SHIMADZU

Rapid Identification of Food-Borne Bacteria Using a Vast Database of Theoretical Prokaryotic Protein Masses Predicted from Genome Sequences for MALDI-MS

Kanae Teramoto[1], Mizuki Niida[1], Eiko Nakashima[2], Nana Mikami[3], Eiki Yamasaki[4], Hirofumi Enomoto[2] and Fumiki Morimatsu[2]

1. Shimadzu Corp., Kyoto (Japan), 2. Tokushima Univ., Tokushima (Japan), 3. Obihiro Univ. GAMRC, Hokkaido (Japan), 4. National Institute of Infectious Diseases (NIID), Tokyo (Japan)

Kanae Teramoto, Ph. D.
General Manager
t_kanae@shimadzu.co.jp
MS Business Unit
Shimadzu Corporation
1, Nishinokyo Kuwabara, Nakagyo, Kyoto 604-8511, Japan

Overview

- ✓ A large-scale genomically predicted protein mass database (GPMsDB) was developed from publicly available bacterial and archaeal genomes for use in MALDI-MS, aiming to expand the range of identifiable taxa.
- ✓ MicrobialTrack is a software (web application) that identifies prokaryotic microorganisms by analyzing MALDI mass spectra using theoretical masses of proteins registered in GPMsDB.
- ✓ More than 20 food-born bacteria and the genus Acinetobacter strains were analyzed to evaluated identification accuracy by comparing results with genetic analyses.
- ✓ Most strains were correctly identified, but strains identified as *L. paraplantarum* by 16S rRNA gene analysis were identified *as L. plantarum* by MicrobialTrack. The number of ribosomal proteins detected in the sample strains was the same for both species. However, the number of peaks assigned as *L. plantarum*'s non-ribosomal proteins was slightly larger, which was identified as *L. plantarum* by MicrobialTrack.

1. Introduction

MALDI-MS is widely used mainly in the field of clinical microbiology because it can identify microorganisms quickly and easily. In food microbiology testing, it is required to identify microorganisms derived from food raw materials and the environment in the production line. Therefore, it has been pointed out that some microbial species cannot be identified by MALDI-MS, which was developed mainly for clinical microbiology testing. Although new species are frequently added and taxa are consolidated or eliminated, conventional MALDI-MS microbial identification systems, which require the actual measurement of microorganisms and their entry into libraries, do not always add up-to-date information. Furthermore, even when libraries are updated, they need to be updated by users at a cost.

A large-scale library (GPMsDB) that registers the theoretical mass of microbial proteins from genomic information were constructed instead of MALDI mass spectra obtained from cultured microorganisms (Y. Sekiguchi et al., Genome Biology, 2023). We developed MicrobialTrack, a software program that identifies microorganisms using GPMsDB. MicrobialTrack allow to identify about 85000 species of Bacteria and Archaea.

In this study, food-borne and the genus *Acinetobacter* which have a high demand for analysis were identified by MicrobialTrack.

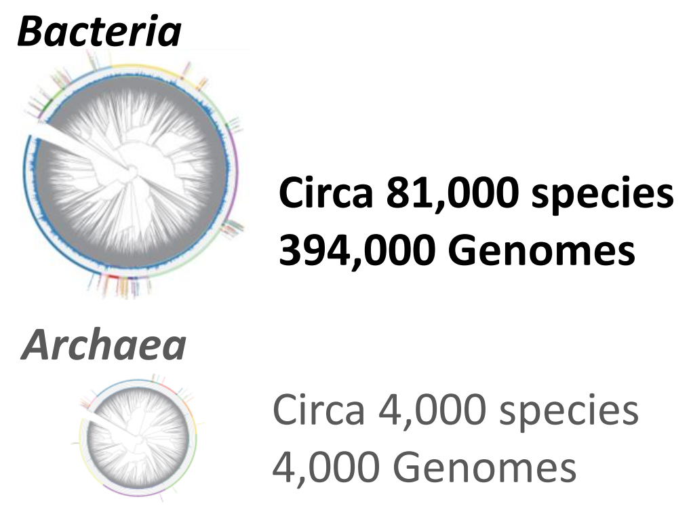
The features of MicrobialTrack™

Review

Results

1. The largest database size for MALDI-MS microbial identification

- **GPMsDB**



Circa 81,000 species
394,000 Genomes

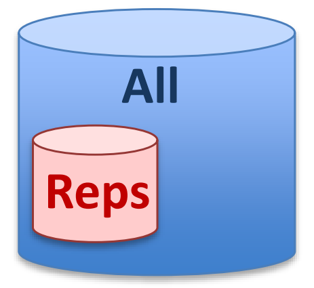
Circa 4,000 species
4,000 Genomes

- ✓ Circa 18,000 prokaryotic species (*Bacteria* and *Archaea*) with valid published taxonomic names
- ✓ Approximately **85,000** prokaryotic species predicted based on genomic information

*Includes difficult-to-culture and uncultured microorganisms.

- ✓ Identification results are displayed with taxonomic names according to the Genome Taxonomy Database (GTDB, https://gtdb.ecogenomic.org).

- **Identification can be performed using either of the two databases**



The "All" database
- ✓ Contains theoretical masses predicted from all genome sequences.
- ✓ Includes protein information from multiple strains (genomes) for most species-level taxa.

85,203 species, 402,130 genomes

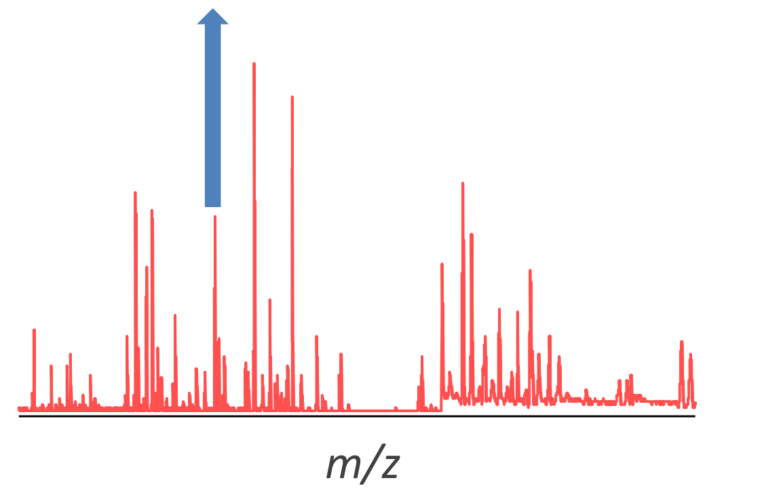
The "Representative (Reps)" database
- ✓ Contains theoretical masses predicted from a single strain for one species.
- ✓ Used for quick estimation of taxonomic identification at the genus and species levels.

85,201 species, 85,201 genomes

Note: The number of data points in the figure is based on information as of March 2025.

2. Proteomics-Based Identification

Observed mass: *m/z* 8754.9 ⟷ Theoretical mass: [M+H]+ 8754.9



Calculate

Amino acid sequence

```
MPKLSAAELR QLSGEELRNK VRELKEELFG LRFQSATGQL ENTARLREVR KDIARVYTVL QERNLNIVDD
PDSTKEA
```

Translate

Gene nucleotide sequence:
Cutibacterium acnes L29

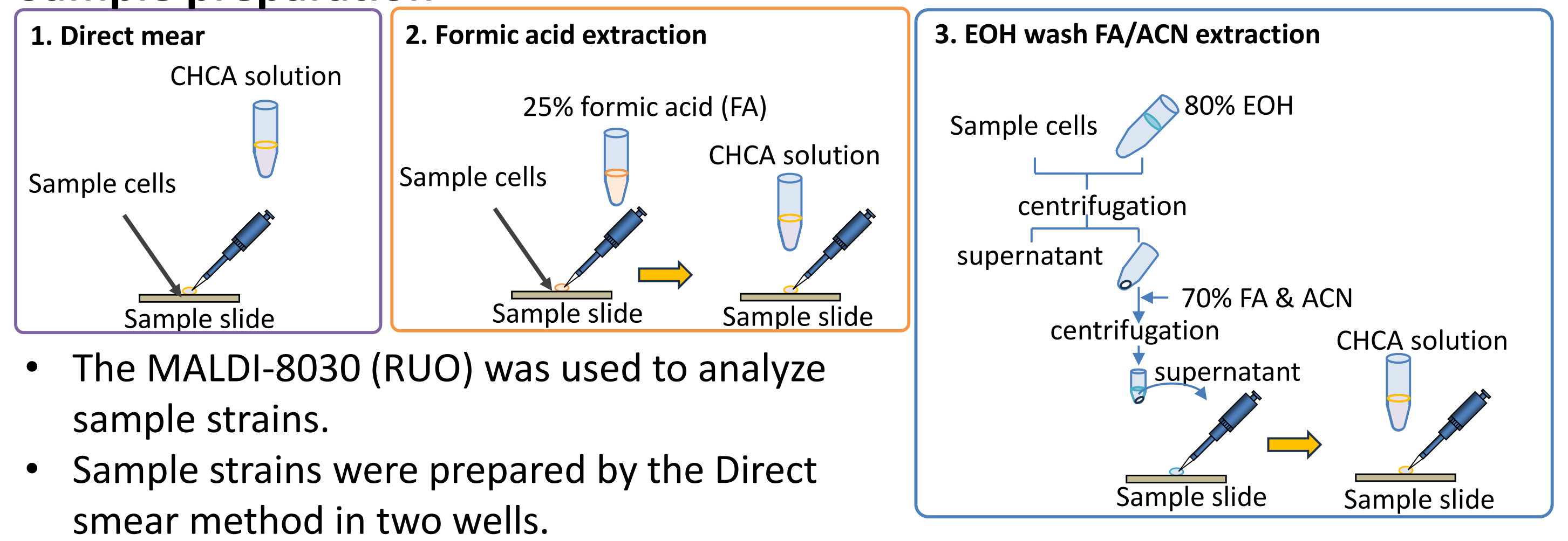
```
ATGCCGAAGC TTTCTGCTGC TGAGCTGCGC CAGCTTTCCG GTGAAGAGTT GCGCAACAAG GTTCGTGAAC
TGAAGGAGGA GCTGTTCGGG CTGCGTTTCC AGTCGGCCAC CGGTCAGCTC GAGAATACGG CTCGCCTGCG
TGAGGTCCGC AAGGATATTG CCCGGGTGTA CACCGTGCTC CAGGAGCGCA ACCTCAATAT CGTCGACGAT
CCCGATTCCA CTAAGGAGGC CTGA
```

3. Cloud service (web application)

- ✓ Low initial cost, no installation required, accessible anywhere
- ✓ Database renewal procedures and costs are not required during the license period because version upgrades are performed on Shimazu's server.

2. Methods

Sample preparation



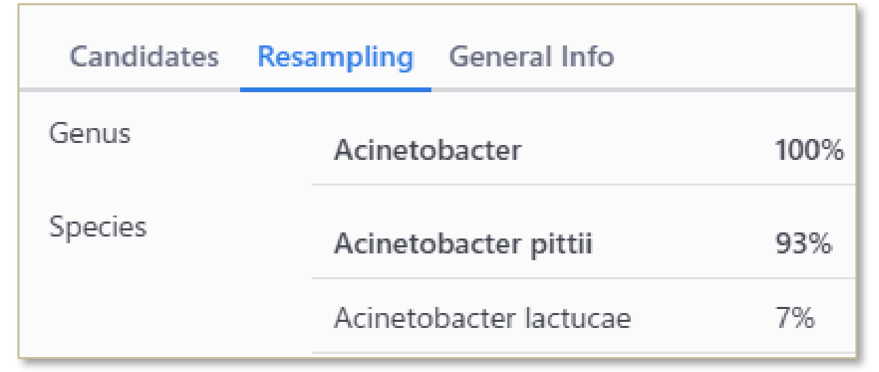
- The MALDI-8030 (RUO) was used to analyze sample strains.
- Sample strains were prepared by the Direct smear method in two wells.
- If the identification results from the direct smear method differed or low reliability (Low or Middle), the samples were re-prepared in pairs of two wells using the FA extraction method. If the results were the same, they were considered the final results.
- If the identification results from the FA extraction method differed, the samples were re-prepared in pairs of two wells using the EOH wash FA/ACN extraction method, and these results were taken as the final identification results.

Identification Results of *Acinetobacter* species by MicrobialTrack

Sample strain		MicrobialTrack: Top Candidate	MicrobialTrack: Resampling*
Acinetobacter baumannii	NBRC 109757^T	*Acinetobacter baumannii*	*Acinetobacter baumannii*
Acinetobacter baylyi	NBRC 112578^T	*Acinetobacter baylyi*	*Acinetobacter baylyi*
Acinetobacter beijerinckii	NBRC 112579^T	*Acinetobacter beijerinckii*	*Acinetobacter beijerinckii*
Acinetobacter bouvetii	JCM 18991^T	*Acinetobacter bouvetii*	*Acinetobacter bouvetii*
Acinetobacter calcoaceticus	NBRC 12552	*Acinetobacter calcoaceticus*	*Acinetobacter calcoaceticus*
Acinetobacter gyllenbergii	NBRC 112581^T	*Acinetobacter gyllenbergii*	*Acinetobacter gyllenbergii*
Acinetobacter haemolyticus	NBRC 109758^T	*Acinetobacter haemolyticus*	*Acinetobacter haemolyticus*
Acinetobacter johnsonii	NBRC 102197	*Acinetobacter johnsonii*	*Acinetobacter johnsonii*
Acinetobacter junii	NBRC 109759^T	*Acinetobacter junii*	*Acinetobacter junii*
Acinetobacter lwoffii	NBRC 109760^T	*Acinetobacter lwoffii*	*Acinetobacter lwoffii*
Acinetobacter nosocomialis	NBRC 110498	*Acinetobacter nosocomialis*	*Acinetobacter nosocomialis*
Acinetobacter parvus	NBRC 112582^T	*Acinetobacter parvus*	*Acinetobacter parvus*
Acinetobacter pittii	NBRC 110505	*Acinetobacter pittii*	*Acinetobacter pittii* 93%/ *A. lactucae* 7%
Acinetobacter radioresistens	NBRC 102413^T	*Acinetobacter radioresistens*	*Acinetobacter radioresistens*
Acinetobacter schindleri	NBRC 112583^T	*Acinetobacter schindleri*	*Acinetobacter schindleri*
Acinetobacter venetianus	NBRC 112585^T	*Acinetobacter venetianus*	*Acinetobacter venetianus*

A. pittii was correctly identified with high reliability using EOH wash FA/ACN extraction .

Resampling results for *A. pittii* NBRC 110505



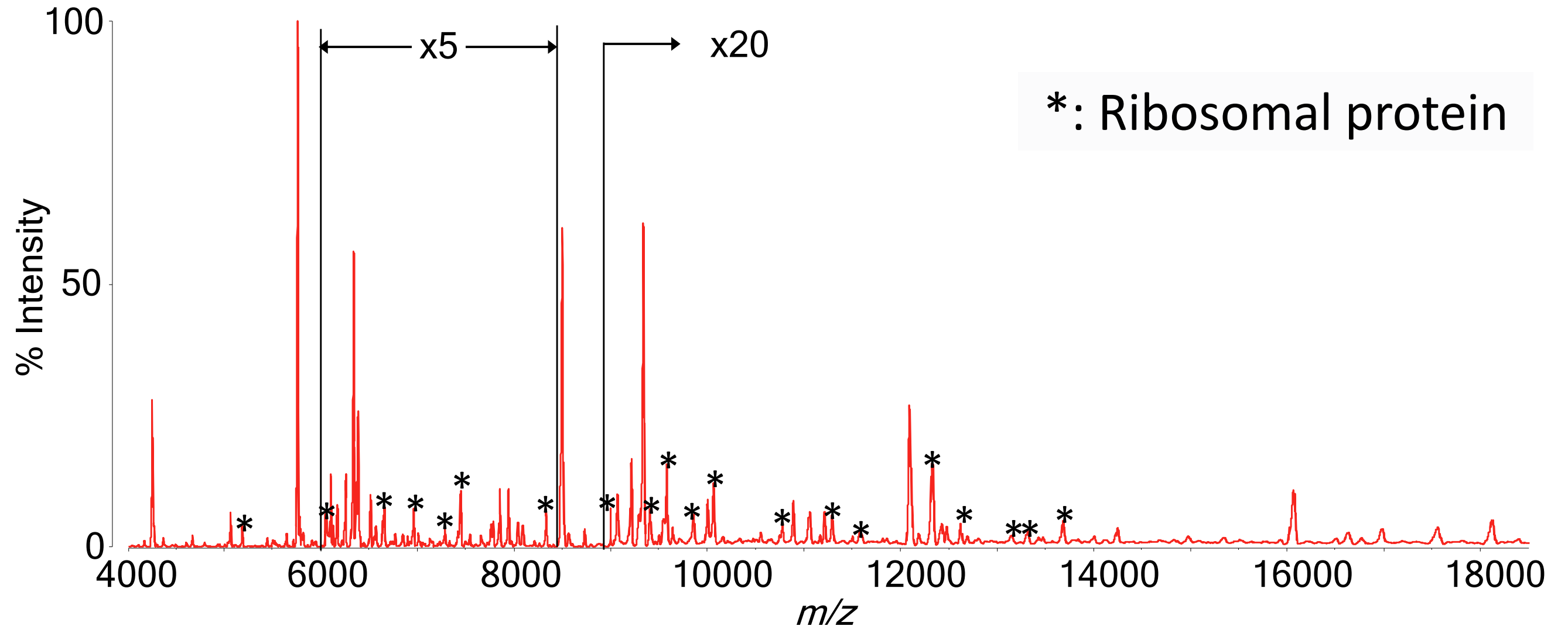
Genus	Acinetobacter	100%
Species	Acinetobacter pittii	93%
	Acinetobacter lactucae	7%

Based on this random resampling, McrobialTrack provides an estimate of the identification certainty level at both the genus and the species levels.

Resampling is a method that involves selecting a subset of randomly observed peaks to repeat the identification process. It displays the genera and species identified as top-hit organisms during this process, along with their frequency of occurrence.

3. Results

MALDI mass spectrum from *Acinetobacter baumannii* NBRC109757^T



- The peaks in the high mass range are successfully detected with high sensitivity using the MALDI-8030.
- Twenty peaks are identified as ribosomal proteins, and 41 peaks are identified as non-ribosomal proteins using MicrobialTrack.

Identification Results of Food-born Bacteria by MicrobialTrack

Sample	16S rRNA (500 bp)	MicrobialTrack: Top Candidate	MicrobialTrack: Resampling*
SB01	*Bacillus amyloliquefaciens/B. velezensis*	*B. amyloliquefaciens/B. velezensis*	*B. amyloliquefaciens* 49%/ *B. velezensis* 36%/ *B. siamensis* 15%
SB02	*Bacillus amyloliquefaciens/B. velezensis*	*B. velezensis*	*B. velezensis*
SB03	*Bacillus licheniformis*	*B. licheniformis*	*B. licheniformis* 98%/ *B. haynesii* 2%
SB04	*Bacillus sonorensis*	*Bacillus sonorensis*	*B. sonorensis*
SB05	*Bacillus subtilis/B. stercoris*	*B. subtilis*	*B. subtilis*
SB07	*Lacticaseibacillus rhamnosus*	*L. rhamnosus*	*L. rhamnosus*
SB08	*Leuconostoc carnosum*	*L. carnosum*	*L. carnosum*
SB09	*Leuconostoc lactis*	*L. lactis*	*L. lactis*
SB10	*Leuconostoc mesenteroides*	*L. mesenteroides*	*L. mesenteroides*
SB11	*Staphylococcus epidermidis*	*S. epidermidis*	*S. epidermidis*
SB12	*Bacillus cabrialesii/B. inaquosorum*	*B. subtilis* complex: *B. inaquosorum*	*B. subtilis*
SB13	*Leuconostoc citreum/L. holzapfelii*	*L. citreum*	*L. citreum*
SB14	*Bacillus zhangzhouensis/B. pumilus/B. safensis*	*B. safensis*	*B. safensis*
SB15	*Leuconostoc holzapfelii*	*L. citreum*	*L. citreum*
SB16	*Lactiplantibacillus paraplantarum/L. plantarum*	*L. plantarum/L. pentosus*	*L. plantarum* 99%/ *L. pentosus* 1%
SB17	*Lactococcus cremoris*	*L. cremoris*	*L. cremoris*
SB18	*Staphylococcus equorum*	*S. equorum*	*S. equorum*
SB19	*Staphylococcus warneri*	*S. warneri*	*S. warneri*
SB20	*Bacillus altitudinis*	*B. altitudinis*	*B. altitudinis* 90%/ *B. safensis* 9%/ *B. safensis* 1%
SB21	*Kocuria salsicia*	*K. salsicia*	*K. salsicia*

- The detection of numerous peaks in the high mass range by MALDI-8030 is contributed to the accurate identification of sample strains.
- Most strains were correctly identified, but strains identified as *L. paraplantarum* by 16S rRNA gene analysis were identified as *L. plantarum* by MicrobialTrack.
  - ✓ The number of ribosomal proteins detected in the sample strains was the same for both species.
  - ✓ However, the number of peaks assigned as *L. plantarum*'s non-ribosomal proteins was slightly larger, which was identified as L. *plantarum* by MicrobialTrack.
- Because partial sequences of the 16 S rRNA gene cannot confirm the species by whole-genome sequencing and reconfirmation of the identification results from MicrobialTrack are necessary.