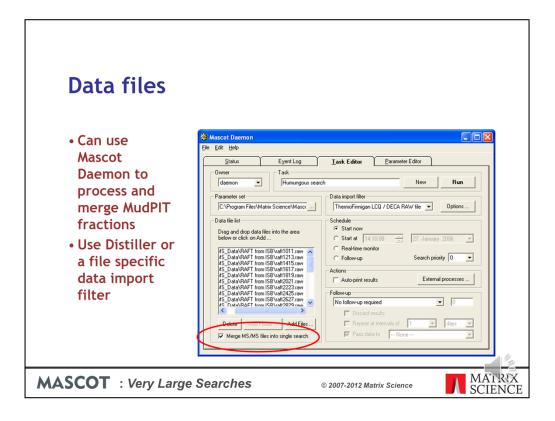


Very large searches present a number of challenges. These are the topics we will cover during this presentation.



The smartest way to merge files, like fractions from a MudPIT run, is using Mascot Daemon. Just tick the box at the bottom left.

The batch can be peak lists or raw files

Note that Mascot Daemon 2.1 had a file size limit of 2 GB. This was lifted in 2.2, and we have successfully merged and searched a 6 GB file, although note that some web servers cannot accept uploads larger than 4 GB

Data	files		
Conc	atenating peak list	:s:	
• D	TA or PKL		
	Download merge.pl from the <pre>http://www.matrixscience.com</pre>		
	Retains filename as scan title	2	
	BEGIN IONS		
	TITLE=raft3031.1706.1706	5.2.dta	
	CHARGE=2+		
	PEPMASS=1243.577388		
	451.1228 5080 487.4352 3283		
	550.4203 5087		
	330.4203 3007		
			12
MASCOT	: Very Large Searches	© 2007-2012 Matrix Science	MAT RX SCIENCE

If you don't want to use Daemon, you can merge peak lists manually.

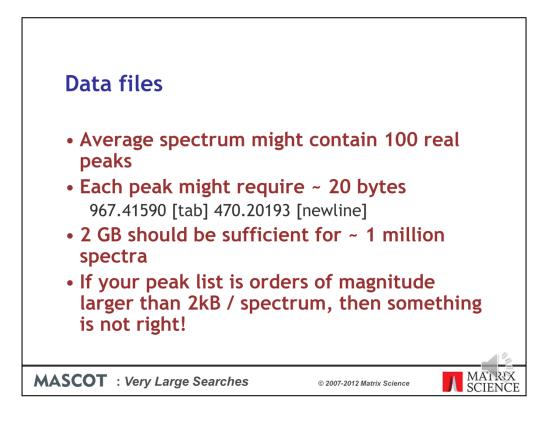
For DTA or PKL, you can download a script from our web site.

A nice feature of this script is that it puts the filename into the scan title, so you can tell which fraction a particular spectrum came from. The scan titles are displayed in the yellow pop-ups on the Mascot result report

Data files	
Concatenating pea	
Windows: copy	c: Command Prompt C:\TEMP>copy *.ngf merge.out
Unix: cat	<pre>Matrix@frill:- [matrix@frill matrix]\$ cat *.mgf > merge.out</pre>
MASCOT : Very Large Search	es © 2007-2012 Matrix Science N SCIENCE

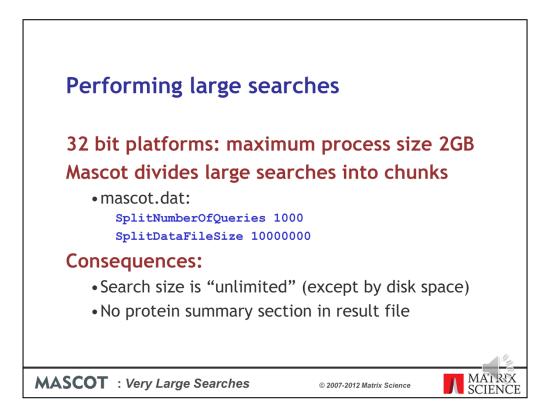
As long as MGF files contain only peak lists, you don't need a script. Just use copy or cat

If the MGF files have search parameters at the beginning, you'll need to remove these before merging the files. Because a number of third party utilities add commands to MGF headers, and these cause a merged search to fail, Mascot Daemon 2.3 and later strips out header lines when merging MGF files.



In talking to Mascot users, it is clear that peak lists files are often much bigger than they should be. In other words, the peak detection is not very good. If you do a back of the envelope calculation, you can see that 2 GB should be enough for approximately 1 million spectra.

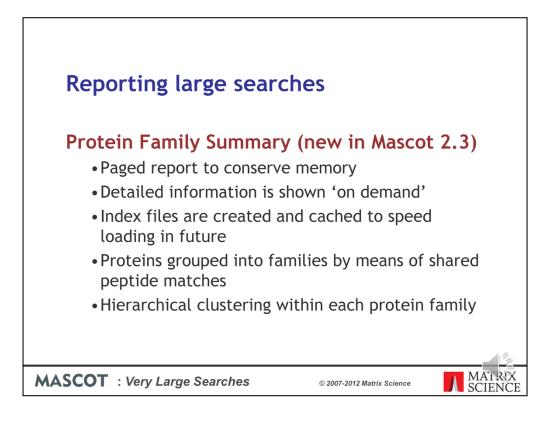
If you intend to do a lot of large searches, its worth getting the peak detection right. Shipping unnecessarily large files around wastes both time and disk space



32 bit platforms have a maximum process size of 2 GB on Windows or 3Gb on Linux. To get around this limit, Mascot divides large searches into smaller chunks, so as to avoid having everything in memory at the same time. The parameters to control this are SplitNumberOfQueries and SplitDataFileSize in the Options section of mascot.dat

One consequence of splitting a search is that there is no protein summary section in the result file. This is not a problem, because no-one wants a protein summary report for a large MS/MS search. However, some old client software gets confused by the missing section. The work around is to increase the values so that large searches never split. Maybe setting SplitNumberOfQueries to 1 million spectra and SplitDataFileSize to 10 billion bytes.

This is OK, but remember to reset these values as soon as you are able to. Otherwise, you might find you run out of memory or address space for your large searches



In Mascot 2.2 and earlier, trying to display result reports for very large searches would often lead to problems with timeouts and running out of memory. To address this, the Protein Family Summary loads most of the information 'on demand'. This requires some index files to be created on the server, and these index files are cached, so that the report loads much faster on the second and subsequent occasions. Proteins are grouped into families by means of shared peptide matches and, within each family, hierarchical clustering is used to illustrate which proteins are closely related and which are more distant.

C sppca	2008 SwissProt Mouse (Mascot Search R	eculte) Windows Internet Fu	alarar			
	 Image: http://bogong/mascot_2_4_0_64/cgi/mascot_2_40_64/cgi/mascot_2_66/cgi/mascot_2_64/cgi/mascot_2_66/cgi/mascot_2_66/cgi/mascot_2_664/cgi/mascot_2_66/cgi/mascot_2_66/cgi/mascot_2_66/cgi/mascot_2		JIOTEI	: [و¢] پ	Canala	
		aster_results_2.pr/me=r901139.dat		· · · · · · · · · · · · · · · · · · ·		
Powermark	s 🏦 \land %					
X X	// IPRG2008 SwissProt Mouse (Mascot Search I	Results)			👌 • 🖶 •	Page • 💮 Tools • 🔞 •
Prot	teins (476) Report Builder Ur	nassigned (27151)				<u>s permalink</u>
		~				
	ein families 1–10 (out of 476					
10	per page 1 2 3 4 5 5	48 Next Expand	I all Collap	se all		
Acces	sion 💙 contains 💙		F	ind		
•1		1::TRY1_BOVIN	1597	TRY1_BOVIN		
· ·						
▶2	—	1 2::CP2CT_MOUSE	1307	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1		
		4 2::CP239_MOUSE	293	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1		
		5 2::CP238_MOUSE	202 535	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1		
		2 2::CP254_MOUSE 6 2::CP270_MOUSE	535	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1 Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2		-
		3 2::CY250_MOUSE	382	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1		
				-,		
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0					
	A A A					
▶3		1 2::GRP78_MOUSE		78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV		
		2 2::HSP7C_MOUSE	353	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1		
	····	3 2::H\$71L_MOUSE	165	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1l PE=2 SV	<i>J</i> =4	
	14 8 8 2 4 4 5					
▶4		2::CYB5_MOUSE	1228	Cytochrome b5 OS=Mus musculus GN=Cyb5a PE=1 SV=2		
▶5		2::PDIA1_MOUSE	1116	Protein disulfide-isomerase OS=Mus musculus GN=P4hb PE=1 SV=1		_
▶6		2::CP1A2_MOUSE	1048	Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1		
▶7	-	1 2::RDH7_MOUSE 2 2::H17B6_MOUSE	1023 612	Retinol dehydrogenase 7 OS=Mus musculus GN=Rdh7 PE=2 SV=1 Hydroxysteroid 17-beta dehydrogenase 6 OS=Mus musculus GN=Hsd17l		
	<u> </u>	2 2	012	nyuroxysteroid 17-beta denyurogenase 6 OS=MUS musculus GR=Hsd1/I	00 PE=2 SV=1	
					67	<u> </u>
Done					Scolar Intranet	100% •
						MANDIN
M	ASCOT : Ver	v Large Seal	rches	© 2007-2012 Matrix Science		MATKA
		,	000	C 2007-2012 Maurix Science		SCIENCE

If there are 300 or more spectra, the Family Summary is the default. This is the appearance of a typical family report immediately after loading. The body of the report consists of three tabs, one for protein families, one for Report Builder, and one for unassigned matches. The report is paged, with a default page size of 10 families. If you wish, you can choose to display a larger number of families on a single page.

Proteins are grouped into families using a novel hierarchical clustering algorithm. If the family contains a single member, the accession string, protein score and description are listed. If the family contains multiple members, the accessions, scores and descriptions are aligned with a dendrogram, which illustrates the degree of similarity between members.

The scores for the proteins in family 2 vary from 1307 down to 69. In the earlier Peptide Summary or Select Summary reports, these would have been at opposite ends of the report. It would have been difficult to recognise that these proteins belonged together, even though they have shared peptide matches and are all cytochrome P450 2C proteins.

	16 %					
🛠 🏉 F	RG2008 SwissProt Mouse (Masci	xt Search Results)				🟠 👻 🖶 👻 🔂 Page 👻 🎯 Tools 🕶
2 -[د د د د د د د د د د د د د د د د د د د	4 2::C 5 2::C 2 2::C 6 2::C 3 2::C	P2CT_MOUSE P239_MOUSE P238_MOUSE P254_MOUSE P270_MOUSE Y250_MOUSE	1307 293 202 535 69 382	Cytochnome P450 2C29 OS=Hur murculus GHI=Cyp2C Cytochnome P450 2C29 OS=Hur murculus GHI=Cyp2C Cytochnome P450 2C28 OS=Hur murculus GHI=Cyp2C Cytochnome P450 2C28 OS=Hur murculus GHI=Cyp2C Cytochnome P450 2C70 OS=Hur murculus GHI=Cyp2C Cytochnome P450 2C50 OS=Hur murculus GHI=Cyp2C	39 PE=2 SV=1 38 PE=2 SV=1 54 PE=2 SV=1 70 PE=2 SV=2
2.1	₫2::CP2CT_MOUSE		61433 86 (8	, , ,	1.47 Cytochrome P450 2C29 OS=Mus musculus (
2.2	d2::CP254_MOUSE		60887 29 (2)		0.87 Cytochrome P450 2C54 OS=Mus musculus (
2.3	2::CY250_MOUSE		61037 25 (2	, , ,	0.87 Cytochrome P450 2C50 OS=Mus musculus C	
2.4	d2::CP239_MOUSE		60932 23 (2	, , , ,	0.33 Cytochrome P450 2C39 OS=Mus musculus C	
2.5	2::CP238_MOUSE		61216 19 (1	, , , ,	0.33 Cytochrome P450 2C38 OS=Mus musculus C	
2.6 Redispla	ef2::CP270_MOUSE	69	61539 5 (5) 4 (4)	0.25 Cytochrome P450 2C70 OS=Mus musculus (SN=Cyp2c70 PE=2 SV=2
	<i>tide matches (34 non-d</i> ît to window	luplicate, 84 duj	olicate)			
		Mr(expt) M			Expect Rank V 1 2 3 4 5 6 Peptide	
Auto-f				535 0 35	0.031 1 R.LPVFDK.A	·
Query	503.8391	1005.6637 10				
✓ Auto-f Query d'3499 d'443	503.8391 9 520.8626	1039.7106 10	39.6157 0.0	949 0 40	0.0079 1 SYLLEK	'SK &
 ✓ Auto-1 Query ∞3499 ∞4433 ∞4444 	503.8391 520.8626 520.8626 521.2416	1039.7106 10 1560.7029 15	39.6157 0.0 59.8187 0.8	949 0 40 842 0 59 0	0.0079 1 SYLLEK .00021 1 U K.NISQSFTNF	
✓ Auto-f Query d'3499 d'443	503.8391 9 520.8626 7 2 521.2416 5 521.3753	1039.7106 10	39.6157 0.0 59.8187 0.8 40.5810 0.1	949 0 40 842 0 59 0 551 0 22	0.0079 1 SYLLEK .00021 1 U K.NISQSFTNF	N + Oxidation (N)
 ✓ Auto-1 Query ∞3499 ∞4433 ∞4443 ∞4444 ∞4460 	503.8391 503.8391 500.8626 520.8626 521.2416 521.3753 525.4566	5 1039.7106 10 5 1560.7029 15 5 1040.7361 10	39.6157 0.0 59.8187 0.8 40.5810 0.1 572.7654 0.5	949 0 40 842 0 59 0 551 0 22 824 0 71 1	0.0079 ▶ 1 SYLLEK .00021 ▶ 1 U K.NISQSFINE 0.031 ▶ 1 U R.FTLMTLR.Y	f + Oxidation (N) FAGR.G
 ✓ Auto-1 Query ∞3499 ∞4433 ∞4443 ∞4444 ∞4466 ∞44703 	503.8391 1 9 520.8626 7 2 521.2416 5 525.4566 1 3 526.2961	i 1039.7106 10 i 1560.7029 15 i 1040.7361 10 i 1573.3479 15	139.6157 0.0 159.8187 0.8 140.5810 0.1 172.7654 0.5 150.5323 0.0	949 0 40 842 0 59 0 551 0 22 824 0 71 1 453 0 35	0.0079 1 SYLLEK .00021 1 U K.NISQSFINE 0.031 1 U R.FTLHILR.) .4e-05 1 U K.EALVDHGEE	N + Oxidation (M) SPAGR.6 C
 ✓ Auto-1 Query ∞3493 ∞4433 ∞4443 ∞4446 ∞4466 ∞44703 ∞44733 	9 503.8391 1 9 520.8626 7 2 521.2416 6 521.3753 5 525.4566 1 3 526.2961 4 8 540.3247	i 1039.7106 10 i 1560.7029 15 i 1040.7361 10 i 1573.3479 15 i 1050.5776 10	139.6157 0.0 159.8187 0.8 140.5810 0.1 172.7654 0.5 150.5323 0.0 178.5385 0.0	949 0 40 842 0 59 0 551 0 22 0 824 0 71 1 453 0 35 0 964 0 54 0	0.0079 1 SYLLEK 0.0021 1 U K.NISQSTIM 0.031 1 U R.FTLMTR.N. 0.46-05 1 U K.ELUDHGEE 0.0005 1 T.CLVEELR.K	N + Oxidation (N) EFAGR.G K.M

If you are interested in family 2, then you click to expand it to show the details. Immediately under the dendrogram is a list of the proteins. The table of peptide matches is similar to that found in the other result reports. Duplicate matches to the same sequence are collapsed into a single row. The columns headed 1, 2, 3, etc. represent the proteins and contain a black square if the peptide is found in the protein. Some matches are shared, but each protein has some unique peptide matches, otherwise it would be dropped as a sub-set.

In this screen shot and the ones that follow, we've set an expect cut-off of 0.05 to simplify the picture by removing low scoring matches

RG2008 SwissProt M	ouse (Mascot Search	h Results)	- Windows	Internet Exp	lorer					
💽 🗸 🙋 http://bogo	ing/mascot_2_4_0_64/cg	gi/master_re:	sults_2.pl?file=I	F981139.dat;_k	gnoreionsscore	below=0.05;_	prefertaxor	iomy=0;_:	sigthreshold=0.05;percolate=0;report=0 💙 👉 🗙 Google	٩
rmarks 👫 🔥 🄏										
🔅 🌈 IPRG2008 Swiss	Prot Mouse (Mascot Sear	rch Results)							🏠 - 🖶 - 🔂 Pa	ge 🔹 🌍 Tools 🔹 🌘
, 'T			::GRP78_M ::HSP7C_N		129				protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	
			::HS71L_M						a protein OS=Mus musculus GN=Hspa8 PE=1 SV=1 1-like OS=Mus musculus GN=Hspa1l PE=2 SV=4	
			_							
140	80 60 20 0									
<		>								
		_								
Threshold (0	i): 0 Cut									
		Score	Mass	Matches	Sequence	as ampa				
3.1 d2::GRP	78_MOUSE	1292	81404	54 (54)	22 (2			glucose-	regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	
3.2 22::HSP	7C_MOUSE	353	78937	23 (23)	9 (9) 0.55	Heatsh	ock cogn	ate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	
3.3 d2::HS7	1L_MOUSE	165	78552	12 (12)	4 (4) 0.25	Heat sh	ock 70 k	Da protein 1-like OS=Mus musculus GN=Hspa1l PE=2 SV=4	
Redisplay All	None									
Auto-fit to windo)W									
Query Dupes	Observed M	r(expt)	Mr(calc)	Delta	M Score	Expect	Rank U	1 2 3	3 Peptide	
2720			974.6004			0.00056	▶1		R.LIGDAAK.N	^
₪3741	508.9092 101					0.035	▶1 U	•	K.IQQLVK.E	
⊠4598 ▶1 ⊠4827 ▶2	523.9268 104 527.4637 105					0.044	1 1	•	R.NTVVPTK.K K.VOVEYK.G	
± 4027 ₽ 2 ± 5924 ▶ 1	546.9979 109					0.0084			K.ITITNDK.G	
z 6994	565.8689 113					0.027	▶1 U		R.LTPEEIER.M	
d7519	573.9761 114	45.9377	1145.6536	6 0.2841	0 38	0.006	▶1 U		R.GTLDPVEK.A	
m ⁹⁰²¹	596.5726 119	91.1306	1190.672	0.4581	0 45	0.0028	▶1 U		R. VMEHFIK. L	
d9459 ▶2	603.8705 120	05.7264	1205.6747	0.0517	0 61	5.5e-05	▶1 U		K.VLEDSDLK.K	
	609.9429 121	17.8713	1217.6486	6 0.2227	0 45	0.0038	▶1 U		K.ITITNDQNR.L	
₫9857) 3				i 0.1872	0 60	0.0001	▶1 U		K.VCNPIITK.L	
	611.4441 123	20.8737	1220.6865	0.1072						
±9857 ≥3 ±9960 ≥2 ±10037 ≥1	612.8115 183	35.4126	1834.8204	0.5923	0 35	0.0074	▶1 U		K.STAGDTHLGGEDFDNR.M	
m 9857 ▶3 m 9960 ▶2 m 10037 ▶1 m 11545	612.8115 183 635.4900 126	35.4126 68.9654	1834.8204 1268.6856	0.5923 0.2799	0 35 0 55	0.00056) 1 U	•	K.ETAEAYLGK.K	
±9857 ≥3 ±9960 ≥2 ±10037 ≥1	612.8115 183	35.4126 68.9654	1834.8204 1268.6856	0.5923 0.2799	0 35 0 55		▶1 U ▶1 U	•	K. ETAEAYLGK. K	
d 9857)3 d 9960)2 d 10037)1 d 11545	612.8115 183 635.4900 126	35.4126 68.9654	1834.8204 1268.6856	0.5923 0.2799	0 35 0 55	0.00056	▶1 U ▶1 U	•	K.ETAEAYLGK.K	10092
m 9857 ▶3 m 9960 ▶2 m 10037 ▶1 m 11545	612.8115 183 635.4900 126 641 5476 199	35.4126 68.9654 81 0806	1834.8204 1268.6856 1980 7996	0.5923	0 35 0 55 0 55	0.00056	▶1 U	•	K. ETAEAYLGK. K	1 A 1 2 2 X

Moving down to family 3, the scale on the dendrogram is ions score, and HSP7C_MOUSE and HS71L_MOUSE join at a score of approximately 30. This represents the score of the significant matches that would have to be discarded in order to make one protein a sub-set of the other. These two proteins are much more similar to one other than to GRP78_MOUSE, which has non-shared peptide matches with a total score of approximately 145. Note that, where there are multiple matches to the same peptide sequence, (ignoring charge state and modification state), it is the highest score for each sequence that is used.

Immediately under the dendrogram is a list of the proteins. In this example, because SwissProt has low redundancy, each family member is a single protein. In other cases, a family member will represent multiple same-set proteins. One of the proteins is chosen as the anchor protein, to be listed first, and the other same-set proteins are collapsed under a same-set heading. There is nothing special about the protein picked for the anchor position. You may have a preference for one according to taxonomy or description, but all proteins in a same-set group are indistinguishable on the basis of the peptide match evidence.

The table of peptide matches is similar to that found in the other result reports. Duplicate matches to the same sequence are collapsed into a single row. Click on the triangle to expand.

The black squares to the right show which peptides are found in which protein. To see the peptides that distinguish HSP7C_MOUSE and HS71L_MOUSE, clear the checkbox for GRP78_MOUSE and choose Redisplay.

marks 🎠 🔥 %										
PRG2008 Swiss	Prot Mouse (Mascot :	Search Results)							🏠 • 🖶 •	🔂 Page 🔹 🍥 Tools
		+ 1 2:	::GRP78_M	OUSE	1292	78 kDa gluco	se-regulated p	orotein OS=Mus musculus GN=Hspa	5 PE=1 SV=3	
-			::HSP7C_M		353			protein OS=Mus musculus GN=Hs		
	-	3 2:	::HS71L_M	OUSE	165	Heat shock 7	0 kDa protein	1-like OS=Mus musculus GN=Hspa	11 PE=2 SV=4	
140 120 100	80 40 80 20 40 80	0								
<		>								
Threshold (0): 0	ut								
		Score			Sequences	; emPAI				
				E4 (E4)	22 (22)	2 17 701	Da alucacere			
	78_MOUSE	1292	81404	54 (54)	22 (22)			egulated protein OS=Mus musculus		
3.2 d'2::HSP	7C_MOUSE	353	78937	23 (23)	9 (9)	0.55 Hea	t shock cogna	te 71 kDa protein OS=Mus musculi	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7	7C_MOUSE					0.55 Hea	t shock cogna		us GN=Hspa8 PE=1 SV=1	
3.2 d'2::HSP 3.3 d'2::HS7 Redisplay All 24 peptide matche	7C_MOUSE 1L_MOUSE None s (11 non-dup	353 165	78937 78552	23 (23)	9 (9)	0.55 Hea	t shock cogna	te 71 kDa protein OS=Mus musculi	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 24 peptide matche Auto-fit to windo	7C_MOUSE 1L_MOUSE None s (11 non-dup	353 165 <i>licate, 13 du</i>	78937 78552 uplicate)	23 (23) 12 (12)	9 (9) 4 (4)	0.55 Hea	ıt shock cogna ıt shock 70 kD	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 1 1 24 peptide matche Auto-fit to windo Query Dupes	7C_MOUSE 1L_MOUSE None s (11 non-dup w Observed	353 165 <i>licate, 13 du</i> Mr(expt)	78937 78552 uplicate) Hr(calc)	23 (23) 12 (12) Delta	9 (9) 4 (4) M Score	0.55 Hea 0.25 Hea Expect Rank	et shock cogna et shock 70 kD t U 2 3 P	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide	us GN=Hspa8 PE=1 SV=1	
3.2 d'2::HSP 3.3 d'2::HS7 Redisplay All 24 peptide matche Auto-fit to windo Query Dupes d'2720 6	7C_MOUSE 1L_MOUSE None s (11 non-dup w Observed 488.3756	353 165 <i>licate, 13 du</i> Mr(expt) 974.7367	78937 78552 <i>uplicate)</i> Mr(calc) 974.6004	23 (23) 12 (12) Delta 0.1363	9 (9) 4 (4) M Score 0 54	0.55 Неа 0.25 Неа С.25 Неа Expect Rank 0.00056 ▶1	t shock cogna t shock 70 kD t U 2 3 P R R	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide . LIGDAAK . N	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 24 peptide matche Auto-fit to windo Query Dupes d2220 d4827 2	7C_MOUSE 1L_MOUSE 400ne s (11 non-dup w 0bserved 488.3756 527.4637	353 165 <i>licate, 13 du</i> Mr(expt) 974.7367 1052.9129	78937 78552 <i>uplicate)</i> Mr(calc) 974.6004 1052.6110	23 (23) 12 (12) Delta 0.1363 0.3020	9 (9) 4 (4) M Score 0 54 0 35	0.55 Hea 0.25 Hea Expect Rank 0.00056 1 0.0094 1	t shock cogna t shock 70 kD t 2 3 P B R V K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide . LIGDAAK, H VyVETX, G	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 24 peptide matche Auto-fit to windo Query Dupes d2220	7C_MOUSE 1L_MOUSE None s (11 non-dup w Observed 488.3756 527.4637 546.9979	353 165 <i>licate, 13 du</i> Mr(expt) 974.7367 1052.9129 1091.9813	78937 78552 (plicate) Mr(calc) 974.6004 1052.6110 1091.6430	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383	9 (9) 4 (4) M Score 0 54 0 35 0 41	0.55 Hea 0.25 Hea 0.25 Hea 0.00056 1 0.00056 1 0.0094 1 0.0084 1	t shock cogna t shock 70 kD t U 2 3 P B R U K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide .LIODAAK.H .VQVEXK.G .TITINOK.G	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 24 peptide matche Auto-fit to windo Query Dupes d2720 d42720 d52720 d52720 d52730 d5274 d5274	7C_MOUSE 1L_MOUSE Jone s (11 non-dup) w Observed 488.3756 527.4637 546.9979 573.9761	353 165 <i>licate, 13 du</i> 974.7367 1052.9129 1091.9813 1145.9377	78937 78552 uplicate) 974.6004 1052.6110 1091.6430 1145.6536	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383 5 0.2841	9 (9) 4 (4) 0 54 0 35 0 41 0 38	0.055 Hea 0.25 Hea 0.0056 1 0.00056 1 0.0094 1 0.0084 1 0.006 1	t shock cogna' t shock 70 kD t 2 3 P m R U R U K U R	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIODAAK.H YQVEXK.G TITNDK.G TITNDK.A	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All 24 peptide matche Auto-fit to windo Query Dupes d5:220 b6 d5:220 b6 d5:224 b1 d75:19 d9960 b2	7C_MOUSE 1L_MOUSE 10000 s (11 non-dup) w Observed 488.3756 527.4637 546.9979 573.9761 611.4441	353 165 <i>licate, 13 du</i> 974.7367 1052.9129 1091.9813 1145.9377 1220.8737	78937 78552 uplicate) Mr(calc) 974.6004 1052.6110 1091.6430 1145.6536 1220.6865	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383 5 0.2841 5 0.1872	9 (9) 4 (4) 0 54 0 35 0 41 0 38 0 60	 0.55 Hea 0.25 Hea 0.25 Hea 0.0056 1 0.0054 1 0.0084 1 0.008 1 0.006 1 0.0001 1 	t shock cogna' t shock 70 kD U R U K U R U K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus L.100AAK.H .VUVETK.G .TITINDK.G .GTLDPVEK.A .VCIPTITK.L	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HS7 Redisplay All [24 peptide matche Auto-fit to windo Query Dupes d12:27) 6 d18:27) 2 d18:27)	7C_MOUSE 1L_MOUSE None s (11 non-dup ww Observed 488.3756 527.4637 546.9979 573.9761 611.4441 612.8115	353 165 <i>licate, 13 du</i> 974.7367 1052.9129 1091.9813 1145.9377 1835.4126	78937 78552 uplicate) Mr(calc) 974.6004 1052.6110 1091.6430 1220.6865 1220.6865	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383 0.2841 0.1872 0.5923	9 (9) 4 (4) 0 54 0 35 0 41 0 38 0 60 0 35	0.55 Head 0.25 Head 0.00056 \$1 0.0094 \$1 0.0084 \$1 0.0086 \$1 0.0006 \$1 0.0006 \$1	t shock cogna' t shock 70 kD U R U K U K U K U K U K U K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIODAAK.H VOVEX.G TITINDK.G VILDEVEK.A VILDEVEK.A VILDEVEK.A STADFUG GEEDEDIR.H	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP Redisplay [All [24 peptide matche] Auto-fit to windo Query Dupes d2:20 b6 d4027 b2 d7519 d7519 d7950 b2 d10037 b1 d1096 b1	7C_MOUSE 1L_MOUSE None s (11 non-dup w Observed 408.3756 527.4637 546.9979 573.9761 611.4441 612.8135 641.5476	353 165 <i>licate, 13 du</i> 974.7367 1052.9129 1091.9813 1145.9377 1220.8737 1220.8737	78937 78552 <i>aplicate)</i> 974.6004 1052.6110 1091.6430 1145.6536 1220.6665 1220.6865 1230.7220	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383 5 0.2841 5 0.1872 0.05923 0.3586	9 (9) 4 (4) 0 54 0 35 0 41 0 38 0 60 0 35 0 55	<pre>0 0.55 Head 0 0.25 Head 0 0.25 Head 0 0.25 Head 0 0.0056 \$1 0.0064 \$1 0.0064 \$1 0.0066 \$1 0.0061 \$1 0.0071 \$1 0.0071 \$1</pre>	t shock cogna t shock 70 kD U 2 3 P U I R U I R U R U I R U I K U I K U I K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LICDAAK. N VQVEYK. 6 TITINK. 6 CILDPVEK. A CILDFVEK. A CILDFVEK. A STAGDTRI.GGEDEDIR. M ETAEATLAG. T	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP Redisplay All [1] 24 peptide matched description Auto-fit to windo description description description	7C_MOUSE 1L_MOUSE None s (11 non-dup) w Observed 488.3756 527.4637 546.9979 573.9761 611.4441 612.8115 641.5376 641.545 647.4422	353 165 <i>licate, 13 du</i> 974.7367 1052.919 1051.9813 1145.9377 1220.8737 1835.4126 1819.3048	78937 78552 uplicate) 974.6004 1092.6110 1091.6430 1145.6536 1220.6865 1834.8220 1818.8253	23 (23) 12 (12) Delta 0.1363 0.33020 0.3303 0.03813 0.0411 0.05923 0.03386 0.03793	9 (9) 4 (4) 0 54 0 35 0 41 0 38 0 60 0 35 0 55	 0.55 Hea 0.25 Hea 0.25 Hea 0.0036 1 0.0034 1 0.004 1 0.004 1 0.001 1 0.0001 1 0.0001 1 3.2e-05 1 	t shock cogna t shock 70 kD 2 3 P 0 8 R 0 8 R 0 8 K 0	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide L.I.GOAAK.N .VQVEX.G .TITIRDK.G .OTLEPVEK.A .STAGOTIL GEOPPDIR.M .EJAPAYLOK.T .ATAGOTIG GEOPPDIR.L	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP Redisplay All 7 T 24 peptide matche 24 uto-ft to windc Query Dupes d2::P20 b6 d::S220 b6 d:S224 b1 d:S24 b1 d:S24 b1 d:S227 b6 d:S227 b1 d:S227 d:S2277 d:S2376	7C_MOUSE 1L_MOUSE None s (11 non-dup w Dserved 488.3756 527.4637 546.9979 573.9761 611.4441 612.8115 641.5476 607.4422 933.0936	353 165 Nicate, 13 du 974,7367 1032.9129 1091.9813 1145.9377 1220.8737 1835.4126 1281.0806 1994.1726	78937 78552 (plicate) 974.6004 1052.6110 1091.6430 1145.6336 1220.6863 1834.8204 1280.7220 1818.8255 1903.9843	23 (23) 12 (12) Delta 0.1363 0.3020 0.3383 0.2841 0.1623 0.35923 0.35963 0.35963 0.35965	9 (9) 4 (4) 0 54 0 0 35 0 41 0 38 0 60 0 38 0 55 0 0 55 0 0 84	D.55 Head 0.25 Head 0.25 Head 0.0056 ▶1 0.0064 ▶1 0.0066 ▶1 0.0061 ▶1 0.00015 ▶1 3.2e-05 ▶1 1.3e-07 ▶1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus muscula a protein 1-like OS=Mus musculus eptide LIODAAK. N VOTEXK. G TITTIDKK. G CILDPVEK. A VCIRPITIK. L STAGDTM.GEOEPDIR. M RIARAVLGK. T ATAGDTM.GEOEPDIR. L STYPEEVS SNV.TK. M	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP 2.3.4 d2::HSP Redisplay All 24 peptide matche 2 Auto-fit to windo Query Dupes d2720 d4827 d5950 d1007 d10960 d10977 d10977 d125277 d25876 d25876	7C_MOUSE 1L_MOUSE Jone s (11 non-dup w Dbserved 408.3756 527.4637 346.9979 573.9761 611.4441 612.8115 641.5476 607.4422 953.0936 650.1325	353 165 <i>licate, 13 du</i> 974, 7367 1052, 9129 1091, 9813 1145, 9377 1220, 0737 1220, 0737 1283, 4126 1281, 0806 1819, 3048 1904, 1726	78937 78552 (plicate) 1052 6104 1052 6110 1091 6430 1145 6336 1220 6665 1220 6665 1234 8240 1284 8240 1284 8251 1903 9845 1947 0920	23 (23) 12 (12) Delta 0.1363 0.3303 0.2441 0.15923 0.3586 0.4793 0.1881 0.1891 0.2836	9 (9) 4 (4) 0 54 (4) 0 35 0 41 0 38 0 60 35 0 35 0 35 0 55 0 84 37	0.55 Head 0.25 Head 0.025 Head 0.0056 1 0.0056 1 0.0064 1 0.0064 1 0.0061 1 0.00015 1 3.2e-05 1 1.3e-07 1 0.013 1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIGDAAK.N VQVETK.G TITITUK.G TIDEVTEK.A VCHPIITK.L STAGOTH.GOEDFDIR.M STAFPEEVSSRULTK.T TIMETTAALTAVGLDK	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP Redisplay All 7 T 24 peptide matche 24 uto-ft to windc Query Dupes d2::P20 b6 d::S220 b6 d:S224 b1 d:S24 b1 d:S24 b1 d:S227 b6 d:S227 b1 d:S227 d:S2277 d:S2376	7C_MOUSE 1L_MOUSE Jone s (11 non-dup w Dbserved 408.3756 527.4637 346.9979 573.9761 611.4441 612.8115 641.5476 607.4422 953.0936 650.1325	353 165 Nicate, 13 du 974,7367 1032.9129 1091.9813 1145.9377 1220.8737 1835.4126 1281.0806 1994.1726	78937 78552 (plicate) 1052 6104 1052 6110 1091 6430 1145 6336 1220 6665 1220 6665 1234 8240 1284 8240 1284 8251 1903 9845 1947 0920	23 (23) 12 (12) Delta 0.1363 0.3303 0.2441 0.15923 0.3586 0.4793 0.1881 0.1891 0.2836	9 (9) 4 (4) 0 54 (4) 0 35 0 41 0 38 0 60 35 0 35 0 35 0 55 0 84 37	D.55 Head 0.25 Head 0.25 Head 0.0056 ▶1 0.0064 ▶1 0.0066 ▶1 0.0061 ▶1 0.00015 ▶1 3.2e-05 ▶1 1.3e-07 ▶1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus muscula a protein 1-like OS=Mus musculus eptide LIODAAK. N VOTEXK. G TITTIDKK. G CILDPVEK. A VCIRPITIK. L STAGDTM.GEOEPDIR. M RIARAVLGK. T ATAGDTM.GEOEPDIR. L STYPEEVS SNV.TK. M	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP 2.3.4 d2::HSP Redisplay All 24 peptide matche 2 Auto-fit to windo Query Dupes d2720 d4827 d5950 d1007 d10960 d10977 d10977 d125277 d25876 d25876	7C_MOUSE 1L_MOUSE Jone s (11 non-dup w Dbserved 408.3756 527.4637 346.9979 573.9761 611.4441 612.8115 641.5476 607.4422 953.0936 650.1325	353 165 <i>licate, 13 du</i> 974, 7367 1052, 9129 1091, 9813 1145, 9377 1220, 0737 1220, 0737 1283, 4126 1281, 0806 1819, 3048 1904, 1726	78937 78552 (plicate) 1052 6104 1052 6110 1091 6430 1145 6336 1220 6665 1220 6665 1234 8240 1284 8240 1284 8251 1903 9845 1947 0920	23 (23) 12 (12) Delta 0.1363 0.3303 0.2441 0.15923 0.3586 0.4793 0.1881 0.1891 0.2836	9 (9) 4 (4) 0 54 (4) 0 35 0 41 0 38 0 60 35 0 35 0 35 0 55 0 84 37	0.55 Head 0.25 Head 0.025 Head 0.0056 1 0.0056 1 0.0064 1 0.0064 1 0.0061 1 0.00015 1 3.2e-05 1 1.3e-07 1 0.013 1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIGDAAK.N VQVETK.G TITITUK.G TIDEVTEK.A VCHPIITK.L STAGOTH.GOEDFDIR.M STAFPEEVSSRULTK.T TIMETTAALTAVGLDK	us GN=Hspa8 PE=1 SV=1	
3.2 d2:HSP 3.3 d2:HSP Redisplay [Al] [24 peptide matche] Auto-fit to winds due-fit to winds direct page direct pa	7C_MOUSE 1L_MOUSE None s (11 non-dup w W 0bserved 488.3756 527.4637 546.9979 573.9761 611.4441 612.8115 641.5376 641.5376 641.5376 641.5376 641.5376 641.5376 641.5376	353 165 <i>licate</i> , 13 du 974.7367 1052.9129 1091.981.145.9377 1220.0737 1231.0806 1819.3048 1904.1726 1947.3756 1947.3756	78937 78552 (plicate) 974.6004 1092.610 1091.6430 1145.6336 1220.6665 1834.8204 1280.7220 1818.8255 1903.9845 1903.9845 1947.0920	23 (23) 12 (12) Delta 0.1363 0.3303 0.2441 0.15923 0.3586 0.4793 0.1881 0.1891 0.2836	9 (9) 4 (4) 0 54 (4) 0 35 0 41 0 38 0 60 35 0 35 0 35 0 55 0 84 37	0.55 Head 0.25 Head 0.025 Head 0.0056 1 0.0056 1 0.0064 1 0.0064 1 0.0061 1 0.00015 1 3.2e-05 1 1.3e-07 1 0.013 1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIGDAAK.N VQVETK.G TITITUK.G TIDEVTEK.A VCHPIITK.L STAGOTH.GOEDFDIR.M STAFPEEVSSRULTK.T TIMETTAALTAVGLDK	us GN=Hspa8 PE=1 SV=1	
3.2 d2::HSP 3.3 d2::HSP 2.3.4 d2::HSP Redisplay All 24 peptide matche 2 Auto-fit to windo Query Dupes d2720 d4827 d5950 d1007 d10960 d10977 d10977 d125277 d25876 d25876	7C_MOUSE 1L_MOUSE None s (11 non-dup w W 0bserved 488.3756 527.4637 546.9979 573.9761 611.4441 612.8115 641.5376 641.5376 641.5376 641.5376 641.5376 641.5376 641.5376	353 165 <i>licate</i> , 13 du 974.7367 1052.9129 1091.981.145.9377 1220.0737 1231.0806 1819.3048 1904.1726 1947.3756 1947.3756	78937 78552 (plicate) 974.6004 1092.610 1091.6430 1145.6336 1220.6665 1834.8204 1280.7220 1818.8255 1903.9845 1903.9845 1947.0920	23 (23) 12 (12) Delta 0.1363 0.3303 0.2441 0.15923 0.3586 0.4793 0.1881 0.1891 0.2836	9 (9) 4 (4) 0 54 (4) 0 35 0 41 0 38 0 60 35 0 35 0 35 0 55 0 84 37	0.55 Head 0.25 Head 0.025 Head 0.0056 1 0.0056 1 0.0064 1 0.0064 1 0.0061 1 0.00015 1 3.2e-05 1 1.3e-07 1 0.013 1	t shock cogna t shock 70 kD U 2 3 P U 8 R U 8 R U 8 K U 8 K	te 71 kDa protein OS=Mus musculu a protein 1-like OS=Mus musculus eptide LIGDAAK.N VQVETK.G TITITUK.G TIDEVTEK.A VCHPIITK.L STAGOTH.GOEDFDIR.M STAFPEEVSSRULTK.T TIMETTAALTAVGLDK	us GN=Hspa8 PE=1 SV=1	007

It can now be seen that HS71L_MOUSE would be a sub-set of HSP7C_MOUSE if it was not for one match, K.ATAGDTHLGGEDFDNR.L. It is the significant score for this match that separates the two proteins in the dendrogram by a distance of 32 (score of 55 - homology threshold score of 23).

You can "cut" the dendrogram using the slider control.

	use (Mascot Se	arch Results)) - Windows Ir	nternet Explo	rer					
0	ng/mascot_2_4_0_6	64/cgi/master_re	sults_2.pl?file=F9	981139.dat;_ign	oreionsscore	below=0.05;_pref	ertaxonom	iy=0;_sig	threshold=0.05;percolate=0;report=0 🛩 47 🗙 Google	٩
ermarks 膬 \land %										
🔅 PRG2008 Swiss	Prot Mouse (Mascot	Search Results)							<u>6</u> • e	🖶 🔹 🔂 Page 🔹 🎯 Tools 👻 🌘
3		1 2	::GRP78_MC	DUSE	129	2 78 kDa glu	cose-regu	lated p	otein OS=Mus musculus GN=Hspa5 PE=1 SV=3	
		^{2 2}	::HSP7C_M	DUSE	35	3 Heat shock	cognate :	71 kDa	protein OS=Mus musculus GN=Hspa8 PE=1 SV=1	
		- 'רי								
140	5 4 6 6	0								
<		>								
		<u> </u>								
Threshold (50	1): 50	Cut								
		Score	Mass N	datches S	equence	s emPAI				
☑ 3.1 Ø2::GRP	8_MOUSE	1292	81404	54 (54)	22 (22		8 kDa glu	cose-re	gulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3	
3.2 d2::HSP	7C_MOUSE	353	78937	23 (23)	9 (9) 0.55 н	eat shock	cognat	e 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=:	1
Redisplay All N	lone									
(reareplay) (ref	iono -									
▼68 peptide matche		olicate, 34 di	uplicate)							
68 peptide matches Auto-fit to windo Query Dupes		Mr(expt)	uplicate) Mr(calc)	Delta I	1 Score	Expect Ra	nk U 1	L 2 Pe	ptide	
Auto-fit to windo	W		Mr(calc)	Delta 1 0.1363 (Expect Ra 0.00056 🕨			ptide LIGDAK.N	
Auto-fit to windo Query Dupes	0bserved 488.3756	Mr(expt)	Mr(calc) 974.6004		54		1 🗖	R.		^
 ✓ Auto-fit to windo Query Dupes ^d2720 ▶6 ^d3741 ^d4598 ▶1 	W 0bserved 488.3756 508.9092 523.9268	Mr(expt) 974.7367 1015.8039 1045.8390	Mr(calc) 974.6004 1015.6633 1045.6375	0.1363 0.1406 0.2015) 54) 34) 27	0.00056 0.035 0.044	1 U 1 U 1 U	R. K. R.	LIGDAAK.N IQQLVK.E NTVVPTK.K	^
 ✓ Auto-fit to windo Query Dupes ∞2720 ▶ 6 ∞3741 ∞4598 ▶ 1 ∞4827 ▶ 2 	Observed 488.3756 508.9092 523.9268 527.4637	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110	0.1363 0.1406 0.2015 0.3020	0 54 0 34 0 27 0 35	0.00056) 0.035) 0.044) 0.0094)	1 🗖	R. K. R.	LIGDAAK.N 100LVK.E NTVVPTK.K VQVEYK.G	^
 ✓ Auto-fit to windo Query Dupes ∞2720 ▶ 6 ∞3741 ∞4598 ▶ 1 ∞44827 ▶ 2 ∞5924 ▶ 1 	Observed 488.3756 508.9092 523.9268 527.4637 546.9979	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430	0.1363 0.1406 0.2015 0.3020 0.3383) 54) 34) 27) 35) 41	0.00056) 0.035) 0.044) 0.0094) 0.0084)	1 U 1 U 1 U 1 U	R. K. R. K.	LIGDAAK.N IQQLYK.E NTVVPTK.K YVVPTK.G ITITNDK.G	^
 ✓ Auto-fit to windo Query Dupes ∞2720 ▶ 6 ∞3741 ∞4558 ▶ 1 ∞4558 ▶ 1 ∞45924 ▶ 1 ∞65924 ▶ 1 ∞6594 	Observed 488.3756 508.9092 523.9268 527.4637 546.9979 565.8689	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131	0 54 0 34 0 27 0 35 0 41 0 32	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.027)	1 U 1 U 1 U	R. K. K. K. R.	LIGDAR.H IQUVK.E HYVPTK.K VQVEX.G ITITHOK.G LITHERER.M	^
✓ Auto-fit to windo Query Dupes d2720 ▶6 d3741 d4598 ▶1 d4827 ▶2 d5924 ▶1 d6994 d7519	Observed 488.3756 508.9092 523.9268 527.4637 546.9979 565.8689 573.9761	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841	0 54 0 34 0 27 0 35 0 41 0 32 0 38	0.00056 0.035 0.044 0.0094 0.0084 0.027 0.006	1 U 1 U 1 U 1 U 1 U 1 U	R. K. K. K. R. R.	I GOAAK, N IQUIVK, E NYVVPIK, K VQVEYK, G ITITIDK, G LIPEELEE, M GTLDPVEK, A	•
✓ Auto-fit to windo Query Dupes d2720 ≥ 6 d3741 d4598 ≥ 1 d*827 ≥ 2 d5924 ≥ 1 d*6994 d*7519 d*9021	W Observed 488.3756 508.9092 523.9268 527.4637 546.9979 565.8689 573.9761 596.5726	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377 1191.1306	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581	0 54 0 34 0 27 0 35 0 41 0 32 0 38 0 45	0.00056 0.035 0.044 0.0094 0.0084 0.027 0.006 0.0028 0.0028	1 U 1 U 1 U 1 U	R. K. K. R. R. R.	LIGGAAK, H IQQLYK, E HYVPTK, K VQVEYK, G ITTHDK, G LIPBELER, H OTLDFVEK, A WHEPTK, L	
✓ Auto-fit to windo Query Dupes d2720 d3741 d4827 23924 d5924 d5934 d7519 d9459 23945	Observed 488.3756 508.9092 523.9268 527.4637 546.9979 555.8689 573.9761 596.5726 603.8705	Mr(expt) 974,7367 1015,8039 1045,8390 1052,9129 1091,9813 1129,7232 1145,9377 1191,1306 1205,7264	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1205.6747	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.0517) 54) 34) 27) 35) 41) 32) 38) 45) 61	0.00056 0.035 0.044 0.0094 0.0084 0.027 0.006 0.0028 5.5e-05	1 U 1 U 1 U 1 U 1 U 1 U 1 U	R. K. K. K. R. R. R.	LIGGARN H IQUIVK.E HYVPTK.K VQVEYK.G ITITHOK.G ITIPEETER.M GTLOPVEK.A VHENTK.L VIENDLK.K	•
✓ Auto-fit to windo Query Dupes d2720 ≥ 6 d3741 d4598 ≥ 1 d*827 ≥ 2 d5924 ≥ 1 d*6994 d*7519 d*9021	W Observed 488.3756 508.9092 523.9268 527.4637 546.9979 565.8689 573.9761 596.5726 603.8705 609.9429	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377 1191.1306	Mr (calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1205.6747 1217.6486	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581) 54) 34) 27) 35) 41) 32) 38) 45) 61) 45	0.00056 0.035 0.044 0.0094 0.0084 0.027 0.006 0.0028 0.0028	1 U 1 U 1 U 1 U 1 U 1 U 1 U 1 U	R. K. K. K. R. R. K. K.	LIGGAAK, H IQQLYK, E HYVPTK, K VQVEYK, G ITTHDK, G LIPBELER, H OTLDFVEK, A WHEPTK, L	
✓ Auto-fit to windo Query Dupes d2720 ▶6 d3741 d4882 ▶1 d4827 ▶2 d5934 ▶1 d5934 ± d5934 d7519 d9021 d9955 ▶2 d3855 ▶3	W Observed 488.3756 508.9092 523.9268 527.4637 546.9979 565.8689 573.9761 596.5726 603.8705 609.9429 611.4441	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377 1191.1306 1205.7264 1217.8713	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1205.6747 1217.6486 1220.6865	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.0517 0.2227) 54) 34) 27) 35) 41) 32) 38) 45) 61) 45) 60	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.027) 0.006) 0.0028) 5.5e-05) 0.0038)	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. R. R. K. K.	LIGGAAK, N IQQLVK, E NYVPTK, K VQVEYK, G ITTHDK, G LIPBELER, H GILDPVEK, A VIEBSDLK, K ITTHDQMF, L	
✓ Auto-fit to windo Query Dupes d2720 ▷ 6 d3741 d4558 ▷ 1 d4529 ▷ 1 d5924 ▷ 1 d5994 d7519 d5994 d7519 d5994 d7519 d5994 d7519 d5996 ▷ 2	W Observed 488.3756 508.9092 523.9268 527.4637 546.9979 553.8689 573.9761 596.5726 603.8705 609.9429 611.4441 612.8115	Mr(expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377 1191.1306 1205.7264 1217.8713 1220.8737	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1205.6747 1227.6486 1220.6865 1834.8204	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.4581 0.0517 0.2227 0.1872) 54) 34) 27) 35) 41) 32) 38) 45) 61) 45) 60	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.0084) 0.0027) 0.0028) 0.0028) 0.0028) 0.0038)	1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	R. K. K. K. R. R. K. K. K.	LIGGARK. N IQQLVK.E NYVVPTK.K VQVEXK.G ITPEEIER.M GTLBVVEK.A WEBETIK.L VLEDSILK.K ITITNDQNR.L VLEDSUK.K	
✓ Auto-fit to windo Query Dupes d2720 b6 d2741 d4825 b1 d6894 d7519 d7519 d7519 d9455 b2 d9855 b2 d9857 b3 d99960 b2 d75097 1	W Observed 488.3756 508.9092 523.9268 527.4637 565.8689 573.9761 596.5726 603.8705 609.9429 611.4441 612.8115 635.4900	Mr (expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7232 1145.9377 1191.1306 1205.7264 1217.8713 1220.8737 1835.4126	Mr(calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1205.6747 1217.6486 1220.6865 1203.6856	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.0517 0.2227 0.1872 0.5923) 54) 34) 27) 35) 41) 32) 38) 45) 61) 60) 35) 55	0.00056) 0.035) 0.044) 0.0084) 0.0084) 0.0027) 0.0028) 5.5e-05) 0.0038) 0.0038) 0.0031)	1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	R. K. K. K. R. R. K. K. K. K.	LIGDAR. N IQULVK. E HYVPTK. K VQVEYK. G ITITNOK. G LIFPECIER. M GTLOPVEK. A VIENGLK. K ITITNOJUR. L VIENGLK. K ITITNOJUR. L STAGDTH. GGEOEPDIR. M	•
✓ Auto-fit to windo Query Dupes d2720 >6 d3741 d4598 >1 d4598 >1 d5994 d7719 d9021 d9857 >3 d9859 >2 d10000 >1 d10000 >1	W Observed 488.3756 508.9092 523.9268 327.4637 546.9979 565.8689 773.9761 596.5726 603.8705 609.9429 611.4441 612.8115 635.4900 641.5476	Hr (expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9813 1129.7222 1145.9377 1191.1306 1205.7264 1217.8713 1220.8737 1268.9654	Mr (calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6336 1190.6725 1205.6747 1217.6486 1220.6665 1834.8204 1266.6856 1280.7220	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.4581 0.2227 0.1872 0.2227 0.1872 0.5923) 54) 34) 27) 35) 41) 32) 38) 45) 61) 60) 35) 55) 55	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.0027) 0.0028) 0.0028) 0.0028) 0.0038) 0.0003) 0.0074)	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. R. R. K. K. K. K.	LIGGAAK, N IQQLVK, E IYVVPTK, K VQVEYK, G ITTINDK, G LIPBETER, M GFLDPVEK, A VIEDSIK, K VIEDSIK, K VIEDSIK, K STAGDTH, GGEDEDRR, M ETABALIG, K	
✓ Auto-fit to windo Query Dupes a2720 b6 a27741 a4589 b1 a4582 b2 a5944 a4595 b2 a5954 a5954 a5955 b2 a5956 b2 a5956 b2 a5956 b2 a5956 b2 a510037 b1 a11345 b1	W Observed 488.3756 508.9092 523.9268 327.4637 546.9979 565.8689 773.9761 596.5726 603.8705 609.9429 611.4441 612.8115 635.4900 641.5476	Hr (expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9413 1129.7232 1145.9377 1191.1306 1205.7264 1217.8713 1220.8737 1835.4126 1268.9654 1268.9654	Mr (calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6336 1190.6725 1205.6747 1217.6486 1220.6665 1834.8204 1266.6856 1280.7220	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.0517 0.2227 0.1872 0.5233 0.2799 0.3586) 54) 34) 27) 35) 41) 32) 38) 45) 61) 60) 35) 55) 55	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.0028) 0.0028) 5.5e-05) 0.0038) 0.0038) 0.0001) 0.00056)	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. R. R. K. K. K. K.	LIGDAR. H IQU.VK. E HYVOPIK. K VQVEYK. G IITHIDO. G LIPPETER. M GTLEPVEK. A WEEDYIK. L VUEEDSLK. K IITHIDOUR. L VIEEDSLK. K IITHIDOUR. L STADTH.GEEDFDHR. M ETAEAYLOK. K	anat D0f
✓ Auto-fit to windo Query Dupes d2720 > 6 d2741 d5989 > 1 d5989 > 1 d5989 > 2 d5984 > 2 d7985 > 2 d9859 > 2 d1845 > 1 d1846 > 1	W Observed 488.3756 508.9092 523.9268 327.4637 546.9979 565.8689 773.9761 596.5726 603.8705 609.9429 611.4441 612.8115 635.4900 641.5476	Hr (expt) 974.7367 1015.8039 1045.8390 1052.9129 1091.9413 1129.7232 1145.9377 1191.1306 1205.7264 1217.8713 1220.8737 1835.4126 1268.9654 1268.9654	Mr (calc) 974.6004 1015.6633 1045.6375 1052.6110 1091.6430 1129.6101 1145.6336 1190.6725 1205.6747 1217.6486 1220.6665 1834.8204 1266.6856 1280.7220	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.2841 0.4581 0.0517 0.2227 0.1872 0.5233 0.2799 0.3586) 54) 34) 27) 35) 41) 32) 38) 45) 61) 60) 35) 55) 55	0.00056) 0.035) 0.044) 0.0094) 0.0084) 0.0028) 0.0028) 5.5e-05) 0.0038) 0.0038) 0.0001) 0.00056)	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. R. R. K. K. K. K.	LIGGARK. N IQQLVK. E IQQLVK. E VVVVTK. K VVVVTK. K ITTINDK. G LTPEELER. M GILDPVER. A VIEMFIK. L VIEMSUK. K VIEMSUK. K STADDTH. GEOEPDNR. M ETADATI. GK. T ETADATI. GK. T ETADATI. GK. T	anet DT
✓ Auto-fit to windo Query Dupes a2720 > 6 a2721 > 6 a4527 + 1 a4589 > 1 a4582 + 1 a4592 + 1 a5924 + 1 a5924 > 1 a5952 + 2 a5955 + 2 a5955 + 2 a5955 + 2 a5957 + 3 a11046 + 1 a11144 + 1 a118194	W Dbserved 408.3756 500.9092 523.9268 527.4637 546.9279 565.8689 573.9761 596.5726 603.8705 603.9426 611.4441 612.8115 633.4900 641.5476 740.3568	Hr (expt) 974.7367 1015.8039 1052.9129 1091.9813 1129.7212 1145.9377 1191.1366 1205.7264 1217.8713 1220.8737 1635.4126 1268.9554 1268.9554 1268.9554	Mr (calc) 974.6004 1015.6633 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1203.6747 1217.6486 1220.6665 1230.6665 1280.7220 1478.8336	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.4381 0.4381 0.4381 0.4381 0.4381 0.4385 0.3455 0.2799 0.3386 0.3455) 54) 34) 27) 35) 41) 38) 45) 61) 45) 60) 35) 55) 55) 42	0.00056 0 0.035 0 0.094 0 0.094 0 0.0027 0 0.0026 0 0.0026 0 0.00056 0 0.00015 0 0.00015 0 0.0001 0 0.0005 0 0.005 0 0	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. K. R. R. K. K. K. K. K. K.	LIGGARK. N IQQLVK. E IYVVPTK. K VQVEYK. 6 LIFPETER. M OTLDFVEK. A VUEDSTK. L VUEDSTK. L VUEDSTK. L VUEDSTK. L VUEDSTK. L STAGDTHLGGEDEPAR. M ETABATLOK. K ETABATLOK. K ETABATLOK. K ETABATLOK. C MEDDEPAR. D	anet Doff
✓ Auto-fit to windo Query Dupes d2720 > 6 d2741 d5989 > 1 d5989 > 1 d5989 > 2 d5984 > 2 d7985 > 2 d9859 > 2 d1845 > 1 d1846 > 1	W Dbserved 408.3756 500.9092 523.9268 527.4637 546.9279 565.8689 573.9761 596.5726 603.8705 603.9426 611.4441 612.8115 633.4900 641.5476 740.3568	Hr (expt) 974.7367 1015.8039 1052.9129 1091.9813 1129.7212 1145.9377 1191.1366 1205.7264 1217.8713 1220.8737 1635.4126 1268.9554 1268.9554 1268.9554	Mr (calc) 974.6004 1015.6633 1052.6110 1091.6430 1129.6101 1145.6536 1190.6725 1203.6747 1203.6747 1217.6486 1220.6665 1230.6665 1280.7220 1478.8336	0.1363 0.1406 0.2015 0.3020 0.3383 0.1131 0.4381 0.4381 0.4381 0.4381 0.4381 0.4385 0.3455 0.2799 0.3386 0.3455) 54) 34) 27) 35) 41) 38) 45) 61) 45) 60) 35) 55) 55) 42	0.00056 0 0.035 0 0.094 0 0.094 0 0.0027 0 0.0026 0 0.0026 0 0.00056 0 0.00015 0 0.00015 0 0.0001 0 0.0005 0 0.005 0 0	1 V 1 V 1 V 1 V 1 V 1 V 1 V 1 V	R. K. K. K. K. R. R. K. K. K. K. K. K.	LIGGARK. N IQQLVK. E IQQLVK. E VVVVTK. K VVVVTK. K ITTINDK. G LTPEELER. M GILDPVER. A VIEMFIK. L VIEMSUK. K VIEMSUK. K STADDTH. GEOEPDNR. M ETADATI. GK. T ETADATI. GK. T ETADATI. GK. T	anet por MATRXX

If we cut the dendrogram at a score of 50, HS71L_MOUSE will be dropped because it is now a sub-set protein. If you compare the matches to HSP7C_MOUSE with those to GRP78_MOUSE, it is clear that these are very different proteins. They are part of the same family because of two shared matches, but many highly significant matches would have to be discarded for either protein to become a sub-set of the other. In summary, we can quickly deduce from the Family Summary that there is abundant evidence that both GRP78_MOUSE and HSP7C_MOUSE were present in the sample. There is little evidence for HS71L_MOUSE. It is more likely that the HSP7C_MOUSE contained a SNP or two relative to the database sequence.

iPRG2008 SwissProt Mo		,		,					
	g/mascot_2_4_0_	64/cgi/master_res	ults_2.pl?file=F9	81139.dat;_ignor	eionsscoreb	elow=0.05;_	preferta	xonon	nomy=0;_sigthreshold=0.05;percolate=0;report=0 💌 47 🔀 Google
rermarks 🕅 🔥 %									
🖗 PRG2008 SwissF	rot Mouse (Mascot	: Search Results)							🛐 👻 👼 👻 🔂 Page 👻 🎯 Tools 🕶
Proteins (445)	eport Builder	Unassign	ied (30350)]					§. permalink
Protein families 4	1 50 (of 445)							
and the second s		2 3 4 5	<u>6</u> <u>7</u> <u>8</u> <u>9</u>	2 10 45	Next	_	pand a		Collapse all
Sequence 🔽 is e	qual to 🔽 MNV	/LADALK				F	nd C	lear	r
41 2::DHI1_	MOUSE	358	B Corticoster	oid 11-beta-del	ydrogena	ise isozyme	1 OS=	Mus r	s musculus GN=Hsd11b1 PE=1 SV=3
42 2::R\$19_	MOUSE	355	5 40S ribosor	mal protein \$19	OS=Mus	musculus (N=Rps	19 PE	PE=1 SV=3
43 2::RS3_M	IOUSE	353	3 40S ribosor	mal protein S3	OS=Mus r	nusculus Gi	I=Rps3	PE=1	=1 SV=1
44 2::RL22_	MOUSE	347	7 60S ribosor	mal protein L22	OS=Mus	musculus (iN=Rpl2	2 PE+	PE=2 SV=2
45 2::RS15/	MOUSE	344	4 40S ribosor	mal protein \$15	a OS=Mu	s musculus	GN=Rp	s15a	5a PE=2 SV=2
		Score		latches Se					
45.1 d2::RS15	A_MOUSE	344		16 (16)	3 (3			iboso	somal protein S15a OS=Mus musculus GN=Rps15a PE=2 SV=2
▼16 peptide matches	(A non-duni	icata 12 dun	licate)						
Auto-fit to windo		icute, 12 dap	incute)						
_									
Query Dupes 23708 ▶ 5		Mr(expt) 1014.7407		Delta M 0.1100 0		Expect 0.00053			Peptide K.IVVNLTGR.L
₫11285 ▼5	631.9663	1261.9180	1261.7308	0.1872 0	77	2.3e-06	▶1	U	R. MIVLADALK . S
£11274		1261.7591		0.0284 0		1.8e-05	▶1		R. MNVLADALK . S
£11275		1261.7682		0.0375 0		9.4e-05	1		R. MIVLADALK . S
d 11283		1261.8686		0.1379 0		0.00012	1		R. MIVLADALK . S
11287		1262.0014		0.2706 0	(42)	0.0063	P1		R. MIVLADALK . S
d 11288		1262.0291		0.2983 0		6.2e-05	1		R. MIVLADALK . S
g11604 🕨 1		1270.9355		0.2452 0	28	0.03	1		K.WQNNLLPSR.Q
		1277.7762		0.0505 0	50	0.00081	1		R. MNVLADALK.S + Oxidation (M)
±11780 ▼1	639.9899	1277.9652	1277.7257	0.2396 0	(48)	0.00054	▶1	U.	R. NNVLADALK.S + Oxidation (10)
±11780 ▼1 ±11790								_	
									Scal intranet
<i>±11790</i>								-	Local Intranet
₫11790) T . 1	Very I	arge	Searc	has		_	_	© 2007-2012 Matrix Science

The family report also includes a text search facility, which is particularly important for a paged report. You can search by accession or description sub-string, or by query, mass or sequence. Here, for example, we searched for a peptide sequence. The display jumps to the first instance of the sequence, expands, and highlights (in green) the target peptides.

C iPRG200	8 Sw	issProt Mous	e (Mascot Search Result	s) - Wind	ows Intern	net Explorer					
00-	6	http://bogong/r	mascot_2_4_0_64/cgi/master_	results_2.pl	?file=F98113	9.dat;_ignoreio	nsscorebe	elow=0.05;_prefert	axonomy	=0;_sigthres	hold=0.05;percolate=0;report=0; 💙 47 🗙 Google
Powermarks		\$									
* * 🖸		2008 SwissProt	Mouse (Mascot Search Result	0							🏠 👻 📾 👻 🔂 Page 🕶 🎯 Tools 🕶 🔞 -
•••••••••••••••••••••••••••••••••••••••				-,							
Protein	ns (4	45) Rep	ort Builder Unassi	gned (30	350)						§ permalink
		s (470 pi									
▶ Colum	ns: §	Standard (1	12 out of 12)								
Filters		200									
Fritters	. (110	ine)									
Export	as C	SV									
† <u>Family</u>	м	DB	Accession	<u>Score</u>	Mass	<u>Matches</u>	Pep (sig)	<u>Sequences</u>	<u>Seq</u> (sig)	emPAI	Description
1	1	cRAP	@1::TRY1_BOVIN	1597	28266	48	48	7	7	2.34	TRY1_BOVIN
2	1	SwissProt	d2::CP2CT_MOUSE	1307	61433	86	86	13	13	1.47	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
2	2	SwissProt	d2::CP254_MOUSE	535	60887	29	29	10	10	0.87	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
2	3	SwissProt	2::CY250_MOUSE	382	61037	25	25	10	10	0.87	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1
2	4	SwissProt	d2::CP239_MOUSE	293	60932	23	23	5	5	0.33	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
2	5	SwissProt	2::CP238_MOUSE	202	61216	19	19	5	5	0.33	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
2	6	SwissProt	2::CP270_MOUSE	69	61539	5	5	4	4	0.25	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
3	1	SwissProt	2::GRP78_MOUSE	1292	81404	54	54	22	22	2.17	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1
3	2	SwissProt	2::HSP7C_MOUSE	353	78937	23	23	9	9	0.55	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE:
3	3	SwissProt SwissProt	2::HS71L_MOUSE	165 1228	78552 16817	12 48	12 48	4	4	0.25	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1l PE=2
4 5	1	SwissProt	2::CYB5_MOUSE 2::PDIA1_MOUSE	1228	64779	48	48	18	6 18	5.00 1.76	Cytochrome b5 OS=Mus musculus GN=Cyb5a PE=1 SV=2 Protein disulfide-isomerase OS=Mus musculus GN=P4hb PE=1 SV=1
6	1	SwissProt	2::CP1A2_MOUSE	1048	63034	38	38	10	10	1.16	Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1
Z	1	SwissProt	2::RDH7_MOUSE	1023	38455	45	45	10	12	2.50	Retinol dehydrogenase 7 OS=Mus musculus GN=Rdh7 PE=2 SV=1
ž	2	SwissProt	2::H1786_MOUSE	612	38949	23	23	7	7	1.03	Hydroxysteroid 17-beta dehydrogenase 6 OS=Mus musculus GN=Hsd
8	1	SwissProt	2::ENPL_MOUSE	1014	103744	66	66	22	22	1.24	Endoplasmin OS=Mus musculus GN=Hsp90b1 PE=1 SV=2
2	1	SwissProt	2::MGST1_MOUSE	833	18595	25	25	3	3	1.96	Microsomal glutathione S-transferase 1 OS=Mus musculus GN=Mgst1
10	1	SwissProt	2::RL7A_MOUSE	771	35860	28	28	8	8	1.37	60S ribosomal protein L7a OS=Mus musculus GN=Rpl7a PE=2 SV=2
11	1	SwissProt	2::RLA0_MOUSE	758	37215	26	26	8	8	1.09	60S acidic ribosomal protein P0 OS=Mus musculus GN=Rplp0 PE=1 S
12	1	SwissProt	g2::ACSL1_MOUSE	751	86050	41	41	19	19	1.24	Long-chain-fatty-acidCoA ligase 1 OS=Mus musculus GN=Acsl1 PE
12	12	SwissProt	221-ACSUS MOUSE	297	84629	15	15	6	6	0.58	Long-chain-fatty-acidCoà ligase 5 OS=Mus musculus GN=àcelS PF
											S Local intranet
M	45	SCO	T:Very	Larg	ge S	earch	ies			© 2	007-2012 Matrix Science

The Report Builder tab is useful when you need a table of proteins suitable for publication. Lets assume we want to drop the 'one hit wonders' and only report proteins that have significant matches to at least 2 different peptide sequences

			results_2.pl	l?file=F98113	39.dat;_ignoreid	insscorebe	elow=0.05;_prefert	axonomy	=0;_sigthres	hold=0.05;percolate=0;report=0; V 47 X Google
vermarks 🕅	1	*								
* 6	IPRG	2008 SwissProt Mouse (Mascot Search Result:	5)							🟠 + 🖶 + 🔂 Bage + 🎯 T <u>o</u> ols + 🔞
Proteins	: (4	45) Report Builder Unassi	aned (30	0350)						S permalink
	-									
Protein	hit	s (470 proteins)								
Column	s: S	tandard (12 out of 12)								
Filters:										
Filters:	(no									
		Family Protein hits	< ~							
		Family		ſ	Filter					
		Member Database			r inter					
		Accession								
Export a	s CS	Score								
		Mass Num. of matches								
Family	м	Num. of significant matches	Score	Mass	Matches	Pep (sig)	Sequences	Seq	emPAI	Description
		Num, of sequences Num, of significant sequences						<u>(sig)</u>		· · · · · · · · · · · · · · · · · · ·
	1	emPAI dha	1597	28266	48	48	7	7	2.34	TRY1_BOVIN
	1	Description U Fixed modifications	1307	61433	86	86	13	13	1.47	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
	2	Methylthio (C)	535	60887	29	29	10	10	0.87	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
2	3	iTRAQ4plex (K)	382	61037	25	25	10	10	0.87	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1
2	4	iTRAQ4plex (N-term) Variable modifications	293	60932	23	23	5	5	0.33	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
	5	Acetyl (Protein N-term)	202	61216	19	19	5	5	0.33	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
	6	GIn->pyro-Glu (N-term Q)	69	61539	5	5	4	4	0.25	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
1	1	Oxidation (M)	1292	81404	54	54	22	22	2.17	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=
2	2	SwissProt 2::HSP7C_MOUSE	353	78937	23	23	9	9	0.55	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PB
1	3	SwissProt 2::HS71L_MOUSE	165	78552	12	12	4	4	0.25	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1l PE=
	1	SwissProt 2::CYB5_MOUSE	1228	16817	48	48	6	6	5.00	Cytochrome b5 OS=Mus musculus GN=Cyb5a PE=1 SV=2
	1	SwissProt 2::PDIA1_MOUSE	1116	64779	55	55	18	18	1.76	Protein disulfide-isomerase OS=Mus musculus GN=P4hb PE=1 SV=1
	1	SwissProt 2::CP1A2_MOUSE	1048	63034	38	38	10	10	1.16	Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1
z I	1	SwissProt 2::RDH7_MOUSE	1023	38455	45	45	12	12	2.50	Retinol dehydrogenase 7 OS=Mus musculus GN=Rdh7 PE=2 SV=1
z I	2	SwissProt 2::H17B6_MOUSE	612	38949	23	23	7	7	1.03	Hydroxysteroid 17-beta dehydrogenase 6 OS=Mus musculus GN=Hs
	1	SwissProt 2::ENPL_MOUSE	1014	103744	66	66	22	22	1.24	Endoplasmin OS=Mus musculus GN=Hsp90b1 PE=1 SV=2
					·				·	
										Local intranet
										A CANDON
		SCOT : Very								

We open up the filters section and add a suitable filter.

C iPRG200	8 Sw	issProt Mous	e (Mascot Search Result	s) - Wind	ows Inter	net Explorer					
00-	0	http://bogong/r	nascot_2_4_0_64/cgi/master_	results_2.pl	?file=F9811:	39.dat;_ignoreic	nsscoreb	elow=0.05;_prefer	axonomy	=0;_sigthres	shold=0.05;percolate=0;report=0; 🛩 47 🗙 Google
Powermarks		۱ %									
* * 6		52008 SwissProt	Mouse (Mascot Search Result	0							🐴 + 📾 + 🔂 Page + 🎯 Tools + 🔞 -
				~/							
Protei	ns (4	45) Rep	ort Builder Unassi	gned (30	<u>350)</u>						<u>S permalink</u>
D		(001									
		ts (231 pi									
Colum	ns: §	Standard (1	12 out of 12)								
Filters	. "Ni	um, of signi	ificant sequences" >	= 2							
7 millions		ann. or sign	incunt sequences >	- 2							
Export	as C	S∨									
							-	1			
† <u>Family</u>	M	<u>DB</u>	Accession	<u>Score</u>	Mass	Matches	Pep (sig)	Sequences	<u>Seq</u> (sig)	<u>emPAI</u>	Description
1	1	CRAP	@1::TRY1_BOVIN	1597	28266	48	48	7	7	2.34	TRY1_BOVIN
2	1	SwissProt	d2::CP2CT_MOUSE	1307	61433	86	86	13	13	1.47	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
2	2	SwissProt	d2::CP254_MOUSE	535	60887	29	29	10	10	0.87	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
2	3	SwissProt	d2::CY250_MOUSE	382	61037	25	25	10	10	0.87	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1
2	4	SwissProt	d2::CP239_MOUSE	293	60932	23	23	5	5	0.33	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
2	5	SwissProt	ef2::CP238_MOUSE	202	61216	19	19	5	5	0.33	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
2	6	SwissProt	d2::CP270_MOUSE	69	61539	5	5	4	4	0.25	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
3	1	SwissProt	2::GRP78_MOUSE	1292	81404	54	54	22	22	2.17	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1
3	2	SwissProt	2::HSP7C_MOUSE	353	78937	23	23	9	9	0.55	
3	3	SwissProt	2::HS71L_MOUSE	165	78552	12	12	4	4	0.25	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1l PE=2
4	1	SwissProt	Z::CYB5_MOUSE	1228	16817	48	48	6	6	5.00	
<u>5</u>	1	SwissProt	2::PDIA1_MOUSE	1116	64779	55	55	18	18	1.76	
<u>6</u>	1	SwissProt	2::CP1A2_MOUSE	1048	63034	38	38	10	10	1.16	Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1
Z	1	SwissProt	2::RDH7_MOUSE	1023	38455	45	45	12	12	2.50	, , ,
Z	2	SwissProt	2::H17B6_MOUSE	612	38949	23	23	7	7	1.03	
8	1	SwissProt	2::ENPL_MOUSE	1014	103744	66	66	22	22	1.24	
2	1	SwissProt	2::MGST1_MOUSE	833	18595	25	25	3	3	1.96	Microsomal glutathione S-transferase 1 OS=Mus musculus GN=Mgst1
10	1	SwissProt	2::RL7A_MOUSE	771	35860	28	28	8	8	1.37	60S ribosomal protein L7a OS=Mus musculus GN=Rpl7a PE=2 SV=2
11	1	SwissProt SwissProt	2::RLA0_MOUSE 2::ACSL1_MOUSE	758 751	37215 86050	26	26	8	8 19	1.09	
<u>12</u> 12	2		2::ACSL1_MOUSE	751 297	84629	41	41 15	19	19	1.24	Long-chain-fatty-acidCoA ligase 1 OS=Mus musculus GN=Acsl1 PE Long-chain-fatty-acidCoA ligase 5 OS=Mus musculus GN=Acsl5 PE
<	1	CowrosPffill	DECOMPOSES MOOSE		64024		15				TO DECEMBER 10 CONTRACT OF A DECEMBER 10 SCHOOL SCH
Done											Local intranet
M	45	SCO	T:Very	Larg	ge S	earch	ies			© 2	007-2012 Matrix Science

Only proteins with significant matches to at least 2 sequences remain. The filtering is very flexible, with lots of useful terms.

6 iPRG2008	-		e (Mascot Search Result ascot 2 4 0 64/cgi/master	· ·			nsscoreb	elow=0.05; prefer	taxonomy	=0: sigthres	hold=0.05;percolate=0;report=0; V 47 X Google
Powermarks 🕅										.,	
			Mouse (Mascot Search Result	s)							🏠 🔹 💀 🔹 🚱 Page 🔹 🛞 Tools 🔹 🚷
Protein	- //		ort Builder								C normalial
Protein	5 (4	ној кер	ort Builder	igned (30	1350)						<u>& permalink</u>
Protein	hit	s (230 pr	oteins)								
Column	ns: 8	tandard (1	2 out of 12)								
Filters:	(N0	DT(Databas	e is cRAP) AND "Nur	m. of sig	nificant	sequences	">= 2	:)			
Export a	as CS	SV									
† <u>Family</u>	м	DB	Accession	<u>Score</u>	Mass	Matches	Pep (sia)	Sequences	<u>Seq</u> (siq)	emPAI	Description
2	1	SwissProt	d2::CP2CT_MOUSE	1307	61433	86	86	13	13	1.47	Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
2	2	SwissProt	2::CP254_MOUSE	535	60887	29	29	10	10	0.87	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
2	3	SwissProt	d2::CY250_MOUSE	382	61037	25	25	10	10	0.87	Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1
2	4	SwissProt	d2::CP239_MOUSE	293	60932	23	23	5	5	0.33	Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
2	5	SwissProt	d2::CP238_MOUSE	202	61216	19	19	5	5	0.33	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
2	6	SwissProt	2::CP270_MOUSE	69	61539	5	5	4	4	0.25	Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
3	1	SwissProt	d2::GRP78_MOUSE	1292	81404	54	54	22	22	2.17	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=
3	2	SwissProt	d'2::HSP7C_MOUSE	353	78937	23	23	9	9	0.55	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE
3	3	SwissProt		165	78552	12	12	4	4	0.25	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa1l PE=2
4	1	SwissProt	2::CYB5_MOUSE	1228	16817	48	48	6	6	5.00	Cytochrome b5 OS=Mus musculus GN=Cyb5a PE=1 SV=2
5	1	SwissProt	2::PDIA1_MOUSE	1116	64779	55	55	18	18	1.76	Protein disulfide-isomerase OS=Mus musculus GN=P4hb PE=1 SV=1
5	1	SwissProt		1048	63034	38	38	10	10	1.16	Cytochrome P450 1A2 OS=Mus musculus GN=Cyp1a2 PE=1 SV=1
z	1	SwissProt	2::RDH7_MOUSE	1023	38455	45	45	12	12	2.50	Retinol dehydrogenase 7 OS=Mus musculus GN=Rdh7 PE=2 SV=1
Z B	2	SwissProt SwissProt	2::H17B6_MOUSE	612 1014	38949 103744	23 66	23 66	7	7	1.03	Hydroxysteroid 17-beta dehydrogenase 6 OS=Mus musculus GN=Hsu
1	1	SwissProt		833	103744	25	25	22	22	1.24	Endoplasmin OS=Mus musculus GN=Hsp90b1 PE=1 SV=2 Microsomal glutathione S-transferase 1 OS=Mus musculus GN=Mgst
2 10	1	SwissProt	2::MGST1_MOUSE	771	35860	25	25	3	3	1.96	60S ribosomal glutathione S-transferase 1 0S=Mus musculus GN=Mgst 60S ribosomal protein L7a OS=Mus musculus GN=Rpl7a PE=2 SV=2
1	1	SwissProt		758	37215	20	26	8	8	1.09	60S acidic ribosomal protein P0 OS=Mus musculus GN=Rplp0 PE=1 S
2	1	SwissProt		751	86050	41	41	19	19	1.09	Long-chain-fatty-acidCoA ligase 1 OS=Mus musculus GN=Acsl1 P
2	2	SwissProt	22::ACSL5_MOUSE	297	84629	15	15	6	6	0.28	Long-chain-fatty-acidCoA ligase 5 OS=Mus musculus GN=AcsIS PI
3	17	SwieeDrot	20-DI 13 MOURE	749	28083	21	21	7	7	1 65	
e						11					S Local intranet
MA	13	SCO	T:Very	Larg	ge S	earch	ies			© 2	007-2012 Matrix Science

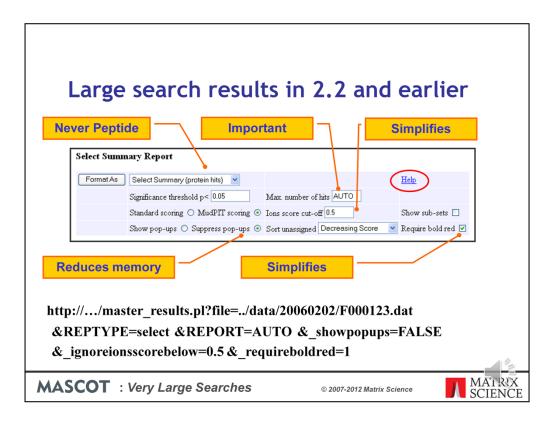
Another thing that you could easily do would be to exclude proteins from the contaminants database

🖉 iPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer		
A the second se	-0: sigthreshold=0.05:percolate=0:report=0;	
Powermarks MA %		
🛠 🖗 🌈 PRG2008 SwissProt Mouse (Mascot Search Results)		🐴 - 📾 - 🔂 Page - 🎯 Tools - 🔞-
Proteins (445) Report Builder Unassigned (30350) Protein hits (230 proteins)		<u>s permalink</u>
▼Columns: Standard (12 out of 12)		
© Saved arrangement Standard (default) V Load Make default © Custom Enabled Available		
Enabled Available Family Member Database Accession Score Num. of significant matches Num. of significant sequences Mass emPAl Image: Content of the sequence sequences Description Image: Content of the sequence seq	l≩	
Save arrangement as Knew> Show column string		
Filters: (NOT(Database is cRAP) AND "Num. of significant sequences" >= 2) Export as CSV		×
2		Local intranet
MASCOT : Very Large Searches	© 2007-2012 Matrix Science	MAL RX SCIENCE

The columns section of Report Manager allows you to choose which columns to include and, if required, change their order

	A28	-	= Filters:									
	A	В	C	D	E	F	G	Н		J	K	
	Preferre Show P	All entries	s									
26	Show P	no										
	Filtore:	Num of s	significant seque	2 = 2 990								
29	i intero.	indini. or a	Significant seque	1003 - 2								
	Family	Member	Database	Accession	Score	Mass	Num. of	Num. of	Num. of	Num. of	emPAI	Description
30							matches	significant matches	sequences	significant sequences		
31	1	1	iPRG 2012	P00925	2140	46942	148	100	53	43	44.71	Enolase 2 OS=Saccharomyces cere
32		2	iPRG_2012	P00924	1059	46844	71	46	35	27	7.47	Enolase 1 OS=Saccharomyces cere
33		1	iPRG 2012	P00549	1933	54909	133	87	56	43	18.28	Pyruvate kinase 1 OS=Saccharomyc
34		1	iPRG_2012	P40150	1613	66668	105	66	66	45	11.76	Heat shock protein SSB2 OS=Sacch
35	3	2	iPRG_2012	P11484	1590	66732	103	65	64	44	11.12	Heat shock protein SSB1 OS=Sacch
36	4	1	iPRG_2012	P10592	1591	69599	107	57	52	32	5.01	Heat shock protein SSA2 OS=Sacch
37		2	iPRG_2012	P10591	1161	69786	85	44	48	26	3.02	Heat shock protein SSA1 OS=Sacch
38		3	iPRG_2012	P16474	233	74479	23	8	17	6	0.32	78 kDa glucose-regulated protein hor
39		1	iPRG_2012	P00330	1453	37282		51	32	25	13.48	Alcohol dehydrogenase 1 OS=Sacch
40		2	iPRG_2012	P07246	101	40743		5	7	3	0.29	Alcohol dehydrogenase 3, mitochonc
41		1	iPRG_2012	P00560	1382	44768	102	58	54	33	12.75	Phosphoglycerate kinase OS=Sacch
42		1	iPRG_2012	P00359	1361	35838	76	54	31	25	12.29	Glyceraldehyde-3-phosphate dehydro
43 44		2	iPRG_2012	P00358 P00360	1242 505	35938 35842	69	48 20	29 14	24 12	9.89 2.47	Glyceraldehyde-3-phosphate dehydro
44		4	iPRG_2012 iPRG_2012	P00360 P04406	41	35842	4	20	4	2	0.21	Glyceraldehyde-3-phosphate dehydro Glyceraldehyde-3-phosphate dehydro
45		1	iPRG 2012	P04408	1289	61685		41	28	26	4.7	Pyruvate decarboxylase isozyme 1 C
40		1	iPRG 2012	P00950	1031	27592		44	32	25	34.97	Phosphoglycerate mutase 1 OS=Sac
48		1	iPRG_2012	P07281	1015	15881	51	38	16	13	22.71	40S ribosomal protein S19-B OS=Sa
49		2	iPRG 2012	P07280	1014	15907		38	16	13	22.71	40S ribosomal protein S19-A OS=Sa
50	11	1	contaminants	P00761	922	25078	37	27	7	6	2.89	SWISS-PROT: P00761 [TRYP PIG Tr
51	12	1	iPRG_2012	P32324	784	93686	49	33	33	23	1.44	Elongation factor 2 OS=Saccharomy
52		1	iPRG_2012	P16521	771	116727		33	47	30	1.52	Elongation factor 3A OS=Saccharom
53		1	iPRG_2012	P05319	765	10739	38	29	10	9	95.65	60S acidic ribosomal protein P2-alph
54		1	iPRG_2012	Q03048	721	15948	28	23	17	14	17.82	Cofilin OS=Saccharomyces cerevisis
55		1	iPRG_2012	P0C0V8	719	9797	42	29	15	12	207.43	40S ribosomal protein S21-A OS=Sa
56	16	2	iPRG_2012	Q3E754	694	9811	41	28	15	12	148.28	40S ribosomal protein S21-B OS=Sa
4 4	P	data_	_20120501_F00	1467_dat_	rf/				1			
Dra	w - <table-cell></table-cell>	(i) Autos	Shapes 🕶 🔪 ` 🔌		4 🙎	۵ - 🛓	<u>∕</u> - <u>A</u> - ≡	■ ☴ ■	-			
Rea	dy											
ea	dy											

Once the list is filtered and the columns arranged as required, there is a button to export the table as CSV, which can be pasted into Excel and formatted to create a suitable figure for dropping into a publication



If you are still using Mascot 2.2 or if you have some application software that requires the results in the earlier format, and you are encountering problems with timeouts and running out of memory, here are some tips:

•Ensure you are using the Select report. If you are using a third party client that has specified Peptide summary or Protein summary, add this to the URL before opening the file: &REPTYPE=select

•Don't specify a huge number of hits 'just in case'. Choose AUTO to display all protein hits that contain at least one significant peptide match: &REPORT=AUTO

•Get rid of the yellow pop-ups: &_showpopups=FALSE

•Setting require bold red and an expect value cut-off will minimise the number of hits: &_ignoreionsscorebelow=0.5&_requireboldred=1

Note that the ions score cut-off is as score threshold when the value is 1 or greater. When the value is between 0 and 1, it is an expect threshold, which is often much more useful. I often set this to 0.5 to get rid of all the junk matches.

🕽 Back 🔹 🕥 👻 😰 🖠	🏠 🔎 Search 🛛 📩 Favorites	🤣 🔊 🖏 🗹 - 🖵 🕄	1	Powermarks 🎼 🦽 🆄
m	aster_results.pl			
	URL	mascot.dat	Value	Description
			peptide	Peptide Summary
			archive	Archive Report
n	eptype		concise	Concise Protein Summary
			protein	Full Protein Summary
			select	Select Summary (hits)
_			unassigned	
re	eport		auto	Report all significant hits
_			N	Report N hits
-	showsubsets	ShowSubSets	1	Set value to 1 to report Peptide Summary hits that match a subset of peptides. Default is 0.
-	requireboldred	RequireBoldRed	1	Set value to 1 to report Peptide Summary hits only if they contain at least one "bold red" peptide. Default is 0.
-	showallfromerrortolerant	ShowAllFromErrorTolerant	1	Set value to 1 to report all hits from an error tolerant search, including the garbage. Default is 0.
-	sigthreshold	SigThreshold	N	Probability to use for the significance threshold. Range is 0.1 to 1E-18. Default is 0.05.
			scoredown	Sort unassigned matches by descending score, (default)
-	sortunassigned	SortUnassigned	queryup	Sort unassigned matches by ascending query number
			intdown	Sort unassigned matches by descending intensity
-	ignoreionsscorebelow	IgnoreIonsScoreBelow	N	Any ions scores below this value are set to 0. Floating point number, default 0.0.
	showpopups		true	Show top 10 peptide matches fro each query in JavaScript pop-up, (default)
			false	Suppress JavaScript pop-ups.
-	alwaysgettitle		1	Set to 1 to force reports to fetch Fasta titles from database when they are not included in the result file. Default is 0.
-	mudpit	Mudpit	N	Number of queries at which protein score calculation switches to large search mode. Default 1000
				Local intranet

If you can't remember these URL parameters, just click on the help link

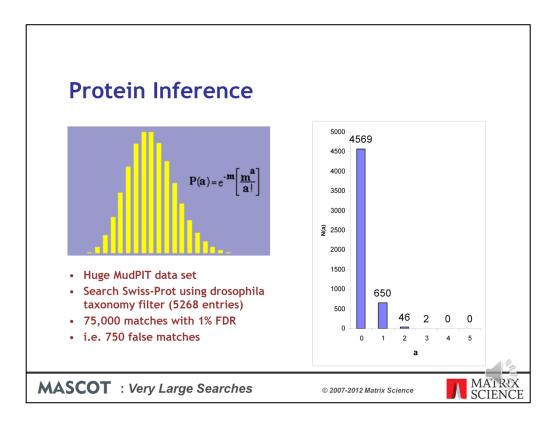
Reporting large searc	h results	
Select Summary Report FormatAs Select Summary (protein hits) Significance threshold p<0.05 Max. r Standard scoring MudPIT scoring Ions so Show pop-ups Suppress pop-ups Sort units		
MASCOT : Very Large Searches	© 2007-2012 Matrix Science	A'I RIX CIENCE

What do we mean by Standard scoring and MudPIT scoring?

Prote	ein Sco	ores	for	٨٨	571		с с	0	arc	hos	
FIUL		7 62	101	///	ווכ	V1.	כו	CC		.1162	
Standar	d protein	score									
• the	e sum of the	e ions sc	ores								
• 620	cluding the	scores f	or du	olicat	te n	hat	ches	w	hich	are shown	in
		3001031	or uu	Juca		au	ciics	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	nen		
•	rentheses										
• cor	rection to	reduce t	he co	ntrib	outio	on d	of lov	w-so	corir	ng random	
	tches									5	
1110	cerres										
342.	2::IPI00023283	Mass: 38328	303 Scor	: 181	Mato	hes:	51(0)	Seme	nces: 4	18 (0)	
	Tax_Id=9606 Gene										
	Query Observed	Mr(expt)	Mr(calc)	ppm 1	Miss S	core	Expect	Rank 1	Unique	Peptide	
	28 359.7341		717.4537		0	7	4.2	5	U	R.LFAIVR.G	
	209 394.2371		786.4599		0	8	13	3	U	K.LTIADVR.A	
	334 411.2073	820.4000	820.3954	5.61	0	3	15	4	U	K.TDSGLYR.C	
	357 413.2642		824.5135	0.48	1	12	1.1	5	U	K.RFLTLR.K	
	715 450.7365		899.4588		0	10		2	U	K.IVDVSSDR.C	
	740 451.7681		901.5233		0	3	24	3	U	R.VTLVDVTR.N	
	840 459.2484		916.4767	5.98	0	2	29	2	U	K.GVEFNVPR.L	
	844 459.7299		917.4454		0	4		6	U	K.ELEETAAR.M	
	1029 473.2757		944.5331	3.97	1	3		3	U	R.EPPSFIKK.I	
	1058 475.7505		949.4869	-0.47	0	4	22	5	U	R.SSVSLSWGK.P	
	1066 476.2790	950.5433	950.5425	0.94	0	1	23	4	ប	R.PLTDLQVR.E	12
MASCOT	: Very Lar	ge Sear	ches				© 2007-:	2012 I	Matrix S	Science	MAT RIX SCIENCE

With standard peptide summary scoring, the protein score is essentially the sum of the ions scores of all the peptides assigned to the protein. Where there are duplicate matches to the same peptide, the highest scoring match is used. A correction is applied based on the number of candidate peptides that were tested. This correction is very small unless it is a very large protein, like here, or a no-enzyme search

Despite this correction, as this example shows, when we have many low scoring matches assigned to the same protein, we can still get a high protein score, even though none of the individual peptide matches are significant



A protein with matches to just a single peptide sequence is commonly referred to as a "one-hit wonder" and is often treated as suspect. This is actually a slight oversimplification. In a search with a large number of spectra and a small database, even though the peptide false discovery rate is low, a protein can pick up multiple false matches by chance. This is easily calculated using a Poisson Distribution, where m is the average number of false matches per protein. In this example, m is 750/5268, and we would expect 650 database entries to be one-hit wonders. However, 46 entries will pick up two false matches and 2 entries will pick up three, which could mean we report 48 false proteins.

The problem isn't limited to large searches. It is the ratio between the number of spectra and the number of entries in the database that matters. So, a small search against a small database can give similar numbers

Drotoir			. f		C	/	c c		~ ~	chac	
Proteir	1 20	ores	S TOI	N	I)		23	be	ar	cnes	
MudPIT p	rotei	n sco	ore								
				of th	o i	ons	scor		ver	the ident	ity or
	logy th							0	ver	the lucht	
	07					uci.	У				
• Plus 1	x the	avera	ge thr	esno	δια						
1249. <u>2::IPI</u>		Mass: 383				tches:	51(0)	Seque	nces:	48 (0)	
Tax_10	=9606 Gene_ Observed	Mr(expt)	Mr(calc)			Score	Expect	Rank 1	lnime	Peptide	
28	359.7341	717.4537	717.4537	-0.09	0	7	4.2	5	U	R.LFAIVR.G	
209	394.2371	786.4596	786.4599	-0.46	0	8	13	3	U	K.LTIADVR.A	
334	411.2073	820.4000	820.3954	5.61	0	3	15	4	U	K.TDSGLYR.C	
357	413.2642	824.5139	824.5135	0.48	1	12	1.1	5	U	K.RFLTLR.K	
715	450.7365	899.4584	899.4588	-0.38	0	10	2.9	2	U	K.IVDVSSDR.C	
740	451.7681	901.5217	901.5233	-1.72	0	3	24	3	U	R.VTLVDVTR.N	
840	459.2484	916.4821	916.4767	5.98	0	2	29	2	U	K.GVEFNVPR.L	
844	459.7299	917.4452	917.4454	-0.24	0	4	15	6	U	K.ELEETAAR.M	
1029	473.2757	944.5368	944.5331	3.97	1	3	21	3	U	R.EPPSFIKK.I	
1058	475.7505	949.4864	949.4869	-0.47	0	4	22	5	U	R.SSVSLSWGK.P	
1066	476.2790	950.5433	950.5425	0.94	0	1	23	4	ប	R.PLTDLQVR.E	12
MASCOT : Ve	ery Lar	ge Sea	arches				© 200	7-2012	2 Matrix	Science	MATRX SCIENCE

To avoid this problem, we use MudPIT protein scoring, in which the score for each peptide match is not its absolute score, but the amount that it is above the threshold. Therefore, matches with a score below the threshold do not contribute to the score. The MudPIT protein score is the sum of the score excess over threshold for each of the matching peptides plus one times the average threshold. For each peptide, the "threshold" is the homology threshold if it exists, otherwise it is the identity threshold.

So, even though a large protein like titin may pick up several random matches, with MudPIT scoring, the protein score is zero, so you don't see it listed in the report unless you specify a huge number of protein hits, as was done here to capture this screen shot.

By default, MudPIT protein scoring is used when the ratio between the number of queries and the number of database entries, (after any taxonomy filter), exceeds 0.001. This default switching point can be moved by changing the value of MudpitSwitch in mascot.dat. You can also switch between the two scoring methods by using the format controls at the top of the report.



At some stage, it is likely that you will want to export the search results to another application or a relational database. If you want to write your own code, we provide a free library called Mascot Parser that provides a clean, object oriented programming interface to the result file. The supported languages are C++, Java, and Perl.

Mascot also includes a flexible export utility.

If you want the XML format, you probably know that this is what you want. If you've no idea what XML is, chances are you don't want it.

Choose CSV if you want to export to Excel - I'll show an example in a moment.

Choose pepXML if you want to export to Protein Prophet from ISB.

mzIdentML is the new, standard format from PSI for search result interchange. Mascot provides a very full implementation of mzIdentML and this is the one to choose if you are writing new application software that will use Mascot results

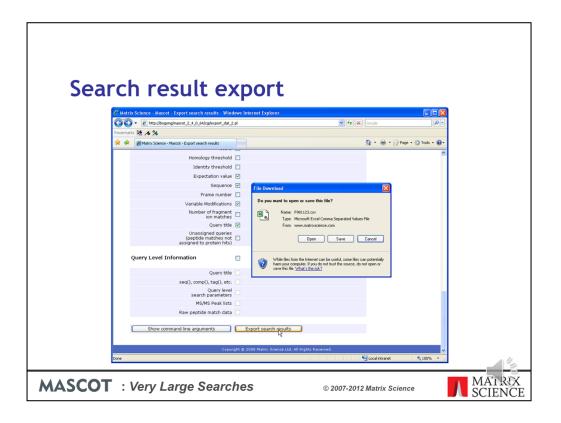
DTASelect is the tab separated format used by David Tabb's DTASelect program

The Mascot DAT file is the raw result file. If you need the result file for some reason, and don't have FTP or SCP access to your Mascot server, this is a convenient way to get the file.

MGF peak list is useful when you have the search result but can't find the peak list.

Search result e	export		
cur en result (
C Matrix Science - Mascot - Export search results	Windows Internet Explorer		
A the second se	t_dat_2.pl?group_family=18_showsubsets=18file=.	%2Fdata%2FF98 🗹 🐓 🗙 Google	
Powermarks 🎼 🔥 %			
🙀 🕸 🌈 Matrix Science - Mascot - Export search results		🏠 👻 🖶 👻 🕞 Page 👻 🎧 Tools 👻 🚱 ۳	
(MATRIX) (SCIENCE)		HOME I MASCOT I HELP Search GO	
Mascot > Export search results		Logged in as jcottrell Edit Logout	
Export search results		Help	
Export fo	ormat XML 💌		
Significance thresho			
Ions score c	ut-off mzIdentML DTASelect		
Threshold	type Mascot DAT File ogy		
Max. number o	f hits AUTO		
	oring Standard O MudPIT 💿		
Include same-set protei (additional proteins that the same set of pept	span 🔲 ides)		
Include sub-set protei (additional proteins that a sub-set of pept	span 1		
Group protein fai	milies 🔽		
Require bol			
Show Percolator s			
Preferred Taxon * Occasionally requires information to be retrieved from external	omy* All entries	M	
A A A A A A A A A A A A A A A A A A A	and the second sec		
Search Information		×	
Done		Secal Intranet 🔍 100% 👻	
			_
T: Very Large Sea	rches	© 2007-2012 Matrix Science	MÂT R SCIEN

If you arrive here from one of the older reports, to begin with, you may need to select the required output format. Different formats have different options further down the page

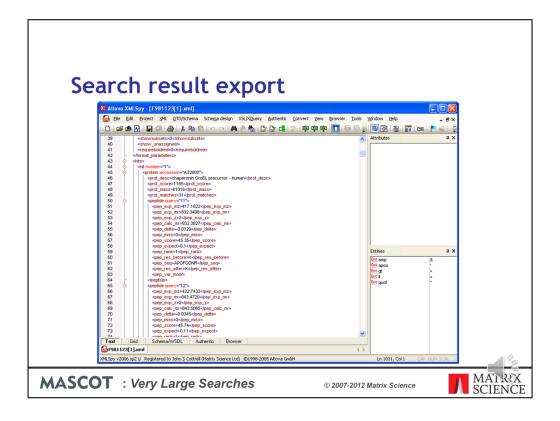


To export to Excel, simply select CSV as the format, and click on the Export Search Results button at the bottom of the page.

You can then click on the Open button to open it into Excel:

ard												
	h r	esι	ilt (ex	oor	t						
Microsoft E												
🖲 Ele Edit V	-				-	21 40-						. a ×
	y 🗋 🗳	👗 🖻 🛍	1 🖉 1 0 -	연 🖌 👹				- 🛛 - 🖯	2 🖬 🔁			
Arial	- 10	- B /	<u>u</u> ≣ :	■ = B	₩% ,	•.0 0.↓ 00.	律律 日	🗄 • 🤷 • 🖌	<u> </u>			
A1	-	=	-	-	-	-				14		_
A 35	В	C	D	E	F	G	Н		J	K	L	-
36 Significand	0.05	i										
37 Max. numb												
38 Use MudP	0	I.										
39 Ions score	0	1										
40 Include sa												
41 Include sul		1										
42 Include una												
43 Require bo) C	1										
45 Protein hit:												
46												
47 prot_hit_nu	prot_acc	prot_desc	prot_score	prot_mass	prot_matcH	pep_query	pep_exp_r	pep_exp_r	pep_exp_z	pep_calc_	pep_delta	pep
	A32800	chaperonir	1195	61016	31		417.1822			832.3827		
49 1							422.7433			843.5065		
50 1 51 1							430.7328 451.2499			859.4837 900.528		_
52 1							451.2499			911.5803		
53 1							480.7447			959.5036		
54 1							595.7855			1189.601		
55 1						25		1205.529		1205.596		
	v]F08112	3[1]/				26	608 3099	1214 605	2	1214.651	-0.0454	
					B #			0				
Draw - 🔓 🕝	AutoShap	oes ▼ ∖ ¥		4	× - <u>≁</u> -	<u> ∼</u> =	···· 🛱 🗖	<i>i</i> -				
Ready												

Much easier and safer than "screen scraping"



For those of you into XML, here is a sample XML file. The schema is available from our web site or your local Mascot installation.

Please read the help for details.

arch	ro	C I	ult o	vn		rt					
arch		σι				I L					
Microsoft Acces	:s										
_	Insert Format	Record	s Tools Windo	w Help						importing xml sch	
	j. Balt - 1 g illar 6 []a. ♥ [] & R	-			44	▶ ▼ ◎ 加					
	/⊡. ∨ & ч		** 🕲 Z 🕯 🗡	• 🏷 🖽 🗸	948		-ψ-	_	_		
F 981123	: Database (Acc	cess 20									
Gilon Mr	esign 🏪 New 🕽	~ 0	NT. 8-9- (000								
u∰ Oben ₩ 1	Sezidu . Will Mem 1	∧ <u>-</u> ₽	57 EE 100								
🔲 peptide : Tabl	le										
pep_exp_mz			pep_calc_mr		pep		pep_expect	pep_			pep 🔺
417.1822	832.3498	2	832.3827	-0.0329	0	45.35	0.1	1	ĸ	APGFGDNR	К
451.2499	900.4853	2	900.5280	-0.0427	0	51.95	0.025	1	K	LSDGVAVLK	V
456.7806	911.5467	2	911.5803	-0.0337	0	59	0.0041	1	K	VGLQVVAVK	A
480.7447 595.7855	959.4748 1189.5565	2	959.5036 1189.6012	-0.0288	0	45.33 56.55	0.11 0.0068	1	R	VTDALNATR EIGNIISDAMK	A K
603.7720	1205.5294	2	1205.5961	-0.0447	0	50.55	0.0066	1	ĸ	EIGNIISDAMK	ĸ
608.3099	1205.5254	2	1214.6506	-0.0666	0	73.21	0.00015	1	ĸ	NAGVEGSLIVEK	
617.2857	1232.5569	2	1232.5884	-0.0315	0	80.63	2.7e-05	1	K	VGGTSDVEVNEK	ĸ
672.8375	1343.6605		1343.7085	-0.0480	0	64.38	0.001	1	R	TVIIEQSWGSPK	V
714.8884	1427.7623	2	1427.8057	-0.0434	0	64.52	0.00086	1	R	GVMLAVDAVIAELK	ĸ
714.8938	1427.7730	2	1427.8057	-0.0327	0	72.61	0.00013	1	R	GVMLAVDAVIAELK	К
722.8849	1443.7552	2	1443.8006	-0.0454	0	72.71	0.00014	1	R	GVMLAVDAVIAELK	К
722.8934	1443.7722	2	1443.8006	-0.0284	0	70.08	0.00025	1	R	GVMLAVDAVIAELK	К
752.8643	1503.7141	2	1503.7490	-0.0349	0	89.56	2.7e-06		K	TLNDELEIIEGMK	F
760.8461	1519.6777	2	1519.7439	-0.0662	0	84.43	8.9e-06	1	K	TLNDELEIIEGMK	F
640.3281	1917.9625	3	1918.0636	-0.1010	0	101.5	1.3e-07	1	K	ISSIQSIVPALEIANAHR	K
960.0327	1918.0509 2037.0067	2	1918.0636 2037.0153	-0.0127	0	87.34 52.42	3.2e-06 0.01	1	K R	ISSIQSIVPALEIANAHR IQEIIEQLDVTTSEYEK	E
1019.5106	2037.0067	2	2037.0153	-0.0086	0	52.42	0.01 4.6e-09	1	R	ALMLQGVDLLADAVAVTMGP	
1065.0399	2128.0653	2	2128.1271	-0.0618	0	68.73	0.00022	1	R	ALMLQGVDLLADAVAVTMGP	
1073.0477	2144.0809	2	2144.1220	-0.0411	0	69.64	0.00018	1	R	ALMLQGVDLLADAVAVTMGP	
789.1062	2364.2968	3	2364.3263	-0.0296	0	55.53	0.0038	1	R	KPLVIAEDVDGEALSTLVLNR	L
1183.1570	2364.2994	2	2364.3263	-0.0269	0	65.46	0.00038	1	R	KPLVIAEDVDGEALSTLVLNR	L
789.1094	2364.3063	3	2364.3263	-0.0200	0	94.59	4.5e-07	1	R	KPLVIAEDVDGEALSTLVLNR	L _1
Record: 14 4	40	2	0.491 30.41 of 40	0.0103	0	17 53	0.02	1	D	TALLDAAQVASLITTAE\AA/TE	
				<u> </u>	-			_	-		
Datasheet View											

XML is ideal for transferring the results to a relational database. Even Microsoft Access can open the XML file directly into database tables

Searc	h re	الناءم	lt exp	ort							
_									7		
C		· · ·	m/help/export_help.html	ner-explored			Google	ا ا			
Pow	ermarks 隆 🔥 %										
*	🔅 🗽 Matrix Scien	ce - Help - Export sea	arch results				<u>6</u> • ∉	• 🕞 Bage • 🍈 Tools • 🔞	-		
Ł	MATRIX) SCIENCEJ		HOME	WHAT'S NEW I MAS	COT HELP PRODUCTS	SUPPORT	RAINING	Search Go	^		
	Help > Export s	earch results									
		Export se	arch results								
	in this page ustom XML and CSV epXML hzIdentML TASelect lascot DAT File IGF Peak List pitional Protein Hit	This usible mables flactor dearth resides to be expected in a velocity of "machine read-bine" format. When used instructively, the finance of excession and outcomised on any and becomes from, docalage of buckness Excession of the second second second second and a second second second second second second second second second second scripts, with the options specified on the command line. Custom XML and CSV The information contained in these two formats is identical. XML is ideal for importing into a relational database. CSV can be oppined in preadheres such a Microsoft Excel.									
C	ormation ommand Line xecution ML Schema	report. For sea peptide matche	lass Fingerprint, the resu rch results that include M is in a similar way to a Pe ation equivalent to a par	S/MS data, you optide Summary	can choose whethe report or a Protein	Family rep	ure the protein lis ort. To create an	t and associated export that			
		Type of search	HTML Report	Threshold type	Protein Scoring	Same- sets	Sub-sets	Group proteins			
		PMF	Concise Protein Summary	N/A	N/A	checked	1	N/A			
		MS/MS	Peptide Summary	Identity	As format controls	checked	As format controls	not checked			
		MS/MS	Protein Family Report	Homology	MudPIT	checked	1	checked			
		schema. The so additional data schema has be	for individual data items, hema introduced with Ma structures for Mascot 2.1 en created: mascot_sea her down this page. Doc	ascot 2.1 is mas 2, including quar ch_results_2.xs	ot_search_results ititation results, wo	_1.xsd, (do build have b). For gener	cumentation). Th roken this schem	e need to add a, so a new	×		
Done	3						🍙 😜 Internet	at 100% 🔹			
SCOT	· Verv I	arge	Searches			0.0007	2012 Matrix	Colonaa	MA		

There is a very detailed help page for all of this.

Search result exp	ort		
C Matrix Science - Mascot - Export search results - Win	dows Internet Explorer		
Co Co + L http://www.matricscience.com/cgi/export_dat_2.		IS.dat8RE V + X Google	ρ - I
File Edit View Favorites Tools Help	🛱 Snagit		
🚖 🕸 🗽 Matrix Science - Mascot - Export search results		🚰 • 📾 • 🔂 Page • 🎯 Tools •	0 -
Homology threshold			~
Identity threshold			
Expectation value			
Sequence			
Frame number			
Variable Modifications			
Number of fragment ion matches			
Query title	V		
Unassigned queries (peptide matched assigned to protein hits)			
Query Level Information			
Query title			
seq(), comp(), tag(), etc.			
Query level search parameters			
MS/MS Peak lists			
Raw peptide match data			
Show command line arguments	Export search results		
· · · · · · · · · · · · · · · · · · ·			
	© 2008 Matrix Science Ltd. All Rights		V
Done		🏹 🚱 Internet 🗮 100%	- 1/2
ASCOT : Very Large Searche	s	© 2007-2012 Matrix Science	MA'I R'X SCIENCI

Which describes how the export script can be called from the command line or a shell prompt, as part of an automated pipeline.

I won't go into any detail here, but this means that it is possible to set up a script that will, for example, automatically convert all of your Mascot results to XML files.

Figuring out the command line arguments from the help can be tricky so, in Mascot 2.3, we added a function to display the command line corresponding to the selected options



By the way, don't delete the original result files after exporting them or your won't be able to view the standard Mascot reports in a browser.