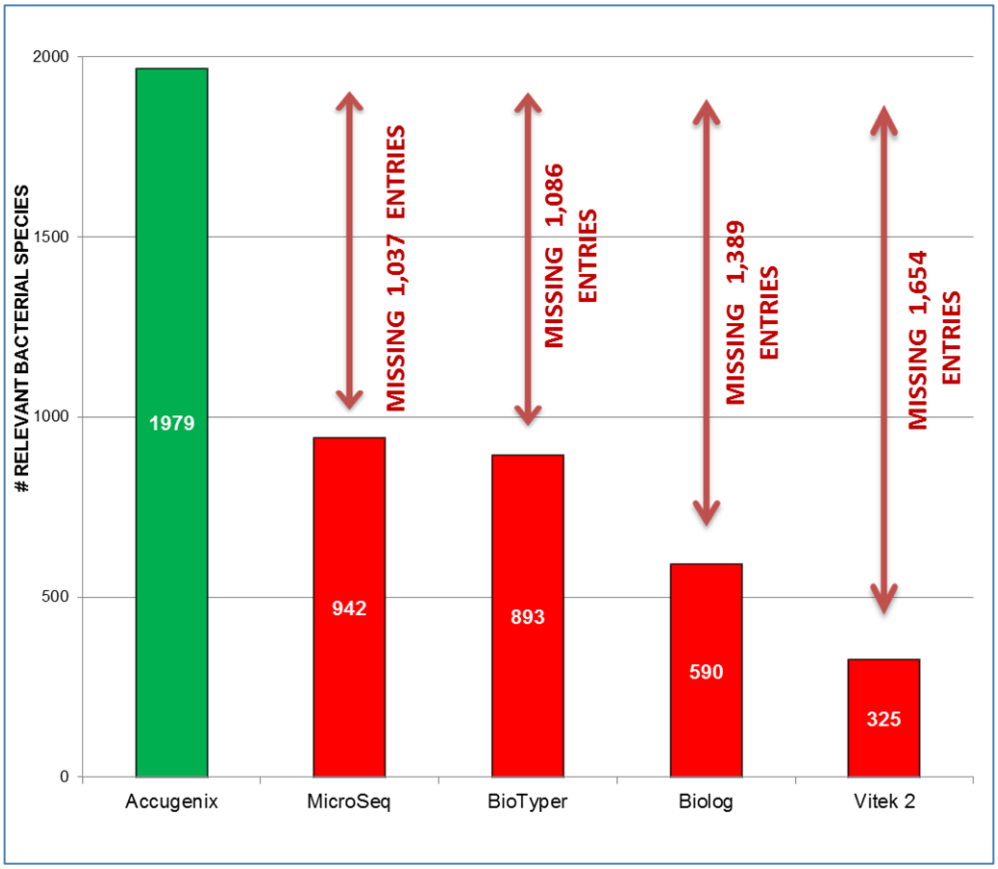


Figure 1. Differences in Number of Relevant Bacterial Species (n=262,699)

Case Study

A large aseptic manufacturing site submitted 1,532 unknown bacterial isolates for identification by AccuGENX-ID™ 16S DNA sequencing. From theses samples, 359 unique species were identified at this facility. Figure 2 illustrates that the other library databases had significant gaps in relevant entries to make a species level identification, ranging from 54% to 21% ability to record a species level identification from the organisms encountered at this facility.

Furthermore, 35% of the 359 unique species encountered at this facility were exclusive to the Accugenix library. When we examined the frequency of occurrence of these bacterial species entries, they represented 25% of the total number of isolates (Figure 3). This demonstrates that even employing two different systems, a species level identification could not be made.

From a cost perspective, at a minimum, this site would have to send 378 samples for further testing, usually to a contract laboratory for sequencing. The options would be to utilize a service that has a cGMP compliant, proprietary library and polyphasic approach in analyzing the data or searching non-GMP compliant public databases to find a matching sequence. Regardless of the approach, the net result would be an additional cost to determine an identification. Using $125 as an average price, it would cost at least $47,250 plus shipping expenses to identify these samples.

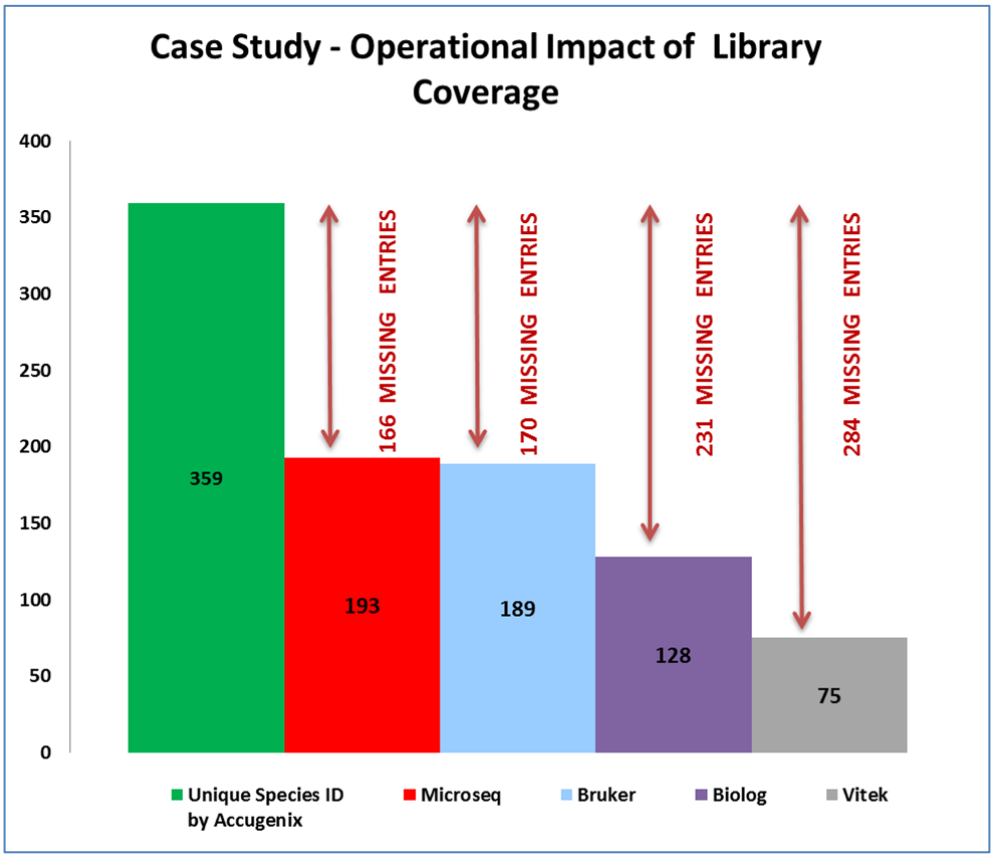
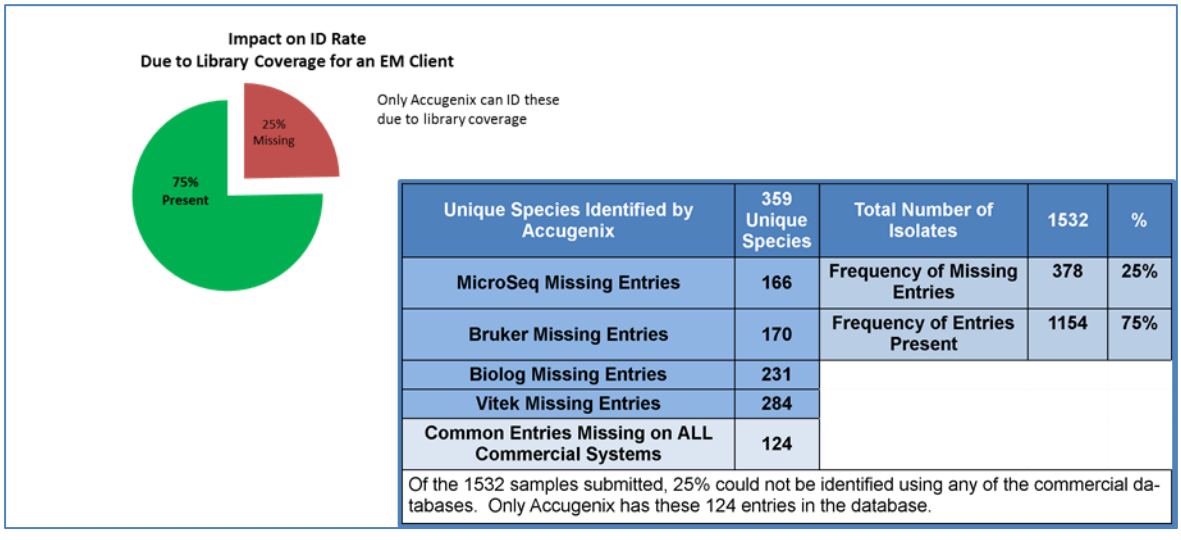


Figure 2. Operational Impact of Library Coverage



Unique Species Identified by Accugenix	359 Unique Species	Total Number of Isolates	1532	%
MicroSeq Missing Entries	166	Frequency of Missing Entries	378	25%
Bruker Missing Entries	170	Frequency of Entries Present	1154	75%
Biolog Missing Entries	231			
Vitek Missing Entries	284			
Common Entries Missing on ALL Commercial Systems	124			

Of the 1532 samples submitted, 25% could not be identified using any of the commercial databases. Only Accugenix has these 124 entries in the database.

Figure 3. Impact on ID Rate Due to Library Coverage for an EM Client

Discussion

According to the FDA guidelines for Sterile Drug Products Produced by Aseptic Processing* the goal of the laboratory is to achieve a species level identity for microorganisms encountered in environmental monitoring programs. The superior coverage and quality of the Accugenix bacterial library has yielded over a 91% species level identification – the highest available for any bacterial identification system. The goal in the evaluation process is finding the right platform for microbiological identification that is accurate, robust, and reliable for the purpose of monitoring the state of environmental control at a reasonable cost. Objectivity is required to analyze and interpret product claims against your own historical performance indices to determine the right solution. Are you seeking comparable performance or improvement in your performance? The greatest impact on the accuracy, reproducibility and cost is the library database used for identification. Databases that are deficient will lead to errors in species level identification, inconclusive results, or no matches against the reference library. This will directly lead to higher costs created by additional testing and time delays. The superior performance, reliability and relevancy, of microbial identification systems require that library updates are perpetually performed to reflect taxonomic changes and inclusion of novel organisms encountered in manufacturing environments. Performance evaluations must include testing the library species coverage by employing a challenge panel that represents the most frequently occurring organisms from raw materials, in-process stages and final production. Panels provided by manufacturers are often pre-selected because they are known to "work" on the system. These panels are suitable for IQ, OQ and PQ testing to verify that the instrumentation is working in meeting operating specifications. Challenge panels are required to enable and quantify the ability of a system to meet your technical specifications for accuracy, reproducibility and robustness to determine fitness of use for your facility.

***September 2004 Pharmaceutical cGMPs.**

Mark Calmann,
Senior Product Manager, Accugenix, Inc.
markcalmann@accugenix.com