

Comparison of Sequencing Identification Platforms used for Investigations and for Routine Water and Environmental Monitoring

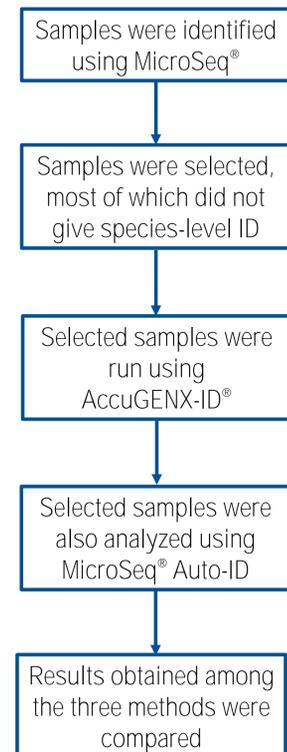
Dominique Joyce V. Foronda

1 ABSTRACT

Accurate and reproducible identification of microorganisms is necessary for conclusive failure investigations and for establishing relevant monitoring programs within a manufacturing facility. DNA sequence analysis is the gold standard for the identification of microorganisms, as the DNA provides relatively objective information compared with phenotypic profiling. The accuracy of the identification obtained and the frequency of a species-level identification are subject to the attributes of the sequencing platform utilized. These attributes include the reference library through which sample sequence is searched against, the method for assembly of the sequence data, and the method for interpretation of the results.

This study compares two sequencing platforms for the identification of microorganisms: AccuGENX-ID® and MicroSeq®. The samples utilized for the comparison were tested using the two sequencing platforms and their respective data analysis methods. This comparison was able to demonstrate a significant difference in the rate of species-level identification obtained between the two platforms and that this difference is relevant to the frequently occurring isolates within the region.

2 METHODOLOGY



Phylogenetic Tree Generated by MicroSeq® ID Data Analysis Software

Data QC and Analysis using AccuGENX-ID® Proprietary Software

Identification using the Auto-ID Function of the MicroSeq® ID Data Analysis Software

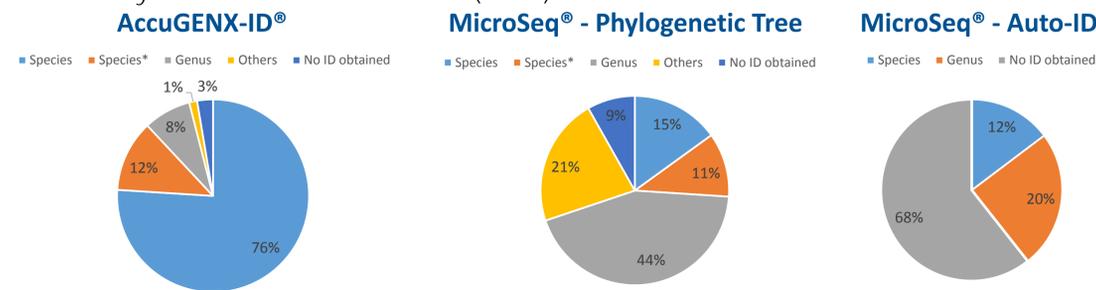
Specimen	Identification	Description	Comment
0362-SP4	Methylobacterium radiotolerans	Auto-ID for Species	

Auto-ID Parameters: Minimum sample score: 30, Minimum specimen score: 37, Minimum consensus length percentage: 80.00%, Minimum inter-species distance: 0.20%, Percent needed for a species match: 99.00%, Percent needed for a genus match: 97.00%

The Auto-ID function of the MicroSeq® ID Data Analysis Software will determine the identification of the unknown based on pre-set parameters in the software.

3 RESULTS

Summary of Identification Results (n=75)



Differences in Identifications Obtained

Out of the 75 samples used for the comparison, 8 samples had mismatches in the identifications obtained using AccuGENX-ID® and MicroSeq®. For the purpose of this study, a mismatch in identification is an occurrence where (1) the methods gave different species-level identifications for the same sample, or (2) the methods gave identifications that belong to different taxonomic groups (within the same taxonomic rank) for the same sample. Examples below show some instances where differences in the identifications were obtained.

Different species-level IDs obtained

Different confidence levels for the IDs obtained

Different IDs and confidence levels obtained

4 DISCUSSION

Comparison of AccuGENX-ID® and MicroSeq®

Property / Attribute	AccuGENX-ID®	MicroSeq®
Target Gene	16S rRNA (Bacteria) ITS2 rRNA (Fungi)	16S rRNA (Bacteria) D2 LSU rDNA (Fungi)
Data QC, Assembly, and Analysis	Semi-automated (manually-assisted) review of data quality, assembly of sequence, and analysis of data	Fully-automated review of data quality, assembly of sequence, and analysis of data based on threshold levels
Unique Bacterial Sequence Entries (as of April 2016)	6395 (3629 relevant* species entries)	1866 (1393 relevant* species entries)
Data Interpretation	Phylogenetic tree + % nucleotide difference + other attributes	Auto-ID function or based on threshold levels of % match or phylogenetic tree analysis

*Relevant species entries are those that are considered to be relevant to FDA-regulated industries.

Confidence Level of ID results

Comparison of Confidence Levels Obtained for the 75 Samples Selected for the Study

		MicroSeq® + Phylogenetic Tree			MicroSeq® Auto-ID		
		Species	Others	No ID	Species	Others	No ID
AccuGENX-ID®	Species (66)	27	34	5	8	14	44
	Others (7)	2	5	0	1	1	5
	No ID (2)	0	0	2	0	0	2

- AccuGENX-ID® was able to identify to species level 5 samples that did not yield any ID using MicroSeq® + phylogenetic tree analysis, and 44 samples with no ID results using MicroSeq® Auto-ID.
- AccuGENX-ID® was unable to identify 2 samples, but these two samples also cannot be identified using MicroSeq®.

Possible Impact on Environmental/Water Isolate Identifications

Environmental / Water Samples Tested in 2015 using MicroSeq® and Phylogenetic Tree Analysis for Interpretation

Summary of MicroSeq® ID Results For 2015

Category	Count	Percentage
NO ID	16	2%
OTHERS	57	7%
GENUS-LEVEL	166	22%
SPECIES-LEVEL	522	69%
TOTAL	761	

Frequently-Occurring Genus-Level IDs from 2015

Genus	Count	Percentage
Leifsonia	29	4%
Microbacterium	26	3%
Aspergillus	16	2%
Bacillus	15	2%
Paenibacillus	15	2%
Cladosporium	7	1%
Penicillium	5	1%

- 28 out of the 75 samples used for the study belong to the same genus as the frequently occurring genus-level IDs from MicroSeq®.
- AccuGENX-ID® was able to identify 26 (93%) of these 28 samples to the species level.
- Most identifications obtained by AccuGENX-ID® are not present in the MicroSeq® library.
- The use of the ITS region for fungal identification was able to give species-level identification for *Aspergillus*, *Cladosporium* and *Penicillium*.

5 CONCLUSION

- The AccuGENX-ID® platform was able to obtain Species-level ID for 88% of the samples used for the comparison.
- There is a high probability that genus-level identifications from MicroSeq® for environmental and water samples would yield species-level IDs if AccuGENX-ID® was used.
- The target gene used for identification, the method for analysis and interpretation, and library size have significant impact on the rate that species-level identification is obtained.