FEBS Letters

FEBS PRESS science publishing by scientists

Detecting the Significant Flux Backbone of *Escherichia coli* metabolism

Oriol Güell[1], Francesc Sagués[1] and M. Ángeles Serrano[2,3,4]

1 Departament de Ciència dels Materials i Química Física, Universitat de Barcelona, Spain
2 Department de Física de la Matèria Condensada, Universitat de Barcelona, Spain
3 University of Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Spain
4 ICREA, Barcelona, Spain

Correspondence
O. Güell, Departament de Ciència dels Materials i Química Física, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
Fax: 934 021 231
Tel: 934 020 138
E-mail: oguell@ub.edu

(Received 18 October 2016, revised 20 March 2017, accepted 1 April 2017, available online 2 May 2017)

doi:10.1002/1873-3468.12650

Edited by Alfonso Valencia

[Correction added after online publication on 2 May 2017: Data Accessibility section added].

The heterogeneity of computationally predicted reaction fluxes in metabolic networks within a single flux state can be exploited to detect their significant flux backbone. Here, we disclose the backbone of *Escherichia coli*, and compare it with the backbones of other bacteria. We find that, in general, the core of the backbones is mainly composed of reactions in energy metabolism corresponding to ancient pathways. In *E. coli*, the synthesis of nucleotides and the metabolism of lipids form smaller cores which rely critically on energy metabolism. Moreover, the consideration of different media leads to the identification of pathways sensitive to environmental changes. The metabolic backbone of an organism is thus useful to trace simultaneously both its evolution and adaptation fingerprints.

Keywords: disparity filter; flux balance analysis; metabolic backbones; metabolic networks

High-quality genome-scale metabolic reconstructions are composed of thousands of reactions and metabolites [1–4]. Due to their complexity, the analysis of these metabolic reconstructions requires computational approaches, like constraint-based optimization techniques [5,6], and methodological frameworks, like complex network science [7,8], to elucidate features of their functional organization and pathway structure. Some of the tools used with this purpose are Elementary Flux Modes [9], Extreme Pathways [10], Minimal Metabolic Behaviors [11], and Minimal Pathway Structures [12], which are based on finding feasible subnetworks to relate them with definite functions. A different approach in this endeavor is provided by the concept of backbone. Backbones maintain significant information while displaying a substantially decreased number of interconnections and, hence, can provide accurate but reduced versions of the whole system. In this direction, the work by Almaas *et al.* [13] introduced a filtering technique that selects the reactions dominating the production and consumption of each metabolite and connects two metabolites if the reaction producing one of them with the highest flux happens to be the reaction consuming the other with the highest flux, which defines a high-flux backbone. This method is able to segregate classical pathways, but the selected high-flux subgraphs present a

Abbreviations
ADP, adenosine diphosphate; ATP, adenosine triphosphate; FAD^+, flavin adenine dinucleotide; FBA, Flux Balance Analysis; LB, Luria–Bertani; NAD^+, nicotinamide adenine dinucleotide; NADPH, nicotinamide adenine dinucleotide phosphate; SCC, strongly connected component; SFB, Significant Flux Backbone; SSFB, Super Significant Flux Backbone.

linear structure with very little interconnectivity and so they necessarily lack the characteristic complex features of real metabolic networks [14–16].

Filtering approaches have also interested researchers working on networks in a more general context. A filtering method for weighted networks based on the disparity measure [17,18] was developed in Ref. [19]. This approach exploits the heterogeneity present in the intensity of interactions (weights) in real networks, both at the global and local levels [20], to extract the dominant set of connections for each element. Typically, the obtained disparity backbones 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, disparity backbones preserve the heterogeneity of the degree distribution, the level of clustering, and the bow-tie structure [21], and other complex features of the original networks [19].

In this work, we use Flux Balance Analysis (FBA) [5] maximizing the biomass production rate to determine reaction fluxes in the metabolic network of *Escherichia coli* (*E. coli*) *i*JO1366 [4] and other microorganisms for comparison. Afterwards, we use the disparity filter [19] to extract its Significant Flux Backbone (SFB) as a one-mode projection in the space of metabolites. In contrast to the filtered linear structures from Ref. [13], SFBs obtained using the disparity filter conserve not only high-flux reactions but also many low-flux reactions—provided that they are significant for the production or consumption of a certain metabolite—, such that the complexity of the original networks is preserved in the backbone. We investigate the obtained SFB in glucose minimal medium for fingerprints of evolution and environmental adaptation finding that its central core is mainly composed of evolutionary conserved reactions in energy metabolism whose fluxes still retain at present a key role in the evolved organisms. This feature is also observed in the central core reactions of the SFB of other bacteria. In *E. coli*, the analysis of the SFB reveals that the synthesis of nucleotides and the metabolism of lipids form smaller cores which rely critically on energy metabolism, but not conversely. We also study how the structure of the SFBs in *E. coli* depends on the composition of the medium, which allows us to identify pathways that are more sensitive to environmental changes and nutrient availability. This leads us to define a Super Significant Flux Backbone (SSFB) by merging the SFBs for the different environments. Finally, we find that the SFB contains most of the highest fluxes in the metabolic network and that the SSFB is enriched with genes in a core proteome of metabolism and gene expression across conditions.

Materials and methods

We use FBA to compute the fluxes of the reactions in a metabolic network which maximize the biomass production rate of the organism. These fluxes are treated as weights by the disparity filter. The large-scale connectivity structure of the obtained SFB is analyzed in terms of connected components and additional media are considered to analyze the environmental sensitivity of the backbone composition. Finally, the composition of the SFB of *E. coli* is compared with that of the high-flux backbone defined in [13] and with the core proteome given in [22].

Flux Balance Analysis

Flux Balance Analysis [5] is a technique which allows to compute metabolic fluxes without the need of kinetic parameters, just by using constrained-optimization. Flux Balance Analysis proceeds by writing the stoichiometric matrix S of the whole network and multiplying it by the vector of fluxes ν. 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 biomass yield of the organism, which means that FBA finds the solution that optimizes the biomass production rate of the organism, which is equivalent to maximize biomass formation. Reversibility of reactions is 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.

We implement FBA using the GNU Linear Programming Kit (GLPK). FBA calculations are performed on four bacteria—*E. coli* *i*JO1366 [4], *Helicobacter pylori* (*H. pylori*) *i*IT341 [23], *Staphylococcus aureus* (*S. aureus*) *i*SB619 [24], and *Mycoplasma pneumoniae* (*M. pneumoniae*) *i*JW145 [25] (we also provide results for *Mycobacterium tuberculosis* (*M. tuberculosis*) *i*NJ661 [26], *Saccharomyces cerevisiae* (*S. cerevisiae*) *i*ND750 [27], and *Methanosarcina barkeri* (*M. barkeri*) *i*AF692 [1] in Supporting Information). The models (see Table S1 for number of metabolites and reactions in the genome-scale reconstructions of the organisms) include the biomass reaction, auxiliary reactions such as exchange or sink reactions, and all cellular compartments which are taken into account. We represent these metabolic networks as bipartite graphs with two kinds of nodes, metabolites, and reactions, and with links containing directionality, which leads to incoming, outgoing, and bidirectional links. For *E. coli*, FBA calculations are performed in a glucose minimal medium with

a maximum uptake of glucose limited to 10 mmol·gDW^{-1}·h^{-1} [4], whereas for *M. pneumoniae*, FBA computations are carried in a defined medium with a maximum glucose uptake of 7.37 mmol·gDW^{-1}·h^{-1} and a supply of D-ribose to simulate the availability of ribosylated bases [25]. For *H. pylori*, we used the minimal medium provided in Ref. [23] needed by *i*IT341 to fulfill the biomass requirement. In the case of *S. aureus*, the calculation are performed in a glucose minimal medium [24].

Disparity filter on metabolic networks

The disparity filter [19] takes advantage of the local heterogeneity present in the fluxes of reactions associated with a given metabolite. The filter is able to retain those fluxes which are significant, meaning that their value is unexpectedly high. Notice that, since we work with directed metabolic networks, we have three kinds of links: incoming, outgoing and bidirectional links. In the filtering procedure, we treat incoming and outgoing connections separately, and so bidirectional links are disassociated into incoming and outgoing links. The filtering method starts by normalizing the k_i^{in} fluxes v_{ij}^{in} associated with the reactions which produce a certain metabolite i, $p_{ij}^{\text{in}} = \frac{v_{ij}^{\text{in}}}{\sum_{j=1,k_i^{\text{in}}} v_{ij}^{\text{in}}}$, and the k_i^{out} fluxes v_{ij}^{out} associated with the reactions which consume metabolite i, $p_{ij}^{\text{out}} = \frac{v_{ij}^{\text{out}}}{\sum_{j=1,k_i^{\text{out}}} v_{ij}^{\text{out}}}$. The key point is that a few incoming links carry a significant value of p_{ij}^{in} and a few outgoing links carry a significant value of p_{ij}^{out}. We characterize the disparity in the local distribution of incoming and outgoing fluxes around i with the disparity measures [17,18]

$$\Upsilon_i(k_i^{\text{in}}) \equiv k_i^{\text{in}} \sum_j (p_{ij}^{\text{in}})^2$$
$$\Upsilon_i(k_i^{\text{out}}) \equiv k_i^{\text{out}} \sum_j (p_{ij}^{\text{out}})^2.$$

Under perfect homogeneity, when all the incoming links share the same amount of the total flux of reactions producing metabolite i, $\Upsilon_i(k_i^{\text{in}}) = 1$ independently of k_i^{in}, whereas for perfect heterogeneity, when one of the links carries all the flux, $\Upsilon_i(k_i^{\text{in}}) = k_i^{in}$. The same is for $\Upsilon_i(k_i^{\text{out}})$. Usually, an intermediate behavior is observed in real systems.

To assess the relevance of the fluxes of reactions associated with a given metabolite, a null hypothesis is used which provides the expectation of the disparity measure when their total flux is distributed at random according to a uniform distribution. The filter then proceeds by identifying which fluxes must be preserved. To do this, we compute the probability α_{ij} that a normalized flux p_{ij} is noncompatible with the null hypothesis. This probability is a P-value which is compared with a significance level α, and thus links that carry fluxes with a probability $\alpha_{ij} < \alpha$ can be considered nonconsistent with the null model and so significant for the metabolite. The null model is applied separately to producing and consuming reactions and the probabilities α_{ij} for incoming and outgoing connections are computed with the expressions (see Ref. [19]).

$$\alpha_{ij}^{\text{in}} = (1 - p_{ij}^{\text{in}})^{k_i^{\text{in}}-1}$$
$$\alpha_{ij}^{\text{out}} = (1 - p_{ij}^{\text{out}})^{k_i^{\text{out}}-1}.$$

Once the significant fluxes have been selected, we construct the significant flux backbone by placing a directed link between two metabolites if there is a reaction whose flux is simultaneously significant for the consumption of one metabolite and for the production of the other, see Fig. 1 for an illustration of the filtering process. The filter cannot decide on nodes with just one connection. According to our maximization-of-nodes-minimization-of-links principle, we use the prescription to preserve the reaction associated with metabolites with only one incoming or outgoing connection.

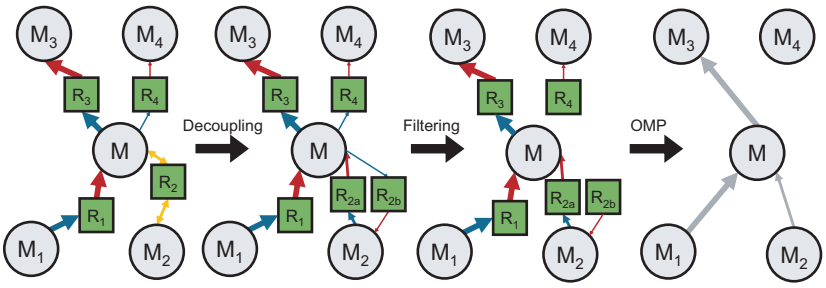

Fig. 1. Disparity filter. Gray circles correspond to metabolites and green squares denote reactions. Incoming connections to metabolites are represented by red arrows, outgoing connections with blue arrows, and bidirectional connections with dark yellow arrows. The width of the arrows is proportional to the value of the fluxes, i.e., the larger the flux, the wider the arrow. After the Filtering step, only links with an unexpectedly high flux remain in the filtered bipartite subnetwork. OMP denotes One-Mode Projection and the links in the final SFB appear in grey.

Connected components

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

Directedness in network connections introduces a rich substructure in the connected components. Inside the connected component of a directed network, the so-called bow-tie structure emerges [21]. 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 any node is reachable from any other 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.

Construction of environments in *Escherichia coli*

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], i.e., a set of minerals salts and four metabolites representing carbon, nitrogen, phosphorus, and sulfur sources, in addition to the following compounds: amino acids, purines and pyrimidines, biotin, pyridoxine, thiamin, and the nucleotide nicotinamide mononucleotide (see Ref. [28] for specific details).

Minimal media

We use the different minimal media defined in Ref. [4]. More precisely, these media contain a set of minerals salts and four extra metabolites representing carbon, nitrogen, phosphorus, and sulfur sources [4]. To determine FBA solutions in different media we change the carbon source while we fix the sources of nitrogen, phosphorus and sulfur to the standard metabolites in each class, which are ammonia, phosphate, and sulfate, respectively. In this way, each carbon source determines a different minimal medium. In the same way, other minimal media are constructed using the same procedure of changing the nitrogen, phosphorus, or sulfur source while keeping the standard metabolite unchanged for the rest of categories; note that, in these cases, the standard carbon source is fixed to glucose. Five hundred and fifty-five media can be constructed using this procedure, 333 of them allowing growth *in silico*.

Results

An important feature of the flux solutions obtained using FBA is that they capture the heterogeneity of the flux distribution within a single flux state [13]. The probability distribution function of the obtained FBA fluxes, insets in Fig. 2 Top, is characterized by a broad scale distribution of values. We disregarded zero-flux reactions, such that the set of active reactions and metabolites is markedly reduced as compared to their number when all reactions in the genome-scale reconstruction are considered, see Table S1.

Identification of the Significant Flux Backbone (SFB) of the metabolic networks

Besides the broad diversity of FBA fluxes at the global scale of the network, heterogeneity is also present at the local level in the fluxes of reactions associated with the production and consumption of a given metabolite. We calculate the disparity measure for every metabolite i participating in k producing or consuming reactions [13,19] as $\Upsilon_i(k) = k\sum_{j=1}^{k} p_{ij}^2$, where p_{ij} is the flux of reaction j normalized by the total flux of reactions consuming or producing metabolite i, $p_{ij} = v_{ij}/\sum_{j=1}^{k} v_{ij}$, (see Materials and methods). We treat separately producing and consuming reactions. Figure 2 Top displays the disparity values for all metabolites as a function of their number of producing and consuming reactions (incoming and outgoing degree) in *E. coli*, *S. aureus*, *H. pylori*, and *M. pneumoniae* (see Fig. S1 in Supporting Information for more organisms). The shadowed areas correspond to disparity values compatible with the null hypothesis that the total flux of incoming or outgoing reactions of a metabolite is uniformly distributed at random among them. The null hypothesis helps to discount local heterogeneities produced by random fluctuations (see caption of Fig. 2). As shown, most metabolites in the different organisms present disparity values that cannot be explained by the null hypothesis, meaning that the local distribution of the fluxes of reactions associated with metabolites is significantly heterogeneous. We conclude then that the disparity filter will be able to efficiently extract the backbone with the most significant connections for the organisms, while preserving the characteristic features of metabolism as a complex network. Notice that significant fluxes are those with values much above the average expectation given by the null hypothesis and that, in absolute terms, they can be high or low (see Materials and methods).

The disparity filter preserves a reaction in the backbone if the probability α_{ij} that its normalized flux p_{ij} is

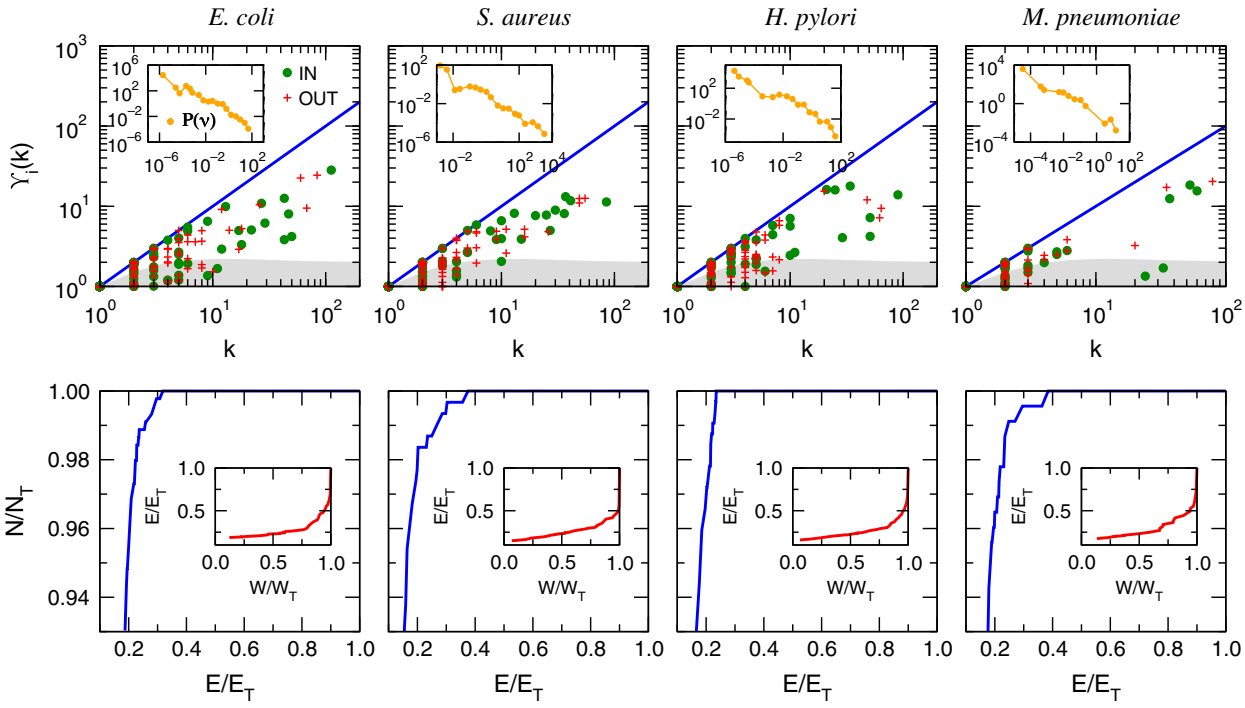

Fig. 2. Heterogeneity of reaction fluxes and statistics of the disparity filter in different bacteria. Top. Disparity measure for fluxes associated with metabolites as a function of their degree k. Each green point/red plus symbol corresponds to the incoming /outgoing fluxes of reactions associated with a different metabolite. The shadowed area corresponds to the average plus 2 standard deviations given by the null hypothesis, meaning that points which lie outside this area can be considered to present a heterogeneous distribution [19]. Note that there may exist an overlap between points for each value of k. Inset: global distribution of fluxes in the FBA solution. Bottom. Fraction of metabolites as a function of the fraction of links in the filtered networks. Inset: remaining fraction of total flux as a function of the fraction of links in the filtered networks. Notice that the curves come from a parametric representation of the different fractions as a function of α.

compatible with the null hypothesis (P-value) is smaller than a chosen significance threshold α, which determines the filtering intensity. For each metabolite i, we compute the P-value α_{ij} for each producing and consuming reaction j and compare the obtained P-value with the significance level α. In this way, all the reaction with fluxes which are significant for the production or consumption of a metabolite can be selected, in contrast to the approach in [13] where only a single most significant flux was selected so that the obtained subgraphs presented an obvious linear structure. The disparity filter can be adjusted by tuning the critical threshold to observe how the metabolic networks of the bacteria are reduced as we decrease α from 1 to 0, $\alpha = 1$ meaning the complete genome-scale reconstruction. Notice that, after applying the filter with a specific value of α, we obtain a bipartite representation of the metabolic backbone. To avoid working with stoichiometrically nonbalanced reactions, we transform the filtered bipartite representation into a one-mode projection of metabolites obtained by placing a directed link between two metabolites if there is a reaction whose flux is simultaneously significant for the consumption of one metabolite and for the production of the other [19] (see Fig. 1).

Next, we compute the links E, metabolites N and total flux W remaining in the one-mode filtered networks as a function of the significance level α. These magnitudes are normalized by dividing them by the corresponding values in the original network, E_T, N_T, and W_T. In Fig. 2 Bottom, we show N/N_T vs E/E_T, and W/W_T vs E/E_T in the insets, for the one-mode projections of the filtered networks for the four bacteria (see Fig. S1 in Supporting Information for more organisms). While the filter can reduce considerably the fraction of links, the corresponding fraction of metabolites is maintained at almost the original value. The number of nodes and the total flux in the backbone only starts to drop appreciably after more than 50% of the links are removed. We take the critical value α_c as the point where the fraction of metabolites starts to decay. This critical value can be seen as an optimal point which greatly reduces the number of links in the network preserving, at the same time, most nodes and therefore much biochemical and structural information as possible. The values for the four

bacteria are reported in Table S1 of Supporting Information. Notice that lowest value of α_c corresponds to *H. pylori*, due to a stronger heterogeneity in the local distribution of fluxes, which allows to reduce further the fraction of links while preserving all the nodes. The backbone of a metabolic network corresponds to the filtered one-mode projection graph using the critical value α_c as the *P*-value threshold. The backbone unveils the pivotal pathways of fluxes in metabolism and at the same time preserves the connectivity structure, the heterogeneity of the degree distribution, and the high level of clustering typical of complex networks.

Large-scale structure and pathway composition of the SFBs

We filter the metabolic networks of *E. coli*, *S. aureus*, *H. pylori*, and *M. pneumoniae*, taking into account only active reactions in glucose minimal medium and using the identified critical values α_c for the significance level. The bipartite subnetwork filtered in each organism retains almost all active reactions in the FBA solution, more than 90% in all bacteria, and by definition all metabolites in active reactions, see Table S1 in Supporting Information. Each filtered subnetwork stands as a bipartite representation with stoichiometrically nonbalanced reactions. To produce the corresponding SFB, we generate for each of them a one-mode projection of metabolites placing a directed link between two metabolites if there is a reaction connecting them in the subnetwork obtained with the disparity filter. In the one-mode projection, we remove metabolites without any connection. The SFB of *E. coli* includes, for instance, 436 metabolites linked by 328 reactions as compared to the 445 metabolites linked by 404 reactions in the bipartite subnetwork before the one-mode projection. For the four bacteria, almost all metabolites in active reactions and a very large fraction of active reactions are preserved in the SFB of each organism, see Table S1 in Supporting Information for the specific values.

Next, we analyze in more detail the structure of the resulting SFBs in terms of connectedness. Metabolic networks have been found to display typical large-scale connectivity patterns of directed complex networks, characterized by a bow-tie structure [21] (see Materials and 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 [14,29]. This is the case also for the genome-scale reconstructions of the bacteria analyzed in this work, whose SCCs contain the largest part of metabolites and whose IN and OUT components are formed, respectively, by nutrients and waste metabolites. Metabolites in the SFB of *E. coli* are arranged in a large connected component of 178 metabolites and 41 disconnected small components. Three different SCCs can be identified in the largest connected component of the SFB, each one with 25%, 10%, and 6% of the metabolites (see Fig. 3 Top). The two smallest SCCs are in the OUT component of the largest SCC. Central compounds of metabolism are identified in these SCCs: protons, water, adenosine triphosphate (ATP), L-glutamate, phosphate, nicotinamide adenine dinucleotide (NAD^+), diphosphate, adenosine diphosphate (ADP) and flavin adenine dinucleotide (FAD^+). These metabolites are highly connected even in the backbone, highlighting the ability of the disparity filter to preserve the same structural features of the complete metabolic network while in a reduced version. However, a comparison of the hubs in the backbone with the hubs in the nonfiltered metabolic network, i.e., considering $\alpha = 1$, shows some deviations. The most important difference happens for L-glutamate. It occupies the 4th position in the ranking of the most connected metabolites, whereas it is located in the 12th position in the nonfiltered network. Besides, the fraction of reactions producing L-glutamate in relation to the total number of reactions in which this metabolite participates is 0.27 in the backbone vs 0.65 in the nonfiltered network. Regarding other metabolites, which in this case cannot be considered as a hub due to its smaller degree, we find that FAD^+ ranks 12th in the backbone, while its position for the nonfiltered case is 22nd. Recall that the number of metabolites in the backbone and in the nonfiltered network is the same. Another interesting case concerns the metabolite glucose 6-phosphate. It contains no outgoing connections in the backbone, whereas its outgoing degree in the nonfiltered network is four.

Since links in the metabolic backbone denote reactions, it is interesting to assess the composition of the backbone of *E. coli* in terms of pathways. First, we start by computing the percentages of pathway participation in the nonfiltered metabolic network disregarding zero-flux reactions. The five pathways with more reactions are Cofactor and Prosthetic Group Biosynthesis (24%), Membrane Lipid Metabolism (7%), Exchange (6%), Cell Envelope Biosynthesis (5%), and Purine and Pyrimidine Biosynthesis (5%). The same analysis in the SFB gives a different composition, with pathways such as Oxidative Phosphorylation and the Citric Acid Cycle gaining weight, as shown in Fig. 3 Bottom and Table S3 in Supporting Information. This

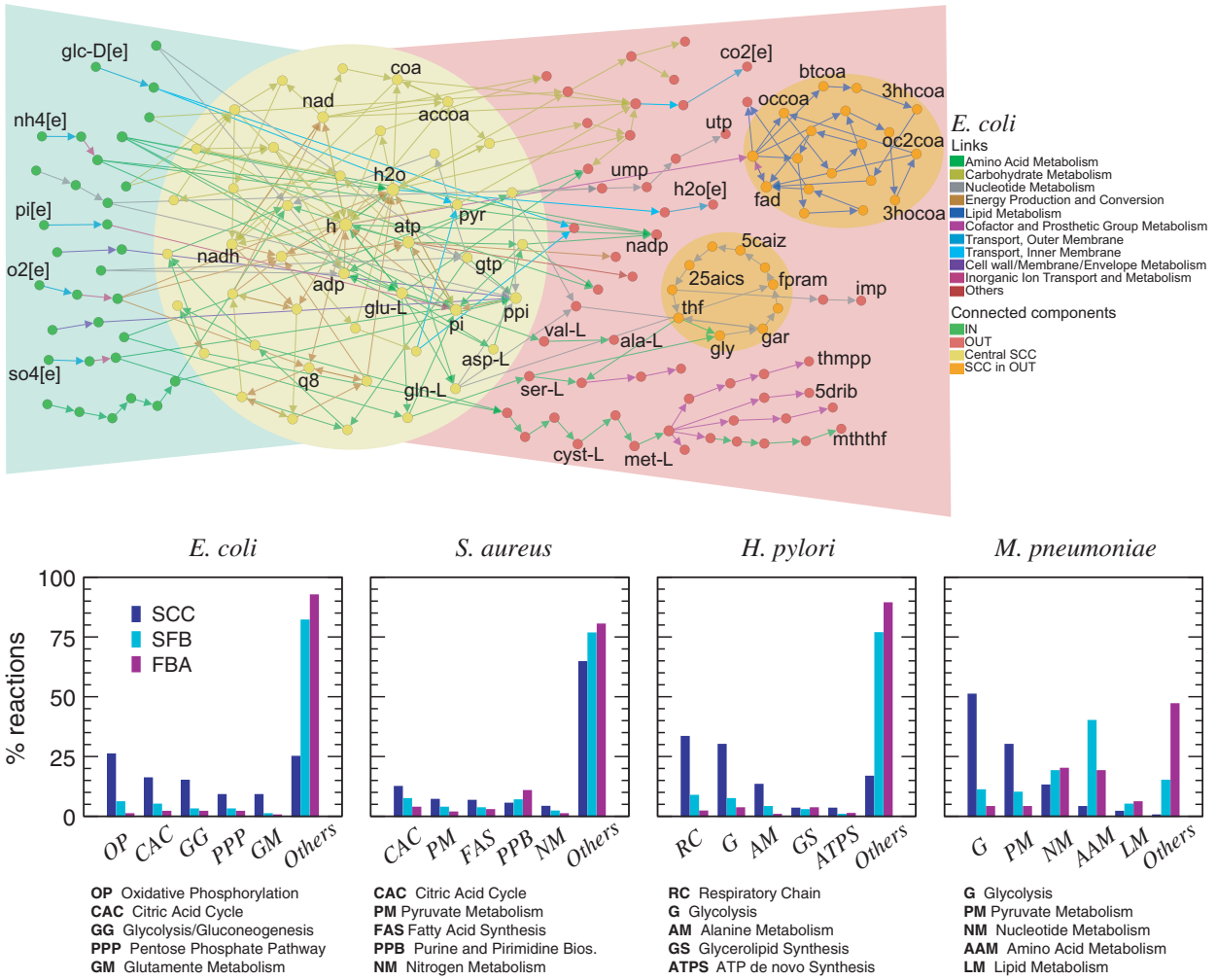

Fig. 3. Structure of the largest connected component in the SFB of Escherichia coli and pathway composition for different bacteria. Top. Largest connected component in the metabolic backbone of *E. coli*. The colors of the metabolites depend on the component each node belongs to. The color of the links, 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. Bottom. Percentage of links in pathways for the largest SCC in the SFB, for the SFB as a whole, and for the nonfiltered metabolic network disregarding zero-flux reactions (numerical values are given in Supporting Information, Tables S3–S6). Notice that the list of acronyms under each graph applies only to the corresponding organism.

is in fact due to the pathway participation in the largest SCC, where the major contribution comes from these two pathways along with Glycolysis/Gluconeogenesis and Pentose Phosphate Pathway. It has been found that Glycolysis (G) and Pentose Phosphate Pathway can take place without the need of enzymes [30]. Concerning the Citric Acid Cycle, it is an ancient pathway that has evolved in order to achieve maximum ATP efficiency [31] by being coupled to Oxidative Phosphorylation and Glycolysis. In addition, this coupling helps the organism to decrease their quantity of reactive oxygen species by modulation of their participating metabolites [32], conforming in this way one of the central pillars of carbon metabolism and energy production. 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 [33]. The second largest SCC contains links that belong mainly to Membrane Lipid Metabolism (97%) and Cofactor and Prosthetic Group Biosynthesis (3%). Membrane Lipid Metabolism supplies the necessary lipids to generate the cell membrane needing the participation of the cofactor $FAD^+/FADH_2$. It has been shown that the pathways involved in lipid metabolism exhibit differences between different lineages in organisms [34],

whereas pathways related to central metabolism are more conserved [34]. Finally, pathways related to the smallest SCC are Purine and Pyrimidine Biosynthesis (91%). Purines and pyrimidines make ribonucleotides deoxyribonucleotides, which serve as activated precursors of RNA and DNA, etc. [35,36]. It has been found that the synthesis of purines and pyrimidines was the first pathway involving enzyme-based metabolism [37]. Interestingly, the other contribution to this SCC is Glycine and Serine Metabolism. Glycine is a precursor of purines and pyrimidines.

When considering α values smaller than the critical one, implying that the filter is more restrictive, we observe that the smallest SCCs disappear from the more stringent sub-backbone. More precisely, it happens for a value of $\alpha = 0.19$. Decreasing even more the significance level to $\alpha = 0.15$, the second largest SCC containing reactions in the Purine and Pyrimidine Biosynthesis pathway still retains 30% of the nodes in the backbone, whereas the largest SCC still contains a 86%. At a value of $\alpha = 0.14$, the second SCC finally disappears and the sub-backbone only remains a single SCC, still preserving 82% of the nodes in the backbone. Hence, the SCC containing links belonging to pathways related to energy metabolism shows a large resistance to get fragmented, even though the filter becomes progressively more and more restrictive, which points to increased levels of local flux heterogeneities and to flux dependencies for the consumption and production of metabolites.

We performed the same analysis in the SFBs of *S. aureus*, *H. pylori*, and *M. pneumoniae*, see Table S2 in Supporting Information for a summary statistics of connected components, and Fig. 3 Bottom and Tables S4–S6 in Supporting Information for pathway composition values. See also Fig. S2 in Supporting Information for a display of the structure of the largest connected component in *M. pneumoniae*. In all cases, the composition of the nonfiltered metabolic networks (disregarding zero-flux reactions) is consistent with the composition of the corresponding SFBs, as in *E. coli*. Another common feature is that the main SCC of the largest connected component is enriched in links related with energy metabolism. The dominant pathways in the SCCs are again the Citric Acid Cycle, Glycolysis and Pyruvate Metabolism and Respiratory Chain. As mentioned above, the CAC is used by aerobic organisms to release stored energy through oxidation processes and its central role in metabolism postulates it as one of the earliest components. Both Glycolysis and Pyruvate Metabolism are pathways that were even present in the earliest stages of life [38], when no oxygen was present in the early atmosphere.

The SFB of *Escherichia coli* encodes potential adaptation capabilities

In the previous section, we analyzed the pathway composition of the significant flux backbone of *E. coli* in glucose minimal medium and found that pathways in the central core are related to energy metabolism. The analysis of the backbone also revealed that the synthesis of nucleotides and the metabolism of lipids form smaller cores which rely critically on energy metabolism; but not conversely. In this section, we study how changes in the environment modify the SFB of *E. coli* in relation to the one obtained in glucose minimal medium, which exposes potential adaptation capabilities.

First, we calculate the FBA fluxes that maximize the biomass production rate of *E. coli* in the rich medium Luria–Bertani (LB) Broth [28,39] (see Materials and methods). Afterwards, we apply the disparity filter to extract the SFB in this new environment, that is obtained for a significance critical 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 188 metabolites and 60 small disconnected components. The connected component contains also three SCCs. However, two of them are tiny with only two metabolites, whereas the largest one encloses 34% of the metabolites in the connected component. Interestingly, the pathway contributing more reactions to this large SCC is Membrane Lipid Metabolism (see Fig. 4A). This fact is in accordance with Ref. [40], 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 relative loss of prominence of Oxidative Phosphorylation and the Pentose Phosphate Pathway. This might seem surprising since the Pentose Phosphate Pathway is typically the main source of nicotinamide adenine dinucleotide phosphate (NADPH). However, in rich medium the functionally significant production of this metabolite takes place in the Citric Acid Cycle pathway [41]. This also evidences the importance of the Citric Acid Cycle to produce NADPH, and so its importance in the synthesis of membrane lipids. Nevertheless, links associated with both Oxidative Phosphorylation and Pentose Phosphate Pathway are also present in the backbone, located outside the SCCs.

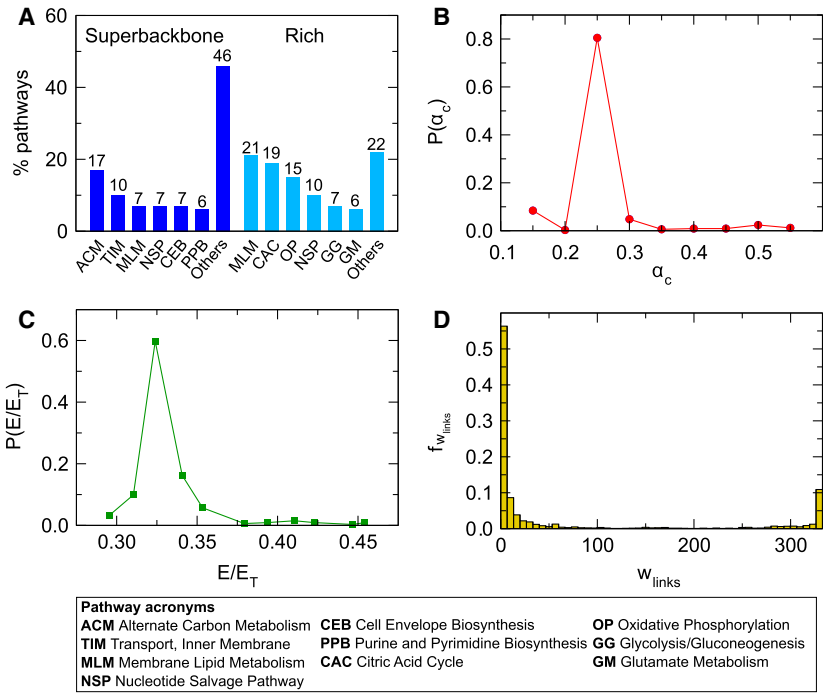

Fig. 4. Dependence of the SFB of *Escherichia coli* with the composition of the environment. (A) Histogram of the fraction of links belonging to each pathway (*x* axis) for the largest SCC of the backbone in the 333 minimal media (blue) and in the rich medium (sky blue). (B) Probability distribution function of α_c for all minimal media. (C) Probability distribution function of the fraction of links in the SFBs for all minimal media. (D) Histogram of relative frequency of weights of links in the SSFB, where the weight of the links counts the number of media in which the corresponding SFB contains the link.

Next, we consider the set of minimal media given in Ref. [4] (see Materials and methods) where different carbon, nitrogen, phosphorus, and sulfur sources are alternated. For each minimal medium, we scan for α_c. In Fig. 4B,C we plot, respectively, the probability distribution function of the collection of α_c values and of the fraction of links remaining in the metabolic backbone for all media. We find that these magnitudes present a characteristic value, 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 value of α_c and the stability of the retained fraction of links in the SFB in the different media motivated us to combine all of them into a single merged metabolic backbone, obtaining a network which we call SSFB. The links in this SSFB 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. 4D, 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 the SSFB contains a large connected component and 11 disconnected components. The connected component with 1090 metabolites is composed by a large SCC containing 43% of its metabolites, 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. 3 Bottom and Fig. 4A). The most prominent pathway is Alternate Carbon Metabolism, in agreement with Ref. [42], 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. [43], where the authors hypothesized that Alternate Carbon Metabolism can adapt to different nutritional environments, and also with results in Ref. [16], 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. [42] and Ref. [16]. This pathway 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 order to prevent the fragmentation of the superbackbone.

Consistency of the SFB of *Escherichia coli*

We compared our SFB definition against two very different approaches which also use genome-scale models to identify the reactions or genes which dominate the functioning of metabolism.

The high-flux backbone of a metabolic network is a one-mode projection in the space of metabolites of a metabolic network for which the reaction fluxes are known. Typically, fluxes are computed using FBA and two metabolites are connected if the reaction producing one of them with the highest flux happens to be the reaction consuming the other with the highest flux [13]. The resulting backbones present a linear structure with very little interconnectivity and so they necessarily lack the characteristic complex features of real metabolic networks. Nevertheless, high-flux backbone contains information about the reaction that dominates the production or consumption of each metabolite in active reactions and, given the local heterogeneity in the distribution of the fluxes consuming or producing metabolites, it is mostly contained in the SFB. We calculated the high-flux backbone of *E. coli* in glucose minimal medium and found that it is formed of 362 metabolites connected by 237 reactions mediating 390 links, of which 355 metabolites (98%), 225 reactions (95%) and 297 links (76%) are also in the SFB. However, we observe that the SFB is not merely a parsimonious relaxation of the high-flux backbone. The SFB includes not only most of the highest fluxes but also other significant (maybe small) fluxes and related metabolites (18% of its metabolites, 69% of its reactions, and 45% of its links are exclusive to it), which leads to a greater chance of more metabolites being connected than in the high-flux backbone but still using a reduced set of connecting links.

A different approach defined the model-based core proteome of *i*OL1554-ME, a metabolism-expression model for *E. coli*, as the minimal set of genes that are consistently used across a large number of environmental conditions [22]. This core proteome is based on a genome-scale model that accounts for metabolism and gene expression under the 333 different environments used in the previous subsection and described in Methods. Genes in the core proteome are those expressed for optimal growth across all simulation conditions and it is defined to be a list of 356 genes, many of them with enzymatic function. We compared the core proteome with the list of genes associated with reactions in the SSFB of *E. coli i*JO1366. The number of genes associated with reactions in the metabolic network of *E. coli i*JO1366 is 1366 and 198 of them overlap with the core proteome. Of the common set, 164 (83%) are contained in the SSFB, 19% of the 842 different genes associated with reactions in the SSFB. We obtain similar percentages if we take the most conservative criterium of only considering the genes or protein complexes of reactions in the SSFB without alternative isoenzymes. In this case, we find that the metabolic network has 974 associated genes and 165 overlap with the core proteome, of which 131 (79%) are present in the SSFB, 23% of the 568 different genes associated with reactions in the SSFB. In conclusion, we found that the SSFB is enriched with genes of the metabolic network in the core proteome.

Discussion

Identifying high-flux routes in metabolic networks has been useful in order to, for example, propose principal chains of metabolic transformations [13,44,45]. In this work, we go beyond the identification of high-flux routes with metabolic pathways. Using FBA maximization of the biomass production rate in a certain environment and a filtering tool which needs no *a priori* assumptions for the connectivity of the filtered network, we are able to extract a significant flux backbone as a reduced version of the metabolic network. The SFB is globally connected and retains the characteristic complex features of the metabolic network as a whole. It contains not only high-flux reactions but also many low-flux reactions which are significant for the production or consumption of a certain metabolite. These reactions can be both intra- and interpathway. As an explanation for the strong dependence between both types in the SFB, one could say that the overall performance depends critically on cross interactions. This fact reinforces the idea that pathways are not isolated identities functioning independently of one another [16].

As stated in Ref. [46], properties that originate from evolutionary pressure should not be observed in random networks. In our investigations, the effect of evolutionary pressure is understood to favor the maximization of the biomass yield of the organism [47–49], in accordance to FBA and its assumptions. It is important to note that the assumed objective function used to compute FBA can have a direct effect on the diversity of fluxes in the metabolic network and so in the proportions of pathways in the filtered backbone and its cores. The analysis of the significant flux backbone of *E. coli*, *S. aureus*, *H. pylori*, and *M. pneumoniae* in glucose minimal medium shows that, for the four organisms, the SCCs are mainly composed by reactions that belong to ancient pathways that were already present at the first stages of life. A very

different composition is obtained when considering the backbone as a whole (see Fig. 3 Bottom) and even more when comparing with the nonfiltered metabolic network. In the SFB of *E. coli*, each of the three SCC has a different and definite metabolic function. The largest SCC contains pathways related to energy metabolism and its composition is dominated by the Citric Acid Cycle, Oxidative Phosphorylation, and Glucolysis/Gluconegenesis, in accordance with the fact that the Citric Acid Cycle has evolved toward an optimal chemical design [31] by coupling to Oxidative Phosporylation and Glycolysis [50]. A second SCCs corresponds to the metabolism of lipids, the most important constituents that compose the cell membrane. The smaller SCC is responsible for the synthesis of purines and pyrimidines, vital for DNA / RNA synthesis. 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 functional requirements. On the one side, the smaller cores need chemical energy to perform their tasks and, on the other side, they need also basic building blocks. These needs are covered by the large SCC, which suggests that the smaller SCCs were added to the OUT component of the largest SCC in later steps of evolution. In contrast, a simpler organism like *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 [25]. More precisely, in *M. pneumoniae* the Citric Acid Cycle and Oxidative Phosphorylation do not take place [25,51], 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 [25].

An analysis of the connectivity of the hubs reveals that L-glutamate acquires in the backbone a functional relevance as a reactant, which emphasizes its importance as the amino group donor for nearly all nitrogen-containing metabolites of the cell, in contrast to its role in the nonfiltered network, where it could be classified rather as a product since in minimal medium it must be synthesized. The case of glucose 6-phosphate is also very interesting in the sense that it presents no outgoing connections in the backbone. Hence, we conclude that, in glucose minimal medium, no route consuming this metabolite can be considered more significant than another. However, it is known that this metabolite has many metabolic fates. Therefore, glucose 6-phosphate is a functional precursor of important metabolites but there is no preference and it is equally important for the production of any of them. At the same time, it is critically produced by others, as confirmed by its presence in the OUT component of the backbone, a fact which magnifies its crucial role in metabolism.

The study of the dependence on the environment of the *E. coli* SFB allows us to identify potential adaptation capabilities. Regarding rich medium, we observe that the critical value of α 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. [40]. Membrane lipid metabolism achieves a high relevance, being the most abundant pathway in the largest core of the rich medium metabolic backbone. This may happen because the proliferation response of *E. coli* to the rich medium increases lipid biosynthesis, since lipids are not present as nutrients even in rich media due to its water insolubility. In contrast, biosynthesis of other components, especially amino acids, is obviously reduced. To satisfy the high lipid demand to generate new cells [40], fast-growing cells must synthesize membrane lipid components more rapidly. This leads to an overexpression of the genes related to membrane lipids which, in terms of metabolism, is observed in terms of a high relevancy of the Membrane Lipid Metabolism pathway. Another feature that we find is the loss of prominence of the Pentose Phosphate Pathway. Although this may suggest to a deficiency of the metabolite NADPH, we find that, in the rich medium backbone, the functionally significant production of this metabolite takes place in the Citric Acid Cycle pathway [41]. This evidences the importance of Citric Acid Cycle to produce NADPH, which is tightly related to the synthesis of membrane lipids. An alternative explanation for the increase of the appearance of the Citric Acid Cycle in the backbone is that more metabolites need to be catabolized in rich medium, and the Citric Acid Cycle is the primary site for catabolism of many nonsugar substrates through anaplerotic reactions.

The analysis of the adaptation of *E. coli* to 333 different minimal media shows that the fraction of links in the significant flux backbone is practically independent on the composition of the nutrients present in these environments (see Fig. 4C). This permits the construction of a super significant flux backbone

that comprises all the links in the SFB in each different media and in which each contributing SFB has uniform representation in fraction of links. This leads to the identification of pathways whose associated reactions are potentially 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 [42,43]. This finding could also be seen as a successful positive control for the methodology applied here. Interestingly, the SSFB is enriched with genes of the metabolic network in a core proteome, defined as the minimal set of genes that are consistently used across environmental conditions on the basis of a genome-scale model that accounts for metabolism and gene expression [22], even if our methodology is only based on gene-scale metabolic information and does not require the processing of gene expression data.

Finally, our computational approach presents some limitations. The connectivity structure and pathway composition of the backbones may in principle depend both on the ability of FBA to predict correctly the flux distribution of real bacteria in certain conditions and on the FBA optimal flux solution considered. On the one hand, in *E. coli* FBA has predicted the uptake and release rates of certain metabolites, cell growth rate under different environmental conditions, and gene essentiality with great success [52–54]. On the other hand, FBA flux vector solutions are not unique in a given condition and it is known that there are a large number of possible flux distributions which maximize the biomass production rate. Nevertheless, the linear high-flux subgraph of the metabolic network of *E. coli* *i*JR904 obtained by applying the method in [13] shows small variation across alternate optima in minimal medium, with only a 3–11% of variability [55]. The fact that the high-flux backbone is conserved across optima and it is mostly contained in the SFB, and the preservation by the filtering method of almost all nodes in the unfiltered network, reinforce the hypothesis that the backbone of *E. coli* *i*JO1366 is mostly conserved in alternate optima, both in terms of constituents and in terms of patterns of connection.

Conclusion

The use of filtering methods usually implies a drastic reduction of the complexity of metabolic maps, which weakens the validity of potentially inferred conclusions. In contrast, the application of the disparity filter, based on a flux significance analysis to produce significant flux backbones, enables to reduce the system while maintaining all significant interactions according to α_c and so it becomes a useful tool to unveil sound biological information. Notice that these reduced versions must be seen as a map of the most significant connections, even though other metabolic reactions must be present to achieve viability and to satisfy the physico-chemical laws governing metabolic networks depending on nutritional conditions and other stresses applied to the cell. Our investigations in different bacteria revealed SFBs mainly composed of a core of reactions belonging to ancient pathways that still retain at present a central role in the evolved metabolism. Besides, in *E. coli*, the analysis of the core reveals a dominant direction with the synthesis of purines and pyrimidines and the metabolism of lipids ensuing after energy metabolism. Our approach can be potentially useful in biotechnology and biomedicine. For instance, the SFB highlights the most significant links connecting metabolites present in the one-mode projection of a metabolic network, which correspond to specific reactions with an associated flux responsible for the significance of the link. Given the relation between metabolic genes and enzymatic reactions, the SFB indicates possible gene knockouts that would significantly modify the associated fluxes so as to obtain a desired effect on the production or consumption of certain metabolite and, at the same time, the SFB allows to assess possible relevant collateral effects on the rest of the network. Another potential application exploiting the capabilities of the SFB is the recognition of pathways and particular reactions more sensitive to environmental changes.

Acknowledgements

OG acknowledges support from a FPU grant funded by MINECO. FS acknowledges support from MINECO projects no. FIS2013-41144P and FIS2016-78507-C2-1-P (AEI/FEDER, UE), and the Generalitat de Catalunya grant No. 2014SGR597. MAS acknowledges support from a James S. McDonnell Foundation 21st Century Science Initiative in Studying Complex Systems Scholar Award, MINECO projects No. FIS2013-47282-C02-01 and No. FIS2016-76830-C2-2-P (AEI/FEDER, UE); and the Generalitat de Catalunya grant No. 2014SGR608.

Data Accessibility

Data Accessibility Research data pertaining to this article is located at figshare.com: https://dx.doi.org/10.6084/m9.figshare.5001869.

Author contributions

MAS designed research, OG performed research, all authors analyzed and discussed results, MAS and OG wrote the paper, and all authors revised the manuscript.

References

1 Feist AM, Scholten JCM, Palsson BØ, Brockman FJ and Ideker T (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol Syst Biol* **2**, 2006.0004.

2 Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V and Palsson BØ (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**, 121.

3 Schellenberger J, Park JO, Conrad TC and Palsson BØ (2010) BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinform* **11**, 213.

4 Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM and Palsson BØ (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism - 2011. *Mol Syst Biol* **7**, 535.

5 Orth JD, Thiele I and Palsson BØ (2010) What is Flux balance analysis? *Nat Biotechnol* **28**, 245–248.

6 Gudmundsson S and Thiele I (2010) Computationally efficient flux variability analysis. *BMC Bioinform* **11**, 489.

7 Barrat A, Barthélemy M and Vespignani A (2008) Dynamical Processes on Complex Networks, 1st edn. Cambridge University Press, New York, NY.

8 Newman M (2010) Networks: An Introduction. Oxford University Press Inc, New York, NY.

9 Schuster S and Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. *J Biol Syst* **2**, 165–182.

10 Price ND, Reed JL, Papin JA, Wiback SJ and Palsson BØ (2003) Network-based analysis of metabolic regulation in the human red blood cell. *J Theor Biol* **225**, 185–194.

11 Larhlimi A and Bockmayr A (2009) A new constraint-based description of the steady-state flux cone of metabolic networks. *Discrete Appl Math* **157**, 2257–2266.

12 Border A, Nagarajan H, Lewis NE, Latif H, Ebrahim A, Frederowicz S, Shcellenberger J and Palsson BØ (2014) Minimal metabolic pathway structure is consistent with associated bimolecular interactions. *Mol Syst Biol* **10**, 737.

13 Almaas E, Kovács B, Vicsek T, Oltvai ZN and Barabási AL (2004) Global Organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843.

14 Ma HW and Zeng AP (2003) The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* **19**, 1423–1430.

15 Barabási AL and Oltvai ZN (2004) Network biology: understanding the cells functional organization. *Nat Rev Genet* **5**, 101–113.

16 Serrano MÁ, Boguñá M and Sagués F (2012) Uncovering the hidden geometry behind metabolic networks. *Mol BioSyst* **8**, 843–850.

17 Herfindahl OC (1959) Copper Costs and Prices: 1870–1957, pp. 1–260. John Hopkins University Press, Baltimore, MD.

18 Hirschman AO (1964) The paternity of an index. *Am Econ Rev* **54**, 761–762.

19 Serrano MÁ, Boguñá M and Vespignani A (2009) Extracting the mutiscale backbone of complex weighted networks. *Proc Natl Acad Sci USA* **106**, 6483–6488.

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

21 Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata S, Tomkins A and Wiener J (2000) Graph structure in the web. *Comput Netw* **33**, 309–320.

22 Yang L, Tan J, O'Brien EJ, Monk JM, Kim D, Li HJ, Charusanti P, Ebrahim A, Lloyd CJ, Yurkovich JT *et al.* (2015) Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. *Proc Natl Acad Sci USA* **112**, 10810–10815.

23 Thiele I, Vo TD, Price ND and Palsson BØ (2005) Expanded metabolic reconstruction of *Helicobacter pylori* (*i*IT341 GSM/GPR): an *in silico* genome-scale characterization of single- and double-deletion mutants. *J Bacteriol* **187**, 5818–5830.

24 Becker SA and Palsson BØ (2005) Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation". *BMC Microbiol* **5**, 8.

25 Wodke JA, Puchałka J, Lluch-Senar M, Marcos J, Yus E, Godinho M, Gutiérrez-Gallego R, dos Santos VA, Serrano L, Klipp E *et al.* (2013) Dissecting the energy metabolism in *Mycoplasma pneumoniae* through genome-scale metabolic modeling. *Mol Syst Biol* **9**, 653.

26 Jamshidi N and Palsson BØ (2007) Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain *i*NJ661 and proposing alternative drug targets. *BMC Syst Biol* **8**, 26.

27 Duarte NC, Herrgård MJ and Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* *i*ND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **14**, 1298–1309.

28 Güell O, Sagués F and Serrano MÁ (2014) Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLoS Comput Biol* **10**, e1003637.
29 Serrano MÁ and De Los Rios P (2008) Structural efficiency of percolated landscapes in flow networks. *PLoS ONE* **3**, e3654.
30 Keller MA, Turchyn AV and Ralser M (2014) Nonenzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol Syst Biol* **10**, 725.
31 Meléndez-Hevia E, Waddell TG and Cascante M (1996) 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.
32 Mailloux RJ, Bériaut R, Lemire J, Singh R, Chénier DR, Hamel RD and Appanna VD (2007) The tricarboxylic acid cycle, an ancient metabolic network with a novel twist. *PLoS ONE* **2**, e690.
33 Fell DA and Wagner A (2000) The small world of metabolism. *Nat Biotechnol* **18**, 1121–1122.
34 Suen S, Lu HHS and Yeang CH (2012) Evolution of domain architectures and catalytic functions of enzymes in metabolic systems. *Genome Biol Evol* **4**, 976–993.
35 Evans DR and Guy HI (2004) Mammalian pyrimidine biosynthesis: fresh insights into an ancient pathway. *J Biol Chem* **279**, 33035–33038.
36 Powner MW, Gerland B and Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242.
37 Caetano-Anolles G, Kim HS and Mittenthal JE (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc Natl Acad Sci USA* **104**, 93589363.
38 Tadege M, Dupuis I and Kuhlemeier C (1999) Ethanolic fermentation: new functions for an old pathway. *Trends Plant Sci* **4**, 320–325.
39 Sezonov G, Joseleau-Petit D and D'Ari R (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J Bacteriol* **189**, 8746–8749.
40 Tao H, Bausch C, Richmond C, Blattner FR and Conway T (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181**, 6425–6440.
41 Sauer U, Canonaco F, Heri S, Perrenoud A and Fischer E (2004) The soluble and membrane bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* **279**, 6613–6619.
42 Lourenço A, Carneiro S, Pinto JP, Rocha M, Ferreira EC and Rocha I (2011) A study of the short and long-term regulation of *E. coli* metabolic pathways. *J. Integr Bioinform* **8**, 183.
43 Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM and Palsson BØ (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain specific adaptations to nutritional environments. *Proc Natl Acad Sci USA* **110**, 20338–20343.
44 Bourqui R, Cottret L, Lacroix V, Auber D, Mary P, Sagot MF and Jourdan F (2009) Metabolic network visualization eliminating node redundancy and preserving metabolic pathways. *BMC Syst Biol* **1**, 29.
45 Faust K, Dupont P, Callut J and van Helden J (2010) Pathway discovery in metabolic networks by subgraph extraction. *Bioinformatics* **26**, 1211–1218.
46 Basler G, Grimbs S, Ebenhöh O, Selbig J and Nikoloski Z (2011) Evolutionary significance of metabolic network properties. *J R Soc Interface* **9**, 1168–1176.
47 Ibarra RU, Edwards JS and Palsson BØ (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189.
48 Blank LM, Kuepfer L and Sauer U (2005) Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* **6**, R49.
49 Llaneras F and Picó J (2008) Stoichiometric modelling of cell metabolism. *J Biosci Bioeng* **105**, 1–11.
50 Ebenhöh O and Heinrich R (2001) Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. *Bull Math Biol* **63**, 21–55.
51 Manolukas JT, Barile MF, Chandler DK and Pollack JD (1988) Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in mollicutes. *J Gen Microbiol* **134**, 791–800.
52 Edwards JS and Palsson BØ (2000) Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinform* **1**, 1.
53 Edwards JS, Ibarra SU and Palsson BØ (2001) *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19**, 125–130.
54 Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D *et al*. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* **6**, 390.
55 Samal A (2008) Conservation of high-flux backbone in alternate optimal and near-optimal flux distributions of metabolic networks. *Syst Synth Biol* **2**, 83–93.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Disparity filter in more organisms.
Fig. S1. Heterogeneity of reaction fluxes and statistics of the disparity filter in different organisms.
Appendix S2. Backbone statistics.
Table S1. Statistics of number of reactions and metabolites for the different bacteria in our study.
Appendix S3. Global connectivity of backbones.
Table S2. Statistics of connected components in the SFB of the different bacteria in our study.
Fig. S2. Structure of the largest connected component in the metabolic backbone of *M. pneumoniae*.
Appendix S4. Pathway composition.
Table S3. Pathway composition for *E. coli*.
Table S4. Pathway composition for *S. aureus*.
Table S5. Pathway composition for *H. pylori*.
Table S6. Pathway composition for *M. pneumoniae*.
Appendix S5. Algorithm for the computation of probabilities and significant links.
Appendix S6. Algorithm for the construction of the one-mode projection.
Appendix S7. Additional files.