Detecting evolution and adaptation fingerprints in bacterial
metabolic backbones

Oriol Güell,1 Francesc Sagués,1 and M. Ángeles Serrano2

1Departament de Química Física, Universitat de Barcelona,
Martí i Franquès 1, 08028 Barcelona, Spain
2Departament de Física Fonamental, Universitat de Barcelona,
Martí i Franquès 1, 08028 Barcelona, Spain

Abstract

The heterogeneity of reaction fluxes present in metabolic networks can be exploited to construct
disparity backbones as reduced version of metabolism. These backbones maintain all relevant in-
formation while displaying a substantially decreased number of interconnections and, hence, they
become a useful tool to unveil important biological information. Here, we apply the disparity filter
method based on a high-flux fluctuation analysis to extract the metabolic backbones of two bac-
terial organisms, Escherichia coli and Mycoplasma pneumoniae, using computationally predicted
fluxes that maximize their growth rate. We find that the metabolic backbones of both organisms
in minimal medium are mainly composed of a core of reactions belonging to ancient pathways,
meaning that reactions in the early stages of the life of the organisms still remain at present signif-
icant for biomass production. At the same time, a comparative analysis of backbones in different
media leads to the identification of pathways sensitive to environmental changes. Backbones, as
networks of metabolic flux superhighways, are thus useful to trace simultaneously both long-term
evolution and short-term adaptation fingerprints in cell metabolism.
I. INTRODUCTION

The large amount of biological data at the cell level gathered to date allows the construction of high-quality genome-scale metabolic reconstructions composed of thousands of reactions and metabolites [1–4]. The development of constraint-based optimization techniques [5, 6] along with complex network science [7, 8] has enabled a huge step further towards the elucidation of important biological information hidden in genome-scale metabolic reconstructions. A useful tool in this endeavor is the concept of backbone. Backbones maintain relevant information while displaying a substantially decreased number of interconnections and, hence, can provide accurate but reduced versions of the whole system. In particular, the work by Almaas et al. [9] introduced a filtering technique that selects the reaction that dominates the production or consumption of each metabolite such that a high-flux backbone can be retrieved. Although this method recovers pathways, the obtained backbones present a linear structure with very little interconnectivity and lack many of the features of real metabolic networks [10, 11].

Filtering approaches have also interested researchers working on networks in a more general context. A filtering method for weighted networks with a high density of connections based on the disparity measure [12, 13] was developed in Ref. [14]. This approach exploits the heterogeneity present in the intensity of interactions in real networks both at the global and local levels [15] to extract the dominant set of connections for each element. Typically, the obtained disparity backbones (DB) preserve almost all nodes in the initial network and a large fraction of the total weight, while reducing considerably the number of links that pass the filter. At the same time, DBs preserve the heterogeneity and cutoff of the degree distribution, the degree of clustering, and the bow-tie structure [16], and other characteristic features of the original networks [14]. Hence, the complex features of the original networks are preserved.

In this work, we use Flux Balance Analysis (FBA) [5] to determine reaction fluxes and the disparity filter [14] to extract the metabolic backbones of two organisms: *Escherichia coli* (*E. coli*) and *Mycoplasma pneumoniae* (*M. pneumoniae*). We investigate these backbones for fingerprints of evolution and adaptation. We find that the metabolic backbones of both organisms in minimal medium are mainly composed of a core of reactions belonging to ancient pathways. This means that the significant fluxes in these bacterial metabolisms are
associated to reactions which have been present from the earliest stages of their life and still remain at present significant for biomass production. We also study how external conditions modify the structure of the backbones, which allows us to identify pathways that are more sensitive to changes in the environment and so prone to short-term adaptation.

II. RESULTS

We use FBA to compute the fluxes of the reactions composing the metabolic networks (see section A in Methods). These fluxes are treated as weights by the disparity filter (see section B in Methods). We work with iJO1366 version of E. coli K-12 MG1655 [4], which contains 1805 metabolites and 2583 reactions (including auxiliary reactions and taking all cellular compartments into account) and with iJW145 version of M. pneumoniae [17], which contains 266 metabolites and 306 reactions. We model these networks as bipartite graphs with two kind of nodes, metabolites and reactions, and incoming, outgoing, and bidirectional links. FBA calculations are performed in glucose minimal medium with a maximum uptake of glucose limited to 10 mmol gDW\(^{-1}\) h\(^{-1}\) for E. coli and 7.37 mmol gDW\(^{-1}\) h\(^{-1}\) for M. pneumoniae (we add D-ribose to enrich the medium for M. pneumoniae). The connectivity structure (see section C in Methods) of the obtained backbones is analyzed from an evolutionary perspective. Additional media are considered to analyze environmental sensitivity (see section D in Methods).

A. Identification of the disparity backbones of metabolic networks

An important feature of flux solutions obtained using FBA is the heterogeneity of the flux distributions. In the same state, fluxes of reactions can span several orders of magnitude [9]. To check this statement, we show the probability distribution function of the obtained fluxes (disregarding zero-flux reactions) in the insets of Figs. 1b and c, confirming that, indeed, fluxes show an heterogeneous distribution at the global level. The set of metabolites in non-zero flux reactions is considerably reduced from the original total number, from 1805 to 445 metabolites for E. coli, and from 266 to 227 metabolites for M. pneumoniae. To characterize the existence of such heterogeneity also at the local level, we calculate the disparity measure [9, 14] for every metabolite \(i\), \(\Upsilon_i(k) = k \sum_{\nu_j \in \Gamma(i)} (\nu_j / \sum_j \nu_j)^2\) (see section
B in Methods), accounting for the $k$ reactions $j$ in its neighboring set $j \in \Gamma(i)$ with corresponding fluxes $\nu_j$. Figures 1b and 1c display the disparity values for all metabolites as a function of their incoming and outgoing degree in *E. coli* and *M. pneumoniae*, respectively. The shadowed areas correspond to values compatible with a random distribution of fluxes among the reactions producing or consuming a metabolite and help to discount local heterogeneities produced by random fluctuations (see caption of Fig. 1). As shown, most metabolites present flux disparity values that cannot be explained by random fluctuations meaning that the local distribution of the fluxes of reactions associated to metabolites is significantly heterogeneous. We conclude then that the disparity filter will be able to efficiently extract a backbone with the most relevant connections for both organisms, while preserving the characteristic features of metabolism as a complex network.

Briefly, the disparity filter works by comparing weights of links with a random assignment. The filter preserves a link in the backbone if the probability that its normalized weight $\alpha_{ij}$ is compatible with the random assignment (p-value) is smaller than a chosen threshold $\alpha$ which determines the filtering intensity (see section B in Methods for more details). We proceed to filter the metabolic networks with fluxes of reactions as weights of the connections between metabolites and reactions. For each metabolite $i$, we compute the $\alpha_{ij}$ of each connection between metabolite $i$ and its neighboring reactions $j$ and compare the obtained p-value with the significance level $\alpha$. The disparity filter can be adjusted by tuning this threshold to observe how the metabolic networks of both *E. coli* and *M. pneumoniae* are reduced as we decrease $\alpha$ from 1 to 0, both of them included, $\alpha = 1$ meaning the complete network. Notice that, after applying the filter, we recover a bipartite representation of the metabolic backbone. To avoid working with stoichiometrically non-balanced reactions, we transform the filtered bipartite representation into a one-mode projection of metabolites placing a directed link between two metabolites if there is a reaction whose flux is simultaneously relevant for the consumption of one metabolite and for the production of the other [9]. In this one-mode projected backbone, we compute how many links $E$, nodes $N$ and total weight $W$ remain. These magnitudes are normalized by dividing them by the corresponding values in the original network, $E_T$, $N_T$, and $W_T$.

In Figs. 1d and e, we show the dependencies $N/N_T$ vs $E/E_T$, and $W/W_T$ vs $E/E_T$ in the associated insets, for the one-mode metabolic projections of the backbones of both *E. coli* and *M. pneumoniae*. While the filter can reduce considerably the fraction of links, the
FIG. 1. Scheme of the application of the disparity filter and measures of the heterogeneity of reaction fluxes in *E. coli* and *M. pneumoniae*. (a) Scheme of the filtering method. Blue nodes are metabolites and green squares denote reactions. Incoming connections to metabolites are represented by red arrows, outgoing connections with blue arrows, and bidirectional connections with dark yellow arrows. OMP denotes one-mode projection. (b) Disparity measure as a function of incoming and outgoing degrees \(k\) in *E. coli*. The shadowed area corresponds to the average plus 2 standard deviations given by the null model, meaning that points which lie outside this area can be considered heterogeneous [14]. Inset: global distribution of fluxes of *E. coli*. (c) Disparity measure as a function of IN and OUT degrees \(k\) for *M. pneumoniae*. Again, the shadowed area corresponds to the average plus 2 standard deviations given by the null model. Inset: global distribution of fluxes of *M. pneumoniae*. (d) Fraction of nodes as a function of the fraction of links in *E. coli*. Inset: remaining weight as a function of the fraction of links in the network. (e) Fraction of nodes as a function of the fraction of links in *M. pneumoniae*. Inset: remaining weight as a function of the fraction of links in the network.
corresponding fraction of nodes is maintained at almost the original value. In addition, the total weight in the backbone only starts to drop appreciably after more than 50% of the links are removed. We take the critical value $\alpha_c$ as the point where the fraction of nodes starts to decay (see Figs. 1d and e). This critical value can be seen as an optimal point which reduces greatly the number of links in the network preserving at the same time most nodes and so as much biochemical and structural information as possible. The values are $\alpha_c = 0.21$ for *E. coli* and $\alpha_c = 0.37$ for *M. pneumoniae*.

**B. Evolutionary signatures in the backbones of metabolites**

We construct the metabolic backbones of both *E. coli* and *M. pneumoniae* using the identified critical values for the significance level. The backbones retain all the 445 and 227 metabolites respectively. Next, we analyze their structure in terms of connectedness. Metabolic networks have been found to display typical large-scale connectivity patterns of directed complex networks, called the bow-tie structure [16] (see section C in Methods), with most reactions in a interconnected core, named the strongly connected component (SCC), together with in (IN) and out (OUT) components formed mainly by nodes directly connected to the SCC component [10, 18]. This is the case of the original metabolic networks of both organisms, whose SCCs contain the largest part of the metabolites and reactions of the network that realize all the metabolic processes, and whose IN and OUT components are formed, respectively, by nutrients and waste metabolites.

Metabolites in the backbone of *E. coli* are arranged in a connected component of 178 nodes and several disconnected small components (51). Three different SCCs can be identified in the connected part of the backbone, each with 25%, 10%, and 6% of the nodes in the connected component (see Fig. 2a). The two smallest SCCs are in the OUT component of the largest SCC. For the three of them, we recover its IN and OUT components and tendrils. Metabolites corresponding to central compounds of metabolism are identified in these SCCs: protons, water, ATP, glutamate, phosphate, NAD$^+$, diphosphate, ADP and FAD$^+$. These metabolites are highly-connected metabolites even in the metabolic backbone, preserving the same structural features of the complete metabolic network.

Since links in the metabolic backbone denote reactions transforming metabolites, it is interesting to identify links with the pathway associated to the corresponding reaction. In this
FIG. 2. SCCs of the backbone of metabolites and corresponding pathways. (a) Connected component in the metabolic backbone of *E. coli*. The colors of the nodes depend on the component each node belongs to (yellow: SCC1, green: IN component of SCC1, red: OUT component of SCC1, violet: tendrils of SCC1, cyan: SCC2, blue: SCC3). The color of the links, and its association given in the legend, depends on the functional categories given in Ref. [4], where each category contains pathways that realize similar tasks. (b) Connected component of the metabolic backbone of *M. pneumoniae*. The color of the nodes denote again the component each node belongs to (red: SCC1, turquoise: IN component of SCC1, green: OUT component of SCC1, dark yellow: tendrils of SCC1, violet: SCC2). The color of the links, and its association given in the legend, depends on the pathway each reaction belongs to. (c) Percentage of links in pathways for the largest SCC in the metabolic backbone of *E. coli*. (d) The same for *M. pneumoniae*. 
way we can count the composition of the three SCCs in terms of pathways. Starting with
the largest SCC (see Fig. 2a), we find that the major contributions are Oxidative Phospho-
rylation 26%, Citric Acid Cycle 16%, Glycolysis/Gluconeogenesis 15%, Pentose Phosphate
Pathway 9%, and Glutamate Metabolism 9% (see Fig. 2c). It has been demonstrated that
these routes are ancient pathways that have been conserved through evolution. More pre-
cisely, Glycolysis and Pentose Phosphate Pathway take place without the need of enzymes in
a mimetic Archean ocean [19]. Concerning the Citric Acid Cycle, it is also an ancient path-
way that has evolved in order to achieve maximum ATP efficiency [20] by being coupled to
Oxidative Phosphorylation and Glycolysis [21], in addition to help the organism to decrease
their quantity of reactive oxygen species by modulation of their participating metabolites
[22]. Another pathway significantly present in the largest SCC is Glutamate Metabolism.
Glutamate has been reported to be one of the oldest amino acids used in the earliest stages
of life [23].

Links in the other two SCCs correspond also to reactions belonging to ancestral pathways.
The second largest SCC contains links that belong mainly to Purine and Pyrimidine Biosyn-
thesis (91%). Purines and pyrimidines serve as activated precursors of RNA and DNA,
glycogen, etc. [24, 25], and it has been found that the synthesis of purines and pyrimidines
was the first pathway involving enzyme-based metabolism [26]. Interestingly, the other con-
tribution to this SCC is Glycine and Serine Metabolism. Glycine is a precursor of purines
and pyrimidines. Pathways related to the third SCC are Membrane Lipid Metabolism (97%)
and Cofactor and Prosthetic Group Biosynthesis (3%). Membrane Lipid Metabolism sup-
plies the necessary lipids to generate the cell membrane needing the participation of the
cofactor FAD+/FADH2. It has been shown that the pathways involved in lipid metabolism
exhibit differences between different lineages in organisms [27], whereas pathways related to
central metabolism are more conserved and are transversal [27].

When considering $\alpha$ values smaller than the critical one, implying that the filter is more
restrictive and more heterogeneity is needed to overcome it, we observe that the smallest
SCCs discussed above disappears. More precisely, it happens for a value of $\alpha = 0.19$.
Decreasing even more the significance level to $\alpha = 0.15$ the SCC containing reactions in the
Purine and Pyrimidine Biosynthesis pathway retains the 30% of the nodes for $\alpha_c = 0.21$,
whereas the largest SCC still contains a 86%, showing the large resistance of this large
core to lose nodes. At a value of $\alpha = 0.14$, the second SCC finally disappears and there
only remains a single SCC, still preserving 82% of the nodes in it for $\alpha_c = 0.21$. Hence, energy metabolism shows a large resistance to get fragmented even though the filter becomes progressively more and more restrictive.

To contrast the obtained results in *E. coli*, we perform the same analysis in *M. pneumoniae*, a simpler organism that has been proposed as a bacterial model due to its simplicity and its reduced genome [17, 28]. For *M. pneumoniae*, the critical value $\alpha_c$ is 0.37 (see Fig. 1). The connected component of its metabolic backbone is shown in Fig. 2b. It contains two SCCs, one of them being irrelevant with only two nodes (see Fig. 2b). The relevant SCC contains 21% of the nodes in the connected component, and the largest part of its links are related also with energy metabolism as in *E. coli*. The dominant pathways in this core are Glycolysis and Pyruvate Metabolism (see Fig. 2d). Along Glycolysis, Pyruvate Metabolism is also an ancestral pathway that was present in the earliest stages of life [29], when no oxygen was present in the early atmosphere.

C. The metabolic backbones of *E. coli* encode its short-term adaptation capabilities

The previous section analyzes the metabolic backbone of *E. coli* in glucose minimal medium in terms of the long-term evolution of the organism. In this section, we study how changes in the environment modify this backbone, which exposes short-term adaptation capabilities. First, we calculate the FBA fluxes that maximize the growth rate of *E. coli* in the rich medium Luria-Bertani (LB) Broth [30, 31]. Afterwards, we apply the disparity filter to extract the metabolic backbone in this new environment, that is obtained for a significance level threshold $\alpha_c = 0.4$. This value is noticeably larger than $\alpha_c = 0.21$ identified for the glucose minimal medium. Interestingly, this rich medium activates 400 reactions, 11 less than in glucose minimal medium. Of them, 279 are active in both media, of which 247 have a larger flux in LB Broth. An analysis of the connected components in the metabolic backbone of *E. coli* in rich medium is also performed. We find that it contains a large connected component with 449 metabolites and 60 small disconnected components. The connected component contains also three SCCs. However, two of them are tiny with only two nodes, whereas the largest one encloses 34% of the nodes in the connected component. Interestingly, the pathway contributing more reactions to this large SCC is Membrane Lipid
FIG. 3. Dependence of the distribution of pathways in the metabolic backbone of *E. coli* with the composition of environment. (a) Histogram of the fraction of links belonging to each pathway (x axis) for the 333 minimal media (left) and in the rich medium (right) (b) Probability distribution function of $\alpha_c$ for all minimal media. (c) Probability distribution function of the fraction of links in the metabolic backbones for all minimal media. (d) Histogram of weights of links in the metabolic superbackbone. Pathway acronyms: ACM: Alternate Carbon Metabolism, TIM: Transport, Inner Membrane, MLM: Membrane Lipid Metabolism, NSP: Nucleotide Salvage Pathway, CEB: Cell Envelope Biosynthesis, PPB: Purine and Pyrimidine Biosynthesis, CAC: Citric Acid Cycle, OP: Oxidative Phosphorylation, GG: Glycolysis / Gluconeogenesis, GM: Glutamate Metabolism.

Metabolism (see Fig. 3a). This fact is in accordance with Ref. [32], where the authors found that the expression of the genes which synthesize fatty acids was generally elevated in rich medium. Another important difference is the loss of prominence of Oxidative Phosphorylation and the Pentose Phosphate Pathways.

Next, we consider the set of minimal media given in Ref. [4] (see section D in Methods) where different carbon, nitrogen, phosphorus and sulfur sources are alternated. For each minimal medium, we scan for $\alpha_c$ as in Figs. 1b and c. In Fig. 3b and c, we plot, respectively,
the probability distribution functions of the tuned $\alpha_c$ values and of the fraction of links remaining in the metabolic backbones for all media. We find that there is a characteristic value of these magnitudes with no outliers, meaning that the flux structure is very similar across media in spite of the difference in the composition of nutrients. The presence of these characteristic values of $\alpha_c$ and the retained fraction of links in the metabolic backbones motivates us to merge all of them into a single merged metabolic backbone. The links in this superbackbone correspond to reactions that passed the filter in any of the external media considered and are annotated with a weight that corresponds to the number of media in which the corresponding metabolic backbone contains the link. The histogram of the distribution of these weights is shown in Fig. 3d, characterized by a clear bimodal behavior. One peak corresponds to links being common to all media, and the other corresponds to the most common situation of links specific to a few media.

An analysis of connectedness shows that this superbackbone contains a large connected component and 11 disconnected components. The connected component is composed by a large SCC containing 43% of its nodes, in addition to three small SCCs containing only two nodes each. A pathway composition analysis in the large SCC indicates that, again, we obtain significantly different results from the glucose minimal medium (see Fig. 3a). The most prominent pathway is Alternate Carbon Metabolism, in agreement with Ref. [33], where the authors found that Alternate Carbon Metabolism is related to genes whose expression depends on external stimuli, particularly on alteration of carbon sources. It is also in agreement with results in Ref. [34], where the authors hypothesize that Alternate Carbon Metabolism can adapt to different nutritional environments, and also with results in Ref. [35], where Alternate Carbon Metabolism is found to be an important intermediate pathway in the network of pathways. The second most abundant pathway corresponds to Transport, Inner Membrane, which again is in agreement with Ref. [33] and Ref. [35]. It is a transversal pathway which is in charge of the transport of metabolites between periplasm and cytosol. Finally, if we retain links present at least in 25% of the minimal media, the network fragments into 40 components with the largest one containing five SCCs, which indicates that links with small weight, i.e., links specific for a few media, have an important role in providing global connectivity to the superbackbone.
III. DISCUSSION

Identifying high-flux routes in metabolic networks has been useful in order to, for example, identify principal chains of metabolic transformations [9, 36, 37]. In this work, we go beyond the mere identification of high-flux routes with metabolic pathways. Using a high-flux fluctuation analysis we are able to identify ancestral pathways and, on the other hand, pathways with capabilities to adapt to short-term external changes. At the core of the high-flux fluctuation analysis, we use a filtering tool which needs no a priori assumptions for the connectivity of the filtered subnetworks but that produces reduced versions which are globally connected and retain the characteristic complex features of the original network. This procedure allows us to extract a metabolic backbone which contains all relevant connections given a set of external nutrients, recovering both intra- and inter-pathway connections which can be understood as the superhighways of metabolism. Further, an evolutionary explanation can also be given for this identification of both intra- and inter-pathway connections since the cooperation between reaction inside and outside pathways implies that the overall performance of a cell will be improved due to a better and more efficient utilization of the available resources. This fact reinforces the idea that pathways are not isolated identities performing their tasks independently of others Ref. [35].

As stated in Ref. [38], properties that originate from evolutionary pressure should not be observed in random networks. Due to the fact that the disparity filter identifies links that deviate from a random null model, it allow us to identify those reactions for which evolutionary pressure has had a large incidence. Since we are using FBA flux solutions, in our investigations the effect of evolutionary pressure is understood to favor the maximization of the growth of the organism [39–41]. The evolutionary analysis of the metabolic backbones of the two considered organisms shows that their SCCs are composed by reactions that belong to ancient pathways. In E. coli, we find that each SCC has different and definite metabolic functions. For both organisms, the largest SCC contains pathways related to energy metabolism, meaning that these organisms have evolved towards maximum efficiency in obtaining chemical energy, something very important in case of nutrient scarcity. A smaller SCC is responsible for the synthesis of purines and pyrimidines, vital for DNA / RNA synthesis. The third SCC corresponds to the metabolism of lipids, the most important constituents that compose the cell membrane. Two findings relating the two small
SCCs deserve also special attention. Firstly, the two small SCCs are located in the OUT component of the large SCC. Secondly, as the filter becomes more restrictive, the small SCCs fragment, while the large SCC still maintains a large part of links and nodes. These features could be explained in terms of the functional requirements of the small SCCs. On the one side, they need chemical energy to perform their tasks and, on the other side, they need also basic building blocks. These tasks are performed in the large SCC by, for example, Glycolysis/Gluconeogenesis or the Citric Acid Cycle. Therefore, it suggests that those SCCs were added to the OUT component of the large SCC in later steps of evolution. A simpler organism, *M. pneumoniae*, has no other relevant SCCs apart from energy metabolism, as a result of its parasitism, which has led to the loss of many metabolic functions [17]. More precisely, in *M. pneumoniae* the Citric Acid Cycle and Oxidative Phosphorylation do not take place [17, 42], meaning that it must rely on organic acid fermentation to obtain energy. Moreover, changes in the growth rate greatly affect the fluxes through Glycolysis and Pyruvate Metabolism [17].

The study of the dependence on the environment of the *E. coli* metabolic backbone allows us to identify short-time adaptation capabilities. Regarding rich medium, we observe that the critical value of $\alpha$ is substantially different than the one in glucose minimal medium, suggesting that this enriched medium modifies significantly the flux structure compared to the glucose minimal medium. The bacterium in rich medium displays less active reactions than in glucose minimal medium since, in minimal medium, many reactions must be active in order to synthesize biosynthetic precursors that in the rich medium can be obtained from the environment, in agreement with Ref. [32]. The pathway called Membrane Lipid Metabolism achieves a high relevancy, being the most abundant pathway in the largest SCC of the rich medium metabolic backbone. This happens because the instantaneous response of *E. coli* to this rich medium, which induces a large increase in the growth rate of the organism due to nutrient abundance, is to synthesize as much as membrane lipids as possible, since fast-growing cells must synthesize membrane components more rapidly to satisfy the high lipid demand to generate new cells [32]. The analysis of the adaptation of *E. coli* to 333 different minimal media shows that the distribution of fluxes is practically independent on the composition of the nutrients present in these environments, allowing to extract characteristic features that describe the backbones of the metabolic network independently of the environment. This permits the construction of a merged backbone
that comprises all the links composing the metabolic backbone in each media. This leads to the identification of pathways whose associated reactions are more sensitive to changes in the environment, unveiling Alternate Carbon Metabolism as the pathway with more capabilities to respond to external stimuli, in accordance with existing works [33, 34].

The use of filtering methods usually imply a drastic reduction of the complexity of metabolic maps, which weakens the validity of potentially inferred conclusions. The application of the disparity filter based on a high-flux fluctuation analysis to produce metabolic backbones enables to reduce the system while maintaining all relevant interactions and so it becomes a useful tool to unveil sound biological information. For instance, the investigation of *E. coli* and *M. pneumoniae* revealed metabolic backbones mainly composed of a core of reactions belonging to ancient pathways, for which the effects of evolutionary pressure are higher. At the same time, this approach can be appropriate in applications exploiting its capability to recognize pathways and particular reactions more sensitive to environmental changes, which makes it potentially useful in biotechnology and biomedicine.

IV. METHODS

A. Flux Balance Analysis

Flux Balance Analysis (FBA) [5] is a technique which allows to compute metabolic fluxes without the need of kinetic parameters, just by using constrained-optimization. FBA proceeds by writing the stoichiometric matrix $S$ of the whole network and multiplying it by the vector of fluxes $\nu$. This stoichiometric matrix contains the stoichiometric coefficients of each metabolite in each reaction of the network. This product is then equal to the vector of the time variation of the concentrations $\dot{c} = S \cdot \nu$. Steady-state is assumed, thus $S \cdot \nu = 0$. Since in general metabolic networks contain more reactions than metabolites, we have an underdetermined system of equations. Hence, a biological objective function must be defined in order to have a biologically meaningful solution. In this work, the chosen objective function is the growth rate of the organism, which means that FBA finds the solution that optimizes the growth of the organism, which is equivalent to maximize biomass formation. Reversibility of reactions are also added in order to constrain the solutions. Since we have a linear system of equations with linear constraints, Linear Programming is used in order
to compute a flux solution in a small amount of time (of the order of 1 s), which implies a computationally cheap method.

B. Disparity filter on metabolic networks

The disparity filter \[14\] takes advantage of the local fluctuations present in the weights of the links between nodes. It is useful to define the strength \(s_i\) of a node \(i\), as the sum of the weights \((\nu)\) of the links associated to this node, \(s_i = \sum_j \nu_{ij}\). The filtering method starts by normalizing the weight of the nodes \(p_{ij} = \frac{\nu_{ij}}{s_i}\), where \(\nu_{ij}\) is the weight of a link \(j\) of the node \(i\), since we need a measure of the fluctuations of the weights attached to a node at the local level. The key point is that a few links have a large value of \(p_{ij}\) being thus more significant than the others, as computed by the disparity measure defined as \(\Upsilon_i(k) \equiv k \sum_j p_{ij}^2\), where \(k\) is the degree of the node and \(p_{ij}\) is the normalized weight of the link between node \(i\) and node \(j\).

In our application to metabolic networks, \(\Upsilon_i(k)\) characterizes the level of local heterogeneity of a metabolite \(i\), and so \(p_{ij}\) stands for the normalized weight of the link between metabolite \(i\) and reaction \(j\), with \(\nu_{ij}\) the flux of reaction \(j\). Under perfect homogeneity, when all the links share the same amount of the strength of the node, \(\Upsilon_i(k)\) equals 1 independently of \(k\), whereas for perfect heterogeneity, when one of the links carries the whole strength of the node, \(\Upsilon_i(k)\) equals \(k\). Usually, an intermediate behavior is observed in real systems.

To assess the deviations of the weights of the links, a null model is used which provides the expectation of the disparity measure of a node in a random case. The null hypothesis consists on the fact that the normalized weights that correspond to a certain node are produced by a random assignment coming from a uniform distribution. Notice that, since we work with directed metabolic networks, we have three kinds of links. Bidirectional links are decoupled into incoming and outgoing links, leading to a network where nodes have incoming and outgoing links. Each kind of links are treated independently, each one having its own probability density function. The filter then proceeds by identifying which links must be preserved. To do this, we compute the probability \(\alpha_{ij}\) that a weight \(p_{ij}\) is non-compatible with the null model. This probability is compared to a significance level \(\alpha\), and thus links that carry weights with a probability \(\alpha_{ij} < \alpha\) can be considered non-consistent with the null model and they are considered significant for the metabolite. The probability
\( \alpha_{ij} \) is computed with the expression \( \alpha_{ij}^{\text{in/out}} = (1 - P_{ij}^{\text{in/out}})^{k_{\text{in/out}}} - 1 \). Note that, for nodes with only one incoming or outgoing connection, we use the prescription to preserve those links.

### C. Connected components

A connected component of an undirected network is a subset of the network in which any two nodes are connected by following paths. Nodes in a component do not share connections with nodes belonging to a different component [8].

Directedness in network connections introduces a rich substructure to the connected components. Inside the connected component of a directed network, the so-called bow-tie structure can be found [16]. Bow-tie structures are formed by a strongly connected component (SCC), IN and OUT components, tubes and tendrils. A SCC is a subset of the connected component where nodes are reachable from any other vertex by a directed path. The IN component contains nodes that can access the SCC but not vice versa. The OUT component is formed by nodes that can be reached from the SCC but that cannot return there. A tube is a sequence of nodes that connect the IN and the OUT component without going through the SCC. Tendrils are composed by nodes that have no access to the SCC and are not reachable from it.

### D. Construction of environments in \textit{E. coli}

#### 1. Luria-Bertani Broth

We consider a rich medium called Luria-Bertani Broth. This nutritionally rich medium contains the set of compounds defining the minimal medium [4], \textit{i.e.}, a set of minerals salts and four metabolites representing carbon, nitrogen, phosphorus and sulfur sources, respectively, in addition to the following compounds: amino acids, purines and pyrimidines, biotin, pyridoxine, thiamin, and the nucleotide nicotinamide mononucleotide (see Ref. [31] for specific details).
2. Minimal media

We use the different minimal media defined in Ref. [1]. More precisely, these media contain a set of minerals salts and four extra metabolites representing carbon, nitrogen, phosphorus and sulfur sources respectively [4]. To determine FBA solutions in sources containing carbon, we allow each carbon source to be consumed one at a time, while we fix the sources of nitrogen, phosphorus and sulfur to ammonia, phosphate, and sulfate respectively. In this way, each carbon source determines a different minimal medium. To determine solutions corresponding to the other sources, the same procedure is applied, noting that for these cases, the carbon source is fixed to glucose. 555 media can be constructed using this procedure, 333 of them allowing growth after computing the FBA solution in each media.

[1] A. M. Feist, J. C. M. Scholten, B. Ø. Palsson, F. J. Brockman, and T. Ideker, “Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri,” Mol. Syst. Biol. 2, 2006.0004 (2006).
[2] A. Feist et al., “A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information,” Mol. Syst. Biol. 3, 121 (2007).
[3] J. Schellenberger, J. O. Park, T. C. Conrad, and B. Ø. Palsson, “BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions,” BMC Bioinformatics 11, 213 (2010).
[4] J. D. Orth et al., “A comprehensive genome-scale reconstruction of Escherichia coli metabolism - 2011,” Mol. Syst. Biol. 7, 535 (2011).
[5] J. D. Orth, I. Thiele, and B. Ø. Palsson, “What is Flux Balance Analysis?” Nat. Biotechnol. 28, 245–248 (2010).
[6] S. Gudmundsson and I. Thiele, “Computationally efficient flux variability analysis,” BMC Bioinformatics 11, 489 (2010).
[7] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani, Dynamical Processes on Complex Networks, 1st ed. (Cambridge University Press, New York, NY, USA, 2008).
[8] Mark Newman, *Networks: An Introduction* (Oxford University Press, Inc., New York, NY, USA, 2010).

[9] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A. L. Barabási, “Global Organization of metabolic fluxes in the bacterium *Escherichia coli*,” *Nature* **427**, 839–843 (2004).

[10] H. W. Ma and A. P. Zeng, “The connectivity structure, giant strong component and centrality of metabolic networks,” *Bioinformatics* **19**, 1423–1430 (2003).

[11] A. L. Barabási and Z. N. Oltvai, “Network biology: understanding the cells functional organization,” *Nat. Rev. Genet.* **5**, 101–113 (2004).

[12] O. C. Herfindahl, *Copper Costs and Prices: 1870-1957* (John Hopkins University Press, Baltimore, MD, USA, 1959) pp. 1–260.

[13] A. O. Hirschman, “The paternity of an index,” *American Economic Review* **54**, 761–762 (1964).

[14] M. Á. Serrano, M. Boguñá, and A. Vespignani, “Extracting the multiscale backbone of complex weighted networks,” *Proc. Natl. Acad. Sci. USA* **106**, 6483–6488 (2009).

[15] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752 (2004).

[16] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, S. Stata, A. Tomkins, and J. Wiener, “Graph structure in the web,” *Computer Networks* **33**, 309–320 (2000).

[17] J. A. H. Wodke *et al*., “Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling,” *Mol. Syst. Biol.* **9**, 653 (2013).

[18] M. Á. Serrano P. De Los Rios, “Structural efficiency of percolated landscapes in flow networks,” *PLOS ONE* **3**, e3654 (2008).

[19] M. A. Keller, A. V. Turchyn, and M. Ralser, “Nonenzymatic glycolysis and pentose phosphate pathwaylike reactions in a plausible Archean ocean,” *Mol. Syst. Biol.* **10**, 725 (2014).

[20] E. Meléndez-Hevia, T. G. Waddell, and M. Cascante, “The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution,” *J. Mol. Evol.* **43**, 293–303 (1996).

[21] O. Ebenhöh and R. Heinrich, “Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems,” *B. Math. Biol.* **63**, 21–55 (2001).
[22] R. J. Mailloux, R. Bériaut, J. Lemire, R. Singh, D. R. Chénier, R. D. Hamel, and V. D. Appanna, “The tricarboxylic acid cycle, an ancient metabolic network with a novel twist,” PLoS ONE 2, e690 (2007).

[23] D. A. Fell and A. Wagner, “The small world of metabolism,” Nat. Biotechnol. 18, 1121–1122 (2000).

[24] D. R. Evans and H. I. Guy, “Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway,” J. Biol. Chem. 279, 33035–33038 (2004).

[25] M. W. Powner, B. Gerland, and J. D. Sutherland, “Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions,” Nature 459, 239–242 (2009).

[26] G. Caetano-Anolles, H. S. Kim, and J. E. Mittenthal, “The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture,” Proc. Natl. Acad. Sci. USA 104, 9358–9363 (2007).

[27] S. Suen, H. H. S. Lu, and C. H. Yeang, “Evolution of domain architectures and catalytic functions of enzymes in metabolic systems,” Genome Biol. Evol. 4, 976–993 (2012).

[28] E. Yus et al., “Impact of genome reduction on bacterial metabolism and its regulation,” Science 326, 1263–1268 (2009).

[29] M. Tadege, I. Dupuis, and C. Kuhlemeier, “Ethanolic fermentation: new functions for an old pathway,” Trends Plant Sci. 4, 320–325 (1999).

[30] G. SezanoV, D. Joseleau-Petit, and R. D’Ari, “Escherichia coli physiology in Luria-Bertani broth,” J. Bacteriol. 189, 8746–8749 (2007).

[31] O. Güell, F. Sagués, and M. Á. Serrano, “Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis,” PLoS Comput. Bio. 10, e1003637 (2014).

[32] H. Tao, C. Bausch, C. Richmond, F. R. Blattner, and T. Conway, “Functional genomics: expression analysis of Escherichia coli growing on minimal and rich media,” J. Bacteriol. 181, 6425–6440 (1999).

[33] A. Lourenço, S. Carneiro, J. P. Pinto, M. Rocha, E. C. Ferreira, and I. Rocha, “A study of the short and long-term regulation of E. coli metabolic pathways,” J. Integr. Bioinform. 8, 183 (2011).

[34] J. M. Monk et al., “Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments,” Proc. Natl. Acad. Sci. USA 110, 20338–20343 (2013).
[35] M. Á. Serrano, M. Boguñá, and F. Sagués, “Uncovering the hidden geometry behind metabolic networks,” Mol. BioSyst. 8, 843–850 (2012).

[36] R. Bourqui et al., “Metabolic network visualization eliminating node redundancy and preserving metabolic pathways,” BMC Syst. Biol. 1, 29 (2009).

[37] K. Faust, P. Dupont, J. Callut, and J. van Helden, “Pathway discovery in metabolic networks by subgraph extraction,” Bioinformatics 26, 1211–1218 (2010).

[38] G. Basler, S. Grimbs, O. Ebenhöh, J. Selbig, and Z. Nikoloski, “Evolutionary significance of metabolic network properties,” J. R. Soc. Interface, rsif20110652 (2011).

[39] R. U. Ibarra, J. S. Edwards, and B. Ø. Palsson, “Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth,” Nature 420, 186–189 (2002).

[40] L. M. Blank, L. Kuepfer, and U. Sauer, “Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast,” Genome Biology 6, R49 (2005).

[41] F. Llaneras and J. Picó, “Stoichiometric modelling of cell metabolism,” Journal of Bioscience and Bioengineering 105, 1–11 (2008).

[42] J. T. Manolukas, M. F. Barile, D. K. Chandler, and J. D. Pollack, “Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in mollicutes,” J. Gen. Microbiol. 134, 791–800 (1988).