Mapping Omics datasets on KEGG Metabolic Pathways

Duarte Velho[1], João Sequeira[2], and Andreia Salvador[2]

[1] School of Engineering, Minho University, Campus de Azurém, 4800-019, Guimarães, Portugal
[2] Centre of Biological Engineering, Minho University, Campus de Gualtar, 4710 - 057, Braga, Portugal, direcao@ceb.uminho.pt
https://www.ceb.uminho.pt/

Abstract. This study addresses the challenges of omics data analysis, focusing on the development of improved methods for mapping and interpreting complex networks of biological interactions using omics data. With the increasing complexity and volume of data produced by omics technologies, mapping methods are essential to simplify and organize this information. This project focuses on improving the KEGGCharter tool, which is a tool for mapping metabolic pathways, by integrating interactive functionalities. Through an original workflow approach, the development of an interactive HTML interface to combine gene expression data and taxonomic data in the same metabolic map was explored. At the moment, with KEGGCharter it is possible to visualize enzyme expression by clicking on the EC number boxes on the map, and directly link the expression to the taxonomic assignment. In this way, it is possible to interactively visualize which organisms express a certain function and compare them with the rest of the community. These new features of KEGGCharter are particularly relevant when analyzing meta-omics data obtained from complex microbial communities. This project not only improves the usefulness of existing omics data mapping tools, but also facilitates the interpretation and visualization of complex data, making a significant contribution to the landscape of metabolic pathway mapping tools.

Keywords: KEGGCharter · Metabolic pathways mapping · Metagenomics · Metatranscriptomics · Differential expression · Dynamic HTML · Interactivity · Bioinformatic tools.

1 Introduction

1.1 Importance of mapping omics data

Omics and meta-omics technologies are powerful approaches for exploring the functions of microorganisms, but the size and complexity of omics datasets often makes analyzing them a difficult task.

Microbial communities are made up of bacteria, archaea, fungi, yeasts, eukaryotes and viruses, which often live together in the same habitat. With this

size and complexity of microbial communities, the omics data that results from them makes analyzing them a difficult task [23]. Therefore, omics data mapping softwares are of great importance for the interpretation of omics results, as the visualization of the results in pathways will allow for a greater understanding of biological processes. However, there are not many resources to do this.

Over the past decade, omics data mapping has emerged as a valuable aid to understanding the data generated by various "omics" technologies, which have gained widespread interest among researchers in recent years due to their complexity and potential to generate new medical knowledge [24]. Consequently, a number of robust software tools, as well as web interfaces for meta-omics analysis, have been developed to support pathway analysis for genomic and proteomic studies [26].

1.2 Current challenges in mapping omics data

Given the need to map omics data for better understanding, software developed for omics and meta-omics analysis, together with knowledge bases that include information on genes, proteins, taxonomic and functional annotation, among other types of information [23], have become powerful resources for omics analysis. Thus, the new panorama facing science requires the development of new software tools [3], which automatically convert raw data into complete and organised information through metabolic maps.

One of the great advantages of omics data mapping software tools is the interactive identification feature, which allows users to interact with specific elements of a pathway for further exploration or to obtain additional data on specific genes, proteins or even taxonomy, in order to identify and relate gene expression data to taxonomic data. This functionality is essential for researchers who want to understand their experimental data in more detail.

Currently, there are still few bioinformatics resources available for meta-omics analysis, and many of them require significant computational knowledge [23] and a great deal of computing power. On the other hand, web interfaces are easier to use, but often have difficulty dealing with large data files [23].

Thus, the general panorama of omics technologies calls for more development of software for analyzing omics data, which is low in computational cost and easy to use for the user, and which makes it possible to map this data onto metabolic pathways, for a better understanding of biological processes.

1.3 Review of omics data analysis tools used in mapping pathways

With an ever-increasing volume of omics data generated by omics and meta-omics technologies, various tools and web interfaces have been developed to facilitate the interpretation of functional annotation results and to represent the genes or proteins identified, as well as taxonomy, in some cases in metabolic pathways. There are at least 13 tools for mapping metabolic pathways. The following paragraphs provide an overview of these mapping tools.

GenMAPP [22] and MetaCore™ [5] are two metabolic pathway mapping tools that support the analysis of differential gene expression. GenMAPP [22] is a tool implemented in Visual Basic 6.0. A unique feature is the dynamic colouring of genes in pathways based on user-defined criteria. On the other hand, MetaCore™ [5] is a paid web platform that analyses omics data and integrates it for visualisation in the context of metabolic maps, containing more than 400 metabolic pathway maps available for mapping omics data. While MetaCore™ [5] could be useful for those with limited computing resources, GenMAPP [22] has the advantage of not being paid for and being useful for an unlimited number of species.

Pathway Tools [14] is designed to simulate and visualise omics data, implemented in Java-based [21]. Over the course of its versions, Pathway Tools has improved its capabilities and expanded its database from 800 genome/pathway databases (PGDB) [13] to more than 20000 [15]. Learning the tool can be difficult for new users. Pathway Tools [14] allows interactivity in the maps generated, interactive visualisations on the web and on the desktop, zoom support [20]. On the other hand, iPath [27] is a web tool for visualising and analysing pathways that provides user-friendly interactive maps for central metabolism. It is possible to customise maps with differential expression data [1], such as GenMAPP [22] and MetaCore™ [5]. For those looking for a practical tool, iPath [27] and Reactome Knowledgebase [6], [8], can be good choices.

DAVID [4], CellDesigner 3.5 [7] and Pathway Tools [14] can escape the spectrum of tools that map metabolic pathways because, although they can map metabolic pathways, they were designed mainly to perform genetic functional annotation and biochemical network modeling, respectively.

Unlike most of the computational tools described, GenMAPP [22] and PathVisio [19] focus on the immediate needs of biologists, allowing them to quickly interpret omics data in a very intuitive and easy-to-use interface. They are great for those looking for practicality and who want to obtain results more quickly. PathVisio [19] also stands out for its ability to interact with other scientific tools such as Cytoscape and Eu.Gene.

CellDesigner [7] and KGML-ED [17] tools do not represent the taxonomy, which makes them invalid for anyone wishing to relate differential gene expression data to the organism's taxonomy. On the other hand, KEGGCharter [23] allows these two levels of information to be associated, which makes it better from the point of view of correlation and comparison of different knowledge data.

KEGGCharter [23] and KEGG Mapper [11] are designed to work directly with metabolic maps from the KEGG database. KEGGCharter [23] stands out for its command-line interface and its ability to integrate differential gene expression and taxonomic information, visualizing data from multiple organisms in metabolic maps. KEGG Mapper [11] is a KEGG user interface that offers interactive features to explore molecular interactions, visualizing positive or negative regulation through color shading or 3D graphics. KEGGCharter [23] will be great for those who want to relate taxonomic and differential expression data, but KEGG Mapper [11], features interactive graphics.

MetPA [26] and Pathview [18] are tools that combine statistical enrichment analysis with features of metabolic pathways. MetPA [26] presents the networks in a robust web interface, great for those with limited computing resources.

The general panorama of omics data mapping tools calls for the development of more tools that are intuitive, save computing resources, don't lead to a steep learning curve for the end user and also allow for an improved experience through the introduction of interactivity.

(Click here to visit the Table)

Fig. 1. The available analytical tools and web servers designed for mapping pathways data, including their functions, advantages, disadvantages, supported languages, types of identifiers, interactivity, sample representation, taxonomy considerations, differential expression analysis capabilities, and input/output formats.

1.4 KEGG (Kyoto Encyclopedia of Genes and Genomes)

In the same way as the tools presented above, the web interface of the KEGG database can also show the results of omics technologies in the form of metabolic pathways on maps. KEGG's metabolic maps provide a visual representation of metabolic pathways, offering an insight into complex biological systems.

The KEGG pathway database is a very valuable information resource for researchers in the life sciences [17]. It contains metabolic processes and regulators in the form of link diagrams, which can be used for navigation and information retrieval, as well as a basis for modeling and simulation. It thus helps to understand biological processes and the higher-order functions of biological systems.

Currently, KEGG uses semi-static visualizations for the presentation and navigation of its information. Although this style of visualization offers a good presentation and navigation of routes, it does not offer some of the possibilities related to dynamic visualizations.

1.5 KEGGCharter

KEGGCharter [23], is a tool that is available via a command line interface, is easy to use and appears as a tool designed to overlay the differential gene expression and taxonomic information of enzymes on KEGG metabolic maps, using Biopython for this manipulation and visualization, which is particularly useful in metagenomic studies where several taxonomies are present. This information is represented in the form of colored boxes, superimposed on the boxes representing enzyme function. Both differential gene expression and taxonomic assignment are represented as separate entities on the maps, i.e. they are not on the same map. This separation limits the usefulness of directly associating microbial identities with gene expression data in a single image, and is currently the major drawback of KEGGCharter [23]. That said, KEGGCharter's PNG graphs would benefit greatly from being made more interactive, for example:

numerical values could be obtained directly from the graph; multiple taxonomic levels could be included; links and cross-references to online databases could be provided.

1.6 Motivation and objective of the study

KEGGCharter offers many advantages compared to tools with the same objectives. For example, it is fully automated, requiring little computer knowledge on the part of the user. However, there is room for improvement. A limitation of KEGGCharter is that it does not combine gene expression with taxonomy information, which is very relevant, particularly when dealing with complex communities with the aim of knowing which organism is expressing what and to what extent. Therefore, the aim of the project is to update KEGGCharter with functionalities for visualizing results in interactive graphs. The work involved developing new interactive representations on top of the graphs currently produced, to allow more information to be included in the metabolic maps, combining taxonomy and expression levels in the same graph.

2 Methods

The interaction of these new features was done using the Visual Studio Code tool, to work with raw code, Notepad++, to make it easier to work with large html and css structures. The tool's new interactive features were implemented using the Python3 language, and the pages were assembled by implementing html elements in the script's python structure, and then organized using CSS.

2.1 Overview of first version of KEGGCharter and files used for the interactive version of KEGGCharter (Interactive KEGGCharter)

The KEGGCharter [23] workflow (Fig. 2) begins with the interconversion of input IDs: from KEGG IDs or EC numbers to KOs and from KOs to EC numbers. This step is carried out to obtain as much information as possible from the KEGG database. The green file icon represents the input file containing KEEG IDs, KOs or EC numbers, along with quantification and taxonomic information; and the blue file icons represent the two results of the KEGGCharter: TSV files with tables containing the IDs obtained in the interconversion step and KEGG metabolic maps in PNG format, containing taxonomic and quantitative information (stored in the first_time_running_KC directory). For each map, a KGML file is obtained (stored in the resources_directory), which is the XML description of the metabolic map [23], and contains the following information: ids of the EC number boxes (i.e. id="36"), KOs of each box (i.e. "ko:K10944 ko:K10945 ko:K10946"), links to KEGG associated with the KOs of each box (i.e. link="https://www_bget?K10944+K10945+K10946") and the coordinates of each box (x="189" y="140" width="46" height="17").

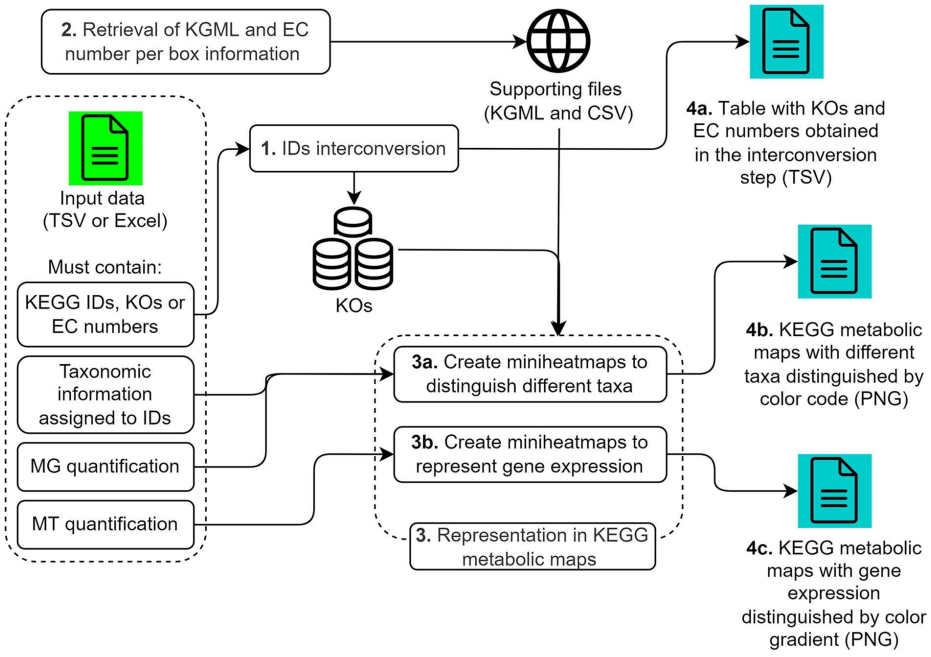

Fig. 2. Schematic representation of the Workflow of KEGGCharter. Diagram taken from the original source [23].

2.2 First step for the developing the Interactive KEGGCharter - Creating a dictionary to store the information for each box on the map effectively

KEGGCharter [23] is made up of two scripts: keggcharter.py, which is responsible for running the tool, and keggpathway_map.py, which is responsible for creating the boxes and maps of differential gene expression and taxonomy. Using KEGGCharter [23], the following files were obtained: box2kos.json, which is a json file that provides information on the k numbers associated with each box; boxes2taxon.json, which is a json file that provides information on the taxonomies associated with each box; and boxes2quant.json, which is a json file that provides information on the colors of each of the four samples in each box.

Through modifications to the tool's code, implemented in Python, new results were generated: the info.json file, which is a tsv file that provides us with information on the quantification of differential gene expression for each of the four samples of each taxonomy, in each box; and taxa2colors.json, which is a json file that gives us the colors associated with each taxonomy. With the information generated by these files: KGML, box2kos.json, boxes2taxon.json, boxes2quant.json, info.json and taxa2colors.json, a dictionary was created in which all the information for each box was stored, thus organizing the information effectively.

2.3 Second step – Turn PNG images interactive

After having the desired results generated by the tool, the next step was to make the PNG images clickable. This is a difficult step because PNG images are static, meaning you can't do anything from them other than visually extract information. To make PNG images clickable, Image maps were defined (Fig. 3), which are extensions that allow different areas of an image to be clickable. In other words, they are a list of coordinates associated with an image that link certain areas of the image to various destinations (for example, to the various html pages). We have created criar_image_maps function, which iterates over the coordinates, stored in the dictionary, and calls the criar_pagina_detalhe function for each ID. This generates interactive areas for the image map and finally saves the HTML.

As the coordinates of each box in the KEGGCharter PNG maps, which are contained in the KGML file, exclude the map legend, the PNG image without the legend was also generated by modifying the tool. This step introduced interactive visualization for the user. Thus, Interactive KEGGCharter now generates, in addition to the outputs generated previously, an HTML file called image_maps_metabolism.html, which is the KEGGCharter metabolic map, without the legend (Fig. 3) and with an interactive user interface, as well as a directory called Image_Maps, which contains the dynamically generated HTML pages for each box on the specific map.

2.4 Create dynamic HTML pages for each map box

Once the images were clickable, dynamic HTML pages were created for each box on the KEGGChater metabolic map, which would open according to the corresponding box. To do this, a system was set up that uses the coordinates of the graphic elements to define interactive areas within the map image. Each area was associated with a detailed HTML page, generated dynamically by the criar_pagina_detalhe function. The result was an interactive user experience, where each box on the map could be clicked to reveal detailed information on a new page, facilitating the exploration and analysis of omics data, combining taxonomy and differential gene expression, all on the same map.

2.5 Enrich the HTML pages with extra information

Finally, the dynamic HTML pages were populated with relevant information: link to KEGG, taxonomy present in the sample, box ID, EC number, K numbers, Map Name, legends and quantification. This information was then distributed coherently throughout the page, so that the user could read it better, by implementing the CSS language in the HTML Style.

3 Results

KEGGCharter provides a view of the results of omics technologies, showing taxonomic and gene expression information of the community. With the introduction

of interactivity in metabolic maps (Fig. 3), this tool is useful for visualizing the genomic potential of microbial communities, and for identifying the microorganisms within communities that can perform this function, all on the same map, improving the user experience.

In Fig. 3), we can see in red, Image Maps, which are the extensions that allowed the different boxes of the metabolic map to be clickable, linking the areas of each box to the various html pages. This was the decisive step in introducing interactive visualization for the user in KEGGChater Interactive.

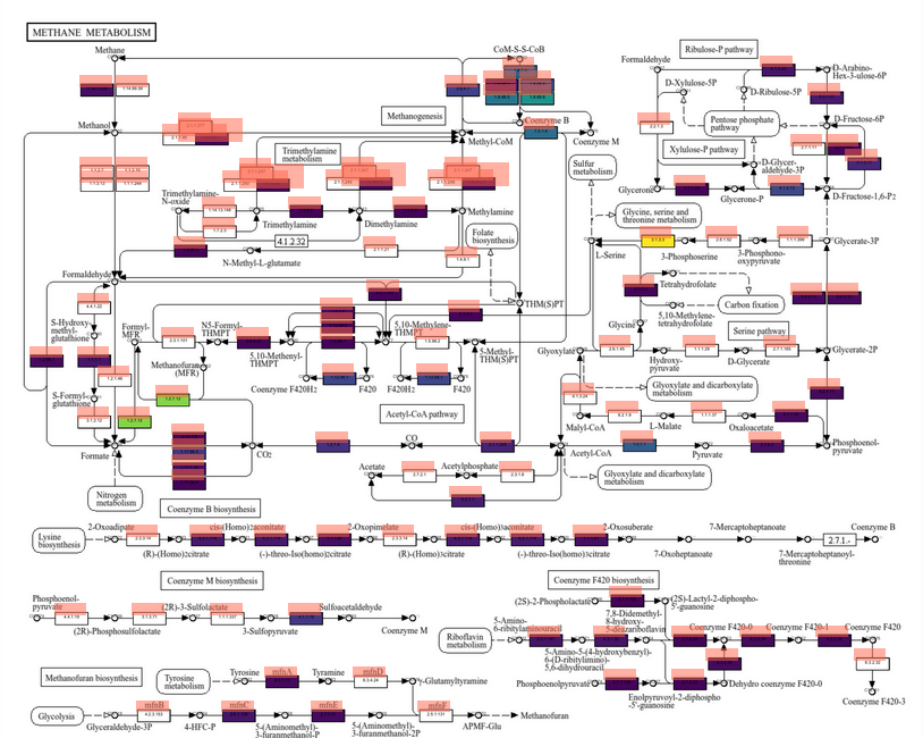

Fig. 3. HTML file called image_maps_metabolism.html, which is the Interactive KEGGCharter metabolic map, without the legend and with an interactive user interface, with an example of a map object, from HTML, placed on map 680 (which represents methane metabolism), to confirm the position of the Image Maps (represented by red areas). The coordinates of the Image Maps have been enlarged to the following scale: 'x'* 2.1; 'y'* 2.1; 'width'* 2.2; 'height'* 2.2, to match the size of the map.

To show the functionalities of the new changes to KEGGCharter, simulated metatranscriptomic datasets representing methanogenic communities were used.

Fig. 4 shows the html page of box 71 of the map obtained with Interactive KEGGCharter, after the modifications introduced to the tool during this study, to generate interactivity in the tool.

In Fig. 4, you can see the color scheme, captioned "Total Expression", which reflects the community's expression of EC number 2.8.4.1. The four horizontal boxes (sample 1, sample 2, sample 3, sample 4) correspond to the four different samples that were analyzed, and the color scheme indicates the greater or lesser expression of the EC number, in this case with samples 1 and 2 being more expressed and samples 3 and 4 being less expressed. Similarly, the vertical boxes, labeled "Expression by Taxonomy", refer to the expression of the microorganisms in the legend, and in this case Methanospirillum was the only microorganism that expressed this enzyme.

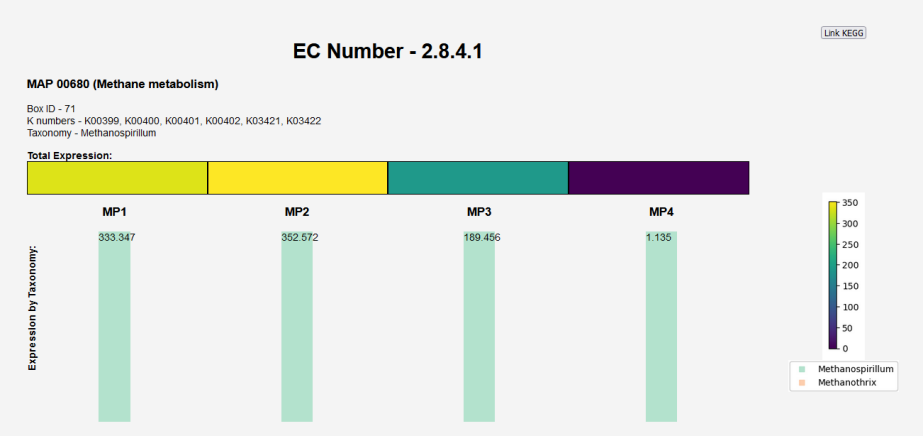

Fig. 4. Example of an html page for the box with ID 71, of the Interactive KEGGCharter output metabolic map obtained from the analysis of the metagenome. This map represents the genomic potential of the community for "Methane metabolism" (map00680 of the KEGG Pathway).

In this html page of the 680 metabolic map, it is possible to view differential gene expression data and taxonomic data, which are related to the methane metabolism pathway. In this omics data set, the microorganisms assigned to *Methanospirillum* and *Methanothrix*, which are colored in blue and pink, with *Methanospirillum* having the greatest expression regarding methane production. On the html pages of each box, up to ten taxonomies can be represented per sample. The page contains the box with the quantification of differential gene expression, supported by the color legend, which ranges from zero to 350. Other microorganisms with less representation in the dataset are identified as "other taxa". It was also possible to express a new type of quantification, associated

with the taxonomy present in each of the four samples (sample 1, sample 2, sample 3, sample 4), which was not previously present in the metabolic graphs in the KEGG database or in the PNG graphs generated by KEGGCharter [23]. The value of this quantification is represented in numerical format for each taxon.

Each of the html pages, such as the one shown in Fig. 4, also contains information about the respective EC number and box ID. It is also possible to access the KEGG database by clicking on the "Link KEGG" button in the right-hand corner of the page. This takes the user to a page in the KEGG database, as shown in Fig. 5, which contains a description of all the K numbers associated with the EC number of the respective box. In this case, the HTML page represented in Fig. 4 contains the K numbers: K00399, K00400, K00401, K00402, K03421, K03422, and upon clicking the "Link KEGG" button, the user is taken to the page that contains the description of each K number. In Fig. 5, we see the table from the KEGG page that describes the K number K00399.The user will then be able to access even more information directly in the KEGG database.

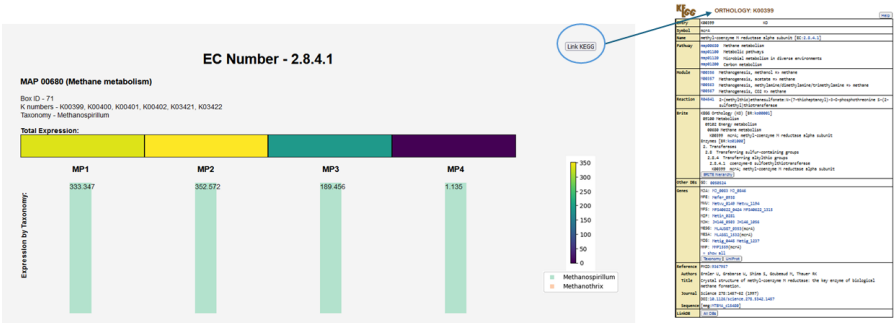

Fig. 5. Html page for EC number 2.8.4.1, which contains information on: Box ID, K numbers, Taxonomy and Total Expression. It is also possible to access the KEGG database by clicking on the "Link KEGG" button in the right-hand corner of the page. On the right-hand side of the picture, you can see a page in the KEGG database, which contains a description of all the K numbers associated with the EC number of the respective box. In this case, we see the table on the KEGG page describing the K number K00399.

In the table describing K number K00399 (Fig. 5), you can see information on Pathway, Module, Reaction, Brite, Other DBs, Genes and Reference (Authors, Title, Journal and Sequence).

With these main results, interactive visualization was introduced into the tool, thus improving the user experience by combining expression data with taxonomic data on the same map.

4 Discussion

This project has contributed to improving the state of the art by enhancing the KEGGCharter [23] tool, a command-line tool for visualizing annotation results on metabolic maps. This tool offers several advantages for end users, for example, it is fully automated as it is run using a single command. In addition, it offers many customization options due to the variety of arguments we can pass to the tool. The results of the metagenomics annotation and gene expression analysis can be visualized using KEGGCharter [23], thus facilitating the interpretation of the metagenomics results, since the expression of these results in metabolic pathways facilitates the understanding of biological processes.

The biggest difficulty with meta-omics studies is interpreting the data generated by omics and meta-omics technologies. To overcome this limitation, it is important that the results of bioinformatics analysis are easily handled, organized and can be easily visualized. Thus, KEGGCharter [23] represents an alternative to existing mapping options, offering the possibility of representing both taxonomic assignment and differential gene expression in metabolic maps for up to 10 samples.

Although the main limitation of KEGGCharter [23] and other mapping tools is the quality of the input data, this project focused on another limitation of most metabolic pathway mapping tools, which was data visualization, that is, the lack of interactivity in the interface of these tools, because in the case of KEGGCharter [23], it was not possible to combine gene expression data with taxonomic data all on the same map, due to the separation of these two entities into separate maps.

With this project, which focused on introducing interactivity to the graphs generated by KEGGChater [23], the tool stood out even more from the rest because it now gives the user the possibility of combining multiple levels of information: the numerical values of taxonomic quantification can be obtained directly from the graph; multiple taxonomic levels have been included; links and cross-references to the KEGG online databases have been made available. In addition, this project has become a very important contribution, as a new type of quantification has been expressed, associated with taxonomy, which until now has not existed in any other tool described. For example, KEGG Mapper [10], which is another tool with similar purposes to KEGGCharter [23], does not automatically represent different taxonomies and differential gene expression results on metabolic maps and requires greater user knowledge. KEGGCharter could become an obvious alternative to KEGG Mapper [10].

Even so, and despite the very significant contribution, which improves the end-user experience, there is room for further improvement in Interactive KEGGCharter. The way in which the HTML pages are generated can cause some problems, especially in terms of being able to slightly increase the tool's execution time in the command line interface, in extreme cases where computing resources are very limited. For possible future contributions, a different approach is recommended, by implementing interactive JavaScript or React elements, through auto spawn, which could slightly decrease the tool's execution time.

JavaScript is a very effective language for creating websites and interactive schemes, but as KEGGCharter [23] has been fully implemented in Python, the integration of the results generated by JavaScript can become the biggest obstacle to this strategy, as the combination of two totally different languages can be complicated.

To make Interactive KEGGCharter more intuitive, I suggest implementing Krona elements in the taxonomy expression, as this will make it possible to organize the various taxonomic levels more effectively. The numerical values of the expression per taxonomy could also be expressed as a percentage of expression within each sample, so it is strongly recommended to update this value in future contributions.

One of the main limitations of the project was its short duration, which prevented all the improvements implemented in the tool, such as those described above, from being perfected. Even so, the main objective of the project was successfully achieved, resulting in a significant contribution to the tool and, overall, to the panorama of tools for mapping metabolic pathways.

5 Conclusion

The project is part of the development of tools for mapping metabolic pathways. In this context, the ability to map omics datasets into metabolic pathways is an indispensable tool for interpreting biological processes. In this work, the KEGGCharter tool was improved with the aim of introducing interactivity into the metabolic graphs generated by the tool, in order to add more information to the metabolic maps, combining taxonomic information with differential gene expression information. With the very significant contribution of this project to improving interactivity in the graphs generated by the tool, users have the unique opportunity, compared to other metabolic pathway mapping tools, to obtain easier conclusions from the metabolic map.

References

1. Darzi, Y., Letunic, I., Bork, P., and Yamada, T.: iPath3.0: interactive pathways explorer v3. Nucleic Acids Research **W510-W513** (2018). https://doi.org/10.1093/nar/gky299.
2. Draghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., and Krawetz, S.A.: Global functional profiling of gene expression. Genomics **81**, 98–104 (2003). https://doi.org/10.1016/S0888-7543(02)00021-6.
3. De Filippo, C., Ramazzotti, M., Fontana, P., Cavalieri, D.: Bioinformatic approaches for functional annotation and pathway inference in metagenomics data. Briefings in Bioinformatics **13**(6), 696–710 (2012). https://doi.org/10.1093/bib/bbs070.
4. 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).
5. Ekins S., Nikolsky Y., Bugrim A., Kirillov E., Nikolskaya T. Pathway mapping tools for analysis of high content data. In: Taylor D. L., Haskins J. R., Giuliano K. A. (eds.) High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery, vol. 356, pp. 319–350. Humana Press, Totowa, NJ (2007).
6. Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Duenas Roca, C., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P.: The Reactome Pathway Knowledgebase. Nucleic Acids Research **46**(D1), D649–D655 (2018). https://doi.org/10.1093/nar/gkx1132.
7. 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.
8. 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.
9. Jin, L., Zuo, X.-Y., Su, W.-Y., Zhao, X.-L., Yuan, M.-Q., Han, L.-Z., Zhao, X., Chen, Y.-D., Rao, S.-Q.: Pathway-based Analysis Tools for Complex Diseases: A Review. Genomics Proteomics Bioinformatics **12**(5), 210–220 (2014).
10. Kanehisa, M., Sato, Y.: KEGG Mapper for inferring cellular functions from protein sequences. Protein Science **29**, 28–35 (2020). https://doi.org/10.1002/pro.3711.
11. Kanehisa, M., Sato, Y., Kawashima, M.: KEGG mapping tools for uncovering hidden features in biological data. Protein Science **31**(1), 47–53 (2022). https://doi.org/10.1002/pro.4172.
12. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., & Tanabe, M.: KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Research **40**(D1), D109–D114 (2012). https://doi.org/10.1093/nar/gkr988.
13. Karp, P.D., Paley, S.M., Krummenacker, M., Latendresse, M., Dale, J.M., Lee, T.J., Kaipa, P., Gilham, F., Spaulding, A., Popescu, L., Altman, T., Paulsen, I., Keseler, I.M., Caspi, R.: Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. Briefings in Bioinformatics **11**(1), 40–79 (2009). https://doi.org/10.1093/bib/bbp043.
14. Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., Spaulding, A., Fulcher,

C., Keseler, I.M., Caspi, R.: Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics **17**(5), 877-890 (2016). https://doi.org/10.1093/bib/bbv079.

15. Karp, P.D., Midford, P.E., Billington, R., Kothari, A., Krummenacker, M., Latendresse, M., Ong, W.K., Subhraveti, P., Caspi, R., Fulcher, C., Keseler, I.M., & Paley, S.M.: Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. Briefings in Bioinformatics **22**(1), 109–126 (2021). https://doi.org/10.1093/bib/bbz104.
16. Khatri, P., Draghici, S., Ostermeier, G.C., Krawetz, S.A.: Profiling Gene Expression Using Onto-Express. Genomics **79**(2), 266–270 (2002). https://doi.org/10.1006/geno.2002.6698.
17. Klukas, C., Schreiber, F.: Dynamic exploration and editing of KEGG pathway diagrams. Bioinformatics **23**(3), 344–350 (2007). https://doi.org/10.1093/bioinformatics/btl611.
18. Luo, W., Brouwer, C.: Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics **29**(14), 1830–1831 (2013). https://doi.org/10.1093/bioinformatics/btt285.
19. van Iersel, M.P., Kelder, T., Pico, A.R., Hanspers, K., Coort, S., Conklin, B.R., & Evelo, C.: Presenting and exploring biological pathways with PathVisio. BMC Bioinformatics **9**, 399 (2008). https://doi.org/10.1186/1471-2105-9-399.
20. Paley, S.M., Karp, P.D.: The Pathway Tools cellular overview diagram and Omics Viewer. Nucleic Acids Research **34**(13), 3771–3778 (2006). https://doi.org/10.1093/nar/gkl334.
21. Rahman, S.A., Advani, P., Schunk, R., Schrader, R., Schomburg, D.: Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). Bioinformatics **21**(7), 1189–1193 (2005). https://doi.org/10.1093/bioinformatics/bti116.
22. 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).
23. 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).
24. Toussaint, P. A., Leiser, F., Thiebes, S., Schlesner, M., Brors, B., Sunyaev, A.: Explainable artificial intelligence for omics data: a systematic mapping study. Briefings in Bioinformatics **25**(1), 1–16 (2024). https://doi.org/10.1093/bib/bbad453.
25. Werner, T.: Bioinformatics applications for pathway analysis of microarray data. In: Current Opinion in Biotechnology **19**, pp. 50–54 (2008). https://doi.org/10.1016/j.copbio.2007.11.005.
26. Xia, J., & Wishart, D.S.: MetPA: a web-based metabolomics tool for pathway analysis and visualization. In: Bioinformatics Applications Note, Vol. 26, No. 18, pp. 2342–2344. Oxford University Press (2010). https://doi.org/10.1093/bioinformatics/btq418.
27. Yamada, T., Letunic, I., Okuda, S., Kanehisa, M., Bork, P.: iPath2.0: interactive pathway explorer. Nucleic Acids Research **39**(suppl_2), W412–W415 (2011). https://doi.org/10.1093/nar/gkr313.