ASSESSMENT OF TRANSCRIPTOMIC CONSTRAINT-BASED METHODS FOR CENTRAL CARBON FLUX INFERENCE

By

SIDDHARTH BHADRA-LOBO

A thesis submitted to the

Graduate School-Camden

Rutgers, State University Camden

In partial fulfillment of the requirements

For the degree of Master of Science

Graduate Program in Computational & Integrative Biology

Written under the direction of

Dr. Desmond S. Lun

And approved by

_____

Dr. Desmond S. Lun

_____

Dr. Guillaume Lamoureux

_____

Dr. Jinglin Fu

Camden, New Jersey

January 2020

THESIS ABSTRACT

Assessment of Transcriptomic Constraint-based Methods for Central Carbon Flux Inference

by Siddharth Bhadra-Lobo

Dissertation Director:

Dr. Desmond S. Lun

Motivation: Determining intracellular metabolic flux through isotope labeling techniques such as ^{13}C metabolic flux analysis (^{13}C-MFA) incurs significant cost and effort. Previous studies have shown transcriptomic data coupled with constraint-based metabolic modeling can determine intracellular fluxes that correlate highly with ^{13}C-MFA measured fluxes and can achieve higher accuracy than constraint-based metabolic modeling alone. These studies, however, used validation data limited to *E. coli* and *S. cerevisiae* grown on glucose, with significantly similar flux distribution for central metabolism. It is unclear whether those results apply to more diverse metabolisms, and therefore further, extensive validation is needed.

Results: In this paper, we formed a dataset of transcriptomic data coupled with corresponding ^{13}C-MFA flux data for 21 experimental conditions in different unicellular organisms grown on varying carbon substrates and conditions. Three computational flux-balance analysis (FBA) methods were comparatively assessed. The results show when uptake rates of carbon sources and key metabolites are known, transcriptomic data provides no significant advantage over constraint-based metabolic modeling (average correlation coefficients, transcriptomic E-Flux2 0.725 and SPOT 0.650 vs non-transcriptomic pFBA 0.768). When uptake rates are unknown, however, predictions obtained utilizing transcriptomic data are generally good and significantly better than those obtained using constraint-based metabolic modeling alone (E-Flux2 0.385 and SPOT 0.583 vs pFBA 0.237). Thus, transcriptomic data coupled with constraint-based metabolic modeling is a promising method to obtain intracellular flux estimates in microorganisms, particularly in cases where uptake rates of key metabolites cannot be easily determined, such as for growth in complex media or *in vivo* conditions.

Acknowledgements

I would like to thank Dr. Desmond S. Lun for all his patience, help, and mentorship.

Also thank you to Dr. Min Kyung Kim for providing the groundwork for this research by providing data and methods.

Finally, thank you to James J. Kelley for his work on Metabolic Optimization and Simulation Tool (MOST).

Table of Contents

List of Figures

1 Introduction

Computational tools integrating transcriptomic data into genome-scale metabolic models can predict system-level and condition specific metabolic flux distributions. Many methods for inferring metabolic fluxes from gene expression data have been, and continue to be, developed [1-3]. However, the comparative performance of these methods lacks diverse experimental flux data for validation. Existing validation was performed exclusively against flux data generated from *E. coli* and *S. cerevisiae* (yeast) cultures grown on glucose as the sole carbon source [3, 4]. Cells cultured on identical substrates produce highly conserved glucose metabolism pathways [5]. This carbon source bias presents significant similarities in the measured metabolic flux distribution across previous validation datasets which may have been inadequate in assessing predictive performance.

Carbon source availability and relative uptake rates influence cellular metabolism. In nature, heterotrophic microorganisms can encounter a wide set of possible carbon sources to support growth, including sugars, polyols, alcohols, organic acids, and amino acids [6]. Heterotrophs such as *E. coli* and *Bacillus subtilis* have been widely studied and cultured on a variety of substrates including monosaccharides (e.g. glucose, fructose, galactose), disaccharides (e.g. sucrose), and two-carbon compounds (e.g. acetate) [7-11]. Thus, under a multitude of possible carbon sources, an incorrectly constrained heterotrophic model can reduce the predictive accuracy of central carbon fluxes from conventional FBA methods. Gene expression may be useful to impute model constraints based on transcript abundance in the absence of specific carbon source and uptake rate data.

Growth condition encompasses the availability of metabolic state-determining metabolites, both organic and inorganic (e.g. glucose, CO_2, photons, NO_3). Missing or incorrect growth condition information can change flux predictions to alternate metabolic states of the cell. Photoautotrophic unicellular metabolic models are generally well characterized and therefore simpler to constrain with respect to carbon source. The depletion of non-carbon metabolites may metabolically adapt the cell to alternate metabolic states. For example, light inhibition can shift metabolism from either autotrophic, heterotrophic, or a combination of both as mixotrophic in *Synechocystis* sp. PCC 6803 [12]. A substrate void of nitrate can induce replenishing of nitrogen from metabolic sinks such as amino acids for *Synechococcus* sp. PCC 7002 [13]. In the lack of

environmental condition specificity, informational deficit may be overcome with gene expression data such as key pathways being allocated flux values based on the upregulation of associated transcripts.

Previous studies [2, 3] have extensively evaluated the predictive capability of *in silico* flux prediction using measured extracellular and intracellular fluxes in multiple experimental conditions, but under single carbon source bias (glucose) in two organisms. To this end, we have compiled an additional 21 experimental conditions of transcriptome measurements coupled with corresponding central carbon metabolic intracellular ^{13}C flux measurements in 4 organisms (8 in *E. coli*, 8 in *Bacillus subtilis*, 3 in *Synechocystis* sp. PCC 6803, and 2 in *Synechococcus* sp. PCC 7002). These conditions were applied to models run using two transcriptomic methods (E-Flux2 and SPOT) [4] and the non-transcriptomic method parsimonious FBA (pFBA) [14], a method present in previous validation studies [2-4] and shown to give good flux predictions. In this study, the generality of E-Flux2 and SPOT have been validated against pFBA using this new dataset of diverse carbon sources and conditional constraints.

In the absence of carbon source and growth condition data, transcriptomic coupled constraint-based modeling is useful in bridging this information gap. If it is even feasible in the experimental condition of interest, the extraction of ^{13}C-labeled isotopes is costly and laborious. The ^{13}C-labeled data also conveys minimal growth condition information as it cannot be directly applied to non-carbon metabolites [15]. In contrast, gene expression data is relatively simple to gather and is obtained from cell culturing experiments regularly. With transcriptomic FBA methods, researchers can utilize their gathered expression data to estimate intracellular metabolism.

2 Materials and Methods

2.1 Gene expression, flux datasets, and metabolic models

All gene expression measurements obtained were not normalized any further past the instrument processed signal. Any log-transformed data was transformed back to their original scale by exponentiation.

2.1.1 Data and model for *E. coli*

For *E. coli*, both the measured gene expression (single color microarray) and ^{13}C flux data were obtained from a previous study by Gerosa et al. [17]. In this study, data were measured from *E. coli* wild type BW25113 cells growing exponentially on eight different carbon sources: glucose, galactose, gluconate, fructose, glycerol, pyruvate, acetate, and succinate. We used iJO1366 [18] as the genome-scale metabolic model.

2.1.2 Data and model for *B. subtilis*

For *B. subtilis*, we used transcriptomic (single color microarray) and ^{13}C flux data published in [19] and [20], respectively. Data were obtained from *B. subtilis* BSB168 cells grown under eight conditions defined by different carbon sources: glucose, fructose, gluconate, succinate + glutamate, glycerol, malate, malate + glucose, and pyruvate. For the genome-scale metabolic model of *B. subtilis*, the model published by Oh et al. [21] was used.

2.1.3 Data and model for Synechocystis sp. PCC 6803

For *Synechocystis* sp. PCC 6803, transcriptomic (RNA-seq) data was graciously provided by Dr. Le You (University of California San Diego, USA) and Dr. Yinjie Tang (Washington University in St. Louis, USA) [12]. The ^{13}C flux data was compiled from three different publications [12, 22, 23]. Data were measured from the strain *Synechocystis* sp. PCC 6803 grown under three different conditions: photoautotrophic (i.e. HCO_3^- (bicarbonate) as the main carbon source) [23], photomixotrophic (i.e. open air CO_2 + glucose) [22], and heterotrophic (i.e. open air CO_2 + glucose, constrained photons) [12], respectively. We used the genome-

scale metabolic model of *Synechocystis* sp. PCC 6803 developed by Knoop et al. [24]. An external pseudo-compartment was added to the model through which metabolites can be exchanged with the external environment via cellular transport reactions.

2.1.4 Data and model for *Synechococcus* sp. PCC 7002

For *Synechococcus* sp. PCC 7002, the transcriptomic (RNA-seq) data was obtained from a previous publication by Ludwig and Bryant [25]. The ^{13}C flux data for this model was gathered from Qian et al. [26]. Data were measured from *Synechococcus* sp. PCC 7002 cells grown photoautotrophically (i.e. CO_2 carbon source and photon uptake) with 10 mM nitrate and with no other nitrogen source. iSyp821 was used for the organism's genome scale-metabolic model [13].

2.2 Computational Prediction and Correlation

2.2.1 Computational metabolic flux prediction

In this study, E-Flux2, SPOT [4], and pFBA [14] were used to predict metabolic flux distributions. Biomass production was set as the objective function for E-Flux2 and pFBA. All FBA methods used in this study are referenced from their original publications [4, 14]. Computations were carried out on the macOS Mojave platform using a personal computer with a 3.1 GHz Intel Core i5 processor with 8GB of RAM. E-Flux2, SPOT and pFBA methods are implemented in MOST (Metabolic Optimization and Simulation Tool) which is available at http://most.ccib.rutgers.edu [27].

2.2.2 Correlation calculations

Validation of the predictive accuracy of the methods used in this study was done by calculating the uncentered Pearson product-moment correlation between *in silico* fluxes and corresponding ^{13}C determined intracellular fluxes as previously described in [4]. A value of the correlation coefficient close to +1 or -1 indicates a strong relationship via a positive or negative scale factor, respectively, between experimentally measured fluxes and computationally predicted fluxes; a value of 0 indicates no such relationship [28]. If a measured reaction corresponds to a set of consecutive reactions in the model that are linked with intermediate

metabolites (AND relationship), then the minimum flux value among the predicted fluxes was used. If a measured flux corresponds to multiple identical reactions (OR relationship), the sum of those predicted fluxes was used to calculate the correlation.

Correlations were calculated between the measured and predicted fluxes per carbon source in MATLAB R2018b (The Math Works Inc., Natick, Mass., USA). The predicted fluxes for the transcriptomic methods (E-Flux2 and SPOT) were generated using the respective carbon source and/or growth condition gene expression profile. pFBA does not use gene expression and was run in two scenarios, one where the carbon source flux was not specified (i.e. maximal uptake allowed) and one where the carbon source flux is specified (for uptake rates used see Supplementary **S1 Table**). Carbon source fluxes were gathered from uptake rates from the respective ^{13}C flux experiments (mmol/g DCW/h).

3 Results

To test generality of E-Flux2 and SPOT, we evaluated predictive accuracy by calculating the uncentered Pearson correlation (Section Methods) between experimentally measured and computationally predicted intracellular fluxes using transcriptomic data, for the compiled 21 experimental conditions. The dataset consists of 8, 8, 3, and 2 conditions of *E. coli*, *B. subtilis*, *Synechocystis* sp. PCC 6803, and *Synechococcus* sp. PCC 7002, respectively (Section Methods provides carbon source information). We expect model choice affects the transcriptional FBA methods more than the non-transcriptional. A less complete model may have reduced constraint mapping from the relevant gene expression data.

We have chosen the uncentered Pearson correlation as a good, goodness-of-fit metric because transcriptomic flux inference, in general, estimates that high transcript count corresponds with high flux, but not the actual flux value. Therefore, the predicted flux values are in arbitrary units. This type of correlation captures predictive accuracy irrespective of the scaling introduced by the gene expression data.

Testing flux prediction under known and unknown carbon sources, *E. coli* and *B. subtilis* fluxes were simulated under different carbon source availabilities, at three different stages.

A. DC: Known carbon source and uptake rate information available, uptake rate is only supplied to the non-transcriptomic method, pFBA.
B. AC: Unknown carbon source and uptake rate, only eight speculative carbon sources without uptake rate data are available to the model.
C. Full AC: No carbon source information available, all possible carbon sources (and any other extracellular metabolites) opened for exchange into the model.

Testing flux prediction under different growth condition in PCC 6803 and PCC 7002. Fluxes were simulated under based on the organism's possible metabolic states, at two different stages. Carbon sources are fewer and simpler to constrain in these photoautotrophic organisms, therefore here AC is the same as Full AC in the previous heterotrophic organisms.

A. DC: Growth condition information and metabolic state are known and uniquely applied to simulate each respective organism's metabolic states. Carbon uptake rate data only supplied to pFBA.

B. AC: No growth condition information is available, all possible carbon and inorganic metabolites available for simulating the mixotrophic condition.

Unknown growth condition was used to demonstrate cases of complex media or *in vivo* growth of cultures. An example of this would be in studying the metabolism of enteric bacteria both pathogenic and commensal/mutualistic, in which the growth medium is complex, and the culture is grown *in vivo*. An example of a pathogenic model is *Mycobacterium tuberculosis*. In tuberculosis, the bacterium live inside of scar tissue of the lung and their metabolism is uncertain. To measure this using conventional ^{13}C-MFA would not be feasible, but extraction of RNA expression data has been shown to be possible [29]. This may also be useful in other cases of bacterial pathogenesis. In the cases of commensalism/mutualism within the gut microbiome, the distribution of bacterial species in the gut has been shown to vary based on diet [30]. It may be possible to determine how metabolism in the microbes shift in the species that continue to persist in the gut during dietary changes, using transcriptomic flux prediction. A hypothetical experiment would be to sample RNA from the gut during a period of one type of host diet, then sample RNA again after a period of time on another diet. Although this is dependent on the expression profiles and metabolic models to be complete enough for prediction of central carbon metabolism.

3.1 Known and Unknown Carbon Source

3.1.1 Central Carbon Flux Correlations in *E. coli*

Under direct carbon source (DC) the *E.coli* models were supplied with only one carbon source each (**Fig 1A**). With complete carbon source information supplied, correlation between the transcriptomic and non-transcriptomic methods are similarly good. pFBA was provided an additional constraint to improve prediction with the experimentally measured carbon source uptake flux (uptake rate) being set within the pFBA runs only (**Fig 1A**). For a speculative set of possible carbon sources, **Fig 1B** shows the measured fluxes of *E. coli* grown on a single carbon source correlated with the predicted fluxes when supplied with all 8

carbon sources (AC) used in the measurements (i.e. glucose, galactose, gluconate, fructose, glycerol, pyruvate, acetate, and succinate) per model. **Fig 1C** simulates the absence of any carbon source and uptake rate information, with the model fully open for exchange with the extracellular environment. Here overall predictive accuracy drops across all methods as the number of available carbon sources increases. E-Flux2 on average performs comparably to SPOT, with slightly worse correlation on average. Models run with all 294 available carbon sources (Full AC) and 30 ion sources, shows that on average E-Flux2 and SPOT generate reasonable correlations (**Fig 1C**). All three methods produce lower correlations for carbon sources found in the TCA cycle (**Fig 1C** Full AC acetate, pyruvate, and succinate). These low correlations were investigated and determined to be due to predicting flux opposite in direction to the measured flux (**Fig S3**). The measured fluxes for glycolysis are negative in reaction direction and the predictions are positive, while the measured fluxes for TCA cycle reactions are positive, and the predicted fluxes are negative. SPOT maintains higher correlations compared to E-Flux2 and pFBA due to predicting the TCA cycle reactions in the correct direction.

3.1.2 Central Carbon Flux Correlations in *B. subtilis*

The *B. subtilis* measured fluxes consist of 8 different carbon sources, with two cases of double carbon sources experiments (**Fig 2** glutamate + succinate and malate + glucose). **Fig 2A** shows the DC correlations from E-Flux2 and pFBA is comparable, with known carbon flux giving the best correlations on average. In speculative carbon sources, **Fig 2B**, all three methods perform similarly on average for AC. pFBA performs similarly poorly to the other methods for the double carbon cases and only marginally better for glutamate + succinate (see Discussion). SPOT performs the best for the TCA cycle single carbon source cases (AC malate, AC pyruvate). The same can be seen in the Full AC models (269 carbon sources, 25 ion sources) (**Fig 2C**) but pFBA on average performs worse, most notably in the TCA cycle single carbon sources.

3.2 Known and Unknown Growth Condition

3.2.1 Central Carbon Flux Correlations in PCC 6803

In *Synechocystis* sp. PCC 6803 autotrophic, mixotrophic, and heterotrophic conditions (**Fig 3A)** E-Flux2 and pFBA produce very similar central carbon flux distributions under the autotrophic condition. These predictions correlate well with the autotrophic measured fluxes, suggesting that both methods are producing nearly identical flux distributions. In the mixotrophic condition, pFBA, with known carbon source and flux, produces a higher correlation than the other methods. All methods predict heterotrophic central carbon metabolism poorly, with SPOT predicting the only positive correlation between measured and predicted fluxes. SPOT produces similar correlation values with the three measured flux distributions, and the only non-negative correlation consistently for all three conditions. **Fig 3B** shows the correlations of fluxes predicted using the three conditional gene expression sets (expression data collected from autotrophic, mixotrophic, and heterotrophic cultures) while under mixotrophic constraints, simulating how predicted fluxes correlate under unknown conditions and guided by transcriptomic data. E-Flux2 and pFBA produce negative correlations for all mixotrophically constrained predictions. SPOT again provides the only positive correlations.

3.2.2 Central Carbon Flux Correlations in PCC 7002

Measured fluxes from *Synechococcus sp.* PCC 7002 in nitrogen replete (10 mM nitrate) and nitrogen deprived (no nitrogen source) conditions correlated well with predicted fluxes under autotrophic constraints. SPOT produced significantly better central carbon flux for the nitrogen deprived condition and the other methods performed similarly across both nitrogen conditions (Fig 4, N replete). Fig 4B shows PCC 7002 in an AC mixotrophic condition not naturally exhibited in PCC 7002 (see Results and discussion, Unknown carbon source and growth condition). Both sets of predicted fluxes are allowed open uptake of all carbon sources as well as NO_3 uptake, simulating unknown carbon source and unknown nitrate availability. SPOT performs well under the set of unknown conditions, while the other methods perform poorly.

4 Discussion

4.1 Known and Unknown Carbon Source

For the *E. coli* and *B. subtilis* models, if carbon source and uptake rates are known, the directly provided carbon source and uptake rate information (DC) produces flux predictions in non-transcriptomic pFBA that are comparable to transcriptomic E-Flux2 (**Fig 1A, 2A**). SPOT provides a reasonable, but lower average prediction for both DC cases. E-Flux2 predicts flux similarly to pFBA except E-Flux2 was not provided any uptake rate information. The effect of gene expression derived reaction bounds predicts central carbon flux well, even without providing respective carbon uptake rates. This suggests that gene expression can serve as a substitute for measured carbon source uptake information, if the carbon source is known.

If carbon source is speculatively known, and uptake rate is unknown, as presented with a relatively small set of 8 possible carbon sources (AC), pFBA predictive power drops significantly (**Fig 2A, 2B**). Without transcriptomic data, pFBA sets fixed proportion uptake rates of the available metabolites in the model across multiple cases. This affects the subsequent central carbon flux prediction as a single flux pattern is being predicted across all conditions. In contrast, the transcriptome coupled methods do not have the same uptake of carbon source per condition, as the gene expression dictates the proportions of carbon source flux for cellular uptake. This suggests that with unknown uptake rates and speculatively known carbon sources, gene expression can still serve as a substitute for measured uptake rate data.

Under both unknown carbon source and unknown uptake rates (Full AC), where the models are allowed uptake of all possible carbon sources present in the model, the pFBA average prediction score drops further while E-Flux2 and SPOT remains similar to their AC correlations (**Fig 1C, 2C).** The E-Flux2 and SPOT average correlation even increases slightly from the *B. subtilis* AC to Full AC cases. A possible explanation is that in the overabundance of carbon sources, the gene expression can mediate the allocation of flux feeding into central carbon metabolism when presenting from multiple metabolic network entry points and thereby predict reaction directionality better (see supplementary **S3 – S5 Fig**). This is in contrast to when flux directionality is set based on a small set of carbon sources, such as the TCA cycle or glycolysis relevant metabolites.

Additionally, in the Full AC model, all ion uptake reactions were open, suggesting the transcriptome can also facilitate ion flux prediction, where ^{13}C data generally does not provide information. For both unknown carbon source and unknown uptake rate conditions (**Fig 1C, 2C**), SPOT performs the best on average. This is likely due to SPOT maximizing correlation with flux prediction and the gene expression set, rather than setting expression-based reaction bounds (as in E-Flux2) which can set a large flux window that can affect predicted directionality in subsequent reactions (see supplementary **S3 – S5 Fig**). The generally higher prediction correlations for E-Flux2 and SPOT suggest that under both unknown carbon source and unknown uptake rates, gene expression data can substitute for carbon source and uptake rate information for central carbon flux prediction.

On average the transcriptomic methods perform better than pFBA under unknown carbon source and uptake, but in one exception of the double carbon source conditions, pFBA predicts central carbon flux with higher accuracy than either transcriptomic method across the DC, AC, and Full AC cases (**Fig 2. A, B,** and **C** glutamate + succinate). This is potentially due to pFBA predicting low flux correctly for a subset of the measured flux values for *B. subtilis*, while E-Flux2 and SPOT allocated different fluxes for these reactions based on the presence of the associated transcripts (see reaction directions in supplementary **S5 Fig**). Hence, when a measured reaction has low flux, but some transcript abundance, the transcriptomic methods may attribute more flux to these reactions.

Additionally, carbon source similarity effects flux predictions. On a carbon source basis, glutamate + succinate measured fluxes are similar to glycerol and pyruvate measured fluxes. The other double carbon source (malate + glucose) exhibits a measured flux distribution very close to the single carbon malate measured distribution (see supplementary **S1 Fig A, S5 Fig**). This suggests that some carbon sources produce similar flux distributions to others, both experimentally and *in silico*. This is supported by the clustering of pFBA flux patterns across all constraints and conditions (supplementary **S1 Fig B, S5 Fig**) which shows similarity between the predicted overall glutamate + succinate distribution to glycerol and pyruvate predicted distributions. This effect has also shown to shift flux predictions away from the measured distribution. In one case, the predicted distribution for malate + glucose more closely resembles the predicted glucose distribution, but in the measured flux patterns the malate + glucose measured flux distributions where more closely resembles the malate flux distribution.

4.2 Known and Unknown Growth Condition

For the cyanobacteria models (PCC 6803 and PCC 7002), carbon source is relatively easy to choose and constrain. The models we assessed are known to fix a single source of inorganic carbon under autotrophic condition, which pFBA can predict well with known carbon source and uptake rates (**Fig 3A**). In the autotrophic growth condition, uptake rate of the inorganic carbon source does not significantly affect central carbon flux prediction (see supplementary **S2 Fig**). But if an organism can increase biomass in multiple possible growth conditions (PCC 6803) then information pertaining to the presence and uptakes rates of inorganic carbon source versus glucose is much more useful.

With unknown growth condition information for PCC 6803, a reasonable approach for modeling the flux distribution is under mixotrophic conditions. That is, allow uptake of both inorganic and organic carbon as well as photon flux and use the associated gene expression to dictate how fluxes should be allocated. **Fig 3B** shows that under such conditions, pFBA and E-Flux2 predict similarly poor central carbon flux. However, SPOT consistently produces positive correlations between the predicted and measured fluxes, across growth conditions. This suggests that with SPOT, gene expression can give some idea of what the condition an organism is growing under using gathered gene expression and the genome-scale metabolic model. A possible explanation for the lower predictive accuracy in both E-Flux2 and pFBA compared to SPOT, is that under glucose availability the typical glycolysis flux distribution is not always found it nature (**S6 Fig**). In PCC 6803, we found fluxes in the pentose phosphate pathway (sees supplementary **Table S2**), which is an alternative metabolic route to glycolysis, has significantly higher flux predicted through it for SPOT in comparison to the other methods. This is further supported by information suggesting that PCC 6803 is merely a facultative heterotroph and therefore only metabolizes exogenous organic carbon when given no other choice [31].

In PCC 7002, the growth condition is only partially known. PCC 7002 is modeled under photoautotrophic conditions, but key secondary metabolite uptake rates are unknown (NO_3 exchange). Here pFBA predicts central carbon flux poorly. By applying different uptake rates of non-carbon metabolites, it is possible to determine whether an organism is in one metabolic state versus another. For example, constraining the uptake of oxygen can produce a flux distribution for anaerobic metabolism [32]. Similarly, in PCC 7002 the presence and depletion of nitrate to the system can lead to different intracellular carbon utilization.

PCC 7002 is known to be an obligate photoautotroph [33]. Therefore, non-transcriptomic methods should be able to perform well in predicting central carbon metabolism, but **Fig 4** shows that pFBA given known carbon source and uptake rate performs worse than the transcriptomic methods in both N replete and N deprived cases. In **Fig 4A** SPOT predicts N deprived central carbon flux better than the other methods. This likely due to the drawing of flux from the nitrogen sinks such as amino acids in order to accommodate for the lack of extracellular nitrate.

4.3 Unknown Carbon Source and Growth Condition

As an extension of our findings, PCC 7002 was constrained under a second artificial growth condition was set to mimic mixotrophic conditions. We attempted to predict flux using the second condition's set of incorrect conditional constraints and see how gene expression might help reduce prediction error. This allows for carbon sources other than CO_2 allowed for uptake as well as unconstrained NO_3 uptake for both the N replete and N deprived cases (**Fig 4B**). This mixotrophic state is not found in nature, and therefore the PCC 7002 central carbon flux distribution correlation was expected to be poor [33]. With the nitrate growth condition unspecified in the model, NO_3 was allowed into the cell freely for both conditions. The correlations for E-Flux2 and pFBA were indeed poor, but SPOT produced strong correlations. This suggests that even in incorrectly constrained models supplied with unrealistic carbon sources and no secondary metabolite information, gene expression can still be used predict central carbon flux well (**S7 Fig**).

4.4 Conclusion

In this study, we compiled 21 experimental conditions and corresponding transcriptomic data for cells grown on various carbon sources and conditions. The predicted fluxes were correlated against experimentally measured fluxes to evaluate the predictive power of E-Flux2 and SPOT compared with the non-transcriptomic method, pFBA. pFBA is a representative method for comparison as it was shown to have good predictions, was used in the previous two validations studies, and does not use transcriptomic data [2, 3].

If carbon source and uptake rate information are accurately known for microorganisms and gene expression data is unavailable, pFBA is a suitable method for central carbon flux prediction (**Fig 1A, Fig 2A**). Even with well-defined carbon source, uptake rate, and growth condition information (other factors on the cell's metabolism, such as light intensity), E-Flux2 performed better than pFBA in 13 of the 21 models. In all of these cases E-Flux2 was not provided any measured uptake rate data, while pFBA was.

If a carbon source or growth condition informational deficit is encountered, then SPOT is the method of choice as it consistently produced good correlations and can account for noncanonical internal metabolism (see Results and discussion, Known and unknown growth condition). Although pFBA can give good predictions, any uncertainty in carbon source or growth condition carries the risk of generating very poor predictions. Even with accurate carbon source and growth condition information pFBA can still produce negative correlations (**Fig 3A**). Gene expression can produce better central carbon flux as the expression data can account for other unknowns in the model, beyond just the carbon source (**Fig 4A,** N deprived).

Based on the findings in this study, we propose a general decision tree to be used in constraint-based modeling for central carbon flux prediction in microorganisms (**Fig 5**). In this figure, if no expression data is available, then pFBA is the default method of choice. If any expression data is available, then a transcriptomic method is suggested as gene expression has been shown to account for additional informational deficits beyond carbon source such as ion exchanges.

Using validated methods like SPOT can minimize the risk of predicting incorrect central carbon flux distributions in the absence of accurate carbon source and growth condition data. Not only does SPOT consistently produce positive correlations in all 21 samples, but also produces low if not the overall lowest, standard deviations in predictive accuracy (**Figs 1-5,** legend σ-values). For future improvement, developing a method for better determining flux directionality based on gene expression values should improve flux transcriptomic flux prediction.

In cells grown on well-defined media, it is relatively easy to determine carbon sources and uptake rates. The carbon source is generally known, while the uptake rate is determined from measuring how fast a culture consumes it. For cells grown *in vivo* or on complex media, where growth condition cannot be fully defined, ^{13}C-labeling may not be even feasible, let alone the cost. Additionally, specifics pertaining to the growth conditions such as inorganic compound exchange may not be available or easily measured. In such cases,

gene expression data can nevertheless be gathered simply and cheaply, and methods to infer intracellular metabolic flux from transcriptomic data (such as E-Flux2 and SPOT) have great utility.

5 Appendix

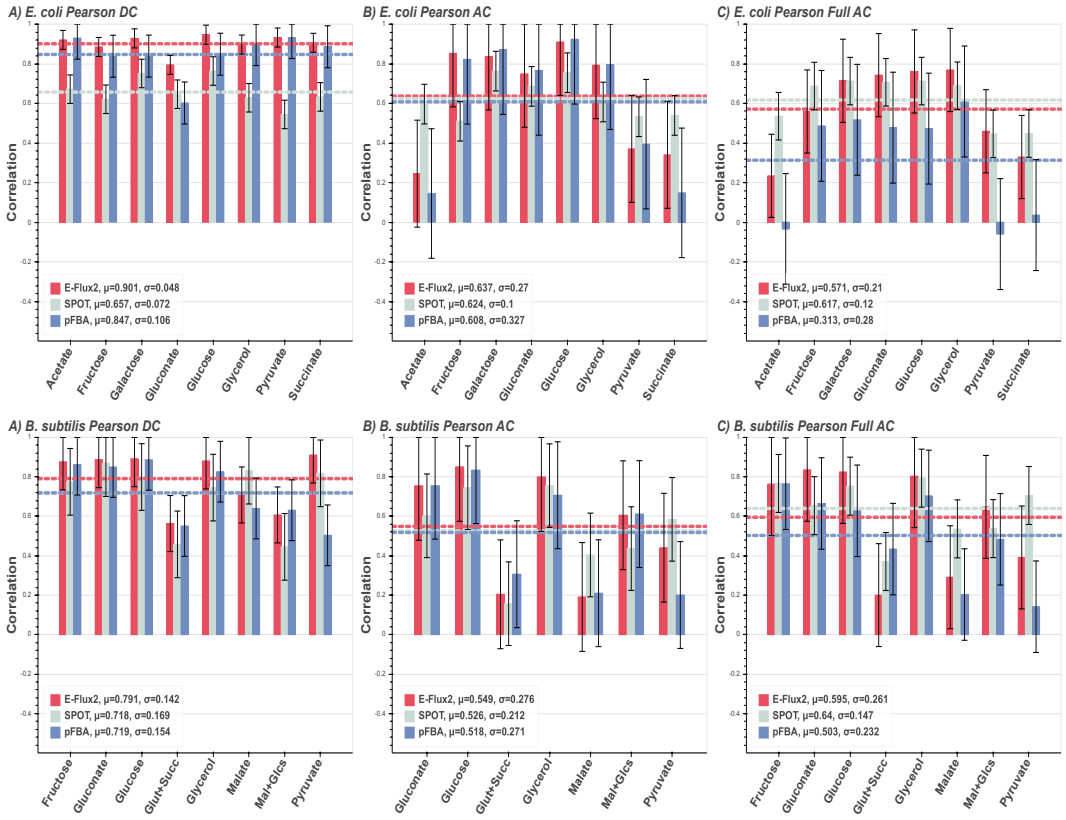


Fig 1. *E. coli* predictions: Correlations between measured and predicted flux of *E. coli* grown on 8 different carbon sources for the three FBA methods E-Flux2 (red), SPOT (gray), pFBA (blue). Horizontal dashed lines are the respectively colored mean correlations per method. The mean (μ) is the average prediction correlation per method. The standard deviation (σ) is the spread of prediction correlation above and below the mean, denoted by the error bars. **(A)** Respective direct carbon source (DC) supplied. pFBA was given the additional constraint of known uptake rate in the single carbon source, while E-Flux2 and SPOT were not. All methods perform consistently across the individual carbon sources. **(B)** All 8 carbon sources supplied and correlated with measured flux from single carbon growth (AC). Correlations drop in all methods, particularly in the TCA cycle carbon sources (Acetate, Pyruvate, and Succinate). **(C)** All possible carbon sources in the model supplied (Full AC). All methods again lose performance, but the transcriptomic methods retain decent correlations. See Supplementary **S1 Table** for uptake rates used.

Fig 2. *B. subtilis* predictions: Correlations between measured and predicted flux of heterotrophic *B. subtilis* grown on 8 different carbon sources for the three FBA methods E-Flux2 (red), SPOT (gray), pFBA (blue). Double carbon sources are denoted as glutamate with succinate (Glut + Succ) and malate with glucose (Mal + Glcs). Horizontal dashed lines are the respectively colored mean correlations per method. The mean (μ) is the average prediction correlation per method. The standard deviation (σ) is the spread of prediction correlation above and below the mean, denoted by the error bars. **(A)** Respective direct carbon source (DC) supplied. pFBA was given the additional constraint of known uptake rate in the single carbon source, while E-Flux2 and SPOT were not. All methods perform consistently across the individual carbon sources, with minor drops in correlation for double carbon sources (Glut + Succ and Mal + Glcs). **(B)** All 8 carbon sources supplied and correlated with measured flux from single carbon growth (AC). Correlations drop in all methods, particularly for Malate. **(C)** All possible carbon sources in the model supplied (Full AC). All methods again lose performance,

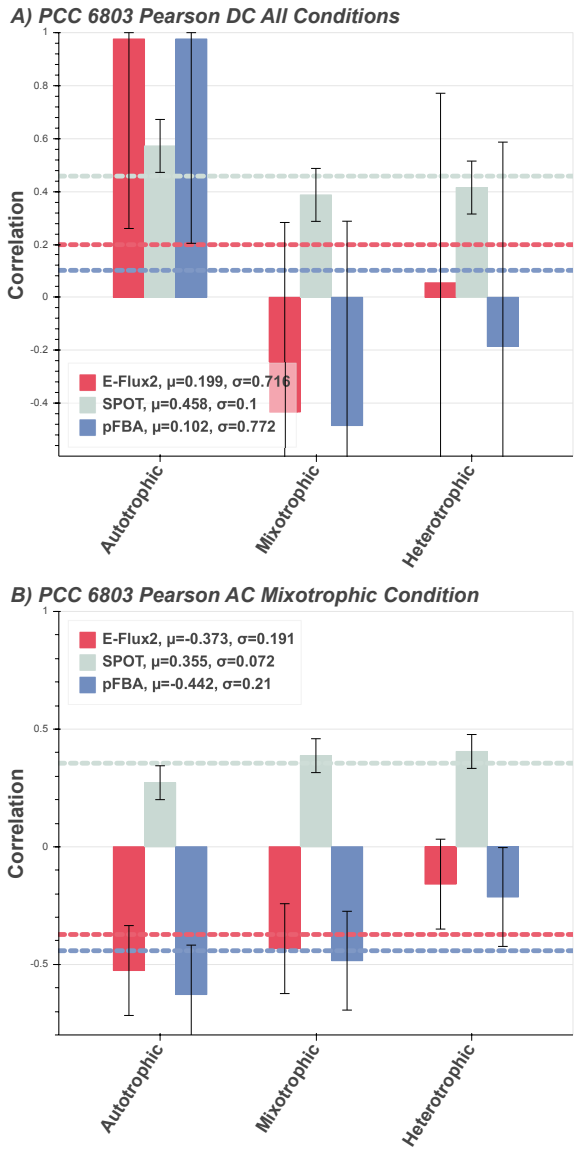


Fig 3. *Synechocystis* sp. PCC 6803 Predictions: Correlations between measured and predicted flux of multitrophic *Synechocystis sp.* PCC 6803 grown in 3 different environment conditions for the three FBA methods E-Flux2 (red), SPOT (gray), pFBA (blue). Horizontal dashed lines are the respectively colored mean μ, correlations per method. The mean (μ) is the average prediction correlation per method. The above and below the mean. A) Autotrophic, mixotrophic, and heterotrophic conditional constraints standard deviation (σ) is the spread of prediction correlation applied and correlated with respective measured fluxes, denoted by the error bars. pFBA was given the additional constraint of known uptake rate in the single carbon source, while E-Flux2 and SPOT were not. B) Mixotrophic condition constraints applied and correlated with all three conditional measured fluxes. See Section Methods for condition specific model constraints and see Supplementary **S1 Table** for uptake rates used. for uptake rates used.

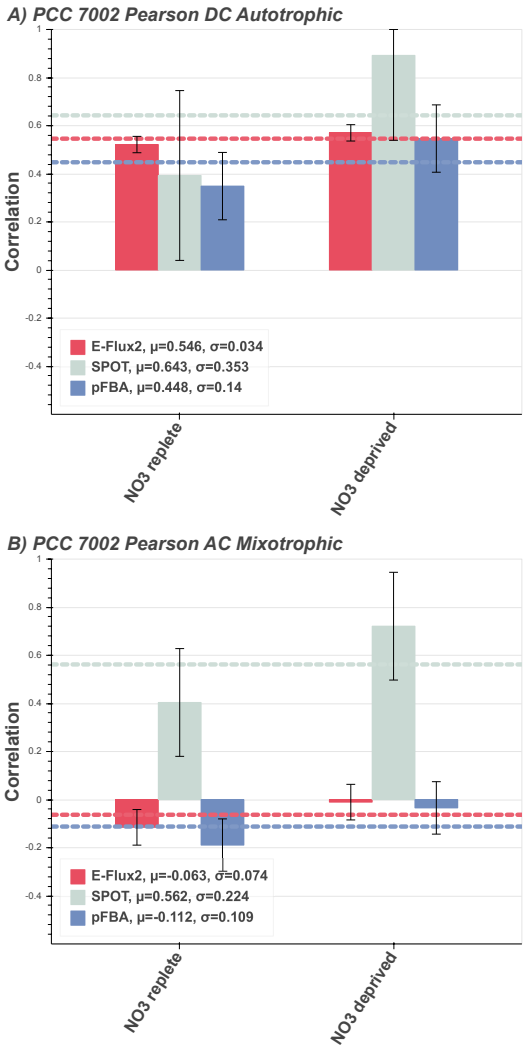


Fig 4. *Synechococcus* sp. PCC 7002 predictions: Correlations between measured and predicted flux of multitrophic *Synechococcus sp.* PCC 7002 grown in autotrophic conditions for the three FBA methods E-Flux2 (red), SPOT (gray), pFBA (blue). Horizontal dashed lines are the respectively colored mean correlations per method. The mean (μ) is the average prediction correlation per method. The standard deviation (σ) is the spread of prediction correlation above and below the mean, denoted by the error bars. **(A)** Autotrophic conditional constraints applied and correlated with N replete and N deprived measured fluxes. Supplementary Materials for uptake rates used. **(B)** PCC 7002 in AC autotrophic condition (mixotrophic, see Discussion) and unconstrained NO3 uptake. See Section Methods for condition specific model constraints and see Supplementary **S1 Table** for uptake rates used. for uptake rates used.

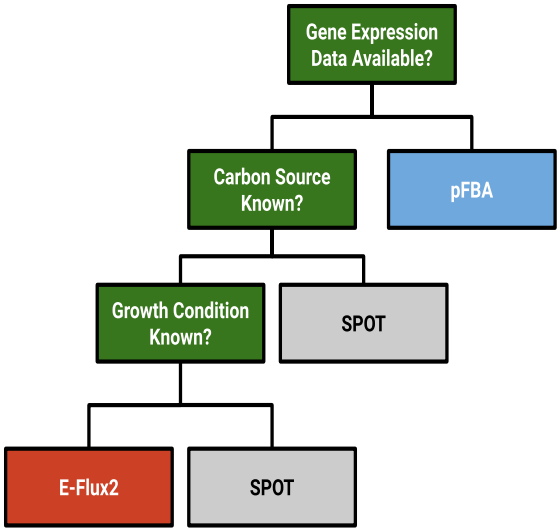
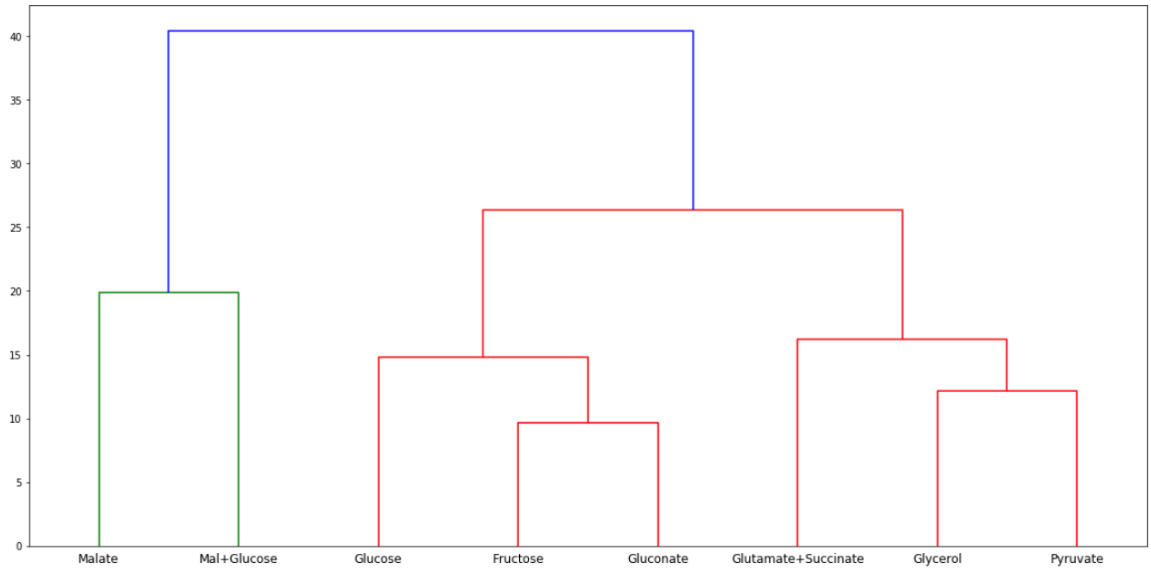
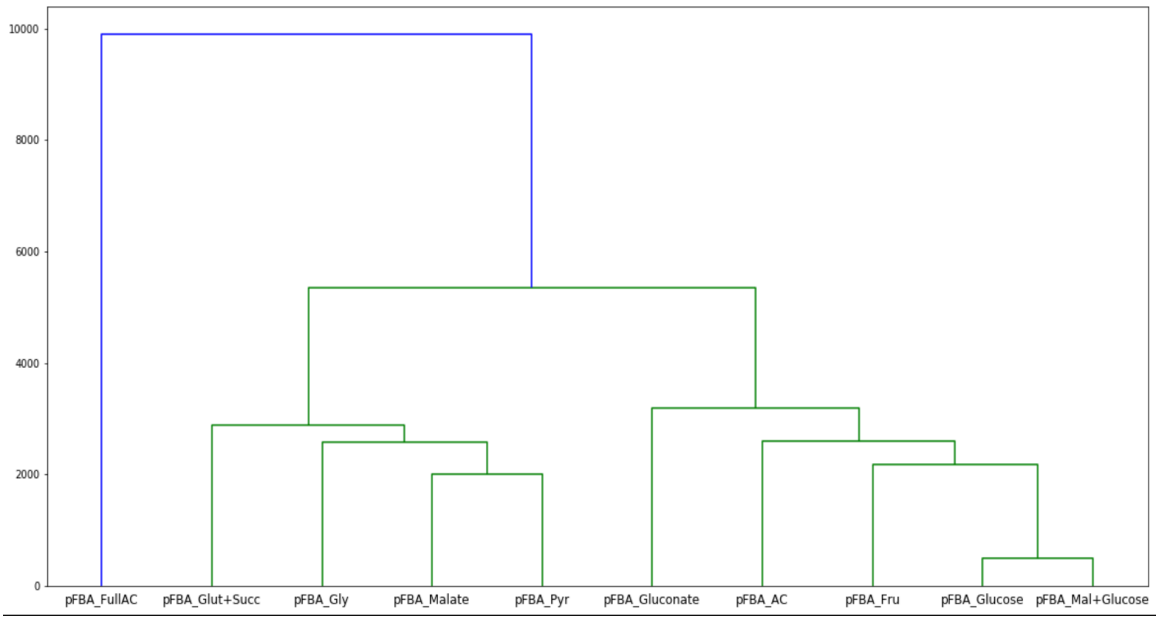


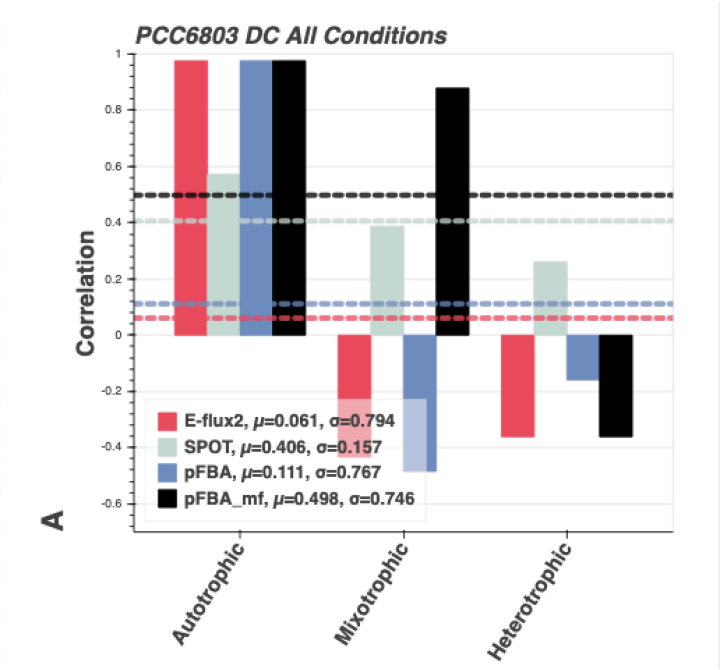
Fig 5. General methods decision tree: The three FBA methods are shown as E-Flux2 (red), SPOT (gray), pFBA (blue). Left branches on the tree indicate a YES decision, right branches indicate a NO decision. Growth condition refers to the availability of inorganic and organic metabolites that can shift metabolism between different states (e.g. photons s, NO_3, CO_2, glucose).



S1 FigA. *B. subtilis* carbon source clustering: *B. subitilis* single and double (Malate+Glucose and Glutamate+Succinate) carbon source measured flux data dendogram. Ward linkage was used and minimizes the sum of squared differences within all clusters based on measured flux values.

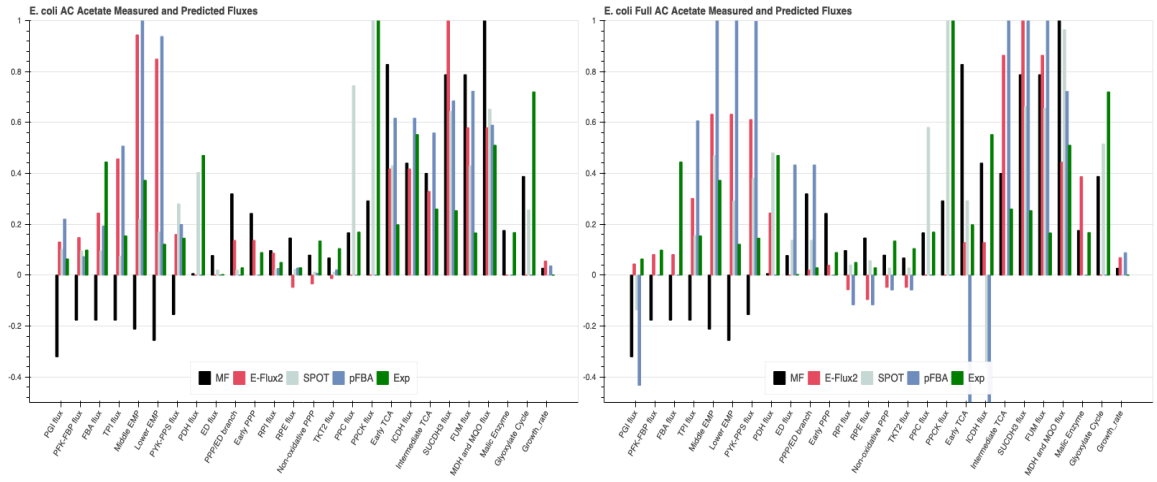


S1 Fig B. *B. subtilis* pFBA carbon source based flux distribution clustering: All *B. subitilis* pFBA flux distributions. AC refers to the single flux pattern produced by pFBA when allowed all 8 carbon sources. FullAC refers to the single pFBA flux distribution produced under all possible carbon sources available in the model. FullAC distribution is the outgroup. Glut+succ looks similar to the other TCA metabolite fluxes, while the AC distribution looks similar to the the sugars in the right subclusters.

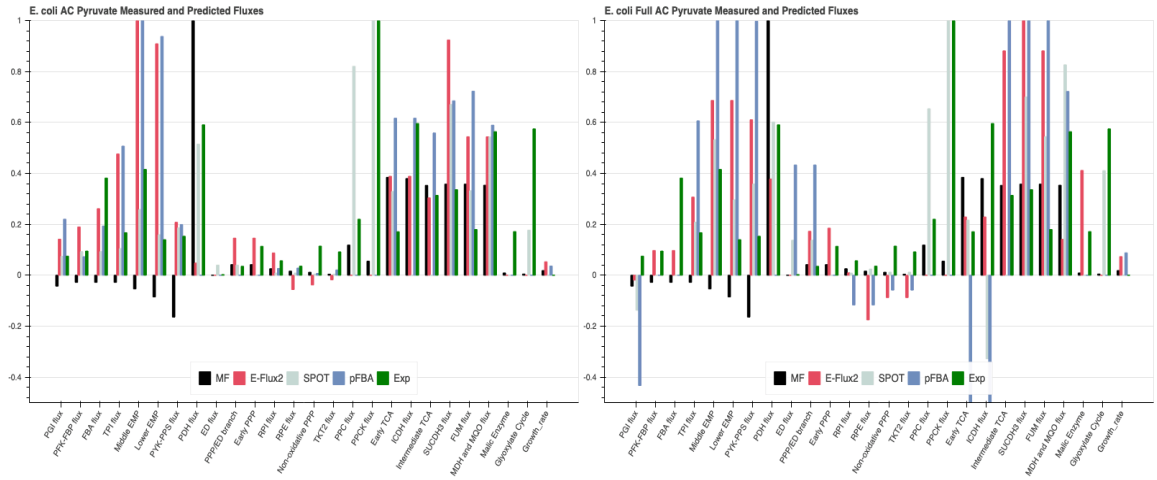


S2 Fig. PCC 6803 pFBA_mf vs pFBA: Similar to Fig 3 in the main text but shows that provided measured uptake rates with pFBA_mf (black) did not significantly improve the autotrophic flux prediction compared to pFBA (blue).

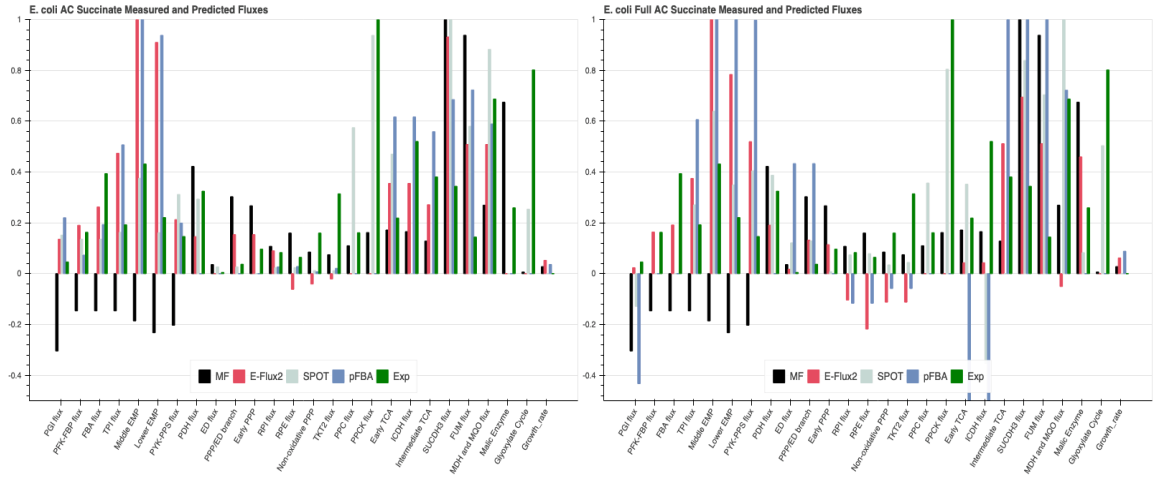
A)


E. coli AC Acetate Measured and Predicted Fluxes
E. coli Full AC Acetate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA
Exp

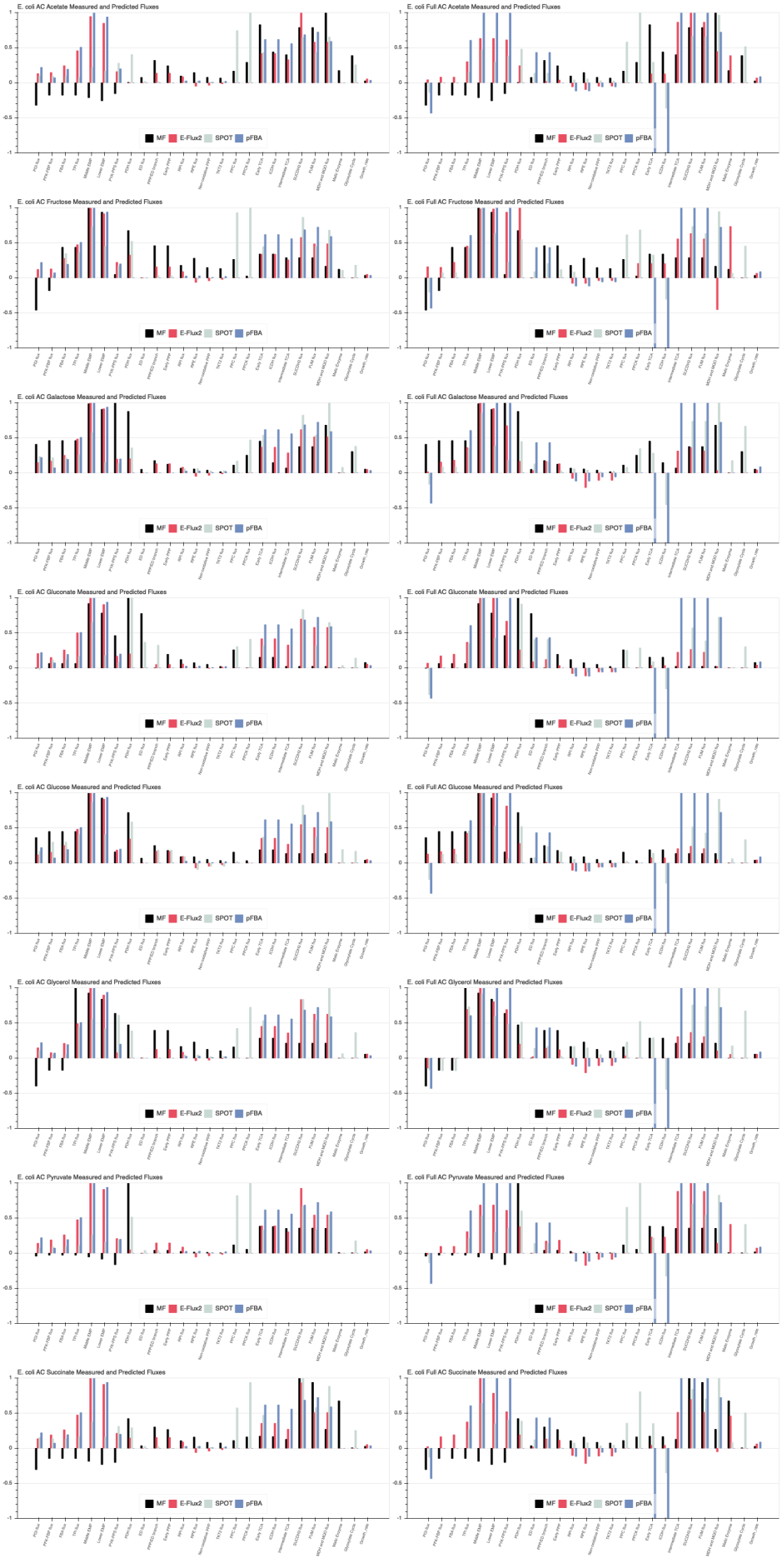
B)


E. coli AC Pyruvate Measured and Predicted Fluxes
E. coli Full AC Pyruvate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA
Exp

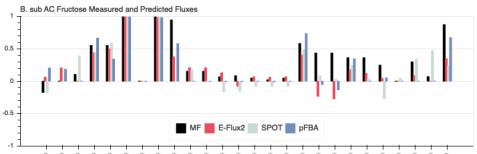
C)

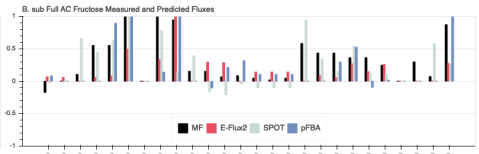

E. coli AC Succinate Measured and Predicted Fluxes
E. coli Full AC Succinate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA
Exp

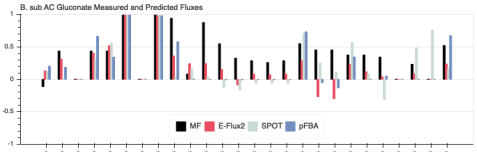
S3 Fig. *E. coli* flux directionality AC vs Full AC: Measured flux (black), E-Flux2 (red), SPOT (gray), pFBA (blue), and gene expression transcript abundance (green); all self-normalized to maximum values 1 and -1. Positive values reflect forward flux and negative values reflection reverse direction flux. The x-axis shows the measured central carbon flux reaction names. **(A)** Acetate data. **(B)** Pyruvate data. **(C)** Succinate data. These carbon sources had the lowest correlation values and these single carbon sources are also found in the TCA cycle. The difference in directionality between the black bars and the prediction bars highlights that glycolysis (PGI flux to PYK-PPS flux) and TCA cycle (RPI flux to TK2 flux) are being predicted in the opposite direction from the measured flux.

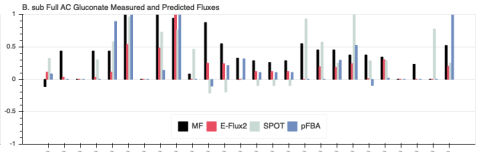

E. coli AC Acetate Measured and Predicted Fluxes
E. coli Full AC Acetate Measured and Predicted Fluxes
E. coli AC Fructose Measured and Predicted Fluxes
E. coli Full AC Fructose Measured and Predicted Fluxes
E. coli AC Galactose Measured and Predicted Fluxes
E. coli Full AC Galactose Measured and Predicted Fluxes
E. coli AC Gluconate Measured and Predicted Fluxes
E. coli Full AC Gluconate Measured and Predicted Fluxes
E. coli AC Glucose Measured and Predicted Fluxes
E. coli Full AC Glucose Measured and Predicted Fluxes
E. coli AC Glycerol Measured and Predicted Fluxes
E. coli Full AC Glycerol Measured and Predicted Fluxes
E. coli AC Pyruvate Measured and Predicted Fluxes
E. coli Full AC Pyruvate Measured and Predicted Fluxes
E. coli AC Succinate Measured and Predicted Fluxes
E. coli Full AC Succinate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

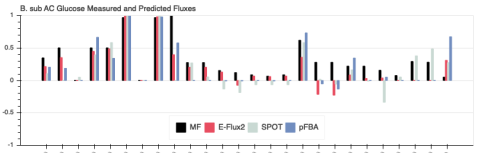
S4 Fig. *E. coli* flux directionality AC vs Full AC all carbon sources: Measured flux (black), E-Flux2 (red), SPOT (gray), pFBA (blue), and gene expression transcript abundance (green); all self-normalized to maximum values 1 and -1. Positive values reflect forward flux and negative values reflection reverse direction flux. The x-axis shows the measured central carbon flux reaction names.

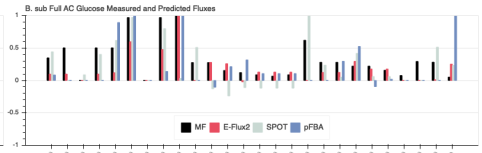

B. sub AC Fructose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

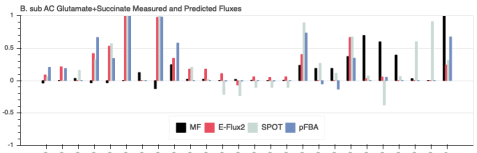

B. sub Full AC Fructose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

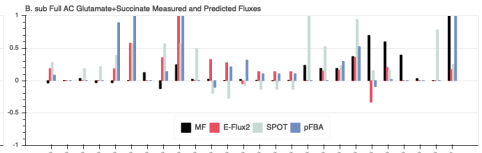

B. sub AC Gluconate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

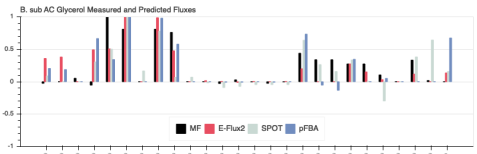

B. sub Full AC Gluconate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

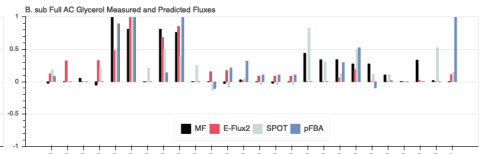

B. sub AC Glucose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

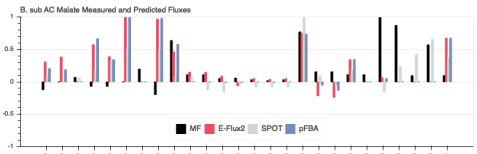

B. sub Full AC Glucose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

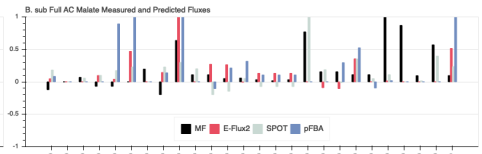

B. sub AC Glutamate+Succinate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

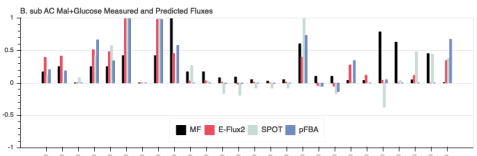

B. sub Full AC Glutamate+Succinate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

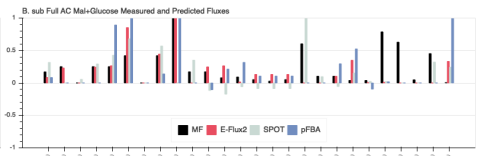

B. sub AC Glycerol Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

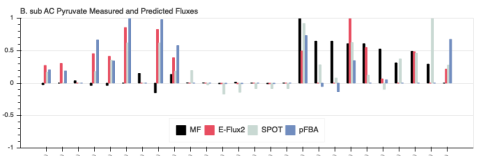

B. sub Full AC Glycerol Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

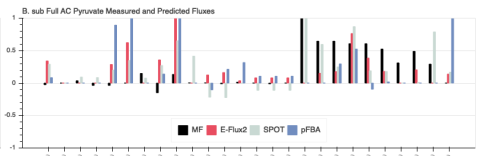

B. sub AC Malate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA


B. sub Full AC Malate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

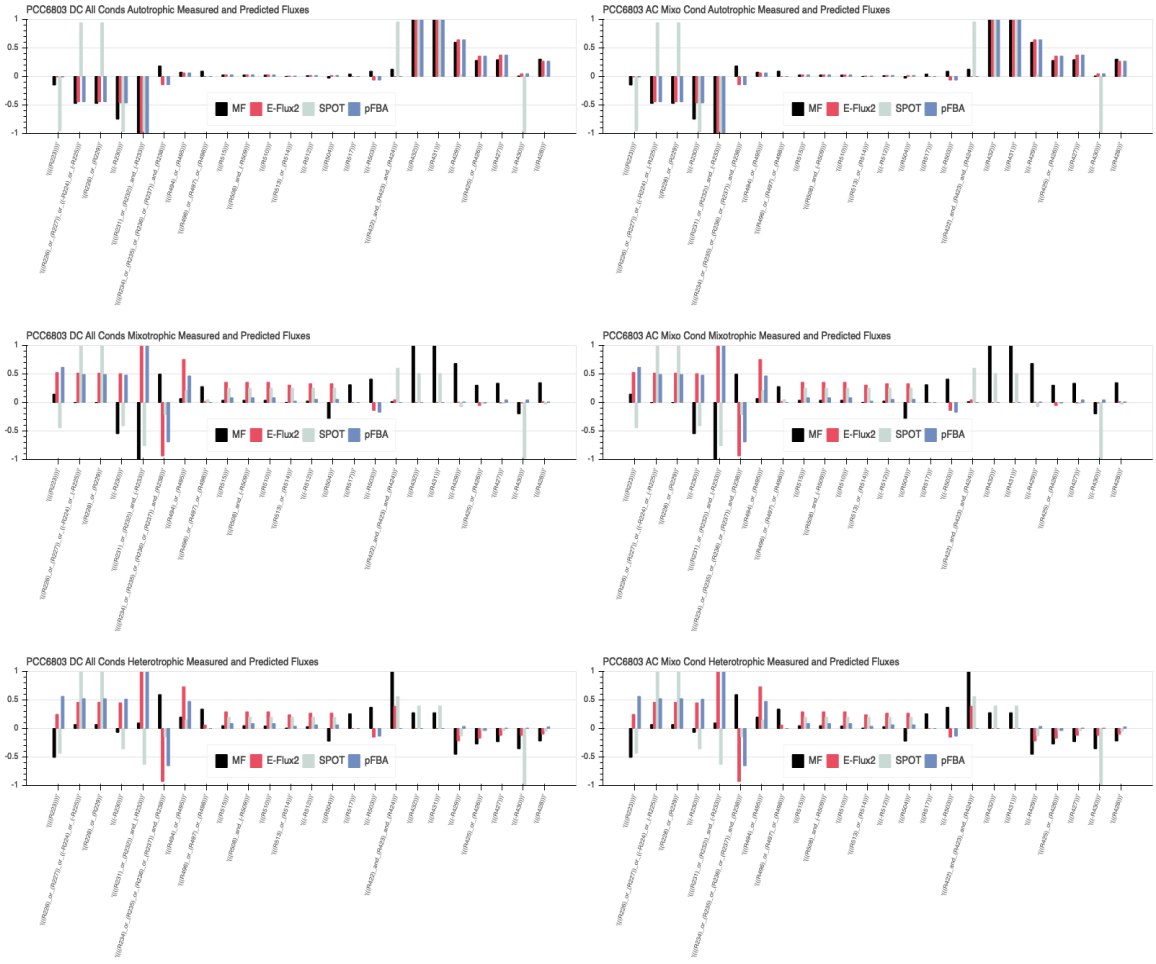

B. sub AC Mal+Glucose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA


B. sub Full AC Mal+Glucose Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

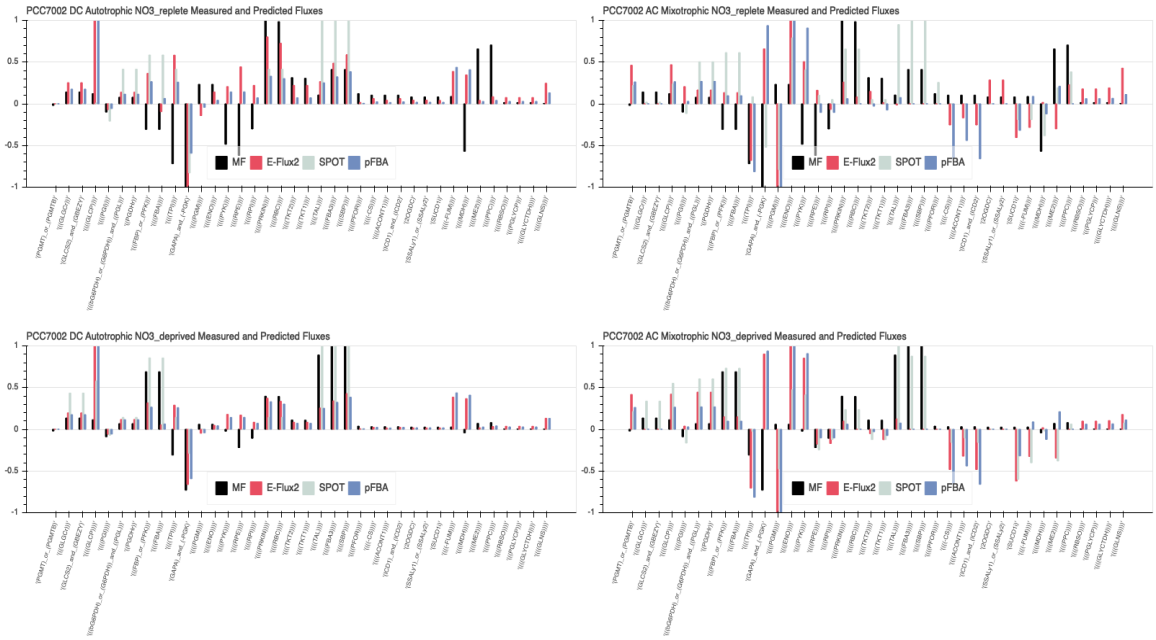

B. sub AC Pyruvate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA


B. sub Full AC Pyruvate Measured and Predicted Fluxes
MF
E-Flux2
SPOT
pFBA

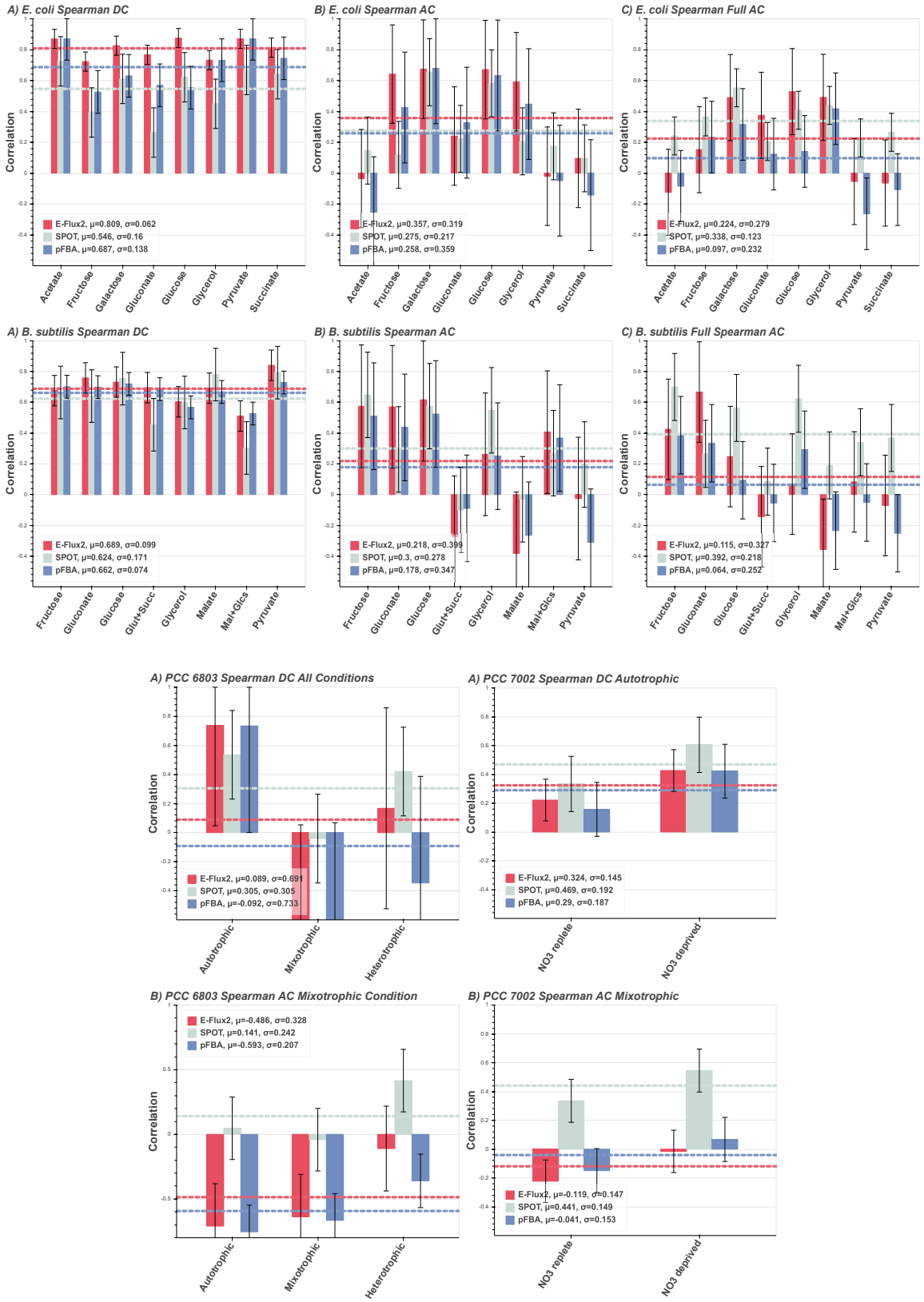
S5 Fig. *B. subtilis* flux directionality AC vs Full AC all carbon sources: Measured flux (black), E-Flux2 (red), SPOT (gray), pFBA (blue), and gene expression transcript abundance (green); all self-normalized to maximum values 1 and -1. Positive values reflect forward flux and negative values reflection reverse direction flux. The x-axis shows the measured central carbon flux reaction names.



S6 Fig PCC 6803 flux directionality DC and AC conditions: Measured flux (black), E-Flux2 (red), SPOT (gray), pFBA (blue), and gene expression transcript abundance (green); all self-normalized to maximum values 1 and -1. Positive values reflect forward flux and negative values reflection reverse direction flux. The x-axis shows the measured central carbon flux reaction names.



S7 PCC 7002 flux directionality DC and AC conditions: Measured flux (black), E-Flux2 (red), SPOT (gray), pFBA (blue), and gene expression transcript abundance (green); all self-normalized to maximum values 1 and -1. Positive values reflect forward flux and negative values reflection reverse direction flux. The x-axis shows the measured central carbon flux reaction names.


A) E. coli Spearman DC
E-Flux2, μ=0.809, σ=0.062
SPOT, μ=0.546, σ=0.16
pFBA, μ=0.687, σ=0.138
B) E. coli Spearman AC
E-Flux2, μ=0.357, σ=0.319
SPOT, μ=0.275, σ=0.217
pFBA, μ=0.258, σ=0.359
C) E. coli Spearman Full AC
E-Flux2, μ=0.224, σ=0.279
SPOT, μ=0.338, σ=0.123
pFBA, μ=0.097, σ=0.232
Acetate
Fructose
Galactose
Gluconate
Glucose
Glycerol
Pyruvate
Succinate
A) B. subtilis Spearman DC
E-Flux2, μ=0.689, σ=0.099
SPOT, μ=0.624, σ=0.171
pFBA, μ=0.662, σ=0.074
B) B. subtilis Spearman AC
E-Flux2, μ=0.218, σ=0.399
SPOT, μ=0.3, σ=0.278
pFBA, μ=0.178, σ=0.347
C) B. subtilis Full Spearman AC
E-Flux2, μ=0.115, σ=0.327
SPOT, μ=0.392, σ=0.218
pFBA, μ=0.064, σ=0.252
Fructose
Gluconate
Glucose
Glut+Succ
Glycerol
Malate
Mal+Glcs
Pyruvate
A) PCC 6803 Spearman DC All Conditions
E-Flux2, μ=0.089, σ=0.691
SPOT, μ=0.305, σ=0.305
pFBA, μ=-0.092, σ=0.733
Autotrophic
Mixotrophic
Heterotrophic
A) PCC 7002 Spearman DC Autotrophic
E-Flux2, μ=0.324, σ=0.145
SPOT, μ=0.469, σ=0.192
pFBA, μ=0.29, σ=0.187
NO3 replete
NO3 deprived
B) PCC 6803 Spearman AC Mixotrophic Condition
E-Flux2, μ=-0.486, σ=0.328
SPOT, μ=0.141, σ=0.242
pFBA, μ=-0.593, σ=0.207
B) PCC 7002 Spearman AC Mixotrophic
E-Flux2, μ=-0.119, σ=0.147
SPOT, μ=0.441, σ=0.149
pFBA, μ=-0.041, σ=0.153
Correlation

S8 Spearman correlations equivalents of main figures 1-4: Figures 1 through 4 from the main paper have been recreated using another correlation calculation, the Spearman rank correlation. The correlations are similar to the Pearson correlation from the main paper but a shift of lower correlation between measured vs predicted flux.

E. coli								
Physiology	**Acetate**	**Fructose**	**Galactose**	**Glucose**	**Glycerol**	**Gluconate**	**Pyruvate**	**Succinate**
Growth rate (h^{-1})	0.29 ± 0.01	0.49 ± 0.01	0.18 ± 0.01	0.65 ± 0.01	0.49 ± 0.01	0.59 ± 0.01	0.39 ± 0.01	0.51 ± 0.01
Uptake rate carbon source (mmol gCDW-1 h-1)	13.58 ± 0.33	8.33 ± 0.29	1.97 ± 0.10	9.65 ± 0.04	10.14 ± 0.15	7.28 ± 0.03	26.71 ± 0.79	15.90 ± 0.31
Acetate secretion (mmol gCDW-1 h-1)	-	3.33 ± 0.33	-	6.83 ± 1.03	0.60 ± 0.20	5.00 ± 0.16	11.91 ± 0.53	3.32 ± 0.31
Lactate secretion (mmol gCDW-1 h-1)	-	-	-	-	-	-	1.16 ± 0.07	-
Fumarate secretion (mmol $gCDW^{-1}$ h^{-1})	-	-	-	-	-	-	-	1.14 ± 0.07
Biomass yield (g g-1)	0.35 ± 0.00	0.33 ± 0.01	0.46 ± 0.03	0.37 ± 0.00	0.47 ± 0.01	0.41 ± 0	0.16 ± 0.01	0.26 ± 0.01

B. subtilis								
Specific Rate	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**	**(mmol g-1 h-1)**
glucose	7.36	-	-	-	-	-	5.95	-
Fructose	-	5.72	-	-	-	-	-	-
gluconate	-	-	5.13	-	-	-	-	-
Succinate	-	-	-	3.35	-	-	-	-
glutamate	-	-	-	2.21	-	-	-	-
Glycerol	-	-	-	-	6.22	-	-	-
malate	-	-	-	-	-	26.51	14.6	-
pyruvate	-	-	-	-	-	-	-	8.26

PCC 6803			
Specific Rate	**Autotroph mmol/g DCW/h**	**Mixotroph mmol/g DCW/h**	**Heterotroph mmol/g DCW/h**
glucose	0	0.24	0.41
CO2	0	1000	1000
HCO3	3.7	0	0
Photons	1000	1000	50

PCC 7002		
Specific Rate	**NO3 replete (mmol g-1 h-1)**	**NO3 depleted (mmol g-1 h-1)**
CO2	-29.5	-9.67
NO3	-1000	0

Table S1. Measured uptake rates: Measured uptake rates for the four organism models (E. coli, B. subtilis, PCC 6803, and PCC7002), sourced from literature (See Materials and Methods 2.1). For *E. coli,* mean values and standard deviations were obtained from 3 biological replicates. Any uptake rate set to 1000 is based on information indicating unconstrained uptake of the respective metabolite. In PCC 6803, the light inhibited culture was PSII chemically inhibited and cultured under aluminum foil. In an attempt to realistically constrain this experimental environment and chemical inhibition on the light reaction, 50 photon units were applied rather than 0. Additionally, this culture was grown in open air and therefore the CO_2 intake is left unconstrained. All uptake rates are present in units of mmol/g DCW/h.

Rxn#	Rxn Metabolites	E-Flux2	SPOT	pFBA
R419	Glycerone phosphate + D-Erythrose 4-phosphate --> Sedoheptulose 1,7-bisphosphate	0.0000	0.1922	0.0000
R420	Glycerone phosphate + D-Erythrose 4-phosphate --> Sedoheptulose 1,7-bisphosphate	0.0000	0.0000	0.0000
R421	H2O + Sedoheptulose 1,7-bisphosphate --> Orthophosphate + Sedoheptulose 7-phosphate	0.0000	0.1922	0.0000
R422	D-Glucose 6-phosphate + NADP+ --> D-Glucono-1,5-lactone 6-phosphate + NADPH + H+	0.0000	0.0225	0.0000
R423	D-Glucono-1,5-lactone 6-phosphate + H2O --> 6-Phospho-D-gluconate	0.0000	0.0225	0.0000
R424	6-Phospho-D-gluconate + NADP+ --> NADPH + H+ + CO2 + D-Ribulose 5-phosphate	0.0000	0.0225	0.0000
R425	D-Ribose 5-phosphate <==> D-Ribulose 5-phosphate	0.0030	0.0002	-0.0031
R426	D-Ribose 5-phosphate <==> D-Ribulose 5-phosphate	0.0030	0.0002	-0.0031
R427	D-Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate <==> D-Xylulose 5-phosphate + D-Ribose 5-phosphate	0.0099	0.0004	0.0011
R428	D-Glyceraldehyde 3-phosphate + D-Fructose 6-phosphate <==> D-Xylulose 5-phosphate + D-Erythrose 4-phosphate	-0.0004	0.0004	0.0050
R429	D-Ribulose 5-phosphate <==> D-Xylulose 5-phosphate	0.0034	-0.0007	-0.0061
R430	D-Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate <==> D-Fructose 6-phosphate + D-Erythrose 4-phosphate	-0.0099	0.1918	-0.0011

Table S2. PCC 6803 Pentose Phosphate Pathway Fluxes: Predicted fluxes for 12 pentose phosphate pathway reactions for PCC 6803 under the heterotrophic growth condition and using heterotrophic gene expression data. Fluxes vectors were normalized by the l2 norm of the genome scale flux vector. Fluxes marked yellow show net flux values that are essentially zero flux (fluxes less than 1e-6). Green shows the net positive flux values. Red shows net negative flux values predicted, suggesting predicting the reversal of the pentose phosphate pathway. In SPOT, higher flux is passing through the pentose phosphate reactions, while in the E-Flux2 and pFBA show significantly less flux.

6 References

1. Beale DJ, Karpe AV, Ahmed W. Beyond Metabolomics: A Review of Multi-Omics-Based Approaches. In: Beale DJ, Kouremenos KA, Palombo EA, editors. Microbial Metabolomics: Applications in Clinical, Environmental, and Industrial Microbiology. Cham: Springer International Publishing; 2016. p. 289-312.

2. Kim MK, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. Computational and structural biotechnology journal. 2014;11(18):59-65.

3. Machado D, Herrgard M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. PLoS Comput Biol. 2014;10(4):e1003580.

4. Kim MK, Lane A, Kelley JJ, Lun DS. E-Flux2 and SPOT: Validated Methods for Inferring Intracellular Metabolic Flux Distributions from Transcriptomic Data. PLoS One. 2016;11(6):e0157101.

5. Peregrin-Alvarez JM, Sanford C, Parkinson J. The conservation and evolutionary modularity of metabolism. Genome Biol. 2009;10(6):R63.

6. Rodrigues F, Ludovico P, Leão C. Sugar Metabolism in Yeasts: an Overview of Aerobic and Anaerobic Glucose Catabolism. In: Péter G, Rosa C, editors. Biodiversity and Ecophysiology of Yeasts. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 101-21.

7. Borkowski O, Goelzer A, Schaffer M, Calabre M, Mader U, Aymerich S, et al. Translation elicits a growth rate-dependent, genome-wide, differential protein production in Bacillus subtilis. Mol Syst Biol. 2016;12(5):870.

8. Brunk E, Chang RL, Xia J, Hefzi H, Yurkovich JT, Kim D, et al. Systemic post-translational control of bacterial metabolism regulates adaptation in dynamic environments. bioRxiv. 2017:180646.

9. Liu M, Durfee T, Cabrera JE, Zhao K, Jin DJ, Blattner FR. Global transcriptional programs reveal a carbon source foraging strategy by Escherichia coli. J Biol Chem. 2005;280(16):15921-7.

10. Otto A, Bernhardt J, Meyer H, Schaffer M, Herbst FA, Siebourg J, et al. Systems-wide temporal proteomic profiling in glucose-starved Bacillus subtilis. Nat Commun. 2010;1:137.

11. Tiwari DP, Chatterjee PM, Rotti H, Chand B, Raval R, Dubey AK. Expression dynamics of the poly-γ-glutamic acid biosynthesis genes of Bacillus subtilis in response to glucose and glutamic acid—a pilot study. FEMS Microbiology Letters. 2018;365(22).

12. You L, He L, Tang YJ. Photoheterotrophic Fluxome in *Synechocystis* sp. Strain PCC 6803 and Its Implications for Cyanobacterial Bioenergetics. Journal of Bacteriology. 2015;197(5):943-50.

13. Qian X, Kim MK, Kumaraswamy GK, Agarwal A, Lun DS, Dismukes GC. Flux balance analysis of photoautotrophic metabolism: Uncovering new biological details of subsystems involved in cyanobacterial photosynthesis. Biochimica et biophysica acta Bioenergetics. 2017;1858(4):276-87.

14. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology. 2010;6(1):390.

15. Crown SB, Antoniewicz MR. Publishing 13C metabolic flux analysis studies: a review and future perspectives. Metabolic engineerig. 2013;20:42-8.

16. Basler G, Fernie AR, Nikoloski Z. Advances in metabolic flux analysis toward genome-scale profiling of higher organisms. Biosci Rep.2018;38(6).

17. Gerosa L, Haverkorn van Rijsewijk BR, Christodoulou D, Kochanowski K, Schmidt TS, Noor E, et al. Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. Cell systems. 2015;1(4):270-82.

18. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular Systems Biology. 2011;7(1):535.

19. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, et al. Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in *Bacillus subtilis*. Science. 2012;335(6072):1103-6.

20. Chubukov V, Uhr M, Le Chat L, Kleijn RJ, Jules M, Link H, et al. Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. Molecular Systems Biology. 2013;9(1):709.

21. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. Journal of Biological Chemistry. 2007;282(39):28791-9.

22. You L, Berla B, He L, Pakrasi HB, Tang YJ. 13C-MFA delineates the photomixotrophic metabolism of Synechocystis sp. PCC 6803 under light- and carbon-sufficient conditions. Biotechnology journal. 2014;9(5):684-92.

23. Young JD, Shastri AA, Stephanopoulos G, Morgan JA. Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis. Metabolic engineering. 2011;13(6):656-65.

24. Knoop H, Gründel M, Zilliges Y, Lehmann R, Hoffmann S, Lockau W, et al. Flux Balance Analysis of Cyanobacterial Metabolism: The Metabolic Network of Synechocystis sp. PCC 6803. PLOS Computational Biology. 2013;9(6):e1003081.

25. Ludwig M, Bryant D. Acclimation of the Global Transcriptome of the Cyanobacterium Synechococcus sp. Strain PCC 7002 to Nutrient Limitations and Different Nitrogen Sources. Frontiers in Microbiology. 2012;3(145).

26. Qian X, Zhang Y, Lun DS, Dismukes GC. Rerouting of Metabolism into Desired Cellular Products by Nutrient Stress: Fluxes Reveal the Selected Pathways in Cyanobacterial Photosynthesis. ACS Synthetic Biology. 2018;7(5):1465-76.

27. Kelley JJ, Lane A, Li X, Mutthoju B, Maor S, Egen D, et al. MOST: a software environment for constraint-based metabolic modeling and strain design. Bioinformatics. 2014;31(4):610-1.

28. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. Crit Care. 2003;7(6):451-9.

29. Talaat AM, Lyons R, Howard ST, Johnston SA. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(13):4602-7.

30. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. Journal of Translational Medicine. 2017;15(1):73.

31. Reyes JC, Crespo JL, Garcia-Dominguez M, Florencio FJ. Electron Transport Controls Glutamine Synthetase Activity in the Facultative Heterotrophic Cyanobacterium Synechocystis sp. PCC 6803. Plant Physiology. 1995;109(3):899-905.

32. Jouhten P, Rintala E, Huuskonen A, Tamminen A, Toivari M, Wiebe M, et al. Oxygen dependence of metabolic fluxes and energy generation of Saccharomyces cerevisiae CEN.PK113-1A. BMC Syst Biol. 2008;2:60.

33. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. Microbiology. 1979;111(1):1-61.