

New

The Protein Family Result Report

MASCOT : *Protein Family Result Report*

© 2010 Matrix Science

MATRIX
SCIENCE

Room for Improvement

Reporting large searches

- HTTP timeouts
- Browser hanging
- Out of memory

Protein Inference

- Need to use ions score cut-off and 'Require bold red' to create minimal list
- Proteins related by shared peptide matches could be widely separated in the report

There were two areas that we really felt needed improvement in the result reports in Mascot 2.2 and earlier.

First, although you could run very large searches on any supported platform, even 32-bit and with limited amounts of memory, trying to display the result report would often lead to problems due to an HTTP timeout or the web browser becoming slow to respond or hanging completely or simply running out of memory on either the server or the client.

Second, protein inference was not smart enough, and left too much of the hard work to the user. To create a minimal list of proteins, you had to tweak the score cut-off and use 'Require bold red'

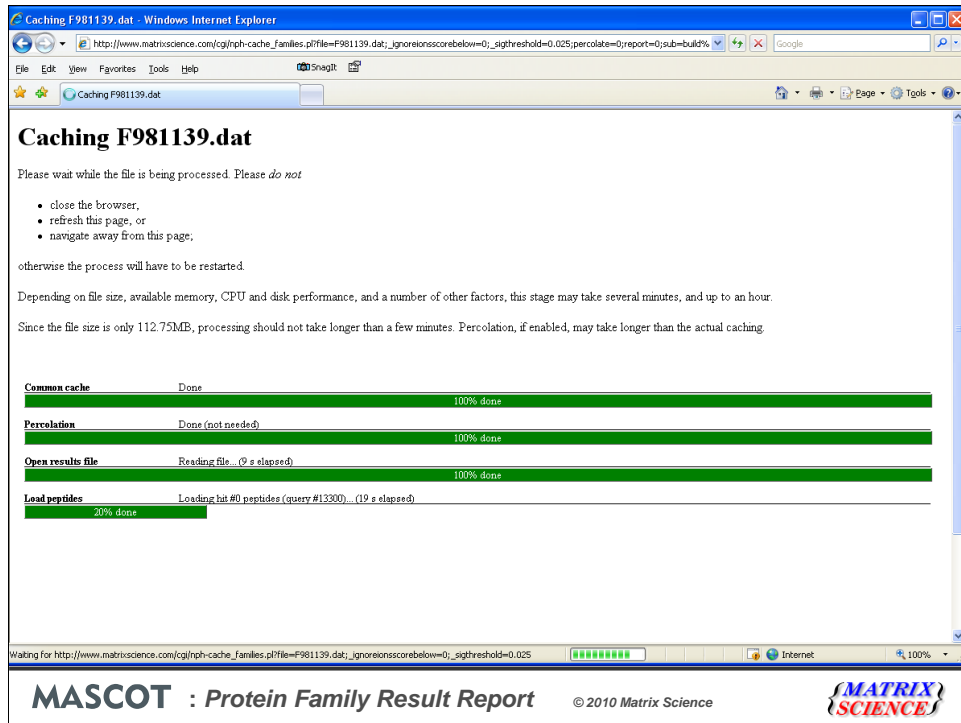
Of course, these two issues are related, because it is with very large search results that protein inference needs to be as automated as possible.

Protein Family Summary

Fixes both problems

- Index files are created and cached to speed loading in future
- Paged report to conserve memory
- Detailed information is shown 'on demand'
- Proteins grouped into families by means of shared peptide matches
- Only significant matches considered
- Hierarchical clustering within each protein family

To address both problems, we have developed a brand new report that loads most of the information 'on demand': the Protein Family Summary. 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. It is a paged report; we don't try to display all the protein hits at once. And, detailed information is shown 'on demand'. 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.



Lets begin by looking at the features that conserve memory and increase the speed of loading.

The Protein Family Summary is the default report whenever there are there are 300 or more spectra. This is controlled by a setting in mascot.dat, so you can easily change this threshold or even make the old report the default, if you prefer.

If you are running the search interactively, before the report loads for the first time, you'll see this progress page while the cache files are created. These cache files make loading much faster on subsequent occasions.

If you run searches using Mascot Daemon, the index files are created automatically and the search results should appear instantly when you click on the link in Daemon

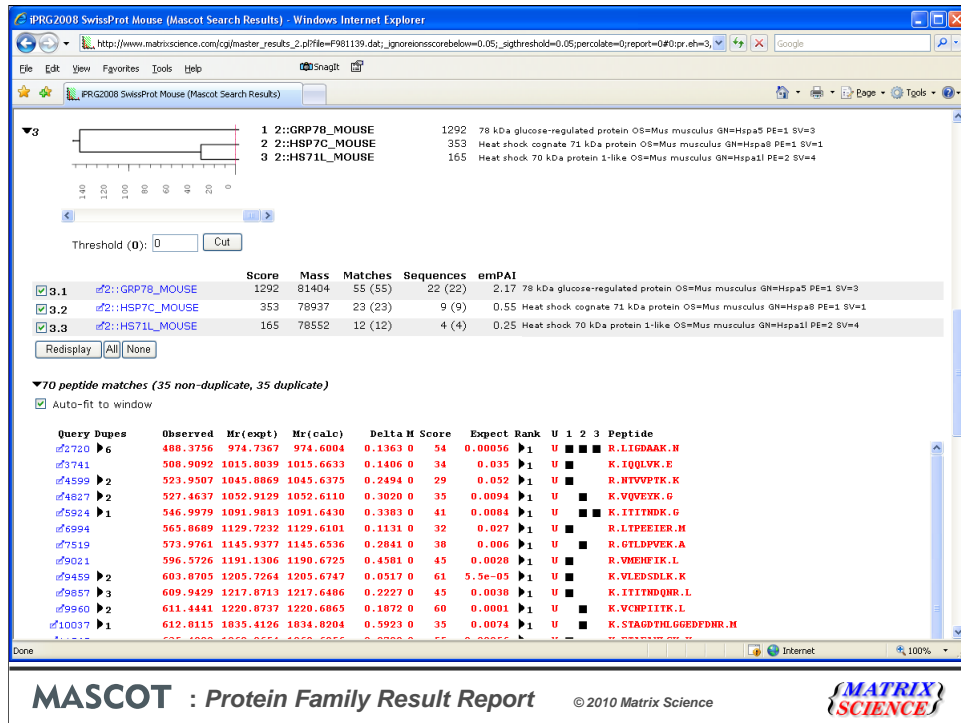
The screenshot shows a web browser window displaying the Mascot search results. The page is titled "iPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer". The browser address bar shows the URL: http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinscorebelow=0.05;_sigthreshold=0.05;_percolate=0;_report=0. The page content is organized into three tabs: "Proteins (482)", "Quantitation (514)", and "Unassigned (30250)". The "Proteins (482)" tab is active, showing "Protein families 1-10 (out of 482)". The page is paginated, showing "10 per page" and "49" families. The families are grouped into five clusters, each with a dendrogram and a list of protein accessions, scores, and descriptions.

Family	Accession	Score	Description
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	Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
	2 2::CP254_MOUSE	535	Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
	6 2::CP270_MOUSE	59	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
3	1 2::GRP78_MOUSE	1292	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
	2 2::HSP7C_MOUSE	353	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
	3 2::H871L_MOUSE	165	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4
4	2::CYB5_MOUSE	1228	Cytochrome b5 OS=Mus musculus GN=Cyb5a PE=1 SV=2
5	1 2::PDIA1_MOUSE	1116	Protein disulfide-isomerase OS=Mus musculus GN=P4hb PE=1 SV=1
	2 2::TXND5_MOUSE	46	Thioredoxin domain-containing protein 5 OS=Mus musculus GN=Txnd5 PE=1 SV=2

The page footer contains the text "MASCOT : Protein Family Result Report" and "© 2010 Matrix Science". The Matrix Science logo is also present.

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 quantitation, 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. We'll return to this later. If a family contains a single member, the accession string, protein score and description are listed. If a family contains multiple members, the accessions, scores and descriptions are aligned with a dendrogram, which illustrates the degree of similarity between members. To see complete information about the proteins in the family and the peptide matches assigned to each family member, click on the family number link. We'll click on family 3



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.

Below this is the table of peptide matches, which contains very similar information to that found in the other result reports. The black squares to the right show which peptides are found in which protein. To see only the peptides that distinguish HSP7C_MOUSE and HS71L_MOUSE, clear the checkbox for GRP78_MOUSE and choose Redisplay.

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinscorebelow=0.05;_sigthreshold=0.05;_percolate=0;_report=0#3:pr.eh=3

File Edit View Favorites Tools Help

IPRG2008 SwissProt Mouse (Mascot Search Results)

▼3

1	2::GRP78_MOUSE	1292	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
2	2::HSP7C_MOUSE	353	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
3	2::HSP7L_MOUSE	165	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Threshold (0): 0 Cut

	Score	Mass	Matches	Sequences	empAI	
<input type="checkbox"/> 3.1	2::GRP78_MOUSE	1292	81404	55 (55)	22 (22)	2.17 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
<input checked="" type="checkbox"/> 3.2	2::HSP7C_MOUSE	353	78937	23 (23)	9 (9)	0.55 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
<input checked="" type="checkbox"/> 3.3	2::HSP7L_MOUSE	165	78552	12 (12)	4 (4)	0.25 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Redisplay All None

▼24 peptide matches (11 non-duplicate, 13 duplicate)

Auto-fit to window

Query Dupes	Observed	Mr(expt)	Mr(calc)	Delta M	Score	Expect	Rank	U	2	3	Peptide
2720 ▶6	488.3756	974.7367	974.6004	0.1363	0	54	0.00056	▶1	U	■	R.LIGDAAK.N
4827 ▶2	527.4637	1052.9129	1052.6110	0.3020	0	35	0.0094	▶1	U	■	K.VQVEYK.G
5924 ▶1	546.9979	1091.9813	1091.6430	0.3383	0	41	0.0084	▶1	U	■	K.IITIDNK.G
7519 ▶1	573.9761	1145.9377	1145.6536	0.2841	0	38	0.006	▶1	U	■	R.GTLDPEK.A
9980 ▶2	611.4441	1220.8737	1220.6865	0.1872	0	60	0.0001	▶1	U	■	K.VCPPIIK.L
10037 ▶1	612.8115	1835.4126	1834.8204	0.5923	0	35	0.0074	▶1	U	■	K.STAGDTHLGGEDFNR.M
11946 ▶1	641.5476	1281.0806	1280.7220	0.3586	0	55	0.00015	▶1	U	■	K.EIAEAYLQK.T
25277 ▶1	607.4422	1819.3048	1818.8255	0.4793	0	55	3.2e-05	▶1	U	■	K.ATAGDTHLGGEDFNR.L
26376 ▶1	953.0936	1904.1726	1903.9845	0.1881	0	84	1.3e-07	▶1	U	■	K.SFYPEEVSQVLTQK.M
26946 ▶1	650.1325	1947.3756	1947.0920	0.2836	0	37	0.013	▶1	U	■	R.IINEPTAAAIAYGLDK
26947 ▶1	974.7142	1947.4139	1947.0920	0.3218	0	43	0.00059	▶1	U	■	R.IINEPTAAAIAYGLDK

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

The number of columns and rows in the table is reduced to the matches to the two selected proteins.

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinsscorebelow=0.05;_sighreshold=0.05;percolate=0;report=0#3:pr.eh=3

File Edit View Favorites Tools Help

PRG2008 SwissProt Mouse (Mascot Search Results)

▼3

1	2::GRP78_MOUSE	1292	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
2	2::HSP7C_MOUSE	353	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
3	2::HS71L_MOUSE	165	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Threshold (0): 0 Cut

	Score	Mass	Matches	Sequences	empAI	
<input type="checkbox"/> 3.1	2::GRP78_MOUSE	1292	81404	55 (55)	22 (22)	2.17 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
<input checked="" type="checkbox"/> 3.2	2::HSP7C_MOUSE	353	78937	23 (23)	9 (9)	0.55 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
<input checked="" type="checkbox"/> 3.3	2::HS71L_MOUSE	165	78552	12 (12)	4 (4)	0.25 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Redisplay All None

▼24 peptide matches (11 non-duplicate, 13 duplicate)

Auto-fit to window

Query Dupes	Observed	Mr(expt)	Mr(calc)	Delta M	Score	Expect	Rank	U	2	3	Peptide
2720 ▶6	488.3756	974.7367	974.6004	0.1363	0	54	0.00056	▶1	U	■	R.LIGDAAK.N
4827 ▶2	527.4637	1052.9129	1052.6110	0.3020	0	35	0.0094	▶1	U	■	K.VQVEYK.G
5924 ▶1	546.9979	1091.9813	1091.6430	0.3383	0	41	0.0084	▶1	U	■	K.VQVEYK.G
7519 ▶1	573.9761	1145.9377	1145.6536	0.2841	0	38	0.006	▶1	U	■	K.VQVEYK.G
9960 ▶2	611.4441	1220.8737	1220.6865	0.1872	0	60	0.0001	▶1	U	■	K.VQVEYK.G
10037 ▶1	612.8115	1835.4126	1834.8204	0.5923	0	35	0.0074	▶1	U	■	K.STAGDTHLGGDFDNR.M
11946 ▶1	641.5476	1281.0806	1280.7220	0.3586	0	55	0.00015	▶1	U	■	K.EIAGYLGK.T
25277 ▶1	607.4422	1819.3048	1818.8255	0.4793	0	55	3.2e-05	▶1	U	■	K.ATAGDTHLGGDFDNR.L
26376 ▶1	953.0936	1904.1726	1903.9845	0.1881	0	84	1.3e-07	▶1	U	■	K.SFYPEVSSHVLTK.M
26946 ▶1	650.1325	1947.3756	1947.0920	0.2836	0	37	0.013	▶1	U	■	R.IINEPTAAAIAYGLDK
26947 ▶1	974.7142	1947.4139	1947.0920	0.3218	0	43	0.00059	▶1	U	■	R.IINEPTAAAIAYGLDK

Score > 39 indicates identity
Score > 27 indicates homology

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinsscorebelow=0.05;_sighreshold=0.05;percolate=0

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

Tooltips are used to display additional information. For example, if the mouse cursor rests over the rank column, the score thresholds are displayed.

IPRG2008 SwissProt Mouse (Mascot Search Results)

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinscorebelow=0.05;_sigthreshold=0.05;_percolate=0;_report=0#4:pr.eh=3

File Edit View Favorites Tools Help

IPRG2008 SwissProt Mouse (Mascot Search Results)

▼ 3

1	2::GRP78_MOUSE	1292	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
2	2::HSP7C_MOUSE	353	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
3	2::HSP71L_MOUSE	165	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Threshold (0): 0 Cut

	Score	Mass	Matches	Sequences	empAI	
<input type="checkbox"/> 3.1	422::GRP78_MOUSE	1292	81404	55 (55)	22 (22)	2.17 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
<input checked="" type="checkbox"/> 3.2	422::HSP7C_MOUSE	353	78937	23 (23)	9 (9)	0.55 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
<input checked="" type="checkbox"/> 3.3	422::HSP71L_MOUSE	165	78552	12 (12)	4 (4)	0.25 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Redisplay All None

▼ 24 peptide matches (11 non-duplicate, 13 duplicate)

Auto-fit to window

Query Dupes	Observed	Mr(expt)	Mr(calc)	Delta M	Score	Expect	Rank	U	Z	Peptide	
42720 ▶6	488.3756	974.7367	974.6004	0.1363	0	54	0.00056	▶1	U	R.LIGDAAK.N	
Locus:19.275.3											
Score > 33 indicates identity											
Score > 27 indicates homology											
4827 ▶2	527.4637	1052.9129	1052.6110	0.3020	0	35	0.0094	▼1	W	K.VQVEYK.G	
	-0.6643	0	12	1.6	2					MPVEYK	
	0.3020	0	10	3.1	3					VEVQYK	
	0.2907	1	9	3.2	4					ADLQYK	
	0.3527	0	8	5	5					VQICEK	
	0.3020	0	7	5.6	6					LNLQYK	
	0.3020	0	7	5.6	6					NIEVYK	
	0.3020	0	6	7.9	8					SVTAPYK	
	0.3570	0	5	9.3	9					MPFPAPGAR + Oxidation (M)	
	-0.7036	1	4	10	10					SGGPRPALR	

Done Internet 100%

MASCOT : Protein Family Result Report © 2010 Matrix Science **MATRIX SCIENCE**

If you click on a marker in the rank column, it expands to show the 'yellow pop-up', containing the top 10 matches for this particular spectrum, except that this is no longer a pop-up. It is being displayed in-line so that you can print it, if you wish. You can also expand the top 10 for more than one match if you want to compare them

Duplicate matches to the same sequence are collapsed into a single row.

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinsscorebelow=0.05;_sigthreshold=0.05;_percolate=0;_report=0#S:pr.ed=3

Rank	Protein	Score	Mass	Matches	Sequences	empAI	Description
1	2::GRP78_MOUSE	1292	81404	55 (55)	22 (22)	2.17	78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
2	2::HSP7C_MOUSE	353	78937	23 (23)	9 (9)	0.55	Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
3	2::HSP71L_MOUSE	165	78552	12 (12)	4 (4)	0.25	Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Threshold (0): 0 [Cut]

Query	Dupes	Observed	Mr(expt)	Mr(calc)	Delta M	Score	Expect	Rank	U	2	3	Peptide
48270	6	488.3756	974.7367	974.6004	0.1363	0	54	0.00056	1	U	■	R.LIGDAAK.N
482700		488.3159	974.6173	974.6004	0.0169	0	(47)	0.0015	1	U	■	R.LIGDAAK.N
482705		488.3317	974.6488	974.6004	0.0484	0	(48)	0.0015	1	U	■	R.LIGDAAK.N
482708		488.3481	974.6817	974.6004	0.0813	0	(36)	0.026	1	U	■	R.LIGDAAK.N
482724		488.3865	974.7585	974.6004	0.1581	0	(32)	0.043	1	U	■	R.LIGDAAK.N
482733		488.4151	974.8156	974.6004	0.2152	0	(35)	0.045	1	U	■	R.LIGDAAK.N
482743		488.4656	974.9166	974.6004	0.3162	0	(41)	0.0094	1	U	■	R.LIGDAAK.N
4827	2	527.4637	1052.9129	1052.6110	0.3020	0	35	0.0094	1	U	■	K.VQVEYK.G

Auto-fit to window

Score > 33 indicates identity
Score > 27 indicates homology

Done

MASCOT : Protein Family Result Report © 2010 Matrix Science **MATRIX SCIENCE**

Usually, we are not very interested in lower scoring duplicate matches. But, if you want to see them, click on the marker in the duplicates column to expand. Obviously, clicking on either of these markers a second time causes the additional rows to collapse back to a single row

MASCOT Search Results

User :
 E-mail :
 Search title : iPRG2008 SwissProt Mouse
 MS data file : D:\PRG2008\img\merged.mgf
 Databases : 1: cRAP 20090731 (111 sequences; 37,180 residues)
 2: SwissProt 57.14 (514,789 sequences; 181,163,771 residues)
 Taxonomy : 1: (none)
 2: Mus. (16,273 sequences)
 Timestamp : 5 Mar 2010 at 14:04:48 GMT
 Warning : No taxonomy indexes for cRAP, taxonomy 'Mus.' ignored. Searching all entries in cRAP

Re-search All Non-significant Unassigned [\[help\]](#) Export As XML

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

Search parameters
 Score distribution
 Legend

Protein Family Summary

Filter Significance threshold p< 0.05 Max. number of families AUTO [\[help\]](#)
 Ions score or expect cut-off 0.05
 Show Percolator scores [\[help\]](#)

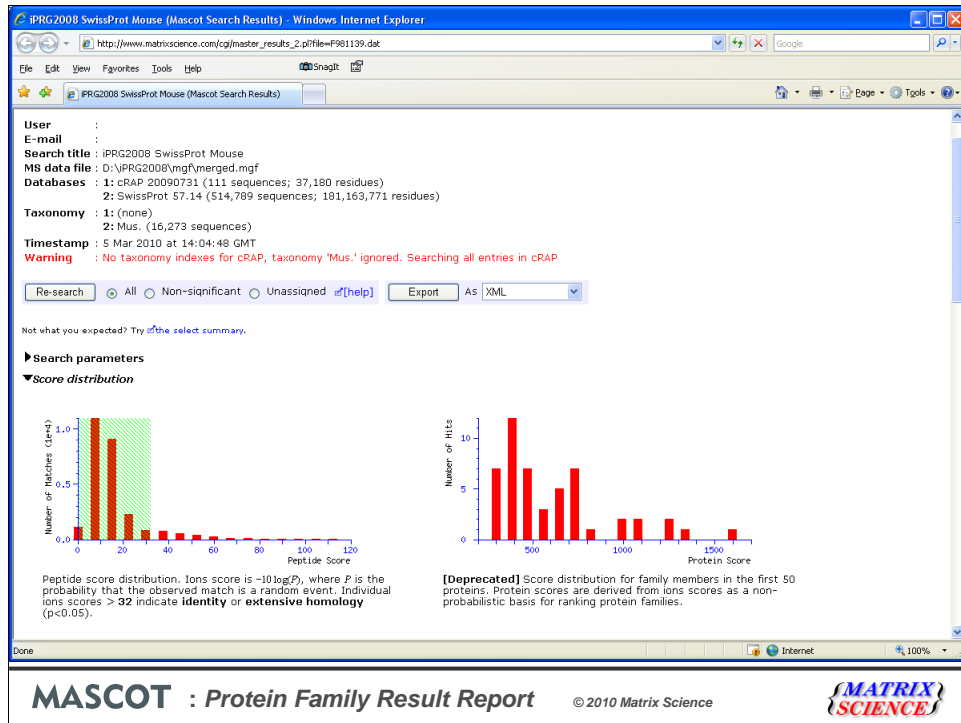
Decoy search summary

Proteins (482) [Quantitation \(514\)](#) [Unassigned \(30250\)](#) [\[link to\]](#)

Protein families 1-10 (out of 482)

MASCOT : Protein Family Result Report © 2010 Matrix Science **MATRIX SCIENCE**

Scrolling back to the top of the report, we can see the header information and format controls. This is mostly similar to the existing reports except that sections can be expanded or hidden, as required. Note that, if you want the older reports, the Select Summary and the Protein Summary, these are still available by clicking on this link.



For example, we can expand to show two score distributions: one for the peptide match scores and one for the protein scores. The old reports just have the protein score distribution, which isn't very meaningful, and is now marked deprecated (a software term meaning obsolete).

Let's take a quick look at the other two tabs. First, quantitation:

Time files like an arrow (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981131.dat;_ignoreionscorebelow=0;_min_precursor_charge=1;_proteinfamilyswitch=0;_quant

Protein Family Summary

Filter: Significance threshold p < 0.05 Max. number of families AUTO [help]
 Ions score or expect cut-off 0

Quantitate: Protein ratio type Weighted Normalisation None [help]
 Min. precursor charge 1 Outlier removal Automatic
 Min. # peptides 2 Peptide threshold Above homology 0.05

Display: Report peptide ratios Report protein ratios

Proteins (29) Quantitation (29) [link to]

Quantitation overview (29 proteins)

	Score	Mass	Matches	Sequences	emPAI	115/114	116/114	117/114		
1	VIT3_DROME	212	50436	19 (12)	13 (11)	1.13	1.740	2.241	3.202	Vitellogenin-3 precursor (Vitellogenin III) (Yolk...
2	BDI_DROME	136	63614	16 (7)	14 (7)	0.46	1.803	3.203	1.712	Protein disulfide-isomerase precursor (EC 5.2.3...
3	UGGG_DROME	132	187242	11 (7)	11 (7)	0.14	1.017	1.184	0.680	UDP-glucose: glycoprotein glucosyltransferase ...
4	BL4_DROME	118	51286	9 (4)	6 (4)	0.31	0.965	1.219	2.152	60S ribosomal protein L4 (L1) - Drosophila me...
5	VIT1_DROME	115	52618	12 (6)	4 (3)	0.22	1.169	1.157	1.544	Vitellogenin-1 precursor (Vitellogenin I) (Yolk...
6	YL_DROME	111	235777	7 (6)	6 (5)	0.08	1.088	1.829	1.178	Putative vitellogenin receptor precursor (Protei...
7	BL13_DROME	97	29639	5 (3)	4 (2)	0.26	1.569	3.054	4.981	60S ribosomal protein L15 (88C1 protein hom...
8	VDAC_DROME	97	34278	11 (5)	6 (3)	0.35	1.039	2.001	3.410	Voltage-dependent anion-selective channel (P...
9	KPYK_DROME	87	63726	5 (3)	4 (2)	0.18	0.853	1.916	3.531	Pyruvate kinase (EC 2.7.1.40) (PK) - Drosophi...
10	ATPB_DROME	87	57821	8 (4)	7 (4)	0.27	0.359	0.828	1.724	ATP synthase subunit beta, mitochondrial prec...
11	VIT2_DROME	84	52892	4 (2)	2 (1)	0.07	1.349	0.829	1.470	Vitellogenin-2 precursor (Vitellogenin II) (Yolk...
12	RS15A_DROME	83	16629	2 (2)	2 (2)	0.50	0.833	1.009	2.162	40S ribosomal protein S15As - Drosophila mal...
13	PERC_DROME	81	98282	2 (2)	2 (2)	0.07	1.260	1.487	1.096	Chorion peroxidase precursor (EC 1.11.1.7) (P...
14	EF11_DROME	77	57566	10 (6)	8 (5)	0.35	0.843	1.653	3.216	Elongation factor 1-alpha (EF-1-alpha) (G0 ID...
15	APLP_DROME	74	417818	28 (5)	26 (5)	0.04	0.911	1.432	1.799	Apolipoprotein precursor (Retinoid- and fatty ...
16	GBLP_DROME	73	38892	5 (4)	5 (4)	0.43	1.049	2.164	2.490	Guanine nucleotide-binding protein subunit be...

MASCOT : Protein Family Result Report © 2010 Matrix Science

The quantitation tab is always populated for a large result report, even if you didn't choose a quantitation method, because emPAI values are calculated automatically. This tab is also useful if you want a simple, single level table of the protein hits. In Mascot 2.3, all MS/MS reports give counts of both the number of peptide matches and the number of distinct sequences. The numbers in parentheses are the counts above the specified significance threshold. If you wanted a table of proteins that conformed to the MCP guideline concerning one-hit wonders, you would exclude any protein that didn't have a count of at least two significant sequences, so VIT2_DROME would be dropped

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreinscorebelow=0.05;_sigthreshold=0.05;_percolate=0;_report=0#7;pr_ed=3

Proteins (482) Quantitation (514) Unassigned (30250) [s.link.to](#)


Unassigned peptides, 1-100 (out of 30250)

100 per page 1 2 3 4 5 6 ... 303 Next Sort by Decreasing score

Query Filter

Peptide matches not assigned to protein families (no details means no match)

Query	Observed	Mr (expt)	Mr (calc)	Delta H	Score	Expect	Rank	Peptide
#220	415.7723	829.5301	829.5143	0.0158	0	35	0.053	▶1 DGLIIR
#4510	522.3610	1042.7075	1042.6257	0.0819	0	35	0.054	▶1 GTVLLSGPR
#7116	567.4544	1132.8943	1132.7059	0.1884	0	34	0.056	▶1 STYALNVK
#17701	731.5877	1461.1608	1460.7932	0.3677	0	34	0.064	▶1 VLPAPHLQYGR + Oxidation (M)
#1340	453.8680	905.7214	905.5789	0.1425	0	34	0.063	▶1 AISISK
#410	424.3200	846.6454	846.5782	0.0672	0	34	0.068	▶1 SLVIK
#1821	466.9290	931.8634	931.6310	0.2324	0	34	0.069	▶1 GILTLK
#4601	523.9779	1045.9413	1045.6375	0.3038	0	33	0.059	▶1 NTVVPTK
#3081	495.5299	989.0653	988.6524	0.4128	0	33	0.061	▶1 LQSIHK
#1896	468.9022	935.7898	935.6048	0.1850	0	33	0.065	▶1 LIPAGK
#6330	555.0061	1107.9977	1107.6168	0.3809	0	33	0.053	▶1 VSAETWK
#3334	501.3746	1000.7346	1000.6888	0.0458	0	33	0.053	▶1 ILAQLVK
#640	433.8702	865.7259	865.4780	0.2479	0	33	0.057	▶1 YVEVGR
#3601	505.4974	1008.9803	1008.6211	0.3591	0	33	0.053	▶1 LFNLSK
#7108	567.4183	1132.8220	1132.7059	0.1161	0	33	0.066	▶1 SGVSLAALK
#8962	595.9551	1189.8956	1189.7274	0.1682	0	33	0.067	▶1 QTATQLLK
#2728	488.3965	974.7784	974.6368	0.1416	0	33	0.071	▶1 NTLIVK
#585	431.8929	861.7712	861.5558	0.2154	0	33	0.081	▶1 LAFVIR
#7790	608.9847	1215.9549	1215.6581	0.2968	0	33	0.059	▶1 LTOEPESIR
#1609	461.3298	920.6451	920.5202	0.1250	0	33	0.067	▶1 FAGVDIR
#3341	501.4032	1000.7918	1000.6888	0.1030	0	32	0.064	▶1 ILAQLVK
#3404	502.3483	1002.6821	1002.6317	0.0504	0	32	0.058	▶1 LVIQDK
#3500	503.8457	1005.6768	1005.6093	0.0675	0	32	0.066	▶1 FSVQILR

MASCOT : Protein Family Result Report © 2010 Matrix Science 

Back to the original report, this is the unassigned peptide match tab. Again, it is paged, with a default of 100 peptide matches per page. As you can see, we have over 30,000 unassigned matches in this particular search, so displaying them all in a single list would use a large amount of memory. Up at the top of both the proteins and unassigned tab, you'll find a text search control

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;_ignoreionscorebelow=0.05;_sigthreshold=0.05;_percolabte=0;_report=0#3;pr_ed=3

Proteins (482) Quantitation (514) Unassigned (30250) [s.link.to](#)

Protein families 1-10 (out of 482)

10 per page 1 2 3 4 5 ... 49 Next Expand all Collapse all

Accession Find Match case

Accession
Description
Query
Observed
Mr(expt)
Mr(calc)
Sequence

1::TRY1_BOVIN 1597 TRY1_BOVIN

1 2::CP2CT_MOUSE 1307 Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1
4 4::CP239_MOUSE 293 Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1
5 5::CP238_MOUSE 202 Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1
2 2::CP254_MOUSE 535 Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1
6 6::CP270_MOUSE 69 Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2
3 3::CY250_MOUSE 382 Cytochrome P450 2C50 OS=Mus musculus GN=Cyp2c50 PE=1 SV=1

1 1::GRP78_MOUSE 1292 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
2 2::HSP7C_MOUSE 353 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
3 3::HSP7L_MOUSE 165 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Threshold (0): Cut

	Score	Mass	Matches	Sequences	empPAI	
<input checked="" type="checkbox"/> 3.1	2::GRP78_MOUSE	1292	81404	55 (55)	22 (22)	2.17 78 kDa glucose-regulated protein OS=Mus musculus GN=Hspa5 PE=1 SV=3
<input checked="" type="checkbox"/> 3.2	2::HSP7C_MOUSE	353	78937	23 (23)	9 (9)	0.55 Heat shock cognate 71 kDa protein OS=Mus musculus GN=Hspa8 PE=1 SV=1
<input checked="" type="checkbox"/> 3.3	3::HSP7L_MOUSE	165	78552	12 (12)	4 (4)	0.25 Heat shock 70 kDa protein 1-like OS=Mus musculus GN=Hspa11 PE=2 SV=4

Redisplay | All | None

MASCOT : Protein Family Result Report © 2010 Matrix Science **MATRIX SCIENCE**

This is particularly important for a paged report. You can search by accession or description sub-string, or by query number, mass or sequence.

Proteins (482) | Quantitation (514) | Unassigned (30250) | [Link to](#)

Protein families 41-50 (out of 482)

10 per page | [Previous](#) | [1](#) | [2](#) | [3](#) | [4](#) | [5](#) | [6](#) | [7](#) | [8](#) | [9](#) | [10](#) | ... | [49](#) | [Next](#) | [Expand all](#) | [Collapse all](#)

Sequence: MNVLADALK | [Find](#) | Match case | [Clear](#)

▶ 41 2::RS3_MOUSE 353 40S ribosomal protein S3 OS=Mus musculus GN=Rps3 PE=1 SV=1

▶ 42 2::RL22_MOUSE 347 60S ribosomal protein L22 OS=Mus musculus GN=Rpl22 PE=2 SV=2

▼ 43 2::RS15A_MOUSE 344 40S ribosomal protein S15a OS=Mus musculus GN=Rps15a PE=2 SV=2

43.1 [2::RS15A_MOUSE](#) **Score** **Mass** **Matches** **Sequences** **empAI**
344 16651 16 (16) 3 (3) 1.24 40S ribosomal protein S15a OS=Mus musculus GN=Rps15a PE=2 SV=2

▼ 16 peptide matches (4 non-duplicate, 12 duplicate)

Auto-fit to window

Query Dupes	Observed	Mr(expt)	Mr(calc)	Delta H	Score	Expect	Rank	U	Peptide
3708 ▶ 5	508.3777	1014.7407	1014.6308	0.1100	0	45	0.00053	▶ 1	U K.IVVHLTGR.L
11285 ▼ 5	631.9663	1261.9180	1261.7308	0.1872	0	77	2.3e-06	▶ 1	U R.MNVLADALK.S
11274	631.8868	1261.7591	1261.7308	0.0284	0	(66)	1.8e-05	▶ 1	U R.MNVLADALK.S
11275	631.8914	1261.7682	1261.7308	0.0375	0	(59)	9.4e-05	▶ 1	U R.MNVLADALK.S
11283	631.9416	1261.8686	1261.7308	0.1379	0	(59)	0.00012	▶ 1	U R.MNVLADALK.S
11287	632.0080	1262.0014	1261.7308	0.2706	0	(42)	0.0063	▶ 1	U R.MNVLADALK.S
11288	632.0218	1262.0291	1261.7308	0.2983	0	(63)	6.2e-05	▶ 1	U R.MNVLADALK.S
11604 ▶ 1	636.4751	1270.9355	1270.6904	0.2452	0	28	0.03	▶ 1	U K.WQHLPSR.Q
11780 ▼ 1	639.8954	1277.7762	1277.7257	0.0505	0	50	0.00081	▶ 1	U R.MNVLADALK.S + Oxidation (O)
11790	639.9899	1277.9652	1277.7257	0.2396	0	(48)	0.00054	▶ 1	U R.MNVLADALK.S + Oxidation (O)

MASCOT : Protein Family Result Report © 2010 Matrix Science

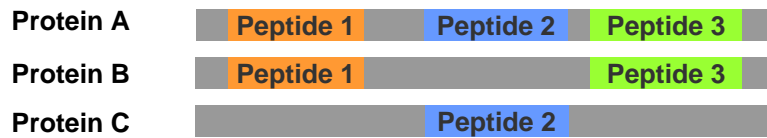
Here, for example, we searched for a peptide sequence. The display jumps to the first instance of the sequence, expands, and highlights (in green) the target peptides.

I hope you'll agree that this new report is a step in the right direction. Using cache files and displaying information 'on demand' allows even the largest search result to be displayed without memory or time-out problems

Protein Inference

General approach is to create a minimal list of proteins.

“Principal of parsimony” or “Occam’s razor”



MASCOT : Protein Family Result Report

© 2010 Matrix Science

MATRIX
SCIENCE

Now let's turn to the other aspect of the new report: improved protein inference.

As we all know, database matching of MS/MS spectra identifies peptides. Using these peptide sequences to deduce which proteins were present in the original sample is surprisingly difficult because many of the peptide sequences in a typical search result can be assigned to more than one protein.

The goal of protein inference is to create a minimal list of proteins. That is, the minimum number of proteins that can account for the observed peptides

Some people call this approach the principal of parsimony, others call it Occam's razor.

Imagine the very simple case we have three peptide matches, which can be assigned to three proteins, as illustrated here. Do we have evidence for all three proteins, or just one?

Mascot will report that the sample contained protein A. Proteins B and C are defined as sub-set proteins, and given an inferior status.

This is certainly a reasonable decision, but there is no guarantee that it is correct. It is possible that the sample actually did contain a mixture of proteins B and C, but not protein A. Another thing to watch for is the possibility that peptide 2 is a very weak match, maybe a random match. If so, then there is nothing to choose between Proteins A and B.

This ambiguity is made worse in a shotgun proteomics or MudPIT experiment, where the proteins from a cell lysate are digested to peptides without any prior fractionation or separation.

Protein Inference

Red indicates the top scoring peptide match for this spectrum

Bold indicates that this is the highest scoring protein in which this peptide is found

Look for protein hits with at least one bold red match

Format As: Peptide Summary Help
Significance threshold $p < 0.05$ Max number of hits: 150
Standard scoring MuSPIT scoring Ions score cut-off: 0 Show only top:
Show pop-ups Suppress pop-ups Sort unassigned: Decreasing Score Require bold red

MASCOT : Protein Family Result Report

© 2010 Matrix Science



In the Peptide Summary and Select Summary reports, the interpretation of the results from a large LC-MS/MS search can be tricky, because it is not always clear which proteins are “real” and which are just alternative groupings of peptide matches that are better assigned to other proteins.

We use red and bold typeface to highlight the most logical assignment of peptides to proteins. Whenever the top scoring peptide match for an MS/MS spectrum appears, it is shown in red. Note that this does not mean that the match is necessarily significant. It is just the best match. A peptide match is shown in bold face when this is the highest scoring protein in which a match to this spectrum has been found. This means that peptide matches that are both bold and red are the most likely assignments of the best matches. Conversely, if a protein hit doesn't have any bold red matches, either the assigned spectra have better scoring matches to other sequences or matches to the same spectra are assigned to higher scoring proteins elsewhere.

Usually, a protein hit without any bold red matches would collapse into a higher scoring protein hit were it not for the presence of one or more weak, random matches. Such hits can be filtered out of the report by ticking the ‘require bold red’ checkbox.

It is also a good idea to set a score cut-off, because this will make eliminate the weak, random matches that prevent proteins collapsing into a minimum number of hits. If the number entered into the Ions score cut-off is a number between 0 and 1 it is treated as a cut-off on the expect value. So, entering 0.05 will remove all non-significant matches.

Protein Inference

20: Q3UWB9	34: Q91WH2	51: Q3UEP4
GAVALNIR		
	GAAVTLNIR	GAAVTLNIR
	AEMWLIR	AEMWLIR
IILDELVOR	IILDELVQR	
FSPGYQIEK	FSPGYQIEK	
DNLENFFIK	DNLENFFIK	
FETFPYSVSK	FETFPYSVSK	
FVDVWTYEMPR	FVDVWTYEMPR	
		WTYEVPR
		IILDELK
		TPATLGPNTR
		FSPGYLEK
WLPQNDLLGHPK	WLPQNDLLGHPK	WLPQNDLLGHPK
ANAIAWALAIQPK		ANAIAWALAIQPK
GHEVTVLRPSAYYVLDPK	GHEVTVLRPSAYYVLDPK	

MASCOT : Protein Family Result Report

© 2010 Matrix Science



Using bold red and a score cut-off to filter the protein list works reasonably well, but it is far from ideal. In particular, the Peptide Summary and Select Summary reports fail to remove intersection proteins. A good example is shown here. Three proteins appear in the report at positions 20, 34, and 51. All have some bold red matches. And yet, when placed side by side, we can see that all the matches in 34 are found in 20 and 51. In other words, there is no independent evidence for protein 34. It's a sub-set of the combined matches to proteins 20 and 51.

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&querylist=at&REPTYPE=select&sigthresh=0.05&REPORT=AUTO&

File Edit View Favorites Tools Help

Select Summary Report (PRG2008 SwissProt Mouse)

31831 854.6325 2560.8736 2560.3563 0.3193 0 63 2.6e-05 1 U K.SIVHPSYNSHTLNDLILTK.L 23401 23406 23411 31632

31702 859.5194 2575.5364 2576.3512 -0.8147 0 (6) 11 3 U K.SIVHPSYNSHTLNDLILTK.L

2. 2:1:CP2CT_MOUSE Mass: 61433 Score: 1307 Matches: 168 (72) Sequences: 30 (11) emPAI: 1.47

Cytochrome P450 2C29 OS=Mus musculus GN=Cyp2c29 PE=1 SV=1

Query	Observed	Hr (expt)	Hr (calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
352	430.8697	839.7249	839.4883	0.2364	0	30	0.12	2	U	K.EALIDR.G 551
1373	454.3348	906.6551	906.5208	0.1350	0	26	0.33	1	U	R.MLQMK.R
1434	455.4326	908.8506	908.4474	0.4033	0	2	68	10	U	R.GEEFAGR.G
1630	461.4196	920.8247	920.5939	0.2308	0	27	0.29	1	U	K.YLIDK.G 1614 1617 1628 1637
1663	462.4237	922.8329	922.5150	0.3180	0	(8)	21	7	U	R.MLQMK.R 1644 1646
3438	502.8658	1003.7170	1003.5906	0.1264	0	28	0.2	1	U	K.LHEWIK.I 3441 3443 3445 3446 3457
4032	513.4571	1024.8997	1024.5861	0.3136	0	(15)	2.7	1	U	R.FTILTLR.N
4047	513.8493	1025.6841	1025.4214	0.2627	0	24	0.29	1	U	R.SPCHQDR.S 4051 4056 4057
4193	516.8977	1031.7808	1031.5369	0.2439	0	32	0.083	1	U	K.VQEEIDR.V 4197
4431	520.8626	1039.7106	1039.6157	0.0949	0	40	0.0079	1	U	K.SYLLEK.I 4409 4412 4420 4421 4423 4425 4429 4433 4439 4441 4442
4466	521.3753	1040.7361	1040.5810	0.1551	0	22	0.56	1	U	R.FTILTLR.N 4451 4476
4467	521.3901	1040.7657	1041.4164	-0.6506	0	(1)	62	10	U	R.SPCHQDR.S
5544	540.3247	1078.6349	1078.5385	0.0964	0	54	0.00028	1	U	R.LCAGEGLAR.M 5530 5539 5540 5542 5548 5549 5550 5566 5573
5605	541.3840	1080.7551	1080.6059	0.1492	0	53	0.00051	1	U	K.YPDVYAK.V 5594 5603 5604 5608 5614 5615 5616
5742	543.9752	1085.9359	1085.6113	0.3246	0	32	0.068	1	U	K.NFVYLR.S 5734 5739
7790	577.9297	1153.8449	1153.6045	0.2404	0	49	0.0011	1	U	R.GSFPFAEK.I 7774 7785 7786 7789 7792 7793 7795 7802
8340	586.0058	1169.9970	1169.5994	0.3976	0	(33)	0.049	1	U	R.GSFPFAEK.I 8271 8320 8321 8323 8326 8328 8332 8333 8338 8339
9079	597.5891	1193.1636	1192.7909	0.3727	1	5	35	1	U	R.KYLIPK.G
9186	599.5059	1196.9971	1196.6021	0.3150	1	8	14	1	U	R.RFTILTLR.N
10395	618.0392	1234.0638	1233.8304	0.2335	0	43	0.0059	1	U	R.YALLLLK.Y
12449	648.4734	1294.9323	1293.7403	-0.0079	0	27	0.22	1	U	K.EFLILHDK.I 12490
17119	681.9583	1442.8535	1442.7963	0.0572	1	23	0.3	1	U	K.VQEEIDRQVQK.H 2412
18138	739.6177	1477.2207	1476.8108	0.4100	0	63	4.7e-05	1	U	R.DPIQTLIK.Q 18083 18095 18098 18100 18105 18108 18117 18123 18124 18125 18126 18127 18128 18129 18130 18131 18132 18133 18134 18135 18136 18137 18138 18139 18140 18141 18142 18143 18144 18145 18146 18147 18148 18149 18150 18151 18152 18153 18154 18155 18156 18157 18158 18159 18160 18161 18162 18163 18164 18165 18166 18167 18168 18169 18170 18171 18172 18173 18174 18175 18176 18177 18178 18179 18180 18181 18182 18183 18184 18185 18186 18187 18188 18189 18190 18191 18192 18193 18194 18195 18196 18197 18198 18199 18200 18201 18202 18203 18204 18205 18206 18207 18208 18209 18210 18211 18212 18213 18214 18215 18216 18217 18218 18219 18220 18221 18222 18223 18224 18225 18226 18227 18228 18229 18230 18231 18232 18233 18234 18235 18236 18237 18238 18239 18240 18241 18242 18243 18244 18245 18246 18247 18248 18249 18250 18251 18252 18253 18254 18255 18256 18257 18258 18259 18260 18261 18262 18263 18264 18265 18266 18267 18268 18269 18270 18271 18272 18273 18274 18275 18276 18277 18278 18279 18280 18281 18282 18283 18284 18285 18286 18287 18288 18289 18290 18291 18292 18293 18294 18295 18296 18297 18298 18299 18300 18301 18302 18303 18304 18305 18306 18307 18308 18309 18310 18311 18312 18313 18314 18315 18316 18317 18318 18319 18320 18321 18322 18323 18324 18325 18326 18327 18328 18329 18330 18331 18332 18333 18334 18335 18336 18337 18338 18339 18340 18341 18342 18343 18344 18345 18346 18347 18348 18349 18350 18351 18352 18353 18354 18355 18356 18357 18358 18359 18360 18361 18362 18363 18364 18365 18366 18367 18368 18369 18370 18371 18372 18373 18374 18375 18376 18377 18378 18379 18380 18381 18382 18383 18384 18385 18386 18387 18388 18389 18390 18391 18392 18393 18394 18395 18396 18397 18398 18399 18400 18401 18402 18403 18404 18405 18406 18407 18408 18409 18410 18411 18412 18413 18414 18415 18416 18417 18418 18419 18420 18421 18422 18423 18424 18425 18426 18427 18428 18429 18430 18431 18432 18433 18434 18435 18436 18437 18438 18439 18440 18441 18442 18443 18444 18445 18446 18447 18448 18449 18450 18451 18452 18453 18454 18455 18456 18457 18458 18459 18460 18461 18462 18463 18464 18465 18466 18467 18468 18469 18470 18471 18472 18473 18474 18475 18476 18477 18478 18479 18480 18481 18482 18483 18484 18485 18486 18487 18488 18489 18490 18491 18492 18493 18494 18495 18496 18497 18498 18499 18500 18501 18502 18503 18504 18505 18506 18507 18508 18509 18510 18511 18512 18513 18514 18515 18516 18517 18518 18519 18520 18521 18522 18523 18524 18525 18526 18527 18528 18529 18530 18531 18532 18533 18534 18535 18536 18537 18538 18539 18540 18541 18542 18543 18544 18545 18546 18547 18548 18549 18550 18551 18552 18553 18554 18555 18556 18557 18558 18559 18560 18561 18562 18563 18564 18565 18566 18567 18568 18569 18570 18571 18572 18573 18574 18575 18576 18577 18578 18579 18580 18581 18582 18583 18584 18585 18586 18587 18588 18589 18590 18591 18592 18593 18594 18595 18596 18597 18598 18599 18600 18601 18602 18603 18604 18605 18606 18607 18608 18609 18610 18611 18612 18613 18614 18615 18616 18617 18618 18619 18620 18621 18622 18623 18624 18625 18626 18627 18628 18629 18630 18631 18632 18633 18634 18635 18636 18637 18638 18639 18640 18641 18642 18643 18644 18645 18646 18647 18648 18649 18650 18651 18652 18653 18654 18655 18656 18657 18658 18659 18660 18661 18662 18663 18664 18665 18666 18667 18668 18669 18670 18671 18672 18673 18674 18675 18676 18677 18678 18679 18680 18681 18682 18683 18684 18685 18686 18687 18688 18689 18690 18691 18692 18693 18694 18695 18696 18697 18698 18699 18700 18701 18702 18703 18704 18705 18706 18707 18708 18709 18710 18711 18712 18713 18714 18715 18716 18717 18718 18719 18720 18721 18722 18723 18724 18725 18726 18727 18728 18729 18730 18731 18732 18733 18734 18735 18736 18737 18738 18739 18740 18741 18742 18743 18744 18745 18746 18747 18748 18749 18750 18751 18752 18753 18754 18755 18756 18757 18758 18759 18760 18761 18762 18763 18764 18765 18766 18767 18768 18769 18770 18771 18772 18773 18774 18775 18776 18777 18778 18779 18780 18781 18782 18783 18784 18785 18786 18787 18788 18789 18790 18791 18792 18793 18794 18795 18796 18797 18798 18799 18800 18801 18802 18803 18804 18805 18806 18807 18808 18809 18810 18811 18812 18813 18814 18815 18816 18817 18818 18819 18820 18821 18822 18823 18824 18825 18826 18827 18828 18829 18830 18831 18832 18833 18834 18835 18836 18837 18838 18839 18840 18841 18842 18843 18844 18845 18846 18847 18848 18849 18850 18851 18852 18853 18854 18855 18856 18857 18858 18859 18860 18861 18862 18863 18864 18865 18866 18867 18868 18869 18870 18871 18872 18873 18874 18875 18876 18877 18878 18879 18880 18881 18882 18883 18884 18885 18886 18887 18888 18889 18890 18891 18892 18893 18894 18895 18896 18897 18898 18899 18900 18901 18902 18903 18904 18905 18906 18907 18908 18909 18910 18911 18912 18913 18914 18915 18916 18917 18918 18919 18920 18921 18922 18923 18924 18925 18926 18927 18928 18929 18930 18931 18932 18933 18934 18935 18936 18937 18938 18939 18940 18941 18942 18943 18944 18945 18946 18947 18948 18949 18950 18951 18952 18953 18954 18955 18956 18957 18958 18959 18960 18961 18962 18963 18964 18965 18966 18967 18968 18969 18970 18971 18972 18973 18974 18975 18976 18977 18978 18979 18980 18981 18982 18983 18984 18985 18986 18987 18988 18989 18990 18991 18992 18993 18994 18995 18996 18997 18998 18999 19000 19001 19002 19003 19004 19005 19006 19007 19008 19009 19010 19011 19012 19013 19014 19015 19016 19017 19018 19019 19020 19021 19022 19023 19024 19025 19026 19027 19028 19029 19030 19031 19032 19033 19034 19035 19036 19037 19038 19039 19040 19041 19042 19043 19044 19045 19046 19047 19048 19049 19050 19051 19052 19053 19054 19055 19056 19057 19058 19059 19060 19061 19062 19063 19064 19065 19066 19067 19068 19069 19070 19071 19072 19073 19074 19075 19076 19077 19078 19079 19080 19081 19082 19083 19084 19085 19086 19087 19088 19089 19090 19091 19092 19093 19094 19095 19096 19097 19098 19099 19100 19101 19102 19103 19104 19105 19106 19107 19108 19109 19110 19111 19112 19113 19114 19115 19116 19117 19118 19119 19120 19121 19122 19123 19124 19125 19126 19127 19128 19129 19130 19131 19132 19133 19134 19135 19136 19137 19138 19139 19140 19141 19142 19143 19144 19145 19146 19147 19148 19149 19150 19151 19152 19153 19154 19155 19156 19157 19158 19159 19160 19161 19162 19163 19164 19165 19166 19167 19168 19169 19170 19171 19172 19173 19174 19175 19176 19177 19178 19179 19180 19181 19182 19183 19184 19185 19186 19187 19188 19189 19190 19191 19192 19193 19194 19195 19196 19197 19198 19199 19200 19201 19202 19203 19204 19205 19206 19207 19208 19209 19210 19211 19212 19213 19214 19215 19216 19217 19218 19219 19220 19221 19222 19223 19224 19225 19226 19227 19228 19229 19230 19231 19232 19233 19234 19235 19236 19237 19238 19239 19240 19241 19242 19243 19244 19245 19246 19247 19248 19249 19250 19251 19252 19253 19254 19255 19256 19257 19258 19259 19260 19261 19262 19263 19264 19265 19266 19267 19268 19269 19270 19271 19272 19273 19274 19275 19276 19277 19278 19279 19280 19281 19282 19283 19284 19285 19286 19287 19288 19289 19290 19291 19292 19293 19294 19295 19296 19297 19298 19299 19300 19301 19302 19303 19304 19305 19306 19307 19308 19309 19310 19311 19312 19313 19314 19315 19316 19317 19318 19319 19320 19321 19322 19323 19324 19325 19326 19327 19328 19329 19330 19331 19332 19333 19334 19335 19336 19337 19338 19339 19340 19341 19342 19343 19344 19345 19346 19347 19348 19349 19350 19351 19352 19353 19354 19355 19356 19357 19358 19359 19360 19361 19362 19363 19364 19365 19366 19367 19368 19369 19370 19371 19372 19373 19374 19375 19376 19377 19378 19379 19380 19381 19382 19383 19384 19385 19386 19387 19388 19389 19390 19391 19392 19393 19394 19395 19396 19397 19398 19399 19400 19401 19402 19403 19404 19405 19406 19407 19408 19409 19410 19411 19412 19413 19414 19415 19416 19417 19418 19419 19420 19421 19422 19423 19424 19425 19426 19427 19428 19429 19430 19431 19432 19433 19434 19435 19436 19437 19438 19439 19440 19441 19442 19443 19444 19445 19446 19447 19448 19449 19450 19451 19452 19453 19454 19455 19456 19457 19458 19459 19460 19461 19462 19463 19464 19465 19466 19467 19468 19469 19470 19471 19472 19473 19474 19475 19476 19477 19478 19479 19480 19481 19482 19483 19484 19485 19486 19487 19488 19489 19490 19491 19492 19493 19494 19495 19496 19497 19498 19499 19500 19501 19502 19503 19504 19505 19506 19507 19508 19509 19510 19511 19512 19513 19514 19515 19516 19517 19518 19519 19520 19521 19522 19523 19524 19525 19526 19527 19528 19529 19530 19531 19532 19533 19534 19535 19536 19537 19538 19539 19540 19541 19542 19543 19544 19545 19546 19547 19548 19549 19550 19551 19552 19553 19554 19555 19556 19557 19558 19559 19560 19561 19562 19563 1956

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&_querylist=alt&REPTYPE=select&sigthreshold=0.05&REPORT=AUTO&

File Edit View Favorites Tools Help

Select Summary Report (PRG2008 SwissProt Mouse)

17997 737.4600 2209.3582 2210.1817 -0.8235 1 3 67 5 U H1RYVASYLLAALGGWSSPSAK.D
 27962 1021.6837 3062.0292 3061.6287 0.4005 0 97 1.9e-08 1 U K.LASVPAAGGAVAVSAAPSAAPAAGSAPAAAEK.K 32832

28. 21:CP254_MOUSE Mass: 60887 Score: 535 Matches: 76 (25) Sequences: 26 (8) emPAI: 0.87
 Cytochrome P450 2C54 OS=Mus musculus GN=Cyp2c54 PE=2 SV=1

Query	Observed	Hr(expt)	Hr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1373	454.3348	906.6551	906.5300	0.1350	0	26	0.33	1		R.NLQWR.R
1563	462.4237	922.8329	922.5150	0.3180	0	(8)	21	7		R.NLQWR.R 1644 1646
1953	470.2918	938.5691	938.3894	0.1797	0	11	4.8	1	U	R.PCHQR.S
3499	503.8391	1005.6637	1005.6103	0.0535	0	35	0.031	1		R.LPVEK.A 2509
4731	526.2961	1050.5776	1050.5323	0.0453	0	35	0.021	1		R.CLVEELR.K 4728 4741 4748 4758
4766	526.4296	1050.8446	1050.6317	0.2129	0	31	0.087	1		R.NFLLEK.I 4752 4773 4775
5303	535.4618	1068.9091	1068.6211	0.2880	0	31	0.082	1		R.NYFIRK.G 5279 5294 5304
5544	540.3247	1078.6349	1078.5385	0.0964	0	54	0.00028	1		R.TCAGEGLAR.N 5530 5539 5540 5542 5548 5549 5550 5566 5573
6164	551.9603	1101.9061	1102.6379	-0.7318	0	5	38	2		K.YPHYAK.V
7014	566.3561	1130.6977	1130.5964	0.1013	1	4	26	5		R.FDYRDK.D
7576	574.8879	1147.7612	1147.6117	0.1495	0	26	0.43	1		K.EFPRPK.F 7534 7538 7568 7573 7578
9880	610.3896	1218.7646	1218.7328	0.0318	1	(8)	25	3		R.GRLPVEK.A 9900
9906	407.3264	1218.9573	1218.7328	0.2245	1	12	8.3	1		R.GRLPVEK.A
10395	618.0392	1234.0638	1233.8304	0.2335	0	43	0.0053	1		R.YALLLLK.Y
17673	731.0732	1460.1318	1460.8159	-0.6841	0	17	2.6	2		R.DFIDYFLIK.G
2918	492.4046	1474.1920	1474.7497	-0.5577	1	(7)	41	10		R.SIEDRVQEAR.C
18043	738.5613	1475.1080	1474.7497	0.3582	1	8	14	2		R.SIEDRVQEAR.C
18114	739.5340	1477.0534	1476.8634	0.1900	0	82	6.3e-07	1	U	K.GTVVITLSLVLR.D 18115 18118 18122 18125 18136 18139 18142 1814
3699	507.4343	1519.2810	1518.8211	0.4599	1	1	1.4e-02	9	U	K.NTRPFLMLLR.N
19465	510.6492	1528.9257	1528.7736	0.1500	0	(59)	9.2e-05	1	U	K.EALVDHGVFAGR.G 19486 19488
19473	765.3187	1529.0298	1528.7736	0.2472	0	93	4.4e-08	1	U	K.EALVDHGVFAGR.G 19488
19851	773.1436	1544.2725	1543.8921	0.4705	0	61	6.6e-05	1	U	R.VDGVVYTLGLR.K 19848 19853
20447	784.5878	1567.1610	1566.7632	0.3978	0	10	8.4	3		K.SDYFHFSTGR.K
20458	923.6711	1567.9915	1567.8813	0.1102	0	27	0.15	1		K.VOREIENIGK.N
20822	528.3086	1581.9040	1582.7581	-0.8541	0	(3)	27	3		K.SDYFHFSTGR.R
20849	792.3833	1582.7520	1582.7581	-0.0061	0	(7)	20	4		K.SDYFHFSTGR.R 20877
23961	580.8898	1739.6475	1738.8592	0.7882	1	4	27	7		K.SDYFHFSTGR.I
33988	581.4760	1741.4063	1741.8889	-0.4819	1	6	18	7	U	K.SDYFHFSTGR.Y 33993

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

If you were to browse down, you'd find another Cytochrome P450 at hit 28

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&querylist=at&REPYPE=select&sigtheshold=0.05&REPORT=AUTO&

Select Summary Report (PRG2008 SwissProt Mouse)

31915 877.9669 2630.8790 2630.3618 0.5173 1 28 0.065 1 U R.KDLYANTLVSGGTYMPLIADR.H

44. 21:CY2S0_MOUSE Mass: 61037 Score: 392 Matches: 76(20) Sequences: 25(7) emPAI: 0.87
 Cytochrome P450 2C50 OsMus musculus GN=Cyp2c50 PE=1 SV=1

Query	Observed	Hr(expt)	Hr(cal)	Delta Miss	Score	Expect	Rank	Unique	Peptide
1373	454.3348	906.6551	906.5200	0.1350	0	26	0.33	1	R.NLQGR.R
1663	462.4237	922.8329	922.5150	0.3180	0	(8)	21	7	R.NLQGR.R 1644 1646
2886	491.4316	980.8487	980.5777	0.2710	0	12	11	3	U R.FSLTLR.N 2869 2878
3499	503.8391	1005.6637	1005.6103	0.0535	0	35	0.031	1	R.LPVFDR.A 3509
4047	513.8493	1025.6841	1025.4214	0.2627	0	24	0.29	1	R.SPCHQDR.S 4051 4056 4057
4467	521.3901	1040.7657	1041.4164	-0.6506	0	(1)	62	10	R.SPCHQDR.S
4731	526.2961	1050.5776	1050.5323	0.0453	0	35	0.021	1	R.CLVEELR.K 4728 4741 4748 4758
4766	526.4296	1050.8446	1050.6317	0.2129	0	31	0.087	1	R.NFLER.I 4752 4773 4775
5303	535.4618	1068.9091	1068.6211	0.2880	0	31	0.082	1	R.NYFIPK.G 5279 5294 5304
5544	540.3247	1078.6349	1078.5385	0.0964	0	54	0.00028	1	R.LCAGEGLAR.M 5530 5539 5540 5542 5548 5549 5550 5556 5573
6164	551.9603	1101.9061	1102.6379	-0.7318	0	5	38	2	R.YPRVTKR.V
7014	566.3561	1130.6977	1130.5964	0.1013	1	4	26	5	R.FDYKDR.D
7576	574.8879	1147.7612	1147.6117	0.1495	0	26	0.43	1	K.EFFNPK.F 7534 7538 7566 7573 7578
9880	610.3896	1218.7646	1218.7328	0.0318	1	(8)	25	3	R.GRLPVFDR.A 9900
9906	407.3264	1218.9573	1218.7328	0.2245	1	12	8.3	1	R.GRLPVFDR.A
10395	619.0392	1234.0638	1233.6304	0.2335	0	43	0.0053	1	R.YALLLLK.Y
13053	637.3789	1312.7271	1312.6948	0.0323	0	3	49	5	R.DFLMLK.L
2918	492.4046	1474.1920	1474.7497	-0.5577	1	(7)	41	10	R.SIEDRVQEAR.C
18043	738.5613	1475.1080	1474.7497	0.3582	1	8	14	2	R.SIEDRVQEAR.C
18443	745.9500	1489.8854	1489.8386	0.0268	0	73	5.0e-06	1	U K.GTWIITLSSVLR.D 18448 18450 18455 18466 18469
20447	784.5878	1567.1610	1566.7632	0.3978	0	10	8.4	3	R.SDYFHFSTGR.R
20458	823.6711	1567.9915	1567.8813	0.1102	0	27	0.15	1	K.VQEIIEHIVGR.H
20594	787.5444	1573.0743	1572.7654	0.3089	0	(43)	0.0039	1	K.EALVDHGEFAGR.G
4705	525.4566	1573.3479	1572.7654	0.5824	0	71	1.7e-05	1	K.EALVDHGEFAGR.G
20822	528.3086	1581.9040	1582.7581	-0.8541	0	(3)	27	3	R.SDYFHFSTGR.R
20849	792.3833	1582.7520	1582.7581	-0.0061	0	(7)	20	4	R.SDYFHFSTGR.R 20877
20899	795.6114	1589.2082	1588.7833	0.4249	0	32	0.057	1	K.DICQSFTHSK.V
23961	580.8898	1739.6475	1738.8592	0.7882	1	4	27	7	K.SDYFHFSTGR.I

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

And another at 44

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&querylist=at&REPYPE=select&sigthresold=0.05&REPORT=AUTO&

File Edit View Favorites Tools Help

Select Summary Report (PRG2008 SwissProt Mouse)

27148 982.7220 1963.4295 1963.0811 0.3484 0 (67) 1.7e-05 1 U K.AAGVSVPEFPGLFAK.A 27149 27150 27152 27157

77. **21:CP239_MOUSE** Mass: 60932 Score: 293 Matches: 97(22) Sequences: 17(6) emPAI: 0.41
 Cytochrome P450 2C39 OS=Mus musculus GN=Cyp2c39 PE=2 SV=1

Query	Observed	Hr(expt)	Hr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1373	454.3348	906.6551	906.5200	0.1350	0	26	0.33	1		R.NLQMRK.R
1663	462.4237	922.8329	922.5150	0.3180	0	(8)	21	7		R.NLQMRK.R 1644 1646
3438	502.8658	1003.7170	1003.5906	0.1264	0	28	0.2	1		K.LNHNVK.I 3441 3443 3445 3446 3457
4047	519.8499	1025.6841	1025.4214	0.2627	0	24	0.29	1		R.SPCMRK.N 4051 4056 4057
4431	520.8626	1039.7106	1039.6157	0.0949	0	40	0.0079	1		R.SVLLK.I 4409 4412 4420 4421 4423 4425 4429 4433 4439 4441 4442
4467	521.3901	1040.7657	1041.4164	-0.6506	0	(1)	62	10		R.SPCMRK.N
5605	541.3848	1080.7551	1080.6059	0.1492	0	53	0.00051	1		K.YPDVTK.V 5594 5603 5604 5608 5614 5615 5616
11467	634.4301	1266.8855	1267.7817	-0.8962	0	9	21	1		R.YALLLHK.V 11517
16922	719.5773	1437.1399	1437.7698	-0.6298	0	10	15	8	U	K.VQEEIDHWI.R.H 16929 16958
17976	491.4253	1471.2541	1471.8642	-0.6101	1	11	6.4	5		K.VYRSYLLK.I
18138	739.6177	1477.2207	1476.8108	0.4100	0	63	4.7e-05	1		R.DFYDYLK.Q 18082 18095 18098 18100 18105 18108 18117 18123 18138
18994	756.5141	1511.0136	1510.7482	0.2653	0	(4)	32	10	U	K.SDHQPFSAK.R
3794	509.7788	1526.3144	1526.7431	-0.4287	0	9	21	1	U	K.SDHQPFSAK.R 19411
19434	764.5613	1527.1080	1526.7431	0.3649	0	(7)	16	1	U	K.SDHQPFSAK.R
20264	781.1019	1560.1892	1559.8187	0.3704	0	25	0.26	2	U	K.NFSQSLTNFSK.A 20249 20252 20256 20257 20262 20265 20267 20268
4447	821.2416	1590.7029	1559.8187	0.8842	0	(24)	0.64	2	U	K.NFSQSLTNFSK.A 20263 20270
21647	810.1993	1618.3920	1617.8709	0.5111	0	13	4	3	U	K.LHGLGIVSMDR.N
22532	554.6824	1661.0254	1660.7814	0.2440	0	48	0.0011	1	U	K.EALIDHGEESDR.G
23907	868.6286	1735.2427	1734.8402	0.4025	0	61	6.6e-05	1		R.VQEAQLVEELR.K
26729	967.1937	1932.3729	1932.0772	0.2957	0	(11)	5.5	2		K.GTTVTSLSLTVLHDSK.E
26736	645.1888	1932.5446	1932.0772	0.4674	0	31	0.054	2		K.GTTVTSLSLTVLHDSK.E 26724 26732 26735 12189 26737 26738 26740 26742
20523	786.4158	2356.2255	2356.2630	-0.0375	1	7	18	7	U	K.YPDVTKVQEEIDHWI.R.H
32013	888.6022	2662.7849	2662.3152	0.4697	1	4	20	2	U	K.EALIDHGEESDRGSLPVEK.I

78. **21:COMT_MOUSE** Mass: 31679 Score: 293 Matches: 37(12) Sequences: 8(5) emPAI: 0.92
 Catechol O-methyltransferase OS=Mus musculus GN=Comt PE=1 SV=1

Query	Observed	Hr(expt)	Hr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
-------	----------	----------	----------	-------	------	-------	--------	------	--------	---------

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

And at 69

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&querylist=at&REPTYPE=select&sigthresh=0.05&REPORT=AUTO&

File Edit View Favorites Tools Help

Select Summary Report (PRG2008 SwissProt Mouse)

17179 482.6943 1445.0610 1445.7708 -0.7098 1 (4) 38 10 U R.SLQSVAAERAGR.H
 26956 650.4782 1948.4126 1947.9822 0.4304 0 36 0.018 1 U R.VLNSYVWGEDSTYK.F

115. 21:CP238_MOUSE Mass: 61216 Score: 202 Matches: 75 (18) Sequences: 19 (4) emPAI: 0.40
 Cytochrome P450 2C38 OS=Mus musculus GN=Cyp2c38 PE=2 SV=1

Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1373	454.3348	906.6551	906.5200	0.1350	0	26	0.33	1		R.NLGRK.R
1663	462.4237	922.8329	922.5150	0.3180	0	(8)	21	7		R.NLGRK.R 1644 1646
3438	502.8658	1003.7170	1003.5908	0.1264	0	28	0.2	1		R.LNENVK.I 3441 3443 3445 3446 3457
4047	513.8493	1025.8041	1025.4214	0.2627	0	24	0.29	1		R.SPCHGR.S 4051 4056 4057
4431	520.8626	1039.7106	1039.6157	0.0949	0	40	0.0079	1		R.STLLEK.V 4409 4412 4420 4421 4423 4425 4429 4432 4439 4441 4442
4467	521.3901	1040.7657	1041.4164	-0.6506	0	(1)	62	10		R.SPCHGR.S
5605	541.3848	1080.7551	1080.6059	0.1492	0	53	0.00051	1		R.YPDVTK.V 5594 5603 5604 5608 5614 5615 5616
8679	590.5231	1179.0315	1178.6209	0.4107	0	5	37	2	U	R.GHPISEK.I
11467	634.4501	1266.8855	1267.7817	-0.8962	0	9	21	1		R.VALLILMK.Y 11517
16093	706.4691	1410.9236	1410.8811	0.0425	1	2	54	8	U	R.STLLEK.V
16519	713.0426	1424.0705	1423.7541	0.3164	0	8	15	4	U	R.VQIEDHVR.H 16490
17673	731.0732	1460.1318	1460.8159	-0.6841	0	17	2.6	2		R.DFIDVLIK.Q
17976	491.4253	1471.2541	1471.8642	-0.6101	1	11	6.4	5		K.YIRSYLLEK.V
18913	755.1673	1508.3200	1507.7132	0.6068	0	2	92	5	U	R.EEAQCLVELR.K
19737	771.0876	1540.1606	1540.7476	-0.5869	0	7	18	3	U	K.SDYFTFSAGK.R
20190	820.1012	1557.2819	1556.7425	0.5394	0	(1)	65	5	U	K.SDYFTFSAGK.R
21325	802.5887	1603.1628	1602.7760	0.3868	0	48	0.0012	2		K.EALIDGEEFSGR.G 21316
21547	530.4897	1612.4474	1611.9592	0.4882	0	29	0.11	1	U	R.PIVVLHGVEAVK.E
7037	566.5113	1696.5121	1696.8487	-0.3366	1	3	1e+02	6	U	K.SDYFTFSAGK.V
26729	967.1937	1932.3729	1932.0772	0.2957	0	(11)	5.5	2		K.GTTVVVTSLSVLHDSK.E
26736	645.1888	1932.5446	1932.0772	0.4674	0	31	0.054	2		K.GTTVVVTSLSVLHDSK.E 26724 26732 26735 12189 26737 26738 26740 2
12778	653.4007	1957.1801	1957.9260	-0.7459	0	(9)	15	2		R.SHPPTDAMIREVQR.F
27266	659.1885	1974.5436	1973.9209	0.6226	0	30	0.072	1		R.SHPPTDAMIREVQR.F
27449	664.4707	1990.3904	1989.9159	0.4745	0	(4)	31	8		R.SHPPTDAMIREVQR.F
31882	874.3484	2620.0233	2619.2842	0.7391	1	1	40	9	U	K.EALIDGEEFSGRGIPSEK.I

116. 34:DL3_MOUSE Mass: 54530 Score: 108 Matches: 24 (13) Sequences: 16 (5) emPAI: 0.37

MASCOT : Protein Family Result Report © 2010 Matrix Science

MATRIX SCIENCE

And 115

Select Summary Report (PRG2008 SwissProt Mouse) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results.pl?file=...%2Fdata%2F981139.dat&_querylist=all&REPTYPE=select&sigthresh=0.05&REPORT=AUTO&

File Edit View Favorites Tools Help

11137 629.9678 1257.9210 1257.7186 0.2024 1 4 43 3 U K.EPHSAKQK.Y
 19883 773.5474 1545.0801 1544.8008 0.2793 0 56 0.0022 1 U K.CHQLTDFILK.F

254. 21:CP270_MOUSE Mass: 61539 Score: 69 Matches: 39 (3) Sequences: 16 (2) emPAI: 0.25
 Cytochrome P450 2C70 OS=Mus musculus GN=Cyp2c70 PE=2 SV=2

Query	Observed	Hr(expt)	Hr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
1452	520.8779	1039.7412	1040.5602	-0.8190	0	(4)	37	6	U	K.SHGELAK.K
1451	521.2913	1040.5680	1040.5602	0.0078	0	8	17	2	U	K.SHGELAK.K
5540	540.2925	1078.5705	1078.5385	0.0320	0	22	0.77	3	U	R.ACTIGELAR.N 5530 5539 5542 5544 5548 5549 5550 5566 5573
7576	574.8879	1147.7612	1147.6117	0.1495	0	26	0.43	1	U	K.EFFNPEK.F 7534 7538 7568 7573 7576
8573	589.4511	1176.8877	1176.6624	0.2253	0	35	0.043	1	U	K.FILEELHR.H
8720	591.4148	1180.8150	1180.6872	0.1278	1	13	5	1	U	R.RFSLVLR.S
9329	602.0796	1202.1447	1201.6524	0.4923	1	7	34	6	U	R.SHGGRK.T 9273
9905	610.4838	1218.9530	1219.8147	-0.8618	0	5	34	8	U	K.YGLLLK.Y
13028	656.9476	1311.8806	1312.7572	-0.8766	1	(6)	21	6	U	K.SHGELAK.Y 13031
13545	443.7078	1328.1015	1328.7521	-0.6507	1	(4)	32	10	U	K.SHGELAK.Y
13585	665.4865	1328.9585	1328.7521	0.2064	1	15	3	3	U	K.SHGELAK.Y
16000	705.3286	1408.6427	1407.7956	0.8471	0	12	5.2	1	U	K.LQELIAVIGR.H 15958 15959
17260	724.4938	1446.9730	1446.7993	0.1737	0	59	0.00013	1	U	R.YIDFVIPSPP.K 17277 17287
18496	490.3042	1491.8907	1491.7602	0.1305	0	24	0.25	1	U	K.FDGGHLEK.G
3666	507.2400	1518.6981	1518.7701	-0.0720	1	8	23	3	U	K.SDYFVAFSAGR.A
24542	592.9386	1775.7939	1775.9340	-0.1401	1	3	35	10	U	R.FSLMVLRSIMGK.K
30746	799.6049	2395.7928	2395.3251	0.4677	0	10	5.3	1	U	R.HEDFLITLILQHFTLK.P
30874	805.5366	2413.5880	2413.0192	0.5689	0	15	1.7	1	U	R.FDYNDYQDFHNFHR.K
31854	872.2805	2613.8196	2613.3013	0.5183	1	29	0.056	1	U	K.EALIDGDEFSKDTSSLLSR.T

255. 21:P823_MOUSE Mass: 18942 Score: 69 Matches: 15 (6) Sequences: 4 (2) emPAI: 0.43
 40S ribosomal protein S23 OS=Mus musculus GN=Rps23 PE=2 SV=3

Query	Observed	Hr(expt)	Hr(calc)	Delta	Miss	Score	Expect	Rank	Unique	Peptide
927	444.8482	887.6818	887.6048	0.0770	0	14	4.6	6	U	R.VQLIK.N 1002
2081	473.8444	945.6741	945.6102	0.0639	0	39	0.014	1	U	K.GIVLEK.V 2066 2068 2084 2086 2090 2095 2096 2097 2100
13646	666.4200	1330.8254	1330.8662	-0.0407	1	3	62	10	U	R.VQLINQK.K
18157	740.0822	1478.1499	1477.9112	0.2387	0	38	0.015	1	U	K.VANVSLALYK.G 18159

MASCOT : Protein Family Result Report © 2010 Matrix Science

And 254. Apart from the descriptions all containing the words Cytochrome P450, did you spot any other relationships? If you did, you're very smart and have a photographic memory. These 6 proteins are linked by a complex web of shared peptide matches. But, spotting this in the current reports is far from easy. This report, which is not particularly large, runs to 691 proteins. Think how many possible families there are in such a set of proteins.

Protein Inference: Family Report

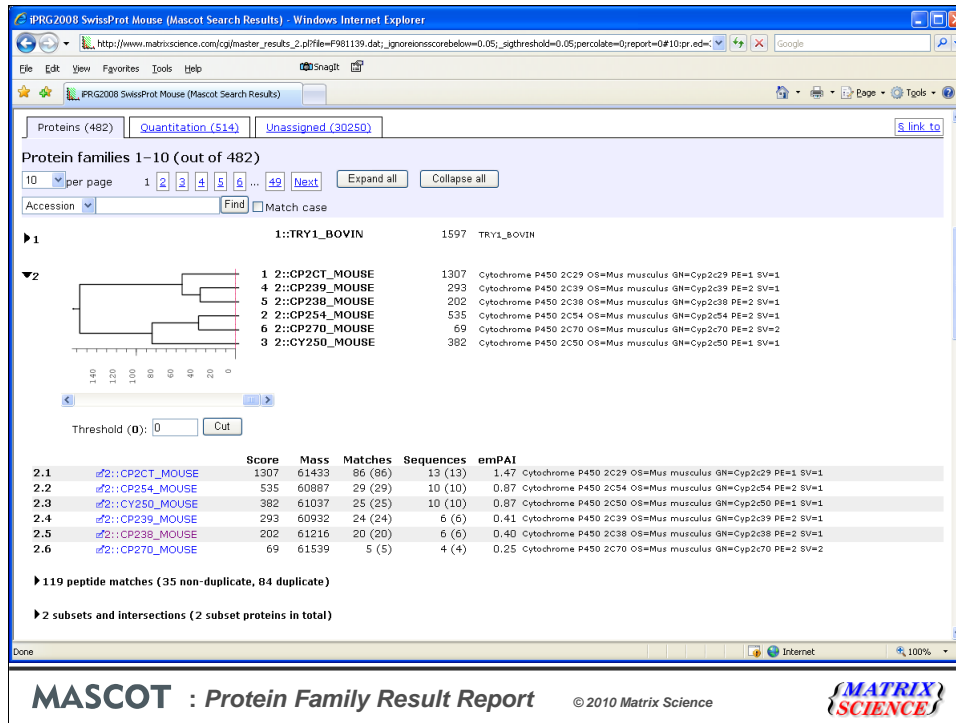
1. From the search results, create an initial list of proteins, ordered by protein score
2. Take the highest scoring protein on the list
3. Find all other proteins in the same family:
 - select all peptide matches with homology score or better
 - for each peptide match, select all the proteins that contain this match and remove from the initial list
 - for each new protein, select all new peptide matches with homology score or better
 - loop until no further proteins or peptide matches remain
4. For each protein in the family, make a list of the distinct peptide sequences. That is, ignoring differences in modification state and precursor charge. Where there are duplicate matches to a sequence, the representative score for the sequence is the highest one
5. Using this set of distinct peptide sequences, divide and group the proteins into same-set proteins and sub-set proteins, which includes intersections.
 - Same-set proteins are collapsed into a single family member
 - Proteins that are sub-sets, including intersections, are relegated to secondary status
 - Perform hierarchical clustering of the family members
6. Loop from step 2 until no more peptide matches remain with homology score or better

MASCOT : Protein Family Result Report

© 2010 Matrix Science



The grouping algorithm in the new, protein family report clusters proteins into families on the basis of shared peptide matches. Only matches with scores above the homology threshold are used, non significant matches are ignored. An iterative method is used to remove intersection proteins as well as sub-set proteins.



This is how the same 6 cytochrome P450 proteins are presented in the new report. Because they have shared, significant matches, they are grouped into a family.

The dendrogram illustrates the degree of similarity between the members. The scale on the dendrogram is ions score, and CP239_MOUSE and CP238_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. Where there are multiple matches to the same peptide sequence, it is the highest score for each sequence that is used to calculate the distance.

IPRG2008 SwissProt Mouse (Mascot Search Results) - Windows Internet Explorer

http://www.matrixscience.com/cgi/master_results_2.pl?file=F981139.dat;ignoreionscorebelow=0.05;sigthreshold=0.05;percolate=0;report=0#11:pr.ed=

PRG2008 SwissProt Mouse (Mascot Search Results)

▼ 119 peptide matches (35 non-duplicate, 84 duplicate)

Auto-fit to window

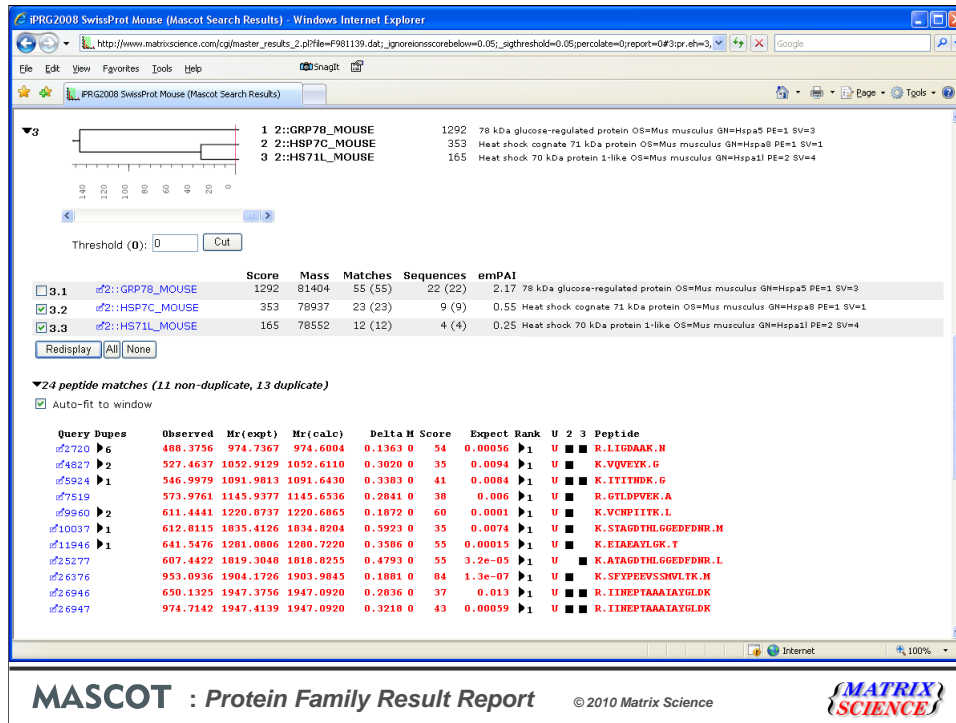
Query Dupes	Observed	Mr(expt)	Mr(calc)	Delta H Score	Expect	Rank	U	1	2	3	4	5	6	Peptide
3499	503.8391	1005.6637	1005.6103	0.0535	0	35	0.031	▶1	U	■	■	■	■	R.LPVFDR.A
4431 ▶9	520.8626	1039.7106	1039.6157	0.0949	0	40	0.0079	▶1	U	■	■	■	■	SYLLEK
4447 ▶2	521.2416	1560.7029	1559.8187	0.8842	0	59	0.00021	▶1	U	■	■	■	■	K.MISQSPHFSA.A
4466	521.3753	1040.7361	1040.5810	0.1551	0	22	0.031	▶1	U	■	■	■	■	R.FTHLYLR.H + Oxidation (H)
4705	525.4566	1573.3479	1572.7654	0.5824	0	71	1.4e-05	▶1	U	■	■	■	■	K.EALVDHGEFAGR.G
4731 ▶3	526.2961	1050.5776	1050.5323	0.0453	0	35	0.0085	▶1	U	■	■	■	■	R.CLYEELR.K
5544 ▶8	540.3247	1078.6349	1078.5385	0.0964	0	54	0.00028	▶1	U	■	■	■	■	R.YAGEGLAR.H
5605 ▶5	541.3848	1080.7551	1080.6059	0.1492	0	53	0.00047	▶1	U	■	■	■	■	K.YPPVIAK.V
7578	574.8892	1147.7639	1147.6117	0.1523	0	25	0.05	▶1	U	■	■	■	■	K.EFPPEK.F
7790 ▶6	577.9297	1153.8449	1153.6045	0.2404	0	49	0.00049	▶1	U	■	■	■	■	R.GSFPMAEK.I
8340 ▶7	586.0058	1169.9970	1169.5994	0.3976	0	33	0.0015	▶1	U	■	■	■	■	R.GSFPMAEK.I + Oxidation (H)
8573	589.4511	1176.8877	1176.6624	0.2253	0	35	0.043	▶1	U	■	■	■	■	K.FILLELR.H
10395	618.0392	1234.0638	1233.8304	0.2335	0	43	0.0053	▶1	U	■	■	■	■	R.YALLLLK.Y
17260 ▶1	724.4938	1446.9730	1446.7993	0.1737	0	59	5.1e-05	▶1	U	■	■	■	■	R.YDFVPIPSPR.K
18114 ▶4	739.5340	1477.0534	1476.8634	0.1900	0	82	6.3e-07	▶1	U	■	■	■	■	K.GTVVITLSSVLR.D
18138 ▶4	739.6177	1477.2207	1476.8108	0.4100	0	63	6.2e-06	▶1	U	■	■	■	■	R.DFDYLR.Q
18443 ▶3	745.9500	1489.8854	1489.8386	0.0268	0	73	3.8e-06	▶1	U	■	■	■	■	K.GTVVITLSSVLR.D
18496	498.3042	1491.8907	1491.7602	0.1305	0	24	0.019	▶1	U	■	■	■	■	K.FDPGHLEK.G
19465 ▶1	510.6492	1528.9257	1528.7756	0.1500	0	59	2.2e-05	▶1	U	■	■	■	■	K.EALVDHGEVFAFR.G
19473	765.5187	1529.0228	1528.7756	0.2472	0	93	4.2e-08	▶1	U	■	■	■	■	K.EALVDHGEVFAFR.G
19851 ▶2	773.1436	1544.2725	1543.8521	0.4205	0	61	5.2e-05	▶1	U	■	■	■	■	R.VYGPVITLYLR.K
20267 ▶8	781.1422	1560.2698	1559.8187	0.4511	0	90	8.2e-08	▶1	U	■	■	■	■	K.MISQSPHFSA.A
20458	523.6711	1567.9915	1567.8813	0.1102	0	27	0.018	▶1	U	■	■	■	■	K.VQEEIHWIK.H
20594	787.5444	1573.0743	1572.7654	0.3089	0	43	0.0039	▶1	U	■	■	■	■	K.EALVDHGEVFAFR.G
20989	795.6114	1589.2082	1588.7833	0.4249	0	32	0.012	▶1	U	■	■	■	■	K.DICQSPHFSA.V
21325	802.5887	1603.1628	1602.7760	0.3868	0	48	0.0012	▶2	U	■	■	■	■	K.EALIDHGEVFAFR.G
21390 ▶10	536.3278	1605.9617	1605.8232	0.1384	1	46	0.0014	▶1	U	■	■	■	■	K.EALIDHGEVFAFR.G

▶ 2 subsets and intersections (2 subset proteins in total)

MASCOT : Protein Family Result Report © 2010 Matrix Science

In the peptides table, we can see the complex relationship between the proteins. Each has at least one significant match to a unique sequence. Otherwise, it would become a sub-set of another family member and be relegated to the sub-sets list. If we want to see the relationship between any particular pair of proteins, we simply uncheck the boxes for the others and choose re-display

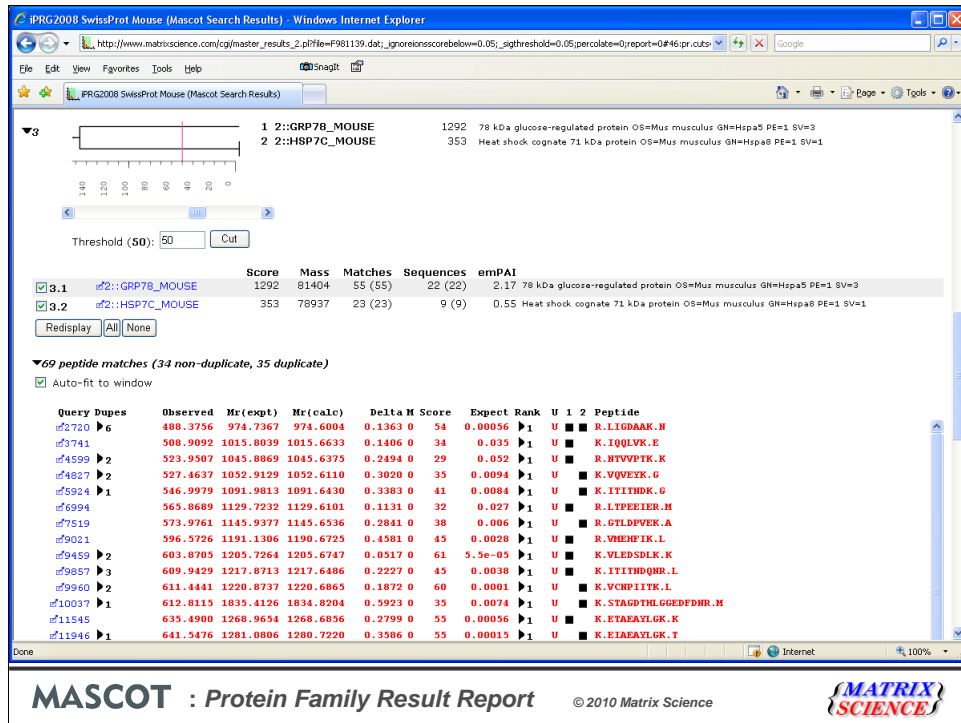
Please note that the use of bold face has changed in the new report. It now means that a match is significant. Not just that it's the first time we have seen a match to that particular spectrum. The meaning of red has not changed, it still indicates a top-ranked match



Here, for example, are a pair of proteins from family 3. You can immediately see 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).

If you look a little more closely, you will notice that HSP7C_MOUSE has a weak match to K.STAGDTHLGGEDFDNR.M. So, the evidence for both proteins being present comes down to a single residue. If S221 in HSP7C_MOUSE was an A, or A223 in HS71L_MOUSE was an S, then HS71L_MOUSE would be a sub-set protein. Interestingly, the stronger match is for the sequence found in the protein with fewer matches. This could be chance or it could be that the analyte sequence was essentially HSP7C_MOUSE but with an A at this position.

If you decide that this is not sufficient evidence for reporting both proteins, you can simplify the report by "cutting" the dendrogram using the slider control.



By cutting the dendrogram at a score of 50, HS71L_MOUSE becomes a subset of HSP7C_MOUSE, and we reduce the family to two members, each of which has a substantial number of unique sequences as evidence. These are different proteins that have sufficient homology to give some shared peptide matches. The strength of the new report is that it gives us a clear and simple overview, which allows us to browse for proteins of interest, as well as a tool for drilling down to the level of detail necessary to make decisions about whether there is evidence for any individual protein.

Protein Family Summary

- Index files are created and cached to speed loading in future
- Paged report to conserve memory
- Detailed information is shown 'on demand'
- Proteins grouped into families by means of shared peptide matches
- Only significant matches considered
- Hierarchical clustering within each protein family

To summarise