MQQC - MAXQUANT QUALITY CONTROL

WRITTEN BY HENRIK ZAUBER

Contents

1. MQQC introduction

MQQC aims in providing easy quality control of LC-MSMS derived proteomic data. It therefore evaluates propabilites of identified peptides returned from automated MaxQuant Analysis[] and returns output in form of .txt, pdf and html output. Prerequisites for MQQC installation is a running Perl and R distribution (see section 4, p.3). MQQC itself is available as an R-package an can be either executed from a running R-console or by a windows executable that can be provided on demand.

1.1. **MQQC algorithm.** MQQC is designed to scan a specified folder for new raw files. If a new raw file is detected it will be moved to a separate analysis folder (Figure 1.1). The Parameter ID given in the name of the raw file will be evaluated to setup the correct MaxQuant settings and the MaxQuant analysis is initiated (Figure 1.2). MaxQuant Result Files are subsequently read and used to perform the quality control analysis (Figure 1.3). PDF and text output are archived, provided temporally on a website and if enabled, an e-mail noticiation is send to the user controlled by a user ID given in the raw file name (Figure 1.4 and .5).
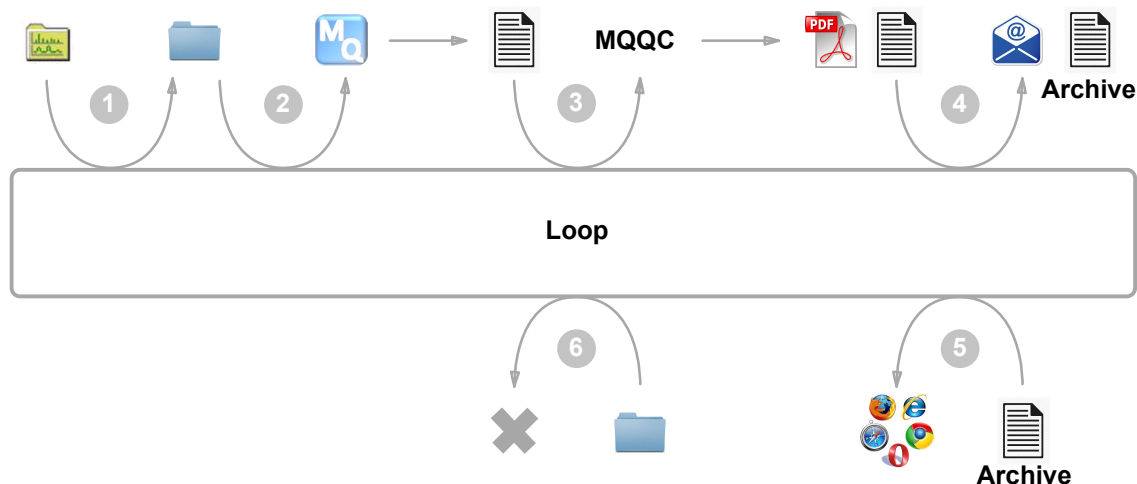
1.2. **MQQC naming nomenclature.**

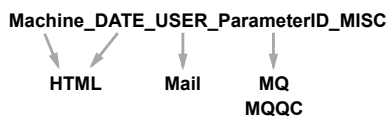FIGURE 1. Sceme of the MQQC Algorithm. Details are given in the text.

FIGURE 2. caption

2. GUI SETTINGS

MQQC provides a graphical user interface for an easy access to user specific settings. This includes required path settings, MQQC parameters such as Parameter IDs and thresholds or Mailing support. The GUI will always launch when MQQC is initiated.

2.1. **General Settings.**

2.2. **MaxQuant Folder.** The MaxQuant Folder setting is the path to a folder containing a working MaxQuant Executable (MaxQuant.exe). The current MQQC 1.22 was tested with the version 1.305. Later versions of MaxQuant might work as well.

2.3. **Analysis Folder.** MQQC will continuously scan this folder for new LC-MSMS raw files and set up a new MaxQuant analyis, if possible. Supported file formats are in principle all files that are supported by MaxQuant. Nevertheless, the current version of MQQC (1.22) works only with the Thermo Raw File format (*.raw). Later versions of MQQC might include support for MzXML, Sciex WIFF and Bruker files.

2.4. **HTML Path.** In the HTML path MQQC will write a HTML output giving quick access to the most recent MQQC runs. Additionally, timelines and comparisson of selected features will be provided as pdf for each set up machine.

2.5. **E-Mail IDs.** This modul can import a list of e-Mail addresses with corresponding user IDs. ¿@ Import¡ will ask for a tab delimited text file containing in the first column user IDs and in a second column the corresponding e-mail addresses. With ¿@ Edit¡ the loaded e-mail list can be inspected, edited and saved.

2.6. **MQQC Parameter Table.** The Parameter Table contains relevant information for each MQQC analysis given by the so called Parameter ID. In MQQC each single analysis is controlled by this Parameter ID. This ID must be given in the name of a raw file. From the Parameter Table MQQC then takes the user specific settings for the MaxQuant analysis (either setting used species in a standard run or using even a separated par.xml setting file) and set thresholds for the subsequent quality control analysis. The differenti entries inf the Parameter Table are given in table 1.

3. Output

3.1.

3.2. HTML-Report.

4. Installation

4.1. Installation of R.
  (1) Download the latest R Version from `http://www.r-project.org`.
  (2) Follow the general R installation procedure.
      Important: R must be installed using **Administrator privileges**. Windows 7 users need to make sure, that R is registered in the Registry (enabled in standard R installation).
  (3) **MacOsX:** Users of MacOsX need additional libraries of Tcl/Tk 8.5.5 for X11 in order to use the GUI. You can get the installation package under the url: `http://cran.r-project.org/bin/macosx/tools/`.

4.2. Installation of Perl *optional. Perl is required for sending mail notifications of finished mqqc runs.*
  (1) Download and install a Perl distribution. Links to available installation packages can be find under the url: `https://www.perl.org/get.html`

4.3. Installation of MQQC.
  (1) Start a new R session.
  (2) Paste the following code in the R console:

5. Startup

MQQC can be started from R by running the command ¿folder.observe()¡. All other settings are set in a graphical user interface. After a fresh installation of MQQC the following window will appear:

In case you have a backup of your MQQC settings you can load it here, if not select no.

WRITTEN BY HENRIK ZAUBER

TABLE 1. Entries in the Parameter ID Table. These Entries are linked to an Parameter ID, given in the name of each file to process with mqqc.

Table Entry	Table Entry name	Entry Explanation	Used Subset	InputType	MQQC Type	Score in MQQC 1.22	Required in MQQC 1.22
Abbreviation	Parameter ID	ID that will be used by MQQC to match it with the ID from the raw file name. Corresponding values in the same row will be used for subsequent MaxQuant and quality control analysis.	-	string	-	-	yes
Full.name	misc	User specific entry about the parameter ID. Not used for MQQC analysis.	-	string	-	-	no
Fasta	MaxQuant Setting	Path to fasta file containing species specific protein sequences.	-	path	-	-	yes
Xml	MaxQuant Setting	Alternative mqpar.xml file for setting up MaxQuant.	-	path	-	-	optional
quan.pep.total	Total Peptide Counts	Threshold for totally identified peptides in each run.	all peptides	Best Value	-	yes	yes
quan.msms.min	Peptide Counts/min	Peptides identified per minute in the inner elution time quantile (25-75 %) of each run.	peptides in 25-75 % elution time quantile	Best Value	MS	yes	yes
mass.error.uncal	uncalibrated mass error	Threshold for tolerated uncalibrated mass error. Input requires 2 values separated by space. First is the optimal expected mass error. Second is the ± distance, used to calculate a confidence range from good to poor.	peptides in 25-75 % elution time quantile	Best Value + range distance	MS	yes	yes
score	Andromeda Score	Best expected median Andromeda Score	all peptides	Best Value	-	yes	yes
msmsEff	MSMS Efficiency	Estimated successfull identification of MSMS spectra.	peptides in 25-75 % elution time quantile	Best Value	MSMS	no, obsolete	yes
quan.duplicates.msms	Duplicated Peptide Identifications	Multiple Identifications of the same peptide species is supposed to indicate bad LC performance. This value is obsolete since MQQC version 1.23	peptides in 25-75 % elution time quantile	Best Value	-	yes	yes
ret.width	Retention time peak width	Best expected Median Retention Time Peak Width.	peptides in 25-75 % elution time quantile	Best Value + range distance	LC	yes	yes
ret.peak.shape	Retention time peak shape	Best expected Median Peak Shape. This Value is based on a ratio of the two peak sections that are separated by the highest point of a peak. The smaller a ratio, the better.	peptides in 25-75 % elution time quantile	Best Value + range distance	LC	yes	yes
quanRet50ratio	Elution Profile Symmetry	Evaluates the symmetry of the inner quantile elution time profile. Based on the Ratio (25 to 50%)/(50 to 75 %). The Smaller the better.	all peptides	Best Value	LC	yes	yes
quanRetRSD	Elution Profile Relative Standard Deviation	Evaluates the stability of the elution profile using the relative standard deviation of the elution time density profile.	peptides in 25-75 % elution time quantile	Best Value + range distance	LC	yes	yes
qunaRetSlope	Elution Profile Slope	Evaluates the stability of the elution profile, by calculating the slope of the density profile in the inner retention time region.	peptides in 25-75 % elution time quantile	Best Value	LC	yes	yes
quanRetSlope	Elution Profile Slope	Evaluates the stability of the elution time profile of the density profile in the inner retention time quantile region.	peptides in 25-75 % elution time quantile	Best Value	LC	yes	yes
msmsQuantile	Peptide Fragment Intensity	Expected Median Intensity of Fragment Scans	peptides in 25-75 % elution time quantile	Best Value + range distance	MSMS	yes	yes
msmsCount	Peptide Fragment Counts	Expected Median Fragment Counts per MSMS	all peptides	Best Value + range distance	MSMS	yes	yes
ProteinCoverage	Protein Coverage	Expected Protein Coverage for the Protein Standard.	all peptides	Best Value	MS (Protein Standard)	yes	yes
Intensities	MS1 Intensity	Median Intensity among all identified MS1 peaks	all peptides	Best Value + range distance	MS (Sample Standard and Sample)	yes	yes

(a) MQQC GUI Main Settings
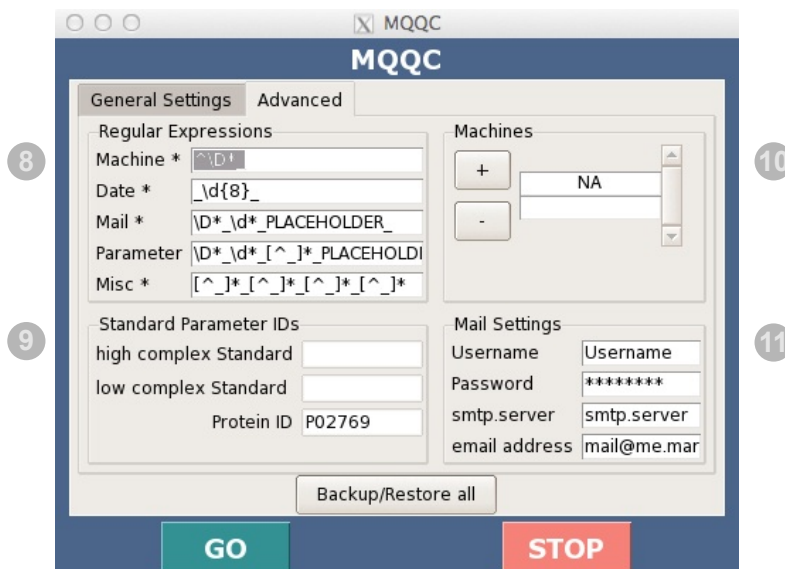
(b) MQQC GUI Advanced Settings

FIGURE 3. MQQC Graphical User Interface. The following options are integrated in MQQC 1.22: 1-3 Required paths pointing at MaxQuant location, MQQC Analysis Folder and Html Output location. 4. Entry option for E-Mail addresses and correspedning user IDs. 5. Control Buttons to modify Parameter ID Table 6. Running options 7. Backup and Restore option of MQQC settings. 8. Regular Expression patterns for matching the raw file name according to a defined nomenclature. 9. Definition of Standard Parameter IDs. 10. Naming of individual Mass Spectrometers used with MQQC. 11. Server settings for e-mail option.