ASMS 2024 Mass Spectral Libraries Interest Group Workshop Report

June 4, 2023, 5:45pm - 7:00pm (PST), Ballroom C

Topic: Spectral Libraries for the Masses

Organizers: Emma Rennie (Agilent)

Douglas Slotta (NIST)

Panelists: Noah Smeriglio (George Washington University, DC)

O. David Sparkman (University of Pacific and NIST)

Andrew McEachran (Agilent of behalf of MassBank team)

Marcus Ludwig (Bright Giant)

The use of mass spectral (MS) libraries has become increasingly popular for its rapid and reliable data analysis capabilities, making it an essential tool in compound identification in various fields, such as metabolomics, health sciences, pharmaceuticals, forensics, food science, agriculture, and environmental research. This successful workshop served as a forum for approximately 170 participants to explore new developments, applications, and software tools related to mass spectral libraries. It also facilitated discussions on critical aspects like the need for library development, enhancing library quality, optimizing search results, and sharing and storing public data. The workshop consisted of 3 parts: an audience survey with immediate feedback, brief informative presentations by each of the panelists, and an interactive Q/A discussion with the panel.

1. The workshop started off with Emma emphasizing the fact this workshop is now in its 2nd year as an official ASMS interest group and asking the audience to select this interest group when they register for ASMS next year. She then presented the audience poll, and moved onto showing some examples from the previous year's poll which directed the topics and selection of presenters for the 2024 workshop:
    - Easy access to banks of highly curated spectral data and orthogonal ID information.
    - More reference library spectra.
    - Theoretical spectra which give good matches.
    - Increased accessibility to cloud databases and libraries with multivendor spectra.

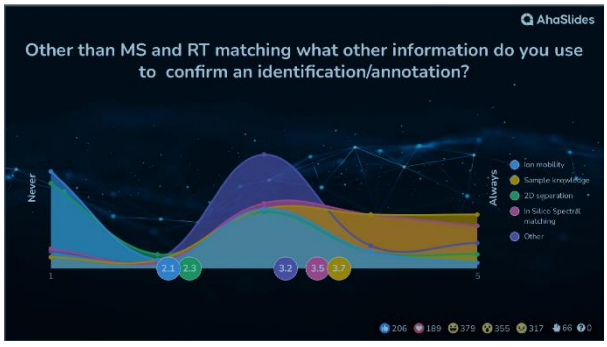
    There was enthusiastic participation from the audience, with attendees actively participating in the poll via the AhaSlides app (see survey feedback below*). The direct and genuine feedback, answers, and suggestions provided by the audience will be very helpful in developing mass spectral libraries of higher quality and practicality.
2. Next, a panel of 4 experts specializing in mass spectral library technology delivered concise presentations on recent developments and practical applications of MS libraries, along with the effective utilization of associated software tools for compound identification.
    - Noah Smeriglio spoke about their contaminant library and fasta publication system which uses GitHub for ease of access for users.
    - O. David Sparkman used the NIST Mass Spectral Search Program's Hybrid Search to identify an unknown that had no spectrum in the search library. This procedure can be used with the

NIST/EPA/NIH EI Mass Spectral Library or the NIST Tandem Library of Product-ion spectra produced by MSMS from precursor ions formed using ESI or APCI. Any library in the NIST MS Search Program format, such as the EI libraries from John Wiley and the product-ion library from MoNA can be used along with the NIST Libraries to assist in this task. More information can be found about NIST Mass Spectrometry Data Center at https://chemdata.nist.gov. A complete list of NIST authorized distributors can be found at: https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:distributors. At least one of these distributors is providing the MoNA library in the NIST format, at no additional charge, when the NIST Tandem Library offering is purchased.

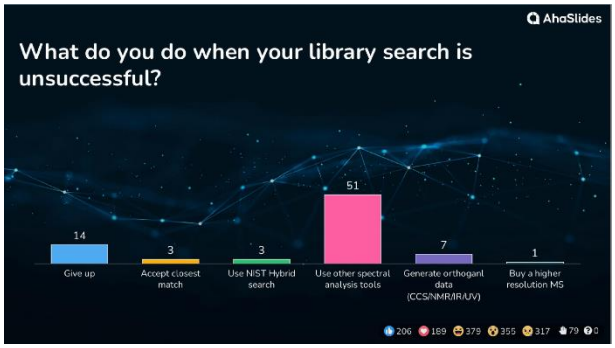
- Andrew McEachran presented on behalf of the MassBank team: MassBank Europe is a public repository for sharing mass spectral data. The data is open source and is accessible via both a web application (https://massbank.eu/MassBank/) and via download from GitHub (https://github.com/MassBank/MassBank-data/). The mass spectral community is encouraged to deposit their spectra to MassBank and can use RMassBank or certain vendor tools like Agilent's ChemVista to annotate and generate MassBank records. Finally, open data enables cross-resource integration and collaborations, as seen with the MassBank-PubChem collaborative effort to create a workflow to display MassBank spectra in PubChem (10.1039/D3EM00181D). Access the slides here: 10.5281/zenodo.8362116
- Marcus Ludwig presented "Integrating spectral matching with structure database search using SIRIUS". SIRIUS searches the query MS/MS spectrum in a molecular structure database and thus can identify compounds missing from spectral libraries. To assess the plausibility of a database hit, we can search for analogues in the spectral library. By using a combinatorial fragmenter, we can annotate peaks in the MS/MS spectrum with potential fragments of the database hit an of the analogue compound - common fragment explanations thus can increase confidence in the database hit. SIRIUS software: https://github.com/sirius-ms/sirius.

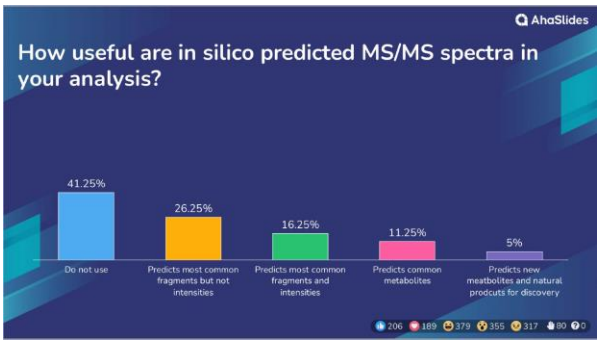
3. Afterwards, there was an interactive Q&A discussion between the panelists and audiences. Through active participation, they shared valuable insights, diverse opinions, helpful suggestions, constructive comments, concerns, and problems related to the MS library's quality, library of retention times or retention indices, library searching, false positive identification, and data sharing and storage, etc.

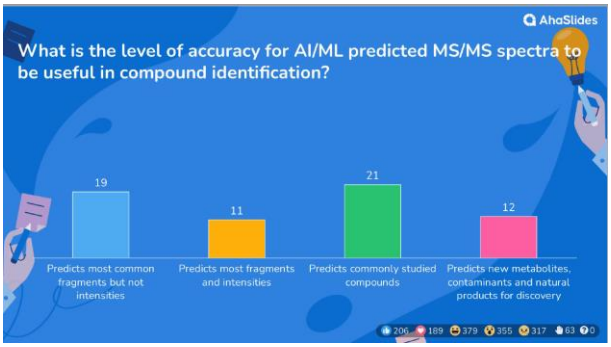
* Some survey feedbacks:

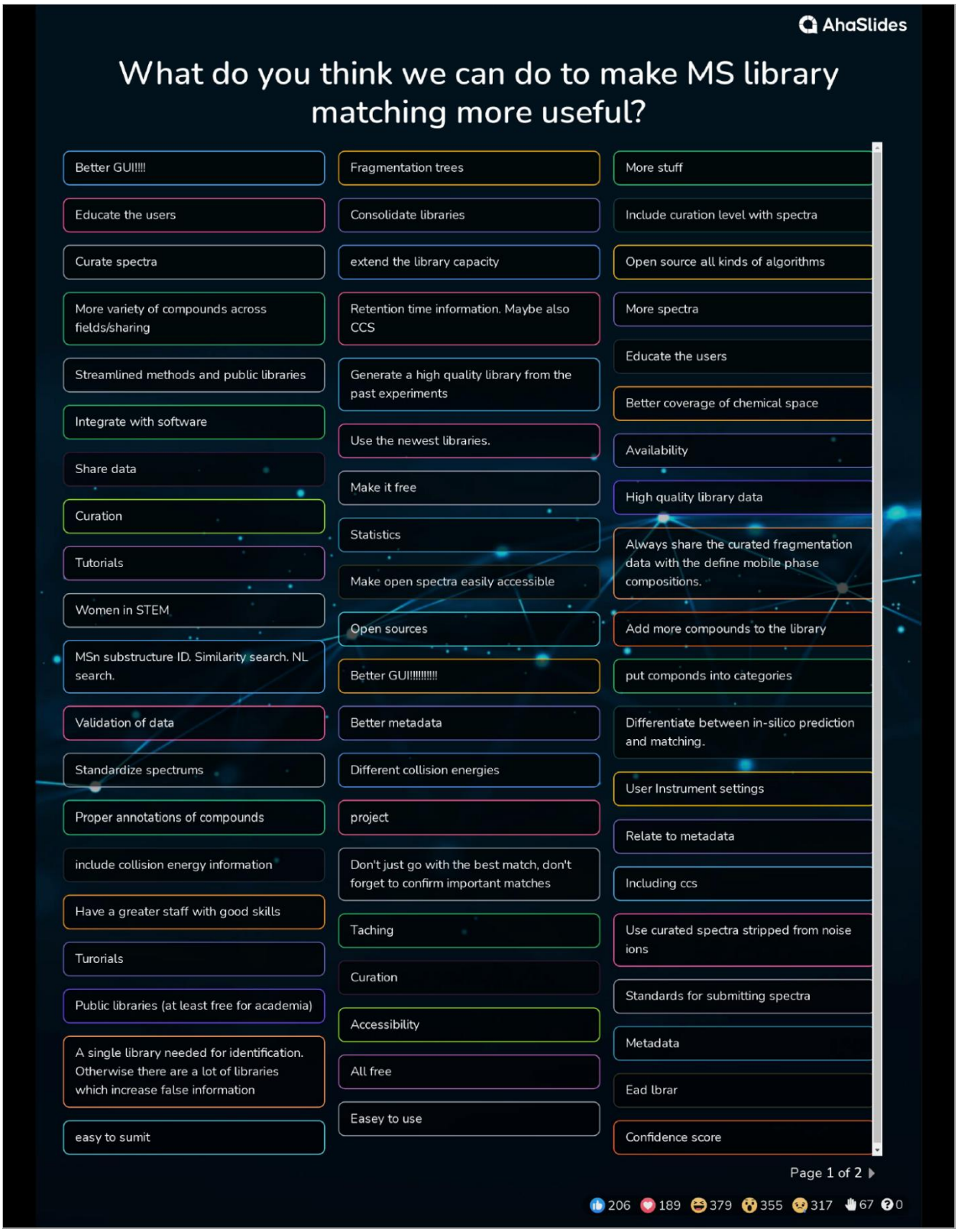


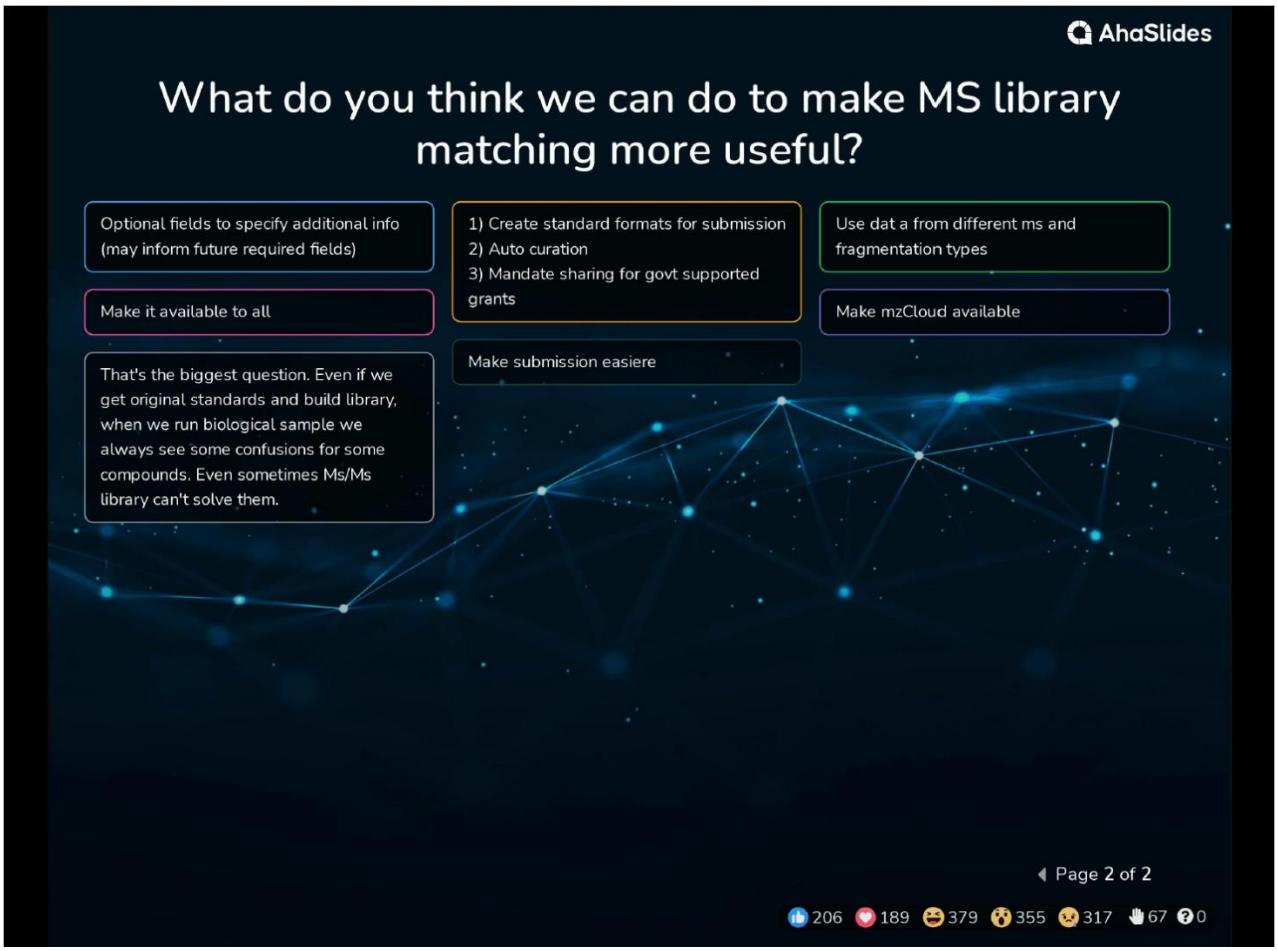

AhaSlides
How many spectra have you shared or contributed to public DB's?
56
8
1
15
0-10
10-50
50-100
>100


AhaSlides
What do you do when your library search is unsuccessful?
14
3
3
51
7
1
Give up
Accept closest match
Use NIST Hybrid search
Use other spectral analysis tools
Generate orthogant data (CCS/NMR/IR/UV)
Buy a higher resolution MS


AhaSlides
How useful are in silico predicted MS/MS spectra in your analysis?
41.25%
26.25%
16.25%
11.25%
5%
Do not use
Predicts most common fragments but not intensities
Predicts most common fragments and intensities
Predicts common metabolites
Predicts new meatbolites and natural prodcuts for discovery


AhaSlides
What is the level of accuracy for AI/ML predicted MS/MS spectra to be useful in compound identification?
19
11
21
12
Predicts most common fragments but not intensities
Predicts most fragments and intensities
Predicts commonly studied compounds
Predicts new metabolites, contaminants and natural products for discovery


AhaSlides
What do you think we can do to make MS library matching more useful?
Better GUI!!!!
Educate the users
Curate spectra
More variety of compounds across fields/sharing
Streamlined methods and public libraries
Integrate with software
Share data
Curation
Tutorials
Women in STEM
MSn substructure ID. Similarity search. NL search.
Validation of data
Standardize spectrums
Proper annotations of compounds
include collision energy information
Have a greater staff with good skills
Turorials
Public libraries (at least free for academia)
A single library needed for identification. Otherwise there are a lot of libraries which increase false information
easy to sumit
Fragmentation trees
Consolidate libraries
extend the library capacity
Retention time information. Maybe also CCS
Generate a high quality library from the past experiments
Use the newest libraries.
Make it free
Statistics
Make open spectra easily accessible
Open sources
Better GUI!!!!!!!!!
Better metadata
Different collision energies
project
Don't just go with the best match, don't forget to confirm important matches
Taching
Curation
Accessibility
All free
Easey to use
More stuff
Include curation level with spectra
Open source all kinds of algorithms
More spectra
Educate the users
Better coverage of chemical space
Availability
High quality library data
Always share the curated fragmentation data with the define mobile phase compositions.
Add more compounds to the library
put componds into categories
Differentiate between in-silico prediction and matching.
User Instrument settings
Relate to metadata
Including ccs
Use curated spectra stripped from noise ions
Standards for submitting spectra
Metadata
Ead librar
Confidence score
Page 1 of 2
206 189 379 355 317 67 0


AhaSlides
What do you think we can do to make MS library matching more useful?
Optional fields to specify additional info (may inform future required fields)
1) Create standard formats for submission
2) Auto curation
3) Mandate sharing for govt supported grants
Use dat a from different ms and fragmentation types
Make it available to all
Make mzCloud available
That's the biggest question. Even if we get original standards and build library, when we run biological sample we always see some confusions for some compounds. Even sometimes Ms/Ms library can't solve them.
Make submission easiere
Page 2 of 2
206 189 379 355 317 67 0

Panelists and workshop organizers: