MAPPING OMICS DATASETS ON KEGG METABOLIC PATHWAYS

MASTER'S PROJECT IN BIOINFORMATICS

DUARTE ALVES VELHO

MAY 2024

Under the guidance of:

Andreia Salvador;

João Sequeira

STATE OF THE ART

- Omics and meta-omics technologies are powerful approaches for exploring the functions of microorganisms;

- software developed for omics and meta-omics analysis is a valuable resource for analysis;

- Visualizing the results in pathways allows for a greater understanding of biological processes, particularly in MT experiments where different genes/proteins and pathways are differentially expressed, but there are not many resources to do so;

- There are several bioinformatics resources for meta-omics analysis, but many of them require significant computational expertise.

Tools	What tools do	Advantages	Disadvantages	Languages	Type of Identifiers	Interac-tivity	Number of Samples	Ta-xonomy	Dif-fe-ren-tial Ex-pres-sion	Input	Output
DAVID	Integrates functional genomic annotations with intuitive graphical	Provides exploratory visualization tools for functional classification	Requirement for internet access	Perl and Java	Gene IDs, protein IDs	Yes	Not specified	Yes	Yes	Txt	HTML
Reactome	It provides an intuitive web-based user interface	Integration with various data analysis tools for ID mapping	Limited by the breadth of its curated data	JavaScript, Perl	UniProt IDs, GenBank/EMBL/DDBJ IDs, RefSeq IDs, EntrezGene IDs	Yes	Not specified	Human and 20 other species	Yes	Txt	SVG, PNG, EPS, PDF, MySQL, BioPAX, SBML, PSI-MI-TAB
CellDesigner	Modeling tool that allows users to visualize, model, and simulate biochemical networks	Offers a comprehensive graphical representation of networks	Requires understanding of SBML	Java	Not specified	Yes	Not applicable	No	Yes	SBML	SBML, PNG,SVG
GenMAPP	Allows users to visualize and analyze genome-scale data	Offers new visualization modes for complex data analysis	Limited primarily to the species and gene ids within its database	Visual Basic 6.0	Ensembl gene IDs, UniProt IDs, Entrez Gene IDs, Affymetrix probe set IDs	Yes	Not specified	Yes	Yes	Microsoft Jet format for (.gex), (.gdb), (.mapp)	HTML, PNG, SVG

Fig. 1 - Example of tools that map metabolic pathways.

Fig. 2 - Representation of MAP00680 in KEGG.

KEGG (KYOTO ENCYCLOPEDIA OF GENES AND GENOMES)

- COMPREHENSIVE DATABASE INTEGRATING GENOMIC, CHEMICAL AND SYSTEMIC FUNCTIONAL INFORMATION;
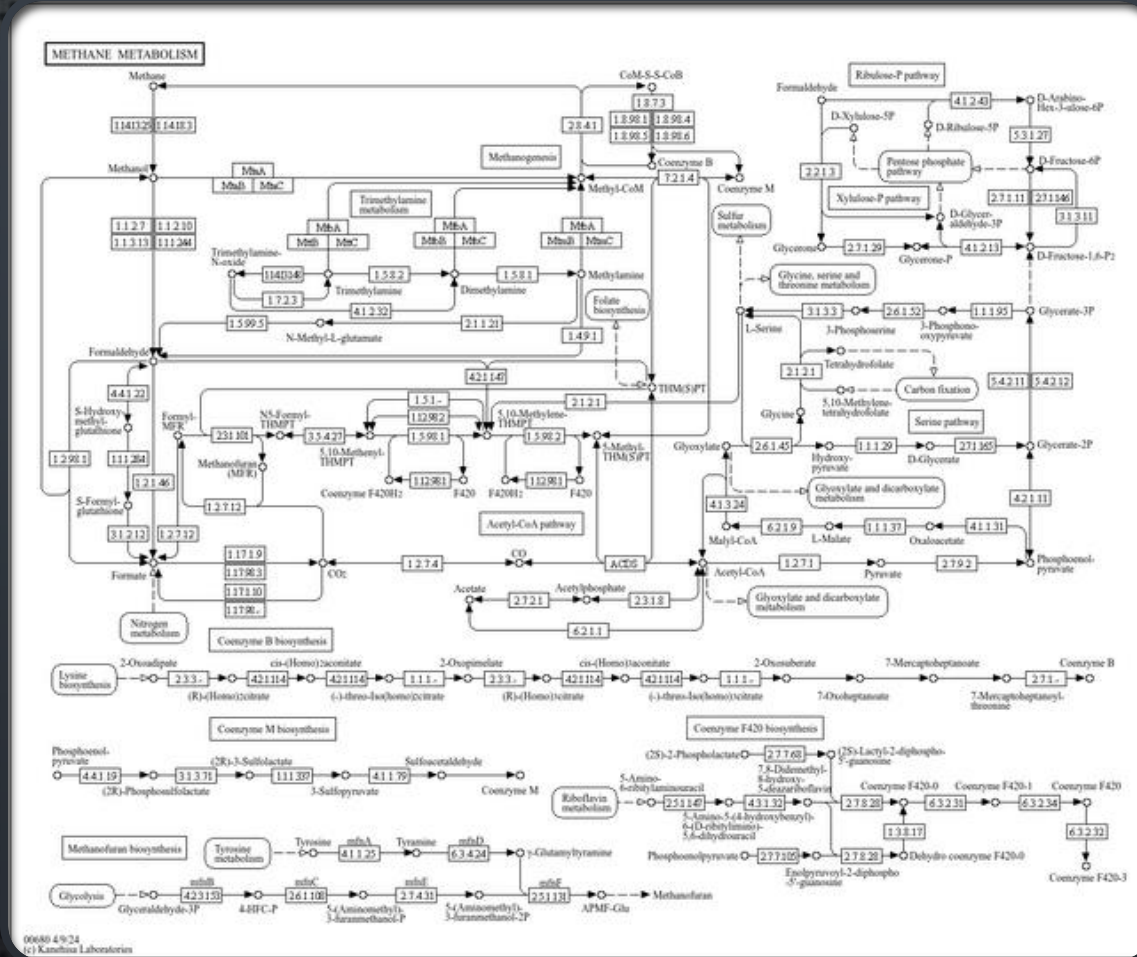
- ALLOWS TO VIEW THE RESULTS IN PATHWAYS.

KEGGCHARTER

- Command line interface;

- Easy to use;

- Tool designed to overlay the differential gene expression and taxonomic information of enzymes on KEGG metabolic maps;
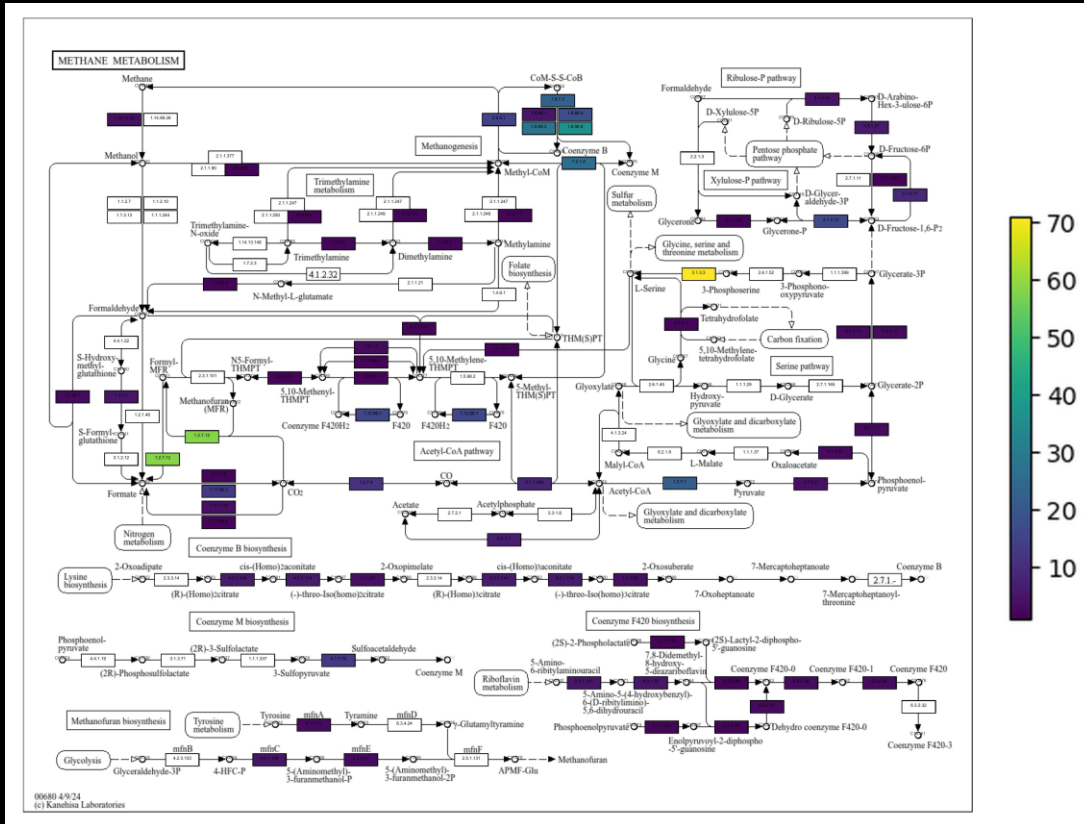
- Based on Biopython.

Fig. 3 - KEGG metabolic map of methane metabolism, with differential analysis of quantified expression for each function from a simulated dataset.
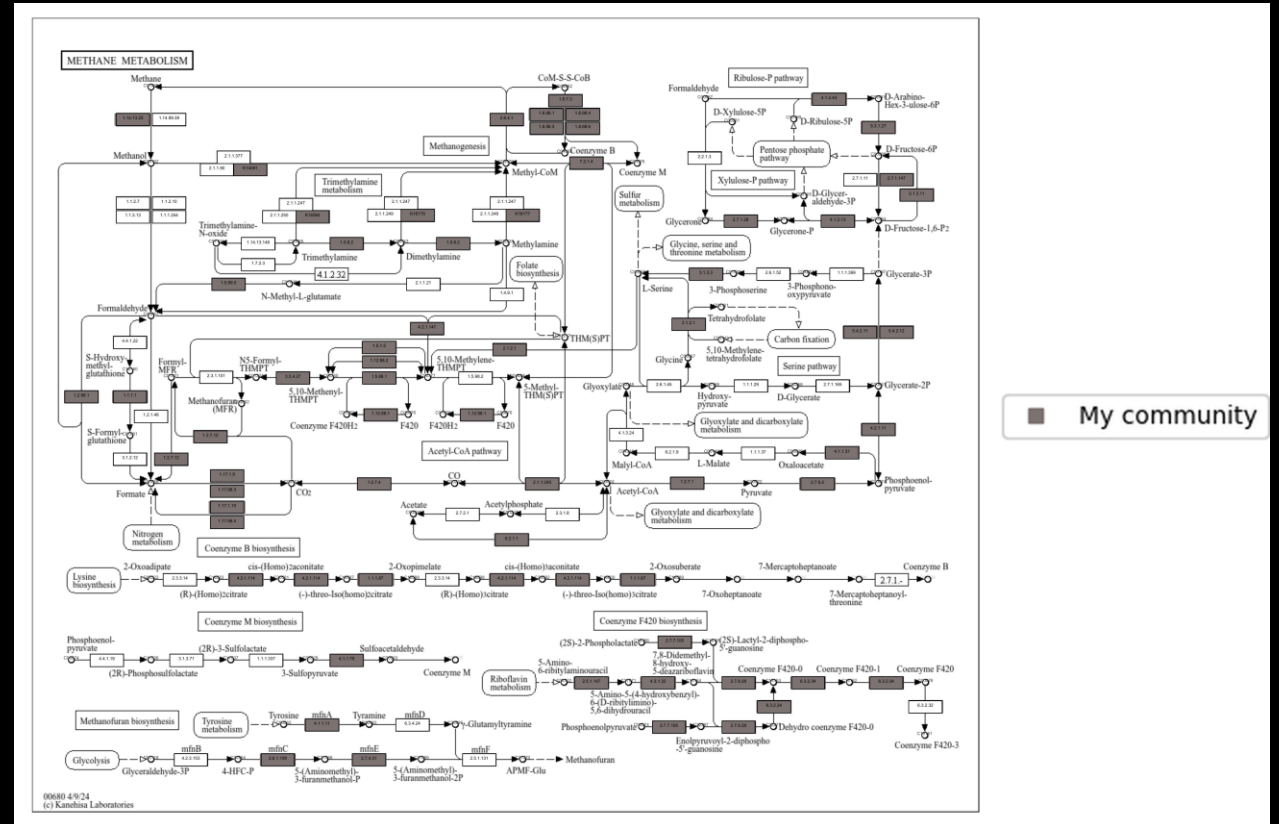
Fig. 4 - KEGG metabolic map of methane metabolism, with identified taxa for each function from a simulated dataset.

• OBJECTIVE OF THIS PROJECT

- Expand the KEGGCharter graphs to include multi-level representation of gene expression information;

- Development of new interactive representations on top of the graphs currently produced, to allow more information to be included in metabolic maps, combining taxonomy and expression levels in the same graph.

• KEGGCHARTER'S WORKFLOW
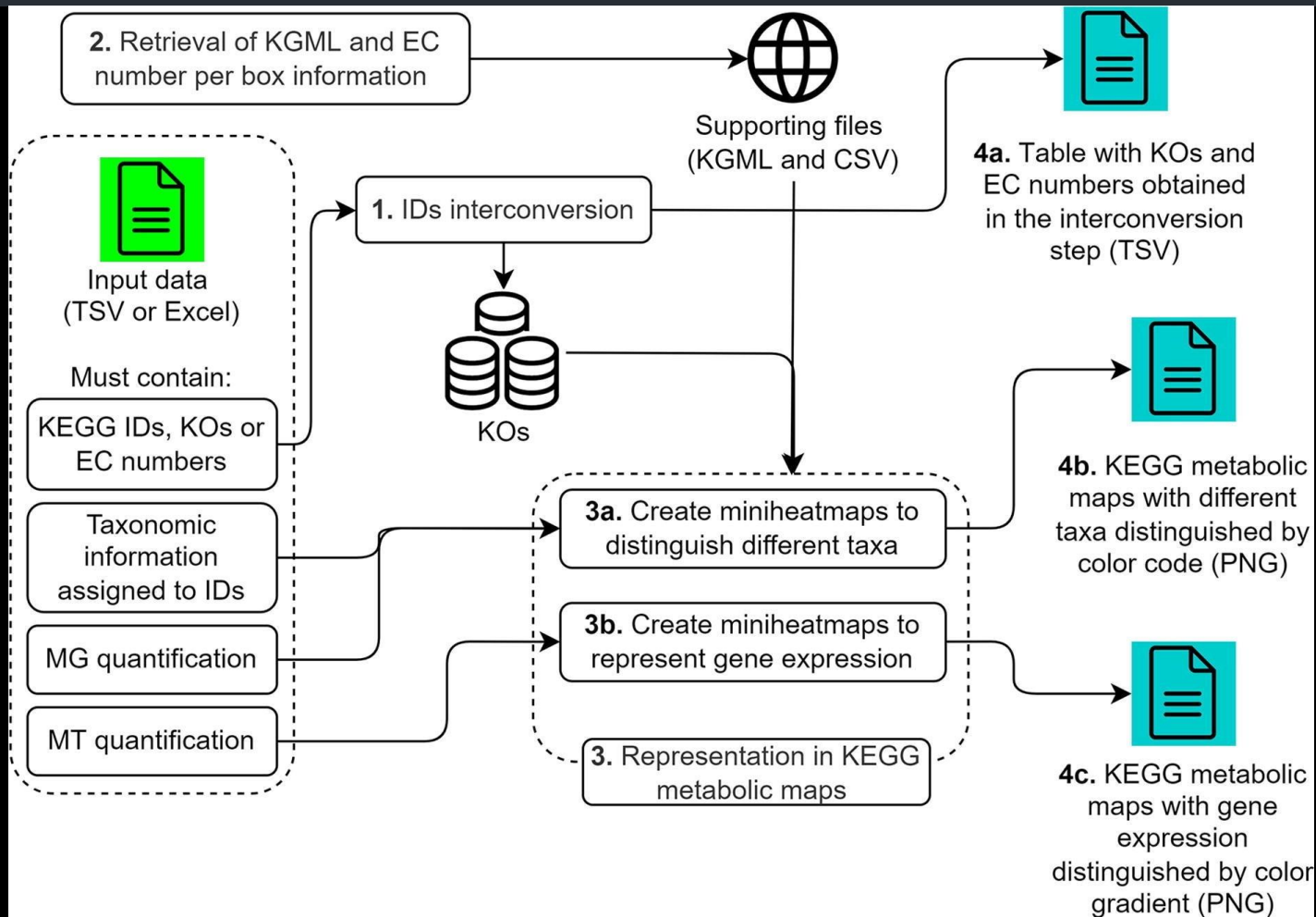
Source: [1] UPIMAPI, reCOGnizer and KEGGCharter: Bioinformatics tools for functional annotation and visualization of (meta)-omics datasets.
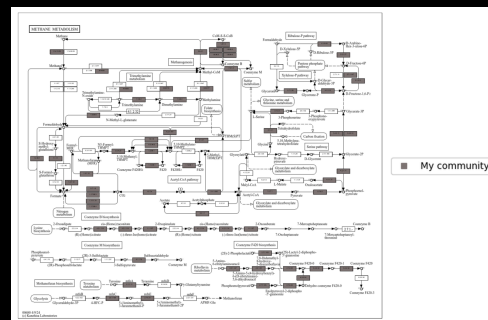
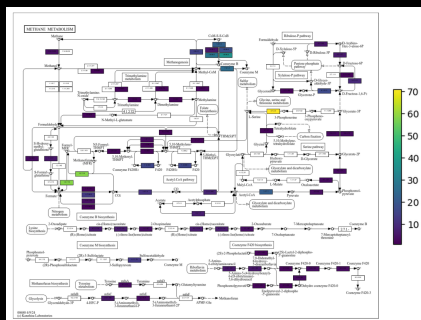# PROJECT MANAGEMENT TO DATE

STEP 1: RUN KEGGCHARTER TO SEE THE OUTPUTS;

```
(base) duartevelho1@DESKTOP-ROLJJ8V:~/Projeto/KEGGCharter$ keggcharter -f keggcharter_input.tsv -rd resource
s_directory -keggc 'KEGG' -koc 'KO' -ecc 'EC number' -cogc 'COG ID' -iq -it "My community" -mm 00680 -o firs
t_time_running_KC
```

KGML

resources_directory
first_time_running_KC

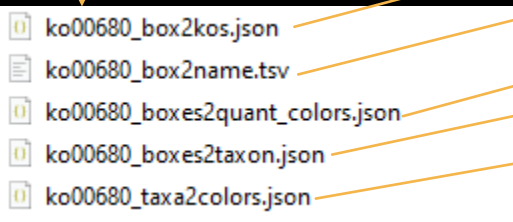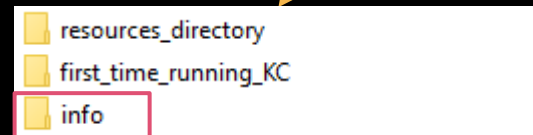Two PNG Maps (differential; potential)


```xml
<!-- Creation date: Apr 9, 2024 11:17:17 +0900 (GMT+9) -->
<pathway name="path:ko00680" org="ko" number="00680" title="Methane metabolism" image="https://www.kegg.jp/kegg/pathway/ko/ko00680.png" link="https://www.kegg.jp/kegg-bin/show_pathway?ko00680">
  <entry id="4" name="path:ko00680" type="map" link="https://www.kegg.jp/dbget-bin/www_bget?ko00680">
    <graphics name="TITLE:Methane metabolism" fgcolor="#000000" bgcolor="#FFFFFF" type="roundrectangle" x="131" y="58" width="181" height="25"/>
  </entry>
  <entry id="33" name="cpd:C01438" type="compound" link="https://www.kegg.jp/dbget-bin/www_bget?C01438">
    <graphics name="C01438" fgcolor="#000000" bgcolor="#FFFFFF" type="circle" x="164" y="95" width="8" height="8"/>
  </entry>
  <entry id="35" name="ko:K16157 ko:K16158 ko:K16159 ko:K16160 ko:K16161 ko:K16162" type="ortholog" reaction="rn:R01142" link="https://www.kegg.jp/dbget-bin/www_bget?K16157+K16158+K16159+K16160+K16161+K16162">
    <graphics name="K16157..." fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="139" y="140" width="46" height="17"/>
  </entry>
  <entry id="36" name="ko:K10944 ko:K10945 ko:K10946" type="ortholog" reaction="rn:R09518" link="https://www.kegg.jp/dbget-bin/www_bget?K10944+K10945+K10946">
    <graphics name="K10944..." fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="189" y="140" width="46" height="17"/>
  </entry>
  <entry id="37" name="cpd:C00132" type="compound" link="https://www.kegg.jp/dbget-bin/www_bget?C00132">
    <graphics name="C00132" fgcolor="#000000" bgcolor="#FFFFFF" type="circle" x="164" y="200" width="8" height="8"/>
  </entry>
  <entry id="38" name="cpd:C00067" type="compound" link="https://www.kegg.jp/dbget-bin/www_bget?C00067">
    <graphics name="C00067" fgcolor="#000000" bgcolor="#FFFFFF" type="circle" x="164" y="410" width="8" height="8"/>
  </entry>
  <entry id="40" name="ko:K17066" type="ortholog" reaction="rn:R00608" link="https://www.kegg.jp/dbget-bin/www_bget?K17066">
    <graphics name="K17066" fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="139" y="274" width="46" height="17"/>
  </entry>
  <entry id="41" name="ko:K14028 ko:K16254 ko:K16255 ko:K14029 ko:K16256 ko:K16257 ko:K16258 ko:K16259 ko:K16260" type="ortholog" reaction="rn:R01146" link="https://www.kegg.jp/dbget-bin/www_bget?K14028+K16254+K16255+K14029+K16256+K16257+K16258+K16259+K16260">
    <graphics name="K14028..." fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="139" y="255" width="46" height="17"/>
  </entry>
  <entry id="42" name="ko:K00093" type="ortholog" reaction="rn:R00605" link="https://www.kegg.jp/dbget-bin/www_bget?K00093">
    <graphics name="K00093" fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="189" y="274" width="46" height="17"/>
  </entry>
  <entry id="44" name="ko:K00148" type="ortholog" reaction="rn:R00604" link="https://www.kegg.jp/dbget-bin/www_bget?K00148">
    <graphics name="K00148" fgcolor="#000000" bgcolor="#BFBFFF" type="rectangle" x="164" y="549" width="46" height="17"/>
```

STEP 2: THROUGH MODIFICATIONS TO THE TOOL, GENERATE NEW RESULTS;

```
(base) duartevelho1@DESKTOP-ROLJJ8V:~/Projeto/KEGGCharter$ python keggcharter.py -f keggcharter_input.tsv -r
d resources_directory -keggc 'KEGG' -o KC_additions -mm 00680 --map-all --taxa-column 'Taxonomic lineage (GE
NUS)' -qcol MP1,MP2,MP3,MP4
```

📁 resources_directory
📁 first_time_running_KC
📁 info

📄 ko00680_box2kos.json
📄 ko00680_box2name.tsv
📄 ko00680_boxes2quant_colors.json
📄 ko00680_boxes2taxon.json
📄 ko00680_taxa2colors.json

- Box id and KOs;
- Box id and EC number or KO;
- Taxon and colors;
- Box id and taxon;
- Box id and colors.

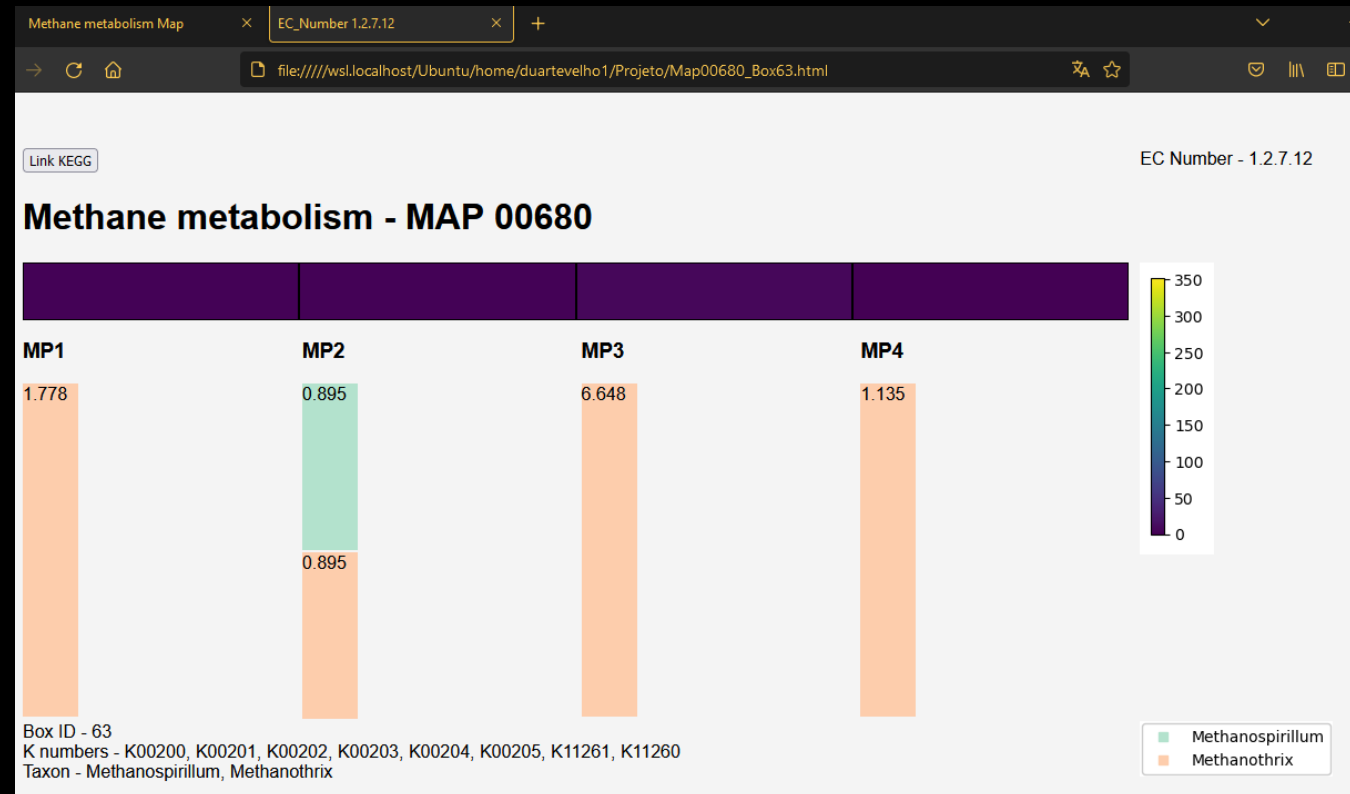# STEP 3: MAKE THE PNG IMAGES CLICKABLE/ INTERACTIVE;
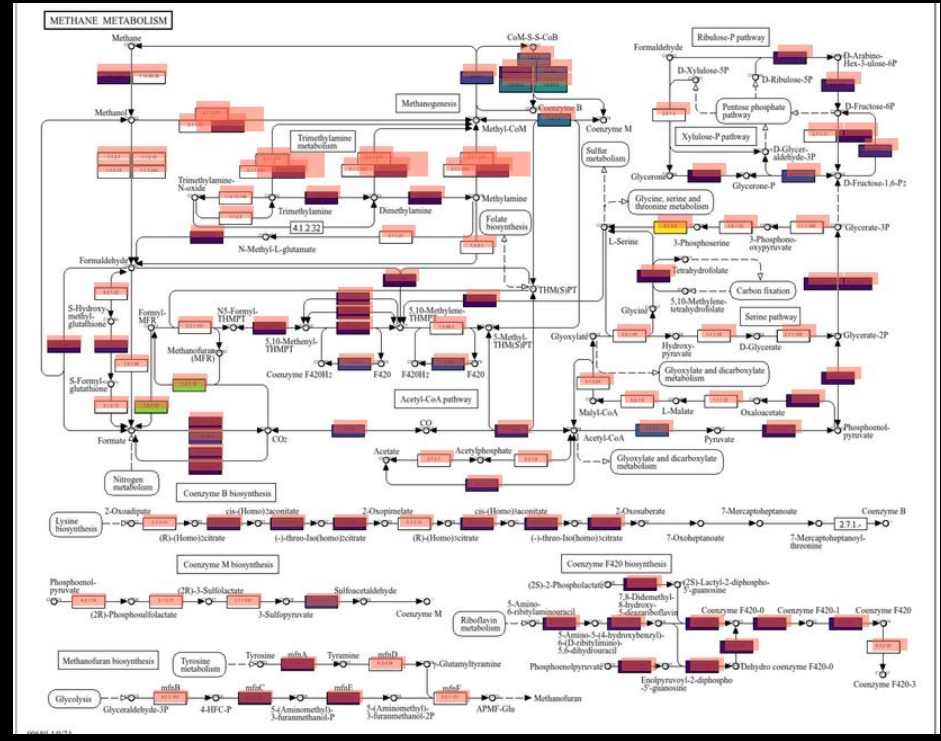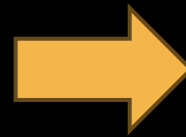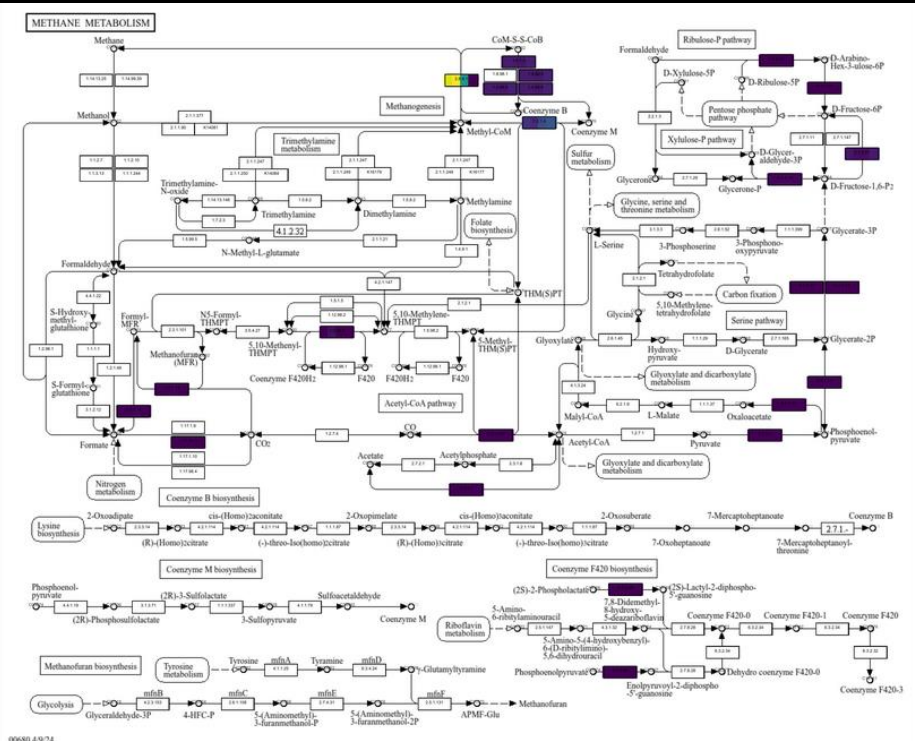
STEP 4: CREATE DYNAMIC HTML PAGES FOR EACH MAP BOX;
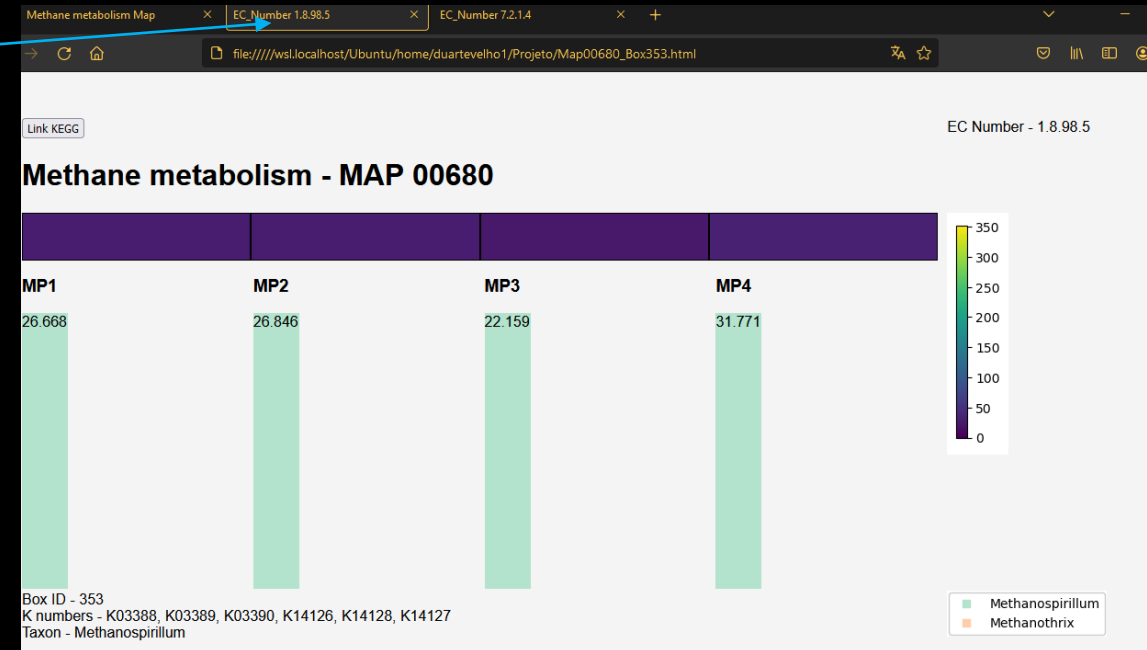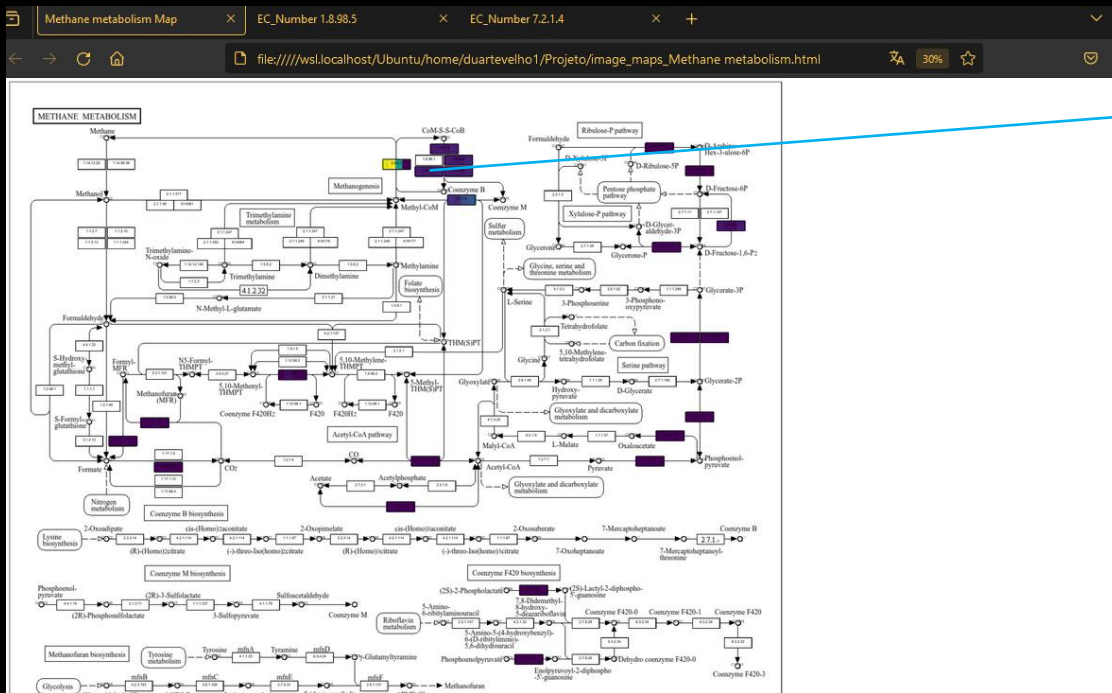
STEP 5: ENRICH THE HTML PAGES WITH RELEVANT INFORMATION.

- What's new: We were able to represent a new quantification, associated with taxonomy (until now it didn't exist)

- # SUMMARY

- STEP 1: RUN KEGGCHARTER TO SEE THE OUTPUTS;

- STEP 2: THROUGH MODIFICATIONS TO THE TOOL, GENERATE NEW RESULTS;

- STEP 3: MAKE THE PNG IMAGES CLICKABLE/INTERACTIVE;

- STEP 4: CREATE DYNAMIC HTML PAGES FOR EACH MAP BOX;

- STEP 5: ENRICH THE HTML PAGES WITH RELEVANT INFORMATION.

Thanks!

BIBLIOGRAPHY

- [1] Sequeira, J.C., Rocha, M., Alves, M.M., Salvador, A.F.: UPIMAPI reCOGnizer and KEGGCharter: Bioinformatics tools for functional annotation and visualization of (meta)-omics datasets. Comput. Struct. Biotechnol. J. 20, 1798–1810 (2022).

- Dennis Jr, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., & Lempicki, R.A.: DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 4(9), R60 (2003).

- Haw, R., Hermjakob, H., D'Eustachio, P., & Stein, L. (2011). Reactome pathway analysis to enrich biological discovery in proteomics data sets. Proteomics, 11(18), 3598–3613.

- Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., Kitano, H.: CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks. Proceedings of the IEEE 96(8), 1254–1265 (2008). https://doi.org/10.1109/JPROC. 2008.925458.

- Salomonis, N., Hanspers, K., Zambon, A.C., Vranizan, K., Lawlor, S.C., Dahlquist, K.D., Doniger, S.W., Stuart, J., Conklin, B.R., Pico, A.R.: GenMAPP 2: new features and resources for pathway analysis. BMC Bioinformatics 8, 217 (2007).