

Supplemental Tables for Comparison to Sequest Cross Correlation Section

The following applies to tables 1, 2, 3, and 4:

^a The values given in the first line of a cell without any brackets are the number of unique peptides identified (“P”). This is a qualitative measure. ^b The values given in curve brackets (“(Q)”) are the number of spectrum identified. This is a quantitative measure. ^c The value given in square brackets (“[X,Y,Z]”) in the second line are the break up of peptide identifications in terms of parent charge states +1, +2 and +3. Note that X+Y+Z should equal the number of unique peptides identified (P). ^d The values given in flower brackets (“{R%}”) are the percentage decrease in number of peptides identified when compared to results from least decoy database size (results in column 3).

Table 1: Performance comparison between Sequest and the MASPIC system for the PSM Test set with deltcN=0. Each cell contains number of unique peptide identifications ^a, number of spectra identified ^b, number of +1 +2 and +3 peptide identifications ^c and percentage decrease with decoy database size ^d.

Table 1:

PSM Test Set	Result Attribute^{a,b,c,d}	Number of Decoy ORFs = 4467	Number of Decoy ORFs = 8933	Number of Decoy ORFs =17400	Number of Decoy ORFs=34023
MASPIC Tryptic	Peptides	259	251	245	234
	Spectra	(322)	(308)	(300)	(287)
	Charge State	[97,107,55]	[91,107,53]	[85,107,53]	[80,105,49]
	% Decrease		{3.2%}	{5.7%}	{10.7%}
Sequest Tryptic	Peptides	245	236	223	219
	Spectra	(315)	(301)	(283)	(279)
	Charge State	[84,104,57]	[76,103,57]	[72,100,51]	[70,98,51]
	% Decrease		{3.8%}	{9.9%}	{11.9%}
MASPIC Non-Tryptic	Peptides	240	220	209	207
	Spectra	(288)	(264)	(250)	(246)
	Charge State	[71,119,50]	[61,111,48]	[57,110,42]	[57,108,42]
	% Decrease		{9.1%}	{14.8%}	{15.9%}
Sequest Non-Tryptic	Peptides	226	207	200	178
	Spectra	(277)	(249)	(241)	(215)
	Charge State	[69,112,45]	[54,109,44]	[54,108,38]	[46,98,34]
	% Decrease		{9.2%}	{13.0%}	{27.0%}

Table 2: Performance comparison between Sequest and the MASPIC system for the Extended PSM test set with $\text{deltCN}=0$. Each cell contains number of unique peptide identifications ^a, number of spectra identified ^b, number of +1 +2 and +3 peptide identifications ^c and percentage decrease with decoy database size ^d.

Table 2 :

Extended PSM Dataset	Result ^{a,b,c,d} Attribute	Number of Decoy ORFs = 4467	Number of Decoy ORFs = 8933	Number of Decoy ORFs = 17400	Number of Decoy ORFs=34023
MASPIC Tryptic	Peptides	413	404	397	388
	Spectra	(491)	(481)	(471)	(458)
	Charge State	[58,211,144]	[53,207,144]	[52,202,143]	[52,197,139]
	% Decrease		{2.22%}	{4.0%}	{6.4%}
Sequest Tryptic	Peptides	377	337	329	310
	Spectra	(448)	(403)	(394)	(371)
	Charge State	[45,195,137]	[21,182,134]	[24,179,126]	[12,172,126]
	% Decrease		{11.9%}	{14.6%}	{21.6%}
MASPIC Non-Tryptic	Peptides	721	668	639	620
	Spectra	(811)	(749)	(718)	(696)
	Charge State	[79,426,216]	[77,403,188]	[62,401,176]	[54,390,176]
	% Decrease		{7.9%}	{12.8%}	{16.3%}
Sequest Non-Tryptic	Peptides	597	568	525	498
	Spectra	(681)	(649)	(601)	(567)
	Charge State	[26,397,174]	[27,370,171]	[16,355,154]	[16,345,137]
	% Decrease		{5.1%}	{13.7%}	{19.9%}

Table 3: Performance comparison between Sequest and the MASPIC system for the ribosomal dataset with $\Delta CN=0$. Each cell contains number of unique peptide identifications ^a, number of spectra identified ^b, number of +1 +2 and +3 peptide identifications ^c and percentage decrease with decoy database size ^d.

Table 3 :

Ribosome Dataset	Result ^{a,b,c,d} Attribute	Number of Decoy ORFs=6141	Number of Decoy ORFs=12489	Number of Decoy ORFs=25249
MASPIC Tryptic	Peptides	900	882	842
	Spectra	(1738)	(1688)	(1620)
	Charge State	[118,481,301]	[115,475,292]	[102,463,277]
	% Decrease		{2.0%}	{6.9%}
Sequest Tryptic	Peptides	795	734	699
	Spectra	(1531)	(1412)	(1345)
	Charge State	[101,426,268]	[94,397,243]	[77,389,233]
	% Decrease		{8.3%}	{13.7%}
MASPIC Non-Tryptic	Peptides	724	661	624
	Spectra	(1362)	(1253)	(1191)
	Charge State	[62,428,234]	[47,405,209]	[30,395,199]
	% Decrease		{9.5%}	{16.0%}
Sequest Non-Tryptic	Peptides	662	609	561
	Spectra	(1284)	(1180)	(1073)
	Charge State	[65,386,211]	[59,344,206]	[50,321,190]
	% Decrease		{8.7%}	{18.0%}

Table 4: Performance comparison between Sequest and the MASPIC system for the Extended PSM proteinase-K digest with $\text{deltCN}=0$. Each cell contains number of unique peptide identifications ^a, number of spectra identified ^b, number of +1 +2 and +3 peptide identifications ^c and percentage decrease with decoy database size ^d.

Table 4 :

Extended PSM Prot-K	Result ^{a,b,c,d} Attribute	Number of Decoy ORFs = 4467	Number of Decoy ORFs = 8933	Number of Decoy ORFs=17400	Number of Decoy ORFs=34023
MASPIC Non-Tryptic	Peptides	313	313	294	263
	Spectra	(325)	(325)	(305)	(273)
	Charge State	[24,185,104]	[25,185,103]	[16,180,98]	[14,168,81]
	% Decrease		{0.0%}	{6.5%}	{19.0%}
Sequest Non-Tryptic	Peptides	282	270	236	218
	Spectra	(295)	(281)	(247)	(228)
	Charge State	[10,191,81]	[10,174,86]	[0,167,69]	[0,161,57]
	% Decrease		{4.44%}	{19.5%}	{29.4%}

Table 5: This table shows the common ^a and unique ^b peptide identifications made by both the algorithms on the four datasets with $\Delta CN=0$.

Dataset and Search type	Parent Charge	Both Methods	MASPIC	Sequest
PSM Test Tryptic	1	68	12	2
	2	96	9	2
	3	41	8	10
PSM Test Non-Tryptic	1	42	15	4
	2	93	15	5
	3	28	14	6
Extended PSM Tryptic	1	12	40	0
	2	171	26	1
	3	123	16	3
Extended PSM Non-Tryptic	1	16	38	0
	2	336	54	9
	3	122	54	7
Ribosome Tryptic	1	70	32	7
	2	373	90	15
	3	207	70	26
Ribosome Non-Tryptic	1	27	3	23
	2	305	90	16
	3	151	48	39
Extended PSM Prot-K Non-Tryptic	1	0	14	0
	2	143	25	18
	3	45	12	36

^a The number of peptides identified by MASPIC and Sequest in the 3 different charge states. These values are presented in the third column. ^b The number of peptide identifications made by one of the systems and not by the other. These values are provided in column 4 for MASPIC and column 5 for Sequest.