SYSTEMS BIOLOGY APPROACHES IN THE INVESTIGATION OF ARTICULATION POINTS IN KEGG METABOLIC PATHWAYS

**Running title: INVESTIGATION OF ARTICULATION POINTS**

Igor Brandão1, Diego Arthur1,Alice Câmara2, Leonardo Campos1, Clovis Reis1, Rodrigo JS Dalmolin1\*

1 Bioinformatics multidisciplinary environment, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brasil

2 Biophysics and Pharmacology Departament, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brasil

\* Corresponding author

E-mail: rodrigo.dalmolin@imd.ufrn.br

# ABSTRACT

The study of proteins essentiality through laboratory methods is expensive, time-consuming and not scalable for large amounts of proteins. Besides, it is relevant to evaluate the essentiality of several proteins of a metabolic pathway as a whole. The metabolic pathways can be analyzed as graphs, which provide several tools to study the topological features such as the articulation points. Nowadays, research in bioinformatics studies the essentiality of proteins based on betweenness and degree metrics, however, graph theory determines that an essential node in a network is characterized by the articulation point. It remains to be determined whether these articulation points are essential in metabolic pathways and what their topological impact on the network. Using network metrics and articulation points, we look for a reliable way to determine the essentiality of proteins by assessing systematically several metabolic pathways. For this purpose, we determine the articulation points in different networks, evaluate the impact of each articulation point, calculate their occurrence and compare them with occurrences of non-articulation points. We consulted KEGG pathways available as KGML (KEGG XML) files. After parsing it, the data were transformed into a graph object. Two centrality parameters including articulation points and degree are determined and the essential proteins based on these parameters are classified. The most part of the articulation points were located in the proteins groups with the highest occurrences. Approximately 34% of metabolic pathways are related to 17% of all studied species. Also, we observed that the metabolic networks length vary between 2-170 proteins. Proteins classified as Hub articulation points and articulation points represented 15% of the proteins. The findings suggest that articulation points are proteins with the highest frequencies. This work contributes to the systematic study of metabolic pathways using computational approaches.

**Key-words:** Articulation points, KEGG, Metabolic pathways, Biological networks, Systems biology.

# INTRODUCTION

Regulatory and interaction networks are important representations of protein networks and are characterized by having directional and non-directional edges, respectively [2]. The study and application of graph theory in the context of biological networks has gained notoriety since it is necessary to systematically understand how molecules interact with each other and how their functions are determined within the complex cell machinery, alone or together with other cells [1]. Metabolic pathways and cycles are reaction chains where chemical products become the substrate for the next step [3]. Besides, structural and functional analysis of genome-based large-scale metabolic networks is important for understanding the design principles and regulation of metabolism at a system level. The metabolic network is conventionally considered to be highly integrated and very complex. A rational reduction of the metabolic network to its core structure and a deeper understanding of its functional modules are important [4].

The main characteristic of metabolic pathways is that reactions are connected by their intermediates. The products of a reaction are the reactants of subsequent reactions [5]. In general, the metabolic pathways can be classified as flow networks, where a specific variable such as mass or energy flow may be conserved at each node. Metabolic networks have unique properties such as the conservation constraints, which have to be satisfied at each node. Another property is that the metabolic networks are represented with nodes as metabolites and the links are reactions that are catalyzed by specific gene products [6].

In terms of graphs, there are multiple ways to build a network from a metabolic model [7]. Network analysis suggests that biological networks have two important structural properties. First, most of these networks, including metabolic networks are scale-free and possess a ‘‘small-world’’ property, that is characterized by a short average path length [1, 4, 6]. Second, scale-free networks are suggested to have high error tolerance (against random failure) and low attack tolerance (vulnerability to the failure of the highly connected nodes) [6]. The topological features of networks can be measured by observing specific characteristics from each metabolic pathway such as betweenness, degree, articulation points (AP), and bridges. Betweenness centrality measures the total number of non-redundant shortest paths going through a certain node or edge [8]. The degree tells how many links the node has to other nodes [1]. A node is an AP or a cut-point if its removal disconnects the network or increases the number of connected components of the network [9, 10]. An edge is a bridge if its removal disconnects the graph, if it is connected, or increase the number of connected components otherwise [11].

A great source of curated functional information is the KEGG database (Kyoto Encyclopedia of Genes and Genomes). It holds a knowledge base on metabolic pathway maps of molecular interaction and orthology relationships between genes/gene products [12]. In KEGG, nodes present enzymes and edges represent the reactions that transform one metabolite into another [13-15]. Although KEGG diagrams are informative and easy to understand, a network topological study about the metabolic pathway can be important to understand specific network characteristics. In this sense, the representation via graphs could stands out because it enables the study of several characteristics such as clusters, articulations points, bridges, an arrangement of proteins in the network, number of protein connections, among other factors. There are many ways to provide visualizations of metabolic pathways, since a reaction may have more than one reactant and/or more than one product, the pathway can be represented by directed graphs. Nodes represent metabolites and edges represent reactions. A directed edge from a compound to a reaction node denotes a reactant while an edge from a reaction to a compound node denotes a product of the reaction [5].

To date, no research systematically evaluated the topology of all metabolic maps as a whole to study the points of articulation. The study of APs is relevant because by identifying these points it is possible to detect the most vulnerable sites in a metabolic pathway. Besides, APs can impact the metabolic pathway differently. The impact is defined as the number of vertices that get disconnected from the main (largest) surviving connected component after the removal of the AP [19].

This work aims to identify the AP in KEGG metabolic pathways and calculate the frequency for each protein in each KEGG metabolic pathways, and to evaluate the impact of each AP on the studied metabolic pathways.

# MATERIALS AND METHODS

## Extraction and processing of metabolic pathway data

Initially, the list of all species available in KEGG was generated along with the list of metabolic pathways for each species. These lists were the reference for the selection of metabolic pathways used in the study. KGML (KEGG XML) file is processed and transformed into a graph object whose nodes represent the combination of EC numbers (enzyme commission number) with the related chemical reaction code. We used KEGGREST R package [20] to load the reference pathway of interest from the REST API provided by KEGG. We used 177 metabolic pathways of all species available on KEGG (totalizing +600k datasets). KEGG data was parsed using a function adapted from the KEGGREST package. After that, compounds and interconnections between metabolic pathways were removed. A way to extract the APs based on depth search algorithms for graphs has been applied. Proteins frequency counting was performed by comparing and analyzing the presence of them in the species-specific metabolic pathways. We used the iGraph R package [21] to extract several features from the graph like communities, betweenness centrality, closeness, and clustering. Proteins were classified according to their metabolic pathway profile into 4 groups: HAP (hub articulation point), HP (hub point), AP and NHNAP (non hub non articulation point). The protein is considered a hub if its degree is within the top 20% of the metabolic pathway.

## Detection of AP

To calculate the AP of metabolic networks, we rely on information described in the tutorial "Articulation Points (or Cut Vertices) in a Graph" [22]. Briefly, the graph was traversed through the depth first search (DFS) algorithm. One of two conditions must be met for the node to be an AP. First, a node must be the root of the DFS and have at least two descendants. Second, a node is not the root and has a descendant whose vertices do not connect with the evaluated node's ancestor.

## Calculation of AP impacts

The APs impact was calculated based on the nodes' number disconnected from the core component of the network after the AP remotion. Besides, to solve the problem regarding the size differences of the different studied networks, an impact normalization was applied. The calculated impact was normalized on a scale from 0 to 1, where 1 represents the highest impact calculated for a given network, 0 represents the lowest impact, and the remaining impacts were allocated to the intermediate values ​​of this scale.

## Statistical analyses

For the statistical analysis, all the datasets processed in the R language were merged. Adherence test (hypergeometric) was performed to verify the distribution and randomness of the APs found in the networks. For this purpose, it was necessary to sort the protein frequency in descending order because it is expected that AP are located in the group of proteins with higher occurrences. Moreover, we normalized the protein frequency by the total species that contain a certain metabolic pathway, since the analyzed pathways shown different numbers of organisms. Then, the highest normalized value was considered 1, the lowest value was considered 0 and others values were within this scale. Finally, proteins with a zero occurrence were removed to avoid biasing the results. P-value < 0.05 was considered significant and the Bonferroni correction method was used to adjust the p-value.

## Metabolic pathways visualization

We provided a graphical visualization of the pathways displaying the main characteristics from its nodes such as: AP status (border in blue), betweenness centrality (background color), frequency (node size), and protein classification (label) using visNetwork R package [23].

## Articulation points curation (in progress)

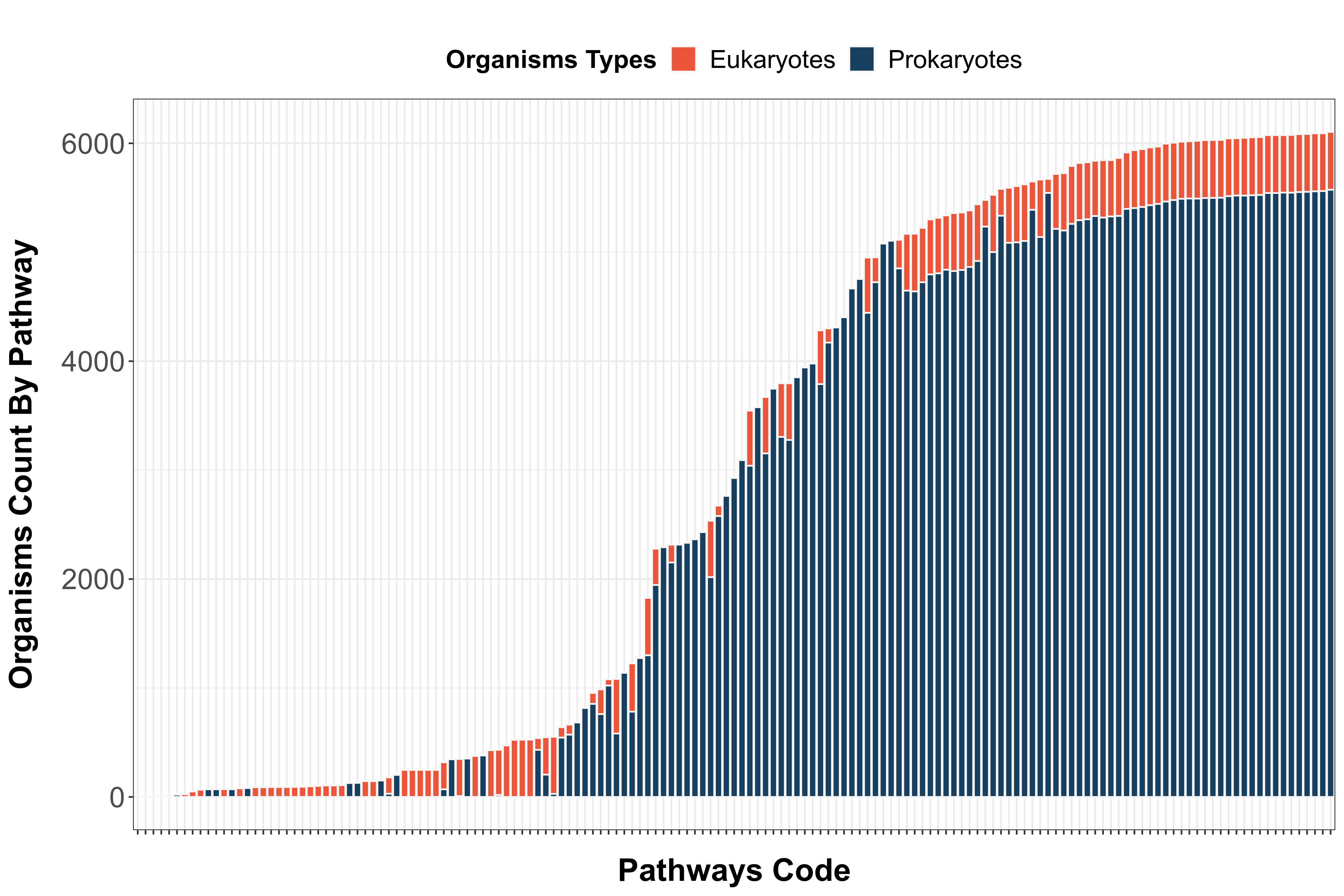
In order to avoid the laborious manual curation for network construction, some methods are developed to automatically reconstruct networks by retrieving interactions or sub-networks from existing maps and models [24, 25]. Combining automatic reconstruction with domain knowledge, manual search of literature and databases would provide a reasonable strategy to construct detailed and fully annotated large-scale biochemical networks [26].

However, establishing these correspondences between different databases (e.g., KEGG and BioCyc) is a non-trivial problem because of the non-standard terminology used in the scientific literature for these three entities (e.g., large numbers of synonyms are used for a given chemical compound) [27].

# RESULTS

**Number of species in metabolic pathways**

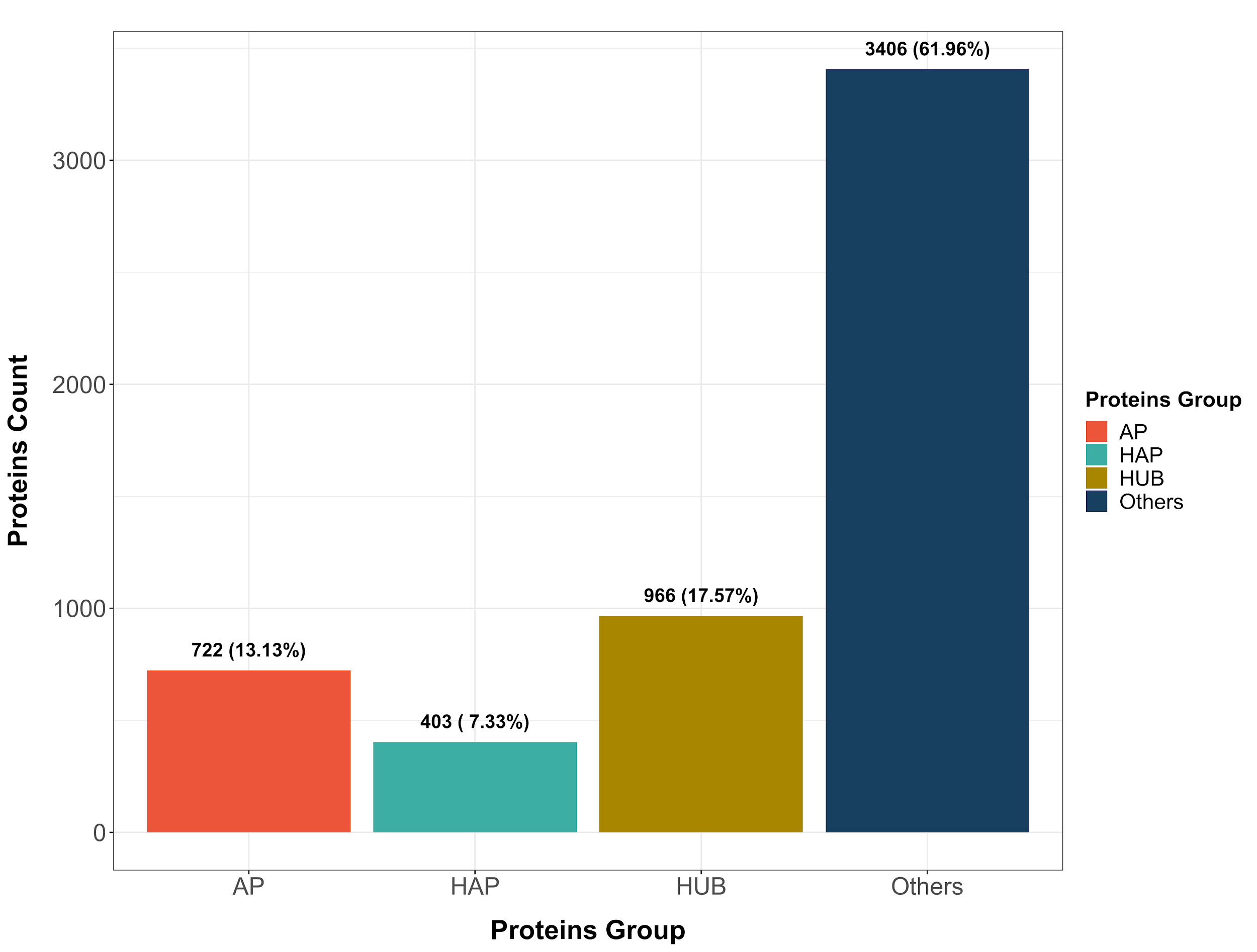
Initially, a script was developed in R language using the KEGGREST package in order to load the list of all organisms with their respective taxonomic information. Through the taxonomy, the organisms were separated according to the kingdoms. Approximately 32% of metabolic pathways are related to less than one thousand species (Figure 2). Additionally, approximately 31% of metabolic pathways are related to at least 5,000 species.



**Figure 2:** Distribution of species quantity by metabolic pathways available in KEGG. X-axis: metabolic pathway. Y-axis: organisms count. The orange bars represent pathways related to eukaryotes. Eukaryotes represent 8.6% (534) of total species (6221). Blue bars represent pathways related to prokaryotes. Prokaryotes represent 91.4% (5687) of total species.

## Nodes classification

Each node is represented by the following entities: enzyme commission number (EC), reaction code, KEGG (x, y) coordinates. The nodes studied (n=5497) were divided into 4 groups (Figure 3) according to their degree and whether or not it is an AP: HAP (hub articulation point), HP (hub point), NHNAP (non hub non articulation point). AP (n=722) represented 13.13% of the nodes studied. Regarding HAP (n=403), a frequency of 7.33% of the analyzed nodes was observed. HUB (n=966) corresponded to approximately 17.57% of total nodes since few connectors are usually present in a metabolic network. Other nodes (n=3406) corresponded to 61.96% of the total of studied nodes.



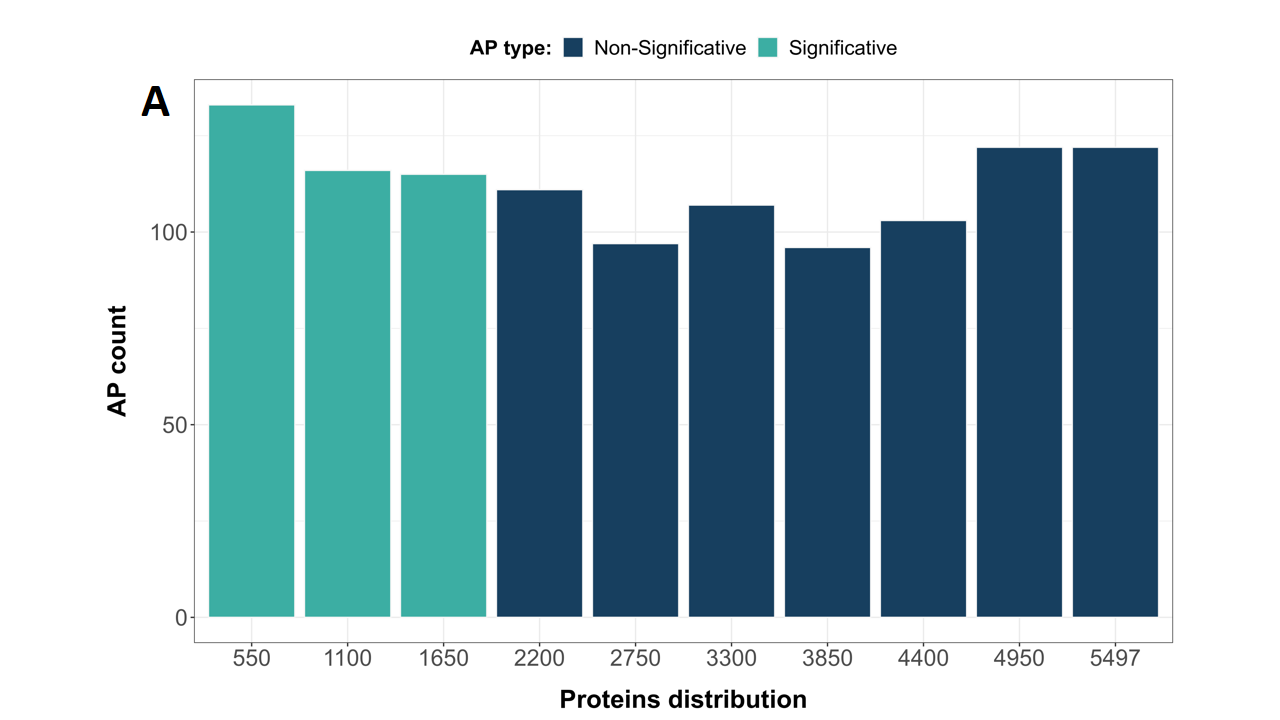
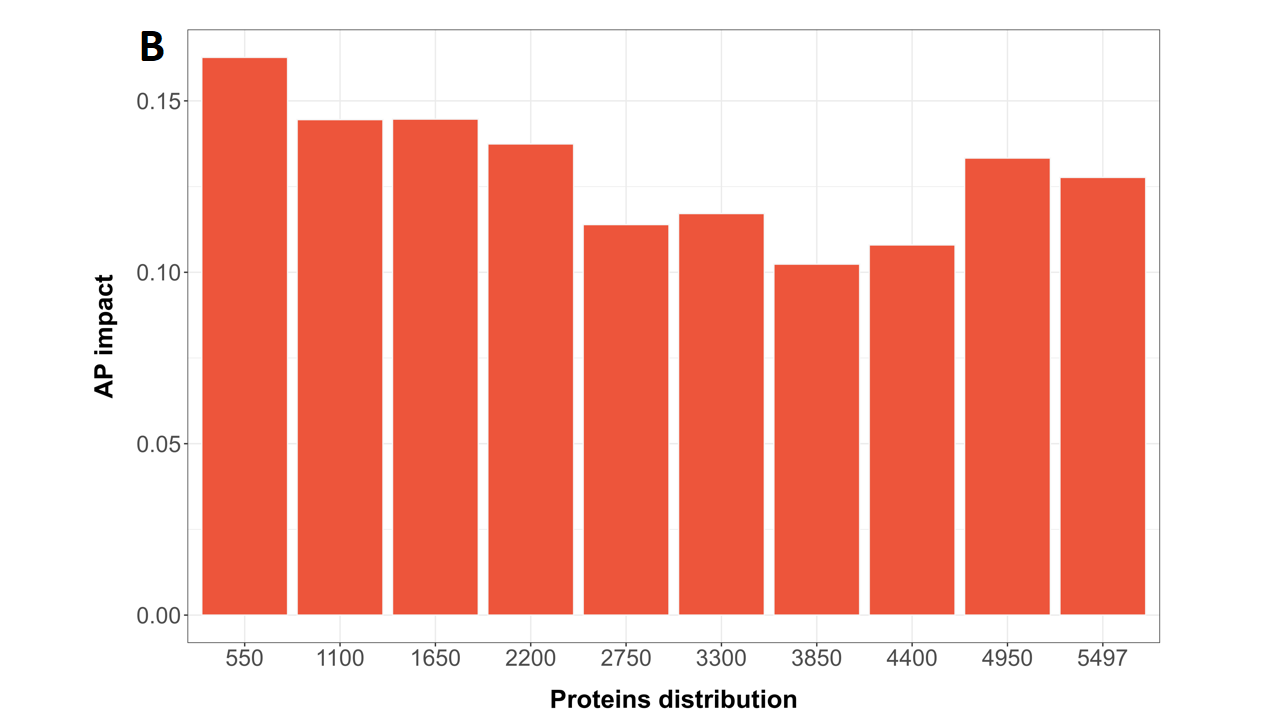
**Figure 3:** Distribution of proteins classifications. Total of studied nodes was 5497.

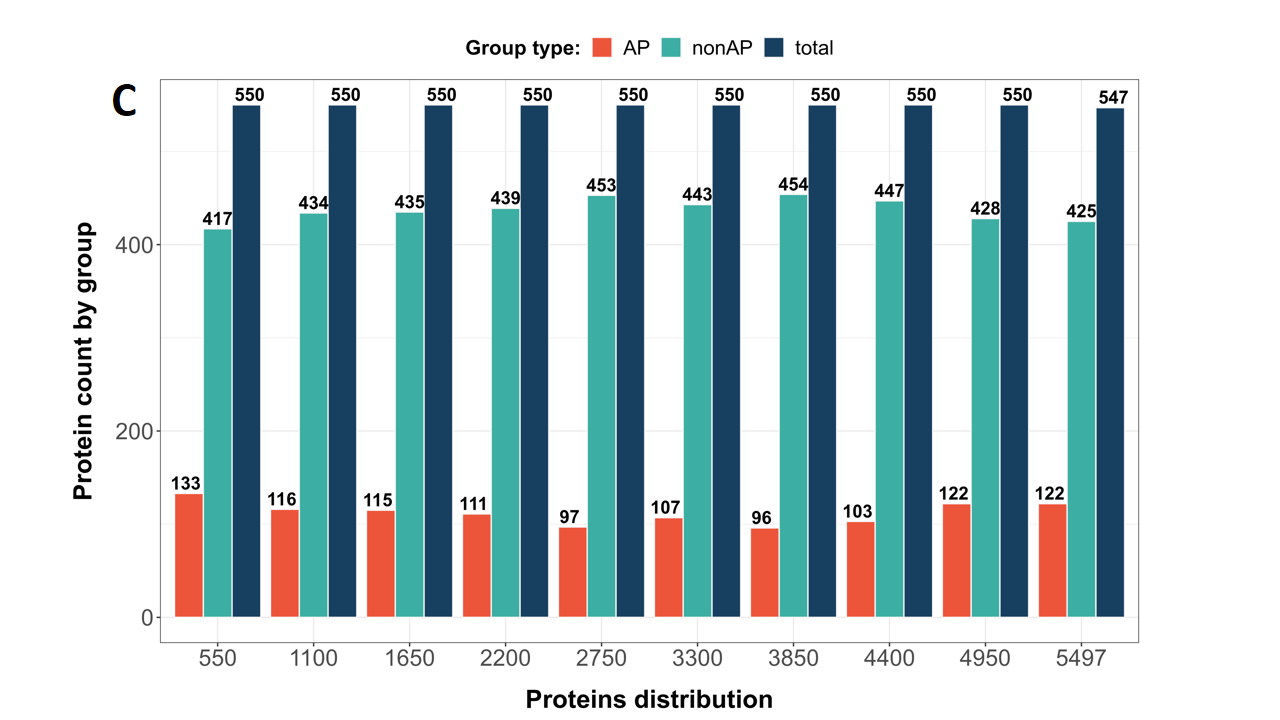
HAP: hub articulation point; HP: hub point; NHNAP: non hub non articulation point.

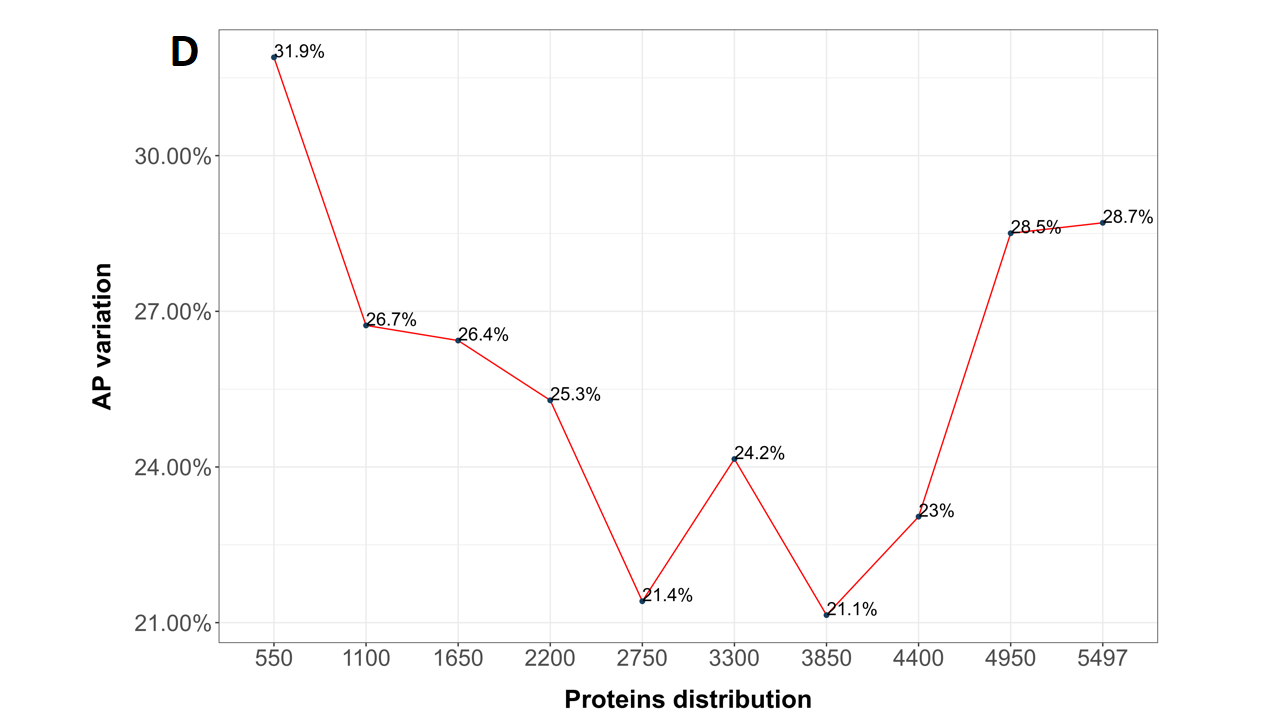
Additionally, we observed that proteins quantity varies from 2 (pathway 00363) to 170 (pathway 00061) in each metabolic pathway.

## AP distribution

Proteins were ordered in decreasing order of normalized frequency (Table1) and divided into 10 ranges, each representing 10% of the proteins. Figure 5A shows que os pontos de articulação são mais essenciais nas três primeiras faixas de grupos de proteínas (p<0.05), pois de uma forma geral apresentaram maiores ocorrências nas rotas metabólicas se comparado aos pontos de articulação das demais faixas. Esses normalmente são os pontos de articulação mais periféricos e não necessariamente estão presentes na maior parte dos organismos que possuem uma determinada rota metabólica. Além disso, APs are more concentrated in the most frequent protein ranges. The range 1 presents the most APs number (24%) and the range 7 presents the lowest APs number (17%). Figure 5B showed that the most frequent AP range has a greater impact on the network when compared to the other protein ranges. In other words, os APS das faixas com maiores frequências tendem a desconectar maior quantidade de vértices na rede. The most impacted range was the first, with an impact of 0.16 (valor normalizado) and the lowest impacted range was the seventh, with an impact of 0.1(valor normalizado). E importante destacar que quanto maior o score de impacto, maior o numero de nos desconectados da rede apos a remocao de um AP. Figure 5C shows the absolute count of AP and non-AP proteins per range. Figure 5D shows the APs proportion in the protein ranges. The group with the highest AP proportion was the first, with 31.9% of APs. The seventh range has the lowest APs proportion, with only 21.1% of APs.

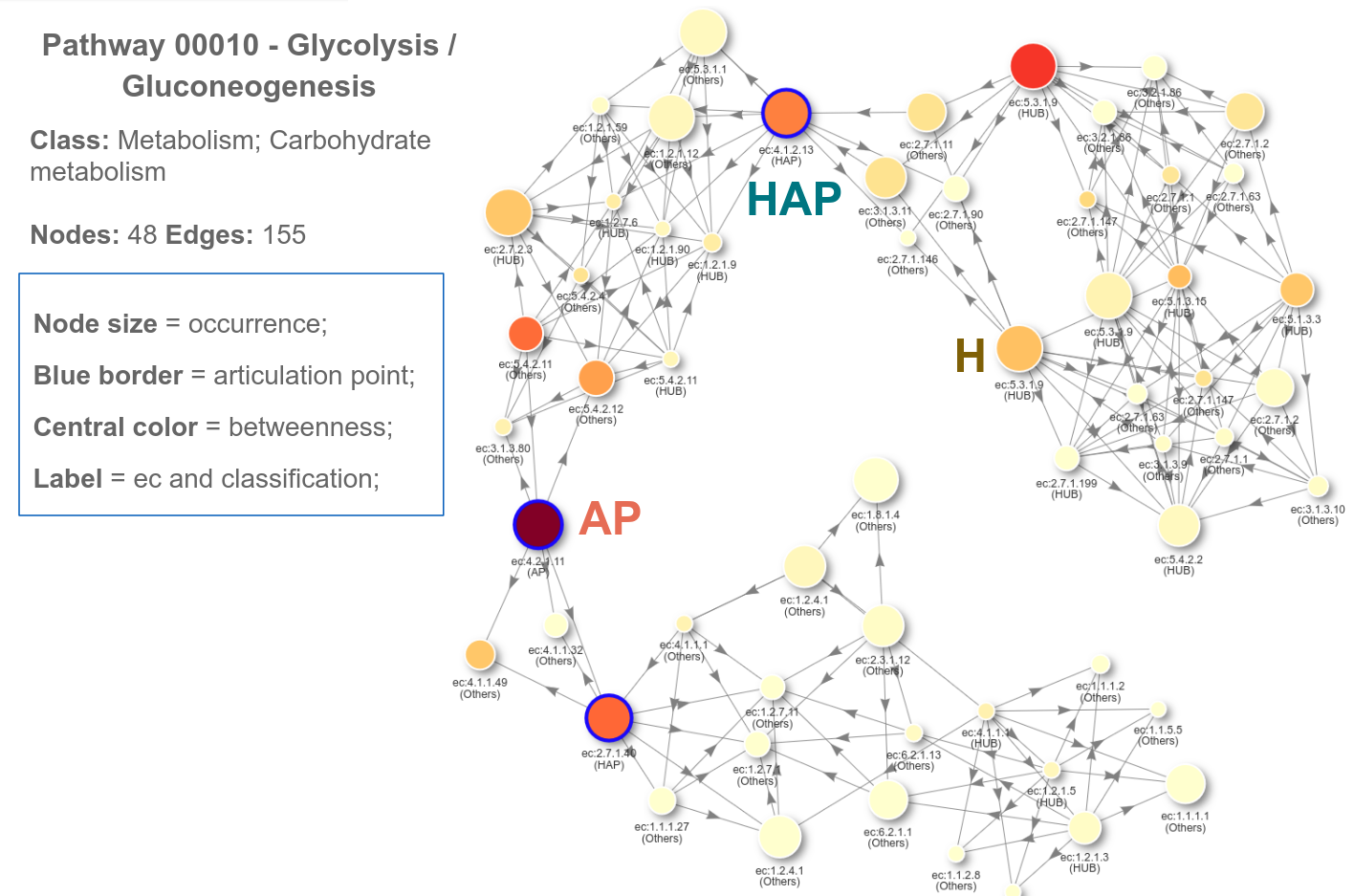






**Figure 5: A)** AP count per protein range. Green bands represent faixas com maiores concentracoes de AP por proteina (p-value <0.05). **B)** Impacto medio causado pela remocao de cada AP em cada faixa de proteina. **C)** Count of AP and non-AP by protein range **D)** Percentage of APs per protein range. Proteins were ordered in decreasing order of normalized frequency and X-axis represents a set of 10% of the proteins evaluated.

## Network visualization

The network elaborated in our study (Figure 6) contemplates protein classifications according to their topological positioning, facilitating the visual identification of the most important proteins of a network (points of articulation). Additionally, our proposal provide dynamic HTML visualizations available at https://igorabrandao.com.br/kegg-pathway-bottleneck, which allow the observation of several metabolic pathways. It is possible to adjust the network size and position, as well as to select proteins according to their classification, making it possible to highlight protein groups of interest. Finally, the AP remotion can disconnect the network and the proteins metrics (impact, degree, and betweenness) can be easily viewed by placing the mouse over the node.

**Figure 6:** Visualization of pathway 0010 (glycolysis and gluconeogenesis). Nodes represent metabolites and edges represent reactions. Node size represents protein frequency, blue borders indicate APs, color in the center of node indicates betweenness level. The classification and EC code of the protein are present on the node label. The network was made using R programming language (visNetwork package).

# DISCUSSION

The present study aimed to investigate the APs for all metabolic pathways available on the KEGG database. To achieve this aim, a pipeline based on R programming language was developed and applied to KEGG XML files for each species (instance) related to each metabolic pathway. We observed that approximately 32% of metabolic pathways are related to less than ⅙ of KEGG species. This might be due to the existence of specific pathways for certain plants, bacteria, and fungi, which are involved with their own cellular functions in these organisms. An example would be Benzoxazinoid biosynthesis route 00402 (Supporting material S2a) responsible for the defense of some plant species (Supporting material S2b). Also, approximately 31% of metabolic pathways are related to at least ⅚ of KEGG species, this may be due these pathways being essential in different types of living beings (Figure 2), such as glycolysis and gluconeogenesis route 00010, present in 6084 species (Supporting material S1).

Besides, we studied the profile of each protein within the metabolic pathways. HAP represented 7% of the proteins studied (Figure 3). This may be due to the topology of metabolic pathways, which are made up of several protein communities that connect to each other through a few HAPs. The remotion of one of these HAPs will cause the network rupture since it is a central point that connects various network complexes and/or peripheral regions. Poranto podemos sugerir que os HAPs sao proteinas essenciais. HUB are approximately 17% of the analyzed proteins. These proteins connect several other proteins of the same community, presenting a high degree. Unlike HAPs, HUBs are not articulations points because a protein cluster has more than one way to connect. As there are usually several protein communities within a network, it is likely that HUBs will be present in larger amounts when compared to HAPs. However, it is important to highlight that the absence of this type of protein does not disconnect the community, but may affect the network in some way.

APs correspond to approximately 13% of total proteins studied. Although they are important points for the network, normally APs are not as centralized as HAPs. In addition, they do not connect with many neighboring proteins (present low degree). Other proteins correspond to 61% of the proteins because, in terms of network topology, they do not play a crucial role and are generally located within the communities and peripheries of the network.

It is possible to evaluate the metabolic pathway profile through the protein types distribution. Is noteworthy that some metabolic pathways were composed mostly of HAPs, such as Benzoxazinoid biosynthesis pathway 00402 (Supporting Information S2), which is composed of only eight proteins, of which six are HAPs. Besides, it is formed by a one-way bridge in terms of mass flow. Others pathways had no HAPs, such as caffeine metabolism pathway 00232 (Supporting Information S3). In this case, there are few communities, which are overcrowded; Some communities are connected to each other through bridges that represent one-way paths. These bridges are composed of APs, which connect with only one protein. Besides, most metabolic pathways are composed of at least three protein classifications. Finally, we highlight the 00061 pathway (Supporting information S4), related to fatty acid synthesis. This pathway is made up of an overcrowded central community, 7 peripheries highly connected to the center and a periphery connected to the center through just one bridge. In addition, this pathway has 3 components disconnected from the main network component, with their own HAPs and APs. This pathway has considerable amount of APs and HAPs. However, although there are several highly connected nodes in the network central area, none of them an AP or HAP due to the connection handles, which guarantee redundancy in this area of the network.

The majority of the proteins classified as APs and HAPs showed higher frequencies than the average (more abundant). It suggests that this set of proteins is essential for its metabolic pathways since they are present in most of the species containing a particular metabolic pathway (Figure 5A). In the hypergeometric analysis, approximately 30% of the most frequent proteins (the first three ranges) were statistically significant, probably due to the higher concentrations of APs when compared with the other ranges (Figure 5B). It is important to highlight that this fact agrees with our hypothesis because we expect that APs to be high-frequency proteins. In the first 550 proteins, for example, we found approximately 133 APs, suggesting that they were strongly present in most species on different metabolic pathways (Figure 5C). The other 70% of the studied protein had no APs with statistical significance. This is probably due to the lower APs concentrations in the protein’s ranges and due to these APs profile, which usually have frequencies below the average. The low frequency may be related to the fact that these APs are connectors in peripheral regions of the metabolic pathways, being responsible only for connecting a few isolated proteins. Besides, we observed that HAPs and APs had the highest impact and frequency rates on the networks (Figure 5B).

Our work considers an essential protein as an AP. One similar study demonstrated that an important protein in a network can be calculated based on the betweenness centrality metric plus its degrees [8]. Other studies suggest performing the identification and removal of APs to provide a new perspective on the organizational principles of complex networks [28]. Finally, to simplify the understanding of the metabolic pathways’ topology, our work proposes the visualization of these pathways as dynamic networks (figure 6). The network visualization is an important aspect of this work since these visualizations provide the possibility to explore network topological features graphically. The choice of a network visualization has *pros* and *cons,* e.g. although the bipartite representation of a metabolic pathway provides richer details, it turns more difficult the network understanding [7].

One study related to inflammatory bowel disease (IBD) used protein-protein interaction network analysis. This study found that there are seven hub-bottleneck proteins in the IBD network responsible to maintain the network integrity [29]. Network-based biomarkers can be associated with animal behavior models, in which a change in behavior resulting from a manipulation of a network biomarker would constitute a strong validation of drivers of symptoms [30]. The network-based biomarker consists of two protein association networks constructed for disease samples and non-disease samples [31].

In summary, we suggested that APs have the potential to cause a great impact on biological networks. New studies will be required to establish more relationships between these key proteins (APs) and topological attributes of the metabolic pathways. Additional studies with metabolomics, which is a powerful technology that allows for the assessment of global metabolic profiles, can be used to distinguish between diseased and non-diseased status information [32]. Another future possibility is the usage of machine learning (ML) techniques to create a predictive model to identify potential essential proteins in various metabolic pathways. Finally, knockout studies can be applied to experimentally evaluate whether specific APs are essential for a given metabolic pathway.

# CONCLUSION

Processing more than 600k datasets of KEGG metabolic pathways, we found that HAPs and APs represented 15% of the proteins. Besides, the findings indicate that most of these APs were placed in a group of proteins with the highest frequencies. Approximately 34% of metabolic pathways are related to less than one thousand species since they are related to organisms with low quantity in KEGG database. Almost 32% of metabolic pathways are related to at least 5,000 species, probably these pathways are present into prokaryote due to its variety in KEGG.

The metabolic pathway assessment can be approached from the top to the bottom, starting from the network’s topological properties and graphical disposal and moving to the proteins' specific functions, and how they interact within the set as a whole. This work contributes to the study of metabolic pathways using computational approaches since nowadays few works are exploring massive data related to curated databases of metabolic pathways and generating analysis to help to understand the big picture of metabolic pathways.

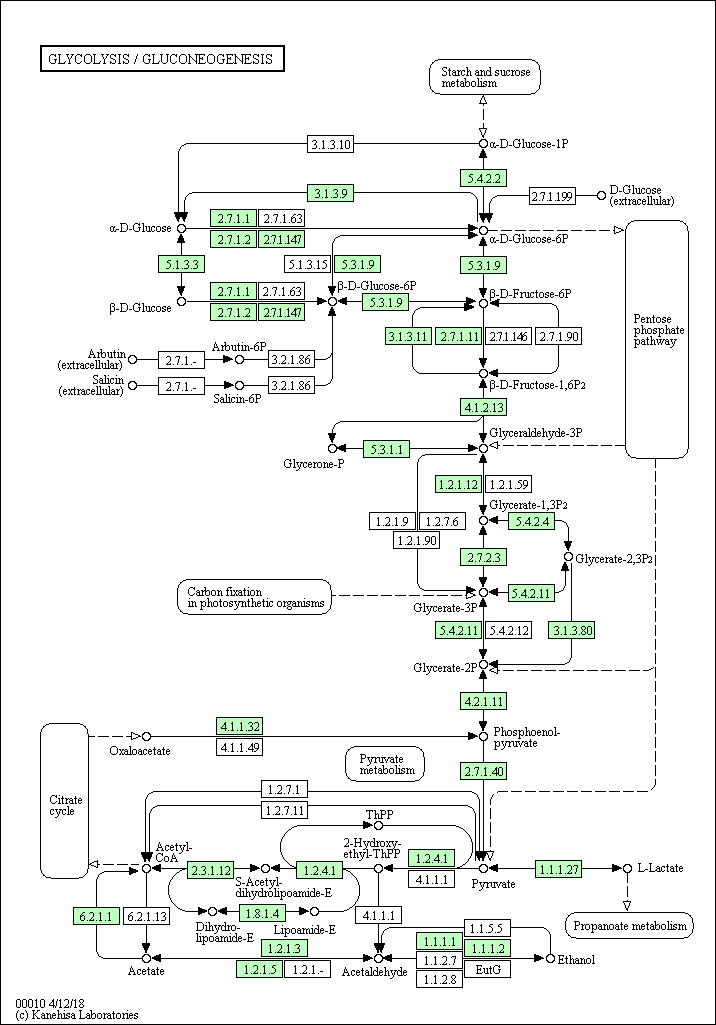
# REFERENCES

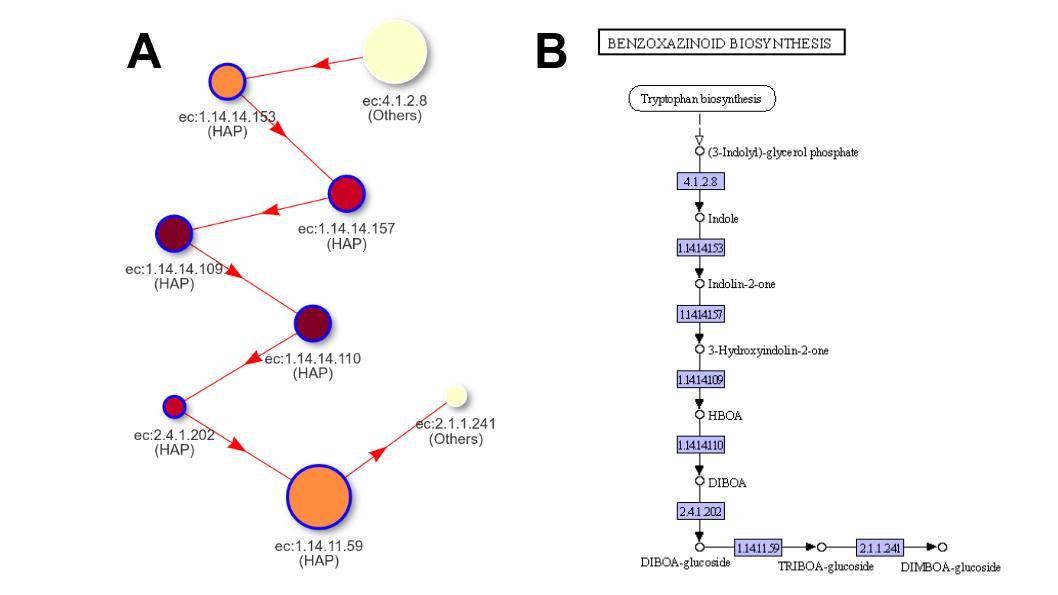
1. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews genetics. 2004; 5(2): 101.
2. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001; 411(6833): 41.
3. Larsen TS. The Scientist's Guide to Cardiac Metabolism: Historical Perspectives (cap.15). 1st ed. Academic Press. Germany: Department of Cardiothoracic Surgery Friedrich-Schiller-University of Jena Jena; 2016: 207-217.
4. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics. 2003; 19(11): 1423-1430.
5. Maniadi EM, Tollis IG. Analysis and visualization of metabolic pathways and networks: A hypergraph approach. 1st ed. Spain: Valencia; 2014: 109-112.
6. Mahadevan R, Palsson BO. Properties of metabolic networks: structure versus function. Biophysical journal. 2005; 88(1): 7-9.
7. Beguerisse-Díaz M, Bosque G, Oyarzún D, Picó J, Barahona M. Flux-dependent graphs for metabolic networks. NPJ systems biology and applications. 2018; 4(1): 32.
8. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS computational biology. 2007; 3(4): e59.
9. Behzad M, Chartrand G. Introduction to the Theory of Graphs. 6st ed. Allyn and Bacon. EUA: Boston; 1972.
10. Harary F. Graph theory. 6st ed. Addison-Wesley. England: London; 1996.
11. Ausiello G, Firmani D, Laura AL. Real‐time monitoring of undirected networks: Articulation points, bridges, and connected and biconnected components. Networks. 2012; 59(3): 275-288.
12. Kanehisa M, Furumichi M., Tanabe M., Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research. 2016; 45(1): D353-D361.
13. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. Nature. 2000; 407(6804): 651.
14. Wagner A, Fell DA. The small world inside large metabolic networks. Proceedings of the Royal Society of London. 2001; 268(1478): 1803-1810.
15. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics. 2003; 19(11): 1423-1430.
16. Alves R, Chaleil RA, Sternberg MJ. Evolution of enzymes in metabolism: a network perspective. Journal of molecular biology. 2002; 320(4): 751-770.
17. Wu X, Qi X. Genes encoding hub and bottleneck enzymes of the Arabidopsis metabolic network preferentially retain homeologs through whole genome duplication. BMC Evolutionary biology. 2010; 10(1): 145.
18. Herrgård MJ, Fong SS, Palsson BØ. Identification of genome-scale metabolic network models using experimentally measured flux profiles. PLoS computational biology. 2006; 2(7): e72.
19. Farina G. A linear time algorithm to compute the impact of all the articulation points. 2015; 1: 1-4.
20. Tenenbaum D. KEGGREST: Client-side REST access to KEGG. R package version. 2016; 1(1).
21. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006; 1695(5): 1-9.
22. Geeks for Geeks. Articulation Points (or Cut Vertices) in a Graph. 2013 May 21 [cited 2019 Dec 10]. Available from: https://www.geeksforgeeks.org/articulation-points-or-cut-vertices-in-a-graph/.
23. Almende BV, Thieurmel B Robert T. visNetwork: Network Visualization using ‘vis. js’ Library, R package version. 2016; 1(1).
24. Ritz A, Poirel CL, Tegge AN, Sharp N, Simmons K, Powell A. et al.. M.. Pathways on demand: automated reconstruction of human signaling networks. NPJ systems biology and applications. 2016; 2: 16002.
25. Supper J, Spangenberg L, Planatscher H, Dräger A, Schröder A, Zell A. BowTieBuilder: modeling signal transduction pathways. BMC Syst Biol. 2009; 3(1): 67.
26. Khan FM, Gupta SK, Wolkenhauer O. Integrative workflows for network analysis. Essays in Biochemistry. 2018; 62(4): 549–61.
27. Altman T, Travers M, Kothari A, Caspi R, Karp PD. A systematic comparison of the MetaCyc and KEGG pathway databases. BMC Bioinformatics. 2013; 14(1): 112.
28. Tian L, Bashan A, Shi D-N, Liu Y-Y. Articulation points in complex networks. Nat Commun. 2017; 8(1): 14223.
29. Asadzadeh-Aghdaee H, Shahrokh S, Norouzinia M, Hosseini M, Keramatinia A, Naghibzadeh B, et al. Introduction of inflammatory bowel disease biomarkers panel using protein-protein interaction (PPI) network analysis. 2016; 9(1): S8-S13.
30. Neylan TC, Schadt EE, Yehuda R. Biomarkers for combat-related PTSD: focus on molecular networks from high-dimensional data. European Journal of Psychotraumatology. 2014; 5(1): 23938.
31. Wang X, Gulbahce N, Yu H. Network-based methods for human disease gene prediction. Briefings in Functional Genomics. 2011; 10(5): 280–93.
32. Wang X, Zhang A, Han Y, Wang P, Sun H, Song G, et al. Urine Metabolomics Analysis for Biomarker Discovery and Detection of Jaundice Syndrome in Patients With Liver Disease. Mol Cell Proteomics. 2012; 11(8): 370–80.

# SUPPORTING INFORMATION

**Supporting information S1:** Visualization of pathway 0010 (glycolysis and gluconeogenesis) in KEGG database. KEGG presents information about the metabolic pathway as a sequential block diagram. In which colored or white boxes represent proteins. Colored boxes are proteins present in the metabolic pathway, while whites are absent proteins in these pathways.

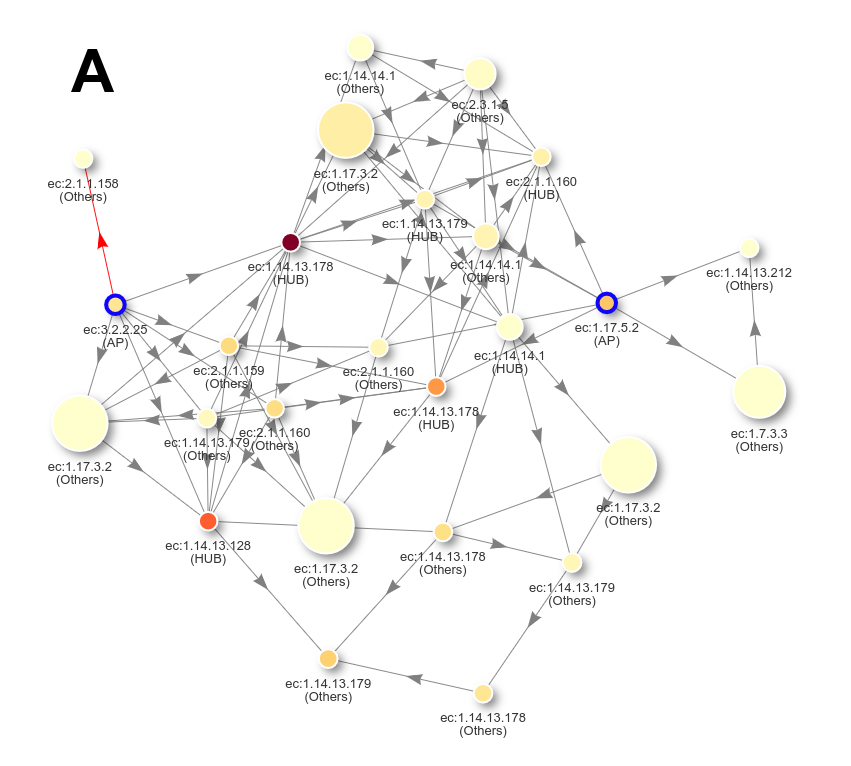
Source: https://www.kegg.jp/kegg-bin/show\_pathway?org\_name=hsa& mapno=00010.

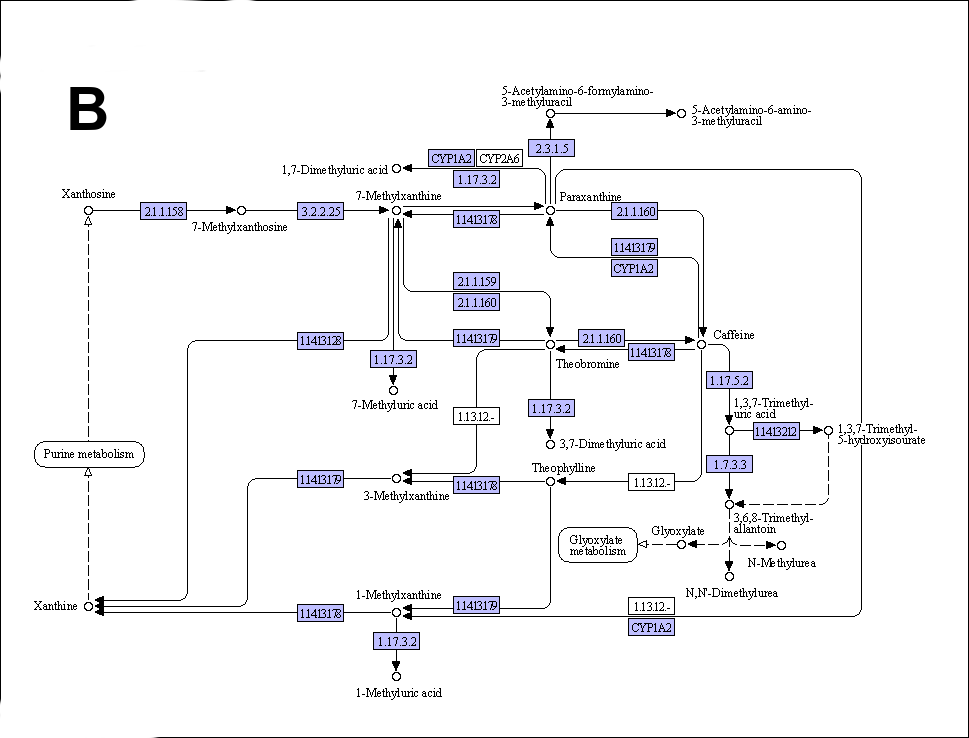


**Supporting information S2:** Pathway 00402 - Benzoxazinoid biosynthesis\*

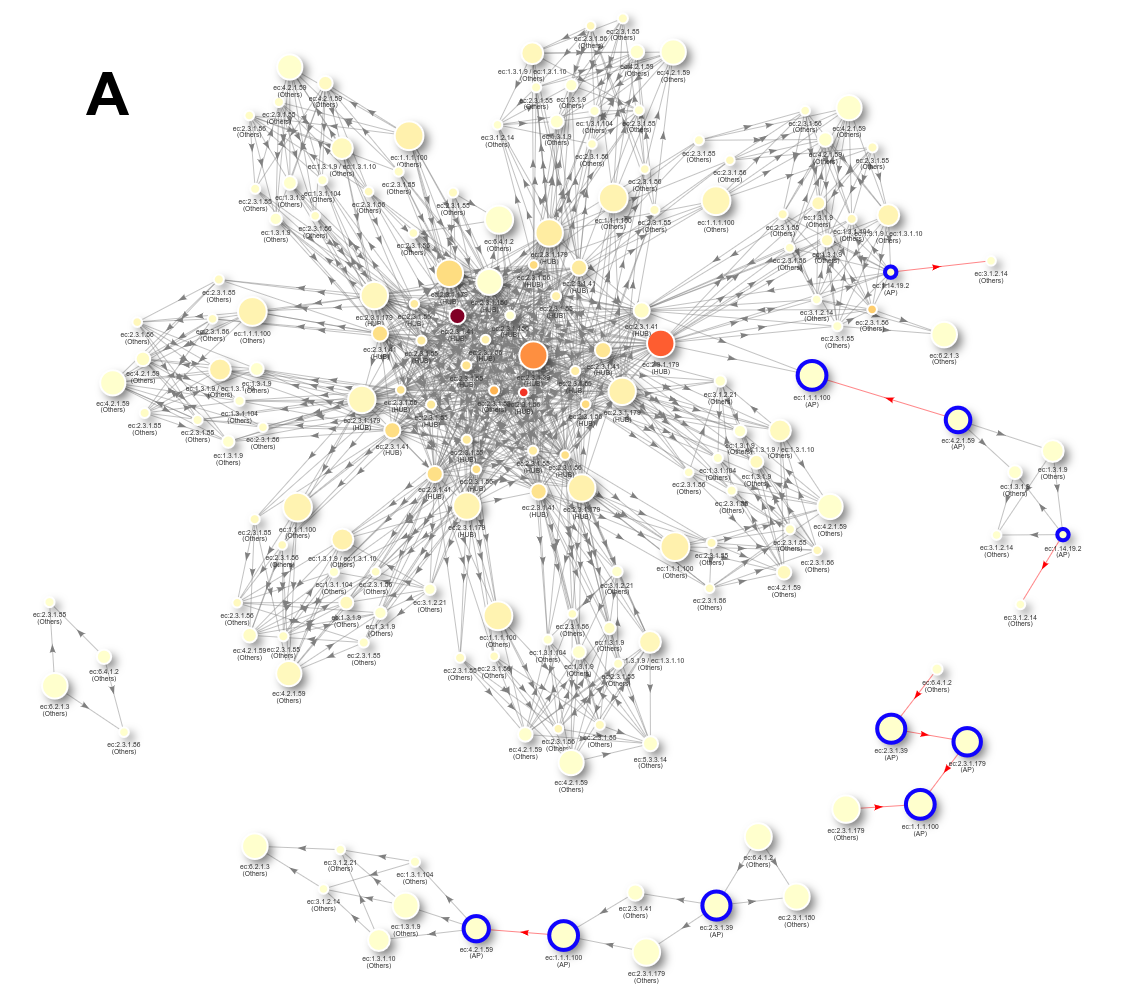
**\*Nodes:** 8 / **Edges:** 7

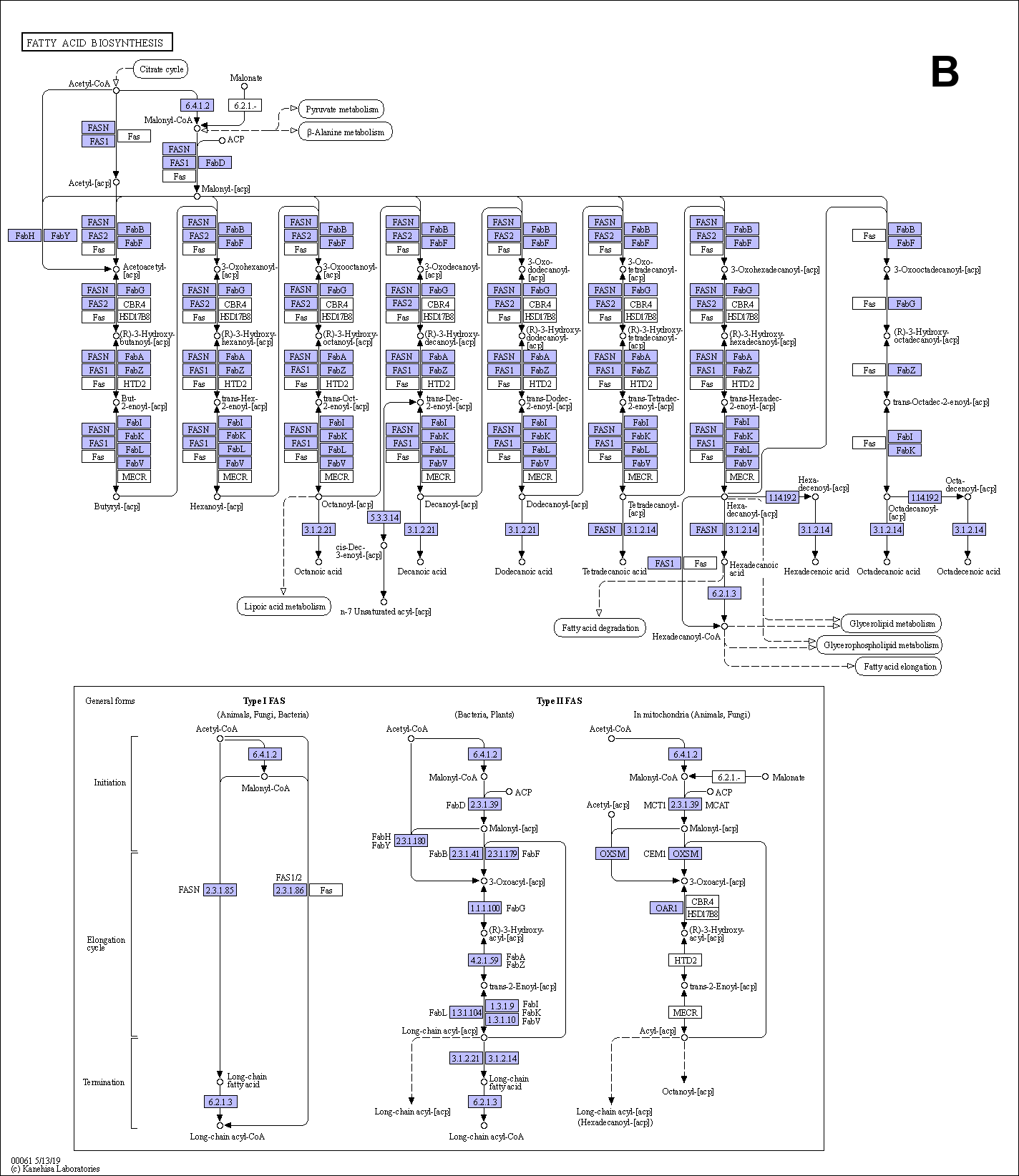
**Supporting information S3:** Pathway 00232 - Caffeine metabolism\*





**\*Nodes:** 26 / **Edges:** 76

**Supporting information S4:** Pathway 00061 - Fatty acid biosynthesis\*

****

**\*Nodes:** 170 / **Edges:** 1063