ENHANCING PEPTIDE IDENTIFICATION IN METAPROTEOMICS THROUGH CURRICULUM LEARNING IN DEEP LEARNING SUPPLEMENTARY DOCUMENT

Shichao Feng
Department of Computer Science and Engineering
University of North Texas, Denton, TX, USA

Bailu Zhang
Department of Computer Science and Engineering
University of North Texas, Denton, TX, USA

Huan Wang
College of Informatics
Huazhong Agricultural University, Wuhan, Hubei, China

Yi Xiong
School of Biological Sciences
University of Oklahoma, Norman, OK, USA

Athena Tian
Department of Mathematics
Emory University, Atlanta, GA, 30322, USA

Xiaohui Yuan
Department of Computer Science and Engineering
University of North Texas, Denton, TX, USA

Chongle Pan
School of Biological Sciences and School of Computer Science
University of Oklahoma, Norman, OK, USA
cpan@ou.edu *

Xuan Guo
Department of Computer Science and Engineering
University of North Texas, Denton, TX, USA
xuan.guo@unt.edu †

*Corresponding author: cpan@ou.edu
†Corresponding author: xuan.guo@unt.edu

Contents

1 Supplementary Figures **6**

2 Supplementary Tables **24**

3 Supplementary Note **43**

3.1 Information of benchmarking datasets 43

4 Supplementary Methods **45**

4.1 Input representation construction 45

4.2 WinnowNet* architecture 45

4.2.1 Spectrum encoder with order-invariant property 45

4.2.2 PSM feature encoder 46

4.3 Training 46

4.4 Analysis of the significance for the features learned by CNN-based WinnowNet 46

4.5 Impact of precursor mass tolerance 46

4.6 Analysis of the significance for the features learned by CNN-based WinnowNet 46

5 Supplementary Discussion **47**

5.1 Assessment of training strategies for curriculum learning 47

5.1.1 Impact of curriculum learning for CNN-based WinnowNet 47

5.1.2 Impact of curriculum learning strategy for self-attention-based WinnowNet 47

5.2 Comparison of WinnowNet and MS^{2}Rescore Using the Entrapment Method on MS-GF+ Outputs . . 48

5.3 Analysis of the probability changes for peptides with adjacent amino acids swapped 48

5.4 Validation of FDR control via synthetic dataset entrapment analysis 49

5.5 Performance evaluation of WinnowNet-integrated protein identification pipelines on timsTOF data . . 49

List of Tables

1 The number of MS/MS spectra and proteins of thirteen proteome datasets 24

2 The list of 27 foreign species used as entrapment 25

3 Filenames of MS datasets and protein databases in this study 26

4 The training and validating accuracies when WinnowNet models converged 27

5 Benchmarking tools used in this study 28

6 11 PSM features used in WinnowNet* (CNN-based WinnowNet) 29

7 Performance comparison for three training strategies for WinnowNet* (CNN-based WinnowNet) . . . 30

8 Performance comparison for four training strategies for WinnowNet (self-attention-based WinnowNet) 31

9 Identification performance of WinnowNet* (CNN-based) integrated with Sipros-Ensemble under different noisy peaks tolerances. 32

10 Number of homologous PSM candidates for different test datasets compared with training datasets . . 33

11 Decrease in identifications after the removal of homologous peptides for various metaproteomes at 1% FDR. 34

12 Decrease in identifications after the removal of homologous peptides for metaproteomes for the mock community at FDR 1% 35

13 PSM/Peptide/Protein identification results on the Human Gut timsTOF dataset using three metaproteomics pipelines at 1% FDR. 36

14 PSM/Peptide/Protein identification results on three datasets using four metaproteomics pipelines at 1% FDR. 37

15 PSM/Peptide/Protein identification results on P1 dataset with 27 foreign species using four metaproteomics pipelines at 1% FDR. 38

16 PSM/Peptide/Protein identification results on marine and soil datasets using three search engineers and eight filters at 1% FDR. 39

17 PSM/Peptide/Protein identification results on mock and human gut datasets using three search engineers and eight filters at 1% FDR. 40

18 Number of identifications reported by the original publications of the datasets used for comparison. . 41

19 Comparison of PSM identification results between WinnowNet and MS^{2}Rescore using the output by MS-GF+. 42

List of Figures

1 The WinnowNet workflow and architecture including both the self-attention-based model and the CNN-based model. (a) Overview of peptide and protein identification using database search engines. (b) Construction of training datasets and the training process. (c) Architecture of WinnowNet and WinnowNet*. 6

2 The architecture of WinnowNet*. It employs a CNN-based backbone to process each PSM candidate. The input includes normalized matrices representing the MS/MS spectrum, the theoretical spectrum, and 11 additional handcrafted features. The spectrum encoder comprises CNN-based transformation and feature extraction networks, while the PSM feature encoder is implemented using a fully connected layer. The classifier, another fully connected layer, receives concatenated outputs from the spectrum and PSM feature encoders and predicts the probability of a PSM being correct. 7

3 Performance comparison of WinnowNet* (CNN-based WinnowNet) with and without the spectrum encoder. The Y-axis represents the gain in identifications achieved by the model compared to Percolator at FDR 1%. 8

4 Visualization of Class Activation Maps (CAMs) of WinnowNet* (CNN-based WinnowNet) for PSMs with the same experimental spectrum. The spectrum above the line $y = 0$ represents the experimental spectrum, while the spectrum below $y = 0$ represents the theoretical spectrum. Fragment ions are color-coded by ion type and charge: b-ions in yellow and y-ions in blue. (a) CAM for a target PSM from the Marine model with peptide sequence SM∼YITYSTK. (b) CAM for a decoy PSM from the Marine model with peptide sequence SEFAEVQSGR. (c) CAM for a target PSM from the ProteomeTools model with the same peptide sequence as in (a), SM∼YITYSTK. Note that the symbol '∼' indicates a variable modification (methionine oxidation). 9

5 Visualization of experimental spectrum, theoretical isotopic distribution, and matching fragment ions. In (c), the spectrum above the line $y = 0$ indicates the theoretical isotope distribution for fragment ions of (blue), and the spectrum below the line $y = 0$ indicates the experimental spectrum. Fragment ions that find matching peaks between the theoretical spectrum and experimental spectrum with different charge states are marked with yellow (b-ion) and blue (y-ion). 10

6 Visualization of aggregation weights of attention maps for WinnowNet (self-attention-based WinnowNet) applied to PSMs with the same experimental spectrum. The spectrum below the line $y = 0$ represents the experimental spectrum, while the spectrum above the line $y = 0$ represents the theoretical spectrum. Fragment ions are color-coded by ion type: blue for b-ions and red for y-ions. (a) Aggregation weights for a top-ranked true positive sample. (b) Aggregation weights for a second-ranked false negative sample. (c) Aggregation weights for a low-ranked false negative sample. 11

7 The ROC curves on the sample test dataset of ProteomeTools dataset for WinnowNet* (CNN-based WinnowNet), Comet, MS-GF+, and MyriMatch. (n= 6594 samples) 12

8 Training and validation loss curves for both CNN-based and self-attention-based WinnowNet. The gaps in the loss curves represent the transition in the curriculum learning process, where training progresses from the easy dataset to the more difficult dataset: (a) The second phase of training began from the checkpoint at epoch 15 of the first phase; (b) The second phase of training began from the checkpoint at epoch 47 of the first phase. 13

9 Training accuracy of WinnowNet* (CNN-based WinnowNet) after each epoch for the best model out of five training rounds using three training strategies. The titles of subfigures present the dataset composition and the best test dataset accuracy. "P1" indicates training models in Phase 1; "P2" indicates training models in Phase 2. 14

10 Training accuracy of WinnowNet* (CNN-based WinnowNet) after each epoch with five different compositions of training samples. The titles in subfigures present the dataset composition and the best test dataset accuracy. "P1" indicates training models in Phase 1; "P2" indicates training models in Phase 2. 15

11 Histograms of score distributions for top-ranked target PSMs in the Soil 2 dataset. The x-axis represents the score values, and the y-axis shows the logarithm of the number of PSMs at each score. (a) PSMs were ranked by WinnowNet, except for the "MS^{2}Rescore-Only" subset, which were initially ranked by MS^{2}Rescore and then evaluated using WinnowNet. The WinnowNet (self-attention-based) score cutoff corresponding to a 1% FDR is 0.85. (b) PSMs were ranked by MS^{2}Rescore, except for the "WinnowNet-Only" subset, which were initially ranked by WinnowNet and then evaluated using MS^{2}Rescore. The MS^{2}Rescore score cutoff corresponding to a 1% FDR is 0.3. 16

12 Histogram of the score decreases for the peptides in which two amino acids were randomly switched from identified peptides at the peptide-level FDR of 1%. Note that self-attention-based WinnowNet was applied here. 17

13 Entrapment analysis using synthetic data from four microorganisms. This figure compares two protein identification pipelines—Sipros Ensemble (SE) with WinnowNet (self-attention-based) and FragPipe—using entrapment experiments with shuffled sequences at a ratio of 1 : 1. All results shown were filtered at 1% FDR. The combined or paired estimated FDR values are plotted against foreign-species-based FDR estimates. The original target database includes proteins from the four microorganisms and foreign species at a 1:212 ratio (1:298 at the peptide level). Due to the small proportion of native peptides in the extended database, the combined and paired curves appear visually overlapping. 18

14 Score distributions for top-ranked PSMs in the mock P2 dataset, derived from WinnowNet (self-attention-based) and MS^{2}Rescore. Original target PSMs are shown in blue, and entrapment and decoy PSMs in yellow. The top panel displays PSM scores from WinnowNet, the left panel shows scores from MS^{2}Rescore, and the central panel illustrates the correlation between WinnowNet scores (x-axis) and MS^{2}Rescore scores (y-axis). Solid lines indicate the 1% PSM-level FDR cutoffs for WinnowNet (0.93) and MS^{2}Rescore (-0.07). 19

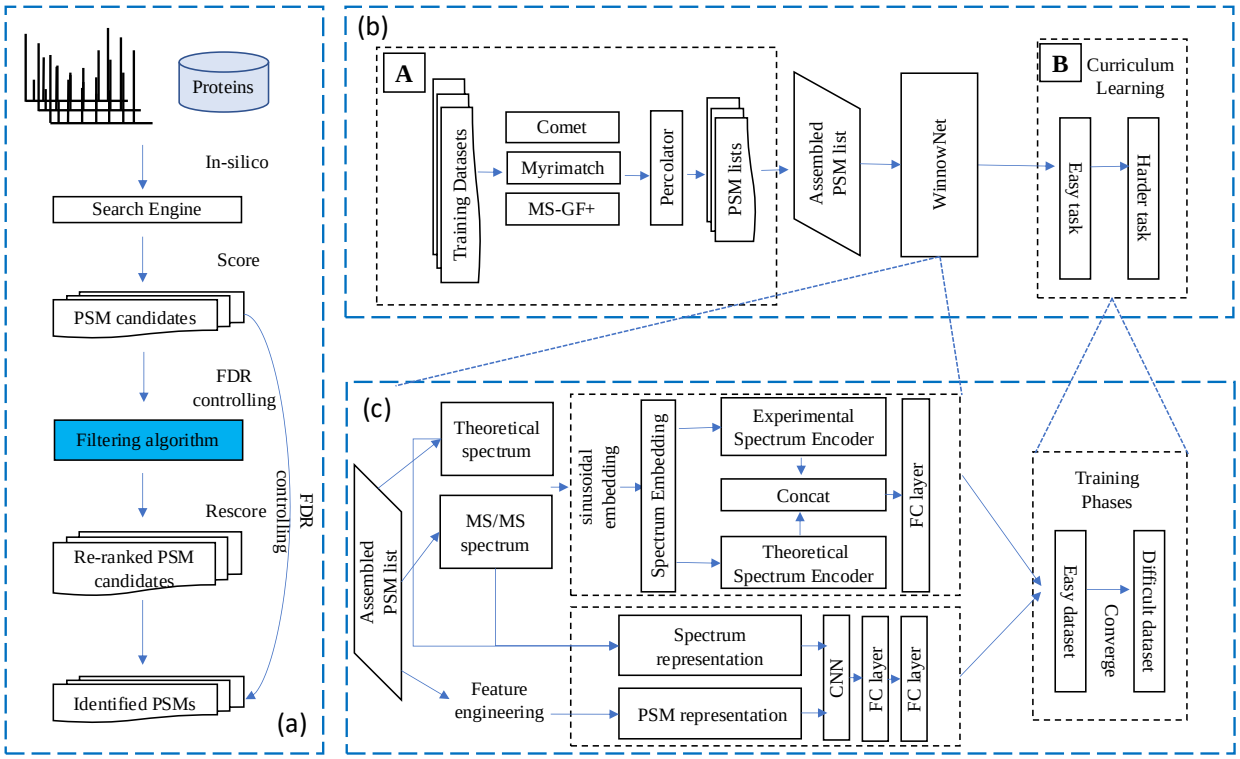
15 Distributions of spectral angle similarity for top-ranked PSMs classified as MS^{2}Rescore-only, WinnowNet-only, or shared by both filtering algorithms. The x-axis represents spectral angle similarity, and the y-axis shows the count of PSMs at each similarity value. 20

16 Distributions of Pearson correlation similarity for top-ranked PSMs classified as MS^{2}Rescore-only, WinnowNet-only, or shared by both filtering algorithms. The x-axis represents the Pearson correlation similarity, and the y-axis shows the count of PSMs at each similarity value. 21

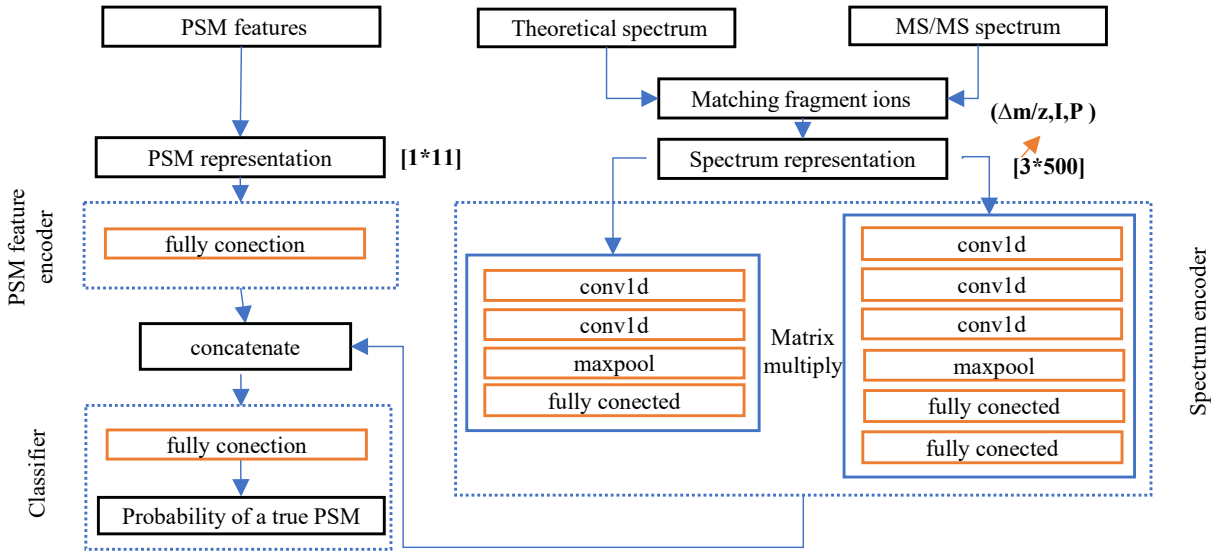
17 Quantile–quantile (QQ) plots comparing spectral angle similarity distributions. The plot is based on top-ranked PSMs across different groups: MS^{2}Rescore-only, WinnowNet-only, and shared PSMs, for the Marine 2 and Soil 2 datasets. Each subplot compares two of the three groups to assess distributional agreement. The x- and y-axes represent spectral angle similarity quantiles for each group. 22

18 Visualizations of four representative PSMs from the Marine 2 and Soil 2 datasets, each annotated with its WinnowNet matching score. These examples illustrate varying levels of spectral match quality, as reflected by the corresponding scores. 23

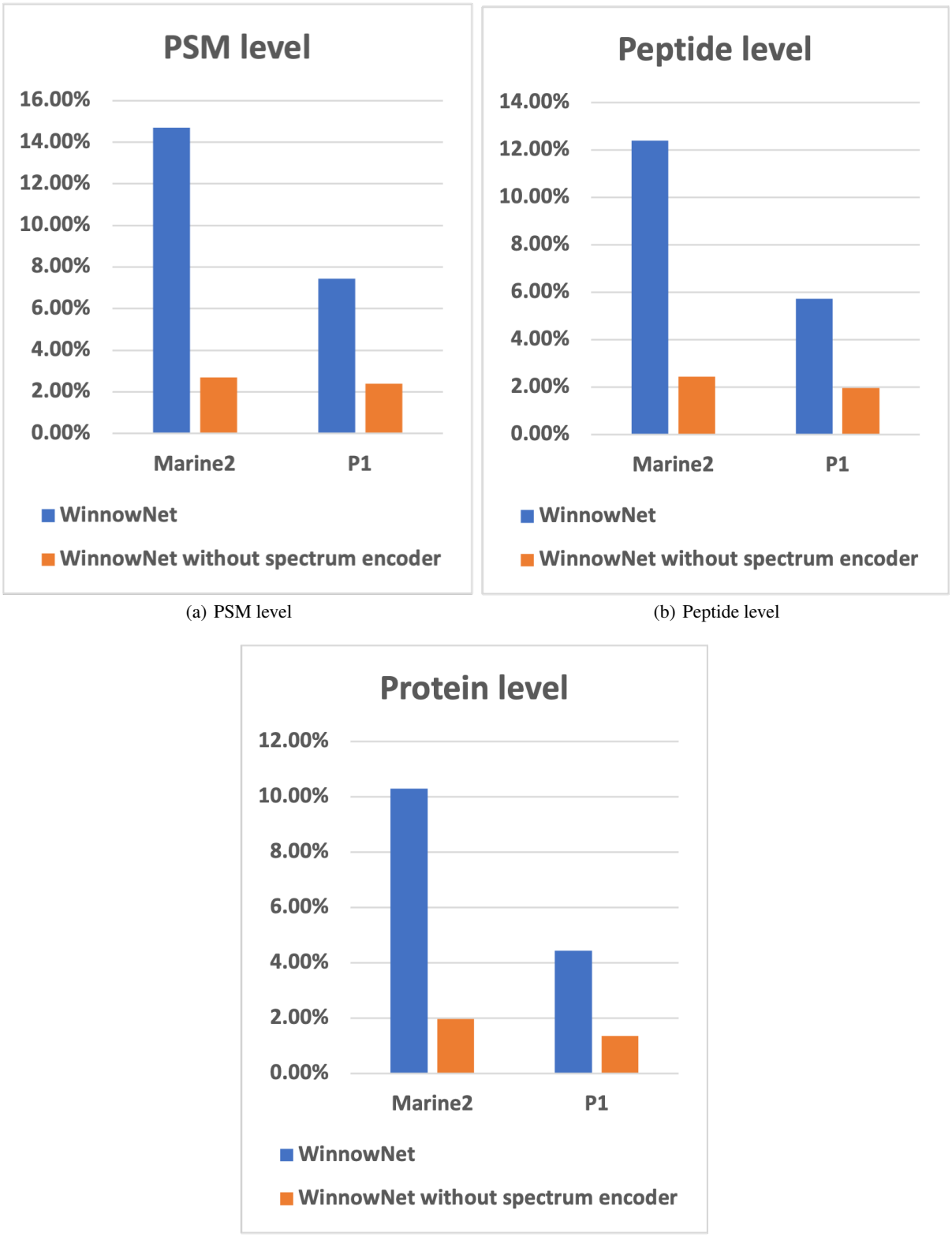
1 Supplementary Figures



Supplementary Figure 1: The WinnowNet workflow and architecture including both the self-attention-based model and the CNN-based model. (a) Overview of peptide and protein identification using database search engines. (b) Construction of training datasets and the training process. (c) Architecture of WinnowNet and WinnowNet*.



Supplementary Figure 2: The architecture of WinnowNet*. It employs a CNN-based backbone to process each PSM candidate. The input includes normalized matrices representing the MS/MS spectrum, the theoretical spectrum, and 11 additional handcrafted features. The spectrum encoder comprises CNN-based transformation and feature extraction networks, while the PSM feature encoder is implemented using a fully connected layer. The classifier, another fully connected layer, receives concatenated outputs from the spectrum and PSM feature encoders and predicts the probability of a PSM being correct.

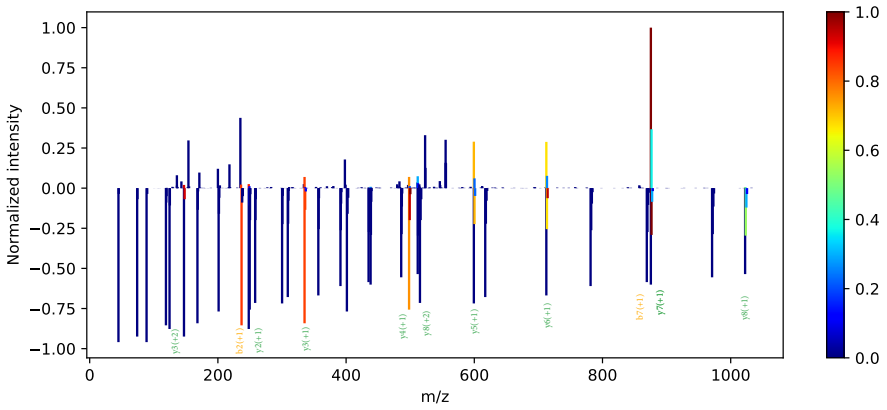


(a) PSM level

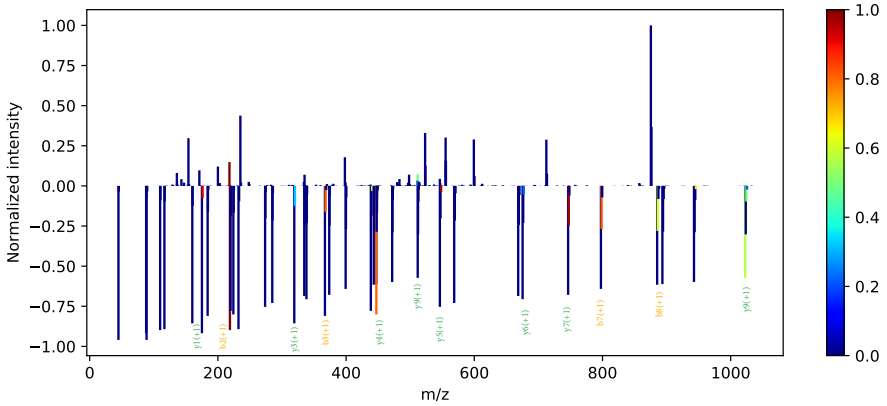
(b) Peptide level

(c) Protein level

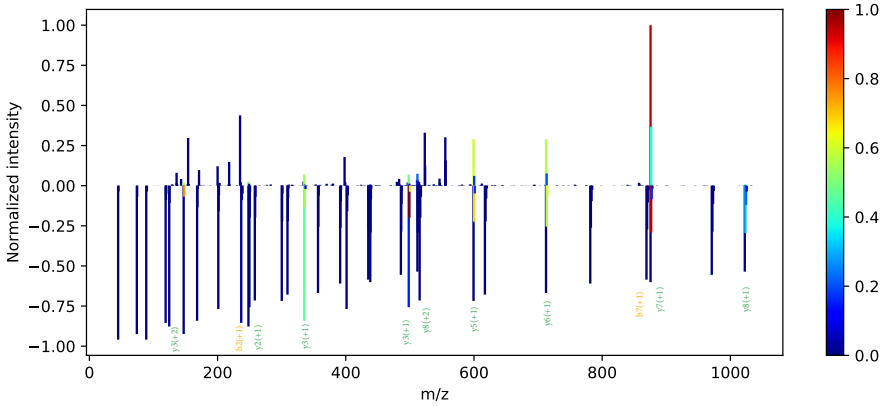
Supplementary Figure 3: Performance comparison of WinnowNet* (CNN-based WinnowNet) with and without the spectrum encoder. The Y-axis represents the gain in identifications achieved by the model compared to Percolator at FDR 1%.



(a) CAM for a target PSM from the Marine model with peptide sequence SM∼YITYSTK.

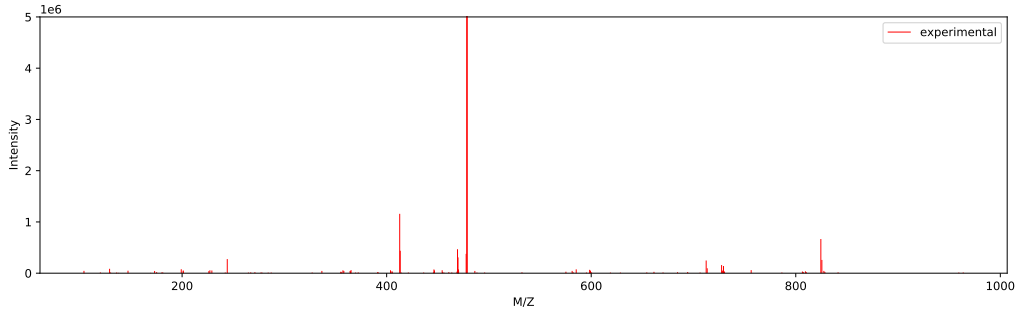


(b) CAM for a decoy PSM from the Marine model with peptide sequence SEFAEVQSGR.

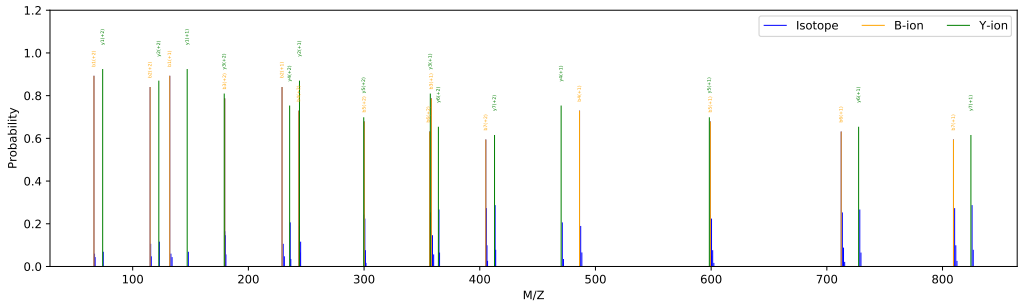


(c) CAM for a target PSM from the ProteomeTools model with peptide sequence SM∼YITYSTK.

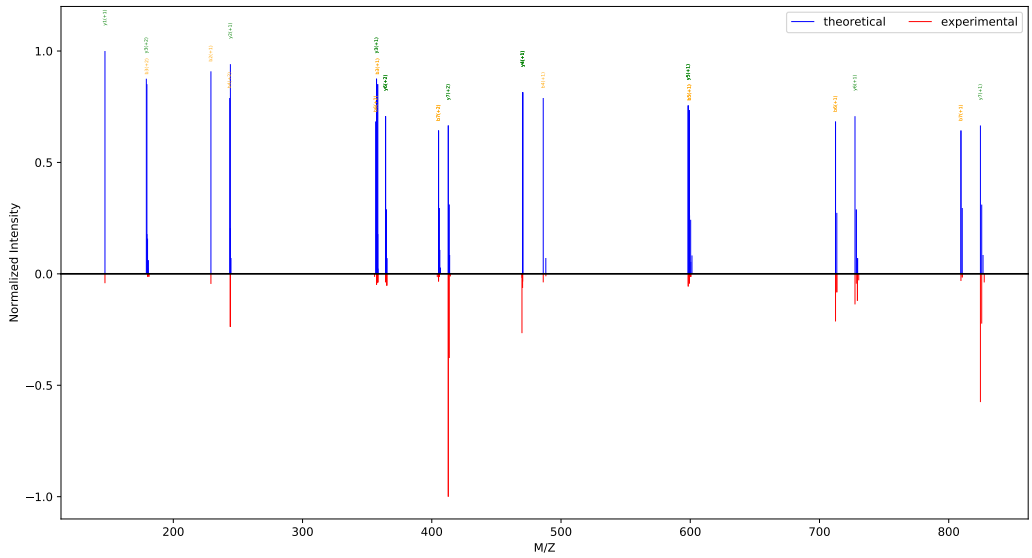
Supplementary Figure 4: Visualization of Class Activation Maps (CAMs) of WinnowNet* (CNN-based WinnowNet) for PSMs with the same experimental spectrum. The spectrum above the line $y = 0$ represents the experimental spectrum, while the spectrum below $y = 0$ represents the theoretical spectrum. Fragment ions are color-coded by ion type and charge: b-ions in yellow and y-ions in blue. (a) CAM for a target PSM from the Marine model with peptide sequence SM∼YITYSTK. (b) CAM for a decoy PSM from the Marine model with peptide sequence SEFAEVQSGR. (c) CAM for a target PSM from the ProteomeTools model with the same peptide sequence as in (a), SM∼YITYSTK. Note that the symbol '∼' indicates a variable modification (methionine oxidation).



(a) Visualization of an experimental spectrum

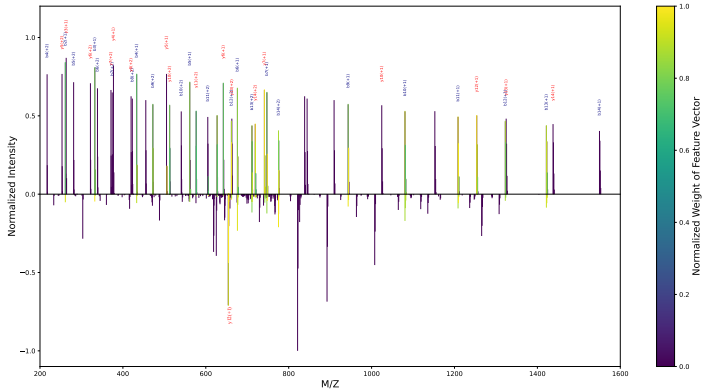


(b) Visualization of the theoretical isotope distribution for the fragment ions of a peptide: MPEKILPK.

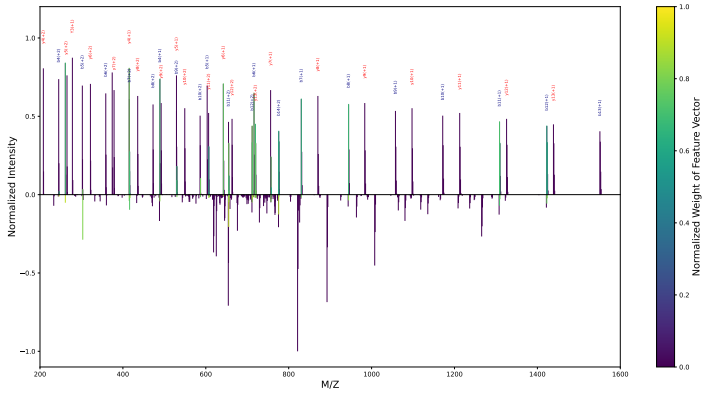


(c) Visualization of the matching ions within 0.2 Da mass tolerance between the experimental spectrum in (a) and the theoretical isotope distribution in (b).

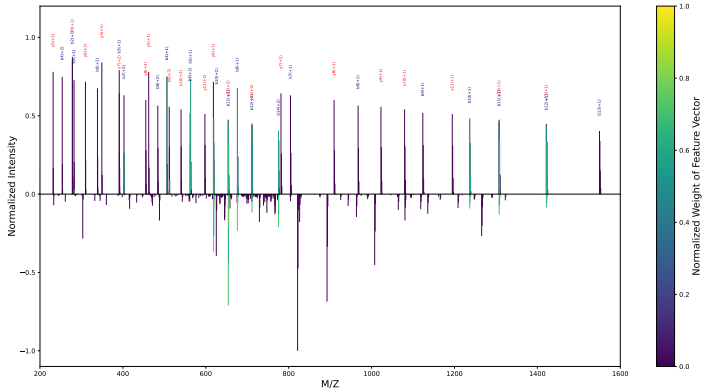
Supplementary Figure 5: Visualization of experimental spectrum, theoretical isotopic distribution, and matching fragment ions. In (c), the spectrum above the line $y = 0$ indicates the theoretical isotope distribution for fragment ions of (blue), and the spectrum below the line $y = 0$ indicates the experimental spectrum. Fragment ions that find matching peaks between the theoretical spectrum and experimental spectrum with different charge states are marked with yellow (b-ion) and blue (y-ion).



(a) Peptide Sequence: FIATQDAPVHQNV

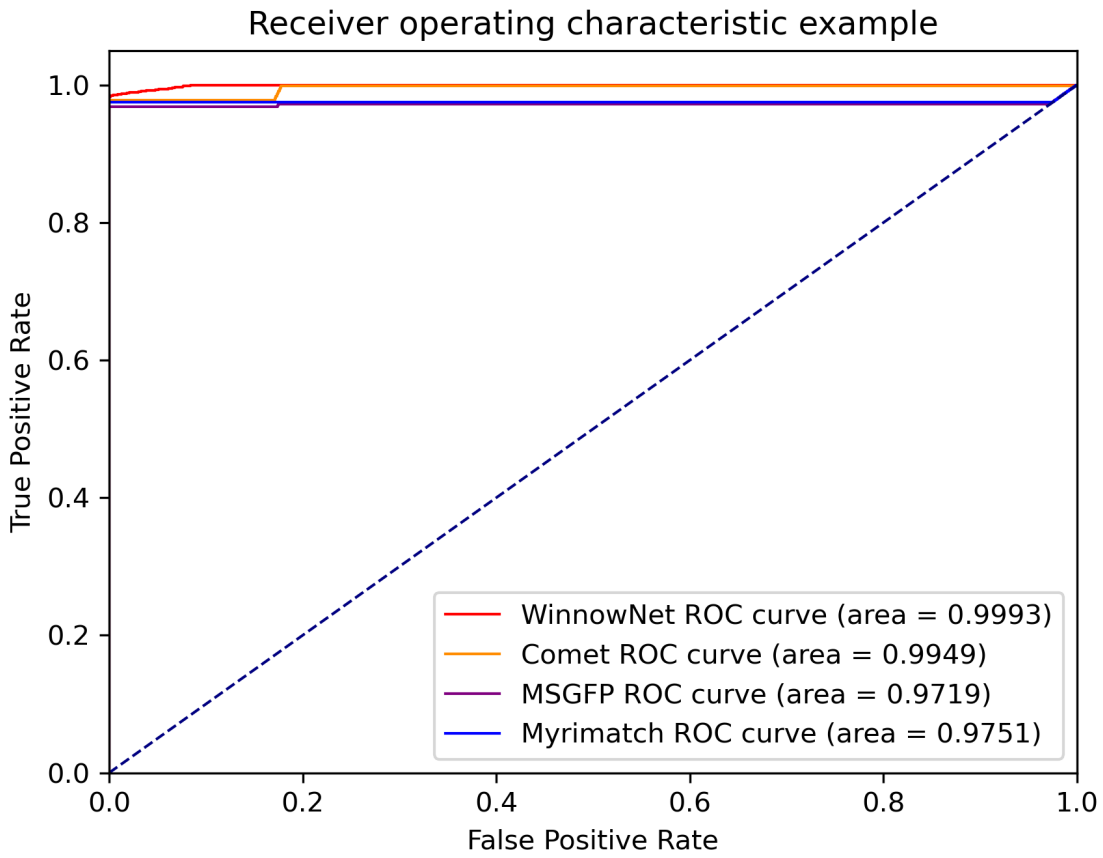


(b) Peptide Sequence: FILDNINNINHNK

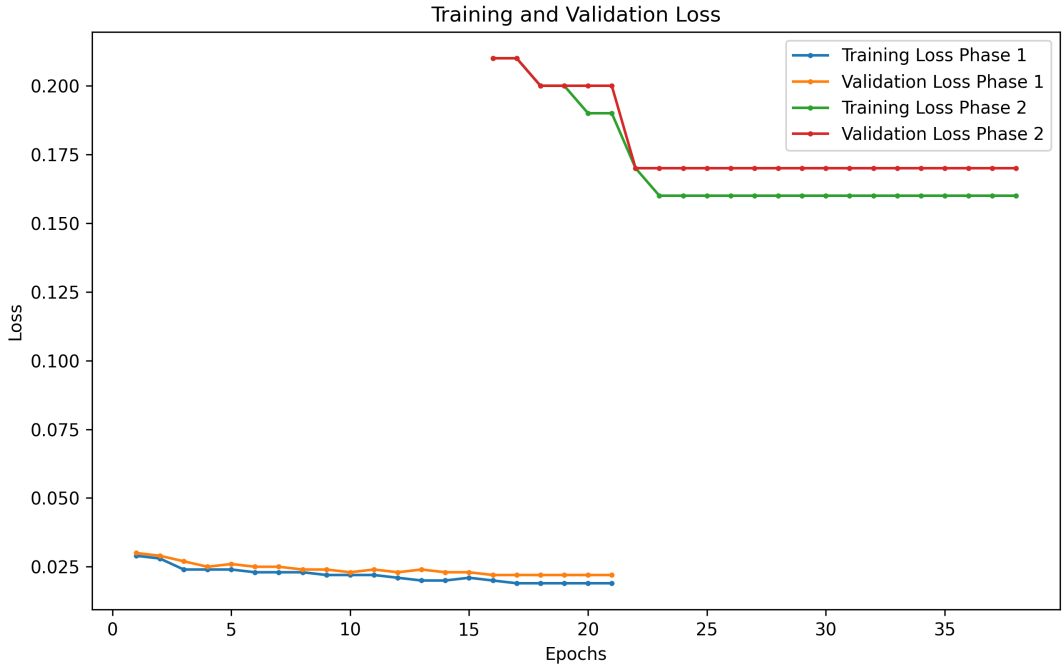


(c) Peptide Sequence: YNIDGLQYRIANK

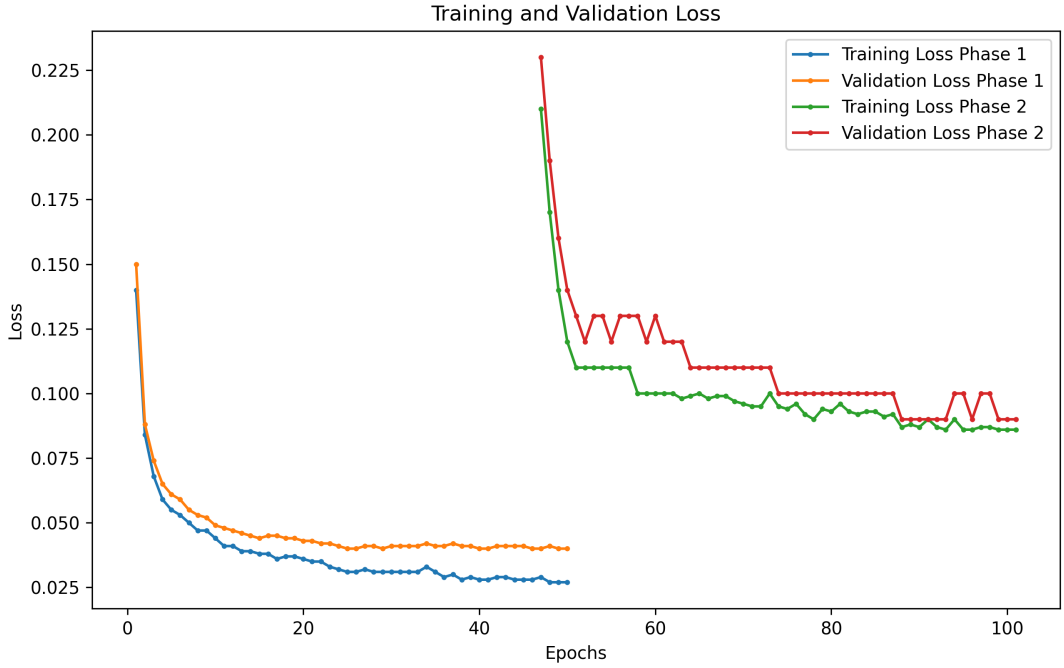
Supplementary Figure 6: Visualization of aggregation weights of attention maps for WinnowNet (self-attention-based WinnowNet) applied to PSMs with the same experimental spectrum. The spectrum below the line $y = 0$ represents the experimental spectrum, while the spectrum above the line $y = 0$ represents the theoretical spectrum. Fragment ions are color-coded by ion type: blue for b-ions and red for y-ions. (a) Aggregation weights for a top-ranked true positive sample. (b) Aggregation weights for a second-ranked false negative sample. (c) Aggregation weights for a low-ranked false negative sample.



Supplementary Figure 7: The ROC curves on the sample test dataset of ProteomeTools dataset for WinnowNet* (CNN-based WinnowNet), Comet, MS-GF+, and MyriMatch. (n= 6594 samples)

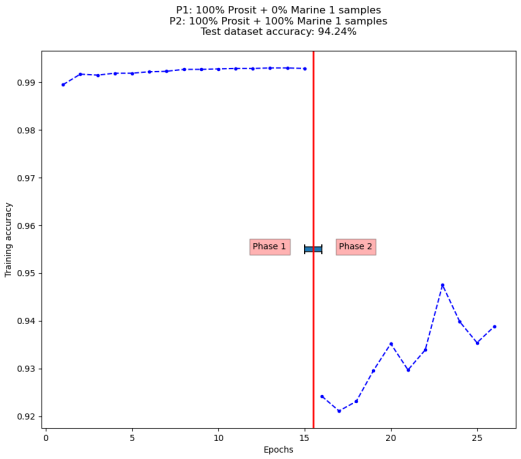


(a) Loss curves for WinnowNet* (CNN-based WinnowNet).

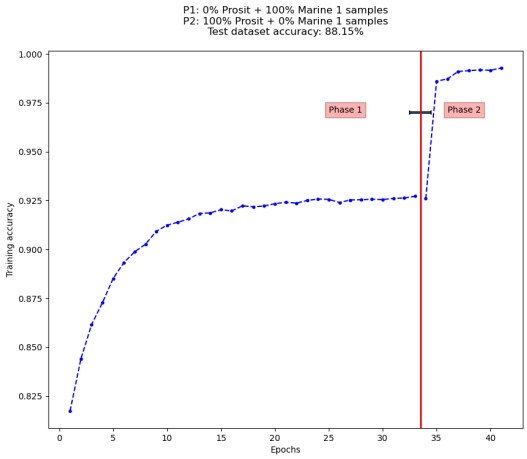


(b) Loss curves for WinnowNet (self-attention-based WinnowNet).

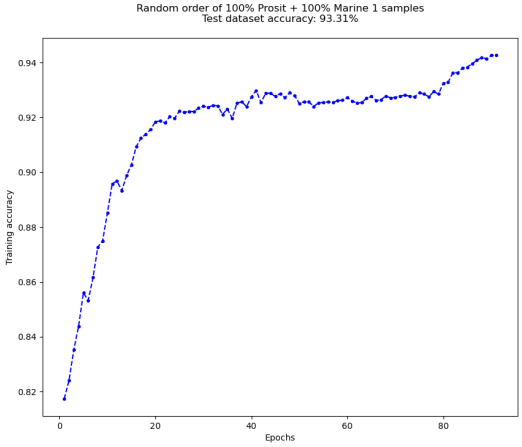
Supplementary Figure 8: Training and validation loss curves for both CNN-based and self-attention-based WinnowNet. The gaps in the loss curves represent the transition in the curriculum learning process, where training progresses from the easy dataset to the more difficult dataset: (a) The second phase of training began from the checkpoint at epoch 15 of the first phase; (b) The second phase of training began from the checkpoint at epoch 47 of the first phase.



(a) Easy-difficult Strategy

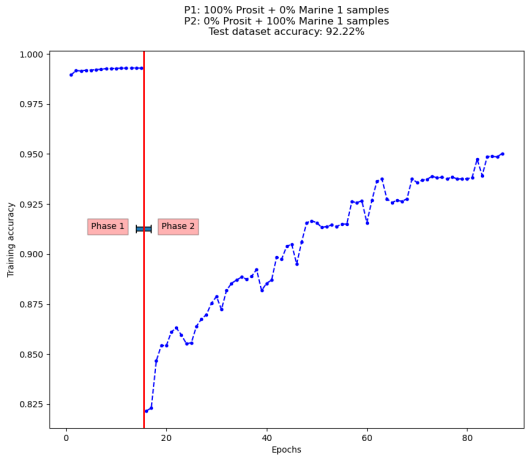


(b) Reverse Strategy

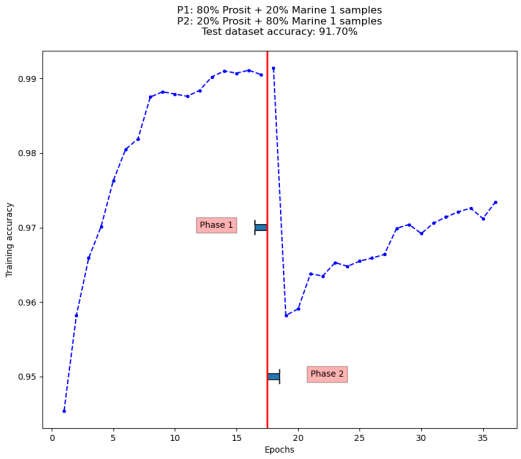


(c) Random Strategy

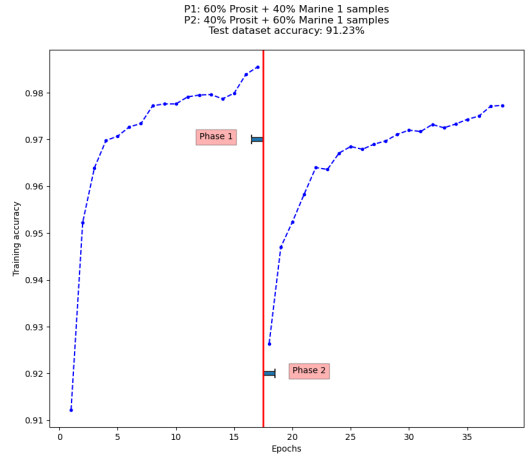
Supplementary Figure 9: Training accuracy of WinnowNet* (CNN-based WinnowNet) after each epoch for the best model out of five training rounds using three training strategies. The titles of subfigures present the dataset composition and the best test dataset accuracy. “P1” indicates training models in Phase 1; “P2” indicates training models in Phase 2.



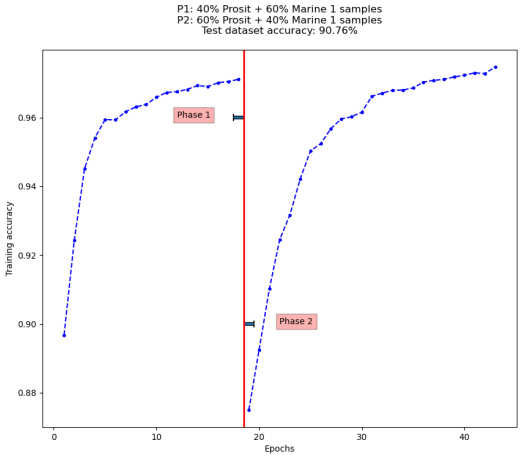
(a)



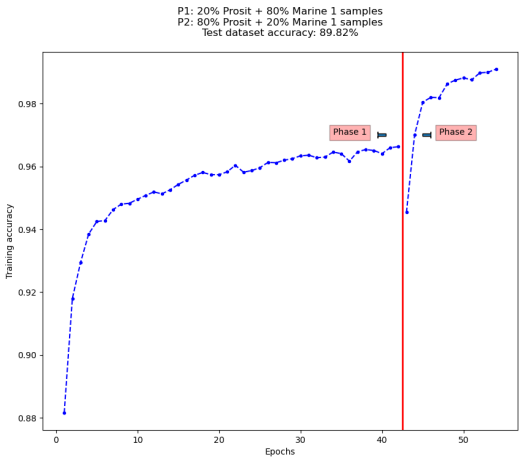
(b)



(c)

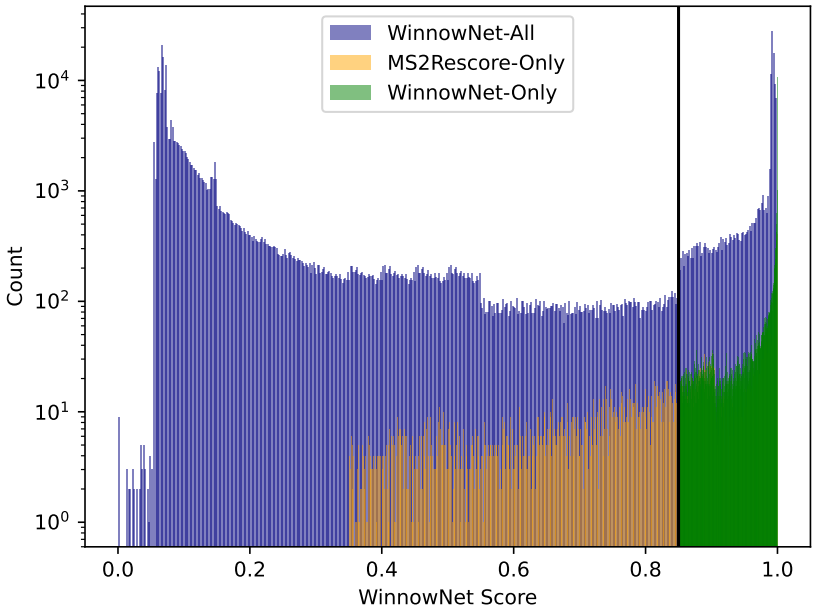


(d)

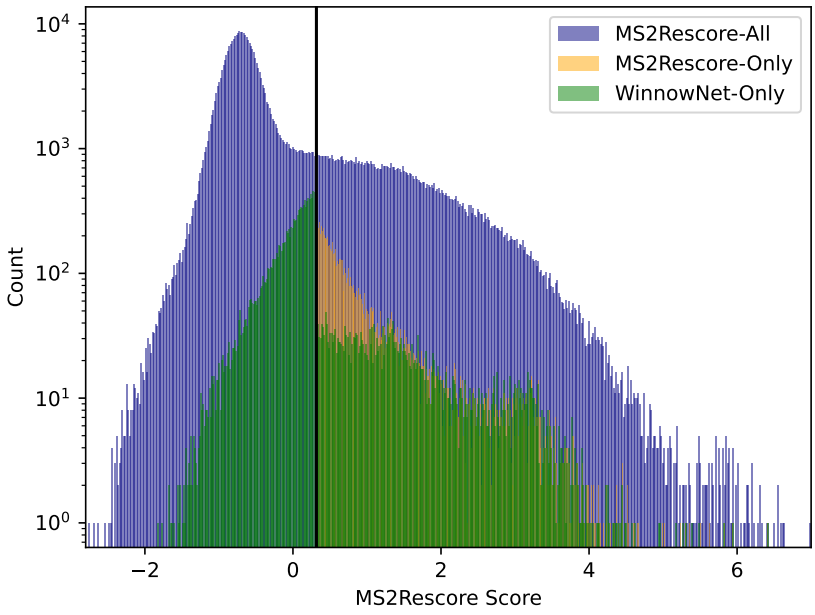


(e)

Supplementary Figure 10: Training accuracy of WinnowNet* (CNN-based WinnowNet) after each epoch with five different compositions of training samples. The titles in subfigures present the dataset composition and the best test dataset accuracy. “P1” indicates training models in Phase 1; “P2” indicates training models in Phase 2.

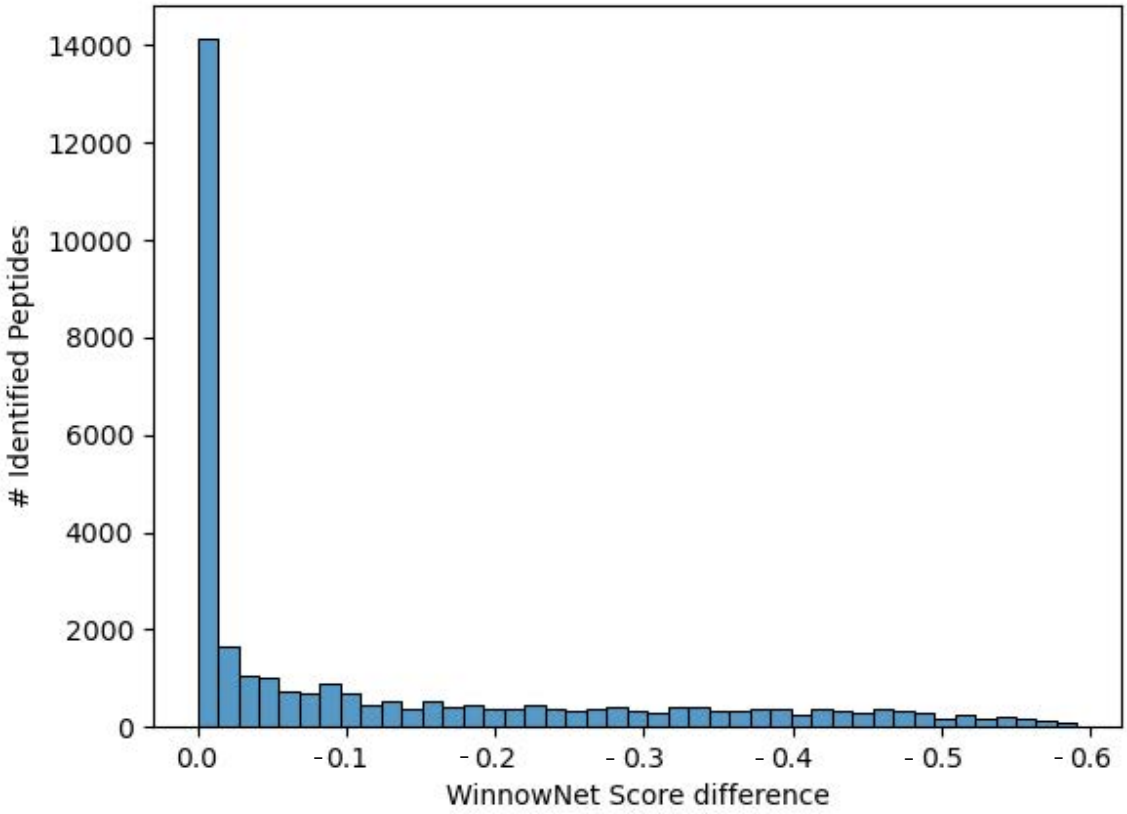


(a) Score distributions for WinnowNet

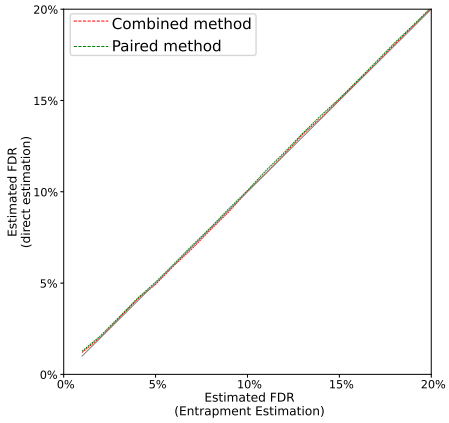


(b) Score distributions for MS^{2}Rescore

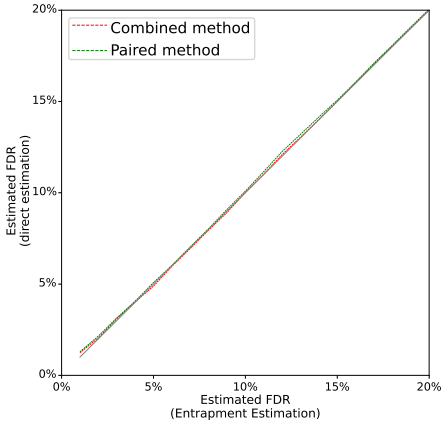
Supplementary Figure 11: Histograms of score distributions for top-ranked target PSMs in the Soil 2 dataset. The x-axis represents the score values, and the y-axis shows the logarithm of the number of PSMs at each score. (a) PSMs were ranked by WinnowNet, except for the "MS^{2}Rescore-Only" subset, which were initially ranked by MS^{2}Rescore and then evaluated using WinnowNet. The WinnowNet (self-attention-based) score cutoff corresponding to a 1% FDR is 0.85. (b) PSMs were ranked by MS^{2}Rescore, except for the "WinnowNet-Only" subset, which were initially ranked by WinnowNet and then evaluated using MS^{2}Rescore. The MS^{2}Rescore score cutoff corresponding to a 1% FDR is 0.3.



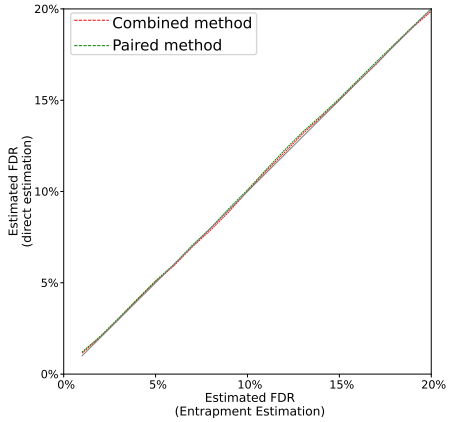
Supplementary Figure 12: Histogram of the score decreases for the peptides in which two amino acids were randomly switched from identified peptides at the peptide-level FDR of 1%. Note that self-attention-based WinnowNet was applied here.



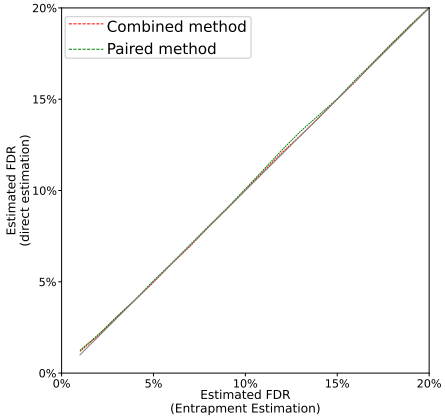
(a) PSM level using SE with WinnowNet



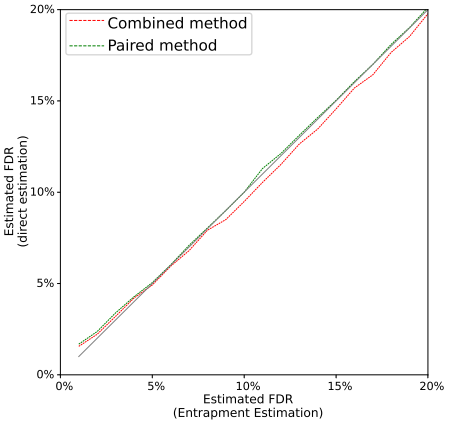
(b) PSM level using FragPipe



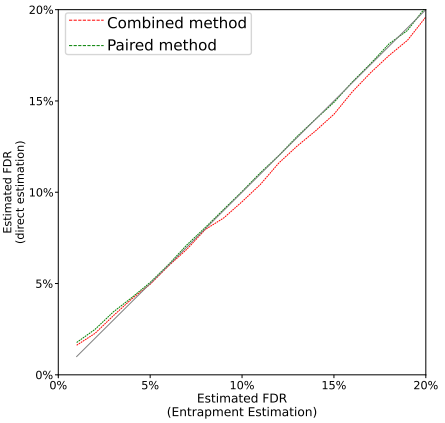
(c) Peptide level using SE with WinnowNet



(d) Peptide level using FragPipe

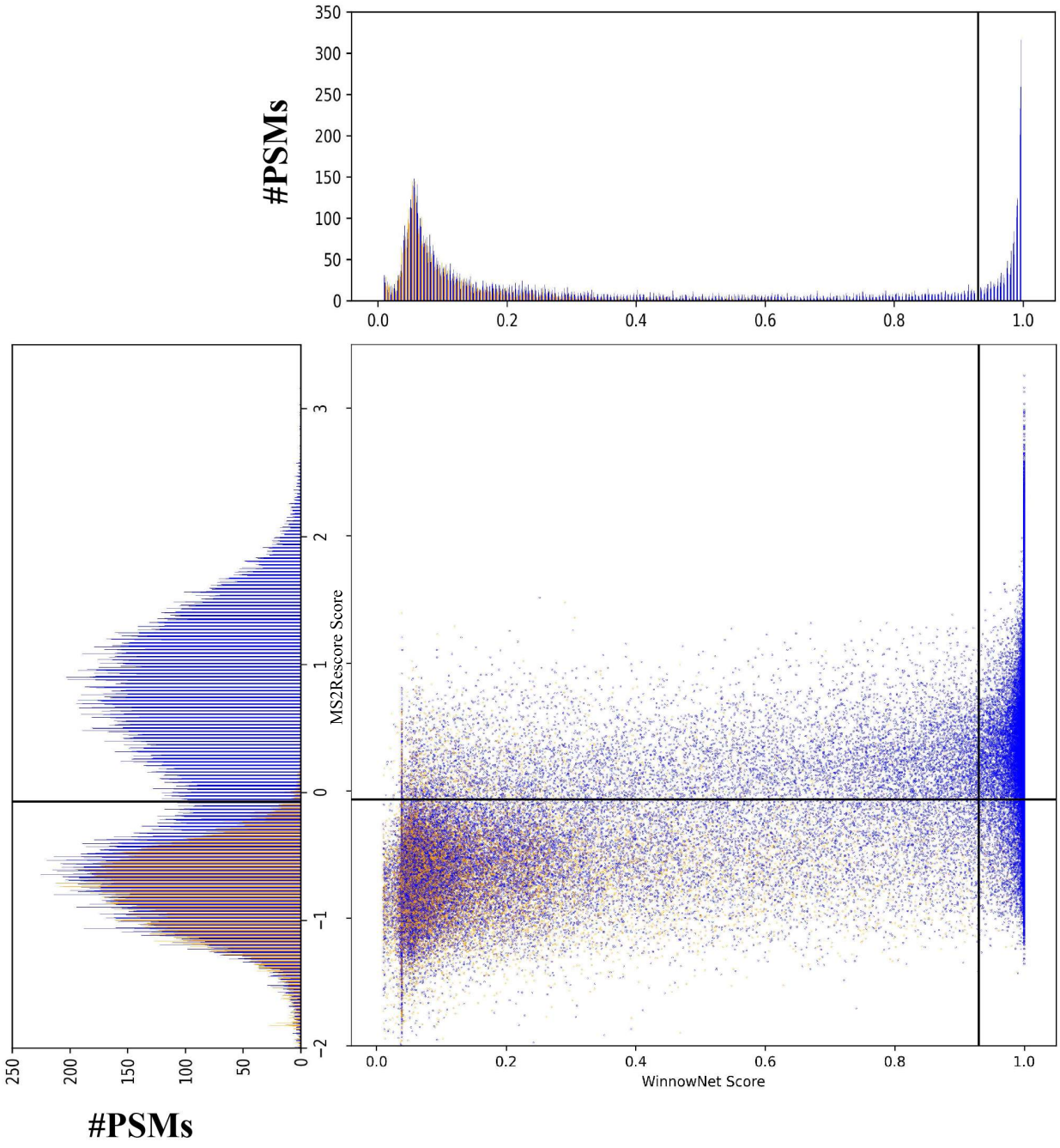


(e) Protein level using SE with WinnowNet

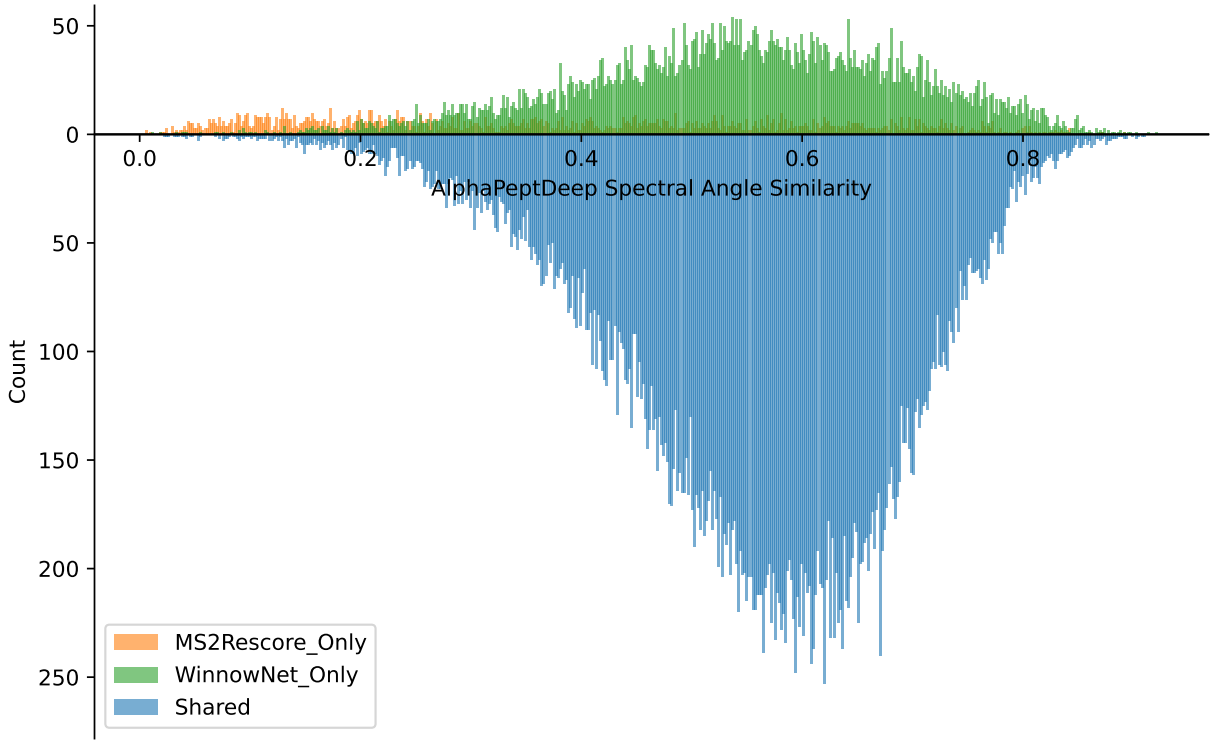


(f) Protein level using FragPipe

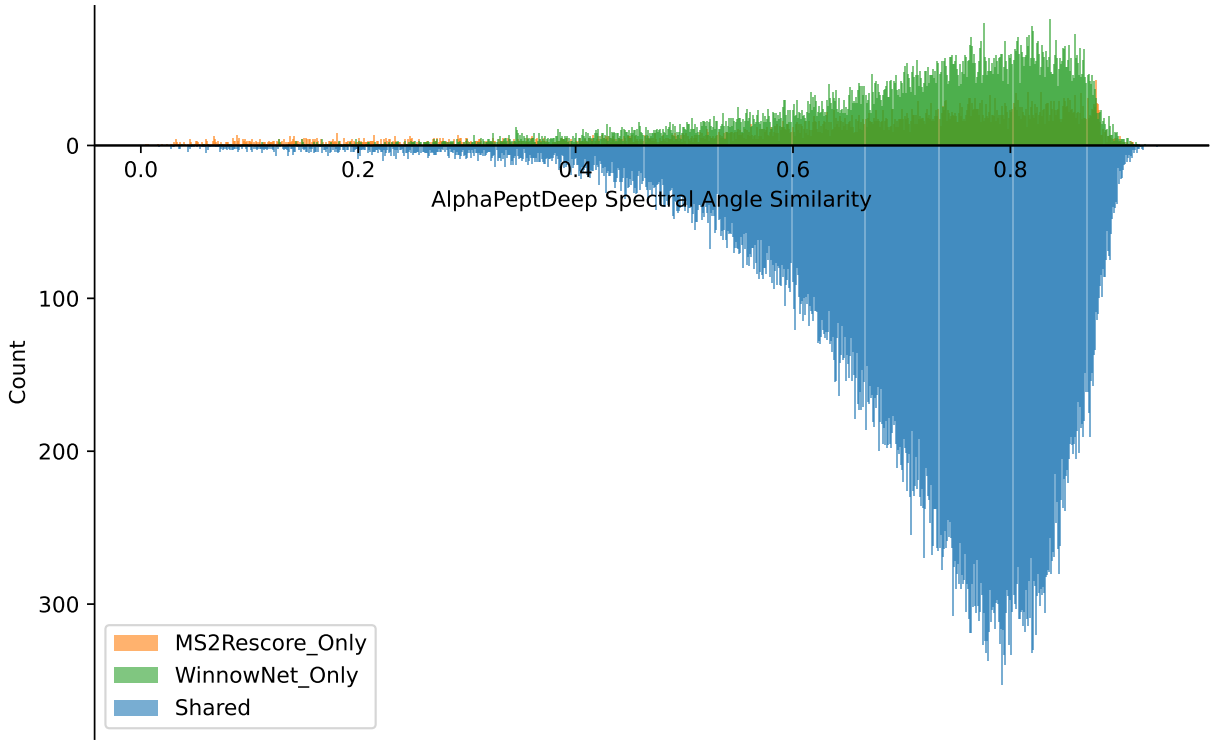
Supplementary Figure 13: Entrapment analysis using synthetic data from four microorganisms. This figure compares two protein identification pipelines—Sipros Ensemble (SE) with WinnowNet (self-attention-based) and FragPipe—using entrapment experiments with shuffled sequences at a ratio of 1 : 1. All results shown were filtered at 1% FDR. The combined or paired estimated FDR values are plotted against foreign-species-based FDR estimates. The original target database includes proteins from the four microorganisms and foreign species at a 1:212 ratio (1:298 at the peptide level). Due to the small proportion of native peptides in the extended database, the combined and paired curves appear visually overlapping.



Supplementary Figure 14: Score distributions for top-ranked PSMs in the mock P2 dataset, derived from WinnowNet (self-attention-based) and MS^{2}Rescore. Original target PSMs are shown in blue, and entrapment and decoy PSMs in yellow. The top panel displays PSM scores from WinnowNet, the left panel shows scores from MS^{2}Rescore, and the central panel illustrates the correlation between WinnowNet scores (x-axis) and MS^{2}Rescore scores (y-axis). Solid lines indicate the 1% PSM-level FDR cutoffs for WinnowNet (0.93) and MS^{2}Rescore (-0.07).

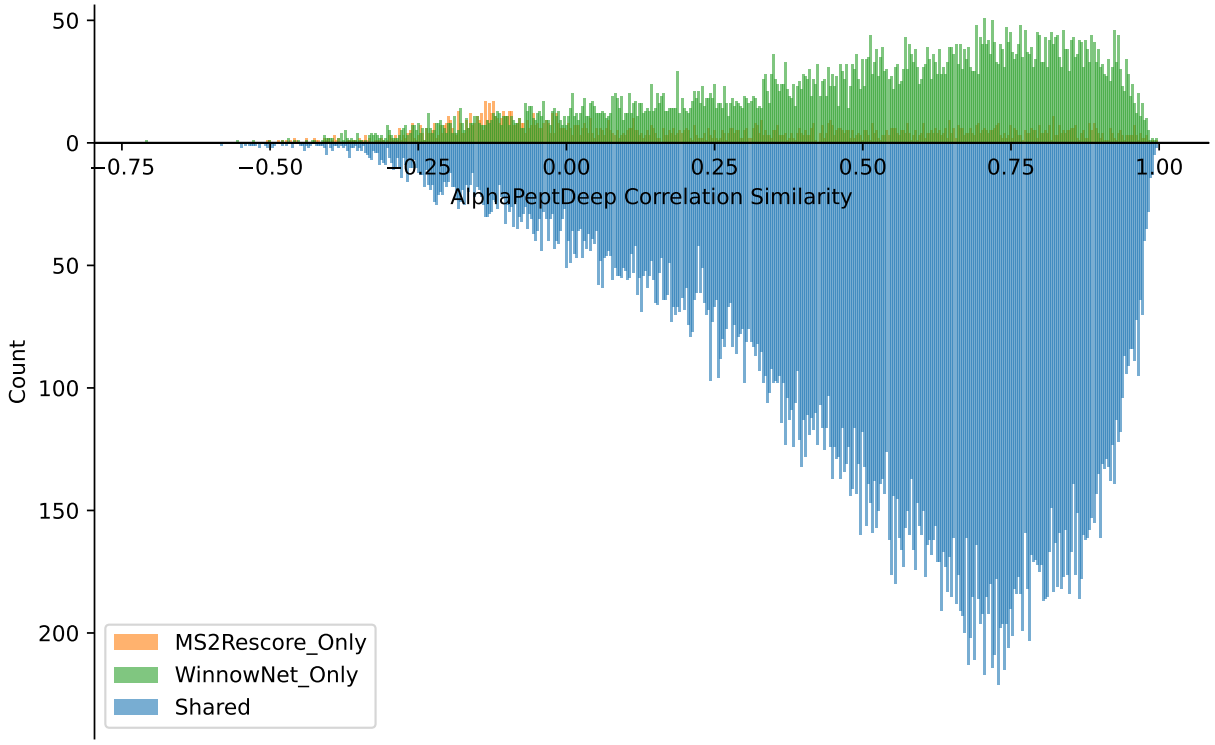


(a) Marine 2 dataset: spectral angle similarity of top-ranked PSMs.

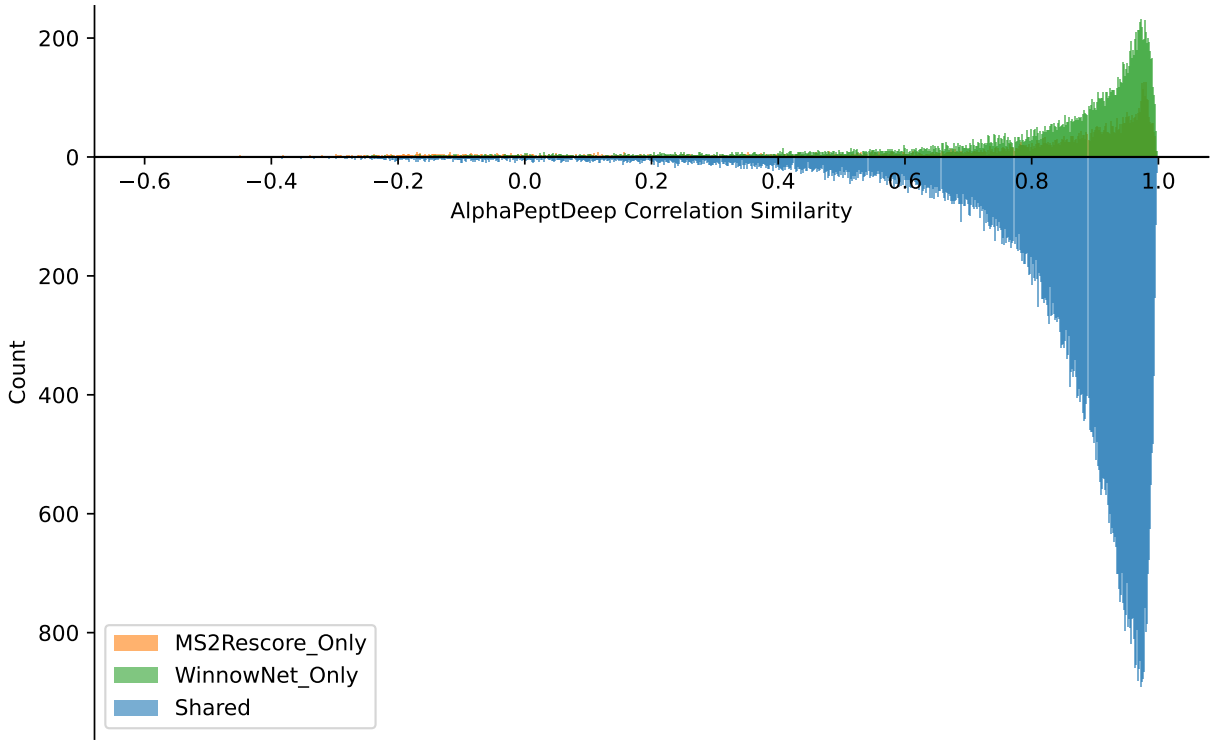


(b) Soil 2 dataset: spectral angle similarity of top-ranked PSMs.

Supplementary Figure 15: Distributions of spectral angle similarity for top-ranked PSMs classified as MS^{2}Rescore-only, WinnowNet-only, or shared by both filtering algorithms. The x-axis represents spectral angle similarity, and the y-axis shows the count of PSMs at each similarity value.

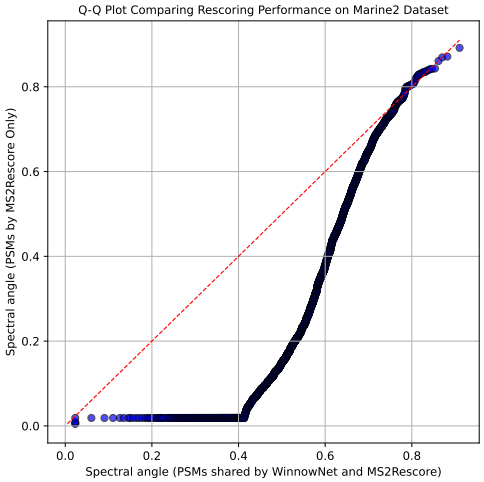


(a) Marine 2 dataset: Pearson correlation similarity of top-ranked PSMs.

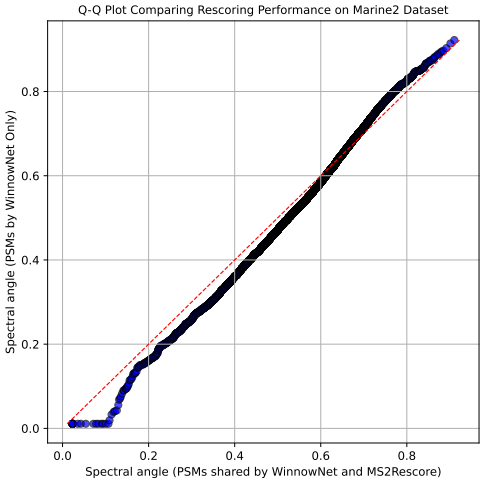


(b) Soil 2 dataset: Pearson correlation similarity of top-ranked PSMs.

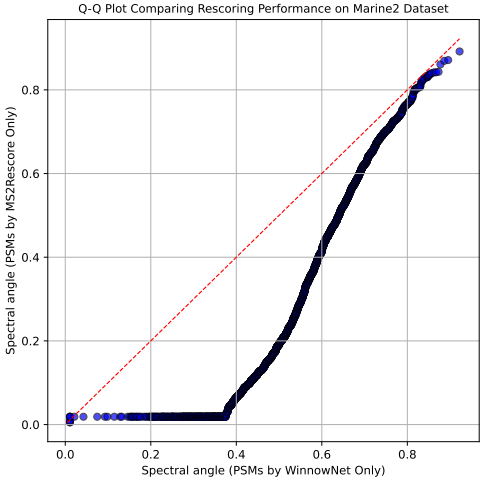
Supplementary Figure 16: Distributions of Pearson correlation similarity for top-ranked PSMs classified as MS^{2}Rescore-only, WinnowNet-only, or shared by both filtering algorithms. The x-axis represents the Pearson correlation similarity, and the y-axis shows the count of PSMs at each similarity value.



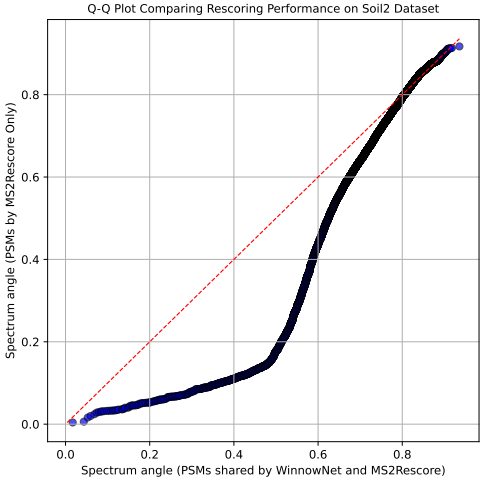
(a) Marine 2: MS^{2}Rescore-only vs. shared PSMs.



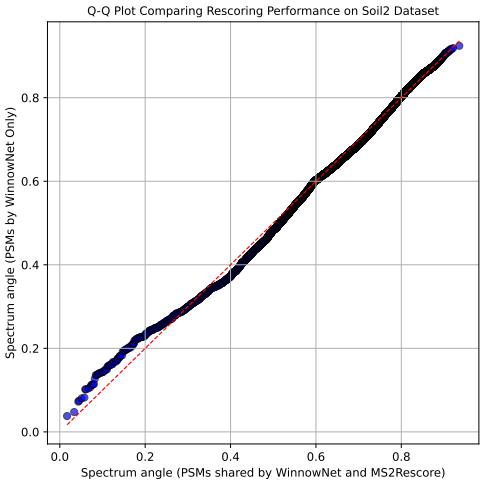
(b) Marine 2: WinnowNet-only vs. shared PSMs.



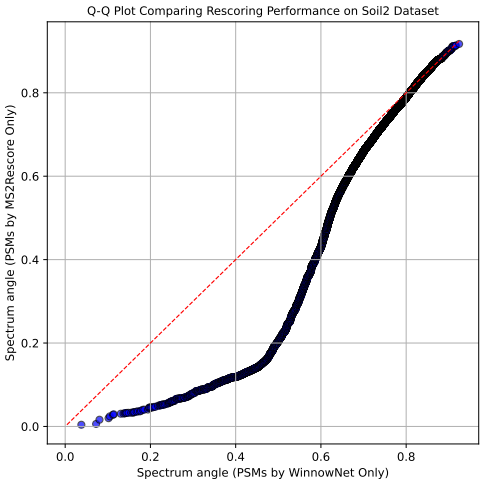
(c) Marine 2: MS^{2}Rescore-only vs. WinnowNet-only PSMs.



(d) Soil 2: MS^{2}Rescore-only vs. shared PSMs.

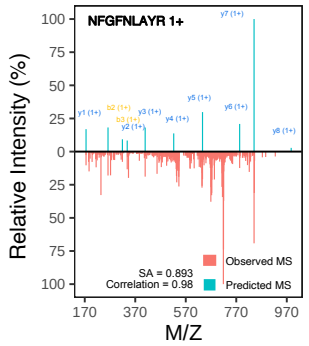


(e) Soil 2: WinnowNet-only vs. shared PSMs.

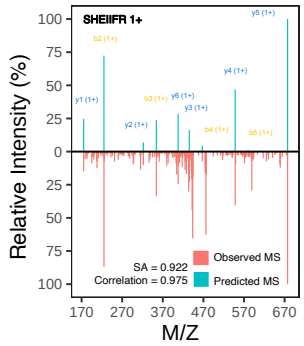


(f) Soil 2: MS^{2}Rescore-only vs. WinnowNet-only PSMs.

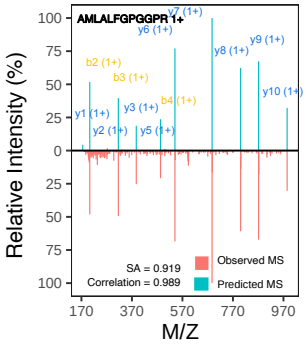
Supplementary Figure 17: Quantile–quantile (QQ) plots comparing spectral angle similarity distributions. The plot is based on top-ranked PSMs across different groups: MS^{2}Rescore-only, WinnowNet-only, and shared PSMs, for the Marine 2 and Soil 2 datasets. Each subplot compares two of the three groups to assess distributional agreement. The x- and y-axes represent spectral angle similarity quantiles for each group.



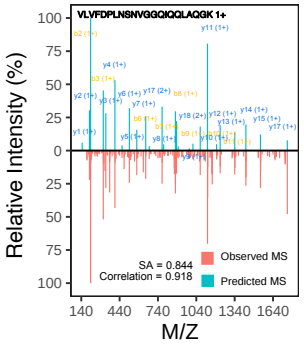
(a) Marine 2 PSM with WinnowNet score of 0.99.



(b) Marine 2 PSM with WinnowNet score of 0.96.



(c) Soil 2 PSM with WinnowNet score of 0.95.



(d) Soil 2 PSM with WinnowNet score of 0.99.

Supplementary Figure 18: Visualizations of four representative PSMs from the Marine 2 and Soil 2 datasets, each annotated with its WinnowNet matching score. These examples illustrate varying levels of spectral match quality, as reflected by the corresponding scores.

2 Supplementary Tables

Supplementary Table 1: The number of MS/MS spectra and proteins of thirteen proteome datasets

Training Datasets	# of spectra	# of proteins	# Training samples	# Validating samples
ProteomeTools	238,012	-[a]	240,000	27,000
Marine 1	139,265	391,847	1,681,198	186,799

Test Datasets	# of spectra	# of proteins[b]
Marine 2	144,841	391,847
Marine 3	128,476	391,847
Soil 1	404,328	3,408,250
Soil 2	494,227	3,408,250
Soil 3	415,234	3,408,250
P1[c]	364,395	123,100
P2	358,674	123,100
P3	352,462	123,100
HG[d]	670,418	4,854,034
HGT[e]	170,709	441,558
Synthetic dataset	1,108,570	17,092

[a] The ProteomeTools dataset comprises a spectral library derived from synthetic peptides representing the human proteome, without the use of full-length protein sequences.
[b] Target proteins excluding entrapment proteins and decoys.
[c] P1, P2, and P3 are the three mock community datasets.
[d] HG: human gut dataset by Orbitrap.
[e] HGT: human gut dataset by timsTOF.

Supplementary Table 2: The list of 27 foreign species used as entrapment

Species Names
Anaerostipes caccae DSM 14662
Brevibacillus laterosporus
Bifidobacterium longum NCC 2705
Bacteroides thetaiotaomicron (strain ATCC 29148 / DSM 2079 / NCTC 10582 / E50 / VPI-5482)
Blautia sp. YL58
Clostridium butyricum E4 str. BoNT E BL5262
Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / DSM 644 / NCIMB 8303)
Enterococcus faecalis
Erysipelatoclostridium ramosum DSM 1402
Haemophilus influenzae
Kocuria rhizophila DSM 348 (ATCC 9341, PCI 1001)
Lactobacillus acidophilus
Lactobacillus plantarum (strain ATCC BAA-793 / NCIMB 8826 / WCFS1)
Lacticaseibacillus casei
Moraxella catarrhalis
Methanospirillum hungatei
Methanothrix soehngenii
Paenibacillus polymyxa
Pasteurella multocida
Pediococcus pentosaceus
Pseudomonas aeruginosa
Rhodotorula glutinis
Ruminiclostridium cellulolyticum (strain ATCC 35319 / DSM 5812 / JCM 6584 / H10)
Saccharomyces cerevisiae
Staphylococcus aureus
Streptococcus pneumoniae (strain ATCC BAA-255 / R6)
Stenotrophomonas rhizophila

Supplementary Table 3: Filenames of MS datasets and protein databases in this study

Datasets	Filenames of mass spectrometry datasets	Filenames of protein databases[a]
ProteomeTools	MS Scans selected from "traintest_hcd.hdf5"	Peptide collected from "traintest_hcd.hdf5"
Marine1 Marine2 Marine3	OSU_D10_FASP_Elite_03202014* OSU_D2_FASP_Elite_02262014* OSU_D7_FASP_Elite_03172014*	Marine_shuffled.fasta
Soil1 Soil2 Soil3	Angelo_08202013_P1_3040cm_MB_FASP_Elite* Angelo_09272013_P1_1020cm_MB_FASP_Elite* Angelo_10022013_P1_3040cm_MB_FASP_Elite*	Soil_shuffled.fasta
P1 P2 P3	P1_run3* P2_run3* P3_run3*	Mock_Comm_RefDB_V3_shuffled.fasta
HG	cdj-DDA*	stool_nrnew_shuffled.fasta
HGT	F08_Rep*	GUT_DB2MG_shuffled.fasta
Synthetic dataset	P1_run3* (PXD035759)	synthetic.fasta with 27 species protein sequences from uniprot

[a] The protein databases utilized in this study can be accessed at [`https://github.com/Biocomputing-Research-Group/WinnowNet4Review`].

Supplementary Table 4: The training and validating accuracies when WinnowNet models converged

Datasets	Training Accuracy	Training Accuracy*[a]	Validating Accuracy	Validating Accuracy*
ProteomeTools	99.49%	99.32%	99.18%	99.05%
Marine 1	96.76%	94.06%	95.11%	93.77%

Datasets	Training Precision	Training Precision*	Validating Precision	Validating Precision*
ProteomeTools	99.46%	99.23%	99.22%	98.85%
Marine 1	95.34%	94.89%	93.58%	92.97%

[a] Column name with "*" presents the accuracy or precision for the CNN-based model; Column name without "*" presents the accuracy for the self-attention-based model.

Supplementary Table 5: Benchmarking tools used in this study

Benchmarking tools	Publication date	Recent software update date (toolkit[a])	Release date (version) used in this study
Comet	2013	2024-10-14	2018-01(V2)
Myrimatch	2007	2020-09-30 (Bumbershoot)	2012 (V2.2)
MS-GF+	2014	2024-03-26	2019-04-08
Sipros-Ensemble	2018	2018-04-13	2018-04-13 (V1.3)
FragPipe	2017	2024-05-29	2024-05-29 (V21.0)
MS^2Rescore	2019	2024-10-02	2024-07-21 (V3.1.0)
Percolator	2007	2024-06-20	2018-02–5 (V3.2)
Q-ranker	2009	2024-10-18 (Crux)	2018-05-20 (V3.2)
PeptideProphet	2002	2024-07-31 (TPP)	2019-05-30 (V5.2.0)
iProphet	2011	2024-07-31 (TPP)	2019-05-30 (V5.2.0)
DeepFilter	2021	2021-11-06	2021-11-06 (V1.0)
Peaks	2023	2024-11-20	2024-11-20 (12.5)
AlphaPept	2024	2023-05-22	2023-05-22 (V0.5)

[a] Toolkit contains the benchmarking tool as it is not available as a standalone version.

Supplementary Table 6: 11 PSM features used in WinnowNet* (CNN-based WinnowNet)

1	Q-value	Expected proportion of false identifications among all identifications with a given score or higher (Eq. 2)
2	ΔQ_n	Fractional difference between current and second best Q-value
3	ΔQ_n^L	Fractional difference between current and the L_{th} best Q-value, L indicates the top-L PSM candidates reported by a search engine
4	$Mass$	The measured mass $[M+H]^+$
5	ΔM	The difference between theoretical and measured mass
6	$abs(\Delta M)$	The absolute value of the difference between theoretical and measured mass
7	$pepLen$	The length of the matched peptide, in residues
8	$enzInt$	Number of missed internal enzymatic (tryptic) sites
9-11	$charge$ 1-3	Three Boolean features indicating the charge state

Supplementary Table 7: Performance comparison for three training strategies for WinnowNet* (CNN-based WinnowNet)

[a]	Easy-difficult Strategy (506 mins)			Reverse Strategy (736 mins)			Random Strategy (1,794 mins)		
Acc[b]	94.05%(±0.19%)			87.98% (±0.17%)			92.98%(±0.33%)		
Prec[c]	95.18%(±0.19%)			80.00%(±0.24%)			92.86%(±0.31%)		
[d]	#PSMs	#peps	#pros	#PSMs	#peps	#pros	#PSMs	#peps	#pros
M2[e]	55,890	38,454	8,982	17,819	10,831	2,587	54,271	37,596	8,576
M3	63,209	41,594	10,120	20,739	12,193	3,104	61,873	40,716	9,650
S2	15,4183	41,006	8,563	41,093	13,237	3,452	150,772	39,207	8,126
S3	140,015	37,976	7,486	38,102	11,230	3,279	136,670	35,870	7,147
P1	122,459	34,076	8,388	37,326	11,035	3,081	121,396	33,315	8,019
P2	132,998	48,719	11,193	38,120	13,296	3,612	128,098	47,016	10,596
P3	107,197	36,589	9,614	30,193	12,031	3,370	105,353	35,911	9,072
HG	282,939	168,564	38,145	63,242	43,888	9,092	278,310	166,104	36,437

[a] Training strategy: Each entry is formatted as strategy (training time consumed).
[b] Acc: Accuracy on test datasets, reported as mean accuracy (± standard error of the mean).
[c] Prec: Precision on test datasets, reported as mean precision (± standard error of the mean).
[d] Number of identifications at the PSM, peptide, and protein levels.
[e] Datasets: M2 and M3 represent two marine metaproteomes; S2–3 represent two soil metaproteomes; P1–3 represent three mock communities; HG represents a human gut metaproteome.

Supplementary Table 8: Performance comparison for four training strategies for WinnowNet (self-attention-based WinnowNet)

	Marine 1 only Strategy (4,111 mins)			Easy-difficult Strategy (1,934 mins)		
Accuracy[a]	94.11%(±0.26%)			96.86%(±0.21%)		
Precision[b]	94.47%(±0.21%)			96.25%(±0.19%)		
	#PSMs	#Peptides	#Proteins	#PSMs	#Peptides	#Proteins
Marine 3	63,150	41,528	10,102	66,432	43,071	10,416
Soil 3	139,209	37,794	7,429	142,711	38,677	7,643
P3	106,653	36,310	9,602	108,922	37,735	9,788
Human Gut	269,301	168,936	37,916	271,272	170,433	38,101

	Reverse Strategy (2,687 mins)			Random Strategy (5,819 mins)		
Accuracy[a]	88.11%(±0.26%)			94.37%(±0.35%)		
Precision[b]	82.12%(±0.28%)			93.92%(±0.32%)		
	#PSMs	#Peptides	#Proteins	#PSMs	#Peptides	#Proteins
Marine 3	23,149	13,975	2,710	62,776	41,329	10,066
Soil 3	40,109	12,277	3,315	139,412	37,941	7,401
P3	29,194	11,219	3,213	104,172	35,114	9,357
Human Gut	66,479	44,827	9,214	269,877	169,121	37,950

[a] Reported as the best accuracy (± standard error of the mean) at 1% FDR.
[b] Reported as the best precision (± standard error of the mean) at 1% FDR.

Supplementary Table 9: Identification performance of WinnowNet* (CNN-based) integrated with Sipros-Ensemble under different noisy peaks tolerances.

	Marine3	Soil3	P3	HG
Noisy Peaks Tolerance	PSM 1%			
1 *Da*	64,132	142,711	108,922	271,212
0.1 *Da*	65,311	143,212	110,136	269,301
	Peptide 1%			
1 *Da*	42,511	37,976	38,047	189,433
0.1 *Da*	43,110	38,239	38,277	188,653
	Protein 1%			
1 *Da*	10,001	6,036	8,729	35,623
0.1 *Da*	9,290	7,172	9,539	38,133

Supplementary Table 10: Number of homologous PSM candidates for different test datasets compared with training datasets

	Comet		MS-GF+		Myrimatch	
Metaproteomes	Total	Homologous	Total	Homologous	Total	Homologous
Soil1	2,922,412	155,76	2,332,004	13,921	2,030,093	12,812
Soil2	2,493,111	13,923	2,104,529	11,083	1,839,341	9,384
Soil3	1,804,726	8,392	1,602,423	7,824	1,452,452	6,193
P1	1,436,893	17,419	1,232,549	16,838	1,031,549	15,412
P2	1,423,349	17,848	1,421,690	17,101	1,385,931	17,012
P3	1,542,123	17,493	1,512,354	17,021	1,412,480	16,948
HG	6,230,490	41,312	5,611,314	36,230	4,134,123	29,113

Supplementary Table 11: Decrease in identifications after the removal of homologous peptides for various metaproteomes at 1% FDR.

	Metaproteomes	Soil 1			Soil 2			Soil 3			Human Gut		
Level[a]	Filter[b] \ Search[c]	C	M	G	C	M	G	C	M	G	C	M	G
PSM	W/O	-14	-20	-18	-55	-20	-83	-4	-6	-2	-115	-249	-8
	P	-18	-25	-21	-61	-24	-91	-7	-11	-8	-129	-255	-12
	Q	-18	-24	-21	-60	-24	-88	-7	-11	-8	-127	-251	-12
	PP	-14	-20	-18	-55	-20	-82	-4	-6	-3	-112	-240	-7
	I	-14	-20	-18	-55	-20	-82	-4	-6	-3	-112	-240	-7
	DF	-26	-33	-21	-61	-29	-88	-10	-17	-4	-133	-261	-17
	W	-31	-37	-24	-65	-32	-91	-12	-18	-6	-141	-270	-18
Peptide	W/O	-4	-4	-5	-8	-8	-14	-3	-4	-2	-48	-99	-4
	P	-6	-6	-8	-10	-10	-19	-5	-5	-4	-50	-103	-5
	Q	-6	-6	-8	-10	-10	-19	-5	-5	-4	-50	-103	-5
	PP	-4	-4	-5	-8	-8	-14	-3	-4	-2	-48	-99	-4
	I	-4	-4	-5	-8	-8	-14	-3	-4	-2	-48	-99	-4
	DF	-8	-8	-11	-12	-12	-21	-7	-8	-5	-55	-108	-6
	W	-10	-10	-12	-13	-13	-23	-8	-9	-6	-61	-112	-8
Protein	W/O	-2	-1	-1	-5	-5	-6	-1	-1	-1	-22	-35	-2
	P	-2	-1	-1	-5	-5	-8	-1	-1	-1	-25	-35	-2
	Q	-2	-1	-1	-5	-5	-8	-1	-1	-1	-25	-35	-2
	PP	-2	-1	-1	-5	-5	-6	-1	-1	-1	-21	-33	-2
	I	-2	-1	-1	-5	-5	-6	-1	-1	-1	-21	-33	-2
	DF	-3	-2	-2	-5	-5	-9	-1	-2	-1	-29	-41	-3
	W	-3	-2	-2	-5	-5	-10	-2	-2	-1	-32	-44	-4

[a] Level: Number of identifications at PSM/Peptide/Protein level at FDR 1%
[b] W/O, identification result of corresponding search engines without any filtering algorithm; P, Percolator; Q, Q-ranker; PP, PeptideProphet; I, Prophet; DF, DeepFilter; W, WinnowNet*
[c] Search engines: C, Comet; M, Myrimatch; G, MS-GF+

Supplementary Table 12: Decrease in identifications after the removal of homologous peptides for metaproteomes for the mock community at FDR 1%

Level[a]	Metaproteomes	P1			P2			P3		
	Filter[b] \ Search[c]	C	M	G	C	M	G	C	M	G
PSM	W/O	-518	-1092	-1129	-1284	-1137	-1568	-376	-778	-978
	P	-620	-1190	-1509	-1419	-1274	-1729	-485	-895	-1111
	Q	-612	-1177	-1485	-1402	-1257	-1701	-471	-882	-1096
	PP	-503	-1075	-1360	-1266	-1118	-1546	-359	-760	-962
	I	-503	-1075	-1360	-1266	-1118	-1546	-359	-760	-962
	DF	-786	-1372	-1695	-1603	-1443	-1895	-647	-1063	-1280
	Wt	-827	-1399	-1712	-1633	-1482	-1935	-688	-1103	-1320
Peptide	W/O	-124	-241	-257	-349	-318	-375	-98	-317	-212
	P	-157	-275	-298	-382	-353	-407	-124	-353	-238
	Q	-155	-272	-295	-380	-352	-405	-123	-350	-236
	PP	-121	-240	-255	-344	-315	-370	-97	-314	-210
	I	-121	-240	-255	-344	-315	-370	-97	-314	-210
	DF	-201	-320	-346	-427	-398	-455	-172	-454	-285
	W	-225	-347	-388	-472	-423	-479	-199	-423	-317
Protein	W/O	-55	-82	-91	-125	-115	-128	-44	-77	-82
	P	-63	-95	-103	-134	-129	-140	-52	-86	-93
	Q	-63	-94	-102	-133	-129	-140	-52	-84	-92
	PP	-53	-82	-90	-121	-113	-126	-44	-75	-82
	I	-53	-82	-90	-121	-113	-126	-44	-75	-82
	DF	-70	-102	-109	-142	-137	-148	-60	-97	-101
	W	-75	-106	-118	-151	-145	-156	-68	-106	-110

[a] Level: Number of identifications at PSM/Peptide/Protein level at FDR 1%
[b] W/O, identification result of corresponding search engines without any filtering algorithm; P, Percolator; Q, Q-ranker; PP, PeptideProphet; I, Prophet; DF, DeepFilter; W, WinnowNet*
[c] Search engines: C, Comet; M, Myrimatch; G, MS-GF+

Supplementary Table 13: PSM/Peptide/Protein identification results on the Human Gut timsTOF dataset using three metaproteomics pipelines at 1% FDR.

Discovery Levels (1% FDR)	Filters	Protein idenfication Pipelines		
		FragPipe	Peaks	AlphaPept
PSM	Original Filter	48,369	52,843	49,731
	WinnowNet	49,660	54,103	51,396
Peptide	Original Filter	16,139	16,839	16,985
	WinnowNet	16,734	17,825	17,732
Protein	Original Filter	4,488	4,633	4,674
	WinnowNet	4,619	4,710	4,786

Supplementary Table 14: PSM/Peptide/Protein identification results on three datasets using four metaproteomics pipelines at 1% FDR.

Metaproteome Datasets	Discovery Levels (1% FDR)	Filters	Protein identification Pipelines			
			Sipros Ensemble	FragPipe	Peaks	AlphaPept
Marine 3	PSM	Original Filter	61,190	47,970	46,727	43,791
		WinnowNet	66,432	53,276	52,789	49,841
	Peptide	Original Filter	40,519	25,658	24,864	23,895
		WinnowNet	43,071	31,769	30,091	29,857
	Protein	Original Filter	9,500	9,909	9,001	8,426
		WinnowNet	10,416	10,277	9,327	8,796
Soil 3	PSM	Original Filter	131,522	107,197	109,765	106,795
		WinnowNet	142,711	110,563	121,985	117,480
	Peptide	Original Filter	34,103	29,372	28,641	27,926
		WinnowNet	38,677	31,144	32,792	31,092
	Protein	Original Filter	6,815	7,061	6,802	6,593
		WinnowNet	7,643	7,369	7,184	7,007
P3	PSM	Original Filter	103,343	100,749	101,218	102,993
		WinnowNet	108,922	105,576	106,388	107,385
	Peptide	Original Filter	35,414	26,023	26,513	27,138
		WinnowNet	37,735	27,176	27,690	27,945
	Protein	Original Filter	8,674	9,276	9,104	9,181
		WinnowNet	9,788	9,613	9,438	9,565
Human Gut	PSM	Original Filter	264,753	255,376	248,241	243,734
		WinnowNet	271,272	263,319	257,943	253,598
	Peptide	Original Filter	158,865	143,482	139,840	136,742
		WinnowNet	170,433	151,297	148,756	145,905
	Protein	Original Filter	36,433	36,852	35,986	35,385
		WinnowNet	38,101	37,271	36,739	36,277

Supplementary Table 15: PSM/Peptide/Protein identification results on P1 dataset with 27 foreign species using four metaproteomics pipelines at 1% FDR.

Metaproteome Dataset	Discovery Levels (1% FDR)	Filters	Protein identification Pipelines			
			Sipros Ensemble	FragPipe	Peaks	AlphaPept
P1	PSM	Original Filter	87,275	84,448	85,033	82891
		WinnowNet	91,271	88,691	89,013	87381
	Peptide	Original Filter	24,633	20,235	21,155	19,243
		WinnowNet	25,827	21,138	22,849	20,195
	Protein	Original Filter	7,126	7,007	7,015	6,715
		WinnowNet	7,389	7,272	7,267	6,942

Supplementary Table 16: PSM/Peptide/Protein identification results on marine and soil datasets using three search engineers and eight filters at 1% FDR.

Datasets	Level[a]	S.[b]	Filters[c]								
			W/O	MS	P	Q	PP	I	DF	Win*	Win
Marine 2	PSM	C	32,744	35,346	34,847	33,984	32,245	33,033	37,883	40,762	42,201
		M	25,976	29,179	27,912	27,603	26,211	26,283	30,513	35,498	37,372
		G	34,875	39,157	37,215	36,304	34,927	35,106	40,247	43,376	44,828
	Pep.	C	23,369	25,955	25,216	24,749	22,825	23,193	26,514	28,639	29,487
		M	18,864	19,518	19,716	19,543	18,623	18,720	21,236	23,675	24,648
		G	23,683	26,068	25,641	24,698	23,601	23,697	27,363	29,271	30,036
	Pro.	C	6,991	7,517	7,498	7,395	6,896	6,912	7,892	8,304	8,455
		M	6,237	6,585	6,449	6,415	6,175	6,178	6,745	7,282	7,436
		G	7,357	8,021	7,838	7,568	7,315	7,315	8,176	8,514	8,723
Marine 3	PSM	C	39,148	42,792	41,862	41,003	38,524	38,736	44,271	47,734	48,911
		M	41,205	44,479	42,185	41,316	41,019	41,276	47,036	48,572	49,860
		G	40,373	43,714	41,538	40,627	39,863	40,094	44,035	46,019	48,373
	Pep.	C	25,361	27,840	27,309	26,574	25,277	25,412	28,977	29,882	31,291
		M	26,138	28,789	28,015	27,114	25,930	26,018	30,011	31,035	32,298
		G	25,947	27,319	26,674	26,005	25,498	25,580	27,895	29,252	29,941
	Pro.	C	7,339	7,886	7,813	7,632	7,003	7,017	8,187	8,695	8,781
		M	6,417	7,958	7,802	7,533	6,408	6,218	8,322	8,596	8,673
		G	7,702	8,104	8,036	7,824	7,617	7,498	8,387	8,681	8,894
Soil 1	PSM	C	71,547	78,950	77,686	76,241	70,391	70,873	83,139	87,016	90,223
		M	72,994	85,750	81,237	79,102	71,889	71,948	87,784	89,848	95,921
		G	86,923	92,058	90,326	87,954	85,633	86,024	95,492	98,439	101,145
	Pep.	C	25,495	29,071	28,557	28,146	25,109	25,984	30,285	31,698	32,061
		M	22,201	26,759	27,323	27,105	22,039	22,114	28,610	29,894	30,836
		G	26,987	29,542	28,552	28,219	26,753	26,716	30,479	31,362	32,181
	Pro.	C	6,613	7,163	7,095	7,453	6,387	6,387	7,385	7,601	7,773
		M	6,016	7,063	6,998	6,991	5,817	5,825	7,389	7,602	7,728
		G	7,386	7,874	7,727	7,646	7,331	7,338	8,009	8,124	8,277
Soil 2	PSM	C	71,699	82,226	80,024	78,490	69,252	70,395	86,749	92,288	93,640
		M	65,329	74,092	71,638	69,032	63,897	64,082	78,104	83,293	84,413
		G	79,410	88,083	84,311	82,179	76,033	77,159	91,282	94,873	95,637
	Pep.	C	23,006	27,706	26,647	26,892	22,973	23,004	28,428	29,726	30,791
		M	23,272	25,177	24,215	23,953	22,845	22,968	25,297	27,503	28,219
		G	24,742	26,500	26,534	26,018	24,319	24,372	28,196	29,271	30,133
	Pro.	C	6,551	7,293	7,274	7,201	6,511	6,405	7,631	7,895	7,982
		M	6,208	6,613	6,501	6,428	6,088	6,091	7,076	7,185	7,312
		G	6,813	7,299	7,326	7,235	6,994	7,001	7,657	7,912	8,066
Soil 3	PSM	C	70,952	80,224	78,732	75,890	66,714	68,101	82,759	87,752	89,768
		M	63,491	72,231	70,328	69,331	62,691	63,725	75,723	81,673	82,076
		G	71,353	79,476	75,814	73,690	68,742	69,793	80,429	87,191	87,204
	Pep.	C	21,355	23,810	24,165	24,357	20,194	20,372	25,287	26,523	27,197
		M	20,930	23,249	21,579	22,091	19,851	20,015	23,074	24,237	24,798
		G	21,772	23,291	23,141	23,763	20,679	20,982	24,179	25,210	26,103
	Pro.	C	5,690	6,203	6,189	6,427	5,513	5,588	6,423	6,668	6,792
		M	5,596	5,876	5,703	5,896	5,485	5,510	6,034	6,215	6,408
		G	6,036	6,697	6,602	6,691	5,921	6,029	6,818	7,057	7,172

[a] Level: Number of identifications at PSM/Peptide/Protein level at FDR 1% using entrapment estimation.
[b] Search engines: C, Comet, M, MyriMatch, G, MS-GF+..
[c] Filters: W/O, identification result of corresponding search engines without any filtering algorithm; MS, MS^{2}Rescore; P, Percolator; Q, Q-ranker; PP, PeptideProphet; I, Prophet; DF, DeepFilter; Win*, CNN-based WinnowNet; Win, Self-attention-based-WinnowNet.

Supplementary Table 17: PSM/Peptide/Protein identification results on mock and human gut datasets using three search engineers and eight filters at 1% FDR.

Datasets	Level[a]	S.[b]	Filters[c]								
			W/O	MS	P	Q	PP	I	DF	Win*	Win
P1	PSM	C	91,800	95,958	97,743	94,908	90,261	89,984	103,115	105,186	106,420
		M	82,626	92,555	93,083	90,985	80,234	80,234	99,617	102,808	104,759
		G	97,073	99,426	97,518	97,050	93,895	93,895	103,269	105,161	106,858
	Pep.	C	25,791	27,187	26,695	26,214	24,856	24,907	27,590	28,377	28,997
		M	24,748	25,406	25,892	25,380	23,596	23,645	27,485	27,897	28,718
		G	26,053	26,379	26,768	23,169	25,088	25,063	27,682	28,227	29,237
	Pro.	C	6,835	7,303	7,228	7,132	6,601	6,601	7,624	7,778	7,904
		M	7,012	7,113	7,080	7,011	6,726	6,831	7,571	7,693	7,755
		G	7,239	7,370	7,341	7,278	6,904	7,012	7,670	7,822	7,935
P2	PSM	C	99,821	106,265	107,062	101,976	99,225	99,225	111,721	113,437	115,459
		M	95,069	104,142	103,804	98,510	94,070	93,877	108,753	111,541	113,770
		G	107,161	107,029	109,240	106,992	105,477	105,477	113,845	115,374	117,387
	Pep.	C	37,940	40,403	39,106	39,024	37,733	37,927	40,203	41,302	41,764
		M	36,830	38,389	38,607	38,488	36,451	36,488	39,303	40,623	41,746
		G	38,590	39,326	39,464	39,456	38,091	38,248	40,766	41,358	42,067
	Pro.	C	9,017	9,545	9,519	9,276	8,955	9,095	9,892	10,116	10,228
		M	9,505	9,655	9,587	9,410	9,375	9,512	10,012	10,208	10,273
		G	9,513	9,611	9,613	9,423	9,371	9,557	10,124	10,246	10,323
P3	PSM	C	78,458	87,382	88,417	87,019	76,798	76,719	91,333	93,596	95,951
		M	72,887	79,596	83,393	90,706	70,956	71,029	86,808	90,117	91,885
		G	86,554	89,189	89,511	89,043	84,530	84,791	91,881	94,625	96,254
	Pep.	C	27,998	30,099	29,605	29,115	26,743	26,771	30,181	30,625	31,585
		M	26,717	28,009	28,683	27,501	25,416	25,390	29,630	30,300	31,208
		G	28,867	29,745	29,835	29,927	27,042	26,987	30,402	31,351	31,957
	Pro.	C	7,904	8,379	8,295	8,167	7,684	7,829	8,694	8,776	8,920
		M	8,177	8,149	8,154	8,126	7,984	8,109	8,590	8,720	8,824
		G	8,371	8,511	8,488	8,512	8,225	8,371	8,842	9,037	9,138
HG	PSM	C	201,411	214,973	215,963	207,687	181,659	181,288	223,854	230,245	237,011
		M	239,899	245,985	247,348	242,170	220,770	220,997	254,552	255,558	258,556
		G	229,934	237,091	239,924	231,644	210,702	211,135	248,967	252,821	256,094
	Pep.	C	140,403	149,870	149,758	148,215	130,288	130,155	153,745	158,103	159,827
		M	147,576	148,956	149,308	147,456	135,919	136,198	151,389	155,691	158,478
		G	148,550	149,482	150,982	149,098	137,400	137,541	155,247	159,829	159,651
	Pro.	C	30,894	33,678	33,268	32,916	29,420	30,035	34,318	35,776	35,564
		M	31,813	34,015	33,325	32,728	30,234	30,771	34,178	35,438	35,503
		G	32,289	34,187	33,699	33,262	30,912	31,396	34,456	36,158	36,143

[a] Level: Number of identifications at PSM/Peptide/Protein level at FDR 1% using entrapment estimation.
[b] Search engines: C, Comet, M, MyriMatch, G, MS-GF+..
[c] Filters: W/O, identification result of corresponding search engines without any filtering algorithm; MS, MS^{2}Rescore; P, Percolator; Q, Q-ranker; PP, PeptideProphet; I, Prophet; DF, DeepFilter; Win*, CNN-based WinnowNet; Win, Self-attention-based-WinnowNet.

Supplementary Table 18: Number of identifications reported by the original publications of the datasets used for comparison.

Datasets\Identifications	PSM	Peptide	Protein	FDR thresholds[a]
Marine2 [1]	22,669	19,800	3,613	Controlled at 1% FDR at peptide level
Marine3	25,157	19,668	3,513	
Soil1 [2]	-[b]	-	4,716	Controlled at 1% FDR at peptide level
Soil2	-	-	3,019	
Soil3	-	-	3,015	
P1 [3]	124,940	-	-	Controlled at 5% FDR at PSM level
P2	131,017	-	-	
P3	111,800	-	-	
HG [4]	-	-	30,062	Controlled at 1% FDR at protein level

[a] The FDR thresholds reported in the original publications.
[b] "-" denotes the number of identifications not reported.

Supplementary Table 19: Comparison of PSM identification results between WinnowNet and MS^{2}Rescore using the output by MS-GF+.

Datasets	Marine 2	Marine 3	Soil 1	Soil 2	Soil 3	P1	P2	P3	HG
#Top-scored unfiltered target PSMs by WinnowNet	98,426	93,266	273,095	313,993	236,346	185,180	187,859	172,120	483,625
#Top-scored unfiltered decoy PSMs by WinnowNet	43,115	32,476	147,624	190,713	130,198	64,139	60,264	63,662	183,196
# Estimated true positive PSMs by WinnowNet[a]	55,311	60,790	125,471	123,280	106,148	121,041	127,595	108,458	300,429
#Top-scored unfiltered target PSMs by MS^{2}Rescore	96,376	90,634	270,398	310,918	234,483	182,459	184,876	169,721	474,128
#Top-scored unfiltered decoy PSMs by MS^{2}Rescore	45,165	35,108	150,321	193,788	132,061	66,860	63,247	66,061	192,693
# Estimated true positive PSMs by MS^{2}Rescore[a]	51,211	55,526	120,077	117,130	102,422	115,599	121,629	103,660	281,435
# PSM identifications obtained by both methods	36,389	41,358	88,957	80,601	73,835	91,036	99,602	82,495	239,093
# PSM identifications obtained by WinnowNet only	8,439	7,015	12,188	15,036	13,369	15,882	14,168	13,759	19,463
# PSM identifications obtained by MS^{2}Rescore only	2,768	2,356	3,101	8,164	5,641	6,482	4,540	6,694	6,892
The union of PSMs identified by both methods	47,596	50,729	104,246	103,119	92,845	113,400	118,310	102,948	265,448

[a] The number of true positive PSMs was estimated by #targets - #decoys.

3 Supplementary Note

3.1 Information of benchmarking datasets

A total of thirteen datasets were used in our experiments, originating from six biological studies:

- **ProteomeTools dataset**: A synthetic human proteome library generated by the ProteomeTools project [5].
- **Marine metaproteomes (Marine 1, 2, 3)** : Three datasets collected from the coastal surface waters of Monterey Bay, California [1].
- **Soil metaproteomes (Soil 1, 2, 3)**: Three datasets obtained from the Angelo Coast Range Reserve meadow [2].
- **Mock community metaproteomes (P1, P2, P3)**: Three datasets derived from an artificially assembled microbial community consisting of 32 species and strains [3].
- **Human gut metaproteome (HG)**: One dataset generated from fecal samples of human patients [4].
- **Human gut metaproteome (timsTOF dataset) (HGT)**: One dataset obtained from fecal samples of human patients, analyzed using a trapped ion mobility spectrometry (timsTOF) + TOF mass spectrometry approach [6].
- **Synthetic microbial mixture**: One dataset constructed from a four-species microbial mixture [7].

Tab. 3 summarizes the filenames of the mass spectrometry datasets and protein databases used in this study.

ProteomeTools dataset: This dataset was derived from a synthetic human proteome library, constructed using tryptic peptides corresponding exclusively to the human proteome. The library comprises 330,286 unique synthesized peptides, collectively covering 19,840 reviewed human genes from the UniProt/SwissProt database (version 2016-07-20) [5]. Each peptide pool and LC-MS run contained approximately 1,000 individually synthesized peptides, produced via the Fmoc-based solid-phase synthesis strategy. For peptides requiring cysteine modification, a carboxyamidomethylated cysteine building block was incorporated during synthesis. Prior to analysis, dried peptide pools were solubilized in 100% DMSO to a concentration of 10 pmol/μl by vortexing for 30 minutes at room temperature, then diluted to 10% DMSO with 1% formic acid in HPLC-grade water. Approximately 200 fmol of each peptide was analyzed using a Dionex 3000 HPLC system (Thermo Fisher Scientific) equipped with in-house packed C18 columns: a 75 μm × 2 cm trap column (5 μm Reprosil Pur ODS-3) and a 75 μm × 40 cm analytical column (3 μm Reprosil Gold 120 C18). The LC system was directly coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Each peptide pool was analyzed with a full MS scan in the Orbitrap at a resolution of 60,000, followed by HCD MS/MS scans at 28% normalized collision energy (resolution 15,000), and CID scans at 35% normalized collision energy. Each sample in our experiments includes up to two technical replicates.

Marine and soil metaproteomes: These datasets were analyzed using the Multidimensional Protein Identification Technology (MudPIT) approach [8]. Proteins were extracted from environmental samples, precipitated overnight with trichloroacetic acid, and washed with acetone. The resulting pellets were resolubilized in 6 M guanidine hydrochloride, reduced with 10 mM dithiothreitol (DTT), and digested with trypsin. For multidimensional separation, a back column was packed sequentially with 3 cm of strong cation exchange (SCX) resin (5 μm, 100 Å Luna SCX, Phenomenex) and 3 cm of reversed-phase (RP) media (5 μm, 200 Å Aqua C18, Phenomenex). Peptide samples were loaded into the column using a pressure cell. The column was equilibrated with HPLC Solvent A (5% acetonitrile, 95% H_2O, 0.1% formic acid), followed by separation on an analytical column containing 15 cm of C18 RP media. Mass spectrometry analysis was performed using an LTQ Orbitrap Elite instrument (Thermo Scientific). Tandem mass spectra were acquired across 11 consecutive online SCX fractions, with MS1 and MS2 scans recorded at resolutions of 30,000 and 15,000, respectively. All mass spectrometry data and corresponding metaproteome databases have been deposited in public data repositories as described previously in the respective studies. The three marine datasets and the three soil datasets represent biological replicates. Each marine dataset includes one technical replicate, whereas each soil dataset includes two technical replicates.

Mock community datasets: These datasets were obtained from the "EQUAL PROTEIN AMOUNT" microbial mock communities [3], comprising 32 species and strains across Archaea, Bacteria, Eukaryotes, and Bacteriophages, each contributing an equal amount of protein to the community. For our experiments, we used metaproteomes that were fractionated into twelve parts via a two-dimensional LC-MS/MS workflow. Protein extracts underwent overnight digestion at 37°C. Peptides were eluted from the strong cation exchange (SCX) column into the C18 pre-column through a series of twelve salt plug injections (0–2000 mM NaCl, 20 μl per injection), delivered by the autosampler. After each salt plug, the pre-column was connected in-line with a 50 cm × 75 μm analytical column for peptide separation using a 120-minute gradient. Mass spectrometry analysis was carried out using a Q Exactive Plus hybrid quadrupole-Orbitrap instrument (Thermo Fisher Scientific). The three mock community datasets are biological replicates, each with one technical replicate.

Human gut dataset: This dataset was derived from microbial cells isolated from fecal samples of colorectal cancer (CRC) patients and healthy individuals [4]. We used data-dependent acquisition (DDA) data for our analyses. Briefly, fecal samples were suspended in 45 mL of phosphate-buffered saline (PBS), vortexed for 45 minutes, and centrifuged at 500 × g for 5 minutes to remove food debris. Proteins were extracted, alkylated with iodoacetamide (IAA), and digested overnight with trypsin. The resulting peptides were purified and concentrated using Pierce C18 spin columns (Thermo Fisher Scientific), and quantified via the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific). Peptides were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) coupled to a Nano ACQUITY UPLC system. Full MS scans were acquired at a resolution of 120,000, followed by HCD MS/MS scans at a resolution of 30,000. The number of technical replicates for the human gut dataset is one.

Human gut metaproteome (timsTOF dataset): This dataset was generated from a natural human fecal microbiome sample collected from a 33-year-old omnivorous, non-smoking woman. The sample was immediately homogenized, treated with RNAlater, aliquoted, frozen, and stored at -20°C until distribution to participating laboratories. Metaproteomic analysis was performed using a reversed-phase liquid chromatography (RP-LC) approach. Protein extraction involved phenol extraction followed by solvent washing and dissolution in a urea solution. Protein concentrations were measured using the Amido Black Assay, and 100 µg of protein was subjected to tryptic digestion using the Filter-Aided Sample Preparation (FASP) method. The resulting peptides were acidified with 0.5% trifluoroacetic acid (TFA) prior to LC-MS analysis. Peptides were separated using an UltiMate 3000 nanoHPLC system (Thermo Fisher Scientific) equipped with a C18 reversed-phase column (Acclaim PepMap RSLC, 75 µm × 50 cm, 2 µm particle size, Thermo Fisher). A 120-minute linear gradient from 5% to 35% acetonitrile in 0.1% TFA was applied at a flow rate of 400 nL/min. Ionization was achieved via captive spray ionization (CSI), and mass spectrometry analysis was carried out on a timsTOF Pro instrument (Bruker). Data acquisition was conducted in data-dependent acquisition DDA mode. The number of technical replicates for this timsTOF dataset is three.

Synthetic microbial mixture dataset: This dataset was derived from a synthetic microbial community composed of four distinct microorganisms [7]: *Ruminiclostridium cellulolyticum H10 (ATCC 35319)*, *Desulfovibrio vulgaris Hildenborough (ATCC 29579)*, *Methanospirillum hungatei JF1 (txid323259)*, and *Methanosaeta concilii (txid990316)*. Mass spectrometry (MS) data were acquired using a Fourier Transform Mass Spectrometry (FTMS) analyzer in profile mode, with a resolution of 240,000 over a mass range of 375–1500 m/z. Following higher-energy collisional dissociation (HCD) activation, MS/MS spectra were recorded in centroid mode using an ion trap analyzer, with a normalized collision energy of 28–31%, adjusted based on charge state and precursor selection. Quad-culture samples were selected for full metaproteomic analysis. Protein extraction was performed following a previously described protocol: cell pellets were washed with nanopure water and lysed in buffer (10 mM Tris-HCl, 1% SDS, 0.1 M DTT) at 60°C for 1 hour. Lysates were centrifuged, and supernatants were collected. Proteins were precipitated overnight with trichloroacetic acid at 4°C, pelleted by centrifugation, washed three times with ice-cold acetone, and then resuspended in guanidine buffer. Protein concentration was determined using the Bicinchoninic Acid (BCA) Assay. For digestion, 20 mg of protein was processed via filter-aided sample preparation (FASP) and treated with a Trypsin/LysC enzyme mixture. Peptides were analyzed using two-dimensional liquid chromatography coupled to tandem mass spectrometry (2D-LC-MS/MS) on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, USA) at the IDeA National Resource for Quantitative Proteomics.

Database search parameters The versions of the database search engines used for benchmarking are provided in Table 5. The precursor and fragment mass tolerances were set to 0.09 Da and 0.01 Da, respectively. Peptides within the mass range of 700 to 7000 Da were considered. Trypsin/P was selected as the proteolytic enzyme, with a maximum allowance of three missed cleavages. More detailed parameters' setting can be found in Github via [`https://github.com/Biocomputing-Research-Group/WinnowNet4Review/tree/main/Performance%20Comparison%20to%20State-of-the-art%20Filtering%20Algorithms`]

4 Supplementary Methods

4.1 Input representation construction

The WinnowNet* input comprises two representations: the spectrum representation and the PSM representation. The spectrum representation is derived from matching fragment ions between the measured spectrum (MS/MS spectrum) and the theoretical spectrum. For each PSM candidate, there exists a measured spectrum and a corresponding theoretical spectrum (visualizations examples are shown in Fig. 5(a) and (b). The peaks of fragment ions were matched between these two spectra, with a tolerance of 0.1 Dalton mass difference to generate as many matches as possible. An example of the matched fragment ions, along with their normalized intensities, is visualized in Fig. 5 (c), and this information is transformed into an $N \times 3$ matrix as the spectrum representation. In this spectrum representation matrix, each row is a 3-dimensional tuple of ($\Delta m/z$, I, P) for a pair of matched fragment ions. Here, $\Delta m/z$ represents the m/z difference between the matching peaks of the measured spectrum and the theoretical spectrum; I is the normalized intensity from the measured spectrum, and P is the normalized isotope distribution from the theoretical spectrum. Up to the top-N matches, ordered by the absolute value of their normalized intensity, are preserved for filling in the spectrum representation matrix. If the number of peak matches is less than N, zeros were filled in the remaining space in the matrix. In our experiments, N was set to 500.

The PSM feature representation vector was constructed by 11 features formulated from the initial PSM score, the measured spectrum, and the peptide sequence from the PSM, as shown in Tab. 6. As for the Q-value of a PSM, we calculated as follows. We used f to denote the score of a PSM reported by a search engine. Then the number of target and decoy identifications with scores better than f are denoted as t and d, respectively. The estimated false discovery rate (FDR) with f as the threshold is given by Eq. 1, and the Q-value for the PSM with the score of f is calculated by Eq. 2.

$$FDR(f) = \frac{d}{t} \tag{1}$$

$$q(f) = \min_{f_i \leq f}(FDR(f_i)) \tag{2}$$

4.2 WinnowNet* architecture

WinnowNet* uses two encoders to capture spectrum presentations and PSM feature presentations, separately. The outputs from these two encoders were concatenated into a 512-dimensional vector and fed into a fully connected layer with a softmax activation function (Classifier). To prevent overfitting, a dropout strategy was employed in the spectrum encoder. The output is a probability, ranging from 0 to 1, that indicates that how likely a PSM is a true match. Further details regarding the two encoders are provided below.

4.2.1 Spectrum encoder with order-invariant property

The spectrum encoder is designed to maintain the order-invariant property for input peaks ([9]), which eliminates the need of a sparse high-dimensional matrix to preserve the m/z information of PSMs as in our previous method, i.e., DeepFilter. The spectrum encoder consists of an affine transform network and a feature extraction network, followed by the design of T-Nets ([9]). The affine transformation network aims to correct geometric distortions caused by the irregular matching features between peaks in measured and theoretical spectra. The convolutional kernel window size was set to 1×3 to ensure the preservation of complete peak information. The affine transformation network comprises two 1D convolutional layers with 64 and 256 kernels, respectively. These are followed by a max-pooling layer and a fully connected layer with 256 input units and 9 output units, which transforms the 9-dimensional vector into a 3×3 matrix. The feature extraction network consists of three 1D convolutional layers with 64, 128, and 1024 kernels, respectively. A max-pooling layer is applied as a symmetric function to aggregate key spectral information. Following this, the first fully connected layer has 1024 input dimensions and 512 hidden units, while the second fully connected layer has 512 input dimensions and 256 output dimensions, both of which utilize the ReLU activation function. The spectrum representation matrix was multiplied with the output matrix from the affine transform network and then fed into the feature extraction network to generate a 256-dimensional vector. This vector serves as the representation matrix for subsequent PSM classification. A dropout layer with a rate of 0.3 was added after the last convolutional layer to prevent overfitting. Additionally, batch normalization was applied to each convolutional layer and two fully connected layers in the feature extraction network.

4.2.2 PSM feature encoder

For the PSM feature encoder, the 11 dimensional vector of PSM feature representation was fed to the PSM feature encoder, which returns a 512-dimensional vector using a single fully connected layer with a ReLU activation function.

4.3 Training

The WinnowNet* model is a binary classifier. To train it, we used a modified cross-entropy loss function (Eq. 3) that takes into account the Q-value (q), which was explained in Section 4.1. In Eq. 3, p_i represents the predicted probability that the i_{th} PSM is a true match. The modified loss function based on the posterior error probability was used in our previous study ([10]). Replacing the posterior error probability with the Q-value in the loss function, we achieved more identifications compared to using the standard cross-entropy loss. WinnowNet was implemented using PyTorch version 1.4.0 and was trained on a workstation with 8 GeForce RTX 2080 Ti GPUs. The learning rate and weight decay were set to 1e-4 and the mini-batch size was set to 256.

$$Loss = -\sum [q_i \log(p_i) + (1 - q_i) \log(1 - p_i)] \tag{3}$$

For the easy task, we employed a ratio of 8:1:1 for training, validation, and testing using the ProteomeTools dataset. For the hard task of training on the Marine dataset, we adjusted the ratio between the training and validation datasets to 9:1. WinnowNet demonstrated convergence on the ProteomeTools dataset with training and validation accuracies of 99.02% and 98.5%, respectively. The Receiver Operating Characteristic Curve (ROC) for the test dataset is illustrated in Fig. 7, indicating that WinnowNet* exhibited robust generalization to the unseen dataset in the easy task, achieving the highest Area under the ROC Curve (AUC) compared to the other three benchmarked database search engines.

4.4 Analysis of the significance for the features learned by CNN-based WinnowNet

To evaluate the significance of the spectrum encoder, we compared the performance of WinnowNet* with the spectrum encoder disabled. We selected two metaproteome datasets, i.e., Marine 2 and P1. Fig. 3 shows the improvements of WinnowNet* and its variants over the best filtering algorithm at PSM/Peptide/Protein level with FDRs controlled at 1% by the target-decoy strategy [11]. We can find that without the spectrum encoder, the PSM identifications of WinnowNet were still slightly better than the existing tools but dropped significantly compared to WinnowNet with the spectrum encoder enabled.

4.5 Impact of precursor mass tolerance

We investigated the impact of varying fragment mass tolerances on the input representations for the CNN-based WinnowNet. Specifically, we evaluated error tolerances of 0.1 Da and 1 Da, with the results summarized in Tab. 9. The findings indicate that increasing the fragment mass tolerance negatively affects the performance of WinnowNet*. This decline in performance may be attributed to the larger tolerance increasing the likelihood of false matches to decoys due to the inclusion of more peaks in each PSM.

4.6 Analysis of the significance for the features learned by CNN-based WinnowNet

To visualize and analyze the features learned by WinnowNet, we applied the class activation mapping (CAM) ([12]) to interpret the learning decision of WinnowNet. CAMs were generated in the spectrum representation to visualize the patterns influencing WinnowNet's prediction of PSMs as true or false. After extracting the CAMs, we tracked back to the original spectra data and reflected the weight from the CAMs to the actual spectra. Subsequently, we visualized the extracted weight from CAMs of the measured and theoretical spectra for a PSM. Fig. 4 presents the CAM visualization for a target PSM and a decoy PSM. One CAM figure originated from the Maine model for the target PSMs, and the other was from the ProteomeTools Model. The legend on the right side indicates that "hotter" colors represent higher weights for a region; i.e., red-colored regions make the PSMs more likely to be true (positive), while blue-colored regions make them more likely to be false (negative).

When scrutinizing the comparison between target PSMs and decoy PSMs (see Fig. 4(a) and Fig. 4(b)), a noticeable trend emerges: the target PSMs exhibit more matching peaks with high intensity, evident in their "hotter" coloration. Intriguingly, WinnowNet not only captures patterns associated with high-intensity ions but also assigns considerable weight to those with low intensity. For instance, while most PSM score functions typically prioritize high-intensity peaks, such as the y3 ion with a 1+ charge, WinnowNet stands out by assigning a high weight to the y3 ion with a 2+ charge, despite its lower intensity (see Fig 4(a)). Moreover, a comparison of CAMs for the same PSM generated by the

Marine model and the Prosit model (see Fig. 4(a) and Fig. 4(c)) reveals a shift towards a more intense red colormap after WinnowNet undergoes fine-tuning with the challenging dataset. This observation underscores the significant impact of curriculum learning in shaping a more effective model, as it refines WinnowNet to better recognize and prioritize matching peaks.

5 Supplementary Discussion

5.1 Assessment of training strategies for curriculum learning

5.1.1 Impact of curriculum learning for CNN-based WinnowNet

To assess both the convergence and generalization ability of the models derived from curriculum learning, we tested different orders of samples in the two training phases based on the difficulty level. Three strategies were involved:

- Easy-difficult Strategy: the model was trained on the Marine 1 dataset dataset first (Phase 1) and then on both Marine 1 dataset and Marine 1 datasets (Phase 2). This strategy, as shown in Section 2.1, follows the order from easy tasks to difficult tasks.
- Reverse Strategy: the model was trained on both Marine 1 dataset and Marine 1 datasets first (Phase 1) and then on the Marine 1 dataset dataset (Phase 2), which is the reverse order of Easy-difficult Strategy.
- Random Strategy: the model was trained on a shuffled dataset obtained by combining the Marine 1 dataset and Marine 1 datasets, which is the random order of Easy-difficult Strategy.

To compare the performance fairly, a hold-out test dataset is randomly selected from the Soil 1 dataset, which consists of 17,757 positive samples and 18,495 negative samples, where the positive and negative samples were defined as described in Section 2.1. Each strategy was applied to train the models five times, and the standard errors for accuracy and precision of the test dataset were calculated. The best model from each strategy's five training rounds was applied to the other eight datasets for comparing PSM/peptide/protein identifications, i.e., after database search by the Sipros-Ensemble, PSMs were filtered with WinnowNet variants of the above three training strategies. In addition to the PSM/peptide/protein identifications, the training time was also measured using the same hyperparameters and devices as described previously and shown in the first row in Tab. 7. The results depicted in Tab. 7 indicate that the Easy-difficult Strategy achieved the best accuracy and precision with a slight standard error (0.19%) in the hold-out test dataset, and it also identified the highest PSMs/peptides/proteins at FDR 1% consistently. The performance of the Reverse Strategy significantly dropped compared to the Easy-difficult Strategy, possibly due to the model converging to fit the easier dataset in the second phase. We also observed the identification improvement is larger than the increase in test dataset precision. For instance, comparing the Easy-difficult Strategy and Random Strategy, the protein identification improved by an average of 5%, whereas the precision increase was 2.2%. This finding reveals that even a slight improvement in precision can significantly boost peptide identification results.

The Random Strategy also produced lower numbers of identified PSMs/peptides/proteins, compared to the Easy-difficult Strategy. Our analysis revealed that the performance of the Easy-difficult Strategy is statistically significantly better than the Random Strategy, with a p-value of 0.0067 in a two-tailed paired t-test for identified proteins. Additionally, we altered the ratio of easy and difficult samples during the first and second training phases. The results were presented in Figs. 8 and 9. Consistent with the findings presented in Tab. 7, the Easy-difficult Strategy in our curriculum learning framework demonstrated the best performance. Moreover, training models in an easy-to-difficult order, which used the least time for training (See Tab. 7), can promote faster convergence. This efficiency is particularly beneficial for the training of extensive models or in transfer learning scenarios, where fine-tuning pre-trained models for specific datasets is required.

5.1.2 Impact of curriculum learning strategy for self-attention-based WinnowNet

To evaluate both the convergence and generalization abilities of models trained via curriculum learning, we investigated four sample ordering strategies across two training phases based on task difficulty. The strategies are as follows:

- **Marine 1 Only Strategy:** The model was trained exclusively on the Marine 1 dataset.
- **Easy-Difficult Strategy:** In Phase 1, the model was trained on the ProteomeTools dataset, and in Phase 2 it was trained on both the ProteomeTools and Marine 1 datasets. As described in Section 2.1, this approach follows a progression from easy to difficult tasks.
- **Reverse Strategy:** Here, the model was first trained on both the ProteomeTools and Marine 1 datasets (Phase 1) and then on the ProteomeTools dataset (Phase 2), effectively reversing the Easy-Difficult order.

- **Random Strategy:** The model was trained on a shuffled combination of the ProteomeTools and Marine 1 datasets.

A hold-out test dataset, randomly selected from the Soil 1 dataset, consisting of 17,757 positive samples and 18,495 negative samples (as defined in Section 2.1), was used for a fair comparison. Each strategy was evaluated across five training runs, and the standard errors for accuracy and precision on the test dataset were computed. The best-performing models were then applied to four datasets—Marine 3, Soil 3, P3, and Human Gut—after being searched by Sipros Ensemble. The target-decoy strategy was applied to control the FDR at 1%. As shown in Tab. 8, the Easy-Difficult strategy achieved the highest accuracy and precision, with a minimal standard error of 0.19%, and consistently identified the most PSMs, peptides, and proteins at 1% FDR. In contrast, the Reverse strategy exhibited a significant drop in performance, likely due to the model converging on the easier dataset during the second phase. Similarly, both the Random strategy and the Marine 1 Only strategy resulted in fewer identified PSMs, peptides, and proteins compared to the Easy-Difficult strategy.

5.2 Comparison of WinnowNet and MS^{2}Rescore Using the Entrapment Method on MS-GF+ Outputs

We compared the score distributions of PSMs to assess gains and losses between WinnowNet and MS^{2}Rescore on the Soil2 dataset. For MS^{2}Rescore, we used the default configuration with mokapot as the machine learning-based rescoring engine. Figure 11 shows histograms of top-ranked target PSMs scored by each tool. We also included PSMs ranked highest by one tool and re-evaluated by the other. At a 1% FDR, the score cutoff for MS^{2}Rescore is 0.3, while for WinnowNet it is 0.85. PSMs missed by WinnowNet tend to have MS^{2}Rescore scores primarily within the range [0, 4], whereas PSMs gained by WinnowNet are mostly within [–1.5, 4]. Similar trends are observed when examining MS^{2}Rescore gains and losses with respect to WinnowNet scores around its threshold.

These discrepancies in tool-specific identifications can be attributed to their fundamentally different modeling approaches: WinnowNet employs a self-attention-based architecture trained solely on spectral similarity, whereas MS^{2}Rescore uses a semi-supervised learning framework incorporating a broad set of engineered features.

To further validate the PSMs uniquely identified by WinnowNet, we assessed their spectral agreement using AlphaPeptDeep predictions on the Marine2 and Soil2 datasets. For each dataset, PSMs were divided into three categories:

1. **WinnowNet-Only:** Confident PSMs identified only by WinnowNet.
2. **MS^{2}Rescore-Only:** Confident PSMs identified only by MS^{2}Rescore.
3. **Shared:** PSMs confidently identified by both tools.

We computed spectral angle and Pearson correlation between the predicted and experimental spectra. The resulting similarity distributions are shown in Figures 15 and 16, with corresponding Q–Q plots in Figure 17. Across both datasets, WinnowNet-only and shared PSMs demonstrate nearly identical levels of spectral agreement, consistently outperforming the MS^{2}Rescore-only group. These results support the conclusion that WinnowNet-only PSMs are not false positives, but rather high-confidence identifications.

Additionally, we manually examined four representative WinnowNet-only PSMs, two from each dataset, by overlaying predicted and experimental spectra (Figure 18). In all cases, there is strong agreement in both peak position and intensity, further reinforcing the quantitative findings.

It is important to note that MS^{2}Rescore incorporates a wide array of features, including search engine-derived scores and metadata, whereas WinnowNet relies solely on spectral features. As a result, a high MS^{2}Rescore score does not necessarily imply strong spectral agreement, which helps explain some of the observed differences in PSM selection.

Table 19 summarizes results across nine benchmark datasets, reporting the estimated number of true positive PSMs (calculated as the number of target PSMs minus decoys) alongside the size of the union of PSMs identified by both tools. In all cases, the union remains below the estimated true positive count, indicating that the methods might identify complementary subsets of PSMs and that neither tool exceeds the limits implied by FDR in terms of unique identifications.

5.3 Analysis of the probability changes for peptides with adjacent amino acids swapped

To assess self-attention-based WinnowNet's sensitivity to sequence alterations, we conducted an experiment where two adjacent amino acids were randomly swapped in identified peptides (controlled at a peptide-level FDR of 1%) from the Marine 3 dataset. The modified peptides were then re-scored using WinnowNet with their corresponding spectra. To ensure meaningful variations, the swaps were randomized but excluded identical adjacent amino acids or

amino acids with modifications. The results, illustrated in the histogram in Fig. 12, show the distribution of WinnowNet score differences between the original and modified peptides. The x-axis represents the magnitude of score differences (i.e., score(modified peptide)-score(original peptide), ranging from 0 to -0.6), while the y-axis indicates the number of peptides. The analysis reveals that the probabilities assigned by WinnowNet generally decrease when adjacent amino acids are swapped, confirming that the originally identified peptides remain the best PSM candidates for their respective spectra. As the score difference increases, the distribution tapers gradually, forming a right-skewed pattern with a long tail. This pattern arises because swapping a single pair of adjacent amino acids affects only one b ion and one y ion, leaving most fragment ions unchanged. These unaffected ions can still be matched to peaks in the corresponding experimental spectrum as before. Consequently, many PSMs exhibit only slight decreases in probabilities after the amino acid swaps.

5.4 Validation of FDR control via synthetic dataset entrapment analysis

To ensure accurate FDR control, we evaluated our entrapment methods using the Synthetic dataset. This dataset consists of MS data from four microbial cultures: *Ruminiclostridium cellulolyticum H10 (ATCC 35319)*, *Desulfovibrio vulgaris Hildenborough (ATCC 29579)*, *Methanospirillum hungatei JF1 (txid323259)*, and *Methanosaeta concilii (txid990316)*. Details of the dataset are provided in Section 3 of this supplementary document. We built an extended "original target" protein database by combining proteins from these four microorganisms with a larger set of proteins from foreign species. The final database contains proteins from the four microorganisms and foreign species at a ratio of 1:212 (1:298 at the peptide level), including 168,467,450 foreign peptides and 3,640,410 foreign proteins. The foreign proteins were randomly selected from the IGC fecal sample database used in [13].

We generated the entrapment database using the tool available at [`https://github.com/Noble-Lab/FDRBench`] [14]. This tool ensured that target proteins have paired entrapment proteins. To assess FDR, we applied both the paired and combined estimation methods using shuffled entrapment sequences. We set the effective ratio of entrapment to target proteins at 1:1. We added the reversed sequences of all target proteins as decoys to the protein database. The paired estimation method is defined in [14], and the combined FDR estimation is defined in Eq. 5 of the main text. We used this double entrapment experiment to compare two analysis pipelines: Sipros Ensemble with WinnowNet and FragPipe.

Since the Synthetic dataset is controlled, any identified foreign proteins are false discoveries, and identifications from the four known microorganisms are true positives. We estimated the FDR by calculating the ratio of foreign and shuffled entrapment identifications to the total number of identifications. The results are shown in Fig. 13. Because the extended database contains very few native proteins, the paired and combined methods produced nearly identical FDR estimates. Both methods also closely matched the FDR values derived directly from the assumption that all foreign and entrapment peptides are false positives. These results confirm the reliability of the FDR control methods used in this study.

5.5 Performance evaluation of WinnowNet-integrated protein identification pipelines on timsTOF data

In addition to the MudPit and Orbitrap datasets, we evaluated the performance of the WinnowNet-integrated protein identification pipelines on a timsTOF dataset derived from a human fecal microbiome sample. Detailed information about this dataset is provided in Section 3 of this Supplementary Document. In this benchmarking experiment, we assessed three widely used pipelines—**FragPipe** [15], **PEAKS** [16], and **AlphaPept** [17]—comparing their default filtering strategies with our WinnowNet-enhanced filtering approach. To rigorously evaluate specificity and robustness, we included entrapment proteins generated by randomly shuffling target protein sequences, enabling the simulation of false targets. All comparisons were performed under a controlled 1% FDR at the PSM, peptide, and protein levels.

Results are summarized in Tab. 13. WinnowNet consistently improved the number of identified PSMs, peptides, and proteins across all pipelines. For example, at the PSM level, FragPipe identifications increased from 48,369 (default filter) to 49,660 with WinnowNet (+3%), and AlphaPept increased from 49,731 to 51,396 (+3%). At the peptide level, PEAKS rose from 16,839 to 17,825 (+6%), and AlphaPept from 16,985 to 17,732 (+4%). At the protein level, identification improvements ranged from 2% to 3% across pipelines. While these gains affirm WinnowNet's effectiveness on timsTOF data, the improvements were slightly less pronounced compared to results obtained with MudPit and Orbitrap systems.

Supplementary References

[1] Samuel Bryson, Zhou Li, Jennifer Pett-Ridge, Robert L Hettich, Xavier Mayali, Chongle Pan, and Ryan S Mueller. Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. Msystems, 1(2):e00027–15, 2016.

[2] Cristina N Butterfield, Zhou Li, Peter F Andeer, Susan Spaulding, Brian C Thomas, Andrea Singh, Robert L Hettich, Kenwyn B Suttle, Alexander J Probst, Susannah G Tringe, et al. Proteogenomic analyses indicate bacterial methylotrophy and archaeal heterotrophy are prevalent below the grass root zone. PeerJ, 4:e2687, 2016.

[3] Manuel Kleiner, Erin Thorson, Christine E Sharp, Xiaoli Dong, Dan Liu, Carmen Li, and Marc Strous. Assessing species biomass contributions in microbial communities via metaproteomics. Nature communications, 8(1):1–14, 2017.

[4] Shuping Long, Yi Yang, Chengpin Shen, Yiwen Wang, Anmei Deng, Qin Qin, and Liang Qiao. Metaproteomics characterizes human gut microbiome function in colorectal cancer. NPJ biofilms and microbiomes, 6(1):14, 2020.

[5] Daniel P Zolg, Mathias Wilhelm, Karsten Schnatbaum, Johannes Zerweck, Tobias Knaute, Bernard Delanghe, Derek J Bailey, Siegfried Gessulat, Hans-Christian Ehrlich, Maximilian Weininger, et al. Building proteometools based on a complete synthetic human proteome. Nature methods, 14(3):259–262, 2017.

[6] Tim Van Den Bossche, Benoit J Kunath, Kay Schallert, Stephanie S Schäpe, Paul E Abraham, Jean Armengaud, Magnus Ø Arntzen, Ariane Bassignani, Dirk Benndorf, Stephan Fuchs, et al. Critical assessment of metaproteome investigation (campi): a multi-laboratory comparison of established workflows. Nature communications, 12(1):7305, 2021.

[7] Dongyu Wang, Kristopher A Hunt, Pieter Candry, Xuanyu Tao, Neil Q Wofford, Jizhong Zhou, Michael J McInerney, David A Stahl, Ralph S Tanner, Aifen Zhou, et al. Cross-feedings, competition, and positive and negative synergies in a four-species synthetic community for anaerobic degradation of cellulose to methane. Mbio, 14(2):e03189–22, 2023.

[8] Michael P Washburn, Dirk Wolters, and John R Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature biotechnology, 19(3):242–247, 2001.

[9] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 652–660, 2017.

[10] Shichao Feng, Ryan Sterzenbach, and Xuan Guo. Deep learning for peptide identification from metaproteomics datasets. Journal of Proteomics, 247:104316, 2021.

[11] Joshua E Elias and Steven P Gygi. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature methods, 4(3):207–214, 2007.

[12] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2921–2929, 2016.

[13] Junhua Li, Huijue Jia, Xianghang Cai, Huanzi Zhong, Qiang Feng, Shinichi Sunagawa, Manimozhiyan Arumugam, Jens Roat Kultima, Edi Prifti, Trine Nielsen, et al. An integrated catalog of reference genes in the human gut microbiome. Nature biotechnology, 32(8):834–841, 2014.

[14] Bo Wen, Jack Freestone, Michael Riffle, Michael J MacCoss, William S Noble, and Uri Keich. Assessment of false discovery rate control in tandem mass spectrometry analysis using entrapment. bioRxiv, 2024.

[15] Andy T Kong, Felipe V Leprevost, Dmitry M Avtonomov, Dattatreya Mellacheruvu, and Alexey I Nesvizhskii. Msfragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nature methods, 14(5):513–520, 2017.

[16] Bioinformatics Solutions Inc. PEAKS Studio 12.5. `https://www.bioinfor.com/peaks-12-5/`, 2024. Accessed: 2025-04-04.

[17] Maximilian T Strauss, Isabell Bludau, Wen-Feng Zeng, Eugenia Voytik, Constantin Ammar, Julia P Schessner, Rajesh Ilango, Michelle Gill, Florian Meier, Sander Willems, et al. Alphapept: a modern and open framework for ms-based proteomics. Nature Communications, 15(1):2168, 2024.