

# **Supplementary Material:**

## **Probabilistic Consensus Scoring Improves Tandem Mass Spectrometry Peptide Identification**

Sven Nahnsen,<sup>\*,†</sup> Andreas Bertsch,<sup>†</sup> Jörg Rahnenführer,<sup>‡</sup> Alfred Nordheim,<sup>¶</sup> and  
Oliver Kohlbacher<sup>†</sup>

*Center for Bioinformatics, Eberhard Karls University, 72076 Tübingen, Germany, Department of  
Statistics, Dortmund University of Technology, 44221 Dortmund, Germany, and Department of  
Molecular Biology, Eberhard Karls University, 72076 Tübingen, Germany*

E-mail: [sven.nahnsen@uni-tuebingen.de](mailto:sven.nahnsen@uni-tuebingen.de)

---

<sup>\*</sup>To whom correspondence should be addressed

<sup>†</sup>Center for Bioinformatics, Tübingen

<sup>‡</sup>Department of Statistics, Dortmund

<sup>¶</sup>Department of Molecular Biology, Tübingen

## Calculation of consensus scores

The excel document *similarity-matrices.xls* contains all three similarity matrices that are important for the assignment of consensus scores to the spectrum in Fig. 1. These matrices have been calculated via pairwise sequence similarity (SeqSim) as described in the methods section. The calculation of the EQSATEQDILKK consensus score is now exemplified:

EQSATEQDILKK was assigned an E-value of 16 by Mascot. Via mixture modeling we assigned a posterior error probability of 97%. The sequence was neither suggested by OMSSA nor by X!Tandem. The first similarity ( $\alpha$ ) is the largest entry in the EQSATEQDILKK row from the Mascot-OMSSA matrix.  $\alpha = 0.36$  reflects the similarity between EQSATEQDILKK and EDNMAIQSIKK (which is the most similar candidate from the OMSSA search results). The second similarity ( $\beta$ ) is the largest entry in the EQSATEQDILKK row from the Mascot-X!Tandem matrix.  $\beta = 0.3$  reflects the similarity between QRESTATDILQK and EQSATEQDILKK. The consensus score is then further calculated:

$$\frac{\text{Mascot}(EQSATEQDILKK) + \alpha \cdot \text{OMSSA}(EDNMAIQSIKK) + \beta \cdot \text{X!Tandem}(QRESTATDILQK)}{(1 + \alpha + \beta)^2}$$
$$= \frac{0.97 + 0.36 \cdot 0.98 + 0.3 \cdot 0.54}{(1 + 0.36 + 0.3)^2} = 0.53$$

## **Suppl. Figure captions**

### **Suppl. Figure 1:**

Correct peptide sequences on different ranks. The dark colors for X!Tandem and Mascot and the grey color for OMSSA correspond to the number of additional peptides.

### **Suppl. Figure 2:**

Presearches with broad tolerance windows to evaluate data quality.

### **Suppl. Figure 3:**

Different sets of spectra are annotated by different search engines. The overlap of spectra that were annotated by one, two or all three engines varies between the different search engines. Here annotated means that the search engine suggests any sequence for a given spectrum.

### **Suppl. Figure 4:**

Comparison of spc and matrix-based similarities between the top hit and sequences at lower ranks.

### **Suppl. Figure 5:**

ConsensusID in comparison to a machine learning method (PepArML) and a heuristic combination method (PepArML:heuristic cominber).

Suppl. Table 1: True peptides on rank one, following individual database searches.

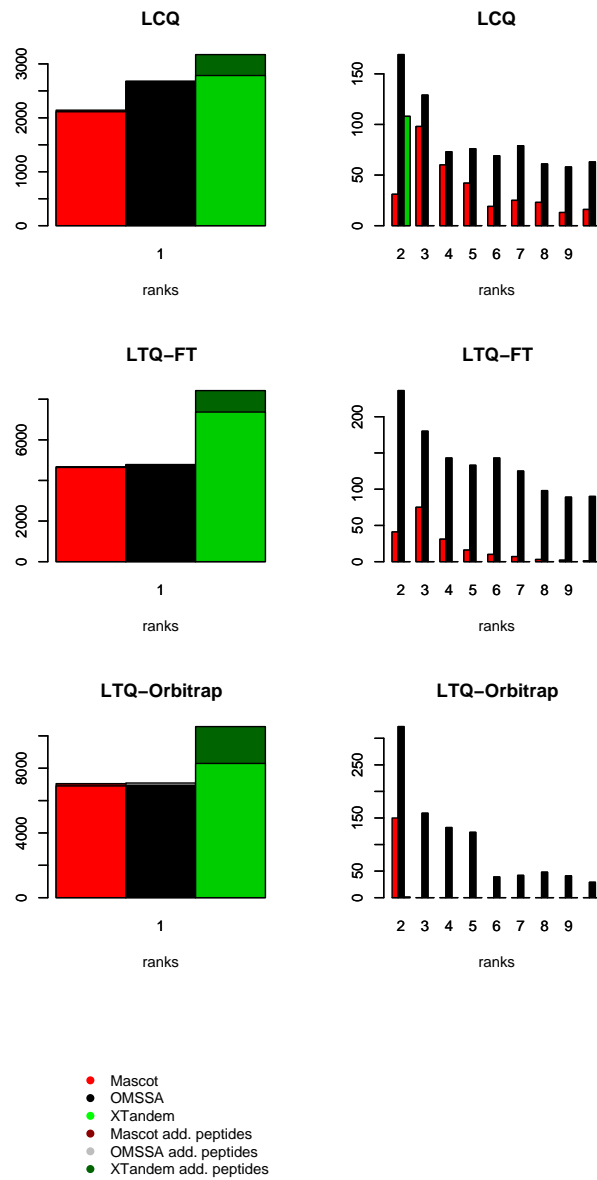
	Orbitrap	FT Ultra	LCQ
Mascot	7046	4675	2142
OMSSA	7086	4783	2679
XTandem	10571	8424	3172

Suppl. Table 2: Number of identified peptides following individual database searches at 1% FDR.

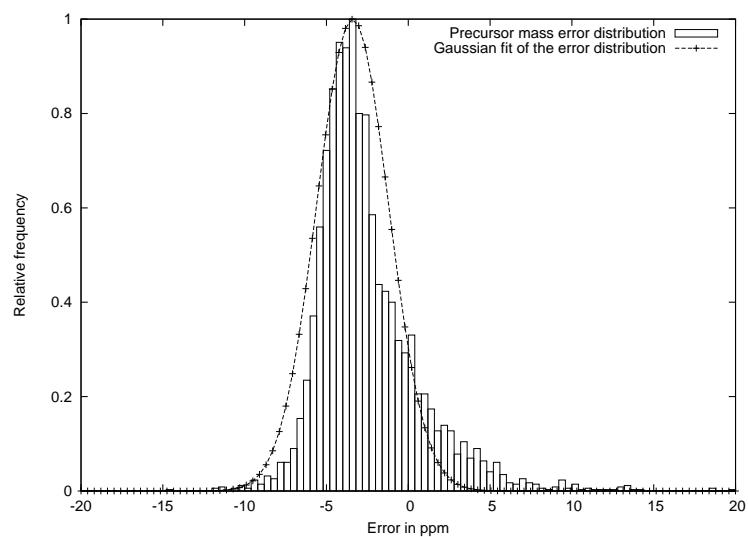
	Orbitrap	FT Ultra	LCQ
Mascot	3919	3466	1314
OMSSA	4719	3729	1786
XTandem	7255	6220	1863

Suppl. Figure 4 was generated using the dataset from the 18 protein mix measured on an LTQ-Orbitrap. It was searched with OMSSA using parameters as described in the experimental procedures. It shows the percentage of sequence similarity of peptide candidates that are suggested by OMSSA at different ranks. The sequence similarity correlates very well with the percentage of overlapping fragment ion masses. At latter ranks the sequence similarity and the percentage of overlapping fragment ions decreases.

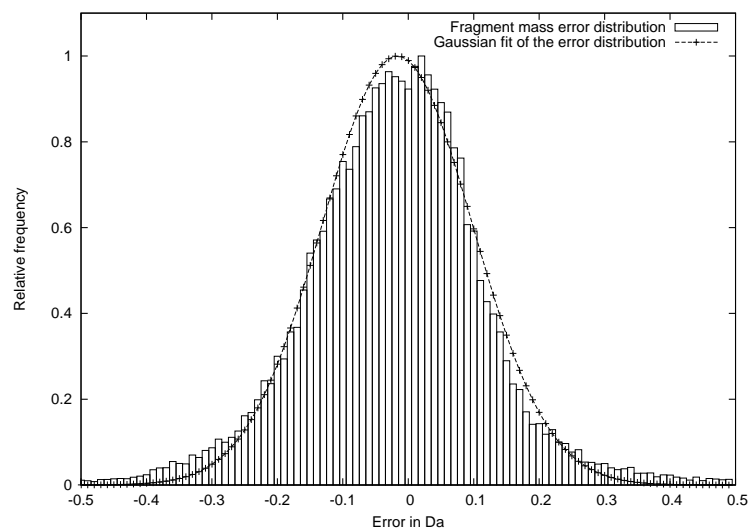
Suppl. Figure 1



Suppl. Figure 2

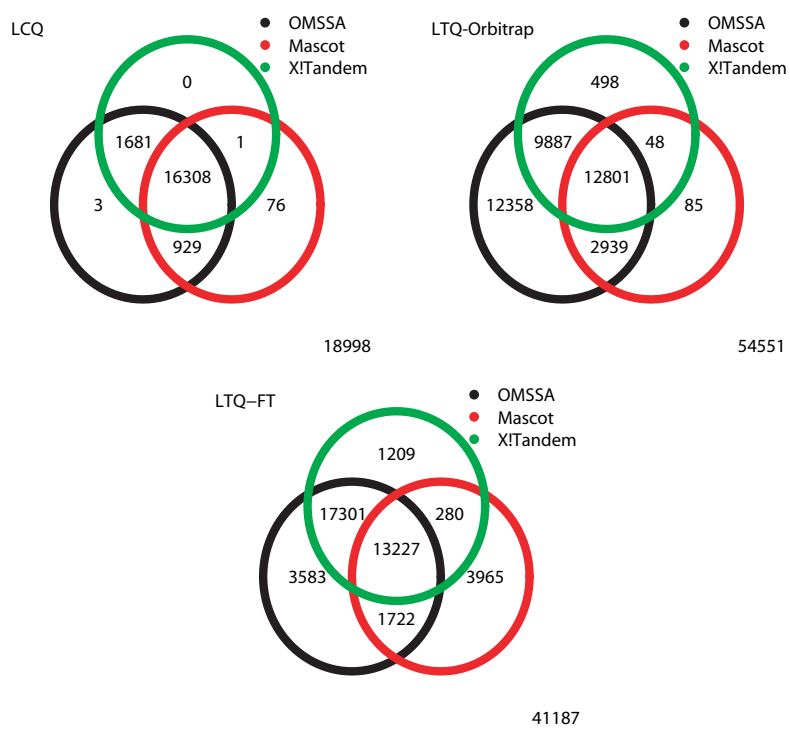


(a)



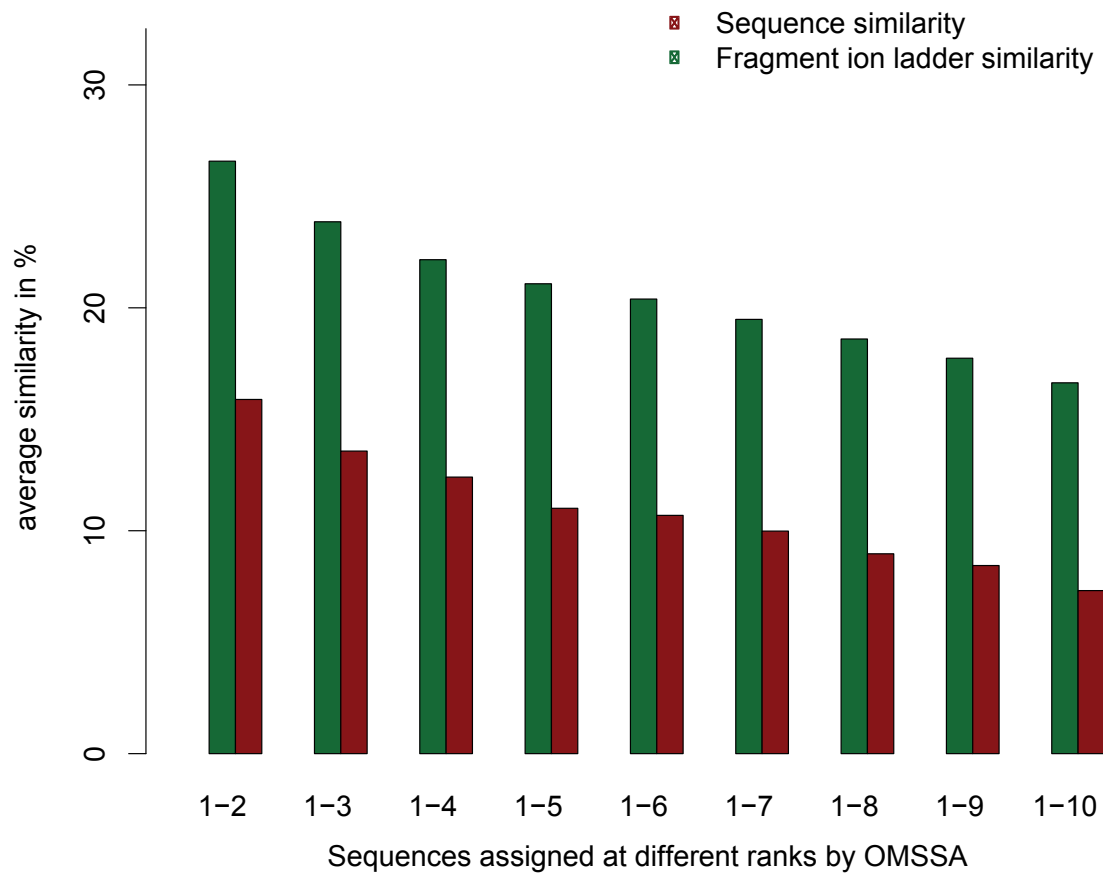
(b)

Suppl. Figure 3





Suppl. Figure 4



Suppl. Figure 5

