Impact of species' diversity in improving proteotypic databases for microbial identification

Cindy Serrato Zavala, Chloé Huyghe, Sujan Timilsina, Bindhu Verghese, Charles River Laboratories, Newark, DE

1 ABSTRACT

Accurate microbial identification remains a crucial component of quality control and safety in the pharmaceutical industry. Among various methods of microbial identification, proteotypic-based MALDI-TOF (Matrix Assisted Laser Desorption/Ionization Time-of-Flight) technology has emerged as a rapid, high-throughput, and cost-effective method for microbial identification. However, the strength of this method hinges on the robustness and representativeness of the database, particularly its ability to capture intra-species variability.

For this study, microbial library entries were generated from microbes isolated from diverse geographic regions to assess the impact of local diversity on identification. Findings showed that identification outcomes by these entries vary by species. For example, approximately 12% of *Sphingomonas colocasiae* and 7% of *Penicillium brevicompactum* identifications relied on intra-species geographic diversity entries. Whereas 60% of *Aspergillus westerdijkiae* and 100% of *Beauveria pseudobassiana* identified matched to similar intra-species entries. We further breakdown the impact of these database entries to identify organisms in their regional zone. Notably, regionally sourced entries assist in providing reliable identification for samples in that geographic region.

This highlights the relevance of incorporating geographically diverse strains into proteotypic libraries. As microbial diversity varies across manufacturing environments, maintaining comprehensive, regionally representative databases is essential for ensuring accurate identification and effective contamination control.

2 BACKGROUND

- Through periodic database updates, entries/spectra are added to the MALDI-TOF library to accurately identify organisms.
- The efficacy of the proteotypic method of identification is influenced by strain diversity.
- In this study, microbial database entries were generated from different strains – isolated from diverse geographic regions – and analyzed to assess the impact of local diversity on identification results.
- This study focuses on entries generated from five unique species processed in Écully, France (Écully site-derived entries) and how their inclusion in the MALDI-TOF library has impacted species-level identification results. These entries were created and added to the Accugenix® MALDI-TOF library throughout 2024.
- Transfer refers to the process where a sample does not receive a species-level identification through our proteotypic databases and thus is transferred over to sequencing for identification results.

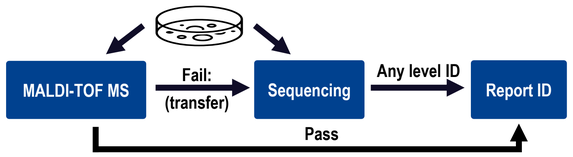

Figure 1. Workflow of Accugenix® microbial identification services. If a sample is submitted for MALDI-TOF MS identification and does not provide a result, the sample is sequenced to provide an identification.

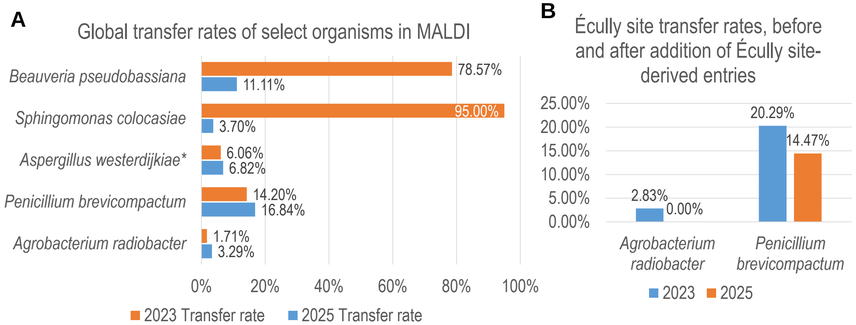

Figure 2. (A) Transfer rate before and after the Écully site-derived entries were added to the Accugenix® MALDI-TOF Library. Transfer rates for each species were analyzed between the months of January through September for 2023 and 2025. All species are reported at a species level identification except for *Aspergillus westerdijkiae* which is reported as *Aspergillus ochraceus / westerdijkiae*, as these two species are too closely related to be reliably differentiated by MALDI-TOF MS platform. Sequencing of ITS / 16S rRNA genes were used to verify species identity prior adding to the database. (**B)** Specific transfer rates reduced for *Penicillium brevicompactum* and *Agrobacterium radiobacter* at the Écully site, although the global transfer were still higher in 2025 compared to 2023, highlighting the impact of encompassing species diversity for proteotypic identification methods.

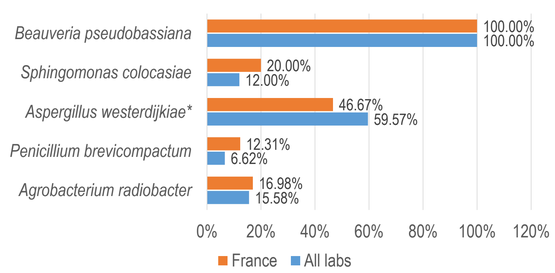

Figure 3. Impact of Écully site-derived entries on identification accuracy across geographic zones. 438 samples, reported between January and September of 2025, were analyzed to assess the impact that the Écully site-derived entries had in the final reporting of the sample once the entries were added to the library. The region-specific entries were relevant in providing identification results beyond the regional labs and helped reduce the global transfer rates resulting in rapid sample identification process through MALDI process.

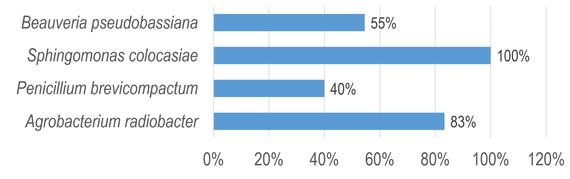

Figure 4: Percentage of potential identifications after addition of Écully site-derived entries. 45 spectra were analyzed, belonging to the listed species as identified by sequencing, that were not able to secure a MALDI identification result between 2021-2024. Following the addition of Écully site-derived entries to Accugenix® databases, these spectra would now be identified to the species level through MALDI processing.

3 RESULTS

- Entries added from samples processed in global lab added to the species diversity in the MALDI library, resulting in significant reduction in transfer rates for organisms such as *Beauveria pseudobassiana* and *Sphingomonas colocasiae* (Figure 2A).
- For organisms where the global transfer rates increased in 2025 compared to 2023, the newer entries were still significant in reducing the transfer rates in the Écully, France lab indicating the strain diversity received in each of Accugenix® global labs (Figure 2B).
- These newly added entries were responsible for providing identification results for majority of these new identification reports provided to the customer, significantly improving the breadth of the MALDI database species coverage (Figure 3).
- Additionally, results indicate that the newly added entry from sample processed in a France lab would have provided an identification result for majority of the previously transferred samples, up to a 100% identification rate in *Sphingomonas colocasiae* (Figure 4).

4 CONCLUSION

- Results from proteotypic methods of identification can vary significantly due to strain diversity, sample growth and environmental condition. Thus, it is important to maintain comprehensive, regionally representative databases providing accurate identification and environmental monitoring.
- Library references are fundamental for identification and the results confirmed that strains sourced from a diverse collection enhance coverage of the database improving identification rates using MALDI-TOF.
- We further plan to continue add and analyze diverse strains to the MALDI-TOF library utilizing the samples collected from global labs to improve the impact of strain diversity in proteotypic identification methods.

5 ACKNOWLEDGEMENTS

A special thank you to our France team – Gabrielle Heussner, Thomas Fourneyron, and Alexandre Pellet – for their overall contributions to library development. We also extend our thanks to Ruvim Kolosey from our Delaware site for his valuable assistance in the continuous library development process.