|  |
| --- |
| Metabolomics data analysis in the context of metabolic networks |
| logo5etExplore |
| **Ludovic COTTRET**  **Florence VINSON**  **Maxime CHAZALVIEL**  **Clément FRAINAY**  **Etienne CAMENEN**  **Yoann GLOAGUEN**  **Fabien JOURDAN**  2019 |

Metabolomics data analysis in the context of metabolic networks

**2019**

logo5

[Fabien.Jourdan@inra.fr](mailto:Fabien.Jourdan@inra.fr)

<https://sites.google.com/site/fabienjourdan/>

[www.metexplore.fr](http://www.metexplore.fr)

[www.facebook.com/metexplore](http://www.facebook.com/metexplore)

@metexplore

# Introduction

**Metabolomics data mapping and sub-network extraction**

As a benchmark, we will use a metabolomics experiment performed on Yeast exposed to cadmium (toxicological compound) (1). Metabolites were identified and quantified in exposed and unexposed conditions. In the article, a list of relevant metabolites is provided. The challenge is then, based on these markers, to provide clues on the potential metabolic reactions involved in the response to this environmental stress.

Metabolic networks stored in databases (like KEGG, BioCyc) or in SBML formats often contain hundreds of reactions and hundreds of metabolites. The challenge is, once your metabolites of interest are identified in the network, to get the information on the reactions connecting them. To do so, the graph structure of the network can be used.

This training aims at guiding you in mining metabolomics data in the context of metabolic networks. Following points will be treated:

* Find relevant metabolite identifiers in databases
* Map data on metabolic pathways and networks
* Identify sub-networks connecting identified metabolites
* Create publication-ready representations

These steps will be performed using two different platforms:

* KEGG
* MetExplore

Madalinski, G., Godat, E., Alves, S., Lesage, D., Genin, E., Levi, P., Labarre, J., et al. (2008). Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Analytical chemistry*, *80*(9), 3291-303.

In the article, we will use **identified metabolites** which intensity ratio **greater or equal to 1.9** and **lower or equal to 0.6**.

This document is constantly evolving, so if you are using it, please contact us to get an updated version:

contact-metexplore@inra.fr

Practice 1

# Finding identifiers - Chemical Translation Service

First challenge when dealing with metabolomics data consists in finding relevant metabolite identifiers based on names provided by experimentalists and the ones contained in the model. In fact, these names come with different spellings, capital letters or not...This makes automatic matching not easy for computers programs.

Fortunately, some online services are able, based on synonym tables and algorithms to automatically retrieve metabolite identifiers from various databases.

|  |
| --- |
| Arginine |
| Reduced glutathione |
| Methionine |
| Threonine |
| Homoserine |
| Glutamylcysteine |
| 5-methylthioadenosine |
| Serine |
| Cystathionine |
| Tyrosine |

Table 1: list of metabolites of interest

For instance, "Chemical Translation Service" (2) allows converting metabolite names into various identifiers.

<http://cts.fiehnlab.ucdavis.edu>/

Other services are available like: <http://biocyc.org/metabolite-translation-service.shtml>

► Go to <http://cts.fiehnlab.ucdavis.edu>/

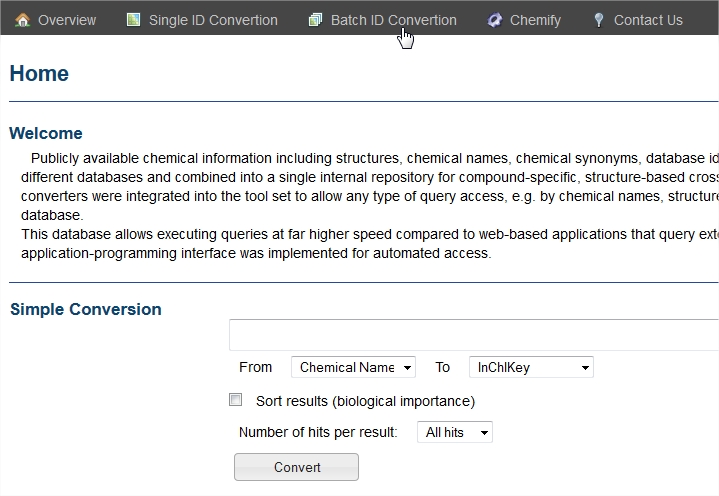


Figure 1: Web interface of CTS

You can do a single conversion or a conversion of several metabolite names.

► Click on the Batch id Conversion menu and then copy paste the list of metabolite names

► Select the database of interest (in our case: KEGG)

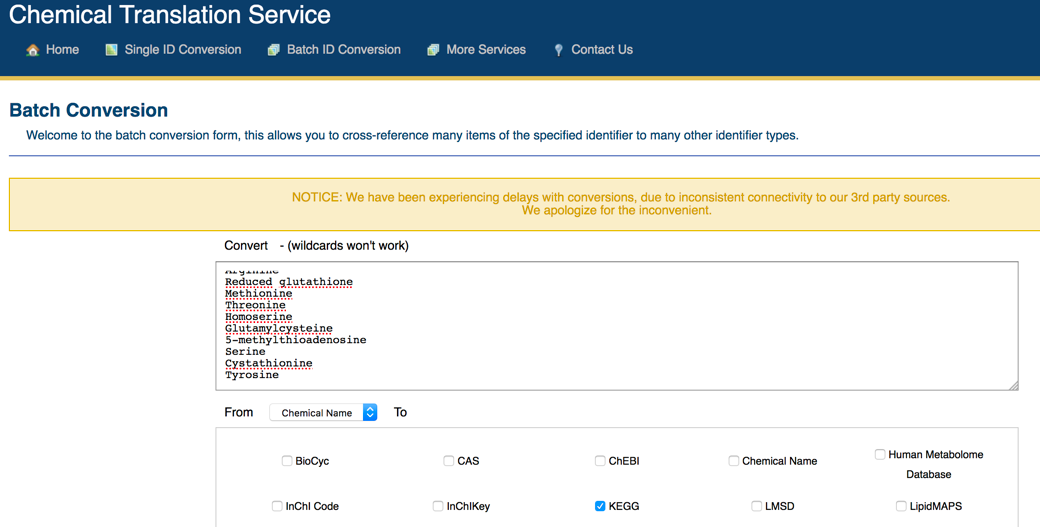


Figure 2: Batch conversion in CTS

►You can then export the resulting mapping in an excel file by clicking on the "excel" button.

Note that some for some compounds there are several identifiers possible like methionine.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  | | --- | --- | | **Search Term** | **KEGG** | | Arginine | C00062 | | Reduced glutathione | C00051 | | Methionine | C01733 | | Threonine | C00188 | | Homoserine | C00263 | | Glutamylcysteine | Not Found | | 5-methylthioadenosine | C00170 | | Serine | C00065 | | Cystathionine | C00542 | | Tyrosine | C00082 | |  |  | |  |  |
|  |  |  |

Table 2: CTS output in Excel. In yellow, metabolites where there are still some identifier issues

Moreover, some metabolites like Glutamylcysteine are not matched to a database entry (“not found”). It thus requires to go back to each database in order to find missing identifiers.

# Finding Metabolite Identifiers in KEGG - http://www.genome.jp/kegg/

KEGG (3) is made of several interconnected databases. The one containing information on metabolites is called: KEGG COMPOUND and is contained in the larger one called LIGAND

►Once you are on KEGG webpage, click on "KEGG COMPOUND"

►Put the name of a compound (e.g. glutamylcysteine) and click "go"

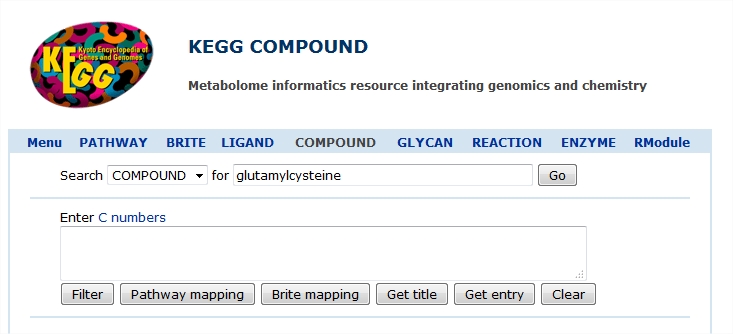


Figure 3: Identifier search in KEGG

A list containing all KEGG's metabolites corresponding to your search is displayed.

Note that several options are possible

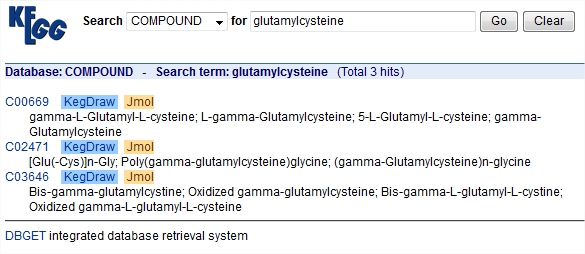


Figure 4: Output of KEGG's metabolite search

Challenge is then to choose the relevant one among all possible identifiers.

It is also good sometimes to cross-check some of the identifiers. In fact, CTS propose C00542 for Cystathionine. But we are going to see it is not the right one!

►Search for cystathionine in KEGG

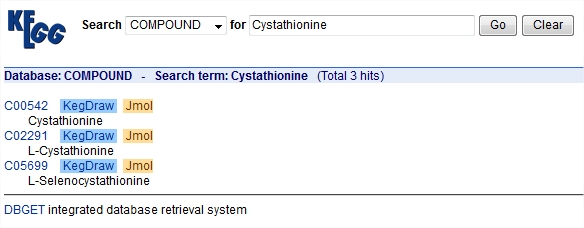


Figure 5: Result for Cystathionine search in KEGG

We get two propositions C00542 and C02291.

►Click on the ID to see the metabolite details

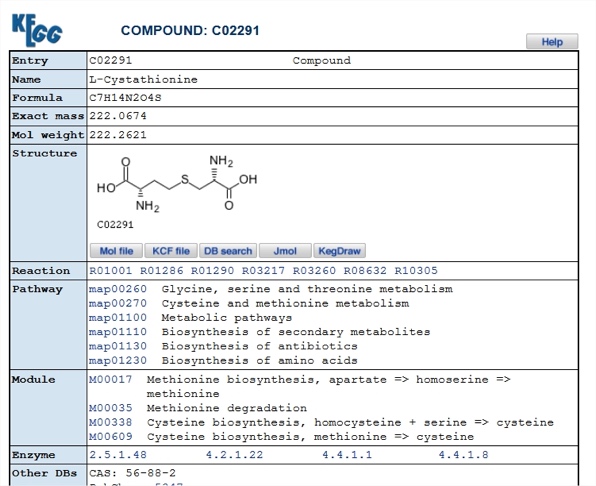
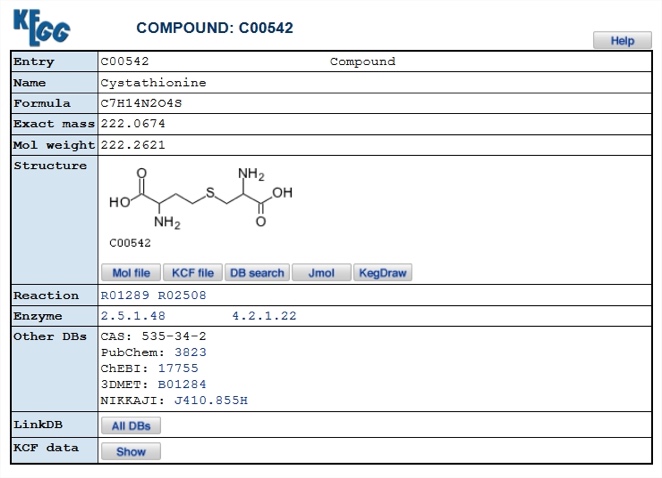


Figure 6: Metabolite information for C00542 and C02291

You will note that one (C00542) is associated to no pathway while the other (C02291) is associated to several pathways. This is a clue that it is the right identifier.

Following table shows the final identifier list.

|  |  |  |
| --- | --- | --- |
| **Name** | **KEGG ID** | **ratio** |
| Arginine | C00062 | 1.9 |
| Reduced glutathione | C00051 | 33.9 |
| Methionine | C00073 | 0.3 |
| Threonine | C00188 | 0.6 |
| Homoserine | C00263 | 0.6 |
| Glutamylcysteine | C00669 | 192.2 |
| 5-methylthioadenosine | C00170 | 11 |
| Serine | C00065 | 0.2 |
| Cystathionine | C02291 | 50.5 |
| Tyrosine | C00082 | 2.9 |

Table 3: Metabolite list and corresponding KEGG identifiers

# Mapping Metabolites on KEGG maps - http://www.genome.jp/kegg/

KEGG allows coloring metabolites (and/or reactions) on global and pathway views of the metabolism.

<http://www.genome.jp/kegg/kegg2.html>

► Click on « KEGG Mapper » and then « Search&Color Pathway».

► Choose the organism, or keep Reference Pathway if you want a non-targeted analysis. (*saccharomyces cerevisiae* -> sce)

**TIP:** In KEGG scrolling menus, organisms are ordered according to their phylogeny which complicates the search. Once the list displayed you can enter the first letter of the organism until you find it.

► Copy/Paste the list of identifiers from the dataset provided in the Metabolite.xls file.

****

Figure 7: KEGG's form for metabolite mapping

►Click the "Exec" button

The list of all the pathways containing at least one metabolite is displayed. Each link is connected to a colored map.

There are several levels of details for pathways in KEGG:

* First line corresponds to the overall representation of the metabolism. Only pathways where at least one metabolite was identified will be shown in color.
* The two next ones correspond to metabolic functions gathering several pathways
* The other ones correspond to individual pathways.



Figure 8: List of all the maps where at least one metabolite of the input list is found.

►Click on map sce00270 "Cysteine and methionine metabolism"



Figure 9: Yeast Cysteine and Methionine metabolism. Green E.C. numbers correspond to reactions with associated genes.

**TIP: The same map can be obtained if you have metabolite identifiers**

Any URL can be built based on the following syntax:

http://www.genome.jp/kegg-bin/show\_pathway?mapid/keggid1/keggid2/.

For instance a map's URL with metabolites C00118 and C00111 colored is:

<http://www.genome.jp/kegg-bin/show_pathway?sce01100/C00118/C00111>

Documentation on this point can be found at: <http://www.kegg.jp/kegg/docs/weblink.html>

Practice 2

# Extracting metabolic paths using KEGG - PathComp

KEGG PathComp tool aims at computing "possible reaction paths between two compounds using the binary relations of substrates and products in known enzymatic reactions. The list of binary relations is generated from the enzyme list that corresponds to the KEGG reference pathways (all known enzymes) or that is found in an annotated genome (organism specific composition of enzymes).

► Go to: http://www.genome.jp/tools/pathcomp/ or in KEGG2 go in the "Search & Compute" and click on the PathComp link.

**Definition**: A path is a set of reactions that can be used to connect two metabolites.

****

Figure 10: Form to launch the computation PathComp to retrieve metabolic path between two compounds.

► Select the organism on which you want to perform the search

► Look for a path between L-Methionine (C00073) and reduced glutathione (C00051)

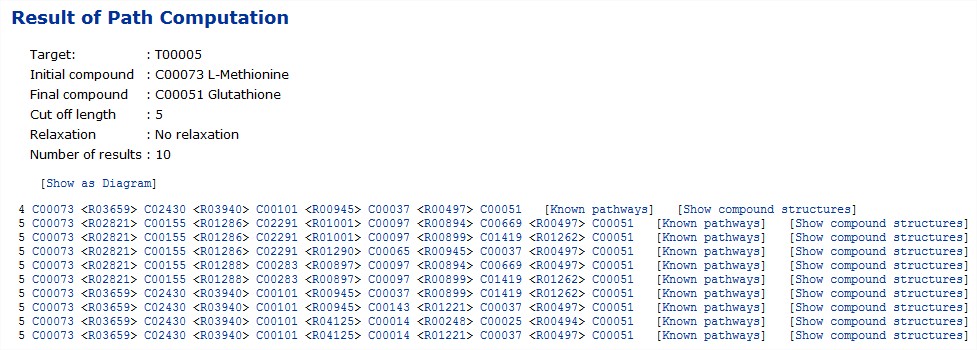


Figure 11: Result of a PathComp search

In the form the "Cut Off Length" value allows to set the size of the maximum length of paths.

► Change the "cut off length" to 10 and compute again

**Definition**: The length of path is the number of reactions belonging to the path between two metabolites.

You will notice that many new paths are found. Moreover, the longer are the paths, more paths are possible. This is due to the structure of the metabolic network that allows many ways to go from a metabolite to another.

Each path is displayed by listing the reactions and metabolites involved. For each path it is possible to display the compounds involved.

► Click on the shortest path line [show compound structure].

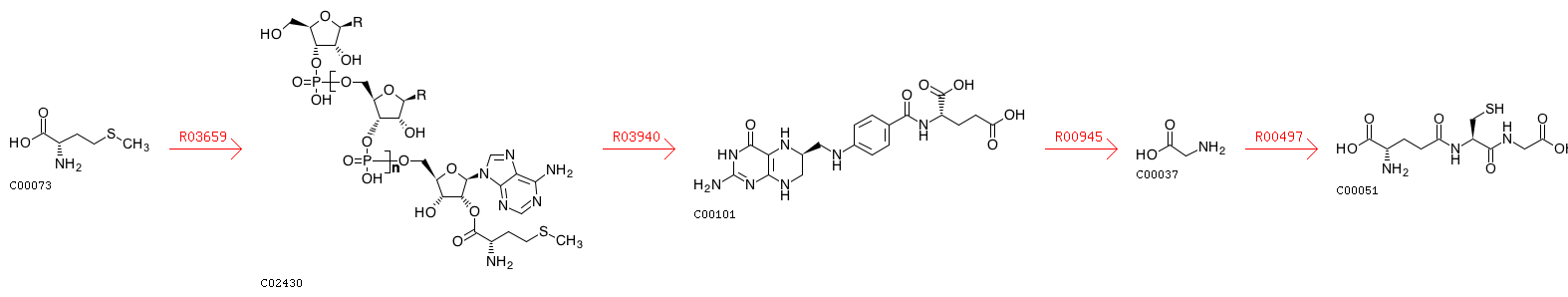


Figure 12 : Predicted shortest metabolic path with metabolite chemical structure displayed

► Click on the shortest path line [show compound structure].

► Look for other solutions and you will find a more relevant one

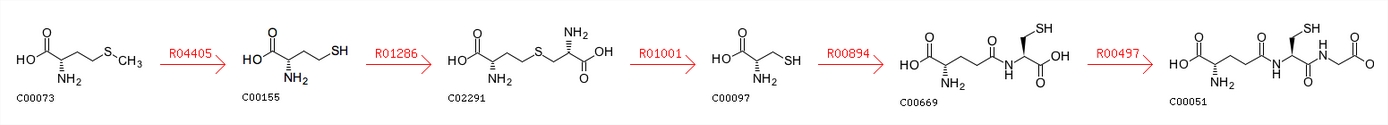


Figure 13: Metabolic path with a more relevant carbon chain conservation

Note that the main carbon chain length is strongly varying between each molecule (e.g. between the third and forth). It can even happen that water is used in a path.

See (4) for an overview of various algorithms to perform path search in metabolic networks.

**Remark**: Shortest paths may create unrealistic results between compounds.

# MetExplore: http://www.metexplore.fr

**It is better to use Chrome web browser.**

MetExplore (5) is a web server dedicated to the analysis of genome scale metabolic networks. A special care was taken to allow analyzing metabolomics data in the context of this network.

►Go on the MetExplore homepage and click on the "START MetExplore" button

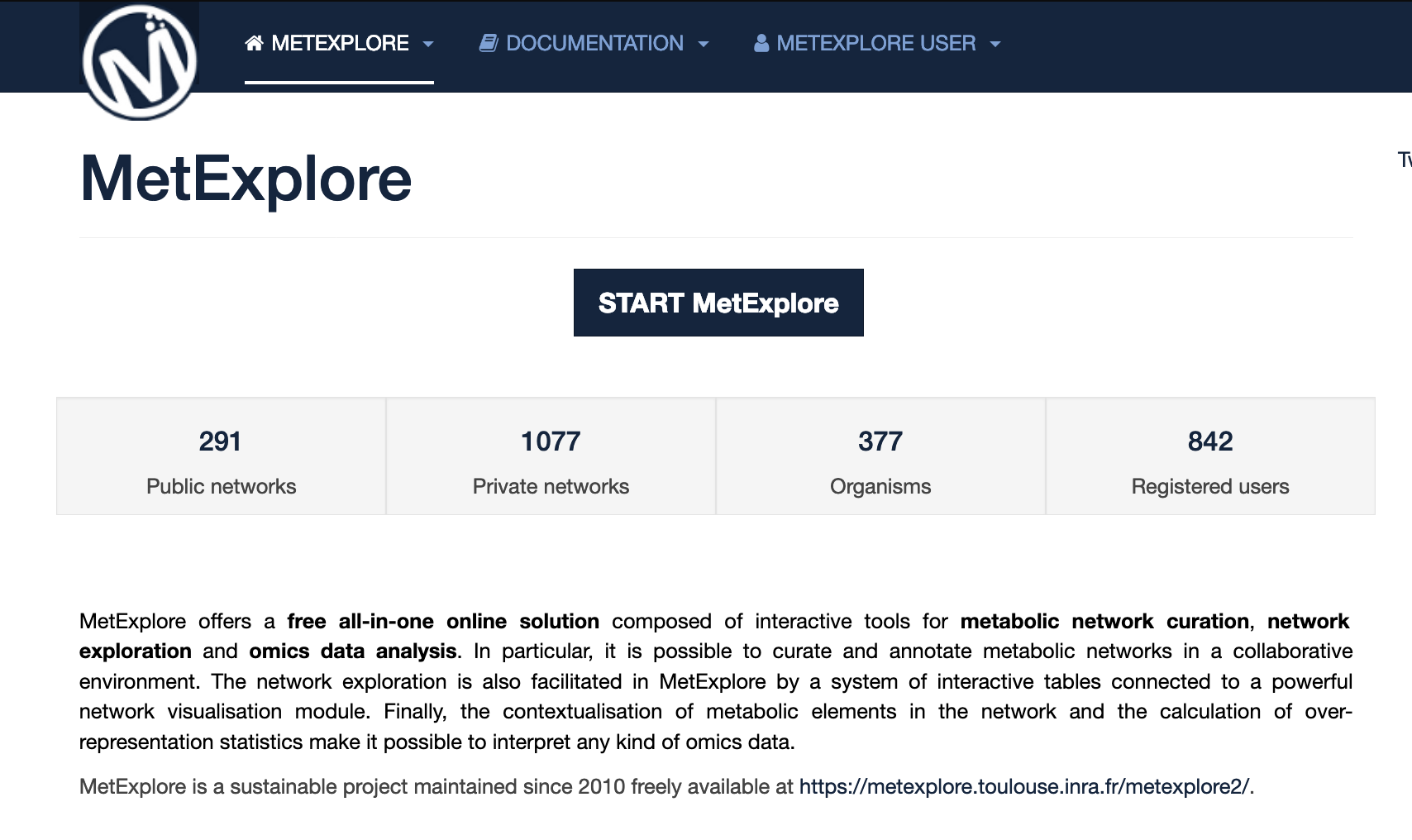
****

Figure 14: MetExplore homepage

MetExplore is providing access to BioCyc networks, KEGG networks and networks imported from SBML files (6, 7). All these files are present in the "public BioSource" repository. Note that you can also register and upload your own SBML files. You will then be the only one to be able to access these networks and will have the opportunity to share them with others.

►In the list of biosources click on: "*Saccharomyces cerevisiae*"

►Right click on the line of KEGG version of the network then click on "select Biosource" (you can also double click on it)

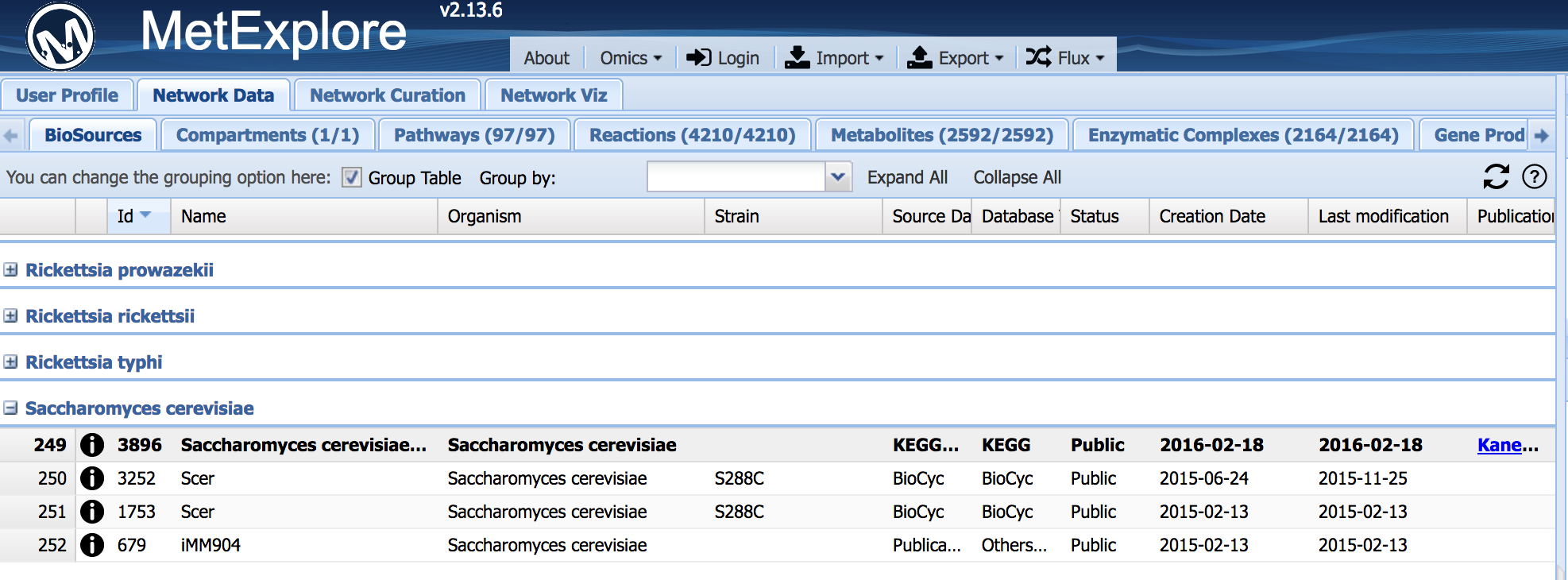


Figure 15: BioSource selection interface in MetExplore

The content of the metabolic network will be displayed in a tabular format.

Information on Compartments, Pathways, Reactions, Metabolites, Enzymes, Proteins and Genes are available.

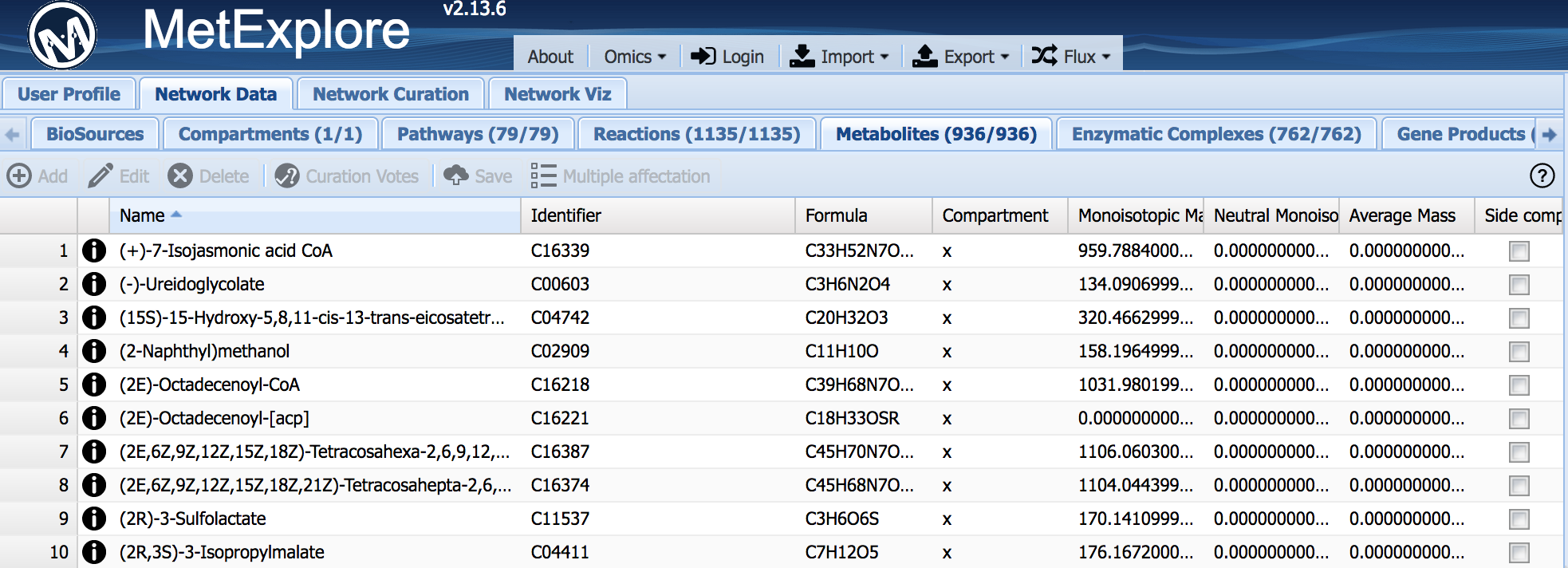
****

Figure 16: MetExplore representation of metabolic network content.

We are going to perform this mapping on yeast and using the metabolite identifiers found in the KEGG database

► Select "Omics->Mapping->New" in the menu at the top of the page.

► Copy (ctrl+C on the excel table) and paste (click on the first line of the MetExplore table and Ctrl+V)

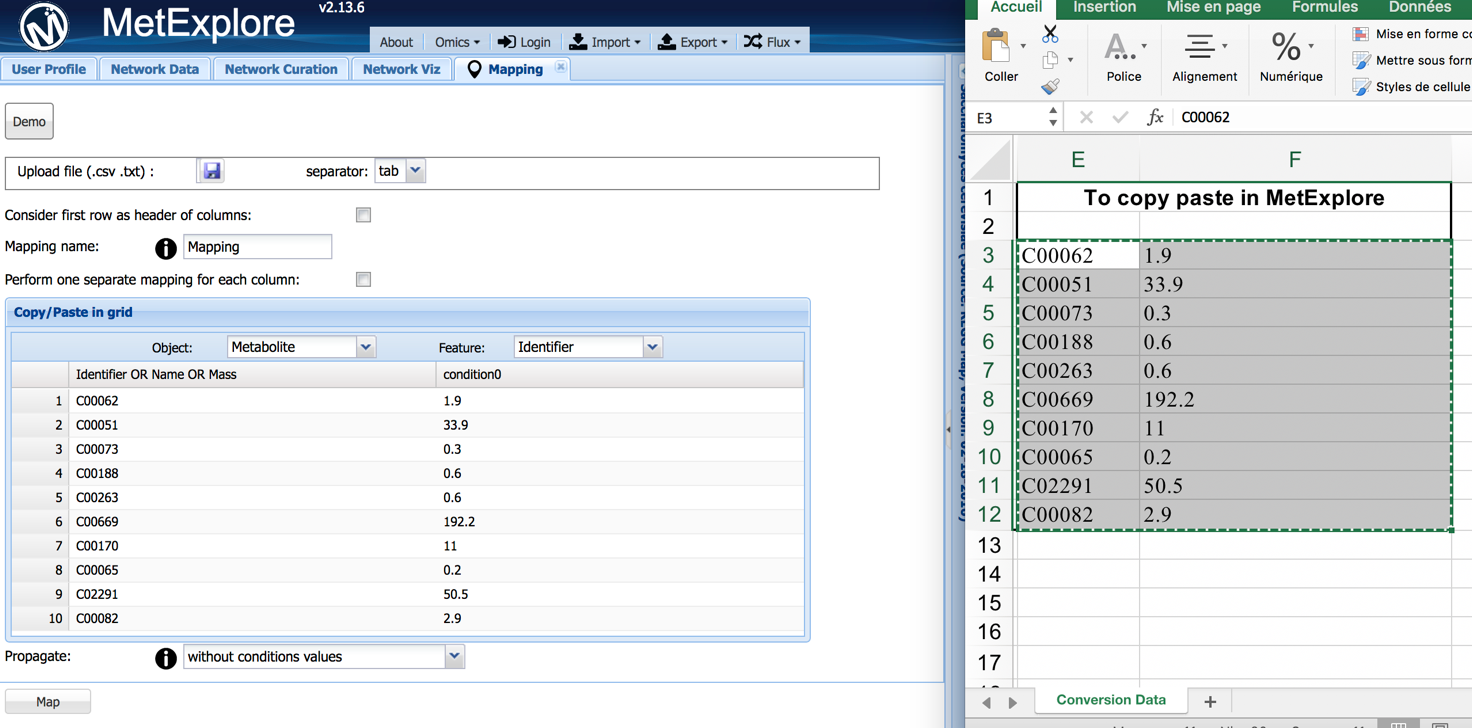


Figure 17: import metabolites in MetExplore

► Click on the "Map" button

Coming back to the tabular view, you will see which metabolites were mapped. (in the metabolite panel)

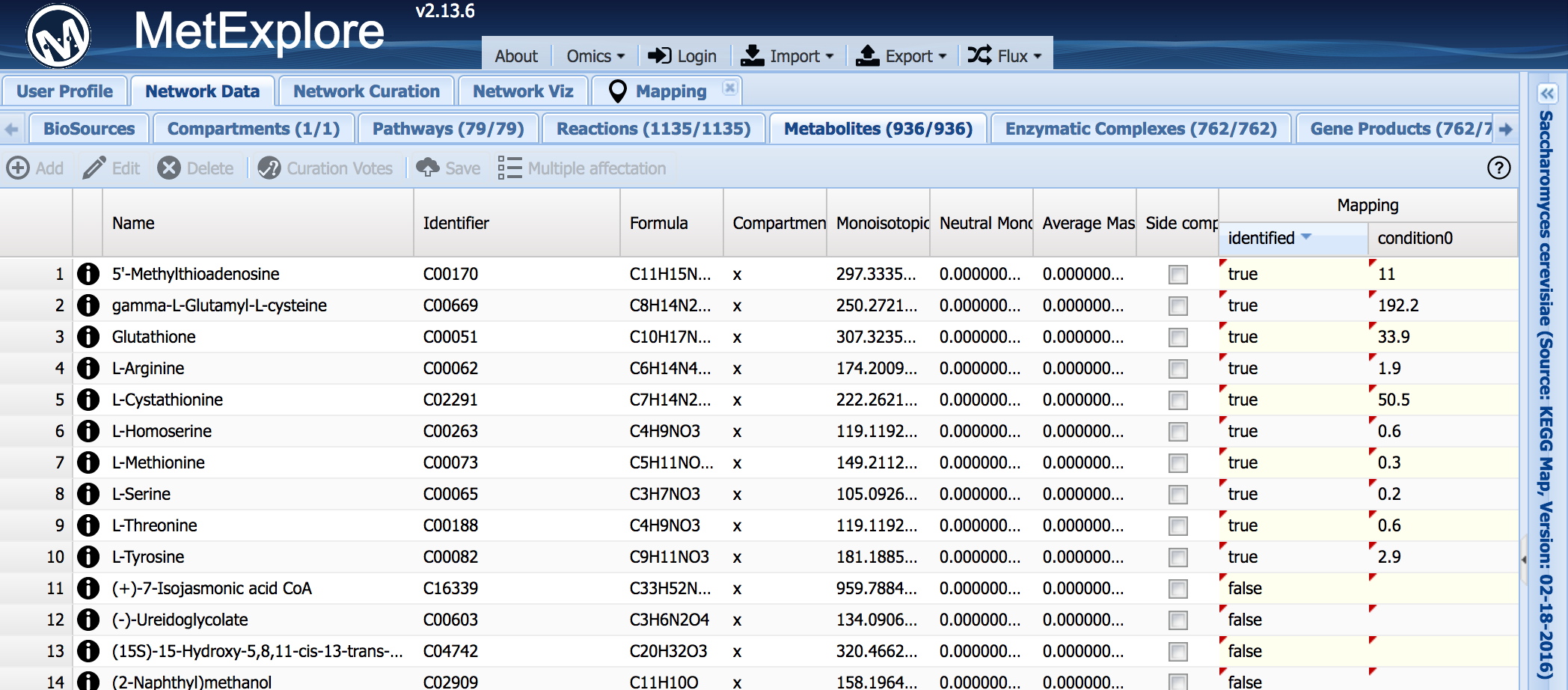


Figure 18: MetExplore output of a mapping.

By selecting lines in a grid it is possible (when right clicking on it) to filter all the other tables based on this selection.

Over-representation analyses are performed, using hypergeometric tests (corrected with Bonferroni or Benjamini Hochberg methods), and reported in the pathway grid to facilitate future interpretation of the mapping results.

► Go to the pathways grid and click on the Bonferroni column to sort pathways based on their p-values.

► Select the 4 first pathways which are above 0.05 threshold.

►Right click and Click on "New Filter on selection"

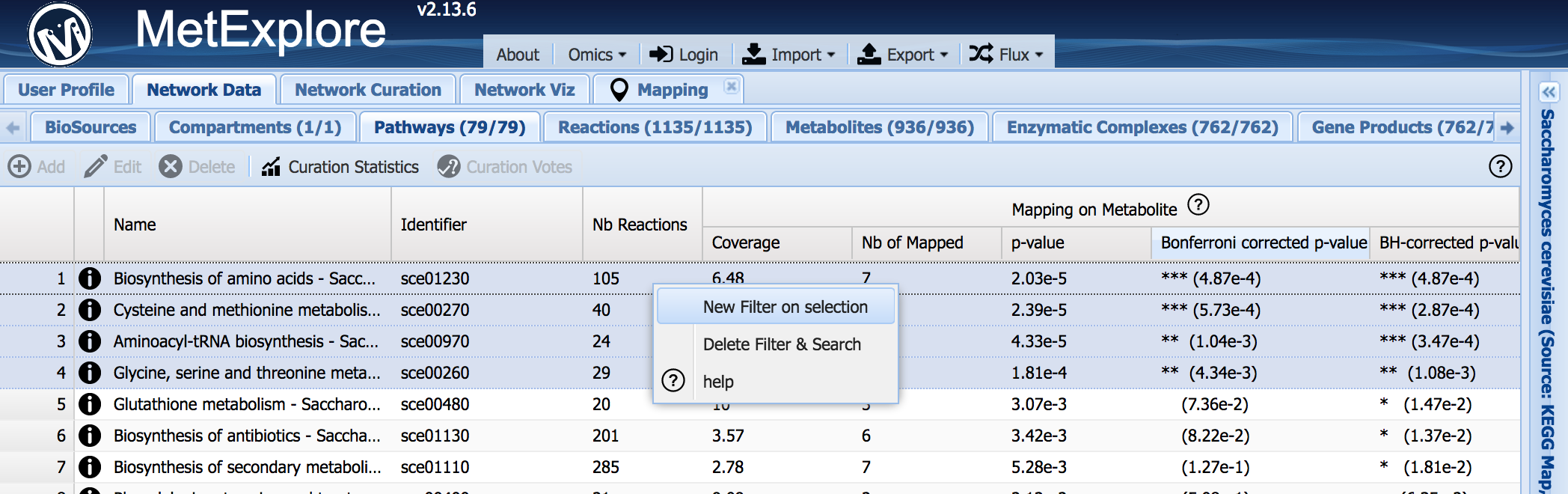


Figure 19: MetExplore filtering on pathways

Now, all the other panels will contain elements (reactions, metabolites...) corresponding to this pathway.

►Go to the reaction panel. Right click on the table and click on "Copy all to cart"

Now all selected reactions are in the car and can be used for visualization

►Select the "Network Viz" panel and cilck on PrtScr capture_1089 button to create the view.

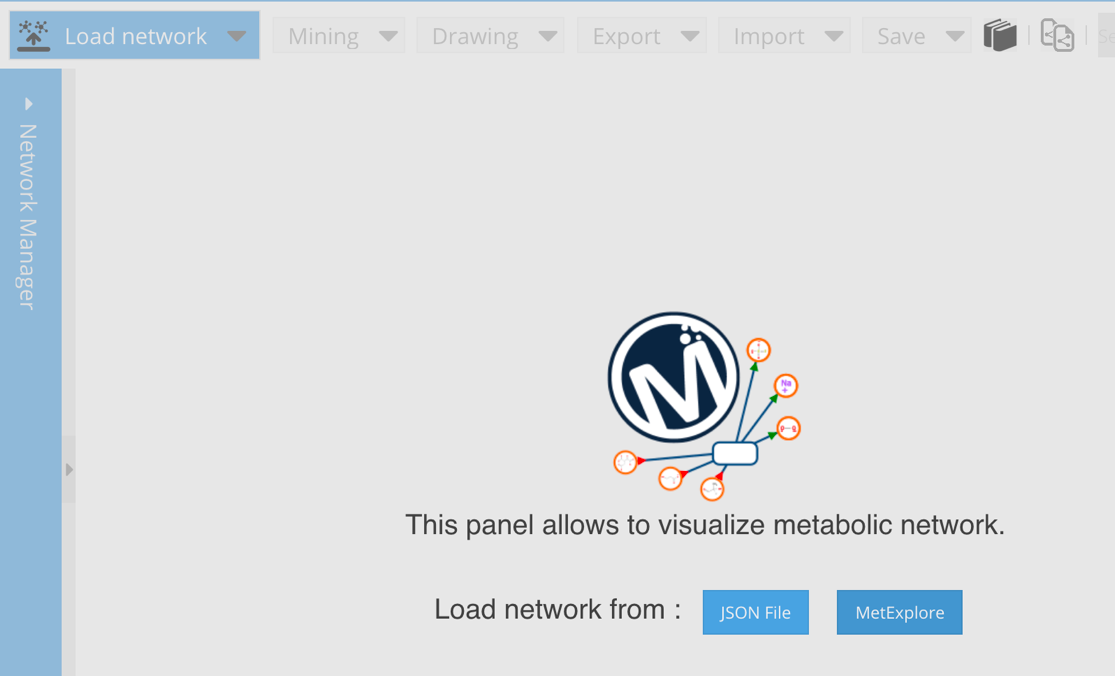


Figure 20: Network visualisation in MetExplore

►Click on the "Omics" menu and then on "Import mapping"

Metabolites mapped our now colored in blue.

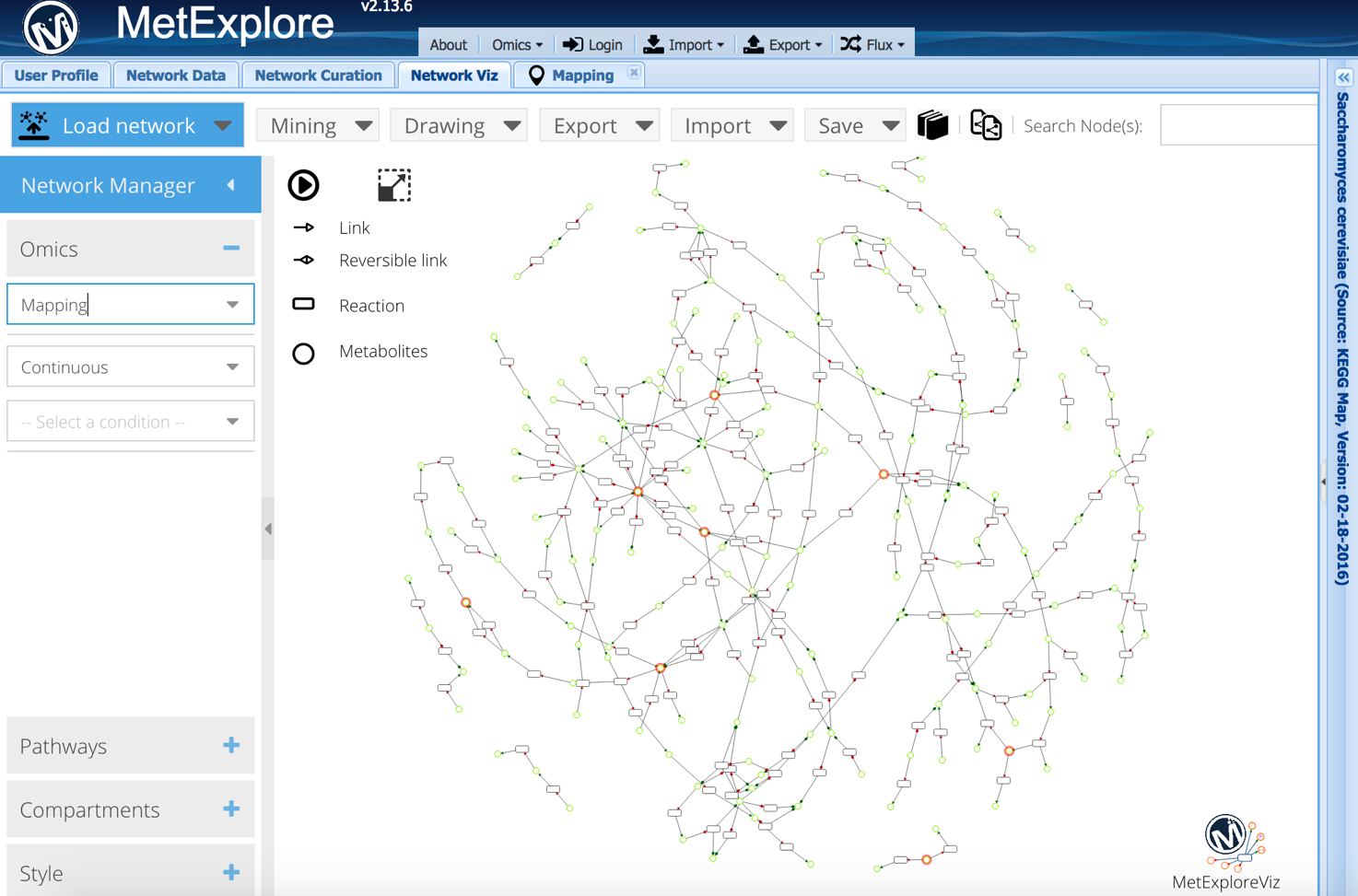


Figure 21: Network based on enriched pathways in MetExplore

►On the left of the panel in the "Omics" box select "Mapping". It will highlight in red metabolites mapped.

►Click on "Mining->Extract Subnetwork->From mapping" to keep only the union of all lightest paths between each pair on identified metabolites

►Select on the "Condition0" in the Omics panel to perform a continuous mapping of data on nodes.

Note that you can change the colors used.

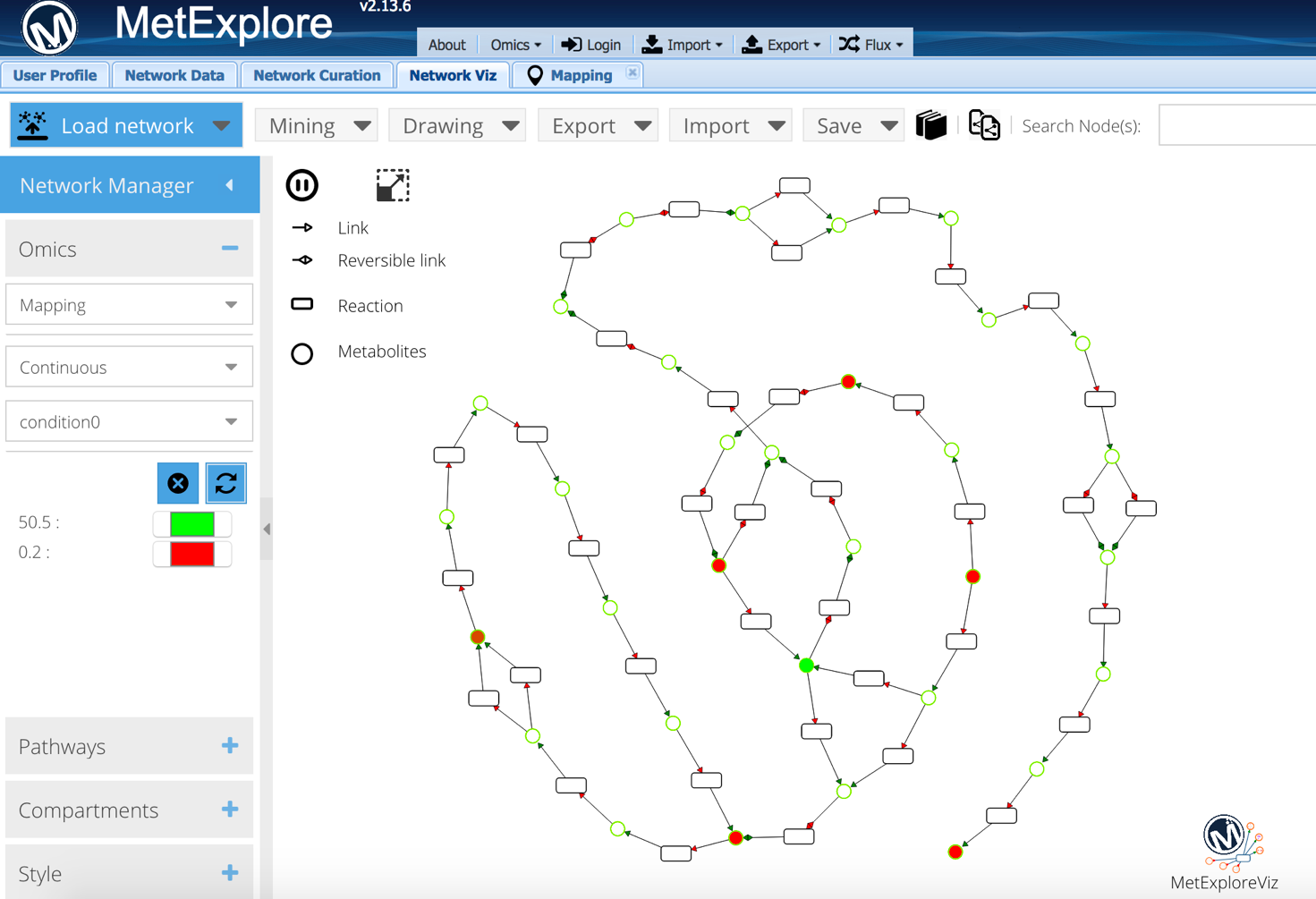


Figure 22: sub-network connecting identified metabolites

Note that 7 out of the 10 metabolites are included in this sub-network. This is due to the fact that we only looked at reactions involved in the selected pathways. Maybe we can connect other metabolites from the fingerprint using more reactions.

► Go back to the Reactions grid in the Network Data part of MetExplore

► Right click and select “Delete Filter & Search”

► Then right click and select “Copy all to cart”

► Go back to Network Viz and click on “Load Network->Load Network From Website”

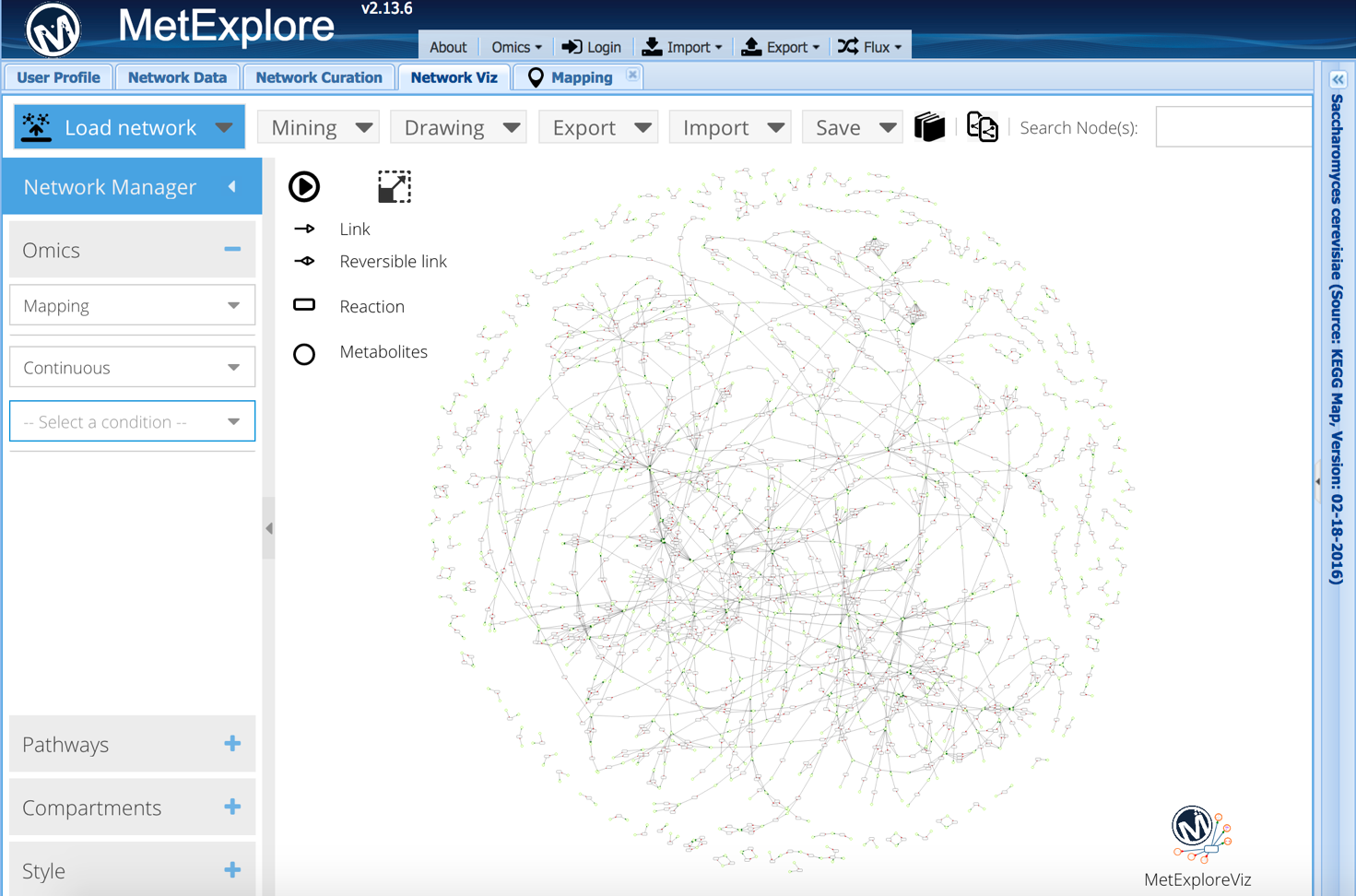


Figure 23 : Yeast metabolic network based on KEGG database.

Now we can map the fingerprint again and perform the sub-network extraction.

► In “Omics” select the mapping

► Extract sub-network using “Mining->Extract Subnetwork->From Mapping”

► Then right click and select “Copy all to cart”

► Go back to Network Viz and click on “Load Network->Load Network From Website”

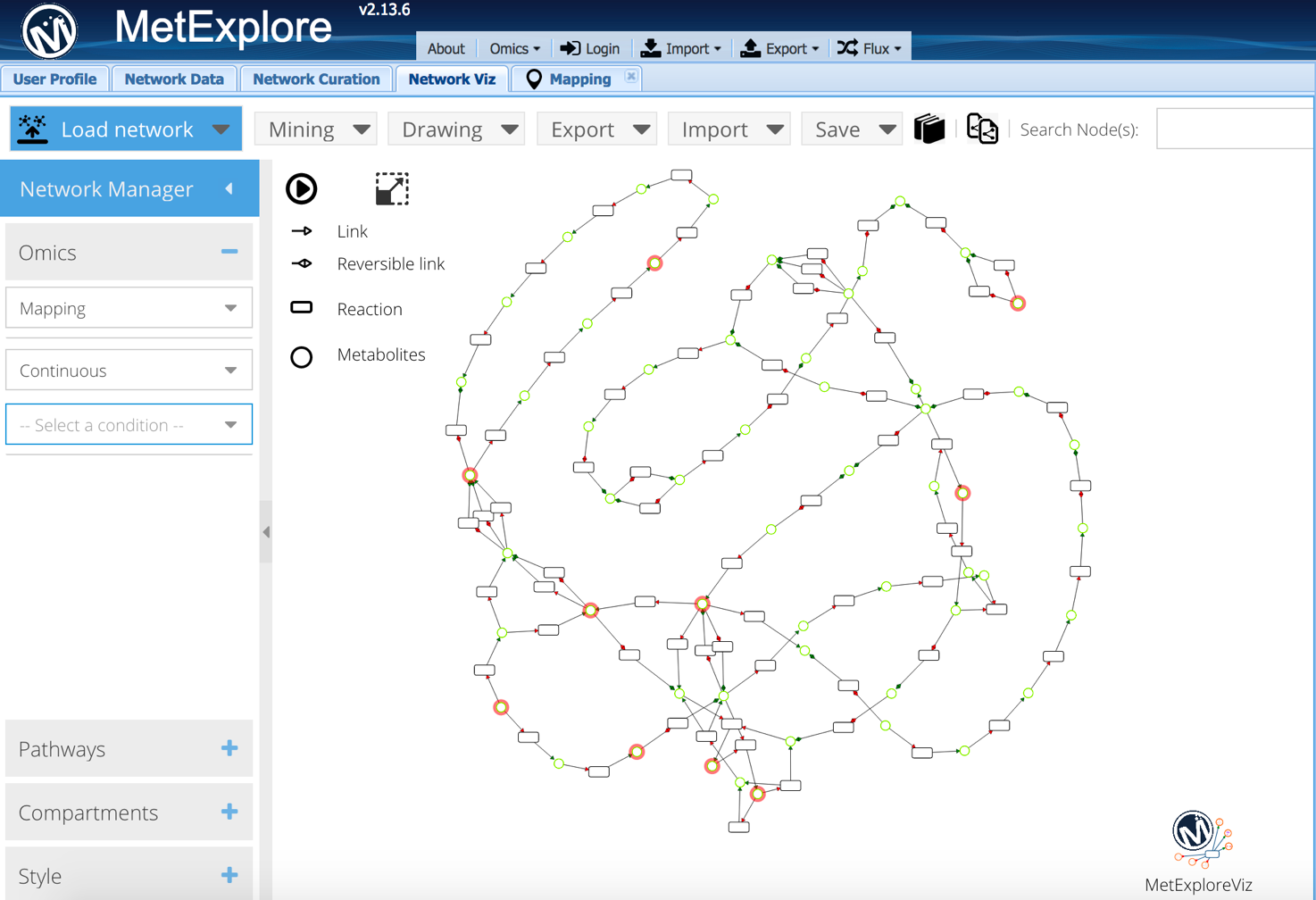


Figure 24 : Sub-network based on the entire Yeast network

Note that now all metabolites from the fingerprint are now included showing that we need to consider all pathways since process can span several pathways.

Finally, you can export the content of the network in excel spreadsheets.

► Click on “Export->Export Network to Excel”

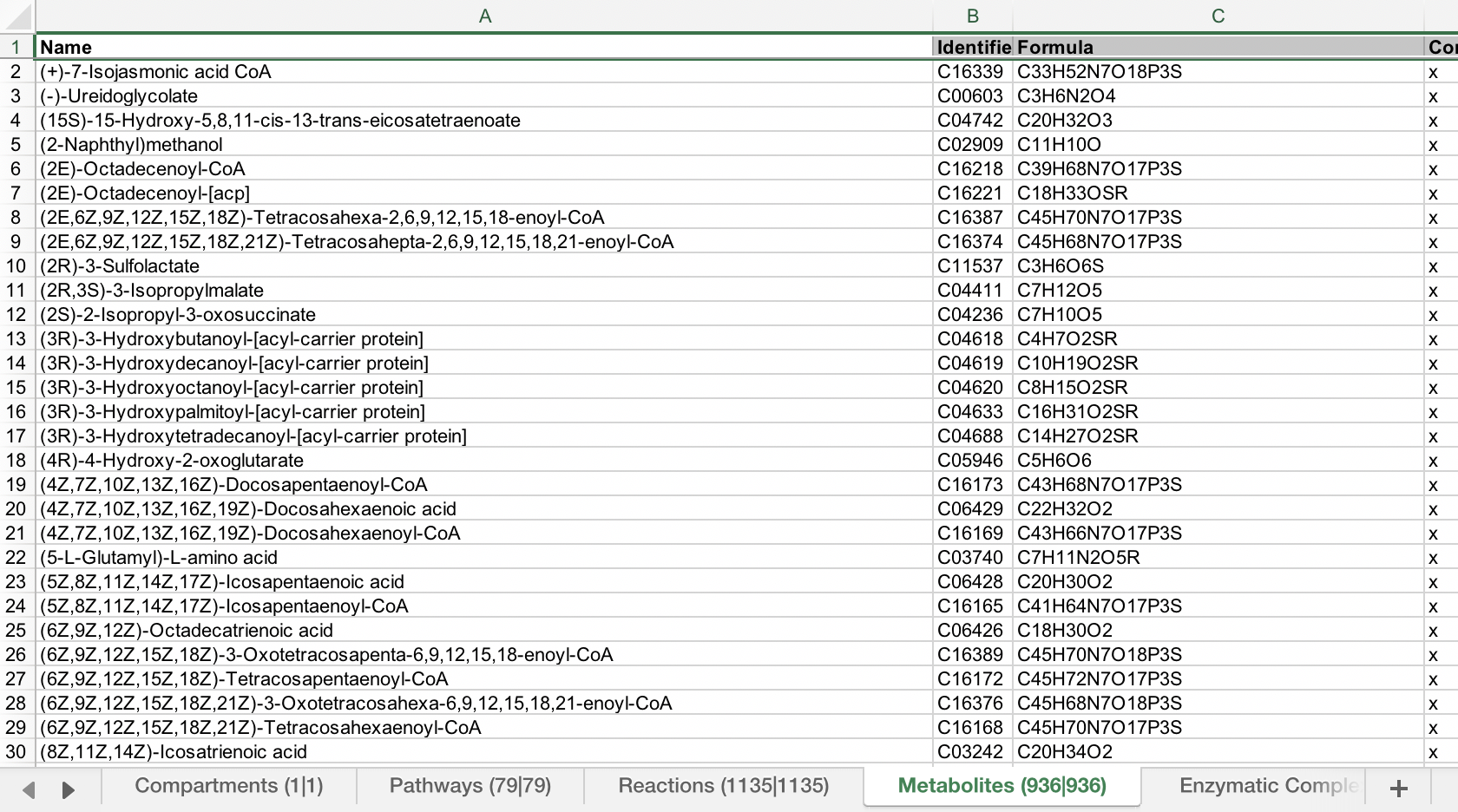


Figure 25 : Network content export in Excel.

# References

1. Madalinski,G., Godat,E., Alves,S., Lesage,D., Genin,E., Levi,P., Labarre,J., Tabet,J.-C., Ezan,E. and Junot,C. (2008) Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal. Chem.*, **80**, 3291–303.

2. Wohlgemuth,G., Haldiya,P.K., Willighagen,E., Kind,T. and Fiehn,O. (2010) The Chemical Translation Service--a web-based tool to improve standardization of metabolomic reports. *Bioinformatics*, **26**, 2647–8.

3. Kanehisa,M., Furumichi,M., Tanabe,M., Sato,Y. and Morishima,K. (2017) KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.

4. Frainay,C. and Jourdan,F. (2017) Computational methods to identify metabolic sub-networks based on metabolomic profiles. *Brief. Bioinform.*, **18**.

5. Cottret,L., Wildridge,D., Vinson,F., Barrett,M.P., Charles,H., Sagot,M.-F. and Jourdan,F. (2010) MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.*, **38**, W132-7.

6. Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D., Cornish-Bowden,A., *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–31.

7. Hucka,M. and Smith,L.P. (2016) SBML Level 3 package: Groups, Version 1 Release 1. *J. Integr. Bioinform.*, **13**, 290.