MaxReportDocumentation

An enhanced proteomic result reporting tool for MaxQuant
Current version: 2.1; Last updated: 2016-02-26

By Dr. Tao Zhou
Contact: zhoutao@njmu.edu.cn

Contents

1. Description

MaxReport is an enhanced proteomic result reporting tool for MaxQuant [1]. Currently, its main missions are to optimize the results of MaxQuant and to provide additional functions for protein N-terminal modifications, isobaric labeling quantification and descriptive statistical analyses.
Highlighted features:

- Results are well organized and indexed.
- Optimized and minimal protein reporting.
- Generates site tables for protein N-terminal modifications.
- Provides general descriptive statistical analyses and figures.
- Supports isobaric labeling quantification at the protein, peptide and site levels.

Change log:

- Supports "I=L" setting and other unknown amino acid in the sequence database.
- Fixesxlsxwriter bug for handling NAN and INF values.
- Supports MaxQuant 1.2.0.18 with no MS/MS IDs in the protein and site files.
- Detects MaxQuant version for bug report automatically.
- Improves multiple files selection.

2. Resources

2.1. To download MaxReport package

Table 2.1 List of MaxReport package

Version	Description	Links
MaxReport 2.1	python script (source code)	Download
MaxReport 2.1	Windows EXE (standalone)	Download

*MaxReport package contains MaxReport_CMD (core program), MaxReport_GUI (GUI shell for MaxReport_CMD),pdf documentation and resource files.
*The current version of MaxReport tool was developed and tested using ActivePython (version 2.7.5.6) under a Windows system. In addition, NumPy (version 1.8.1) was used for mathematical calculation. Pygal (version 1.7.0) was used for creating statistical figures in scalable vector graphics (SVG) format. XlsxWriter (version 0.6.4) was used to generate combined result tables in Excel format (xlsx). wxPython (version 3.0) was applied for GUI development. PyInstaller (version 2.1) was used to distribute the MaxReport script into a standalone executable program under Windows. For evaluating the development of MaxReport tool, a package of the above software or modules is provided: [Download].
*Please note: MaxReport is a free and open source software. Please read the MaxReport license if you wish to modify and redistribute the codes.

2.2. Reporting configuration files for different versions of MaxQuant

Table 2.2 List of reporting configuration files

MaxQuant Version	Links
1.2.x	Download
1.3.x	Download
1.4.x	Download
1.5.x	Download

2.3. Fasta header parsing rule files for extracting protein accession, gene name and protein description

Table 2.3 List of header parsing rule files

Protein header format	Links
UniProt	Download
RefSeq	Download
Ensembl	Download
Vertical bar separated	Download
Space separated	Download
Single accession	Download

2.4. Correction matrix files for isobaric labeling quantification

Table 2.4 List of correction matrix files

Isobaric labeling	Links
TMT6	Download
iTRAQ4	Download
iTRAQ8	Download
iTRAQ8 (Phe)	Download

3. CMD usage

Since Python is a cross-platform programming language, the MaxReport_CMD script (source code) could be directly used in multiple operating systems (including Windows, Linux and Mac OS X) provided that Python and the above listed modules are installed. MaxReportcan be easily used via command line interface (CLI), which will facilitate the processes of multiple MaxQuant projects. By defining a specific reporting configuration file, MaxReprot can support different versions of MaxQuant. The optimized reporting configuration files, header rule files and correction matrix files can be downloaded in the above tables of resources.

Commond line usage:
MaxReport_CMD.py [-h] -r FRULES [FRULES ...] -s FSEQS [FSEQS ...] [-e EXPTMP] [-a] [-c CORMTX] [-m MINITH] [-x] [-v] sdirrcfg
The detailed description of the parameters are listed in Table 3.1.

Table 3.1 List of arguments

Argument	Type	Default value	Description
sdir	positional, required	#	source folder for a MaxQuant project
rcfg	positional, required	#	configuration file for result reporting
-r, --frules	required	#	rules files for parsing fasta sequences
-s, --fseqs	required	#	protein sequence files
-e, --exptmp	optional	*summary.txt， #	Experimental design template file
-a, --allpep	optional	False	Use all peptides for quantification
-c, --cormtx	optional	None， #	correction matrix file for isobaric quantification
-m, --minith	optional	5	minimum threshold value for isobaric quantification
-x, --xlsxop	optional	False	write combined identification and quantification results to xlsx files
-v, --version	optional, standard	-	show the version information and exit
-h, --help	optional, standard	-	show the help message and exit

*By default, MaxReport extracts experimental design information in summary.txt. However, earily versions of MaxQuant did not write experiments in this file. Thus, a specified experimental template file is recommended.
#Note: all input files should be given as absolute paths.

4. GUI usage

In MaxReport package, MaxReport_GUI provides a GUI (graphical user interface) shell to run MaxReport_CMD (the core program in command line interface). As shown in Figure 4.1, MaxReport_GUI contains 12 components for four required arguments, five optional arguments and three control buttons:

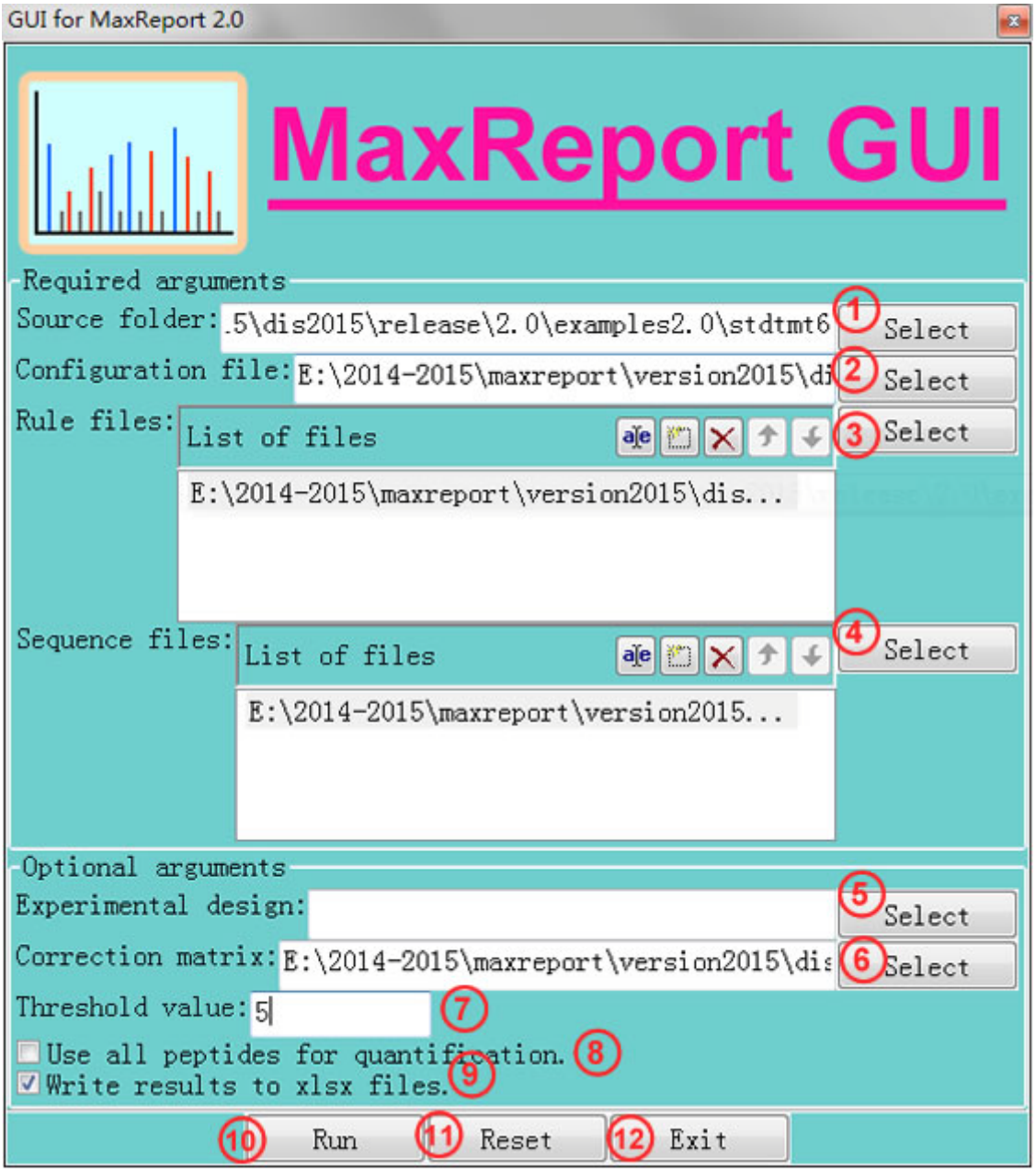

Figure 4.1. Main frame of MaxReport_GUI

1. Press "Select" to choose the directory of a MaxQuant project.
2. Press "Select" to choose the configuration file for MaxQuant reporting.
3. Press "Select" to choose rule files for parsing protein sequence files.
4. Press "Select" to choose sequence database files. Make sure the sequence files and the rule files are sorted consistently.

5. (Optional) Press "Select" to choose the experimental design file.
6. (Optional) Press "Select" to choose the correction matrix file for isobaric labeling quantification.
7. (Optional) Type the minimum threshold value for isobaric labeling quantification.
8. (Optional) Check to use all peptides (only unique peptides are used in default) for isobaric labeling quantification.
9. (Optional) Check to write combined results to xlsx format.
10. Press "Run" to send the above arguments to MaxReport_CMD.
11. Press "Reset" to clear all settings.
12. Press "Exit" to close the window.

After all required fields are filled and the "Run" button is pressed, the main window will be hidden and a pop-up window (Figure 4.2) will show up, which indicates that MaxReport_CMD starts successfully.

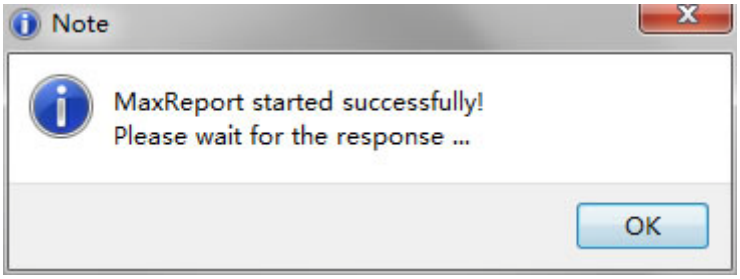

Figure 4.2. Start message

Meanwhile a console window (Figure 4.3) with black background will also be shown that indicates the running status of MaxReport_CMD. Please do not close this window!

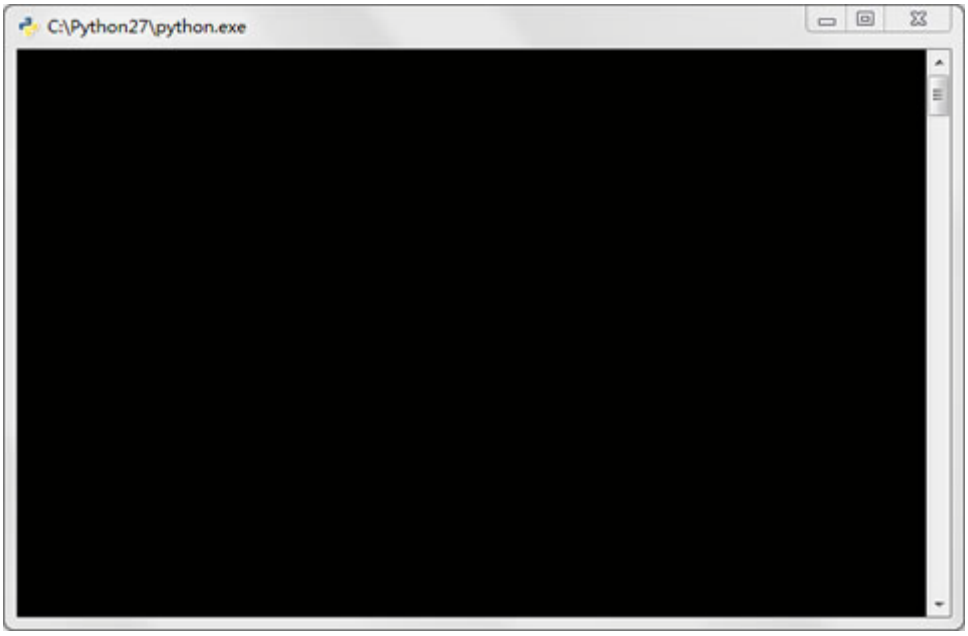

Figure 4.3. Running window of MaxReport_CMD

After the black console window is closed automatically. A message dialog (Figure 4.4) will be shown to tell that MaxReport is finished.

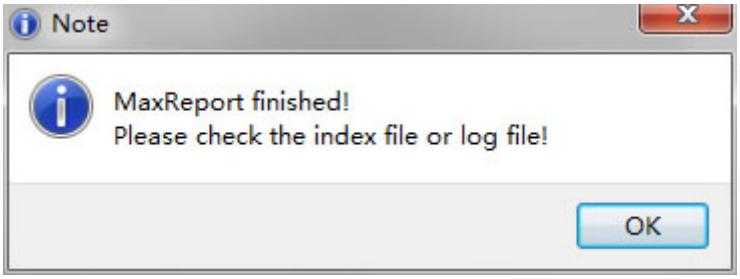

Figure 4.4. Finish dialog

If all results are correctly generated, a html based index page will be opened automatically (Figure 4.5). However, if there is no index file generated, the users should check the log file for simple errors. For complex errors, please contact the developers via the online feedback system.

Figure 4.5. Index page

5. Examples

Table 5.1 List of examples

Name	Description	Links
Macaque sperm	Our previous published large-scale proteome of macaque sperm [2]	Browse
Standard TMT6	A TMT6 labeling standard experiment spiked with four exogenous proteins with known expression ratios. (Data accession: PXD000001 at ProteomeXchange.) [3]	Browse

Examples of different versions of MaxQuant:
V1.2.0.18V1.2.2.5V1.3.0.5V1.4.1.2V1.5.2.8V1.5.3.8
Example of Ile-Leu setting:
I=L setting

6. FAQs

1. Does MaxReport support different versions of MaxQuant?
Response: MaxReport supports MaxQuant 1.2.x ~ 1.5.x, and will update timely if new versions of MaxQuant are released. Since the scoring method and table titles are various among different versions of MaxQuant, the corresponding reporting configuration file must be specified for MaxReport. The protein and site files of MaxQuant 1.2.0.18 do not contain MS/MS IDs.Thus,MaxReport 2.0 and the early versions do not support MaxQuant 1.2.0.18. MaxReport 2.1 fixes this bug and suppportMaxQuant 1.2.0.18.

2. Does MaxReport support multiple sequence databases in a project?
Response: The users can add multiple fasta files for database search in MaxQuant. MaxReport 1.0 only supports single sequence database (with only one fasta header format). MaxReport 2.x supports multiple databases with different fasta header formats. Pleas provide fasta parsing rule files and sequence files in the same order.

3. What is the "KeyError"?
Response: MaxReport loads fasta files based on the corresponding parsing rule files. A "KeyError" error indicates that a wrong parsing rule file is provided. Please check the header format and choose the correct rule file.Moreover,please make sure that the rule as applied in MaxQuant itself is the same with MaxReport.Old versions of MaxQuant may extract 'sp|A0AV96|RBM47_HUMAN' asUniProt protein ID which should be 'A0AV96' (the rule for MaxQuant could be modified via the embedded Andromeda configuration). Thus, for this case, the rule file of 'sp.rules.cfg' (up to first space) should be used to obtain protein ID like 'sp|A0AV96|RBM47_HUMAN' in MaxReport.

4. Which modification does MaxReport support?
Response: MaxReport supports all modifications as defined in MaxReport. Thus, MaxReport will automatically export all modifications as considered in a MaxQuant project.

5. How to resolve otherunpredictable errors?
Response: MaxReportlog file records every steps of analyses. First, please check the log file and narrow down the problem to sequence database, configuration rule or search results. Make sure all the parameters are correct. If unknown errors still occurs, please submit the log file to our online feedback systemfor professional solutions.
*To report bugs or other comments, please also visit our feedback system!

7. References

1.MaxQuant publication:

[1] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008 Dec;26(12):1367-72. doi: 10.1038/nbt.1511. Epub 2008 Nov 30. PubMed PMID: 19029910.

2.Example datasets:

[2] Zhou T, Wang G, Chen M, Zhang M, Guo Y, Yu C, Zhou Z, Si W, Sha J, Guo X. Comparative analysis of macaque and human sperm proteomes: Insights into sperm competition. Proteomics. 2015 May;15(9):1564-73. doi: 10.1002/pmic.201400248. Epub 2015 Jan 23. PubMed PMID: 25545774.

[3] Gatto L, Christoforou A. Using R and Bioconductor for proteomics data analysis. BiochimBiophysActa. 2014 Jan;1844(1 Pt A):42-51. doi: 10.1016/j.bbapap.2013.04.032. Epub 2013 May 18. Review. PubMed PMID:23692960.