

Text S1

MS-H typing of *Salmonella* flagella. Reference strains were cultured overnight without motility induction. Colonies from one plate were harvested on a 10 μ l inoculation loop and transferred to a 1.5 ml centrifuge tubes (Eppendorf) containing 1 ml de-ionized water (dH₂O) and 2 mg lysozyme (Sigma). The suspension was gently mixed by a pipette tip and left at room temperature for 10 minutes. After this, the mixture was vigorously vortexed for 20 seconds three times, with a break of 2 minutes between vortexing. The mixture was then centrifuged for 20 minutes at 16,000 x g and the supernatant collected using a 1 ml syringe. It was then passed through a 13 mm diameter filter with 0.22 μ m pore size and low protein binding capability (PALL). The filter was washed with 3 ml of dH₂O and filled with a solution containing 100 ng/ μ l sequence grade trypsin (Promega) after the residual water was expelled using a new syringe. The on-filter digestion was performed at 37°C for 2 hours and the digest collected by flushing the filter with 0.6 ml dH₂O. The tryptic digest from the filter was expelled completely using the syringe, and 25 μ l of HPLC 2X buffer A (0.2% formic acid, 2% Acetonitrile) was added to 25 μ l digest. Ten μ l of the sample, containing 1/120 of the total tryptic digest, was loaded onto a 0.1 x 3 mm self-packed C18 pre-column for 8 minutes and the pre-column automatically switched to a self-packed C18 nano-LC column. Nano-LC (Proxeon) separation was run at 250 nl/min with a 35 minute acetonitrile gradient from 5-30% followed by a 20 minute flush with 95% acetonitrile. Columns were equilibrated for 10 minutes with buffer A before loading the sample. Mass spectrometry data were collected from LTQ-Orbitrap XL system (ThermoFisher) with a data-dependent acquisition method for peptide ion scanning and fragmentation. For multiple samples runs together in high throughput mode, two blank runs were performed after each sample, the first being called “jigsaw” to wash the nano-column vigorously with several cycles of high concentration acetonitrile, and the second using the same gradient as a true sample run to re-flush and equilibrate the column for the next sample. The mass spectrometry data were searched with Mascot 2.3 (Matrix Science) against a custom *Salmonella* flagella database containing 385 different sequence entries. Search parameters included two missed cleavages of trypsin digestion with no fixed modifications of proteins and possible modifications of methionine oxidation and glutamine and asparagine deamidation.

MASCOT Search Results

User : keding
 E-mail : chengkeding@yahoo.com
 Search title : flagellin
 MS data file : C:\Xcalibur\data\20131210-002-0088-01806-OR\Raw\25-20131206-S-1714.raw
 Database : Flagellin_Salm 20131007 (384 sequences) 188,399 residues
 Taxonomy : Bacteria (Eubacteria) (304 sequences)
 Timestamp : 12 Dec 2013 at 22:13:21 GMT

Not what you expected? Try [the select summary](#).

- ▶ Search parameters
- ▶ Score distribution
- ▶ Legend

Protein Family Summary

Significance threshold p< Max. number of families
 Ion score or expect cut-off Dendrograms cut at
 Preferred taxonomy

Protein family 1 (out of 1)

per page



Threshold (0):

	Score	Mass	Matches	Sequences	emPAI
1.1 gi 38049959	6070	52950	109 (103)	44 (43)	32.70
phase 1 flagellin [Salmonella enterica Enteritidis flagellin antigen g,m]					
▶ 4 samecats of gi 38049959					
1.2 gi 38049971	5546	52974	105 (99)	42 (41)	41.10
phase 1 flagellin [Salmonella enterica Montevideo flagellin antigen g,m,p,s]					
1.3 gi 38054795	4357	52871	88 (81)	33 (31)	16.16
flagellin [Salmonella enterica subsp. houtenae 45:g,z51:- flagellin antigen H:'g,z51']					
▶ 2 samecats of gi 38054795					
1.4 gi 38050013	4001	52773	82 (75)	33 (30)	14.29
phase 1 flagellin [Salmonella enterica Travis flagellin antigen g,z51]					
1.5 gi 38050067	1272	51510	29 (23)	12 (9)	1.10
phase 1 flagellin [Salmonella enterica IIIb 50:k:z53 flagellin antigen k]					
1.6 gi 46359071	1161	51431	20 (21)	13 (9)	0.98
phase 1 flagellin [Salmonella enterica IIIb 61:k:1,5 flagellin antigen k]					
1.7 gi 38054717	877	45025	17 (13)	9 (6)	0.76
flagellin [Salmonella enterica subsp. enterica Duesseldorf flagellin antigen H:'s4,z24']					
▶ 1 samecat of gi 38054717					

151 peptide matches (92 non-duplicate, 59 duplicate)

Query Dupon	Observed	Mr (exp)	Mr (calc)	ppm	M	Score	Expect	Rank	Peptide
74 ▶ 1	344.7045	721.2944	721.2924	2.52	0	41	0.00014	▶	K.GLTQARR-N
95 ▶ 1	377.6940	753.3734	753.3731	0.45	0	24	0.011	▶	K.TMFDK-T
100	380.7214	759.4283	759.4137	20.6	0	6	0.51	▶	K.ADQRLVK-Y
102	392.1452	762.3159	762.3144	1.95	0	25	0.0034	▶	K.TGDDGRR-V
111 ▶ 1	395.6914	769.3682	769.3680	0.22	0	42	0.00019	▶	K.TMFDK-T + Oxidation (M)
183 ▶ 1	416.7244	831.4343	831.4338	0.62	0	48	3.9e-05	▶	K.BATVGDLE-S
242	466.2750	930.5370	930.5306	-1.73	1	22	0.0063	▶	K.LQLLQDR-G
354 ▶ 1	473.2475	944.4804	944.5035	-24.9	0	52	1.8e-06	▶	K.DQLGQIKR-F
355	475.7502	949.4859	949.4869	-1.04	0	2	0.62	▶	K.IYVGAARR-L + Deamidated (N)

Fig. S1 Snapshot of MS-H identification of monophasic *Salmonella* Enteritidis [9,12:g,m:-] flagella. Original MS data of flagella trypsin digest was searched against all 385 known *Salmonella* flagellin sequences. The dendrogram was produced using the sequence data of all hits (only the top seven hits are shown), its units representing “peptide ions score” which is correlated with the confidence of the identified peptides. The vertical red line on the dendrogram shows the peptide ion score cut-off value for grouping identified proteins, which was set to “0” here. Individual rows display the results of each protein search summary by way of accession number, protein score, protein molecular weight (mass), number of matches (redundant) and number of distinct sequences (non-redundant). The numbers outside parentheses in both the “Matches” and “Sequences” columns represent the total count of identified peptides, while the numbers in parentheses denote the number of matches above the significance threshold. “emPAI” denotes the relative quantitation [i.e. ten to the power of (the number of observed peptides divided by the number of observable peptides), minus 1]. The top hit (1.1), showing the highest score and emPAI value, is regarded as the correct flagella type (g,m).

MASCOT Search Results

User : Keding
 E-mail : chengkeding@gmail.com
 Search title : 20120628-gehua
 MS data file : C:\mass_data\Raw data\20131022-001-0088-01753-OR\04-20131018-S-0444.RAW
 Database : Flagellin_Salm 20131007 (384 sequences; 188,399 residues)
 Taxonomy : Bacteria (Eubacteria) (384 sequences)
 Timestamp : 25 Oct 2013 at 13:59:45 GMT

Not what you expected? Try [the select summary](#).

- ▶ Search parameters
- ▶ Score distribution
- ▶ Legend

Protein Family Summary

Significance threshold p < 0.05 Max. number of families AUTO
 Ions score or expect cut-off 0 Dendrograms cut at 0
 Preferred taxonomy All entries

Protein family 1 (out of 1)

10 per page 1



Rank	Accession	Score	Mass	Matches	Sequences	emPAI
1.1	gi 38049945	4284	52223	83 (75)	32 (28)	13.80
1.2	gi 308054827	3475	52504	72 (65)	29 (25)	10.41
1.3	gi 308054833	3417	52468	71 (64)	27 (23)	9.73
1.4	gi 38049712	3387	52646	70 (63)	27 (23)	9.05
1.5	gi 38049808	3381	52135	71 (63)	26 (23)	9.25
1.6	gi 38049812	2708	52135	57 (52)	20 (18)	5.28

Fig. S2 Snapshot of MS-H identification of diphasic *Salmonella* Newport [6,8:e,h:1,2] flagella. Original MS data of flagella trypsin digest was searched against all 385 known *Salmonella* flagellin sequences. The dendrogram was produced using the sequence data of all hits (only the top 6 hits are shown), its units representing “peptide ions score” which is correlated with the confidence of the identified peptides. The vertical red line on the dendrogram shows the peptide ion score cut-off value for grouping identified proteins, which was set to “0” here. Two distinct groupings, signifying phase 1 and phase 2 antigens, can easily be seen. Individual rows display the results of each protein search summary by way of accession number, protein score, protein molecular weight (mass), number of matches (redundant) and number of distinct sequences (non-redundant). The numbers outside parentheses in both the “Matches” and “Sequences” columns represent the total count of identified peptides, while the numbers in parentheses denote the number of matches above the significance threshold. “emPAI” denotes the relative quantitation [i.e. ten to the power of (the number of observed peptides divided by the number of observable peptides), minus 1]. The top hits (1.1 and 1.2), showing the highest scores and emPAI values, are regarded as the correct flagella types (e,h and 1,2 representing phase 1 and phase 2 antigens, respectively).

TABLE S1 Results of MS-H analyses using different loading amounts of flagellin digest produced from *Salmonella enterica* serovar Newport (6,8:e,h:1,2)

Loading amount^a	Phase 1 H antigen top hit	Phase 2 H antigen top hit
1/5000	e,h	1,7
1/500	e,h	1,5
1/50	e,h	1,2 ^b

^aLoading amount represents the fraction of the 600 µl flagellin digest used for MS-H.

^b1,2 denotes the correct H type for the strain under study.

TABLE S2 Results of MS-H analyses using different loading amounts of flagellin digest produced from *Salmonella enterica* Infantis (6,7:r:1,5)

Loading amount^a	Phase 1 H antigen top hit	Phase 2 H antigen top hit
1/5000	b	1,5
1/500	r,i	1,5
1/50	r ^b	1,5

^aLoading amount represents the fraction of the 600 µl flagellin digest used for MS-H.

^b“r” denotes the correct H type for the strain under study.