

## Introduction

Comparative DNA sequencing of the 16S rRNA gene in bacteria has been proven to be the most accurate and reproducible method for identifying unknown organisms. The technology is inherently stable, which allows the data to be reproducible, and has an accumulated mutation rate which is very close to the rate at which species diverge into new, daughter species. Although the science behind the technology is widely accepted, there are still a number of variables which may impact the implementation and usage of this technology to achieve an accurate identification. While the technology is not perfect, most scientists agree this approach will usually provide a species level identification, and when it does not, you still have a unique identifier of the organism...its DNA sequence. Accugenix has identified three main areas of variability and concentrated on developing a process which significantly reduces this variability, resulting in more accurate and reproducible answers.

## The Accugenix Library

The first area which could affect the identification result is the library of known microbial 16S rRNA gene sequences. In order to correctly identify a large percentage of the unknown isolates in a Pharmaceutical Manufacturing environment, the library must contain DNA sequences for the organisms likely to be encountered. Accugenix has spent the past 10 years documenting the organisms found in Pharmaceutical and Biotech manufacturing facilities around the world. *Our validated, cGMP compliant library contains more species relevant to these industries than any other commercially available library in the world* (Figure 1).

Accugenix also continuously updates our library. Every day new organisms are being discovered and named. Some of these new organisms may be present in manufacturing facilities and could pose a risk to your product. Also, organisms are continuously being re-classified and re-named. This is a direct result of the new DNA sequencing technologies which are commonly available. With these new, more accurate molecular technologies, scientists are correcting past classification errors. Accugenix is committed to updating our library to reflect the current taxonomy.

## Our Scientists

The analysis of DNA sequence data is a complicated process. There are several steps which require an experienced scientist to make decisions about DNA sequence quality, DNA sequence assembly and phylogenetic interpretation, in order to correctly analyze and interpret the data. Attempts to automate this process have not been entirely successful, and much higher quality results are obtained when an experienced Microbial Phylogeneticist analyzes and interprets the data.

The Microbial Phylogeneticists at Accugenix first confirm the data is of a high enough quality to generate an identification report. The initial data quality check is performed automatically to classify the data as acceptable, not acceptable or requires a manual verification. Based on our experience, we have learned that some data will automatically be rejected for inappropriate reasons. For example, data which results from an organism that contains multiple copies of the 16S rRNA gene which are not the same length would be rejected by automated software. The reason for this is that the data resulting from this Indel (Insertion/Deletion) event would appear very similar to DNA sequences derived from a mixed culture. However, our experienced Microbial Phylogeneticists can easily tell the difference between mixed and Indel sequences and can use a proprietary software program to correct the Indel events.

Accugenix Bacterial Library 01Jun09	MicroSeq Bacterial Library V. 2.1
Bacillus fordii	
Bacillus fortis	
Bacillus fumaroli	
Bacillus funiculus	Bacillus funiculus
Bacillus galactosidilyticus	
Bacillus gelatini	
Bacillus gibsonii	Bacillus gibsonii
Bacillus ginsengihumi	
Bacillus halmapalus	Bacillus halmapalus
Bacillus halodurans	Bacillus halodurans
Bacillus halophilus	Bacillus halophilus
Bacillus hemicellulosilyticus	
Bacillus herbersteinensis	
Bacillus horikoshii	Bacillus horikoshii
Bacillus horti	Bacillus horti
Bacillus humi	
Bacillus hwajinpoensis	
Bacillus idriensis	
Bacillus indicus	
Bacillus infantis	
Bacillus infernus	
Bacillus insolitus	Bacillus insolitus

Figure 1: Comparison of Accugenix to MicroSeq

## **Expertise in Data Assembly and Interpretation**

Next, our scientists will ensure that our analysis uses the entire sequence generated from the PCR product. No nucleotides will be omitted from our analysis. Automated analysis programs will remove nucleotides from the ends of the sequence until a certain quality limit is reached. Depending on the overall quality of the DNA sequence, the number of nucleotides removed can range from 40 to 450 of the approximately 500 nucleotides in the sequence. Our experience has proven that the most accurate distance measurements, and therefore the most accurate identifications, are generated using as much of the DNA sequence as possible, and the more of the sequence that is analyzed, the higher confidence we have in the answer. *This is why we never produce an identification report using anything other than the entire sequence.*

Once the DNA sequence data has been correctly analyzed and compared to a relevant, validated library of known organisms, the identification report needs to be interpreted. This is a critical part of the analysis process, since there are no interpretation rules which can be applied universally. Instead, the education and experience of the Microbial Phylogeneticists is required to assign the correct identification. For example, if the DNA sequence difference between your unknown microorganism and the closest known microorganism in the Accugenix Library is 2%, should this be considered close enough to be a good Species level identification? If the closest match is a *Paenibacillus*, which typically are separated from one another by ~6%, it probably is a good match. On the other hand, if the closest match is an *Aeromonas*, which on average are separated from one another by only 1%, your isolate probably belongs to an unknown Genus.

In addition to the percentage of nucleotide differences, there are other attributes our Microbial Phylogeneticists consider before making an identification. The most important tool is the Phylogenetic Tree. A Phylogenetic Tree is a visual representation of the genetic variability within the most closely related organisms to the unknown. Both the distribution and the branching order indicate how organisms actually relate to one another and are equally important in the interpretation.

Accugenix understands the limits of 16S sequencing for identifying closely related organisms and developed a mechanism to report these organism clusters. We indicate a Species\* level of confidence to known groups of organisms including *E. coli/ Shigella*, the *Bacillus cereus Group* and the *Bacillus subtilis Group*. Groups such as these are recognized by clinical microbiologists and taxonomists, who agree cannot be differentiated by 16S sequencing alone.

Lastly, our Microbial Phylogeneticists are very experienced in where the problems may lie which would make the interpretation confusing. The phylogenetic classification of many organisms is currently incorrect. This can create a very complex situation which must take into consideration the genetic variability and branching order of a group of organisms. However, based on our experience of identifying over 300,000 unknown microorganisms, Accugenix has accumulated the experience and knowledge to turn this information into routine identifications.

## **Conclusion**

Comparative DNA sequencing for Bacterial and Fungal identification is a technology which has become the Gold Standard for microorganism classification and identification. However, as with most newer technologies, this information has not yet been incorporated into automated identification systems. Appropriate use of the data is critical in assigning correct identifications. Over the past 10 years, Accugenix has identified more microorganisms using this technology than any other laboratory in the industry. This experience, along with our improved process and relevant library allows Accugenix to provide the highest quality results to our customers.