# Supplementary Material for "TaxIt: An iterative and automated computational pipeline for untargeted strain-level identification using MS/MS spectra from pathogenic samples"

Mathias Kuhring<sup>1,2,3,4</sup>, Joerg Doellinger<sup>5,6</sup>, Andreas Nitsche<sup>6</sup>, Thilo Muth<sup>1</sup>, Bernhard Y. Renard<sup>1,\*</sup>

<sup>1</sup> Bioinformatics Unit (MF 1), Department for Methods Development and Research Infrastructure, Robert Koch Institute, Berlin, Germany

<sup>2</sup> Core Unit Bioinformatics, Berlin Institute of Health (BIH), Berlin, Germany

<sup>3</sup> Berlin Institute of Health Metabolomics Platform, Berlin Institute of Health (BIH), Berlin, Germany

<sup>4</sup> Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany

<sup>5</sup> Centre for Biological Threats and Special Pathogens, Proteomics and Spectroscopy (ZBS 6), Robert Koch Institute, Berlin, Germany

<sup>6</sup> Centre for Biological Threats and Special Pathogens, Highly Pathogenic Viruses (ZBS1), Robert Koch Institute, Berlin, Germany

\*Corresponding author

## S1 Search Parameters

#### **Cowpox Sample**

The cowpox sample of strain *Cowpox virus (Brighton Red)* was acquired in-house, as described in *S2 Cowpox Sample Acquisition*. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE <sup>1</sup> partner repository with the dataset identifier PXD014913. Spectra were analyzed applying a tryptic search with parent ion mass tolerance of 10 ppm, fixed modification cysteine carbamidomethylation (+57 Da) as well as an additional variable modification methionine oxidation (+16 Da).

### **Bronchitis Sample**

Bronchitis samples of the strain *Avian infectious bronchitis virus (strain Beaudette CK)* were downloaded from PRIDE (PXD002936) and the sample "BeauR2.raw" was randomly selected for analysis. The raw file was converted to an mgf file using ProteoWizard's MSConvert GUI (3.0.8764)<sup>2</sup>. Spectra were analyzed with default settings including a tryptic search with fixed modification cysteine carbamidomethylation (+57 Da) and parent ion mass tolerance of 100 ppm.

### **Bacillus Sample**

The bacillus sample of the strain *Bacillus subtilis* subsp. *subtilis str. 168* was download from PRIDE (PXD007242, file "614\_NG4\_BSN238\_Urea-Trp\_1ug\_SR-LFQ\_4h\_161201.mgf"). Spectra were analyzed with default settings including a tryptic search with fixed modification cysteine carbamidomethylation (+57 Da). However, parent ion mass tolerance was set to 10 ppm in accordance with the original publication.

<sup>&</sup>lt;sup>1</sup> Vizcaíno, J. A.; Csordas, A.; del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 Update of the PRIDE Database and Its Related Tools. Nucleic Acids Res 2016, 44 (D1), D447–D456. https://doi.org/10.1093/nar/gkv1145.

<sup>&</sup>lt;sup>2</sup> Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; et al. A Cross-Platform Toolkit for Mass Spectrometry and Proteomics. Nat Biotech 2012, 30 (10), 918–920. https://doi.org/10.1038/nbt.2377.

# S2 Cowpox Sample Acquisition

## **CPXV-infection**

One day prior infection, 5x105 HEp-2 cells were seeded into 6 well plates with 4 mL cell culture medium (DMEM supplemented with 10 % FCS and 2 mM L–Glutamine) each and kept in an incubator at 37°C for 16 h. The medium was removed and cells were infected with CPXV Brighton Red (BR) (ATCC<sup>®</sup> VR-302<sup>™</sup>) using a multiplicity of infection (MOI) of 0.1 in 1 mL cell culture medium per well for 1 h. Again, the medium was removed and cells were washed once with 5 mL phosphate-buffered saline (PBS) before they were further incubated in 4 mL cell culture medium for 24h. Supernatant was removed and mixed 1:1 with 4 % SDS, 0.1 M Tris-HCl (pH 7.6), 10 mM Tris(2-carboxyethyl)phosphine (TCEP) and 40 mM 2-chloroacetamide (CAA) at 99°C for 5 min. DNA was sheared by sonication for 3 x 1 min on ice using a 2″ Sonifier<sup>™</sup> Cup Horn in a Cell Disruptor (Branson Ultrasonics Corporation, Danbury, CT, USA) and lysates were clarified by centrifugation at 16,000 × g for 10 min.

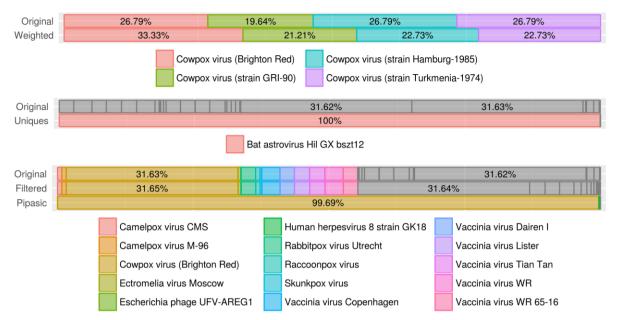
### Filter-aided sample preparation (FASP)

Samples were prepared using the filter-aided sample preparation (FASP) protocol with minor modifications. Briefly, 20 µg of each sample were processed using Vivacon 500 Centrifugal Ultra Filter (Sartorius, Goettingen, Germany) with a molecular weight cut-off (MWCO) of 30 kDa. SDS was depleted 4 x with 200 µL 8 M Urea in 50 mM Tris-HCl, pH 8.5, before the urea concentration was reduced using 3 x 100 µL 50 mM Tris-HCl, pH 8.5. The proteins were digested for 16 h at 37°C in 50 mM Tris-HCl, pH 8.5 using Trypsin Gold, Mass Spectrometry Grade (Promega, Fitchburg, WI, USA) at a protein/enzyme ratio of 100:1. The peptides were collected by centrifugation and two washing steps with 40 µL 50 mM Tris-HCl, pH 8.5. The peptides were desalted using 200 µL StageTips packed with four Empore<sup>™</sup> SPE Disks C18 (3M Purification, Inc., Lexington, USA) and concentrated using a vacuum concentrator but not dried completely. Samples were filled up to 16 µL with 0.1% formic acid and peptides were quantified by measuring the absorbance at 280 nm using a Nanodrop 1000 (Thermo Fisher Scientific, Rockford, IL, USA).

## nLC-MS/MS

Single-run bottom-up proteome analysis was performed on an Easy-nanoLC (Proxeon, Odense, Denmark) coupled online to an LTQ Orbitrap Discovery<sup>™</sup> mass spectrometer (Thermo Fisher Scientific, Rockford, IL, USA). 2 µg peptides were loaded directly on a Reprosil-Pur 120 C18-AQ, 2.4 µm, 300 mm x 75 µm fused silica capillary column (Dr. Maisch, Ammerbuch-Entringen, Germany), which was kept at 60°C using a butterfly heater (Phoenix S&T, Chester, PA, USA). Peptides were separated using a linear 240 min gradient of acetonitrile in 0.1 % formic acid and 3 % DMSO from 0 to 29 % at 200nL/min flow rate. The mass spectrometer was operated in a data-dependent manner in the m/z range of 400 – 1,400 with a resolution of 30,000 in the orbitrap. Up to the seven most intense 2+ and 3+ charged ions were selected for low-energy CID type fragmentation in the ion trap with a normalized collision energy of 35 % using an activation time of 10 ms. The m/z isolation width for MS/MS fragmentation was set to 2 Th. Once fragmented, up to 500 isolated peaks were dynamically excluded from precursor selection for 90 s within a 20 ppm window. The ion selection threshold for  $MS^2$  spectra was 1,000 counts, and the maximum allowed ion accumulation times were 500 ms for full scans and 100 ms for  $MS^2$  spectra. Automatic gain control was set to a target value of 1e6 for full scans and 5e3 for  $MS^2$ . Peptides were ionized using electrospray with a stainless steel emitter, I.D. 30  $\mu$ m, (Proxeon, Odense, Denmark) at a spray voltage of 2.1 kV and a heated capillary temperature of 275°C. The background ion signal intensities were reduced using an ABIRD device (ESI Source Solutions, Woburn, MA, USA).

## S3 Additional Figures



**Figure S1: Relative counts of cowpox.** Relative counts are illustrated for Taxlt (top), uniques-(middle) and Pipasic-based search strategies (bottom). Original, filtered (if applicable) and corrected relative counts are summarized by means of one vertical stacked bar each. Taxa are labeled and color-coded based on a limit of 15 final top candidates (i.e. after correction) with a relative count greater zero. Furthermore, ratios greater than 0.05 are highlighted as percentages within bars. Selecting unique PSMs resulted in only one PSM for *Bat astrovirus Hil GX bszt12* (taxid 1748291).

Original	34.43%		34.43	1%	31.15%				
Weighted	34.25%		34.25	%	31.49%				
	Avian infectious bronchitis virus (strain Beaudette CK)								
		Avia	n infectious bronchitis vi	rus (strain Beaudette	e US)				
	Avian infectious bronchitis virus (strain Beaudette)								
Original Uniques									
		Aviar	i infectious bronchitis vir	us (strain 6 82)					
		Aviar	n infectious bronchitis vir	us (strain Arkansas	99)				
		Aviar	infectious bronchitis vir	us (strain Beaudette	CK)				
		Aviar	n infectious bronchitis vir	us (strain Beaudette	US)				
		Aviar	i infectious bronchitis vir	us (strain Beaudette	)				
		Aviar	Avian infectious bronchitis virus (strain D1466)						
		Avian infectious bronchitis virus (strain D274)							
		Avian infectious bronchitis virus (strain DE072)							
		Avian infectious bronchitis virus (strain GRAY)							
		Aviar							
		Aviar							
		Avian infectious bronchitis virus (strain KB8523)							
		Aviar	Avian infectious bronchitis virus (strain M41)						
		Aviar	Avian infectious bronchitis virus (strain SAIB20)						
		Aviar	Avian infectious bronchitis virus (strain Vic S)						
Original	5.22% 5.26% 5.26% 5.26% 5.22% 5	.24% 5.24	% 5.25% 5.22% 5.22% 5	.22% 5.25% 5.24% 5	5.22% 5.24% 5.24% 5.24% 5.24% 5.24%				
	5.22% <mark>5.26%</mark> 5.26% <mark>5.26% 5.22%</mark> 5	.24% 5.24	<mark>%</mark> 5.25% 5.22% <mark>5.22%</mark> 5	.25% 5.24% 5.24% 5	5.24% 5.24% 5.24% 5.24% 5.22% 5.22%				
Pipasic	27.98%	14	% 19.61	%	<b>19% 5.44%</b> 10.26%				
	Avian infectious bronchitis virus (strain 6 82)								
	Avian infectious bronchitis virus (strain Beaudette CK)								
	Avian infectious bronchitis virus (strain Beaudette US)								
	Avian infectious bronchitis virus (strain Beaudette)								
	Avian infectious bronchitis virus (strain D274)								
		Avian infectious bronchitis virus (strain DE072)							
		Avian infectious bronchitis virus (strain H52)							
	Avian infectious bronchitis virus (strain M41)								
	Avian infectious bronchitis virus (strain UK 142 86)								
	Avian infectious bronchitis virus (strain UK 167 84)								

**Figure S2: Relative counts of bronchitis.** Relative counts are illustrated for Taxlt (top), uniques-(middle) and Pipasic-based search strategies (bottom). Original, filtered (if applicable) and corrected relative counts are summarized by means of one vertical stacked bar each. Taxa are labeled and color-coded based on a limit of 15 final top candidates (i.e. after correction, except for uniques) with a relative count greater zero. Furthermore, ratios greater than 0.05 are highlighted as percentages within bars.

Original Weighted			% 5.55% 5.57% 5.55% 5.55% 5.41%   % 5.81% 5.81% 5.81% 5.65%				
Weighted	0.0070	7.0070 0.0					
			Bacillus subtilis subsp. inaquosorum KCTC 13429				
			Bacillus subtilis subsp. natto BEST195				
			Bacillus subtilis subsp. spizizenii ATCC 6633				
			Bacillus subtilis subsp. spizizenii str. W23				
			Bacillus subtilis subsp. spizizenii TU-B-10				
			Bacillus subtilis subsp. subtilis 6051-HGW				
			Bacillus subtilis subsp. subtilis str. 168				
			Bacillus subtilis subsp. subtilis str. AG1839				
			Bacillus subtilis subsp. subtilis str. BAB-1				
			Bacillus subtilis subsp. subtilis str. BSP1				
			Bacillus subtilis subsp. subtilis str. JH642 substr. AG174				
			Bacillus subtilis subsp. subtilis str. OH 131.1				
	Bacillus subtilis subsp. subtilis str. RO-NN-1						
	Bacillus subtilis subsp. subtilis str. SC-8 Bacillus subtilis subsp. subtilis str. SMY						
Original							
Uniques	9.09% 9.09%	9.09%	<b>9.09% 9.09% 9.09% 9.09% 9.09% 9.09% 9.09%</b>				
			lobacteria bacterium RIFCSPLOWO2_12_FULL_59_11				
	Anaerobacillus alkalilacustris						
	Bacillus lentus						
	Bacillus niacini						
	Bacillus sp. Marseille-P2384						
	Bacillus subtilis subsp. subtilis str. 168						
		Bac	teroides luti				
	Candidatus Glassbacteria bacterium RIFCSPLOWO2_12_FULL_58_11						
	cyanobacterium TDX16						
	Sporolactobacillus laevolacticus						
		Stre	ptomyces griseus				
Original							
Filtered							
Pipasic	11.42% 1	1.89% 8	34% 27.51% 5.79% 8.94% 13.06% 9.82%				
	Bacillus cereus	BAG2O-3	Bacillus cereus E33L Bacillus cereus SJ1 Bacillus cereus VD133				
	Bacillus cereus	BAG2X1-2	Bacillus cereus FRI-35 Bacillus cereus str. Schrouff Bacillus cereus VD140				
	Bacillus cereus BAG3X2-2 Bacillus cereus MSX-A1 Bacillus cereus VD045 Bacillus cereus W						
	Bacillus cereus	BAG4X12-1	Bacillus cereus Rock4-2 Bacillus cereus VD102				

**Figure S3: Relative counts of bacillus 1k.** Relative counts are illustrated for TaxIt (top), uniques-(middle) and Pipasic-based search strategies (bottom). Original, filtered (if applicable) and corrected relative counts are summarized by means of one vertical stacked bar each. Taxa are labeled and color-coded based on a limit of 15 final top candidates (i.e. after correction) with a relative count greater zero. Furthermore, ratios greater than 0.05 are highlighted as percentages within bars.

Original Weighted			% 5.38% 5.37% 5.36% 5.36% 5 9% 5.7% 5.67% 5.68% 5.73°	
Weighted	5.71% 6.84%	Bacillus s Bacillus s	w 5.7% 5.67% 5.68% 5.73   subtilis subsp. inaquosorum KC subtilis subsp. natto BEST195   subtilis subsp. spizizenii ATCC subtilis subsp. spizizenii ATCC   subtilis subsp. spizizenii TU-B- subtilis subsp. spizizenii TU-B-   subtilis subsp. subtilis 6051-HC subtilis subsp. subtilis 6051-HC   subtilis subsp. subtilis str. 168 subtilis subsp. subtilis str. AG11   subtilis subsp. subtilis str. BSP subtilis subsp. subtilis str. BSP   subtilis subsp. subtilis str. JH64 subtilis subsp. subtilis str. OH 1   subtilis subsp. subtilis str. RO-N subtilis subsp. subtilis str. SC-8   subtilis subsp. subtilis str. SC-8 subtilis subsp. subtilis str. SC-8	CTC 13429 6633 23 10 3W 839 -1 1 12 substr. AG174 131.1 NN-1 8
Original Uniques		Dacinus		
	B B B B B B B B C C C C C C C C C C C C	acillus subtilis subs acillus subtilis subs acillus wakoensis andidatus Buchana racilibacillus lacisa aenisporosarcina c araliobacillus ryuky alsuginibacillus koo	KCTC 13613 e-P2384 sp. subtilis str. BSP1 sp. subtilis str. RO-NN-1 sp. subtilis str. SC-8 JCM 9140 anbacteria bacterium RIFCSPL alsi quisquiliarum yuensis	_OWO2_01_FULL_46_12
Original Filtered Pipasic			100%	
	Bacillus amyloliquefac Bacillus amyloliquefac Bacillus amyloliquefac Bacillus cereus ATCC	ciens KHG19 ciens TA208	Bacillus cereus B4264 Bacillus cereus BAG1X1-1 Bacillus cereus Rock3-42 Bacillus subtilis E1	Bacillus subtilis subsp. subtilis str. BAB-1 Bacillus subtilis subsp. subtilis str. SC-8 Bacillus velezensis YAU B9601-Y2

**Figure S4: Relative counts of bacillus all.** Relative counts are illustrated for TaxIt (top), uniques-(middle) and Pipasic-based search strategies (bottom). Original, filtered (if applicable) and corrected relative counts are summarized by means of one vertical stacked bar each. Taxa are labeled and color-coded based on a limit of 15 final top candidates (i.e. after correction) with a relative count greater zero. Furthermore, ratios greater than 0.05 are highlighted as percentages within bars.