ASMS 2023 Mass Spectral Libraries Interest Group Workshop Report

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

Topic: Exploring the World of Mass Spectral Libraries, Library Search Software and Their Applications

Organizers: Emma Rennie (Agilent)

Xiaoyu Yang (NIST)

Melinda McFarland (FDA)

Panelists: Michael MacCoss (University of Washington)

Stephen Stein (NIST)

Lloyd Sumner (University of Missouri)

Arpana Vaniya (UC Davis)

Mingxun Wang (UC Riverside)

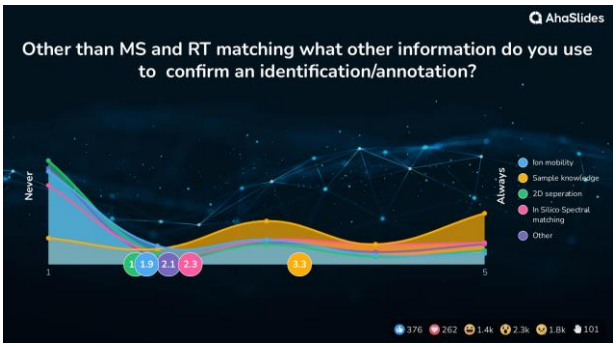
The use of mass spectral (MS) libraries has become increasingly popular for its rapid and reliable data analysis capabilities, making it an essential tool in compound identification in various fields, such as metabolomics, health sciences, pharmaceuticals, forensics, food science, agriculture, and environmental research. This successful workshop served as a forum for 300 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: exceptional survey response, brief informative presentations, and interactive Q/A discussion.

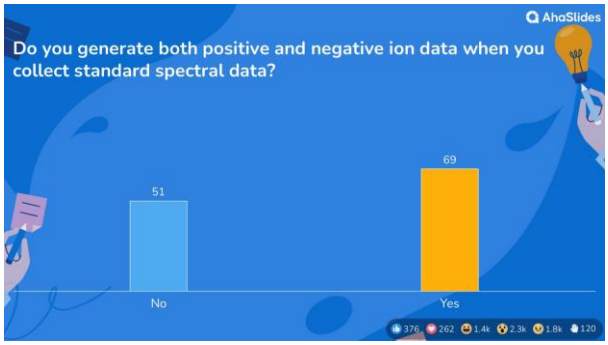
1. The workshop kicked off with Emma at the front and enthusiastic participation from the audience, with attendees actively responding to critical questions and participating in polls via the AhaSlides app (see representative survey feedback below*). The direct and genuine feedback, answers, and suggestions provided by the audience will be very helpful in developing mass spectral libraries of higher quality and practicality.
2. Next, a panel of 5 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. Professor Lloyd Sumner provided insights into metabolite identification, focusing on the utilization of mass spectral libraries, computational prediction, and empirical UHPLC-MS-SPE-NMR techniques. He also highlighted the need for spectral sharing to advance the development of comprehensive mass spectral libraries for the mass spectrometry community. Professor Michael MacCoss focused on the integration of retention time in mass spectral libraries in the Skyline software package, enabling more accurate peptide identification. Dr. Arpana Vaniya talked about the advantages and utility of software tools like MS Dial, Entropy searching software, and MassBank for compound identification in untargeted metabolomics. Professor Mingxun Wang discussed how analog searching in a tandem mass spectral library improved the identification rate in small molecule identification and key hurdles in Machine Learning for spectral prediction. Finally, Dr. Stephen Stein introduced a novel library searching strategy called "Prior Probability" that aims to enhance compound identification in

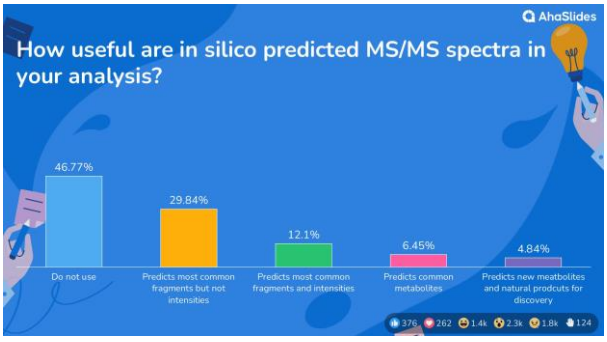
library searches. He also mentioned the newly released NIST23 mass spectral libraries as a valuable resource for more accurate compound identification.

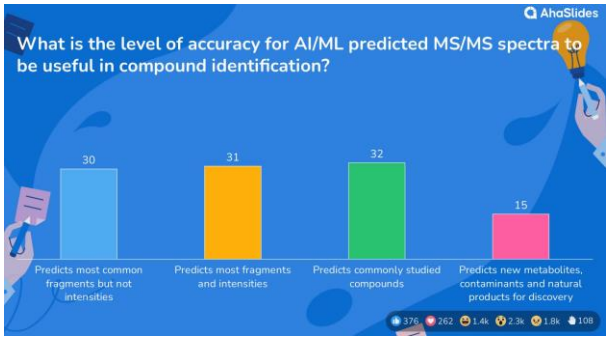
3. Afterwards, there was an interactive Q&A discussion between the panelists and audiences. Through active participation, they shared valuable insights, diverse opinions, helpful suggestions, constructive comments, concerns, and problems related to the MS library's quality, library of retention time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

* Some survey feedbacks:

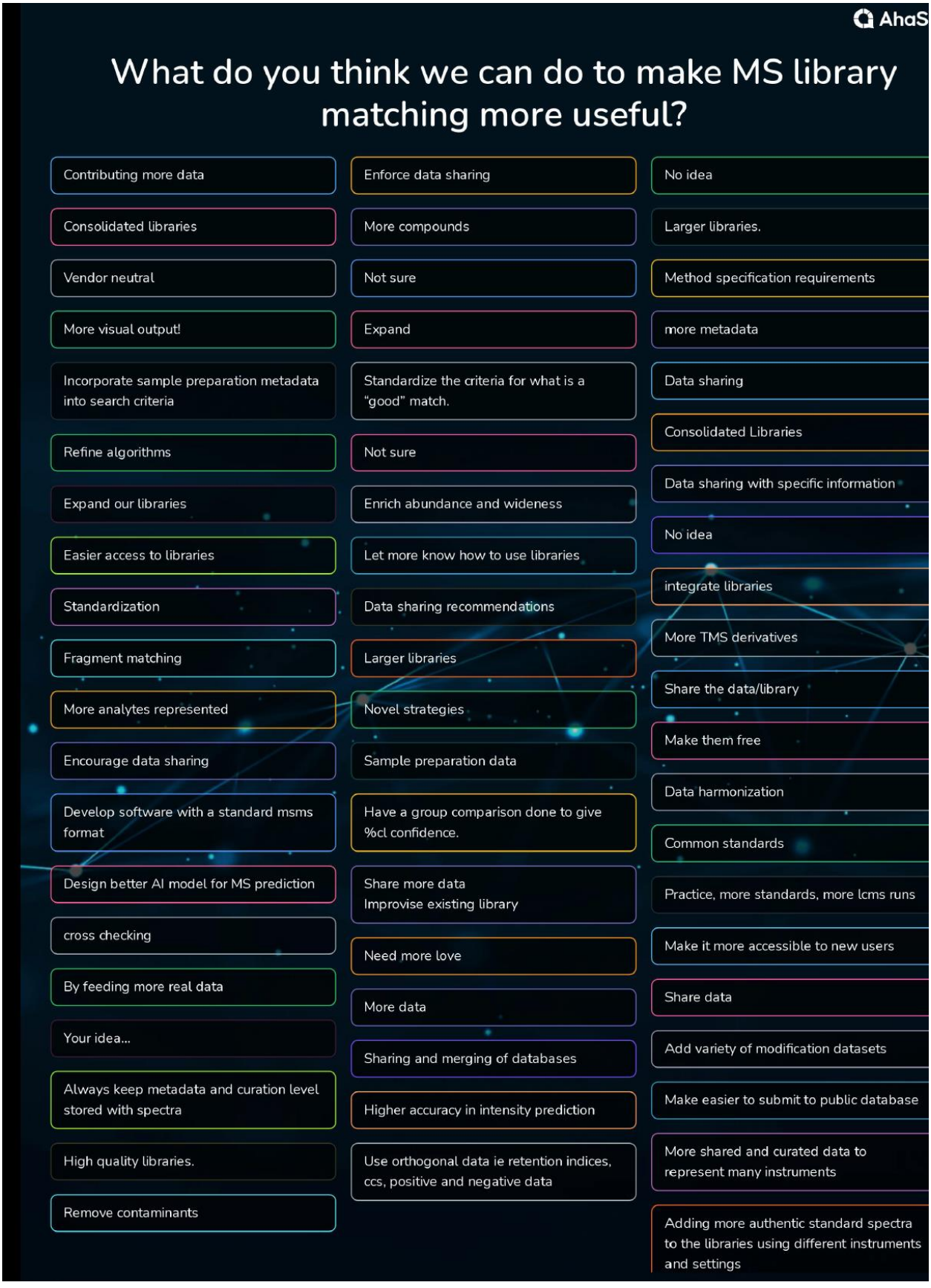


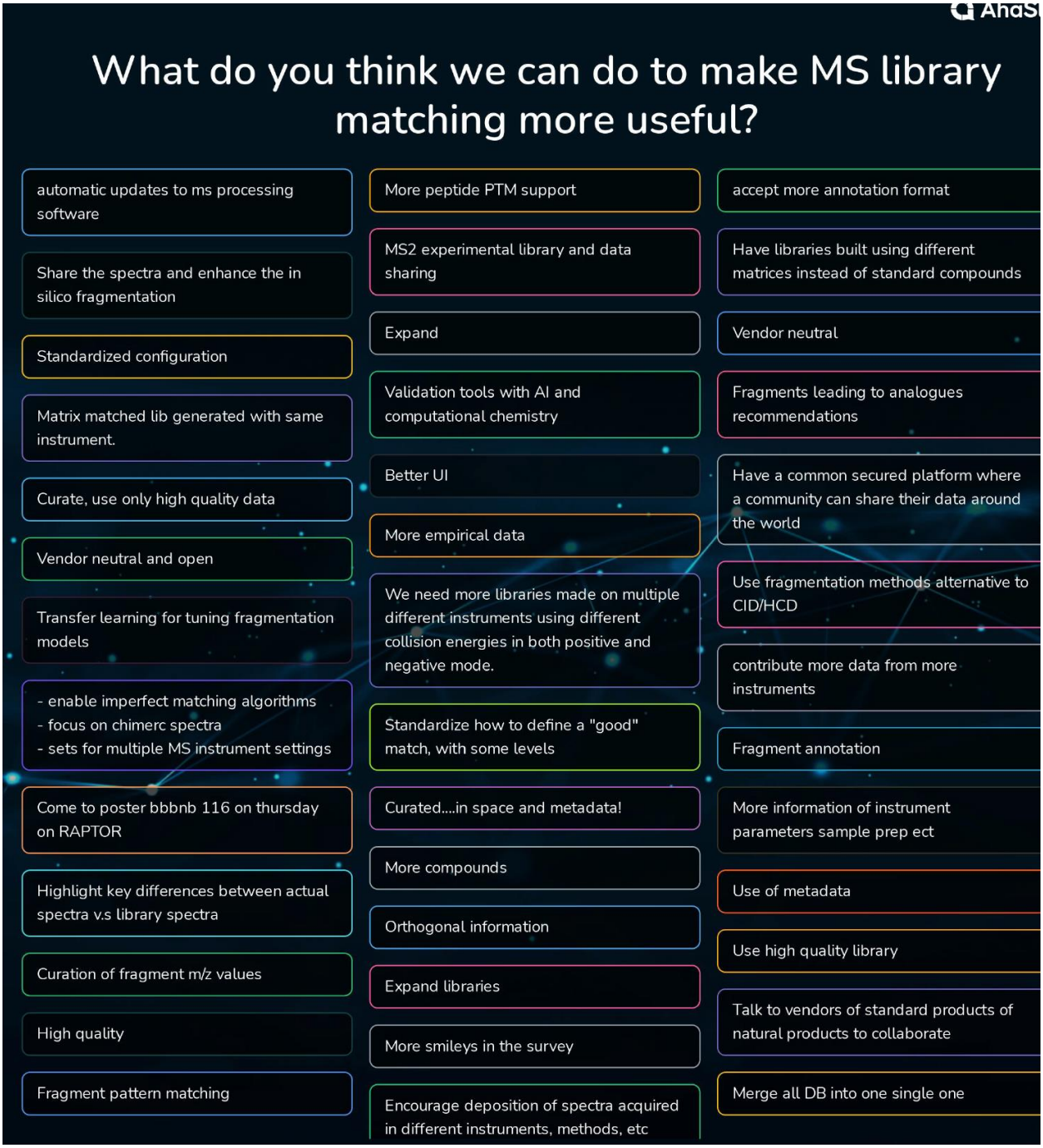







AhaSlides
What do you do when your library search is unsuccessful?
33
29
23
26
8
Give up
Accept closest match
Use molecular networking tools
Generate orthogonl data (CCS/NMR/IR/UV)
Buy a higher resolution MS
376
262
1.4k
2.3k
1.8k
119


AhaS
What do you think we can do to make MS library matching more useful?
Contributing more data
Consolidated libraries
Vendor neutral
More visual output!
Incorporate sample preparation metadata into search criteria
Refine algorithms
Expand our libraries
Easier access to libraries
Standardization
Fragment matching
More analytes represented
Encourage data sharing
Develop software with a standard msms format
Design better AI model for MS prediction
cross checking
By feeding more real data
Your idea...
Always keep metadata and curation level stored with spectra
High quality libraries.
Remove contaminants
Enforce data sharing
More compounds
Not sure
Expand
Standardize the criteria for what is a "good" match.
Not sure
Enrich abundance and wideness
Let more know how to use libraries
Data sharing recommendations
Larger libraries
Novel strategies
Sample preparation data
Have a group comparison done to give %cl confidence.
Share more data
Improvise existing library
Need more love
More data
Sharing and merging of databases
Higher accuracy in intensity prediction
Use orthogonal data ie retention indices, ccs, positive and negative data
No idea
Larger libraries.
Method specification requirements
more metadata
Data sharing
Consolidated Libraries
Data sharing with specific information
No idea
integrate libraries
More TMS derivatives
Share the data/library
Make them free
Data harmonization
Common standards
Practice, more standards, more lcms runs
Make it more accessible to new users
Share data
Add variety of modification datasets
Make easier to submit to public database
More shared and curated data to represent many instruments
Adding more authentic standard spectra to the libraries using different instruments and settings


What do you think we can do to make MS library matching more useful?
automatic updates to ms processing software
Share the spectra and enhance the in silico fragmentation
Standardized configuration
Matrix matched lib generated with same instrument.
Curate, use only high quality data
Vendor neutral and open
Transfer learning for tuning fragmentation models
- enable imperfect matching algorithms
- focus on chimerc spectra
- sets for multiple MS instrument settings
Come to poster bbbnb 116 on thursday on RAPTOR
Highlight key differences between actual spectra v.s library spectra
Curation of fragment m/z values
High quality
Fragment pattern matching
More peptide PTM support
MS2 experimental library and data sharing
Expand
Validation tools with AI and computational chemistry
Better UI
More empirical data
We need more libraries made on multiple different instruments using different collision energies in both positive and negative mode.
Standardize how to define a "good" match, with some levels
Curated....in space and metadata!
More compounds
Orthogonal information
Expand libraries
More smileys in the survey
Encourage deposition of spectra acquired in different instruments, methods, etc
accept more annotation format
Have libraries built using different matrices instead of standard compounds
Vendor neutral
Fragments leading to analogues recommendations
Have a common secured platform where a community can share their data around the world
Use fragmentation methods alternative to CID/HCD
contribute more data from more instruments
Fragment annotation
More information of instrument parameters sample prep ect
Use of metadata
Use high quality library
Talk to vendors of standard products of natural products to collaborate
Merge all DB into one single one

Panelists and workshop organizers: