OVERVIEW OF CURRENT IDENTIFICATION SYSTEMS AND DATABASES
MICROBIAL IDENTIFICATION METHODS

**DNA**
- Sequencing of ribosomal RNA regions of bacteria and fungi

**RNA**
- MALDI-TOF mass spectrometry analysis of ribosomal proteins

**Protein**
- Analysis of biochemical reactions, acid and salt tolerance, membrane characteristics, etc. - VITEK®2 Compact, Biolog, API Strips, FAME

**Expression**
- Genotypic
- Proteotypic
- Phenotypic

- Catalase
- Fatty Acid
- Fermentation
- Gram Stain

**Accuracy & Reproducibility**
Differentiate between organisms based on the results of biochemical tests such as sugar fermentation, salt or pH tolerance.

Differentiate based on patterns of cellular fatty acids that are extracted, methylated and separated by gas chromatography.

Systems have been in use for decades, primarily in clinical settings leading to skewed reference libraries.
Convenient packaging for standard biochemical assays.
**Identification Panels**

Gram negative  
Gram positive  
Yeast  
Bacillus

http://www.biomerieux.com/
Identifications based on carbon utilization
Provides metabolic fingerprint
Can ID filamentous fungi

http://www.biolog.com
FATTY ACID (FAME)

The Sherlock® Microbial Identification System (MIDI)

Identification is based on patterns of cellular fatty acids.
Fatty acids are extracted, methylated and separated by GC.

http://www.midi-inc.com/
PHENOTYPIC ID - VARIABILITY

Requires live, healthy organism
Manual steps - Gram stain, dilution
Additional tests delay results
Technician errors
Subjective interpretation
Gram stain variability
Mixed culture
Limited and non-relevant entries in the reference libraries
Complicated workflows and decision trees
Physiologically stressed microorganisms
Incorrect Gram stains - Customer study: 331/1586 had wrong Gram stain (21%)

<table>
<thead>
<tr>
<th>Genus</th>
<th>Total Samples</th>
<th>Gram Rxn</th>
<th># Incorrect</th>
<th>% Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudomonas</td>
<td>531</td>
<td>GN</td>
<td>136</td>
<td>26</td>
</tr>
<tr>
<td>Enterococcus</td>
<td>97</td>
<td>GP</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Aerococcus</td>
<td>48</td>
<td>GP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Stenotrophomonas</td>
<td>38</td>
<td>GN</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>Bacillus</td>
<td>37</td>
<td>GP</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Microbacterium</td>
<td>32</td>
<td>GP</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>Acinetobacter</td>
<td>29</td>
<td>GN</td>
<td>17</td>
<td>59</td>
</tr>
<tr>
<td>Mycobacterium</td>
<td>16</td>
<td>GP</td>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>Serratia</td>
<td>16</td>
<td>GN</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>Methylobacterium</td>
<td>15</td>
<td>GN</td>
<td>9</td>
<td>60</td>
</tr>
<tr>
<td>Psychrobacter</td>
<td>7</td>
<td>GN</td>
<td>6</td>
<td>86</td>
</tr>
</tbody>
</table>
PHENOTYPIC ID

Still have a role in a microbial quality program, but need accurate identification to utilize the information.

Determine the biochemical activity of an organism on the product and the resulting stability of the product in the presence of that organism.

Understanding the nutrient requirements of an organism provides insight into controlling or eliminating the organism or preserving the product.
Matrix-Assisted Laser Desorption/Ionization (MALDI) Time of flight (TOF) Mass Spectrometry
PROTEOTYPIC ID

MALDI-TOF analysis

Spectra Unknown (primarily ribosomal proteins)

Library Comparison

Escherichia coli

Enterobacter aerogenes

Bacillus niacini

Identification
PROTEOTYPIC ID - VARIABILITY

Requires a live, healthy organism

Temperature exposure and age of culture

Sample preparation

Transcription/technician errors

May require additional tests for an ID

Limited and non-relevant reference libraries

Interpretation of test results
GENOTYPIC ID

DNA-based Identifications using information contained in the Ribosome
The use of rDNA sequences for bacterial and fungal taxonomic classification has been in practice for many decades.

Both highly conserved and highly variable regions of sequence which allow for species level resolution.

Information content is high because the complexity is in the data.

Results are phylogenic.

Results are “portable”.
ABILITY OF A SYSTEM TO DISCRIMINATE IS BASED ON THE POTENTIAL INFORMATION CONTENT

Genotypic methods
AccuGENX-ID®, 460 bp informative, 4 bases..................$4^{460} = 8 \times 10^{276}$
MicroSEQ® 2.0, 400 bp max informative, 4 bases.....$4^{400} = 7 \times 10^{240}$
RiboPrinter, 64 Fragment sizes, 10 levels .............$10^{64} = 1 \times 10^{64}$

Proteotypic methods
AccuPRO-ID®, 700 Mass frag. sizes, yes/no..........$2^{700} = 5 \times 10^{210}$

Phenotypic methods
Fatty Acids, 128 FAs, 4 quantitative levels...............$4^{128} = 1 \times 10^{77}$
Biolog Carbon utilization, 95 tests, yes/no..............$2^{95} = 4 \times 10^{28}$
VITEK®2 Compact Biochemical, 40 tests, yes/no......$2^{40} = 1 \times 10^{12}$
API® Biochemical, 32 tests, yes/no.......................$2^{32} = 4 \times 10^{9}$
Genotypic ID

Bacterial Ribosome

Genome organization

16S  23S  5S
SSU  LSU  ITS1  ITS2

70S

Small Subunit

16S rRNA
21 proteins

Large Subunit

5S and 23S rRNAs
34 proteins
ITS2 (Internal Transcribed Spacer) is used for the identification of fungal species via DNA sequence signatures by the European Consortium for the Barcode of Life

ITS2 is more variable than both ITS1 and D2

Providing more species level differentiation

18S rRNA
33 proteins

5S, 5.8S and 28S rRNAs
49 proteins
GENOTYPIC ID

Isolate DNA from a single, pure colony

PCR amplify target sequences and subject to cycle sequencing

Analyze, assemble and interpret the data

Compare DNA sequence to a library

Sequence Data

1  GATGAAATGG CTGGCGGCTGC TTAAACATCG CAAAGTCAAGAC GATGAAAGCC AGCTGGCTGG
61  GTGAGTTAGT GCGGACAGGG TGGTAAACAC GTGAGTAAACC TGCCCTAAAC TCTGGGATAA
121  GCCCTGGAAA CTTGCTCTAT ACCGAGTATGG AGCCTGTACC GCAATGTTGG TGTTGAAAG
181  ATTTATGGTT TTTGATGCG CTGCCCCGCTG ATCAGGCTTGC TGGTGGTTGA ATGGTCACC
241  AAGGCGACGG CGGTAGACGG GCCGAGAAGGG GTGACGAGCC ACGAAGGACC CGGAGCACGG
301  CCCAGAATCC TACGGAGAGG AGCAGTGAGG AATATCGCAG AATGCCGAGA AGCTGGAGGC
361  AGGCGACGCG CGTGAGGGA TGGCAGCCTT CGGGTTGGTA CCGTTGCTTGGT GTCGGTTG
421  AGGAAATCGG AGCCTACTCG CAGAGAAGG ACCGGCTAAC GATTTGCAG CAGCGCCGCTG
481  AATACGTAGG GTGCGAGCG TATCAAGAGAT TATGGCTGGT AAGACCTGG TAGGGCGTTT

Raw Sequence Electropherogram
GENOTYPIC ID - VARIABILITY

Very labor intensive

Lot-to-lot variations of reagents

Instrumentation (thermalcycler, sequencer, pipettes)

Data/sequence quality

Multi loci variability (insertions, deletions, polymorphism)

Transcription/technician errors

May require additional tests for an ID (especially if not in database)

Limited and non-relevant reference libraries to make the ID
## CONSIDERATIONS OF ID PLATFORM

<table>
<thead>
<tr>
<th></th>
<th>16S sequencing</th>
<th>MALDI-TOF MS</th>
<th>Automated phenotypic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Investigations</td>
<td>Routine</td>
<td>Routine</td>
</tr>
<tr>
<td>Accuracy</td>
<td>High</td>
<td>High-Med</td>
<td>Med-Low</td>
</tr>
<tr>
<td>Organism range (bacteria, fungi)</td>
<td>Broad</td>
<td>Broad-Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Assay throughput</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Assay time</td>
<td>Medium</td>
<td>Fast</td>
<td>Medium</td>
</tr>
<tr>
<td>Capital costs</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Consumables cost</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Operational skill</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
</tbody>
</table>
DIVERSITY AND UNIQUE NUMBER OF SPECIES ROUTINELY IDENTIFIED AT ACCUGENIX
BACTERIAL LIBRARY COMPARISONS

Demonstrating the importance of having a relevant library focused on EM programs and not the clinical setting

Accugenix® Bacterial Library 07 Apr 2016
Bruker MALDI Biotyper® v.5.0.0.0 Bacterial Library 2015
MicroSEQ® ID 16S rDNA 500 Library v2013
Biolog GEN III® Bacterial Library 2013
MIDI DNA Bacterial Library Jul 2011
MIDI FAME Library Jul 2011
bioMérieux VITEK® V2 MS 2013
bioMérieux VITEK® 2 Compact Bacterial Library Feb 2009
BACTERIAL LIBRARY COMPARISONS

<table>
<thead>
<tr>
<th>Technology</th>
<th>Unique Species Entries</th>
<th>Relevant Species Entries</th>
<th>Excluded Species Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCUGENIX®</td>
<td>6395</td>
<td>3629</td>
<td>0</td>
</tr>
<tr>
<td>Biotyper®</td>
<td>1916</td>
<td>1343</td>
<td>2286</td>
</tr>
<tr>
<td>MicroSEQ®</td>
<td>1866</td>
<td>1393</td>
<td>2236</td>
</tr>
<tr>
<td>Biolog</td>
<td>1534</td>
<td>1161</td>
<td>2468</td>
</tr>
<tr>
<td>MIDI DNA</td>
<td>1184</td>
<td>923</td>
<td>2706</td>
</tr>
<tr>
<td>MIDI FAME</td>
<td>969</td>
<td>789</td>
<td>2840</td>
</tr>
<tr>
<td>VITEK® MS</td>
<td>834</td>
<td>710</td>
<td>2919</td>
</tr>
<tr>
<td>VITEK®2</td>
<td>428</td>
<td>405</td>
<td>3224</td>
</tr>
</tbody>
</table>

Insufficient library coverage impacts the number of reportable results.

Employing multiple commercial ID systems increases the cost but will not guarantee a higher rate of species level identifications.

Library coverage must be reflective of the organisms observed in your environment and include current taxonomic information.
Identification and confidence level

Taxonomic information based on current IJSEM taxonomy, e.g.

Visually see how the identified organism relates to its closest neighbors with a % distance measure and a phylogenetic tree

cGMP and 21 CFR Part 11 Compliant

Identification is routine for >9000 bacteria and fungi
**AccuPRO-ID®: Accugenix MALDI-TOF Solution**

Matrix-Assisted Laser Desorption/Ionization (MALDI) Time of flight (TOF) – Mass Spectrometry

Bacterial ID from MALDI-TOF technology and Accugenix MS library

If MALDI-TOF cannot provide an ID, then sent for 16S sequencing

Sequence ID provided at no additional cost
STRAIN TYPING
**Multi and Single Locus Sequence Typing**

- DNA extraction
- MLST/SLST gene amplification
- Sequencing
- Phylogenetic analysis

**Ribotyping**

- DNA extraction
- Restriction digest
- DNA electrophoresis and transfer to membrane
- Probe with a region of the rRNA operon, image and analysis
16S rDNA was useful for species-level ID; however, it failed to give enough resolution at strain level

(Colors refer to the same Ribogroups)

32 isolates tested clustered into 7 groups showing very limited utility of 16S rDNA for strain level differentiation
CASE STUDY: STRAIN TYPING OF *M. LUTEUS*

Sequencing 3 targets, combining and aligning the data, results in a high level of strain level discrimination with the identification of 31 different sequence types.