Ecological Modelling 313 (2015) 259–265

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

A comparison between functional frequency and metabolic flows framed by biogeochemical cycles in metagenomes: The case of "El Coquito" hot spring located at Colombia's national Nevados park

Maria A. Zamora[a], Andres Pinzón[b], Maria M. Zambrano[c], Silvia Restrepo[d], Linda J. Broadbelt[e], Matthew Moura[e], Johana Husserl Orjuela[f], Andrés F. González Barrios[a,*]

[a] Grupo de Diseño de Productos y Procesos (GDPP), Departamento de Ingeniería química, Universidad de los Andes, Bogotá, Colombia
[b] Bionformatics and Systems Biology Group, Institute of Genetics, National University of Colombia, Bogotá, Colombia
[c] GEBIX—Centro Colombiano de Genómica y Bioinformática, Bogotá, Colombia
[d] Laboratorio de Micología y Fitopatología, Departamento de Ciencias biológicas, Universidad de los Andes, Bogotá, Colombia
[e] Department of Chemical and Biological Engineering, McCormick School of Engineering and Applied Sciences, Northwestern University, Evanston, IL, USA
[f] Grupo de Investigación en Ingeniería Ambiental (CIIA), Departamento de Ingeniería Civil y Ambiental, Universidad de los Andes, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 20 February 2015
Received in revised form 16 June 2015
Accepted 29 June 2015

Keywords:
Metabolic reconstruction
Network topology analysis in ecosystems
Ecosystem FBA

ABSTRACT

The study of metagenomic samples is crucial for understanding microbial communities in ecosystems. In this study, genomic samples of the "El Coquito" acidic hot spring, located at the Nevados Natural National Park in Colombia, were analysed to evaluate the relationship between its metabolic functionality, and metabolic flows distribution considering the thermodynamic restrictions and the biogeochemical cycles through a flux balance analysis. The metabolic functionality of the metagenome was established by assigning reactions and enzymes to the metabolic routes. On one hand, we found that relevant metabolic groups contribute in a different fashion depending on the objective function of the ecosystem from the amount of metabolic reactions perspective. On the other hand, from the contribution to the total metabolic flow point of view, it seems the metabolic distribution plasticity conserves, as there exists several common groups that significantly contribute to the total metabolic flux of the system. Nevertheless, this plasticity is not maintained for all objective functions due to the differences between the nitrogen and the rest of the cycles, clearly demonstrating that the microbial adaptation capability to the ecosystem is materialized in the distribution of the metabolic flows instead of the frequency of appearance of the same reactions.

1. Introduction

Metagenomics consists of the genome-based analysis of entire communities in diverse ecological contexts (Dupré and O'Malley, 2007) and it can be defined as the genomic analysis of microbial communities found in an environmental sample such as soil, ocean, rivers, and hot springs (Dupré and O'Malley, 2007; Kakirde et al., 2010; Gilbert and Dupont, 2011; Keshri et al., 2013), among others, or communities associated to other organisms like microorganisms of the human or animal gut (Palackal et al., 2007; Karlsson et al., 2011). Metagenomic comparative approaches have been used to study a wide variety of viral and microbial environmental sequences to elucidate the functional potential of nine biomes (Dinsdale et al., 2008) or the influence of marine microbes on the biogeochemical cycles in the planet (Yilmaz et al., 2011). Understanding how microbial activities influence the biochemical cycles of the planet requires the identification of the metabolic pathways, enzymes, and reactions involved in global processes (Chen et al., 2012), leading to the development of metabolic reconstructions (Thiele and Palsson, 2010). The number of algorithms capable of reconstructing metagenomic networks has therefore increased over time with pipelines such as MGRAST (Meyer et al., 2008) or those proposed by Pinzón et al. (2011).

Once the metabolic reconstruction is obtained, one can obtain information about the metabolic phenotype of the microbiome from different levels. First, it is possible to study the richness, the

* Corresponding author. Tel.: +57 13394949.
E-mail addresses: ma.zamora72@uniandes.edu.co (M.A. Zamora), ampinzonv@unal.edu.co (A. Pinzón), mzambrano@corpogen.org (M.M. Zambrano), srestrep@uniandes.edu.co (S. Restrepo), broadbelt@northwestern.edu (L.J. Broadbelt), MatthewMoura2015@u.northwestern.edu (M. Moura), andgonza@uniandes.edu.co (A.F. González Barrios).

http://dx.doi.org/10.1016/j.ecolmodel.2015.06.041
0304-3800/

frequency, and the functionality of the ecosystem; second, one can evaluate the response to different perturbations utilizing networks physics approaches. Nevertheless, having an idea of the flow of metabolites in the network is not possible departing from these two levels. To do so, it is necessary to use a different approach that involves linear programming optimization methods called flux balanced analysis (FBA), which enables researchers to obtain a quantitative understanding of the metabolic flows and their distribution. Using FBA results in a third level on the understanding of the phenotype.

There are several factors that must be considered in a metabolic reconstruction including variations in sample composition (Keshri et al., 2013; Mocali and Benedetti, 2010) and ecosystem's environmental conditions, e.g., extreme temperatures (Lewin et al., 2013), salinity (Jeffries et al., 2011), variations in pH (Jones et al., 2012), as well as the compartmentalization of the system (Roze et al., 2011) and the thermodynamic feasibility of the reactions (Henry et al., 2007). Thermodynamics allows researchers to understand metabolic networks and their regulations. Metabolic reactions could appear possible from the point of view of the balance calculations, which is the base of FBA, but could be thermodynamically unfeasible, depending on metabolite concentrations due to positive free energy values. Using this tool for analyzing the reactions of the ecosystem, might result powerful for predicting maximal possible yields for the biosynthesis of valuable products, and it could also be used for assessing the feasibility of engineering novel metabolic pathways (Maskow and von Stockar, 2005).

The evaluation of the thermodynamic feasibility of a metagenome derived from experimental data is highly restrained due to the lack of data available for biological systems. The inclusion of thermodynamic restrictions has not yet been studied for metagenomic samples. Nevertheless, there are approximations based on classical thermodynamics that could be applied to biological systems. Several authors (Stephanopoulos et al., 1998; Mavrovouniotis, 1991) have utilized the contribution method to calculate the standard Gibbs free energy change of reaction $\left(\Delta_r G'^o\right)$ and the standard Gibbs free energy of formation $\left(\Delta_f G'^o\right)$ of the majority of the compounds present in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and in a model belonging to *Escherichia coli* (Henry et al., 2006). Moreover, there are expanded group contribution methods available that reduce the uncertainty of the calculation (Henry et al., 2007). Some of the studies are aimed at investigating the metabolic potential along with determining the ecological role of the microorganisms present in the sample (Jones et al., 2012).

Lately, the study of metabolic networks has received great attention (Dinsdale et al., 2008). Metabolic fluxes constitute a fundamental determinant in cell physiology, primarily because they provide a measure of the degree of engagement of various pathways in overall cellular functions and metabolic processes. Knowing this information for a metabolic network reconstructed from a metagenome could provide valuable and novel information regarding the engagement within the ecosystem. FBA can provide the necessary knowledge on metabolic pathway fluxes. Using this methodology, researchers can calculate intracellular fluxes by using stoichiometric models for the major intracellular reactions and apply mass balances for intracellular metabolites (Stephanopoulos et al., 1998; Kauffman et al., 2003; Çalık et al., 2011). Genomic comparative tools have been used for the analysis of a variety of metagenomic samples (Dinsdale et al., 2008), however the FBA offers a new insight for extreme environments ecosystems. Also, FBA of a metagenomic sample can improve the knowledge regarding the reactions involved in the metabolic network and as a result improve the precision of the reconstructed models (Filippo et al., 2012).

During the development of a FBA, an objective function is used to quantitatively define how much each reaction contributes to the phenotype. The objective function corresponds to the biological objective that is relevant to the problem being studied: for example maximization of biomass production and nutrient uptake. In the case of metagenomes the objective functions are not always obvious. Combining the mathematical representations of the metabolic reactions and the objective function, it is possible to define a system of linear equations. These equations are solved using linear programming (Orth et al., 2010). There are several computational linear programming algorithms available that are capable of identifying optimal solutions for the set of linear equations, such as the COBRA Toolbox (Becker et al., 2007) or e Xpress MP (Guéret et al., 2002).

Systems biology approaches based on network topology analysis can provide important information when analyzing ecological niches from the microbiome perspective through metagenomic reconstructions. The biochemical reactions in cellular metabolism can be integrated into a metabolic network in which fluxes are regulated by enzymes catalyzing the reactions (Albert, 2005). There are few studies at a metagenomic level that take into account the topological analysis, for example, a metagenomic systems biology approach that goes beyond traditional comparisons and reveals shifts associated with obesity and inflammatory bowel disease is reported in communities such as the human gut microbiome (Greenblum et al., 2012).

During a topological analysis, researchers can predict the response of the ecosystem to external perturbations and the degree of evolution that results from changes in the environment. These aspects can be described using parameters such as the clustering coefficient, characteristic path length, average number of neighbors, and number of nodes. If the preferred directionality of a reaction is known, only the largest strongly connected component has a well-defined average path length (Bourqui et al., 2007). Knowing this directionality allows a better reconstruction of the network, reducing the number of reactions and restraining some pathways.

Based on these observations the aim of this work is to compare the functional frequency and the metabolic flows in the metagenome of a microbial planktonic community of an acidic hot spring "El Coquito", located at the Nevados National Park in Colombia, based on the metabolic network reconstruction of the ecosystem and consequent analysis of its topology and metabolic flows distribution using FBA. Here we depart from a functional description of the metagenome that has already been reported (Bohorquez et al., 2012). In this study we identified the metabolic pathways, the reactions, and the metabolites associated with the sample. Results obtained were used to conduct a FBA with three different objective functions associated with nitrogen, sulfur, and carbon cycling. Finally, a topological comparative analysis of the network was carried out considering the thermodynamic restrictions of the existing genomic sequenced samples from the same hot spring.

2. Methods

2.1. Sample collection, analysis and processing

Superficial running stream water was collected at the spring "El Coquito" and samples were processed as described (Bohorquez et al., 2012). Essentially, 16 L were collected in 5-L sterile plastic containers, transported to the laboratory at 4 °C and processed within 18 h for DNA extraction and physicochemical analyses (SO_4, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Fe^{2+}, Fe^{3+}, $CaCo^{3+}$, NO^{3+}, Chloride, PO^{4+} and total dissolved solids). DNA was extracted by filtering water (10 L) through 5.0 μm cellulose and 0.22 μm polycarbonate filters, as