Meta4P

MetaProteins-Peptides-PSMs Parser

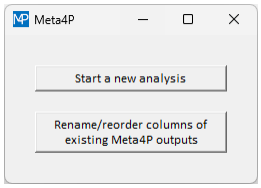
User guide

v.1.5.6 (February 2026)

0. Launch the application and start a new analysis

Double-click on the Meta4P.exe (Windows) file to launch the application. The first time the program is launched, the operating system will ask for your explicit approval to execute the application. Click on "More info" to proceed with security checks and then on "Run anyway".

The opening window (see the image below) allows you to "**Start a new analysis**". Alternatively, if you have previously downloaded Meta4P outputs and just want to rename and/or reorder sample column headers in them, click on "**Rename/reorder sample columns of existing Meta4P outputs**" to go directly to the last Meta4P window (see section 7).

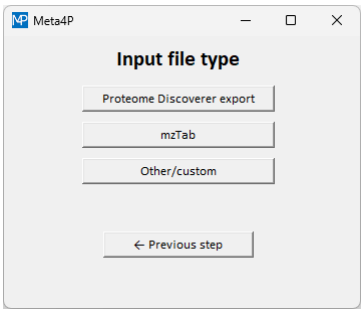


1. Identification/quantification input

As a first input, you have to provide a file containing identification and (label-free) quantification information. Meta4P can handle different file types and data levels, for a total of nine combinations, as detailed below.

1.1. Identification/quantification input: file type

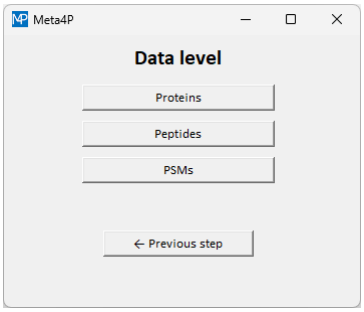
This window allows you to select the desired file type, clicking on one of the following buttons: "**Proteome Discoverer export**", "**mzTab**" or "**Other/custom**" (see the image below).



The three types of files accepted are: Proteome Discoverer files (in xlsx or txt format), mzTab files (one of the standard formats for proteomic data exchange) and generic tabular files (in xlsx, txt or another tab-separated format).

1.2. Identification/quantification input: data level

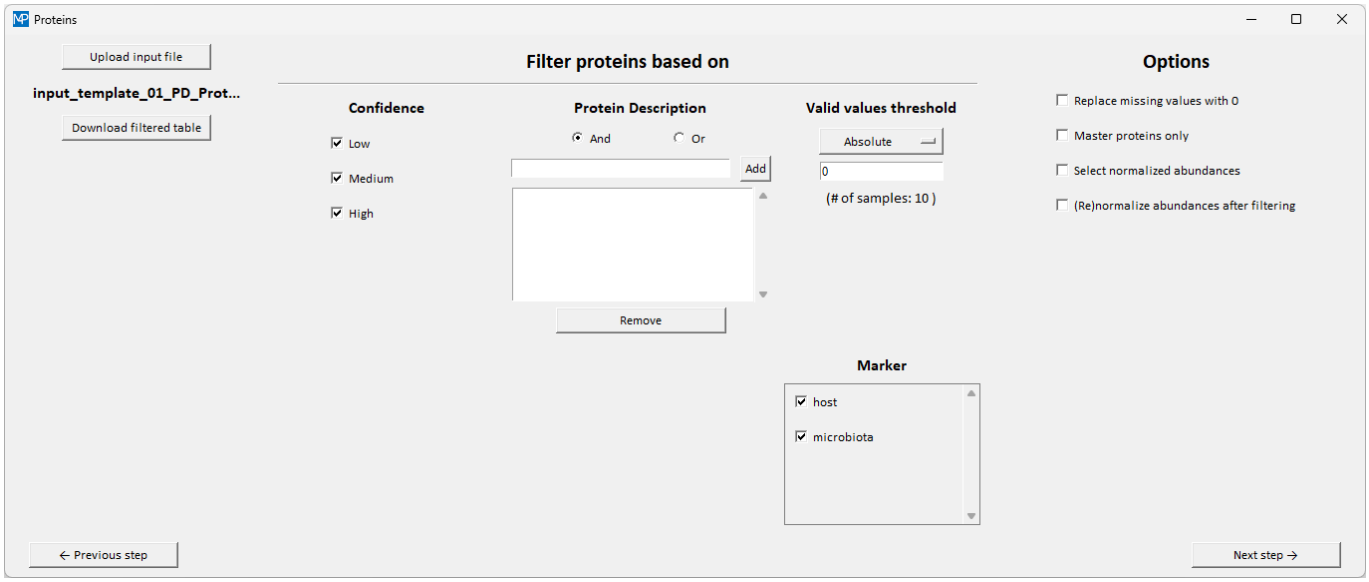
Once the file type has been selected, another window allows you to choose the data level (see the image below). The following buttons are shown, each corresponding to a different level of identification and quantification data: "**Proteins**" (protein identifications with MS1-based quantitative data), "**Peptides**" (peptide identifications with MS1-based quantitative data) and "**PSMs**"(peptide-spectrum matches to be used for spectral counting quantification).



1.3. Proteome Discoverer export - Proteins

When the "Proteome Discoverer export" and "Proteins" options are sequentially selected, protein identification and quantification data are retrieved from a "Proteins" file exported from Proteome Discoverer, available in one of the following formats: xlsx (Microsoft Excel) or txt (tab-separated values). The input file must contain one "Accession" column and at least one "Abundance" column (whose header must start with "Abundance: F"). If you try to upload an input file with the wrong structure or format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_01_PD_Proteins.xlsx* is available for download). Once the file is uploaded, the file name is shown under the upload button and the window is populated with several filtering options based on the file content (see the image below).



Proteins can be filtered based on the number/percentage of valid values ("**Valid values threshold**"), so that only proteins with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all proteins pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.

Other optional filters might also be available (if the corresponding columns are present in the input file):

- **Confidence**: only proteins with the selected level(s) of statistical confidence (low, medium and/or high) are kept in the output, based on the checkbox(es) checked.
- **Protein Description**: to select only those proteins which contain a specific text (e.g., a protein name or an organism name) in their "Description" column, type the text of interest in the textbox (be aware

that the filter is case sensitive) and click on "**Add**". Multiple texts can be typed and added sequentially; in this case, you can choose between two boolean operators, "**And**" and "**Or**", to determine whether all the texts added or only one of them must be present in the string, respectively, so that a protein passes the filter.

- **Marker**: marker names included in the "Marked as" column of the input file, usually indicating which of the protein database(s) used for identification contained that protein sequence, are retrieved by the software and shown next to their respective checkboxes; only proteins annotated with the checked marker names are kept in the output.

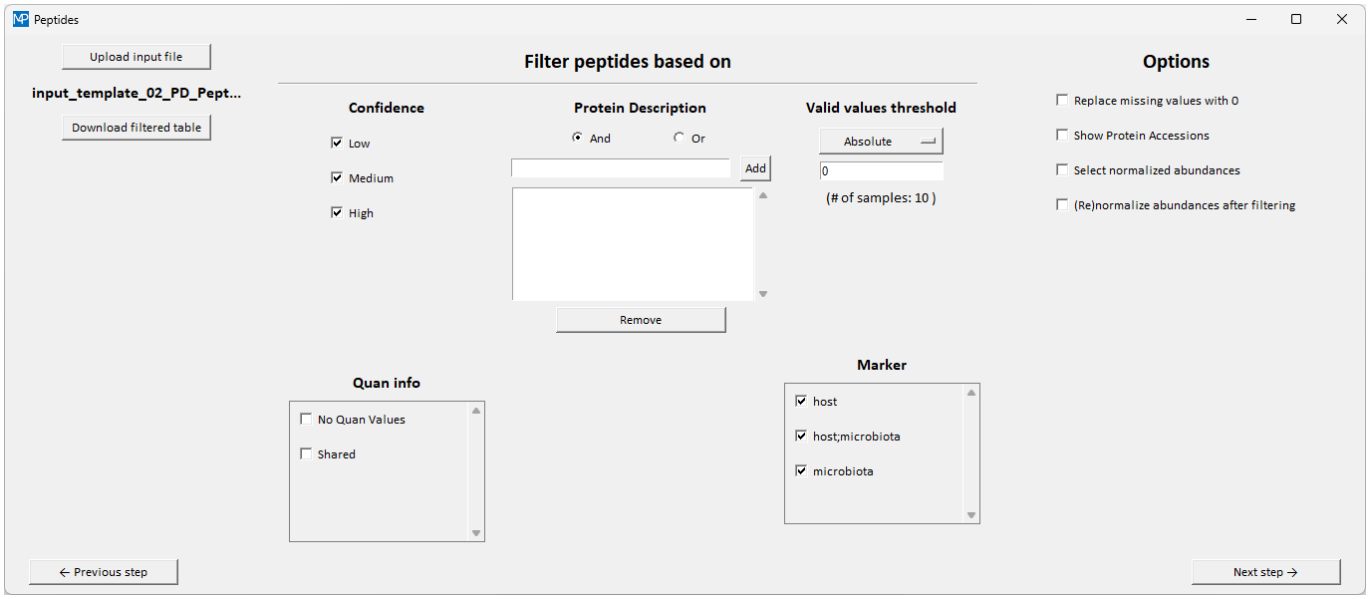
Furthermore, you can choose between the following visualization and calculation options:

- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.
- **Master proteins only**: if selected, only proteins designated as "Master Protein" – i.e., a protein identified by a set of peptides that are not included (all together) in any other protein group – are kept in the output (only available when the corresponding column is present in the input file).
- **Select normalized abundances**: if selected, normalized abundance values will be reported in the table (only available when also normalized abundance values are included in the input file).
- **(Re)normalize abundances after filtering**: if selected, once the chosen filters are applied and the filtered protein list is obtained, the abundance value measured for a protein in a given sample is divided by the total protein abundance measured in that sample and multiplied by 10^{10}.

At the end, the (filtered) identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.4. Proteome Discoverer export - Peptides

When the "Proteome Discoverer export" and "Peptides" options are sequentially selected, peptide identification and quantification data are retrieved from a "Peptide Groups" file exported from Proteome Discoverer, available in one of the following formats: xlsx (Microsoft Excel) or txt (tab-separated values). The input file must contain one "Sequence" column, one "Master Protein Accessions" column and at least one "Abundance" column (whose header must start with "Abundance: F").



Note that when the same peptide sequence is listed more than once (e.g., when presenting different modifications or charge states), Meta4P will report it once and sum its related abundance values. If you try to upload an input file with the wrong structure or format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_02_PD_PeptideGroups.xlsx* is available for download). Once the file is uploaded, the file name is shown under the upload button and the window is populated with several filtering options based on the file content (see the image above).

Peptides can be filtered based on the number/percentage of valid values ("**Valid values threshold**"), so that only peptides with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all peptides pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.

Other optional filters might also be available (if the corresponding columns are present in the input file):

- **Confidence**: only peptides with the selected level(s) of statistical confidence (low, medium and/or high) are kept in the output, based on the checkbox(es) checked.
- **Master Protein Description**: to select only those peptides which belong to a Master Protein containing a specific text in its description (e.g., a protein name or an organism name), type the text of interest in the textbox (be aware that the filter is case sensitive) and click on "**Add**". Multiple texts can be typed and added sequentially; in this case, you can choose between two boolean operators, "**And**" and "**Or**", to determine whether all the texts added or only one of them must be present in the string, respectively, so that the peptide passes the filter.
- **Quan info**: only peptides belonging to the selected quantification categories are kept in the output.
- **Marker**: marker names included in the "Marked as" column of the input file, usually indicating which of the protein database(s) used for identification contained that peptide sequence, are retrieved by the software and shown next to their respective checkboxes; only peptides annotated with the checked marker names are kept in the output.

Furthermore, you can choose between the following visualization and calculation options:

- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.
- **Show Protein Accessions**: if selected, the "Protein Accessions" column (i.e., the column indicating the accession number of all the protein entries matching with a peptide, including non-master proteins) is included in the filtered table (only available when this column is present in the input file).
- **Select normalized abundances**: if selected, normalized abundance values will be reported in the table (only available when also normalized abundance values are included in the input file).
- **(Re)normalize abundances after filtering**: if selected, once the chosen filters are applied and the filtered peptide list is obtained, the abundance value measured for a peptide in a given sample is divided by the total peptide abundance measured in that sample and multiplied by 10^{10}.

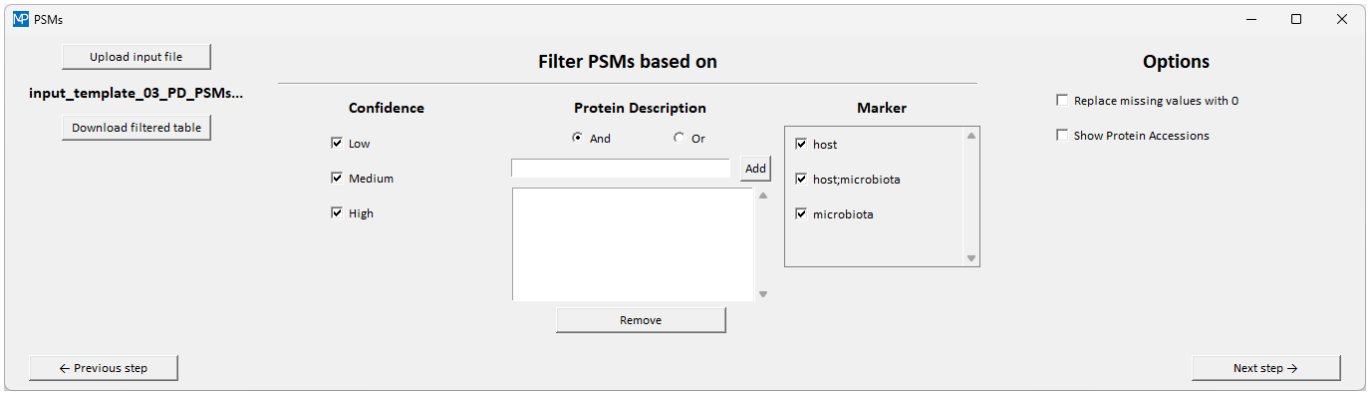
At the end, the (filtered) identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.5. Proteome Discoverer export - PSMs

When the "Proteome Discoverer export" and "PSMs" options are sequentially selected, PSM data are retrieved from a "PSMs" file exported from Proteome Discoverer, available in one of the following formats: xlsx

(Microsoft Excel) or txt (tab-separated values). The input file must contain one "Sequence" column, one "Master Protein Accessions" column and one "File ID" column. If you try to upload an input file with the wrong structure or format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing PSM data (a template file named *input_template_03_PD_PSMs.xlsx* is available for download). Once the file is uploaded, the file name is shown under the upload button and the window is populated with several filtering options based on the file content (see the image below).



Optional filters might be available (if the corresponding columns are present in the input file):

- **Confidence**: only PSMs with the selected level(s) of statistical confidence (low, medium and/or high) are kept in the output, based on the checkbox(es) checked.
- **Master Protein Description**: to select only those PSMs which belong to a Master Protein containing a specific text in its description (e.g., a protein name or an organism name), type the text of interest in the textbox (be aware that the filter is case sensitive) and click on "**Add**". Multiple texts can be typed and added sequentially; in this case, you can choose between two boolean operators, "**And**" and "**Or**", to determine whether all the texts added or only one of them must be present in the string, respectively, so that the PSM passes the filter.
- **Marker**: marker names included in the "Marked as" column of the input file, usually indicating which of the protein database(s) used for identification contained that PSM sequence, are retrieved by the software and shown next to their respective checkboxes; only PSMs annotated with the checked marker names are kept in the output.

Furthermore, you can choose between the following visualization and calculation options:

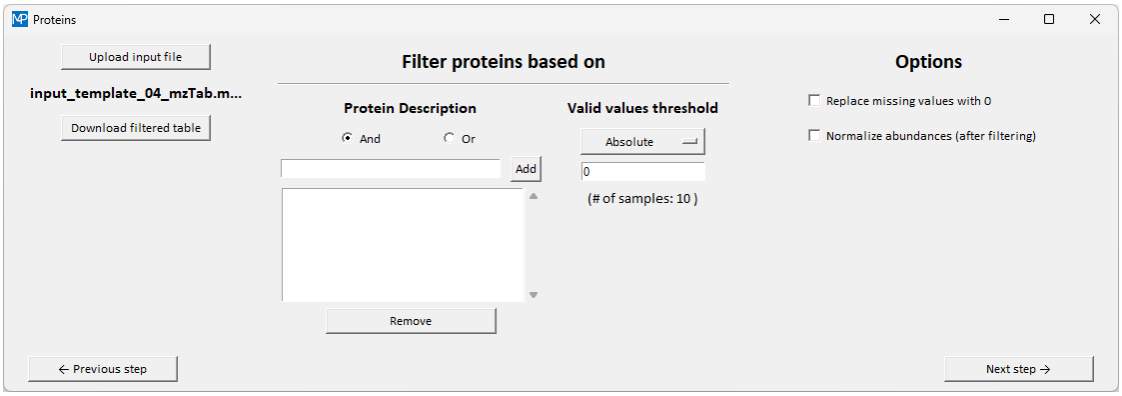
- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.
- **Show Protein Accessions**: if selected, the "Protein Accessions" column (i.e., the column indicating the accession number of all the protein entries matching with a PSM, including non-master proteins) is included in the filtered table (only available when this column is present in the input file).

At the end, Meta4P calculates the number of PSMs detected per sample and reports the corresponding values (next to the corresponding peptide sequence) in a tabular format. The (filtered) PSM table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.6. mzTab - Proteins

When the "mzTab" and "Proteins" options are sequentially selected, protein identification and quantification data are retrieved from a standard mzTab input file. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_04_mzTab.mzTab* is available for download). Once the file is uploaded, the file name is shown under the upload button and the window is populated with several filtering options based on the file content (see the image below).



Proteins can be filtered based on:

- the number/percentage of valid values ("**Valid values threshold**"), so that only proteins with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all proteins pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.
- the presence of a specific text within the protein name/description ("**Protein Description**"), so that only proteins containing that specific text are kept. To do so, type the text of interest in the textbox (be aware that the filter is case sensitive) and click on "**Add**". Multiple texts can be typed and added sequentially; in this case, you can choose between two boolean operators, "**And**" and "**Or**", to determine whether all the texts added or only one of them must be present in the string, respectively, so that a protein passes the filter.

Furthermore, you can select the following options:

- **Normalize abundances (after filtering)**: if selected, once the chosen filters are applied and the filtered protein list is obtained, the abundance value measured for a protein in a given sample is divided by the total protein abundance measured in that sample and multiplied by 10^{10}.
- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.

At the end, the identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

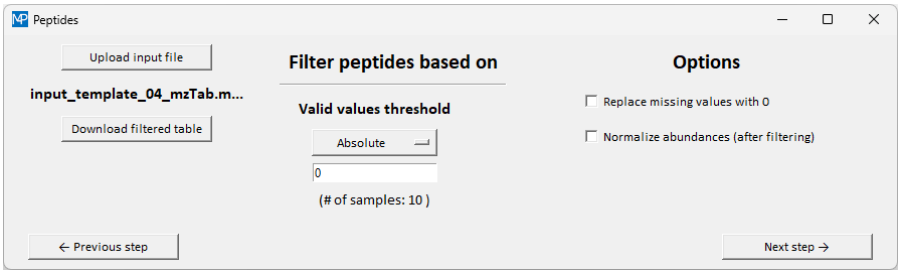
1.7. mzTab - Peptides

When the "mzTab" and "Peptides" options are sequentially selected, peptide identification and quantification data are retrieved from a standard mzTab input file.

Note that when the same peptide sequence is listed more than once (e.g., when presenting different modifications or charge states), Meta4P will report it once and sum its related abundance values. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_04_mzTab.mzTab* is available for download). Once the file is

uploaded, the file name is shown under the upload button and the window is populated with several filtering options based on the file content (see the image below).



Peptides can be filtered based on the number/percentage of valid values ("**Valid values threshold**"), so that only peptides with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all peptides pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.

Furthermore, the following options can be selected:

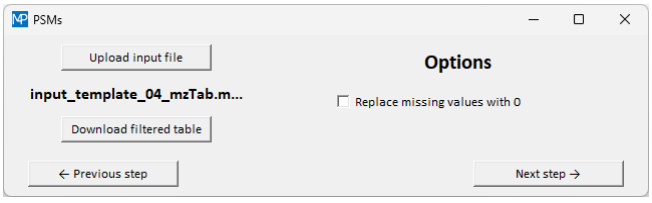
- **Normalize abundances (after filtering)**: if selected, once the chosen filters are applied and the filtered peptide list is obtained, the abundance value measured for a peptide in a given sample is divided by the total peptide abundance measured in that sample and multiplied by 10^{10}.
- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.

At the end, the identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.8. mzTab - PSMs

When the "mzTab" and "PSMs" options are sequentially selected, PSM data are retrieved from a standard mzTab input file. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing PSM data (a template file named *input_template_04_mzTab.mzTab* is available for download). Once the file is uploaded, the following window is shown:



If the "**Replace missing values with 0**" option is selected, missing values (empty cells) are replaced by 0. This selection will be applied to all the following output tables.

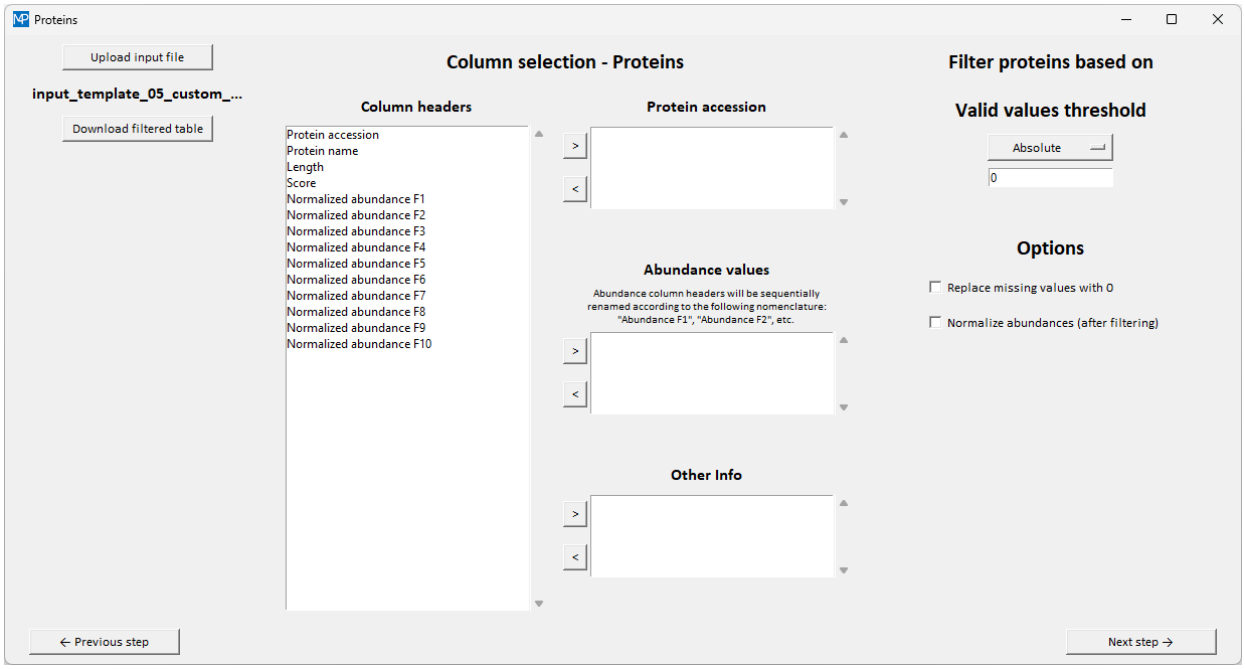
At the end, Meta4P calculates the number of PSMs detected per sample and reports the corresponding values (next to the corresponding peptide sequence) in a tabular format. The (filtered) PSM table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.9. Other/custom - Proteins

When the "Other/custom" and "Proteins" options are sequentially selected, protein identification and quantification data can be retrieved from any tabular input file, in one of the following formats: xlsx, txt or

generic tab-separated format. Column headers must be contained in the first row. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_05_custom_proteins.txt* is available for download). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "**Column headers**" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing protein accession numbers and one or more columns with abundance values must be selected and moved to the "**Protein accession**" and "**Abundance values**" boxes, respectively; other possible columns might remain unselected or be moved to the "**Other info**" box (in case you want them to be kept in the output table). Column reporting information useful for grouping (see "Marked as" column of Proteome Discoverer output) must have "Marked as" as header and semicolon as separator between multiple annotations.

Proteins can be filtered based on the number/percentage of valid values ("**Valid values threshold**"), so that only proteins with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all proteins pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.

Furthermore, the following options can be selected:

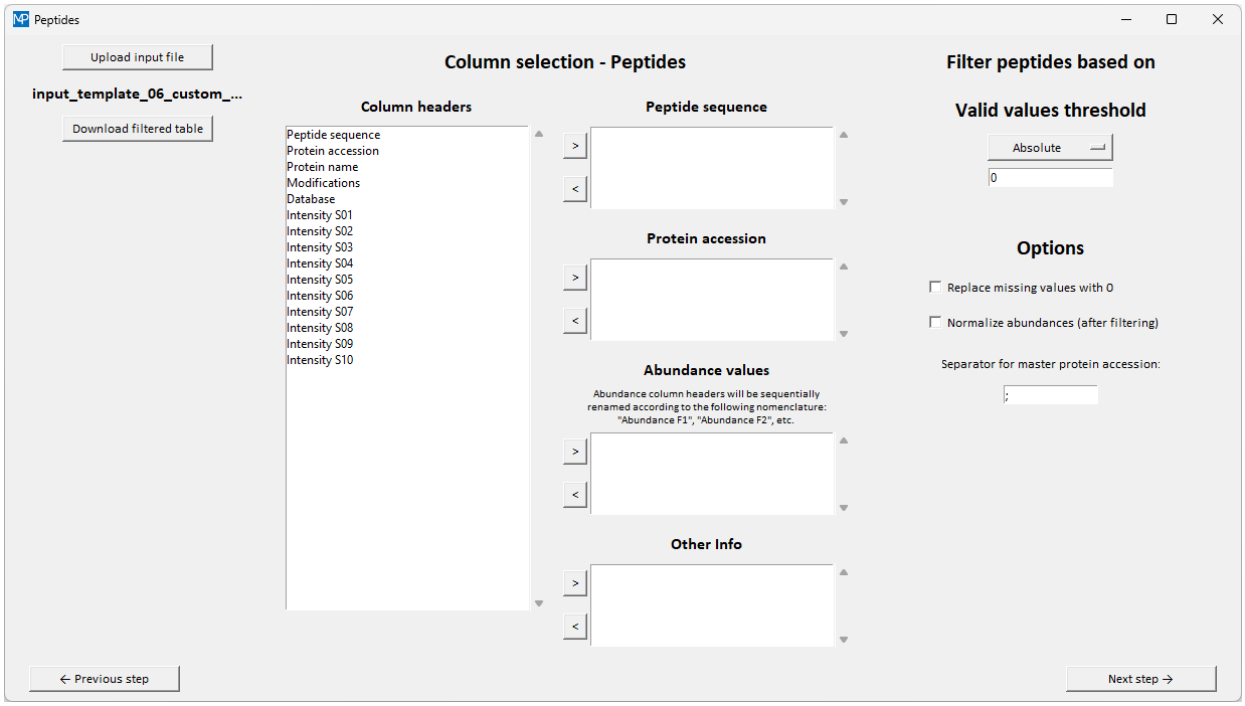
- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.
- **Normalize abundances (after filtering)**: if selected, once the chosen filters are applied and the filtered protein list is obtained, the abundance value measured for a protein in a given sample is divided by the total protein abundance measured in that sample and multiplied by 10^{10}.

At the end, the identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.10. Other/custom - Peptides

When the "Other/custom" and "Peptides" options are sequentially selected, peptide identification and quantification data can be retrieved from any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format. Column headers must be contained in the first row. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing protein identification and quantification data (a template file named *input_template_06_custom_peptides.txt* is available for download). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "**Column headers**" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing peptide sequences, one column listing protein accessionsS, and one or more columns with abundance values must be selected and moved to the "**Peptide sequence**", "**Protein accession**" and "**Abundance values**" boxes, respectively; other possible columns might remain unselected or be moved to the "**Other info**" box (in case you want them to be kept in the output table). Column reporting information useful for grouping (similarly to the "Marked as" column of Proteome Discoverer output) must have "Marked as" as header and semicolon as separator between multiple annotations.

Peptides can be filtered based on the number/percentage of valid values ("**Valid values threshold**"), so that only peptides with a number of valid values (i.e., non-missing values) greater than or equal to the selected threshold are kept. You can choose between indicating absolute or percentage values by selecting the corresponding option in the drop-down menu. Leaving the default value (0) means that all peptides pass the filter. The total number of samples in the dataset is shown in brackets under the textbox.

Furthermore, the following options can be selected:

- **Replace missing values with 0**: if selected, missing values (empty cells) are replaced by 0; this selection will be applied to all the following output tables.

- **Normalize abundances (after filtering)**: if selected, once the chosen filters are applied and the filtered peptide list is obtained, the abundance value measured for a peptide in a given sample is divided by the total peptide abundance measured in that sample and multiplied by 10^{10}.

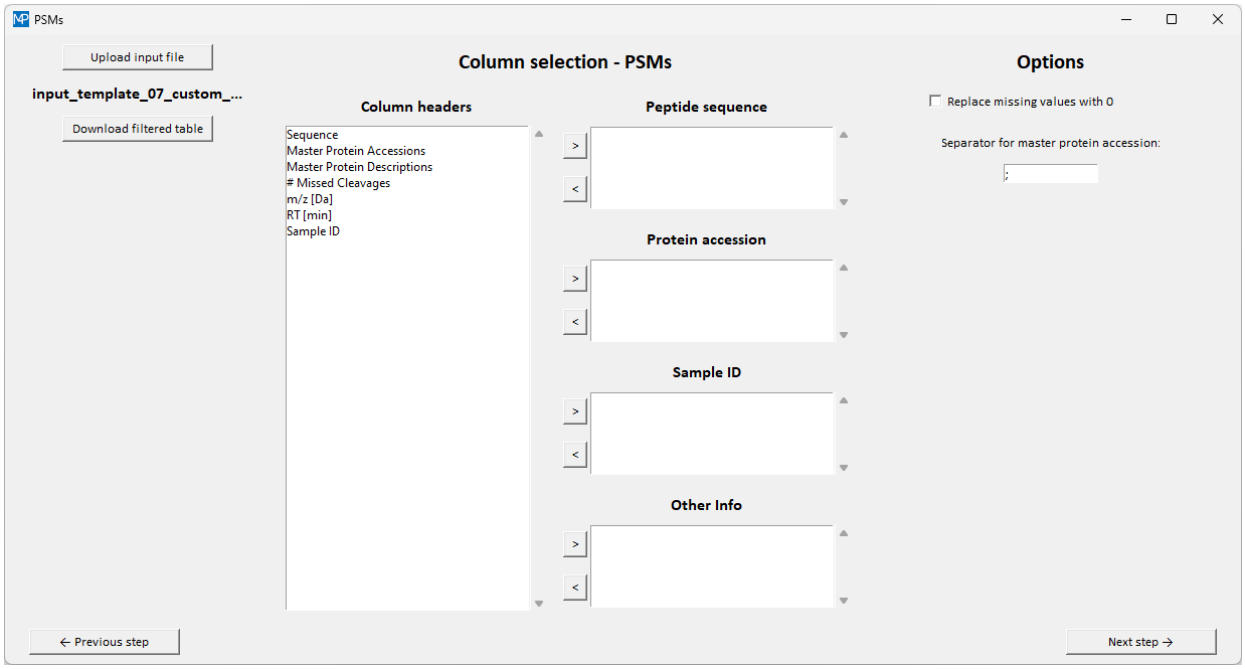
Moreover, it is possible to specify which separator (including comma, semicolon and space) has been used in the "Protein accession" column of the input file in case of multiple protein accessions.

At the end, the identification and quantification table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

1.11. Other/custom - PSMs

When the "Other/custom" and "PSMs" options are sequentially selected, PSM data can be retrieved from any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format. Column headers must be contained in the first row. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload input file**" to select and upload the file containing PSM data (a template file named *input_template_07_custom_PSMs.txt* is available for download). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "**Column headers**" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing peptide sequences, one column listing protein accessions ad one column with sample IDs must be selected and moved to the "**Peptide sequence**", "**Protein accession**" and "**Sample ID**" boxes, respectively; other possible columns might remain unselected or be moved to the "**Other info**" box (in case you want them to be kept in the output table). Column reporting information useful for grouping (similarly to the "Marked as" column of Proteome Discoverer output) must have "Marked as" as header and semicolon as separator between multiple annotations.

If the "**Replace missing values with 0**" option is selected, missing values (empty cells) are replaced by 0. This selection will be applied to all the following output tables.

Moreover, it is possible to specify which separator has been used in the "Protein accession" column of the input file in case of multiple protein accessions.

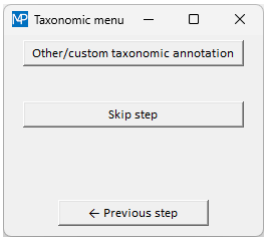
At the end, Meta4P calculates the number of PSMs detected per sample and reports the corresponding values (next to the corresponding peptide sequence) in a tabular format. The (filtered) PSM table can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download filtered table**".

2. Taxonomic annotation

In this (optional) step, a second input file can be uploaded to retrieve taxonomic annotation data and include them in the table containing identification and quantification data.

2.1. Proteins

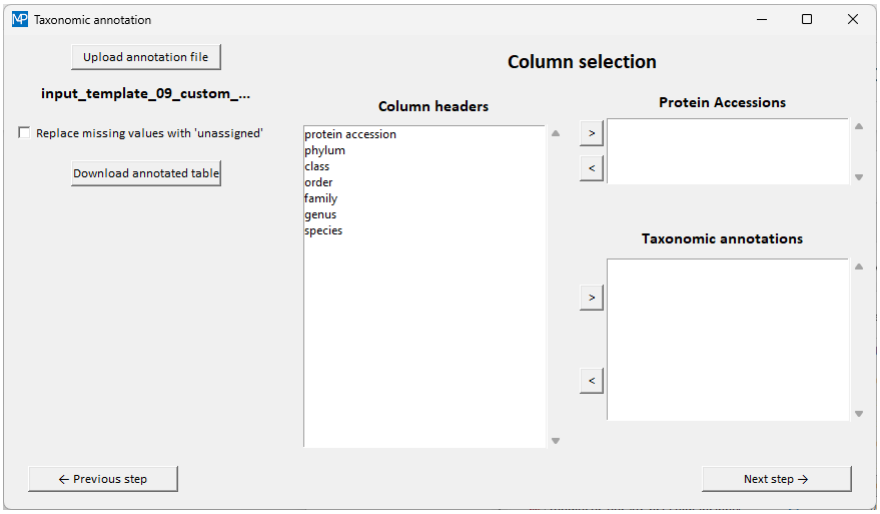
After loading protein identification and quantification data as first input file and clicking on "**Next step**", a "Taxonomic menu" window is shown (see the image below).



Go ahead with the upload of a taxonomic annotation file clicking on "**Other/custom taxonomic annotation**". Alternatively, click on "**Skip step**" if no taxonomic annotation is available.

2.1.1. Other/custom taxonomic annotation

This window allows you to retrieve taxonomic annotation data by uploading any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format. Column headers must be contained in the first row. If you try to upload an input file with the wrong format, an error message will be shown.



Click on "**Upload annotation file**" to select and upload the file containing protein taxonomic annotation data (a template file named *input_template_09_custom_protein_taxonomic_annotation.xlsx* is available for download). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "Column headers" box (see the image above).

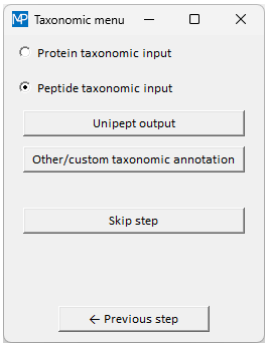
Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing protein accessions and one column listing taxonomic annotations must be selected and moved to the "**Accession**" and "**Taxonomic annotations**" boxes, respectively. Unnecessary columns, if any, should remain unselected.

Additionally, selecting the "**Replace missing values with 'unassigned'**" option allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

An annotated output table, combining taxonomic annotations with identification and quantification data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

2.2. Peptides/PSMs

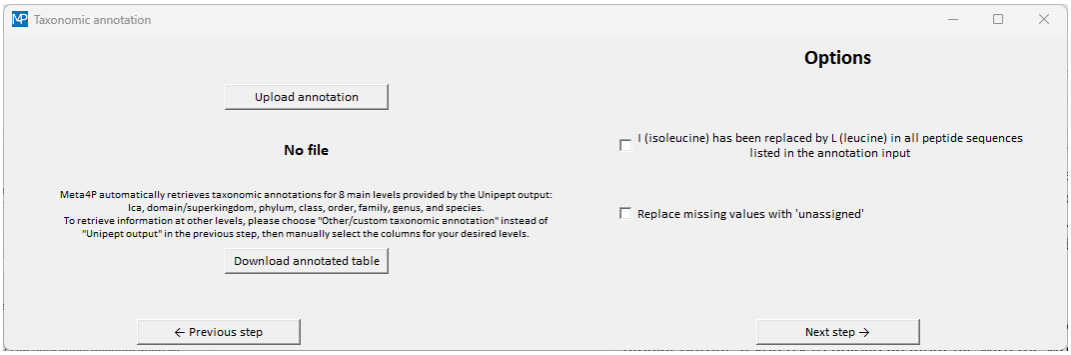
After loading peptide/PSM identification and quantification data as first input file and clicking on "**Next step**", a "Taxonomic menu" window is shown (see the image below).



First, select if the taxonomic input is at protein or peptide level. Then, choose the file type: standard Unipept output table ("**Unipept output**" button, which is available only when the peptide level has been selected) or generic functional annotation tabular file ("**Other/custom functional annotation**" button). Alternatively, if no taxonomic annotation is available, this step can be skipped by clicking on "**Skip step**".

2.2.1. Unipept output

This window (see the image below) allows you to retrieve taxonomic annotation data from a standard Unipept tabular output. If you try to upload an input file with the wrong format, an error message will be shown.



Click on "**Upload annotation**" to select and upload the file containing peptide taxonomic annotation data (a template file named *input_template_08_Unipept_peptide_annotation.csv* is available for download). As specified, Meta4P will automatically retrieve taxonomic annotations for 8 main levels provided by the Unipept output (namely, LCA, domain/superkingdom, phylum, class, order, family, genus, and species). To retrieve

information at other levels, please choose "**Other/custom taxonomic annotation**" in the previous step (see section 2.2), then manually select the columns for your desired levels (see section 2.2.2).

When isoleucine (I) and leucine (L) have been treated as identical amino acids for taxonomic annotation and I has been replaced by L in all peptide sequences listed in the annotation input, please select the corresponding option.

Another option named "**Replace missing values with 'unassigned'**" allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

An annotated output table, combining taxonomic annotations with identification and quantification data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

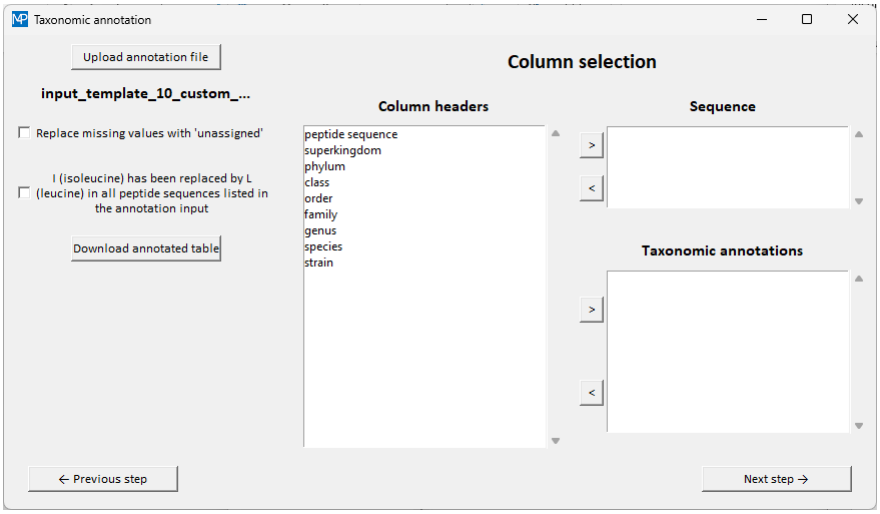
2.2.2. Other/custom taxonomic annotation

This window allows you to retrieve protein taxonomic annotation data by uploading any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format.

If you selected "Protein taxonomic input", windows and options are those described in sections 2.1.1.

If you selected "Peptide taxonomic input", a window is shown that allows you to retrieve peptide taxonomic annotation data by uploading any tabular input file. Column headers must be contained in the first row. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload annotation file**" to select and upload the file containing peptide taxonomic annotation data (a template file named *input_template_10_custom_peptide_taxonomic_annotation.tab* is available for download). Once the file is uploaded, the file name is shown under the upload button and file column headers are listed in the "Column headers" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing peptide sequences and one column listing taxonomic annotations must be selected and moved to the "**Sequence**" and "**Taxonomic annotations**" boxes, respectively. Unnecessary columns, if any, should remain unselected.

Another option named "**Replace missing values with 'unassigned'**" allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

Moreover, when isoleucine (I) and leucine (L) have been treated as identical amino acids for taxonomic annotation and I has been replaced by L in all peptide sequences listed in the annotation input, please select the corresponding option.

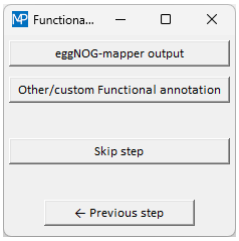
An annotated output table, combining taxonomic annotations with identification and quantification data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

3. Functional annotation

In this (optional) step, another input file can be uploaded to retrieve functional annotation data and include them in the table containing identification and quantification (and optionally taxonomic annotation) data.

3.1. Proteins

After loading taxonomic annotation and clicking on "**Next step**" (or after skipping taxonomic annotation), a "Functional menu" window is shown (see the image below). Here, you can choose between two file types: standard eggNOG-mapper output ("**eggNOG-mapper output**" button) or generic functional annotation tabular file ("**Other/custom functional annotation**" button). Alternatively, if no functional annotation is available, this step can be skipped by clicking on "**Skip step**".



3.1.1 eggNOG-mapper output

This window (see the image below) allows you to retrieve taxonomic annotation data from a standard eggNOG-mapper output. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload annotation**" to select and upload the file containing protein functional annotation data (a template file named *input_template_11_eggNOG-mapper_protein_functional_annotation* is available for download). As specified, Meta4P will automatically retrieve functional annotations for 8 main levels provided by the eggNOG-mapper output (namely, COG category, GO category, EC number, KEGG KO, KEGG Pathway, KEGG Module, KEGG Reaction, and CAZy). To retrieve information at other levels, please choose "**Other/custom functional annotation**" in the previous step (see section 3.1), then manually select the columns for your desired levels (see section 3.1.2).

As an option, to retrieve and include in the table (as supplementary columns) the annotation names provided by the KEGG database for all KEGG categories, click on "**Retrieve KEGG name**". As this information is retrieved from the KEGG website, a working internet connection is needed in order that this operation is performed; this operation may take up to a few minutes. In case a protein has multiple functional annotations, their names will be separated by a comma as well as their codes (except for the "KEGG Module" annotation names, for which a vertical bar is used to avoid mistakes). In case a peptide is associated with multiple Master Proteins (usually separated by a semicolon), codes and names of functional annotations assigned to different Master Proteins are consistently separated by a semicolon (except for the "KEGG Reaction" annotation names, for which a vertical bar is used to avoid mistakes). When two analogous "KEGG Pathway" annotations are reported (i.e., having the same numeric code but two different prefixes, namely 'map' and 'ko'), the code with the 'ko' prefix is removed.

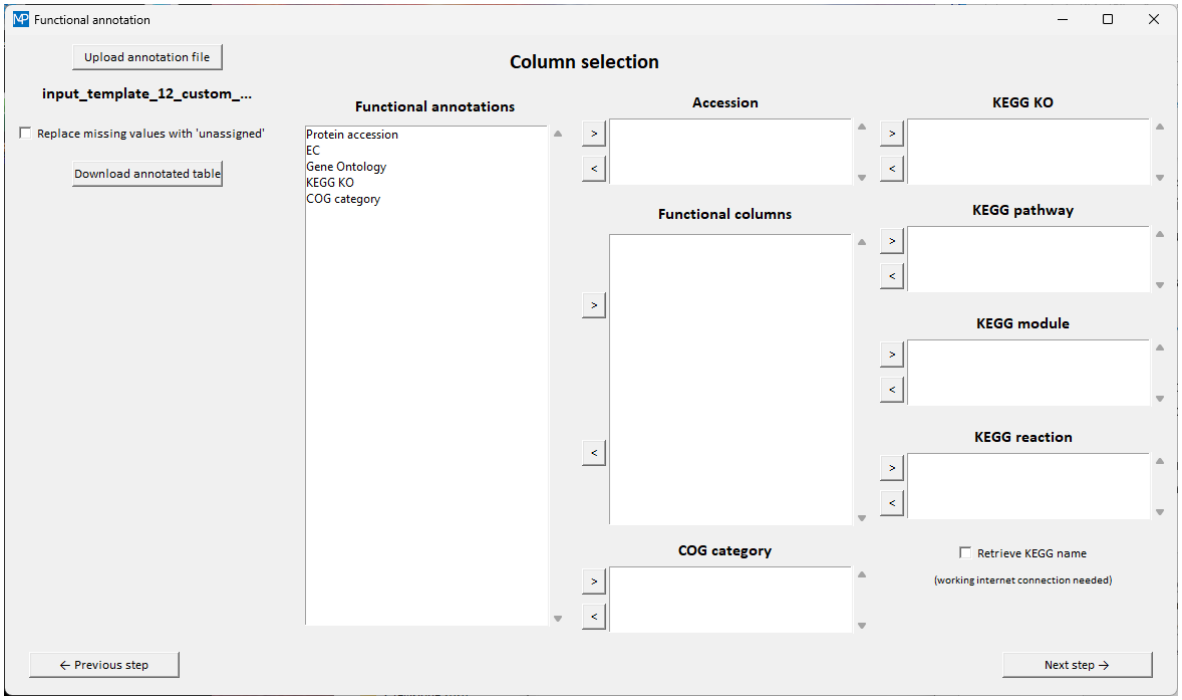
A further option, named "**Replace missing values with 'unassigned'**", allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

An annotated output table, combining functional annotations with identification and quantification (and possibly taxonomic annotation) data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

3.1.2. Other/custom functional annotation

This window allows you to retrieve functional annotation data by uploading any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format. Column headers must be contained in the first row. Multiple functional annotations must be separated by a comma. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload annotation file**" to select and upload the file containing protein functional annotation data (a template file named *input_template_12_custom_protein_functional_annotation.tsv* is available for download). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "Column headers" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing protein accessions and one column listing functional annotations must be selected and moved to the "**Accession**" box and to one of the functional annotation boxes, respectively; when columns listing KEGG or COG annotations are present in the input file, these have to be moved to the specific box corresponding to their category. Unnecessary columns, if any, should remain unselected.

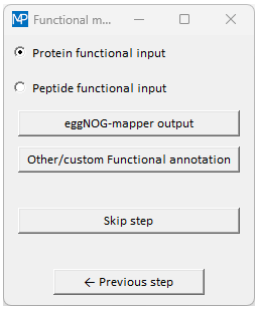
As an option, to retrieve and include in the table (as supplementary columns) the annotation names provided by the KEGG database for all KEGG categories, click on "**Retrieve KEGG name**". As this information is retrieved from the KEGG website, a working internet connection is needed in order that this operation is performed; this operation may take up to a few minutes. In case a protein has multiple functional annotations, their names will be separated by a comma as well as their codes (except for the "KEGG Module" annotation names, for which a vertical bar is used to avoid mistakes). In case a peptide is associated to multiple Master Proteins (usually separated by a semicolon), codes and names of functional annotations assigned to different Master Proteins are consistently separated by a semicolon (except for the "KEGG Reaction" annotation names, for which a vertical bar is used to avoid mistakes). When two analogous "KEGG Pathway" annotations are reported (i.e., having the same numeric code but two different prefixes, namely 'map' and 'ko'), the code with the 'ko' prefix is removed.

A further option, named "**Replace missing values with 'unassigned'**", allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

An annotated output table, combining functional annotations with identification and quantification (and possibly taxonomic annotation) data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

3.2. Peptides/PSMs

After loading taxonomic annotation and clicking on "**Next step**" (or after skipping taxonomic annotation), a "Functional menu" window is shown (see the image below). First, select if the functional input is at protein or peptide level (in the former case, windows and options are those described in sections 3.1.1 and 3.1.2; for the latter case, see sections 3.2.1 and 3.2.2). Then, choose the file type: standard eggNOG-mapper output ("**eggNOG-mapper output**" button) or generic functional annotation tabular file ("**Other/custom functional annotation**" button).



Alternatively, if no functional annotation is available, this entire step can be skipped by clicking on "**Skip step**".

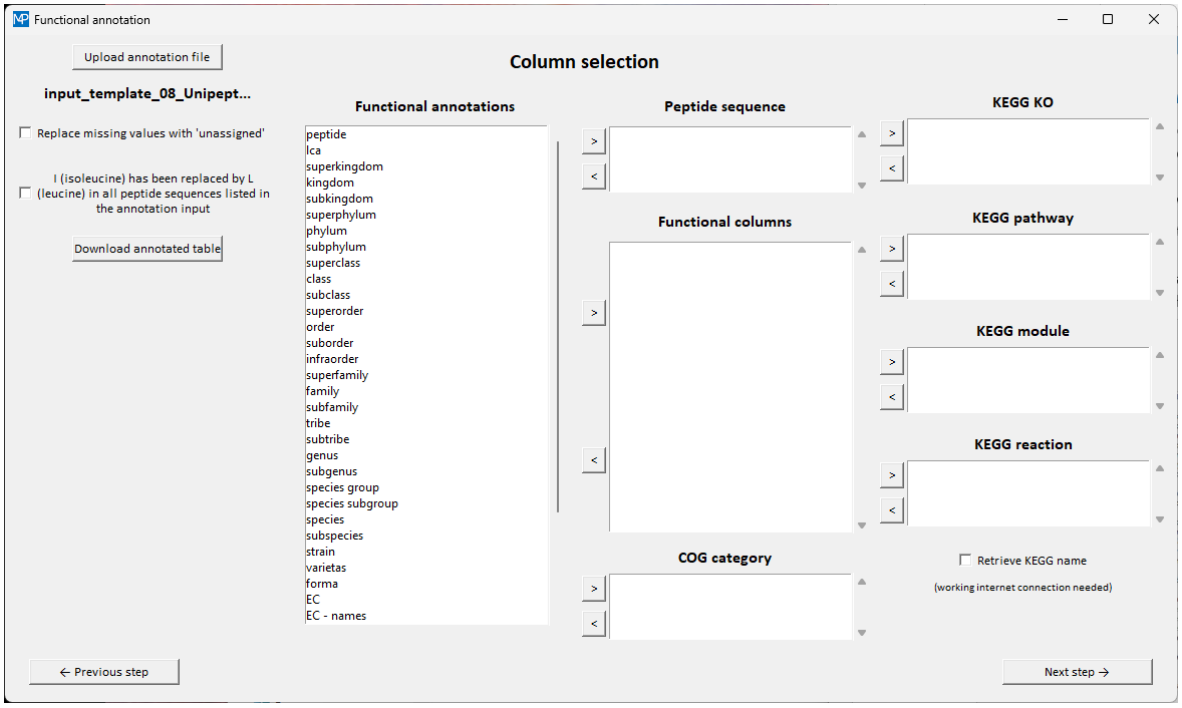
3.2.1 eggNOG-mapper output

Window and options are identical to those described in section 3.1.1. The only difference lies in the presence of peptide instead of protein sequences in the "query" column of the input.

3.2.2 Other/custom functional annotation

This window allows you to retrieve functional annotation data by uploading any tabular input file, in one of the following formats: xlsx, txt or generic tab-separated format. Column headers must be contained in the first row. Multiple functional annotations must be separated by a comma. If you try to upload an input file with the wrong format, an error message will be shown.

Click on "**Upload annotation file**" to select and upload the file containing peptide functional annotation data (the template file named *input_template_08_Unipept_peptide_annotation.csv* also contains this kind of information). Once the file is uploaded, the file name is shown under the upload button and the file column headers are listed in the "Column headers" box (see the image below).



Select the columns of interests and move them to the corresponding box by using the arrows next to the boxes. At least one column listing peptide sequences and one column listing functional annotations must be selected and moved to the "**Peptide sequence**" box and to one of the functional boxes, respectively; when columns listing KEGG or COG annotations are present in the input file, these have to be moved to the specific box corresponding to their category. Unnecessary columns, if any, should remain unselected.

As an option, to retrieve and include in the table (as supplementary columns) the annotation names provided by the KEGG database for all KEGG categories, click on "**Retrieve KEGG name**". As this information is retrieved from the KEGG website, a working internet connection is needed in order that this operation is performed; this operation may take up to a few minutes. In case a protein has multiple functional annotations, their names will be separated by a comma as well as their codes (except for the "KEGG Module" annotation names, for which a vertical bar is used to avoid mistakes). In case a peptide is associated to multiple Master Proteins (usually separated by a semicolon), codes and names of functional annotations assigned to different Master Proteins are consistently separated by a semicolon (except for the "KEGG Reaction" annotation names, for which a vertical bar is used to avoid mistakes). When two analogous "KEGG Pathway" annotations are reported (i.e., having the same numeric code but two different prefixes, namely 'map' and 'ko'), the code with the 'ko' prefix is removed.

A further option, named "**Replace missing values with 'unassigned'**", allows you to designate cells with missing annotations as "unassigned". When this option is selected, quantitative values related to missing annotations are also considered in the next data aggregation step.

Moreover, when isoleucine (I) and leucine (L) have been treated as identical amino acids for taxonomic annotation and I has been replaced by L in all peptide sequences listed in the annotation input, please select the corresponding option.

An annotated output table, combining functional annotations with identification and quantification (and possibly taxonomic annotation) data, can be downloaded in xlsx, txt or generic tab-separated format by clicking on "**Download annotated table**".

4. Protein/Peptide/PSM metrics

The next "Summary metrics" windows allows you to download a summary table listing the main metrics of the identification, quantification and annotation data. More specifically:

- when "Proteins" data are concerned ("**Download protein metrics**" button), the summary table reports the number of quantified proteins and their total abundance for each annotation category, for each sample and for the whole dataset;
- when "Peptides" data are concerned ("**Download peptide metrics**" button), the summary table reports the number of quantified peptides and their total abundance for each annotation category, for each sample and for the whole dataset;
- when "PSMs" data are concerned ("**Download PSM metrics**" button), the summary table reports the number of identified peptides and PSMs for each annotation category, for each sample and for the whole dataset.

Note that a protein/peptide with a missing annotation in a specific category does not contribute to the summary metrics for that particular annotation category. This also applies when the "**Replace missing values with 'unassigned'**" option has been selected for taxonomic and/or functional annotations.

5. Data aggregation

In this step, abundance data can be aggregated based on taxonomic, functional and/or combined taxonomic-functional annotations; in other words, the abundances of all proteins/peptides/PSMs sharing the same annotation are summed for each sample.

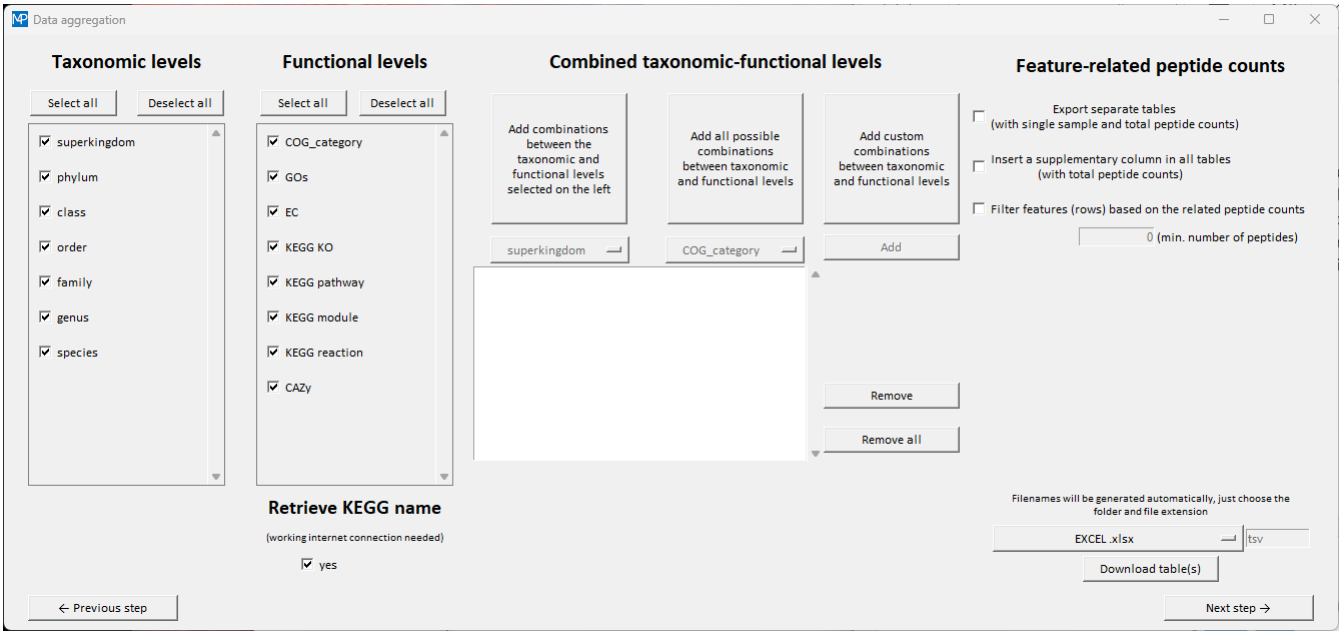
The layout of the "Data aggregation" window is shown in the image below.

If available, taxonomic and functional levels of interest can be selected by checking their respective boxes. The "**Select all**" and "**Deselect all**" buttons are also available at the top.

Additionally, users can combine taxonomic and functional levels in a customized way:

- to add all combinations of the selected taxonomic and functional levels in the "Taxonomic levels" and "Functional levels" menus on the left to the list, click the left button ("**Add combinations between the selected taxonomic and functional levels**").
- to add all combinations between all taxonomic and functional levels (including those not selected) listed in the "Taxonomic levels" and "Functional levels" menus, click the central button ("**Add all possible combinations between taxonomic and functional levels**").

- to add custom combinations of taxonomic and functional levels to the list, click the right button ("**Add custom combinations between taxonomic and functional levels**"), select the desired taxonomic level in the left drop-down menu and the desired functional level in the right drop-down menu, then click "**Add**".

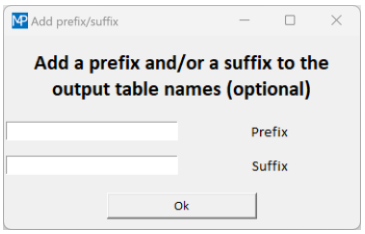


As an option, to retrieve and include in the table(s) (as a supplementary column) the annotation names provided by the KEGG database for all the selected KEGG categories, check the "**Retrieve KEGG name**" checkbox.

For each taxonomic, functional or taxonomic-functional aggregated table selected, **feature-related peptide (or protein) counts** can also be retrieved, i.e., the number of peptides (or proteins) contributing to the summed abundance (showed as aggregated value in the table) for which an abundance value was measured for each measured feature. By checking the corresponding checkbox(es), this type of information can be **exported in separate tables** (having the same name of the corresponding tables reporting the aggregated abundance data, plus the suffix "_proteincounts" for protein level inputs and "_peptidecounts" for peptide/PSM level inputs) and/or **inserted as a supplementary column** (named "Total peptide count") in all aggregated tables. Furthermore, by checking the third checkbox, it is possible to **filter out** all features that do not reach a minimum peptide (or protein) count (the threshold must be typed in the box).

To download the tables for each annotation level selected, select the desired format (xlsx, txt or generic tab-separated) from the drop-down menu, then click on "**Download tables**". As a result, each table will be saved to the selected directory using the name of the corresponding annotation level.

A further window (see the image below) will then appear enabling the addition of a prefix and/or a suffix to the names of all output files. Type the desired prefix and/or suffix in the corresponding box or leave the boxes blank if no prefix or suffix is needed.



Output template files (aggregated tables at various taxonomic, functional and taxon-specific functional levels) are also available for download.

6. Annotation metrics

The next "Summary metrics" windows allows you to download a summary table listing the main metrics of the annotation data, based on the previously downloaded aggregated tables. Note that this option is only available when at least one table has been downloaded in the previous step and that its metrics are specifically referred to the previously downloaded tables. More specifically:

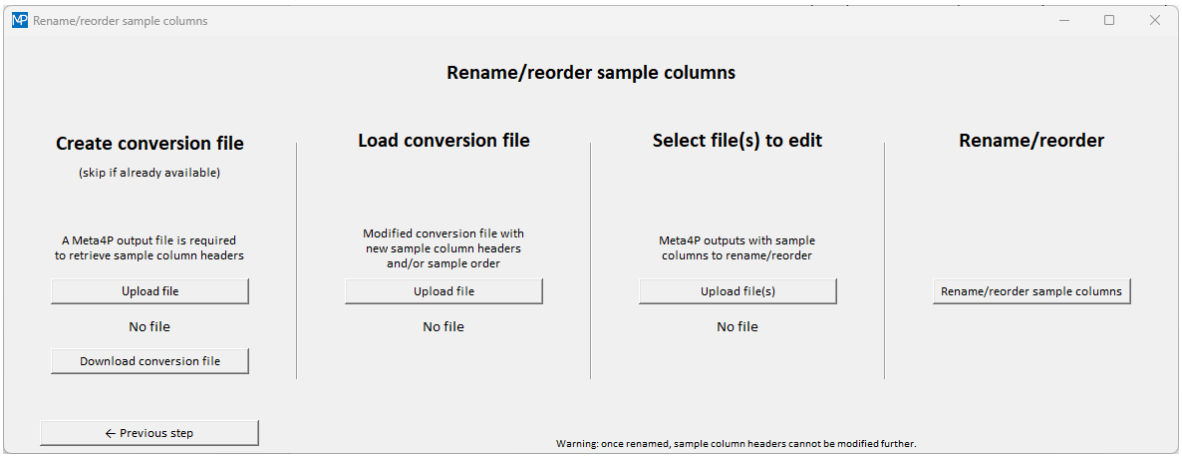
- when "Proteins" data are concerned, the summary table reports the number of quantified features and their total abundance for each annotation level (taxa, functions and/or taxon-specific functions) selected;
- when "Peptides" data are concerned, the summary table reports the number of quantified features and their total abundance for each annotation level (taxa, functions and/or taxon-specific functions) selected;
- when "PSMs" data are concerned, the summary table reports the number of identified features and related PSMs for each annotation level (taxa, functions and/or taxon-specific functions) selected.

Note that unassigned features do not contribute to the annotation metrics, even when the "**Replace missing values with 'unassigned'**" option has been selected for taxonomic and/or functional annotations; in the case of a taxon-specific feature, the presence of a single unassigned term (taxonomic or functional) implies that that given feature will not be taken into account for the calculation of the annotation metrics.

7. Rename/reorder sample columns

In this step, you can customize names Sand order of sample columns, based on a conversion file.

To create a conversion file ("**Create conversion file**" section; see the image below), upload one of the output tables generated by Meta4P in the previous steps (sample column headers must not have been modified in any way) by clicking on "**Upload file**". Based on this input, the software creates a conversion file, i.e., a tabular file containing a first column (header "Old Name") reporting the sample list of the original input file (one sample per row) and a second, empty column (header "New name"). Click on "**Download conversion file**", open the downloaded file and type the new sample names in the "New name" column. If useful, sample order can also be changed and Meta4P will change the column order in the output tables accordingly.



Once the conversion file has been filled in, upload it by clicking on "Upload file" in the "**Load conversion file**" section.

Then, select which of the output table(s) generated and downloaded in the previous steps need(s) to be subjected to renaming/reordering of sample columns ("**Select file(s) to edit**" section, "**Upload file(s)**" button).

Finally, click on "**Rename/reorder sample columns**" ("**Rename/reorder**" section) to complete the process.