Boosting metaproteomics identification rates and taxonomic specificity with MS²Rescore

Tim Van Den Bossche[1,2], Arthur Declercq[1,2], Ralf Gabriels[1,2], Tanja Holstein[1,2], Bart Mesuere[3], Thilo Muth[4], Pieter Verschaffelt[1,3] and Lennart Martens[1,2,5]*

(1) VIB - UGent Center for Medical Biotechnology, VIB, 9052 Ghent, Belgium
(2) Department of Biomolecular Medicine, Faculty of Medicine and Health Sciences, Ghent University, 9052 Ghent, Belgium
(3) Department of Mathematics, Computer Science and Statistics, Ghent University, 9000 Ghent, Belgium
(4) Data Competence Center MF 2, Robert Koch Institute, Berlin, Germany
(5) BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC UMR 7178, University of Strasbourg, CNRS, ProFI FR2048, Strasbourg, France

* Corresponding author: Lennart Martens (Lennart.Martens@UGent.be)

Abstract

Background: Metaproteomics, the study of the collective proteome within microbial ecosystems, has gained increasing interest over the past decade. However, peptide identification rates in metaproteomics remain low compared to single-species proteomics. A key challenge is the limited sequence resolution of current identification algorithms, which were primarily designed for single-species analyses. To address this, we applied the machine learning-driven MS²Rescore tool to multiple metaproteomics datasets from diverse microbial environments and benchmark studies.
Results: We demonstrate that MS²Rescore outperforms typical metaproteomics identification workflows. It significantly increases peptide identification rates compared to standard database searches, even when compared to Percolator rescoring. Moreover, it enables lowering the false discovery rate (FDR) to 0.1% with minimal sensitivity loss, a substantial improvement over the 1% or 5% FDR thresholds commonly used in metaproteomics, in turn leading to greater specificity in downstream taxonomic annotation with Unipept.
Conclusions: Our findings show that MS²Rescore substantially improves peptide identification sensitivity as well as specificity in metaproteomics, and delivers improved taxonomic specificity. This advancement results in a more reliable downstream taxonomic analysis, reinforcing the potential of machine learning-based rescoring in metaproteomics research.

Background

Metaproteomics, the study of the collective proteome within microbial ecosystems, complements metagenomics and metatranscriptomics by providing direct functional insights into complex microbial communities. It has been applied across diverse environments, including the human gut [1], biogas plants [2], and soil [3]. However, the field continues to suffer from lower identification rates compared to single-species proteomics. The primary challenge lies in the limited sequence resolution and statistical validation of current identification algorithms, which were originally designed for single-species datasets [4,5]. Additionally, as datasets increase in complexity, taxonomic annotation becomes more difficult due to sequence homology issues [6]. This necessitates a more refined peptide-level analysis to improve resolution.

To address these challenges, we applied the in-house developed machine learning-based MS²Rescore algorithm [7] to multiple multi-species metaproteomics datasets. MS²Rescore enhances standard rescoring by integrating Percolator's search engine-dependent features with machine learning-derived features, such as MS2 peak intensity predictions from MS²PIP [8] and retention time predictions from DeepLC [9]. This approach not only increases peptide-spectrum match (PSM) identification rates, as previously demonstrated [10,11], but also enables a more stringent FDR threshold of 0.1%, improving upon the commonly used 1% or 5% thresholds in metaproteomics.

In addition to improving PSM identification rates, MS²Rescore enhances taxonomic annotation by reducing false positives while increasing the detection of true positive taxa. These advancements highlight the potential of machine learning-based rescoring to overcome key limitations in metaproteomics data analysis, ultimately improving the accuracy of taxonomic classification [12].

Methods & Implementation

Datasets

To evaluate MS²Rescore for metaproteomics, we conducted three experiments using different datasets. First, we reanalyzed mass spectrometry files from the CAMPI study (PXD023217) [6], the first multi-laboratory benchmark study in metaproteomics. This experiment assessed how MS²Rescore performs compared to traditional search engines in terms of peptide identification rates. Second, we analyzed four raw files (termed "Mixes") from the iPRG 2020 (Phase I) metaproteomics study (PXD034795) [13] to determine the taxonomic specificity of MS²Rescore.

Third, we reanalyzed publicly available datasets from PRIDE [14]: the inflammatory bowel disease (IBD) dataset consisting of 77 raw files (3.3 million spectra) (PXD010371) [1], the biogas plant (BGP) dataset with 441 raw files (14.3 million spectra) (PXD009349) [2], and the soil dataset containing 29 raw files (1.4 million spectra) (PXD005447) [3]. These datasets were selected to assess MS²Rescore's performance across three commonly studied metaproteomics environments (human gut, biogas plants, and soil) and across different fragmentation methods, with trap-type collision induced dissociation (CID) for the IBD and BGP datasets, and beam-type CID — also known as higher-energy collisional dissociation (HCD) — for the soil dataset.

Data processing and search parameters

All raw files were converted to peak lists in Mascot Generic Format (MGF) using ThermoRawFileParser-GUI v1.2.1 [15] with the settings “native Thermo library peak picking” and “Ignore missing instrument properties”. Searches were performed using X!Tandem (Vengeance version 2015.12.1) [16] using SearchGUI version 4.0.22 [17]. For the first and third experiments, peptide identification was performed against a concatenated target/decoy database including all sequences from the original studies [1–3,6]. Since the composition of the iPRG 2020 samples was unknown (second experiment), we searched against UniRef50 [18] (39,232,797 sequences). The database therefore had to include sequences from all domains (Archaea, Bacteria, and Eukarya), as well as from viruses and prions. The UniRef50 database provides clustered sets of sequences from the UniProt Knowledgebase (including isoforms), and selected UniParc records, in order to obtain as complete as possible a coverage of the sequence space at several resolutions, while hiding redundant sequences. To ensure all contaminant proteins were included, the UniRef50 database was concatenated with the common repository of adventitious proteins (cRAP) database (116 proteins, https://www.thegpm.org/crap/). Decoy sequences were generated by reversing target sequences and added to the search databases.

For all databases, decoy sequences were generated by reversing target sequences. Identification settings were identical to those described in the original manuscripts: cleavage with trypsin with a maximum of two missed cleavages (all datasets); 10.0 ppm as MS1 tolerance (all datasets); 0.02 ppm (soil, CAMPI), 0.5 Da (IDB, BGP) or 0.6 Da (iPRG) as MS2 tolerance; Carbamidomethylation of C as fixed modification (all datasets); Oxidation of M as variable modification (all datasets); Acetylation of protein N-termini, Pyrolidone from E and Q as variable modifications during the refinement procedure of X!Tandem (all datasets). X!Tandem was used because of its memory and CPU efficiency, which are very useful assets when searching against

large databases. X!Tandem XML files were extracted from the SearchGUI results, and were used for post-processing with MS²Rescore.

MS²Rescore post-processing

Here, we used MS²Rescore v2.1.3 for post-processing, which follows several steps. First, all rank 1 PSMs, including decoy PSMs, were extracted from X!Tandem XML files using the Pyteomics tandem parser [19]. Next, search engine-dependent features were parsed using tandem2pin from Percolator, and MS²PIP [8] was employed to generate predicted MS2 spectra to extract intensity-based features [20]. Additionally, DeepLC [9] was used to compute retention time-based features based on observed and predicted retention times. The extracted search engine, MS²PIP, and DeepLC features were then combined into a custom Percolator input (.pin) file and rescored using Percolator.

MS²Rescore is available as a Python package at https://github.com/compomics/MS2Rescore and includes a Graphical User Interface (GUI). The tool is open-source under the Apache 2.0 license.

Data analysis

To compare X!Tandem search results with rescoring approaches, q-values were computed using Pyteomics [19]. Pyteomics and Percolator outputs were analyzed using an interactive Python notebook for statistical evaluations and visualization. For the first experiment, unique peptides identified by MS²Rescore and the search engines used in the CAMPI study were visualized using UpSet plots (UpSetPlot package: https://upsetplot.readthedocs.io). Peptides were compared without modifications, with leucine (L) and isoleucine (I) considered equivalent (converted to J). Peptide identifications from the CAMPI study were retrieved from Supplementary Data 15 and 16 [6].

For all experiments, the taxonomic analysis was performed based with Unipept v5.0.8 (https://unipept.ugent.be) [21–23] with the following settings: Equate I/L, no filtering of duplicate peptides, no advanced missed cleavage handling.

Results and Discussion

MS²Rescore improves peptide identification in metaproteomics

In this first experiment, we analyzed all samples using X!Tandem and MS²Rescore, then compared the identification (ID) rates at 0.1% and 1% false discovery rate (FDR) thresholds. As

in the CAMPI study, the samples were searched against both the reference database (REF) and the multi-omic database (MO). Across all samples, MS²Rescore consistently increased the identification rate. Notably, the datasets F03, F08, F09, S13, and S14 were acquired using Bruker instruments (.d files), for which specific MS²PIP models have been trained [24]**.** Despite this, even when applying Orbitrap HCD models, MS²Rescore still improved identification rates (**Figure 1**).

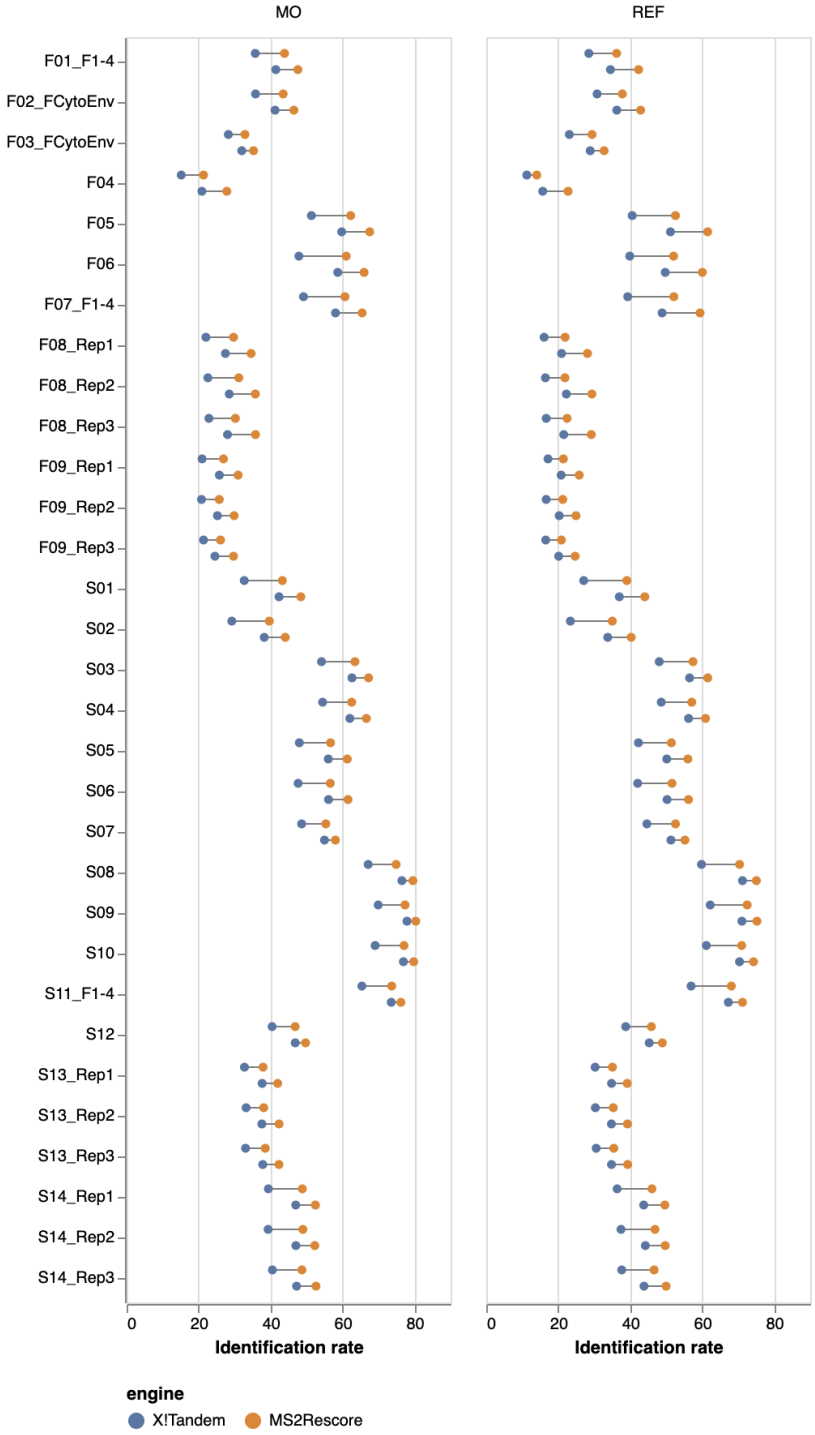

Figure 1. Identification rates for all CAMPI study samples using the multi-omic database (MO) and reference database (REF). The blue dots represent identification rates with X!Tandem, while the orange dots represent those obtained with MS²Rescore. Identification rate is calculated as the percentage of spectra that yielded a peptide identification, with values shown at 0.1% FDR (top bar) and 1% FDR (bottom bar).

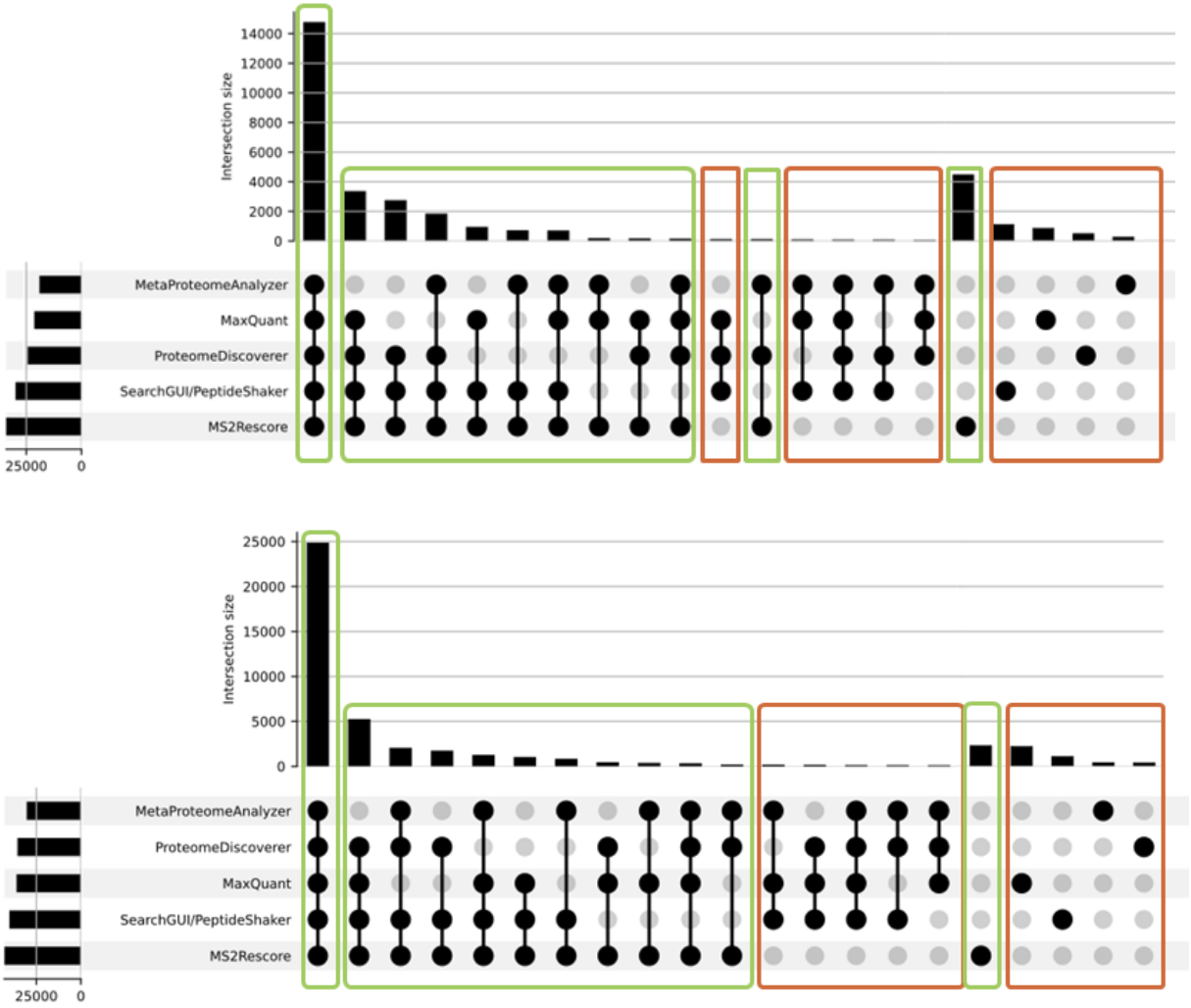

Figure 2. Comparison of MS²Rescore with the four original bioinformatics pipelines used in the CAMPI study for two sample types: the SIHUMIx sample S11 (upper panel) and fecal sample F07 (lower panel). The x-axis lists different bioinformatic pipelines: MetaProteomeAnalyzer (MPA, using X!Tandem and OMSSA), Proteome Discoverer (PD, using SequestHT), MaxQuant (MQ, using Andromeda), and SearchGUI/PeptideShaker (PS, using X!Tandem, OMSSA, MS-GF+, and Comet). Pipelines are sorted by increasing set size, where the set size represents the total number of peptides identified per tool at 1% FDR. Intersection size indicates the number of peptides shared between approaches. Green boxes highlight overlapping peptides found by MS²Rescore, whereas red boxes indicate peptides uniquely identified by other pipelines.

The primary objective of this experiment was to benchmark MS²Rescore against widely used search engines in metaproteomics, as described in the CAMPI study [6] (**Figure 2**). These pipelines include MetaProteomeAnalyzer (X!Tandem and OMSSA), Proteome Discoverer (SequestHT), MaxQuant (Andromeda), and SearchGUI/PeptideShaker (X!Tandem, OMSSA, MS-GF+, and Comet). In line with the CAMPI study, the S11 sample was searched against the reference proteome database, and the F07 was searched against the multi-omic database.

Consistent with the CAMPI study, approximately 51.5% of the identified peptides in the SIHUMIx S11 sample were detected by all bioinformatic pipelines, while this percentage dropped to 39% for the more complex fecal F07 sample. Interestingly, MS²Rescore identified 89.5% of all peptides in S11 and 90.8% in F07.

It is also noteworthy that MS²Rescore outperformed the best-performing pipeline from the CAMPI study. In the original study, SearchGUI/PeptideShaker yielded the highest number of unique peptide identifications by integrating results from four search engines (X!Tandem, OMSSA, MS-GF+, and Comet). However, our results demonstrate that MS²Rescore further increases peptide identifications, reinforcing prior suggestions from the CAMPI study that machine learning-based search engines could significantly enhance metaproteomics identification rates. The green boxes in **Figure 2** indicate peptides identified by both MS²Rescore and other pipelines, while the red boxes represent peptides uniquely found by other pipelines, potentially suggesting false positives.

Higher confidence in identifications improves taxonomic annotation specificity

In this second experiment, we evaluated the impact of MS²Rescore on taxonomic annotation. We therefore compared the number of confidently identified spectra across different q-value or FDR thresholds for three approaches: (1) No rescoring (original X!Tandem results), (2) Percolator rescoring using default search engine features, and (3) MS²Rescore, which incorporates search engine features alongside MS²PIP and DeepLC-derived features.

As in the first experiment, MS²Rescore consistently outperformed both X!Tandem and Percolator in terms of peptide identification rates. Without rescoring, X!Tandem achieved average identification rates of 2.06%, 8%, and 14.2% at 0.1%, 1%, and 5% FDR, respectively. These

relatively low rates are likely due to the vast search space, containing nearly 40 million target sequences, combined with the lower resolution of MS² spectra. By contrast, rescoring with MS²Rescore significantly improved identification rates, reaching 22.4%, 26.5%, and 29.0% at the same FDR thresholds. Figure 3 illustrates these identification rate improvements across multiple runs and methods, further demonstrating MS²Rescore's ability to improve peptide identifications even in challenging, large-scale metaproteomics datasets.

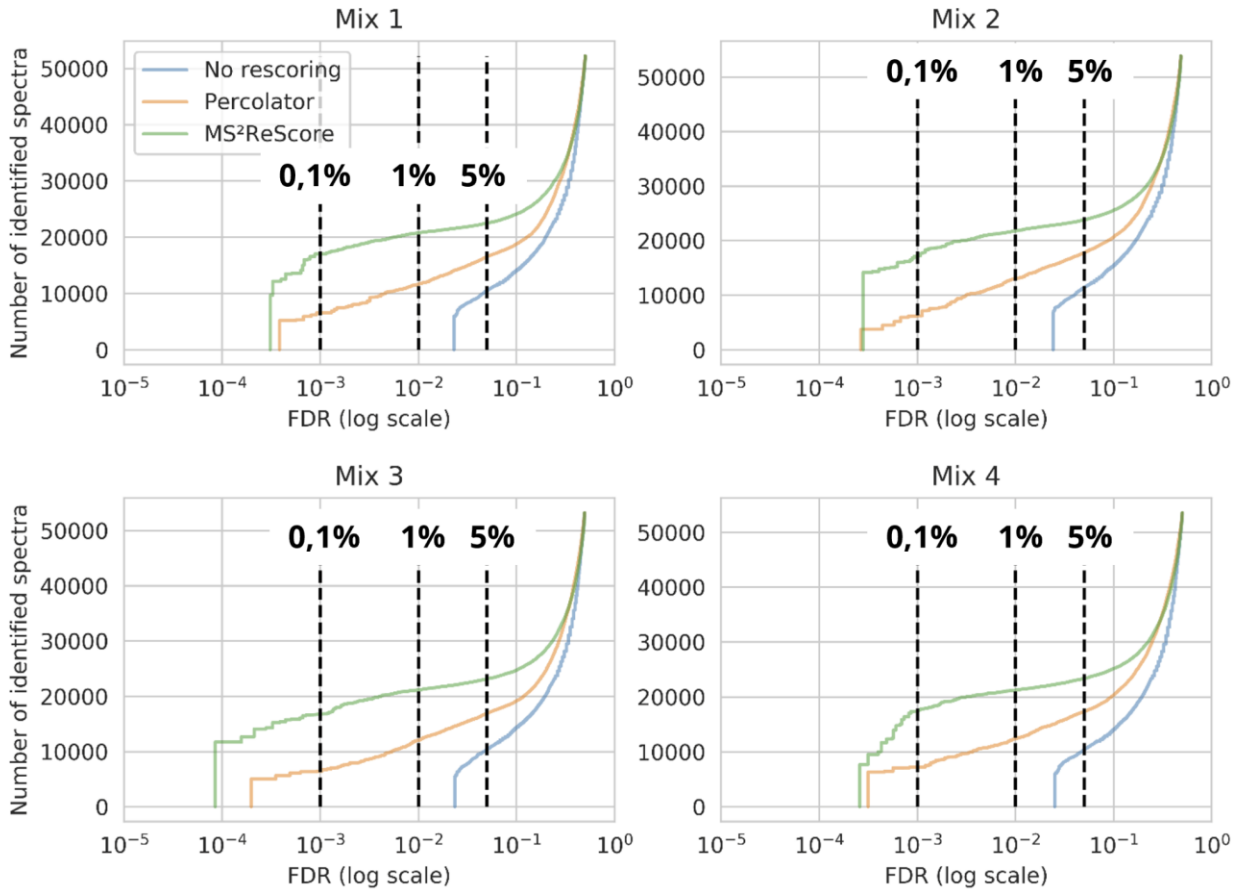

Figure 3. Number of identified spectra across different FDR thresholds for X!Tandem (no rescoring), Percolator, and MS²Rescore. Vertical dotted lines indicate FDR thresholds of 0.1%, 1%, and 5% from left to right.

To assess the effect of improved peptide identification on downstream taxonomic annotation, we analyzed the PSMs identified at both 0.1% and 1% FDR thresholds using Unipept. At 1% FDR, a high number of unexpected and biologically implausible taxa were detected (**Figure 4**), suggesting that many assignments were false positives.

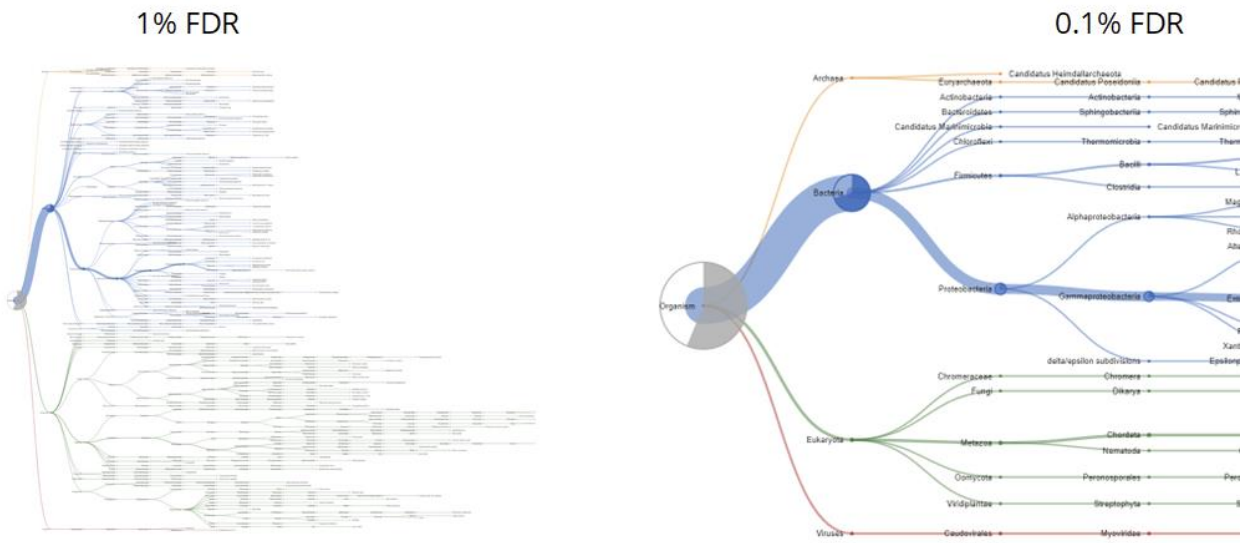

Figure 4. Unipept taxonomic annotations at different confidence levels. The left panel shows the Unipept treeview for Mix 1 at 1% FDR, while the right panel represents Mix 1 at 0.1% FDR. These treeviews are presented for illustrative purposes, demonstrating how a stricter FDR threshold reduces the number of identified taxa, improving specificity while maintaining biologically relevant taxa.

For a more robust comparison, we analyzed the genera and species present in each sample and recorded how frequently peptides were assigned to each taxonomic level. The lowest common ancestor had to be consistent across at least three of the four replicates for a taxon to be retained. **Table 1** presents the final taxonomic results, revealing that the four biological replicates were highly consistent in their taxonomic composition.

After Phase I of the iPRG 2020 study, only four species and a few contaminant proteins were confirmed to be present in the sample (indicated in green in **Table 1**): *Escherichia virus T4, Bacillus subtilis, Escherichia coli*, and *Salmonella enterica*. MS²Rescore accurately recovered these expected taxa while minimizing false-positive identifications. Contaminant proteins, indicated in yellow in **Table 1**, from *Homo sapiens* and *Sus scrofa* were detected. Additionally, peptides matching *Rattus norvegicus* were identified; however, BLAST analysis revealed that these peptides also map to *Homo sapiens* keratins. This is likely due to the way sequences are clustered in the UniRef50 database, potentially leading to misassignments in cases of high sequence similarity. Only a small number of peptides were incorrectly assigned to species not present in the sample, indicated in gray in **Table 1**, further demonstrating the higher specificity of taxonomic annotations obtained with MS²Rescore.

Table 1. Taxa detected in at least three of the four Mixes, with spectral counts for each Mix and the total across all four Mixes. Species highlighted in green were confirmed as present in the samples, yellow indicates contaminants, and gray denotes taxa not present in the sample.

genus	species (strain)	Mix_1	Mix_2	Mix_3	Mix_4	SUM
Salmonella	*Salmonella enterica (subsp. Enterica)*	476	471	501	511	1959
Rattus	*Rattus norvegicus*	99	112	95	110	416
Tequatrovirus	*Escherichia virus T4*	91	82	91	103	367
Bacillus	*Bacillus subtilis*	86	93	85	99	363
Escherichia	*Escherichia coli*	21	29	19	22	91
Klebsiella	*Klebsiella pneumoniae*	4	5	8	7	24
	Candidatus Poseidoniales archaeon	3	2	3	5	13
Homo	*Homo sapiens*	7	1	0	1	9
Sus	*Sus scrofa*	2	1	1	3	7
Parastagonospora	*Parastagonospora nodorum*	1	3	0	1	5
Pantoea	*Pantoea agglomerans*	1	1	1	1	4
	Candidatus Marinimicrobia bacterium	1	2	0	1	4

MS²Rescore increases identification rates in large, publicly available metaproteomics datasets from diverse environments

The first two experiments demonstrated a clear increase in identification rates across all assessed FDR levels (0.1%, 1%, and 5%). In this third experiment, we further evaluated MS²Rescore's performance in three large-scale studies, each representing one of three widely studied metaproteomics samples: human gut (IBD), biogas plant (BGP), and soil. **Figure 5** presents the number of confidently identified spectra across different q-values or FDR thresholds for the three approaches: X!Tandem, Percolator, and MS²Rescore. The identification rates for each dataset are summarized in **Figure 6**.

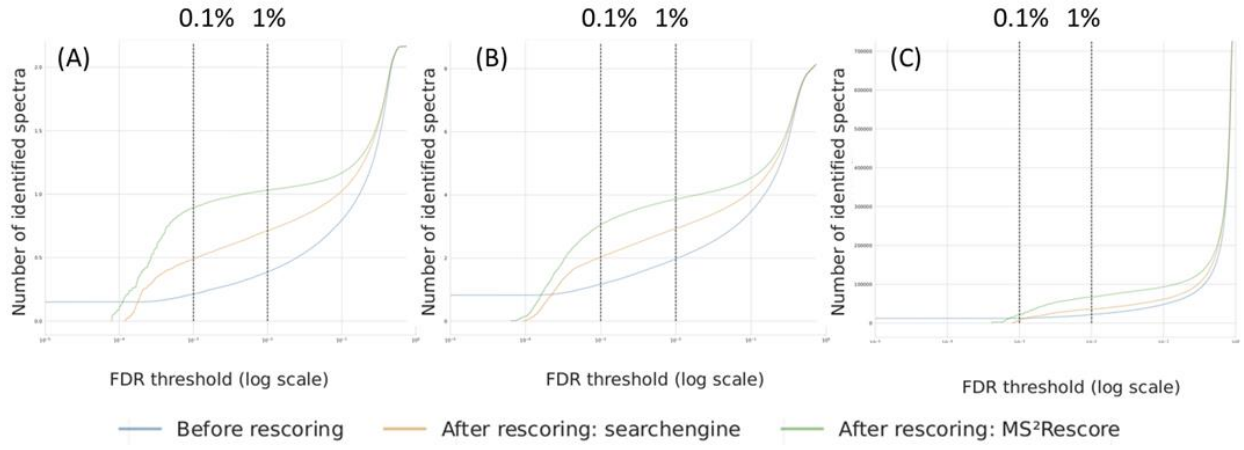

Figure 5. q-value plots for three large-scale metaproteomics datasets representing distinct microbial environments: human gut (IBD dataset, A), biogas plant (BGP, B), and soil (C). The number of identified spectra is plotted against each potential FDR threshold for three methods: X!Tandem (blue), Percolator rescoring with search engine features (orange), and MS²Rescore incorporating search engine, MS²PIP, and DeepLC features (green). Vertical dotted lines indicate FDR thresholds of 0.1% and 1%.

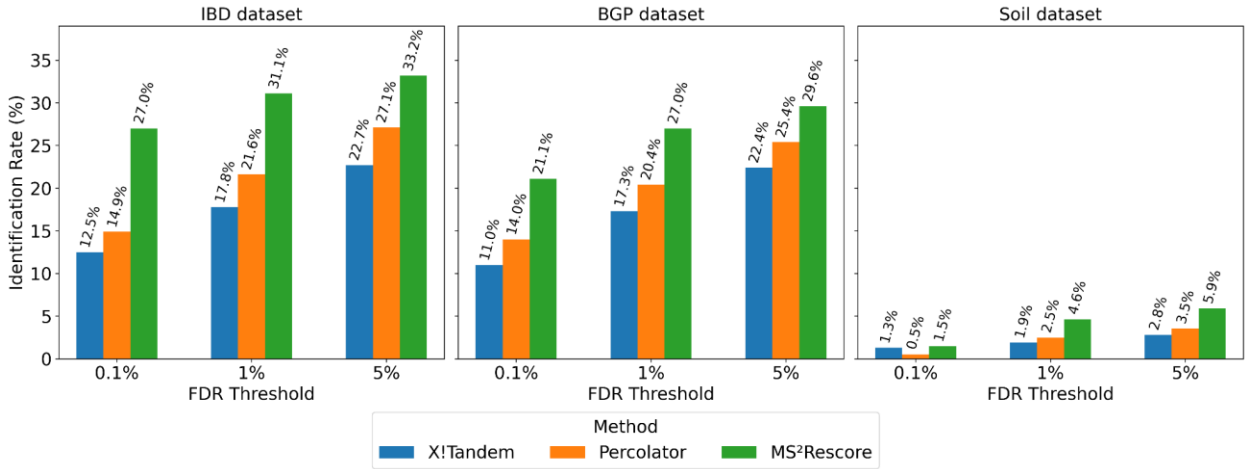

Figure 6. Identification rates for metaproteomics datasets at different FDR thresholds. The bar plots show the identification rates (%) for the IBD, BGP, and soil datasets at 0.1%, 1%, and 5% FDR thresholds using X!Tandem (blue), Percolator (orange), and MS²Rescore (green).

In the IBD dataset, rescoring with Percolator increased identification rates by approximately 20% at all FDR levels compared to X!Tandem alone. MS²Rescore further improved identification rates, yielding a 116% increase at 0.1% FDR, 75% more at 1% FDR, and 46% more at 5% FDR compared to X!Tandem; and 81%, 44%, and 23% more, respectively, compared to Percolator.

This demonstrates that while Percolator improves identification rates by incorporating search engine-dependent features, MS²Rescore significantly improves this even further by leveraging predicted fragment ion intensities (MS²PIP) and retention times (DeepLC), particularly at stricter FDR thresholds.

A similar trend was observed for the BGP and soil datasets. In the BGP dataset, Percolator improved identification rates by 27% at 0.1% FDR, 18% at 1% FDR, and 13% at 5% FDR over X!Tandem, while MS²Rescore provided further gains of 51%, 32%, and 17% over Percolator. The soil dataset, known for its high complexity, showed a more variable trend: Percolator decreased identification rates at 0.1% FDR but improved them by 32% at 1% FDR and 26% at 5% FDR, while MS²Rescore led to substantial improvements of 200%, 84%, and 67% over Percolator, demonstrating its robustness even in complex microbiome datasets.

To conclude, MS²Rescore substantially increases peptide identification rates in metaproteomics, addressing a critical challenge in the field. Notably, for most datasets, we were able to apply a more stringent 0.1% FDR threshold without sacrificing identification sensitivity, with the sole exception of the highly complex soil dataset. Even at this stricter threshold, MS²Rescore continued to outperform standard search engines, demonstrating its ability to provide high-confidence peptide identifications. This increased confidence in peptide identifications, in turn, enhances the reliability of downstream taxonomic analyses, ultimately improving metaproteomics-based microbiome studies.

Acknowledgements

T.V.D.B. acknowledges funding from the Research Foundation Flanders (FWO) [1286824N]. A.D. acknowledges funding from the Research Foundation Flanders (FWO) [1SE3724N]. R.G. acknowledges funding from the Research Foundation Flanders (FWO) [G028821N] T.H. acknowledges funding from the Joachim-Herz Foundation. P.V. acknowledges funding by Ghent University [BOF/01P10623]. L.M. acknowledges funding by the Research Foundation Flanders (FWO) (G010023N, G028821N), a Ghent University Concerted Research Action (BOF21/GOA/033), and the European Union's Horizon Europe Programme (101080544, 101103253, 101195186, and 10119173).

References

1. Lehmann T, Schallert K, Vilchez-Vargas R, Benndorf D, Püttker S, Sydor S, et al. Metaproteomics of fecal samples of Crohn's disease and Ulcerative Colitis. J Proteomics. 2019;201:93–103.

2. Heyer R, Schallert K, Siewert C, Kohrs F, Greve J, Maus I, et al. Metaproteome analysis reveals that syntrophy, competition, and phage-host interaction shape microbial communities in biogas plants. Microbiome. 2019;7:69.

3. Bastida F, Torres IF, Andrés-Abellán M, Baldrian P, López-Mondéjar R, Větrovský T, et al. Differential sensitivity of total and active soil microbial communities to drought and forest management. Glob Chang Biol. 2017;23:4185–203.

4. Colaert N, Degroeve S, Helsens K, Martens L. Analysis of the resolution limitations of peptide identification algorithms. J Proteome Res. 2011;10:5555–61.

5. Muth T, Kolmeder CA, Salojärvi J, Keskitalo S, Varjosalo M, Verdam FJ, et al. Navigating through metaproteomics data: a logbook of database searching. Proteomics. 2015;15:3439–53.

6. Van Den Bossche T, Kunath BJ, Schallert K, Schäpe SS, Abraham PE, Armengaud J, et al. Critical Assessment of MetaProteome Investigation (CAMPI): a multi-laboratory comparison of established workflows. Nat Commun. 2021;12:7305.

7. Declercq A, Bouwmeester R, Hirschler A, Carapito C, Degroeve S, Martens L, et al. MSRescore: Data-Driven Rescoring Dramatically Boosts Immunopeptide Identification Rates. Mol Cell Proteomics. 2022;21:100266.

8. Gabriels R, Martens L, Degroeve S. Updated MS^2PIP web server delivers fast and accurate MS^2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. Nucleic Acids Res. 2019;47:W295–9.

9. Bouwmeester R, Gabriels R, Hulstaert N, Martens L, Degroeve S. DeepLC can predict retention times for peptides that carry as-yet unseen modifications. Bioinformatics. bioRxiv; 2020. p. 321.

10. Potgieter MG, Nakedi KC, Ambler JM, Nel AJM, Garnett S, Soares NC, et al. Proteogenomic Analysis of Mycobacterium smegmatis Using High Resolution Mass Spectrometry. Front Microbiol. 2016;7:427.

11. Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J, et al. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat Methods. 2019;16:509–18.

12. Schiebenhoefer H, Van Den Bossche T, Fuchs S, Renard BY, Muth T, Martens L. Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis. Expert Rev Proteomics. 2019;16:375–90.

13. Jagtap PD, Hoopmann MR, Neely BA, Harvey A, Käll L, Perez-Riverol Y, et al. The Association of Biomolecular Resource Facilities Proteome Informatics Research Group Study on Metaproteomics (iPRG-2020). J Biomol Tech [Internet]. 2023;34. Available from:

http://dx.doi.org/10.7171/3fc1f5fe.a058bad4

14. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 2019;47:D442–50.

15. Hulstaert N, Shofstahl J, Sachsenberg T, Walzer M, Barsnes H, Martens L, et al. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion [Internet]. 2020. Available from: https://pubs.acs.org/doi/10.1021/acs.jproteome.9b00328

16. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004;20:1466–7.

17. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics. 2011;11:996–9.

18. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015;31:926–32.

19. Levitsky LI, Klein JA, Ivanov MV, Gorshkov MV. Pyteomics 4.0: Five Years of Development of a Python Proteomics Framework. J Proteome Res. 2019;18:709–14.

20. C. Silva AS, Bouwmeester R, Martens L, Degroeve S. Accurate peptide fragmentation predictions allow data driven approaches to replace and improve upon proteomics search engine scoring functions. Wren J, editor. Bioinformatics. 2019;35:5243–8.

21. Mesuere B, Devreese B, Debyser G, Aerts M, Vandamme P, Dawyndt P. Unipept: Tryptic Peptide-Based Biodiversity Analysis of Metaproteome Samples [Internet]. Journal of Proteome Research. 2012. p. 5773–80. Available from: http://dx.doi.org/10.1021/pr300576s

22. Singh RG, Tanca A, Palomba A, Van der Jeugt F, Verschaffelt P, Uzzau S, et al. Unipept 4.0: Functional Analysis of Metaproteome Data [Internet]. Journal of Proteome Research. 2019. p. 606–15. Available from: http://dx.doi.org/10.1021/acs.jproteome.8b00716

23. Verschaffelt P, Van Thienen P, Van Den Bossche T, Van der Jeugt F, De Tender C, Martens L, et al. Unipept CLI 2.0: adding support for visualisations and functional annotations. Martelli PL, editor. Bioinformatics [Internet]. 2020; Available from: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btaa553/5850861

24. Declercq A, Devreese R, Scheid J, Jachmann C, Van Den Bossche T, Preikschat A, et al. TIMSRescore: A Data Dependent Acquisition-Parallel Accumulation and Serial Fragmentation-Optimized Data-Driven Rescoring Pipeline Based on MSRescore. J Proteome Res [Internet]. 2025; Available from: http://dx.doi.org/10.1021/acs.jproteome.4c00609