Proposal for metabolic flux pathways comparison

Propuesta para la comparación de flujos metabólicos

Esteban Arias-Méndez[1], Alonso Montero-Marín[2], Francisco J. Torres-Rojas[3]

Arias-Méndez, E; Montero-Marín, A; Torres-Rojas, F. Proposal for metabolic flux pathways comparison. *Tecnología en Marcha*. Edición especial 2020. 6th Latin America High Performance Computing Conference (CARLA). Pág 31-37.

https://doi.org/10.18845/tm.v33i5.5072

1 Escuela de Computación, Instituto Tecnológico de Costa Rica, Campus Cartago, Costa Rica. Profesor. Correo electrónico: esteban.arias@tec.ac.cr.
https://orcid.org/0000-0002-5600-8381

2 Instituto Tecnológico de Costa Rica, Campus Cartago, Costa Rica. Estudiante. Escuela de Computación. Correo electrónico: luimontero@ic-itcr.ac.cr.
https://orcid.org/0000-0002-8206-5312

3 Escuela de Computación, Instituto Tecnológico de Costa Rica, Campus Local San José, Costa Rica. Profesor. Correo electrónico: ftorres@tec.ac.cr.
https://orcid.org/0000-0001-7075-3305

Keywords

Graph traversal; weighted graphs; Needleman-Wunsch algorithm; global alignment; Smith-Waterman algorithm; local alignment.

Abstract

Metabolic flux pathway analysis can provide important information for a better understanding of life and all its processes directly benefiting areas like medicine, agronomy, pharmacy and others alike. Some of the main tools used to study and analyze metabolic pathways have been based on the idea of pathway comparison, using graph data structures. Some of those tasks are considered hard computational problems. On the other hand, those comparisons have not yet taken into consideration the metabolic flux as part of the pathway or metabolic process. It means, to consider how much of a metabolite passes through a reaction system over time. We propose here a simple way to compare metabolic pathways using its related flux information by a simple metabolic pathway comparison method introduced in 2017 and adjusting it to weighted graphs. The algorithms analyze the associated weighted graphs of metabolic flux pathways and provide a fast scoring of its flux similarities in the first place and a list of similarities and differences between the given flux pathways, listed as pathways. We provide some insights into the analysis follow to get a good score system when comparing metabolic pathways related weighted graphs in a low-cost computation.

Palabras clave

Recorrido de grafos; grafos ponderados; algoritmo Needleman-Wunsch; alineamiento global; algoritmo Smith-Waterman; alineamiento local.

Resumen

El análisis de flujos metabólicos puede proporcionar información importante para una mejor comprensión de la vida y todos sus procesos, beneficiando directamente a áreas como la medicina, la agronomía, la farmacia y otras. Algunas de las principales herramientas utilizadas para estudiar y analizar las rutas metabólicas se han basado en la idea de la comparación de rutas metabólicas, utilizando estructuras de datos como grafos. Algunas de esas tareas se consideran problemas computacionales difíciles. Por otro lado, esas comparaciones aún no han tenido en cuenta el flujo metabólico como parte de la vía o proceso metabólico. Es decir, considerar cuánta cantidad de un metabolito pasa a través de un sistema de reacción con el tiempo. Proponemos aquí una forma simple de comparar rutas metabólicas utilizando su información de flujo relacionada mediante un método simple de comparación de rutas metabólicas introducido en 2017 y ajustándolo a grafos ponderados. Los algoritmos analizan los grafos ponderados asociados de las rutas de flujo metabólico y proporcionan una puntuación rápida de sus similitudes de flujo en primer lugar y una lista de similitudes y diferencias entre las rutas de flujo dadas, enumeradas como rutas. Proporcionamos algunas ideas sobre el análisis a continuación para obtener un buen sistema de puntuación al comparar grafos ponderados relacionados con las rutas metabólicas en un cálculo de bajo costo.

Introduction

Metabolic flux is the passage of a metabolite through a reaction system over time, and flux analysis is the combination of time-course methodologies in metabolomics and computational modeling of pathways [1]. A metabolic pathway is an ordered sequence of biochemical reactions between various actors named metabolites, these are substrates that are transformed into a product through a series of reactions catalyzed by enzymes [2], [3].

In graph theory, a graph is a structure consisting of a set of objects in which some pairs of the objects are in some sense "related". The objects correspond to mathematical abstractions called vertices (also called nodes or points) and each of the related pairs of vertices is called an edge (also called link or line) [4]. Additional info can be added to a graph by assigning a weight to each edge of the graph. Weighted graphs are used to represent structures in which pairwise connections have some numerical values, usually representing a value of the strength or cost of the relationship between a pair of nodes.

Graph comparison is usually referred by the formal notion of *graph isomorphism* which captures the informal notion that some graphs have *a similar structure*. In graph theory, an isomorphism of graphs G and H is a bijection between the vertex sets of G and H, such that any two vertices u and v of G are adjacent in G if and only if f(u) and f(v) are adjacent in H. A problem with isomorphism alone is that it is usually specified for graphs with similar count of nodes and edges.

That being said, it means that graphs of different sizes cannot be well compared using this measure; however, in this present work we would like to present an application to compare weighted graphs representing metabolic fluxes, which had not been previously considered in many analysis about the comparison of metabolic pathways itself. Weighted graphs are very closely related to the topic of metabolic flux pathways, there is a great interest in its study and analysis, and weighted graphs have been used to describe the result of the flux discovery or prediction [5]. Flux analysis can be use also in other fields measuring different processes and its related behavior inside them. We present weighted graph comparison from the point of view of fluxes, determining how similar (structurally: nodes and its related weighted relations) a given pair of flux pathways are.

Many different techniques have been developed for the alignment and comparison of general graphs and the interesting routes inside them. Graph comparison is a computationally difficult task [6], [7], most of these comparisons can be represented as problems in the NP-Complete complexity class, which means there are currently no efficient algorithms for solving them. In some works, like [8] [9], [10], some heuristic techniques have been applied in order to reduce the time taken by graph-alignment algorithms. However, none have considered weighted graphs.

Algorithms

On a previous work [11], two different low-cost approaches were developed as simple mechanisms for the comparison of two metabolic pathways that can be used as a previous step to a deeper and more time-consuming analysis to be applied for the graph comparison associated to the pathways. Later, as an extension of the original algorithms, they were applied to a more general kind of graphs [12]. We propose here a simple way to compare metabolic flux pathways, using weighted graphs, based on the ideas of the previous works. These later works took into consideration the way to compare graphs with different base structures and allow us to look for subgraphs or parts of graphs similar between them.

Methodology

Graphs representing any metabolic pathway or metabolic flux pathway are special graphs, for two key reasons: 1) since the information on the graph is not represented just as a regular directed or weighted graph but representing the order of a reaction system or process. 2) the nodes on each graph are not any but distinguished nodes and we use this to simplify the extraction of information from the corresponding graphs and the comparison process.

Algorithm A - Transformation from 2D to 1D and alignment: we must convert the 2D metabolic flux pathway related graph into a linear version of it, now taking into consideration the weights related. For this graph traversal we don't look for the first metabolite in the reaction, but we start with the nodes associated to the edge with the higher value and then continuing with the lower values while writing down the nodes visited following this path in a decreasing order. The process is applied to both metabolic flux pathways associated graphs to be compared and then regular alignment algorithms are applied (global and local). Exactly the opposite is applied when the interest is to start from the edge with less weight or lower value to the heaviest or highest one. In this case we don't look necessarily to visit all the nodes in the graph, but he most important (heaviest) pathway.

Algorithm B - List the Differences: we look for the differences between the metabolic flux pathways. What we do is to take out the edges with equal value between the same couple nodes in both graphs, so we can say it is the same relation. A correctness adjustment value can be introduced in this comparison in order to allow a gain or loss of an amount or value indicated by the user, like a +/- 5 value for instance. So, if the edge is similar inside that weight range (and its related nodes) then the relation is considered similar and anotated. Also, the user can indicate if the weights of the graphs should be normalized or no; this means, all the weights in the graph must sum 100.00 or 1.00 so, all the weights represent its correspondent percentage of the total value of the graph. This is to make possible to compare metabolic flux pathways with different weight scales, but that might have similar flux structures. A list of the found differences (weighted reactions) is given to the user.

Discussion

To provide insights on how our algorithms works we provide a sample case to show the outputs of the algorithms and its implementations. Let's consider a pair of metabolic flux pathways shown on figure 1 and figure 2, in an intuitive way, they look very similar and we want to reinforce that with the data extracted.

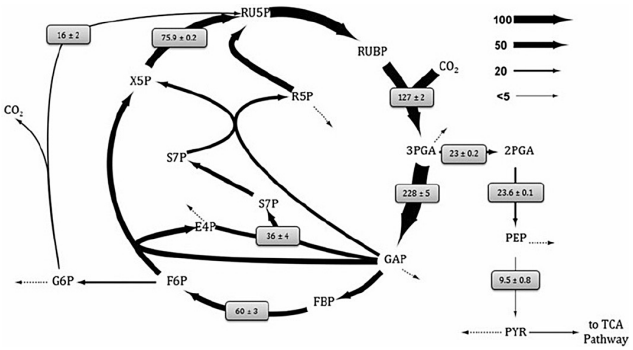

Figure 1. Metabolic Flux Pathway: MFP 1. This flux map shows the estimated fluxes associated with glycolysis and the Calvin cycle for a Synechocystis INST-MFA study. From [5].

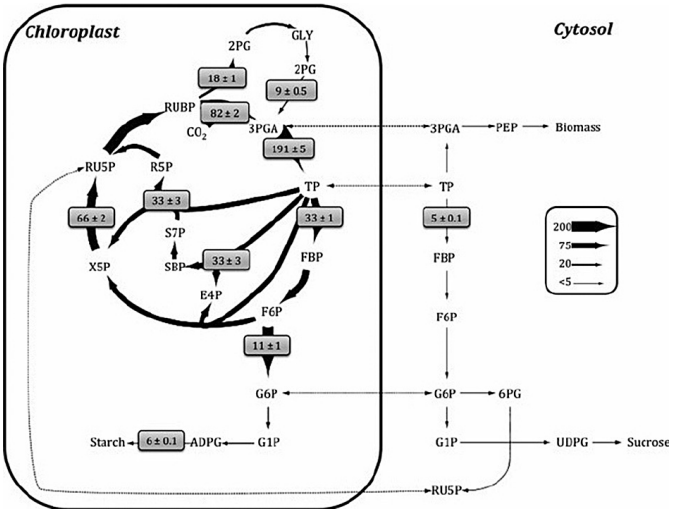

Figure 2. Metabolic Flux Pathway: MFP 2. This flux map shows a hypothetical flux map associated with a plant INST-MFA study involving multiple subcellular compartments. From [5].

For the first algorithm we show the traversal outputs of the associated graphs to the metabolic flux pathways MFP 1 and MFP 2 based on the weights only, from the highest value to the minimum. Not all the values of the original pathway may be considered, since some of the metabolite flux passages can produce cycles or some other fluxes might go to some other reactions. We want to focus on the major flux.

```
MFP 1 traversal: RUSP – RUBP – 3PGA – 6AP – FBP – F6P – X5P
```

```
MFP 2 traversal: RUSP – RUBP – 3PGA – TP – S7P – X5P
```

Then, for these outputs the corresponding Global and Local alignments of pathways MFP 1 and MFP 2, using standard values match=+1, missmatch=-1, gap=-2, we obtained a Global score = 0, and Local score = 3, with alignment: RUSP RUBP 3PGA. We can observe here that the local score value shows a similarity between pathways MFP 1 and MFP 2 between the nodes RUSP to 3PGA.

For the second algorithm we want to find the differences between the weighted paths in the graphs, to discover the broader similar subgraph from the point of view of the associated weights to each pathway, so we can compare better the flux between the given graphs. Let's consider that here we focus in the similar relationship between the nodes and its associated weight, so that an edge connecting two nodes is considered relevant for the comparison if it shows equal o similar weight for the same relation in the other graph. We consider in our algorithm a delta value as a range difference permitted when comparing a couple edge's weights. So, if the same relation is found between the same two nodes in both graphs, but the weights associated to each edge in each graph are not equal or they differ on a maximum delta value given, we don't consider this edge as similar.

Comparing then the pathways MFP 1 and MFP 2 in this way we get the outputs below:

```
MFP 1 - MFP 2 — Similar weighted paths: RUSP—RUBP—3P6A, FBP—F6P—
X5P—RUSP, S7P— X5P—RUSP, R5P—RUSP
```

```
MFP 1 - MFP 2 — Differences: 3PGA—2PG—GLY, 3PGA—2PGA—PEP—PYR, 3PGA—
GAP—RSP, 3PGA—GAP—S7P, 3PGA—GAP—E4P, 3PGA—GAP—FBP—F6P—G6P—RUSP
```

Conclusions

As we can see in the outputs obtained, we don't look to propose which metabolic flux is more efficient of try to get a perfect match between a pair of given pathways, but to provide some insights on the most relevant information on a couple hopefully related processes. This means provide the part of the reaction (sub-graph) more similar between the pathways that matches the related metabolites nodes and its associated weights, for the first case, and the list of main different parts of the pathways, for the second algorithm.

As explained before, the fact that the graphs represents metabolic information is considered: the specific order in the reactions (edges) and corresponding metabolites (nodes), and that each node represents a distinguished metabolite; all this is used to minimize the amount of comparisons made and simplify the extraction of information from the graphs, and to analyze that information. For metabolic flux pathway comparison, we established that the mechanism proposed by the first approach transforming a 2D structure to 1D structure following its weights for later alignment and evaluation can be used as a testing evaluation to predict good comparison results in case a deeper analysis is desired. For the second approach we want to offer to the expert an additional point of view for evaluations about the metabolic flux pathways being compared. In this case, no score is provided but the listed similarities and differences.

The proposed algorithms are fast and of relatively low computational cost, it is possible to provide relevant information for the comparison study about metabolic flux pathways of interest and some other derived analyzes. On the other hand, there are currently not many information available about metabolic flux pathways on databases, most of this information comes on papers analyzing and interpreting the meaning of the fluxes on a metabolic process in study. So, it's our interest to provide this proposal for a future necessity of comparing more discovered metabolic flux pathways.

Acknowledgment

We would like to thank the support provided by the Computer Science School of Instituto Tecnológico de Costa Rica and the Vicerectory of Research and Extension.

References

[1] S. Padmanabhan, *Handbook of pharmacogenomics and stratified medicine.* Academic Press. 2014.

[2] J. M. Lee, E. P. Gianchandani, J. A. Eddy, & J. A. Papin, "Dynamic analysis of integrated signaling, metabolic, and regulatory networks". *PLoS computational biology*, e1000086, 2008.

[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, & P. Walter, "Molecular biology of the cell." *Garland Science*. 2007.

[4] R. J. Trudeau, R. J. (2013). *Introduction to graph theory.* Courier Corporation.

[5] J. D. Young, A. A. Shastri, G. Stephanopoulos, & J. A. Morgan, "Mapping photoautotrophic metabolism with isotopically nonstationary 13C flux analysis". *Metabolic Engineering*, pp. 656-665, 2011.

[6] G. Abaka, T. Bıyıkoglu, & C. Erten, "CAMPways: constrained alignment framework for the comparative analysis of a pair of metabolic pathways." *Bioinformatics*, pp. 145-153. 2013.

[7] F. Ay, M. Kellis, & T. Kahveci, "SubMAP: aligning metabolic pathways with subnetwork mappings". *Journal of Computational Biology*, pp. 219-235, 2011.

[8] O. Kuchaiev, & N. Przulj, "Integrative network alignment reveals large regions of global network similarity in yeast and human". *Bioinformatics*, pp. 1390-1396, 2011.

[9] R. Patro, & C. Kingsford, "Global network alignment using multiscale spectral signatures". *Bioinformatics*, pp. 3105-3114, 2012.

[10] R. Y. Pinter, O. Rokhlenko, E. Yeger-Lotem, & M. Ziv-Ukelson, "Alignment of metabolic pathways". *Bioinformatics*, pp. 3401-3408, 2005.

[11] E. Arias-Mendez, & F. Torres-Rojas, “Alternative low cost algorithms for metabolic pathway comparison.” *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI),* Funchal, Portugal: IEEE. pp. 1-9, 2017.

[12] E. Arias-Mendez, A. Montero-Marin, D. Chaves-Chaves, & F. J. Torres-Rojas, “Simple Graph Comparison Inspired on Metabolic Pathway Correlation”. *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)* San Carlos, Costa Rica: IEEE. pp. 1-8, 2018.