X!Tandem Pipeline

Automated analyses, filtering and export of X!Tandem MS/MS results

Benoit Valot and Olivier Langella
valot@moulon.inra.fr; langella@moulon.inra.fr
PAPPSO - http://pappso.inra.fr/

22 June 2012

Abstract

X!Tandem is an open-source software performing peptide/protein identification from MS/MS mass spectra. X!Tandem is fast and accurate, but the Global Proteome Machine (GPM) is relatively limited regarding the processing of identification results. **X!Tandem Pipeline** is an alternative to the installation of the GPM on local servers.

X!Tandem Pipeline performs database searching and matching on a list of MS/MS runs in one shot, using a list of easily user selected paramaters and databases.

X!Tandem Pipeline also performs filtering of data according to statistical values at peptide and protein levels. Moreover, redundancy of protein databases are fully filtered as follows :

- proteins identified without specific peptides compared to others are eliminated;
- proteins identified with the same pool of peptides are assembled;
- proteins are grouped by function (identified with at least one common peptide), and the specific peptides for each sub-group of proteins are indicated.

X!Tandem Pipeline allows to view and edit the filtered results, compute the false discovery rate, ... The results can be exported into TSV (Tab Separated Values) files.

Contents

1 Installation **2**
1.1 License . . . 2
1.2 Requirements . . . 2
1.3 Third party softwares for Windows and Mac . . . 2
1.4 Third party softwares for Linux . . . 2
1.5 Start X!Tandem pipeline . . . 3
1.6 Configuration . . . 3

2 X!Tandem analysis **4**
2.1 Parameters . . . 4
2.2 Running analysis . . . 4
2.3 Peak-lists . . . 4
2.4 Databases . . . 5

3 Processing the results **6**
3.1 Three modes of analysis . . . 6
3.2 Filter parameters . . . 6

4 View and edit the results **7**
4.1 Proteins List 7
4.2 Protein Details 8
4.3 Peptides List 8
4.4 Peptides Details 8

5 Save and Load X!Tandem Pipeline project **9**

6 Exporting the results **10**
6.1 Export parameters 10
6.2 Files *protein.txt 10
6.3 Files *peptide.txt 11
6.4 Files *compar.txt 12
6.5 Files *fdr.txt 13

7 Changelog **14**
7.1 "Myosine" branch 14
7.2 "Tubuline" branch 14
7.3 "Kératine" branch 14

1 Installation

1.1 License

Copyright (C) 2010 Valot Benoit and Olivier Langella
X!Tandem Pipeline program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.
This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

1.2 Requirements

X!Tandem Pipeline works on all platforms (Linux, Windows and Mac). Java 1.6 must be installed (it can be found : here).

1.3 Third party softwares for Windows and Mac

Download the X!Tandem executable from the GPM site.

1.4 Third party softwares for Linux

Ubuntu

- Add this repository.
- Install the *xtandem-cyclone* package.
- You can also install the *xtandempipeline* package to run **X!Tandem Pipeline**rather than used jlnp link.

Other distributions

- Download the sources and follow the instruction of compilation.

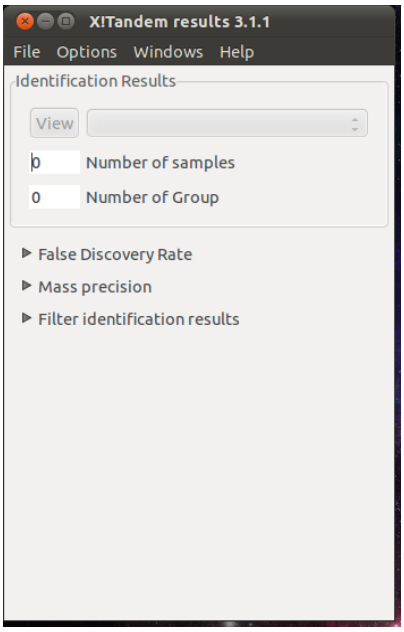


Figure 1: Principal window

1.5 Start X!Tandem pipeline

To run **X!Tandem Pipeline**, simply :

- Open X!Tandem pipeline by using this link
- Wait for the program to execute
- The principal window will appear (Fig 1)

1.6 Configuration

At the firt start, the application open the configuration path window:

- Open the menu *Option* → *Configuration Path* (Fig 2).
- Define the path to the X!Tandem executable
- Choose the folder where to store the X!Tandem parameters (or used default one).
- Choose the folder where the MS/MS data, the protein databases and the X!tandem results are stored

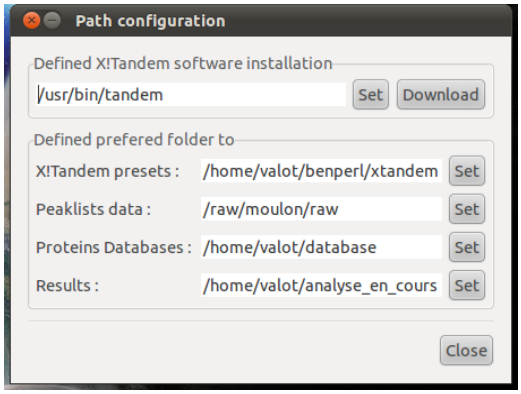


Figure 2: Configuration window

2 X!Tandem analysis

X!Tandem Pipeline allows you to analyze peak-lists files by searching a list of protein databases using the X!Tandem software. Three successive graphical boxes help you select first the mzXML files or other peak-lists, then the protein databases and finally the folder where the results will be stored. The databases must be protein ones, X!Tandem does not work on DNA databases.

2.1 Parameters

To perform database searching, you must create or edit a model XML file (stored in the xtandem models folder). Open the menu *Option* → *X!Tandem preset* (Fig 3).

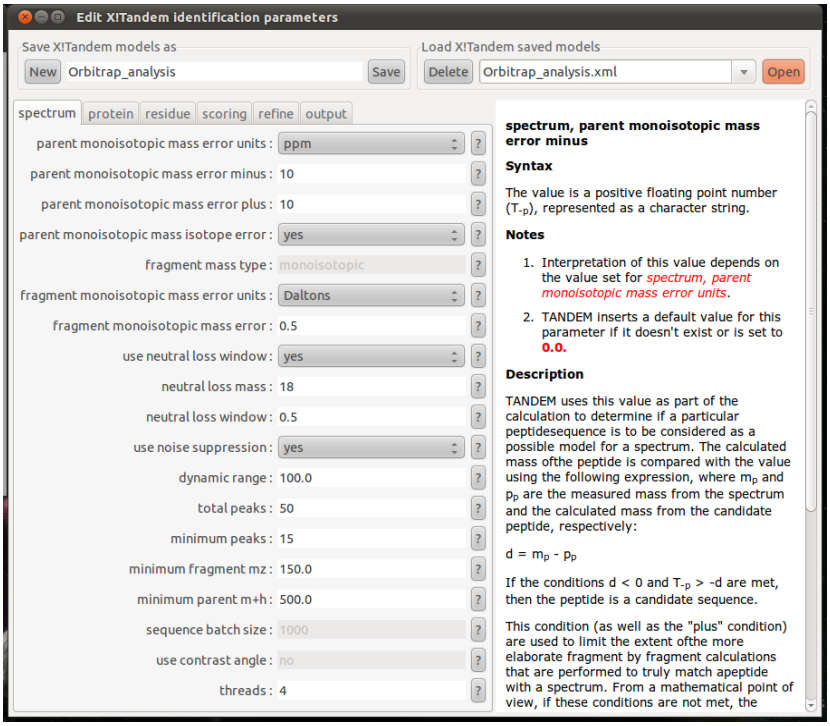


Figure 3: X!Tandem preset window

To use complete performance of your computer, specify the number of CPU in the model : spectrum → threads.

2.2 Running analysis

To perform analysis, start the menu *File* → *X!Tandem* → *Analysis*. Select on the window (Fig 4) :

1. Peak-list files to be analyzed (See 2.3)
2. Database files to be searched (See 2.4)
3. Searching parameters model (See 2.1)
4. Folder where to store the result files

2.3 Peak-lists

X!Tandem works with open peak-list files like mzXML, mgf, mzData, mzML or pkl files.

2.4 Databases

X!Tandem software uses only protein databases in fasta format. It doesn't work with EST[1] sequences. You can transform your database using our application *Protein database manager*, available here, or you can directly run it here.

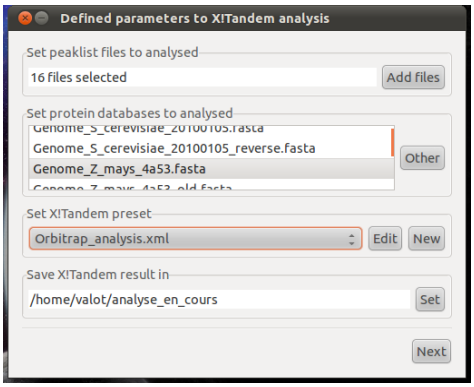


Figure 4: X!Tandem parameter window

[1]Expressed Sequenced tag

3 Processing the results

Warning: To process results, **X!Tandem Pipeline** needs to have X!Tandem result files (.xml) or Mascot result files (.dta). The names of the files are used as **sample names**.

3.1 Three modes of analysis

You can filter the MS/MS identification results and export them in three different modes : (menu *File* → *Load Result*)

Individual mode
Each MS/MS result file is processed individually.
You cannot perform comparison by using this process.

Combined mode
The MS/MS result files are combined in one result file, and this file is filtered / exported.
This mode is useful to compare different results.

Phosphopeptide mode
Same as the combined mode analysis except that only phosphopeptides are conserved and the result is oriented in order to validate phosphosites.

In all modes, you have to defined the filter parameters.

3.2 Filter parameters

The filter window (Fig 5) defines the automated filtering process parameters :

Add files
At this stage, you can add other MS/MS result files to the analysis. If two files have the same name, they are combined in one result file. Interesting if one wants to combine X!Tandem and/or Mascot results of the same LC-MS/MS run using different modification parameters or protein databases.

Peptide E-value
Defines the E-value above which a peptide is considered as valid.

Peptide number
Defines the number of valid unique[2] peptides necessary to validate a protein.

Protein E-value
Defines the E-value above which a protein is considered as valid.

- The protein E-value is the product of its valid unique peptide E-values and it is different from the protein E-values determined by X!Tandem.
- The values are expressed in log(E-value).

Sum to all
Defines how protein filter is performed when MS/MS results are combined :

No To validate a protein, the 2 parameters (peptide number and protein E-value) must be valid in at least one result. Interesting if one wants to compare SDS-PAGE-LC-MS/MS results, where peptides from a protein are in the same LC-MS/MS run.

Yes To validate a protein, the 2 parameters (peptide number and protein E-value) must be valid in the sum of all results. Interesting if one wants to compare 2DLC-MS/MS results, where peptides from a protein are split in different LC-MS/MS runs.

Contaminants
When you perform an analysis using different fasta databases, you can remove the result from one database by selecting this database. Interesting because it allows you to always include the same contaminant proteins during the database search, and because it removes the contaminant proteins from the results.

[2]Unique peptides are defined as peptides with different sequences. This excludes peptides with different modifications.

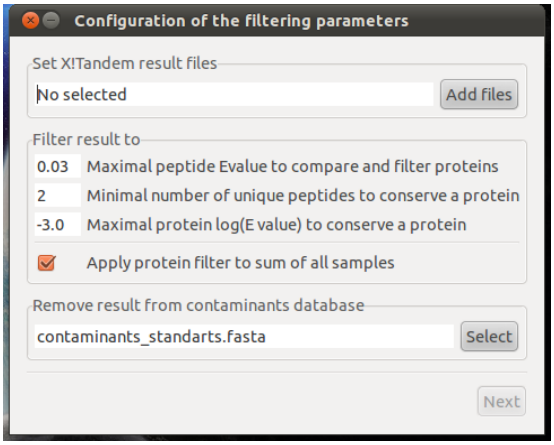


Figure 5: Filter window

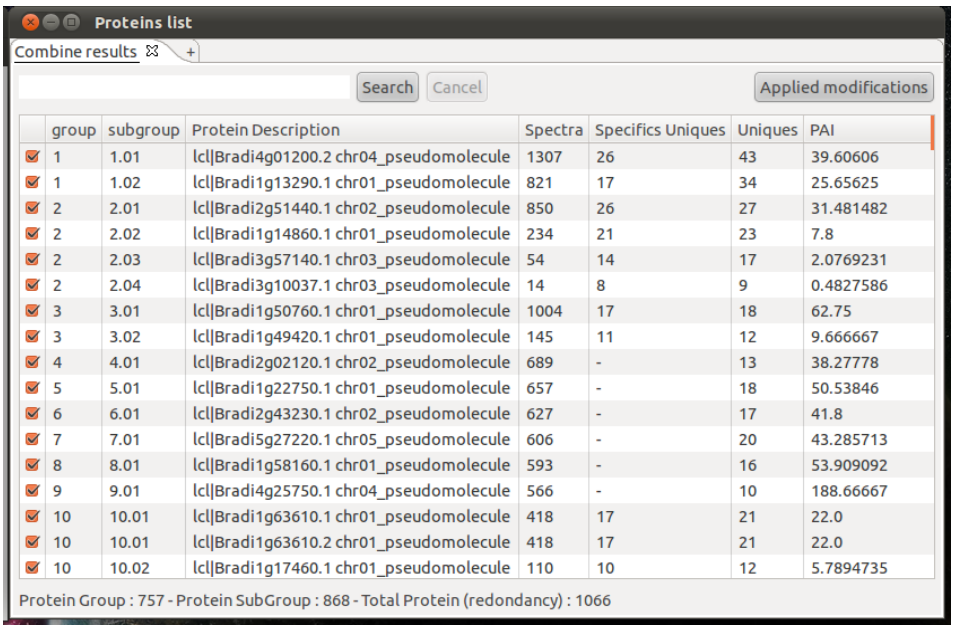
4 View and edit the results

After loading the results, you can select the result to view in the principal window (Fig 1). After this selection, you can navigate in this result in four different windows listed in the menu *Windows* :

4.1 Proteins List

View the list of protein identified on the result. For more details on column see Fig 11.

- Filter the protein by description;
- Clic on a protein to view the corresponding peptides list (see 4.3) and protein details (see 4.2);
- Unchecked protein for unvalidate the corresponding peptides;
- **Applied modification** to validate the edition.



group	subgroup	Protein Description	Spectra	Specifics Uniques	Uniques	PAI
1	1.01	lcl|Bradi4g01200.2 chr04_pseudomolecule	1307	26	43	39.60606
1	1.02	lcl|Bradi1g13290.1 chr01_pseudomolecule	821	17	34	25.65625
2	2.01	lcl|Bradi2g51440.1 chr02_pseudomolecule	850	26	27	31.481482
2	2.02	lcl|Bradi1g14860.1 chr01_pseudomolecule	234	21	23	7.8
2	2.03	lcl|Bradi3g57140.1 chr03_pseudomolecule	54	14	17	2.0769231
2	2.04	lcl|Bradi3g10037.1 chr03_pseudomolecule	14	8	9	0.4827586
3	3.01	lcl|Bradi1g50760.1 chr01_pseudomolecule	1004	17	18	62.75
3	3.02	lcl|Bradi1g49420.1 chr01_pseudomolecule	145	11	12	9.666667
4	4.01	lcl|Bradi2g02120.1 chr02_pseudomolecule	689	-	13	38.27778
5	5.01	lcl|Bradi1g22750.1 chr01_pseudomolecule	657	-	18	50.53846
6	6.01	lcl|Bradi2g43230.1 chr02_pseudomolecule	627	-	17	41.8
7	7.01	lcl|Bradi5g27220.1 chr05_pseudomolecule	606	-	20	43.285713
8	8.01	lcl|Bradi1g58160.1 chr01_pseudomolecule	593	-	16	53.909092
9	9.01	lcl|Bradi4g25750.1 chr04_pseudomolecule	566	-	10	188.66667
10	10.01	lcl|Bradi1g63610.1 chr01_pseudomolecule	418	17	21	22.0
10	10.01	lcl|Bradi1g63610.2 chr01_pseudomolecule	418	17	21	22.0
10	10.02	lcl|Bradi1g17460.1 chr01_pseudomolecule	110	10	12	5.7894735

Figure 6: Proteins List

4.2 Protein Details

View the protein sequence and coverage on a identified protein. To view this window, you must open it in the menu *Windows* → *Protein details*.

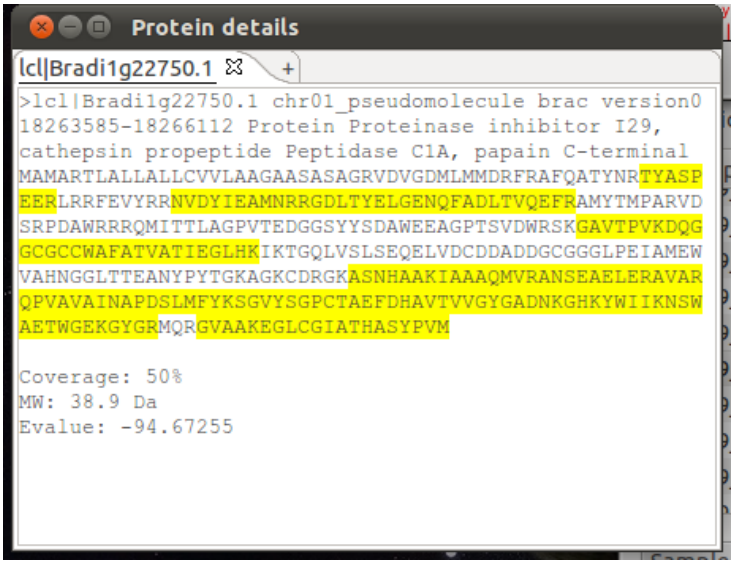


Figure 7: Protein details

4.3 Peptides List

View the peptides identified a protein. For more details on column see Fig 12.

- Filter the peptide by different options;
- Clic on a peptide to view the corresponding MS/MS spectra (see 4.4);
- Unchecked peptide for unvalidate it.

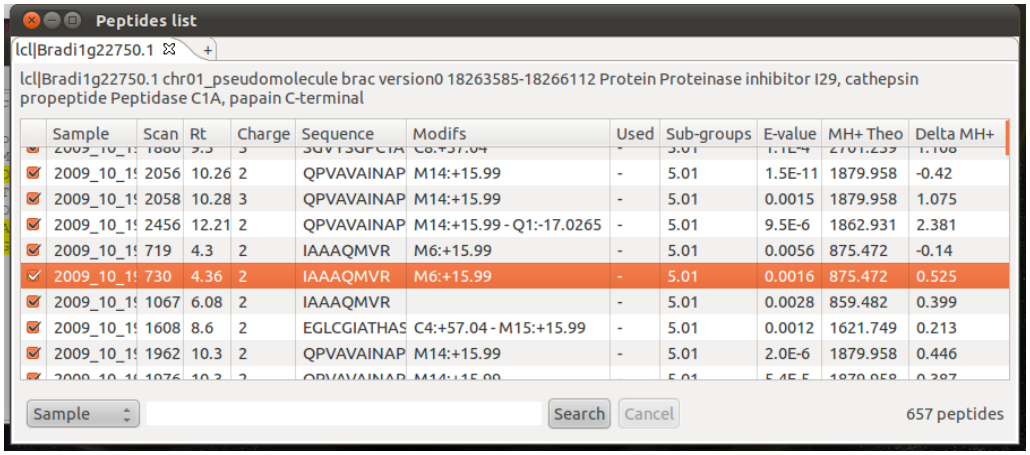


Figure 8: Peptides List

4.4 Peptides Details

View the MS/MS spectra of an identified peptide.

- Clic on spectra to zoom.
- Save MS/MS annotated spectra on png or svg.

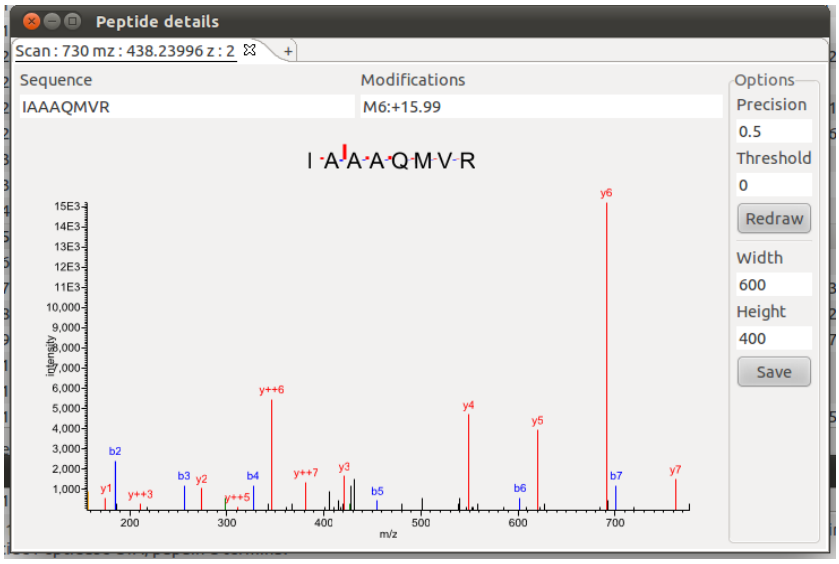


Figure 9: Peptides Details

5 Save and Load X!Tandem Pipeline project

You can save all the current results using menu *File* → *Save Project*, or load an previous one using menu *File* → *Load Project*. The extension of created files is **.xpip*.

6 Exporting the results

You can export the result in different formats in menu *File* → *Export*.

6.1 Export parameters

The export window (Fig 10) shows the different types of available exports :

Default
Creates TSV files containing identification results for proteins (*protein.txt) and peptides (*peptide.txt). When you perform a combined analysis, a *compar.txt file is created that contains the results of comparison between samples.

Fasta
Creates a fasta file for valid proteins.

PepNovo
Creates a XML file containing the peptide results to be removed for an automated *De Novo* interpretation in sequence using our DeNovo pipeline.

FDR
Creates two tabulated files containing the number of valid peptides or valid proteins for the different E-values in each database. Allows you to determine the E-value above which FDR value is acceptable.

Protic
Creates a PROTICdb compatible XML file, so you can store results in the PROTICdb proteomic database.

MassChroQ
Creates a MassChroQ compatible XML file, so you can perform quantitative analysis using our home-made software **MassChroQ** (to be released soon).

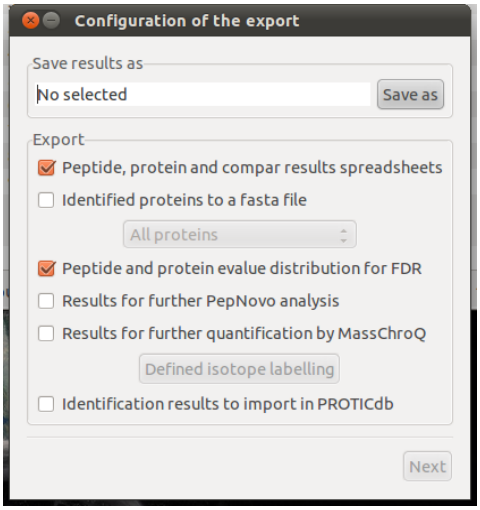


Figure 10: Export window

6.2 Files *protein.txt

The identified proteins are represented by sample (individual mode) or for all samples (combine/phosphopeptide modes) (Fig 11). Proteins are grouped by function.

Group Group to which the protein belongs. All the proteins in a group have at least one peptide in common.

Sub-group Sub-group to which the protein belongs. All the proteins in a sub-group are identified with the same valid peptides.

Description Protein description as it appears in the header of the fasta file.

log(E value) Protein E-value expressed in log.

- Statistical value representing the number of times this protein would be identified randomly.

- Calculated as the product of unique peptide E-values in the sample.

Coverage % of protein coverage.

MW Molecular weight of the protein expressed in KDa.

Spectra Total number of MS/MS spectra identified for the protein

Specifics Number of MS/MS spectra that are specific to the protein, compared to the other proteins of the same group (individual and phosphopeptide mode, see 3.1).

Specific uniques Number of unique peptide sequences specific to the protein, compared to other proteins of the same group (combined mode, see 3.1).

Uniques Number of unique peptide sequences identified for the protein.

PAI Protein Abundance Index :

- PAI estimates the relative abundance of the protein.
- PAI is calculated as the number of identified spectra divided by the number of theoretical peptides[3] of the protein.

Redundancy Number of proteins identified with the same pool of spectra. When there is redundancy, the above described parameters are shown only for the first protein of the subgroup (arbitrary chosen).Only the description of the other members of the subgroup is shown.

Position Position(s) of the phosphosite in the protein. This value is only reported in phosphosite mode (see 3.1).

Group	Sub-group	Description	log(E value)	Coverage	MW	Spectra	Specific uniques	Uniques	PAI	Redundancy
1	1.01	YHR174W ENO2 'Enolase II, a phosphopyruvate hydrat	-515.44476	85	46.8	1341	31	75	83.8125	
1	1.02	YGR254W ENO1 'Enolase I, a phosphopyruvate hydrata	-408.2716	82	46.7	969	19	63	74.53846	
2	2.01	YGR192C TDH3 'Glyceraldehyde-3-phosphate dehydro	-443.95404	90	35.6	980	28	62	70.0	
2	2.02	YJR009C TDH2 'Glyceraldehyde-3-phosphate dehydrog	-372.37628	90	35.7	892	14	52	59.466667	
2	2.03	YJL052W TDH1 'Glyceraldehyde-3-phosphate dehydrog	-297.6118	90	35.6	714	22	46	42.0	
3	3.01	YAL038W CDC19 'Pyruvate kinase, functions as a hom	-380.46527	87	54.4	835	55	60	32.115383	
3	3.02	YBL087C RPL23A 'Protein component of the large (60S	-46.771473	55	14.4	62	8	9	6.888889	* 2
3	3.02	YER117W RPL23B 'Protein component of the large (60S	-46.771473	55	14.4	62	8	9	6.888889	* 2
3	3.03	YOR347C PYK2 'Pyruvate kinase that appears to be mo	-22.56686	17	55.1	50	2	7	1.9230769	
4	4.01	YCR012W PGK1 '3-phosphoglycerate kinase, catalyzes	-330.4942	85	44.6	749	-	54	29.96	
5	5.01	YLR044C PDC1 'Major of three pyruvate decarboxylase	-382.53018	79	61.4	633	33	52	26.375	
5	5.02	YLR134W PDC5 'Minor isoform of pyruvate decarboxyl	-234.60945	74	61.8	300	24	36	13.043478	
5	5.03	YGR087C PDC6 'Minor isoform of pyruvate decarboxyl	-115.297455	29	61.4	218	1	18	9.083333	

Figure 11: Protein results

6.3 Files *peptide.txt

Identified peptides are grouped by group (Fig 12). One line corresponds to one MS/MS spectrum identifying one peptide that can be present in one or more proteins.

Group Group of the proteins containing this peptide.

Description Protein description if the peptide is specific to this protein.

Sample Name of the MS/MS run file.

Scan Scan number of the MS/MS run analysis.

Rt Retention time of the peptide.

Sequence Sequence of the peptide.

Modifs Modifications on the peptide. [4]

Valid Indicates whether the peptide was validated by the filter parameters or not.

Used Number of protein sub-groups in which the peptide is present.

on a total of Total number of protein sub-groups in the group.
Rq : If the peptide is specific, there is only $'-'$.

[3]Theoretical peptides correspond to the peptides resulting from the theoretical digestion of the protein sequence by trypsin and that are visible in mass spectrometry ($800 < MH < 2500$)

[4]For example, M2:+15.99 means that the mass of the second amino acid, which is a methionine, is increased by 15.99. This mass increase indicates that the peptide is oxidized.

Sub-groups Protein sub-groups where the peptide is present.

E-value Peptide E-value.

- Statistical value representing the number of times this peptide would be identified randomly.
- Calculated by X!Tandem with an empiric model.

Charge Charge level of the precursor.

MH+ Obs Monoisotopic observed mass for the peptide + one proton (MH^+)

MH+ Theo Monoisotopic calculated mass for the peptide + one proton (MH^+)

DeltaMH+ Error in the precursor mass between observed and theoretical data (Da)

Delta-ppm Error in the precursor mass between observed and theoretical data (ppm)

Position Position(s) of the phosphosite in the protein. This value is only reported in phosphosite mode (see 3.1).

Description	Sample	Scan	Rt	Sequence	Modifs	Valid	Used	on a total of	Sub-groups	E-value	Charge	MH+ Obs	MH+ theo	DeltaMH+
YHR174W EN(	LTQx_1_1p5	1786	10.4	AVYAGENFHHGDK		yes	-	-	1.01	4.4E-10	2	1444.4927	1444.661	-0.168
YHR174W EN(	LTQx_1_1p5	2146	12.31	IEEELGDK		yes	-	-	1.01	0.029	2	932.4527	932.4575	-0.0048
	LTQx_1_1p5	2714	15.25	GSEVYHNLK		No	2	2	1.01 1.02	0.073	2	1046.3328	1046.527	-0.194
	LTQx_1_1p5	2718	15.27	IGSEVYHNLK		yes	2	2	1.01 1.02	3.3E-5	2	1159.4777	1159.611	-0.133
	LTQx_1_1p5	2814	15.78	IGSEVYHNLK		yes	2	2	1.01 1.02	1.0E-4	2	1159.4528	1159.611	-0.158
YHR174W EN(	LTQx_1_1p5	3091	17.2	YAGENFHHGDKL		yes	-	-	1.01	7.0E-7	2	1387.5728	1387.639	-0.067
YHR174W EN(	LTQx_1_1p5	3138	17.44	AVYAGENFHHGDKL		yes	-	-	1.01	7.3E-9	2	1557.5267	1557.745	-0.218
	LTQx_1_1p5	3163	17.58	GENFHHGDKL		yes	2	2	1.01 1.02	0.0039	2	1153.4677	1153.539	-0.071
	LTQx_1_1p5	3198	17.76	KAADALLLK		yes	2	2	1.01 1.02	0.0023	2	942.5914	942.5987	-0.0073
YHR174W EN(	LTQx_1_1p5	3234	17.95	AVYAGENFHHGDKL		yes	-	-	1.01	2.7E-8	2	1557.5327	1557.745	-0.212
YHR174W EN(	LTQx_1_1p5	3554	19.61	DGKYDLDFKNPESDKSK		yes	-	-	1.01	2.9E-7	3	1987.1355	1985.946	1.19
YHR174W EN(	LTQx_1_1p5	3559	19.64	DGKYDLDFKNPESDKSK		yes	-	-	1.01	1.2E-7	2	1986.6127	1985.946	0.667
YGR254W EN(	LTQx_1_1p5	3715	20.46	TFAEALR		yes	-	-	1.02	0.027	2	807.49274	807.436	0.056
YGR254W EN(	LTQx_1_1p5	3758	20.69	DGKYDLDFKNPNSDK		yes	-	-	1.02	1.2E-6	2	1755.5728	1755.819	-0.246
YHR174W EN(	LTQx_1_1p5	3788	20.84	AVYAGENFHHGDKL		yes	-	-	1.01	0.0014	2	1557.5527	1557.745	-0.192
YGR254W EN(	LTQx_1_1p5	3806	20.93	DGKYDLDFKNPNSDK		yes	-	-	1.02	3.3E-4	3	1755.6554	1755.819	-0.163
YHR174W EN(	LTQx_1_1p5	3896	21.4	YDLDFKNPESDKSK		yes	-	-	1.01	4.2E-7	2	1685.5503	1685.802	-0.252
YGR254W EN(	LTQx_1_1p5	4202	23.0	YDLDFKNPNSDK		yes	-	-	1.02	3.4E-9	2	1456.5527	1455.675	0.877
YHR174W EN(	LTQx_1_1p5	4210	23.05	DGKYDLDFKNPESDK		yes	-	-	1.01	9.8E-5	3	1772.8455	1770.819	2.027
YHR174W EN(	LTQx_1_1p5	4216	23.08	DGKYDLDFKNPESDK		yes	-	-	1.01	3.1E-9	2	1770.6127	1770.819	-0.206
YGR254W EN(	LTQx_1_1p5	4240	23.19	YDLDFKNPNSDK		yes	-	-	1.02	7.8E-4	3	1455.3254	1455.675	-0.35

Figure 12: Peptide results

6.4 Files *compar.txt

All identified proteins are represented in a list: one protein per row, and one sample per column (Fig 13). The list of proteins is repeated 4 times, corresponding to the 4 parameters that are used to compare samples (see Type for details).

Group Protein group. Groups roughly correspond to the different functions.

Sub-group Protein sub-group. All the proteins of a sub-group are identified with the same valid peptides.

Description Protein description extracted from the fasta file.

MW Molecular weight of the protein (KDa).

log(E value) The log of protein's E-value.

- Statistical value representing the number of times this protein would be identified randomly.
- Calculated as the product of unique peptide E-values in all sample.

Type The item that is compared between samples.

Spectra Number of MS/MS spectra identified for the protein.

Specifics Number of specific MS/MS spectra identified for the protein compared to the other proteins belonging to the same group.

Uniques Number of unique peptide sequences identified for this protein.

PAI Protein Abundance Index (6.2).

Position Position(s) of the phosphosite in the protein. This value is only reported in phosphosite mode (see 3.1).

Groupe	Sub-group	Description	Redondancy	MW	Evalue	LTQx_1	LTQx_2	LTQx_3	LTQx_4	LTQx_5	LTQx_6	LTQx_7	Type
1	1.01	YHR174W ENO2 'Enolase II, a pl	-	46.8	-515.44476	134.0	146.0	148.0	146.0	172.0	166.0	168.0	Spectra
1	1.02	YGR254W ENO1 'Enolase I, a ph	-	46.7	-408.2716	100.0	109.0	105.0	110.0	120.0	117.0	122.0	Spectra
2	2.01	YGR192C TDH3 'Glyceraldehyde	-	35.6	-443.95404	101.0	112.0	111.0	117.0	121.0	124.0	122.0	Spectra
2	2.02	YJR009C TDH2 'Glyceraldehyde-	-	35.7	-372.37628	95.0	101.0	105.0	115.0	109.0	98.0	99.0	Spectra
2	2.03	YJL052W TDH1 'Glyceraldehyde	-	35.6	-297.6118	68.0	79.0	78.0	85.0	71.0	84.0	83.0	Spectra
3	3.01	YAL038W CDC19 'Pyruvate kina	-	54.4	-380.46527	80.0	91.0	92.0	89.0	98.0	93.0	90.0	Spectra
3	3.02	YBL087C RPL23A 'Protein comp	2*	14.4	-46.771473	5.0	8.0	3.0	4.0	6.0	3.0	4.0	Spectra
3	3.03	YOR347C PYK2 'Pyruvate kinase	-	55.1	-22.56686	4.0	5.0	7.0	3.0	5.0	4.0	5.0	Spectra
4	4.01	YCR012W PGK1 '3-phosphoglyc	-	44.6	-330.4942	79.0	90.0	85.0	77.0	82.0	91.0	90.0	Spectra
5	5.01	YLR044C PDC1 'Major of three p	-	61.4	-382.53018	69.0	71.0	75.0	69.0	69.0	79.0	70.0	Spectra
5	5.02	YLR134W PDC5 'Minor isoform	-	61.8	-234.60945	30.0	32.0	39.0	31.0	36.0	38.0	33.0	Spectra
5	5.03	YGR087C PDC6 'Minor isoform	-	61.4	-115.29745	21.0	23.0	28.0	23.0	29.0	27.0	27.0	Spectra
6	6.01	YKL060C FBA1 'Fructose 1,6-bis	-	39.5	-386.50046	76.0	74.0	78.0	80.0	80.0	76.0	77.0	Spectra
7	7.01	YDR385W EFT2 'Elongation fact	2*	93.1	-399.65784	50.0	48.0	51.0	49.0	49.0	48.0	50.0	Spectra
7	7.02	YLR048W RPS0B 'Protein compo	-	27.9	-67.70055	8.0	9.0	8.0	9.0	8.0	7.0	7.0	Spectra
8	8.01	YER091C MET6 'Cobalamin-inde	-	85.7	-368.73746	54.0	52.0	50.0	47.0	46.0	57.0	43.0	Spectra
9	9.01	YOL086C ADH1 'Alcohol dehydr	-	36.7	-210.06001	48.0	46.0	47.0	56.0	55.0	50.0	59.0	Spectra
9	9.02	YMR303C ADH2 'Glucose-repres	-	36.6	-87.34199	17.0	20.0	21.0	29.0	29.0	29.0	28.0	Spectra
9	9.03	YMR083W ADH3 'Mitochondrial	-	40.3	-74.741035	9.0	9.0	5.0	9.0	7.0	6.0	6.0	Spectra
9	9.04	YBR145W ADH5 'Alcohol dehyd	-	37.5	-25.602282	5.0	4.0	3.0	6.0	6.0	5.0	3.0	Spectra
10	10.01	YLL024C SSA2 'ATP binding prot	-	69.3	-327.67297	41.0	33.0	38.0	31.0	38.0	41.0	35.0	Spectra
10	10.02	YAL005C SSA1 'ATPase involved	-	69.5	-305.5065	33.0	28.0	33.0	29.0	32.0	37.0	29.0	Spectra
10	10.03	YER103W SSA4 'Heat shock pro	-	69.5	-111.07699	12.0	12.0	14.0	12.0	16.0	15.0	12.0	Spectra
10	10.04	YJL034W KAR2 'ATPase involved	-	74.3	-115.55334	10.0	13.0	12.0	9.0	8.0	11.0	12.0	Spectra
11	11.01	YBR118W TEF2 'Translational el	2*	49.9	-175.64958	42.0	37.0	39.0	44.0	44.0	42.0	40.0	Spectra
12	12.01	YDR050C TPI1 'Triose phosphate	-	26.7	-150.63466	34.0	38.0	39.0	46.0	42.0	42.0	42.0	Spectra
13	13.01	YHR183W GND1 '6-phosphoglu	-	53.4	-261.45218	35.0	33.0	35.0	33.0	37.0	34.0	33.0	Spectra
14	14.01	YDL185W TFP1 'Subunit A of the	-	118.4	-293.85382	31.0	34.0	38.0	32.0	30.0	34.0	27.0	Spectra
15	15.01	YLR355C ILV5 'Acetohydroxyaci	-	44.3	-209.51605	26.0	25.0	28.0	29.0	31.0	27.0	29.0	Spectra
16	16.01	YOR375C GDH1 'NADP(+)-depe	-	49.4	-240.86363	26.0	23.0	26.0	27.0	24.0	23.0	25.0	Spectra

Figure 13: Comparison results

6.5 Files *fdr.txt

This result file indicates the number of peptides with an E-value less than the E-value indicated in the fist column (Fig 14). You just have to divide the number of peptides in the reverse or decoy database by the number of peptides in the normal database to obtain the false discovery rate at each E-value level.
This method could be performed if :

- normal and reverse databases must be saved in different fasta files;
- X!tandem analysis have been performed with reverse option.
  In this case, the column corresponding to the normal and reverse search are indicated as *xtandem normal* and *xtandem reverse*, respectively.

FDR on peptide identification		
Evalue	Normal.fasta	Reverse.fasta
−14.5	0	0
−14	0	0
−13.5	1	0
−13	1	0
−12.5	1	0
−12	3	0
−11.5	3	0
−11	4	0
−10.5	4	0
−10	6	0
−9.5	6	0
−9	7	1
−8.5	7	1

Figure 14: FDR results

7 Changelog

7.1 "Myosine" branch

3.3.0 Grouping of sub-group have been changed to enhanced rapidity and corrected over-grouping on large dataset (thanks to M. Blein)
If you have very large dataset, we recommand to reload xtandem results to corrected error.

7.2 "Tubuline" branch

3.2.2 Corrected report of input parameter on X!Tandem output result (thanks to T. Greko).

3.2.1 Add new X!Tandem paramaters for multiple search of modifications in one analyse and calculation can now be performed on z ¿ 3.

3.2.0 Identification from Mascot dat file can now be imported and filtered. All work as X!Tandem result excepts that protein sequence can not be retrieved : PAI and coverage are absent.
Correction of FDR calculation from Reverse/Decoy search.

7.3 "Kératine" branch

3.1.5 Add support of phosphorylation neutral loss and enhanced ETD detection on MS2 spectra.
Correction of MassChroQ export.

3.1.4 Add support for viewving ETD spectra after automatic detection.

3.1.3 Corrected bug of xtandem preset. Refine analysis was never start instead refine param is set to yes.
Adds a new annotated spectrum renderer and bug fix on ODS export.

3.1.2 Add export results on Open Document Spreadsheet (.ods) file.
Correction of bugs (Grouping, PepNovo export, ...).

3.1.1 FDR computation are now compatible with reverse option of X!Tandem.

3.1.0 Algorithm of grouping have been completly rewritten :

- Older project must be refiltered to be properly grouped.
- Phosphopeptide filtering have been enhanced to correspond to :
  - SubGroup represents the number of phosphosites
  - Group represents the number of phosphoproteins
- Configuration file have been modified and must be parameter again