EuBIC-MS Annual report 2022-2023

The European Bioinformatics Community for mass spectrometry (EuBIC-MS) is the EuPA initiative for mass spectrometry (MS)-related bioinformatics. Its aim is to improve bioinformatics for MS-based research through the setup of community-driven dynamics, thereby improving collaboration, funding, publication, and training activities. Through its different activities, EuBIC-MS aims to raise awareness for the benefits of open science and continuously supports open software. It was started in November 2015 (Vaudel, 2016) to bring together the MS and bioinformatics communities in an open, collaborative and constructive environment. The initiative is carried enthusiastically by the bioinformatics community and welcomes everyone willing to help.

EuBIC-MS organizes a dedicated yearly conference, bringing together researchers in the fields of mass spectrometry and bioinformatics, junior scientists and industry partners from across Europe. Every year, this conference alternates between a Winter School with keynote talks and workshops, and a Developers Meeting consisting of collaborative project sessions with the participants (hackathons). EuBIC-MS also actively contributes to bioinformatics hubs, and provides various workshops and presentations during international MS (e.g. annual conference of American Society of Mass Spectrometry) and proteomics (e.g. Human Proteome Organization congress, EuPA congress) conferences. Additionally, EuBIC-MS coordinates research projects conducted by the community.

Here is the report of our activities and achievements between October 2022 and September 2023.

Winter Schools

The EuBIC-MS Winter Schools on computational MS are sponsored by the European Proteomics Association (EuPA) and several other parties, including but not limited to companies with interest in proteomics and mass spectrometry. They bring together scientists from both academia and industry to present and discuss their research in workshops, keynote lectures, flash talks and poster presentations. We hosted three Winter Schools, the first one in Semmering, Austria in 2017 (Willems et al., 2017), the second in Zakopane, Poland in 2019 (Kopczynski et al., 2019), and the third in March 2022 at the Instituto Gulbenkian de Ciência in Oeiras

(Portugal). We are currently planning the next Winter School that will take place in Winterberg (Germany) the 15th-19th of January 2024. It is sponsored by BSi, MSAID, Evosep, Matrix Science and de.NBI.

Developers Meetings

The EuBIC-MS Developers Meetings are dedicated to computer scientists and developers in the field of bioinformatics applied to MS, where they can discuss and work together in an open and constructive spirit. The program is split between keynote lectures and multiple hackathon sessions where the participants develop bioinformatics tools and resources that address outstanding needs in the bioinformatics community and among biologists using MS data. Our first Developers Meeting took place in Ghent, Belgium in 2018 (Willems et al., 2018), the second took place in Nyborg (Denmark) in 2020 (Ashwood et al., 2020), and the third took place this year from 15-20 January at the ETH meeting platform Congressi Stefano Franscini (CSF), Monte Verità, Ticino (Switzerland).

Organization of the next meetings

This year, we set up an online repository for potential organizers to propose the location of the next EuBIC-MS meeting. A proposition has already been made to organize the Developers Meeting 2025 in the Netherlands, stay tuned for more information.

Community achievements and outreach

Guidelines for Reproducible MS-based Experiments

There is a growing concern about a lack of reproducibility in science that also affects bioinformatics and mass spectrometry (MS)-based technologies. EuBIC-MS and others endeavored to establish a set of guidelines that should be followed to ensure transparency when publishing MS results (eubic.github.io/ReproducibleMSGuidelines). These contain the minimal information required in the bioinformatic methods section of such manuscripts, and a well-defined checklist of requirements to help authors, reviewers, and editors to assess bioinformatic analyses. They also provide the required extended information data (supplementary) that needs to accompany any manuscript in order to ensure its reproducibility.

EuBIC-MS Annual report 2022-2023

After a very productive start, this project has been inactive for the last four years. Guidelines are still very much needed in the field, and we plan to amend/expand these in the year to come.

The SDRF-Proteomics format

The EuBIC-MS community collaborated with the Human Proteome Organization Proteomics Standards Initiative (HUPO-PSI) (Deutsch et al., 2017) to define a standard for metadata annotation of proteomics data: The Sample and Data Relationship Format for Proteomics file format (SDRF-Proteomics) (Dai et al., 2021). This should increase data interpretability, reproducibility, as well as facilitating data reuse in an automated fashion. Unfortunately, community-wide adoption has been slow due to its complexity and the lack of user-friendly annotation tools. To address this, the CompOmics group developed this year a browser-based, intuitive application for metadata annotation called lesSDRF (Claeys et al., 2023). This tool allows users with minimal knowledge on SDRF-Proteomics to correctly annotate their data prior to submission to a public repository. In order to achieve this community-wide adoption of SDRF-Proteomics, several tool developers have been approached to include SDRF compatibility in their tools. So far the following tools have joined forces: FragPipe, MSStats, MSqRob, Mascot and PeptideShaker.

In addition to this, we want to further develop SDRF-Proteomics to keep up with the ever-evolving field of proteomics. Currently, there are active discussions ongoing on inclusion of metaproteomics, metabolomics and DIA acquisition on the dedicated GitHub repository (github.com/bigbio/proteomics-sample-metadata). This will expand to single-cell proteomics and immunopeptidomics so that comprehensive metadata annotation for niche topics in proteomics can be assured.

Rusteomics

The aim of Rusteomics is to build, in the Rust programming language, a collaborative and community-driven toolbox to efficiently process MS data. The 2023 edition of the EuBIC Developers Meeting provided an opportunity to establish a task force for the Rusteomics project. The hackathon jump started the project with the development of an application that creates a spectral library (NIST MSP format) from a peak list file (MGF format) and peptide-spectrum match (PSM) data (psm_utils TSV format). The prototyped tool reannotates MS/MS spectra

based on matched peptide sequences. This Rust implementation proved to be significantly faster than its Python equivalent, with potential for further improvements. Initial attempts to integrate the C# ThermoRawFileReader were inconclusive. Organizational groundwork was laid, including setting up software modules, a GitHub organization, and essential project files. With a new team of maintainers, Rusteomics is poised for future expansion. However, despite the strong interest of project contributors, the project hasn't made significant strides over the past year. The project management needs enhancements to facilitate regular meetings and outline the necessary tasks to be performed.

ProteoBench

ProteoBench is a community-curated benchmark platform for comparing MS data analysis workflows. It started early 2023, and grew to gather more than thirty members on its dedicated EuBIC-MS slack channel. In February, we gathered for a very productive 5-days workshop organized by Matthias Mattanovich and Holda Anhagho in Copenhagen (Denmark). It was hosted in the University of Copenhagen and fully funded by the Danish Data Science Academy (DDSA). We also met for two days in September, just before the EuBIC-MS organizational meeting.

ProteoBench has a very active github repository (github.com/Proteobench/ProteoBench) that contains a soon-to-be working version of our first benchmark module focused on the comparison of workflows dedicated to the quantitative analysis of DDA data. It also contains detailed documentation on how to contribute to the project, how to use the platform, and how to propose new modules. We plan to present the project at the next EuBIC-MS Winter School and have an official launch for beta testing soon after in a webinar.

EuBIC-MS seminars 2023

In 2023, EuBIC-MS was present at the BSPR/EuPA conference in Newcastle. On the first day (17th, 2023), YPIC organized a YPIC Workshop inviting initiatives and other speakers. EuBIC-MS presented an introduction to mass spectrometry and basic concepts of data analysis from a bioinformatics point of view in a 90-minute block. This presentation was split into two parts, a theoretical background presentation describing concepts of data analysis and a more practical presentation, listing various tools from members of EUBIC-MS as well as from other

developers. The slides have been made publicly available. The YPIC Workshops were well attended and offered a great start for the BSPR/EuPA conference.

EuPA early career researcher (ECR) day 2023

The EuPA ECR day took place online on the 5th of October 2023. Several EuBIC-MS members, including Wout Bittremieux as winner of the EuPA Bioinformatics for Mass Spectrometry award, interacted with participants of the ECR day present in a dedicated space in Proteorealm (set up with gathertown.org).

Our website and resources

In partnership with the EuPA Educational Committee (EC), the Proteomics Academy web resource (www.proteomics-academy.org) has become the central communication portal for EuBIC-MS since its creation in 2015. To increase the visibility of EuBIC-MS itself, we also maintain a website dedicated to the initiative at eubic-ms.org. There, people find all information regarding past and upcoming EuBIC-MS activities, including this annual report. We recently added information on how we operate and how to acknowledge EuBIC-MS in publications.

This year, we officially transferred the proteomics Q&A page that was a google group to a dedicated discussion page on the EuBIC-MS github (github.com/orgs/EuBIC/discussions/categories/q-a).

How to get involved?

The EuBIC community is open to everybody working in computational proteomics and mass spectrometry. You can join us by using the "become a member" form available on the eubic-ms.org website. An automatically generated email will be sent out to invite you to join the EuBIC-MS Slack workspace (http://eubic.slack.com), which is our extensively used platform to interact with other members. Additionally, you can follow and retweet our latest activities from our Twitter account (@EuBIC_ms). By joining us, you will be part of a community effort, and thus contribute to the sustainability of these scientific activities, which we expect to foster the development of the computational MS field.

Acknowledgments

References

Ashwood C, Bittremieux W, Deutsch EW, Doncheva NT, Dorfer V, Gabriels R, Gorshkov V, Gupta S, Jones AR, Käll L, Kopczynski D, Lane L, Lautenbacher L, Legeay M, Locard-Paulet M, Mesuere B, Perez-Riverol Y, Netz E, Pfeuffer J, Sachsenberg T, Salz R, Samaras P, Schiebenhoefer H, Schmidt T, Schwämmle V, Soggiu A, Uszkoreit J, Van Den Bossche T, Van Puyvelde B, Van Strien J, Verschaffelt P, Webel H, Willems S. Proceedings of the EuBIC-MS 2020 Developers' Meeting. EuPA Open Proteomics. 2020, volume 24. https://doi.org/10.1016/j.euprot.2020.11.001.

Claeys, T., Van Den Bossche, T., Perez-Riverol, Y. et al. lesSDRF is more: maximizing the value of proteomics data through streamlined metadata annotation. Nat Commun 14, 6743 (2023). https://doi.org/10.1038/s41467-023-42543-5

Dai C, Füllgrabe A, Pfeuffer J, Solovyeva EM, Deng J, Moreno P, Kamatchinathan S, Kundu DJ, George N, Fexova S, Grüning B, Föll MC, Griss J, Vaudel M, Audain E, Locard-Paulet M, Turewicz M, Eisenacher M, Uszkoreit J, Van Den Bossche T, Schwämmle V, Webel H, Schulze S, Bouyssié D, Jayaram S, Duggineni VK, Samaras P, Wilhelm M, Choi M, Wang M, Kohlbacher O, Brazma A, Papatheodorou I, Bandeira N, Deutsch EW, Vizcaíno JA, Bai M, Sachsenberg T, Levitsky LI, Perez-Riverol Y. A proteomics sample metadata representation for multiomics integration and big data analysis. Nat Commun. 2021 Oct 6;12(1):5854. doi: 10.1038/s41467-021-26111-3. PMID: 34615866.

Deutsch EW, Orchard S, Binz PA, Bittremieux W, Eisenacher M, Hermjakob H, Kawano S, Lam H, Mayer G, Menschaert G, Perez-Riverol Y, Salek RM, Tabb DL, Tenzer S, Vizcaíno JA, Walzer M, Jones AR. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work. J Proteome Res. 2017 Dec 1;16(12):4288-4298. doi: 10.1021/acs.jproteome.7b00370. Epub 2017 Sep 15. PMID: 28849660; PMCID: PMC5715286.

Kopczynski, D., Bittremieux, W., Bouyssié, D., Dorfer, V., Locard-Paulet, M., Van Puyvelde, B., Schwämmle, V., Soggiu, A., Willems, S., Uszkoreit, J., 2019. Proceedings of the EuBIC Winter School 2019. EuPA Open Proteomics. https://doi.org/10.1016/j.euprot.2019.07.002

Luo X, Bittremieux W, Griss J, Deutsch EW, Sachsenberg T, Levitsky LI, Ivanov MV, Bubis JA, Gabriels R, Webel H, Sanchez A, Bai M, Käll L, Perez-Riverol Y. A Comprehensive Evaluation of Consensus Spectrum Generation Methods in Proteomics. J Proteome Res. 2022 Jun;21(6) 1566-1574. doi:10.1021/acs.jproteome.2c00069. PMID: 35549218; PMCID: PMC9171829.

Vaudel, M., 2016. The EuPA Bioinformatics Community (EuBIC) initiative. EuPA Open

Proteomics 11, 34. https://doi.org/10.1016/j.euprot.2016.03.009

Willems, S., Bouyssié, D., David, M., Locard-Paulet, M., Mechtler, K., Schwämmle, V., Uszkoreit, J., Vaudel, M., Dorfer, V., 2017. Proceedings of the EuBIC Winter School 2017. J. Proteomics 161, 78–80. https://doi.org/10.1016/j.jprot.2017.04.001

Willems, S., Bouyssié, D., Deforce, D., Dorfer, V., Gorshkov, V., Kopczynski, D., Laukens, K., Locard-Paulet, M., Schwämmle, V., Uszkoreit, J., Valkenborg, D., Vaudel, M., Bittremieux, W., 2018. Proceedings of the EuBIC developer's meeting 2018. J. Proteomics 187, 25–27. https://doi.org/10.1016/j.jprot.2018.05.015