Medicon Microbiology
Volume 1 Issue 4 May 2022
Research Article

Genome-Scale Metabolic Model-Based Reactome-Phenome Map of Synechocystis sp. PCC 6803, A Potential Biofuel Producer

Yik Yew Wee, Xander Kng, Si Xian Sor and Maurice HT Ling*

School of Applied Sciences, Temasek Polytechnic, Singapore

***Corresponding Author:** Maurice HT Ling, School of Applied Sciences, Temasek Polytechnic, Singapore.

Received: April 05, 2022; **Published:** April 21, 2022

Abstract

Synechocystis sp. PCC 6803 is a potential producer of lipids, alcohols, and biofuels. Genome-scale models (GSM) has been used to examine potential knockout to optimize specific metabolite (such as, ethanol) production. Besides from a metabolic production perspective, GSMs can also be used examine the effects of genes from the perspective of genotype-phenotype relationship. However, most GSMs are reaction-based rather than gene-based. Hence, GSMs can be used for reactome-phenome mapping where each reaction may be the result of one or more genes. In this study, we examine the reactome-phenome map of *Synechocystis sp.* PCC 6803 using its GSM model, iJN678, by performing single knockouts to each of its 863 reactions. Our results suggest that 37.3% to 39.7% (322 to 343 reactions) of the knockouts have minimal impact on the phenome as they were clustered together with wildtype phenotype and 53.5% (462 reactions) are essential. The rest of the 58 to 79 reactions can be clustered into 9 to 33 phenotypic clusters. Moreover, the fluxome variation within wildtype cluster is significantly larger than that of essential reaction cluster ($t \geq 3.26$, p-value $\leq$ 1.3E-3). This suggests that individual reaction knockout may have measurable effects on the fluxes; which may be useful in metabolic engineering.

Keywords: Genome-scale model; Genotype-phenotype map; Reactome-phenome map; Fluxome; Cameo; AdvanceSyn Toolkit; Flux-balance analysis

Introduction

The cyanobacterium *Synechocystis sp.* PCC 6803 [1] has been explored for used in biofuel production due to its phototrophic [2] property. This strain has been explored for lipid [3] and ethylene [4] production. Computational modelling and simulation are important to explore suitability of organisms and evaluate engineering approaches to increase production of biofuels [5-8]. Genome-scale metabolic models (GSMs), which is based on steady-states of metabolites [9], have been used inform many metabolic engineering requirements [10, 11]. For example, Zhang et al. [12] used GSM to examine bottlenecks in ethanol production by *Caldicellulosiruptor bescii* while Nguyen and Lee [13] used GSM to design improvement strategies to increase the conversion of methane to putrescine by *Methylomicrobium alcaliphilum*.

Underpinning these computational approaches is the relationship between genotype and phenotype, commonly known as genotype-phenotype relationship [14-17], where genomic perturbations (such as knockouts) results in changes in the fluxome. Fluxome can be defined as the set of metabolite conversion rates in a metabolic network [18-20]. This results in changes in the metabolome, leading to phenotypic changes. Conversely, GSMs can be useful to identify genotype-phenotype relationships [21].

In this study, we aim to elucidate the genotype-phenotype relationship of *Synechocystis sp.* PCC 6803 using its GSM, iJN678 [22]. However, GSMs are based on reaction stoichiometries [23]; hence, reaction is the atomic unit rather than gene [24]. Therefore, GSM-

based reactome-phenome *Synechocystis sp.* PCC 6803 is elucidated where each reaction is encoded by one or more genes. Our results show that 37.3% to 39.7% (322 to 343 reactions) of the knockouts have minimal impact on the phenome and 53.5% (462 reactions) are essential with the rest of the 58 to 79 reactions clustered into 9 to 33 phenotypic clusters.

Materials and Methods

GSM for *Synechocystis sp.* PCC 6803, iJN678 [22], was obtained from BiGG database [25]. Growth rate on native media given as proxy as output from the objective function [26] and fluxes after flux balance analysis [27] using Cameo [28], which was available via `cameo-fba` command from AdvanceSyn Toolkit [29]. The entire set of predicted fluxes obtained from a GSM is known as a predicted fluxome. Single reaction knockouts [21] were performed using `cameo-mutant-fba` command. The number of phenotypic clusters were determined by elbow method on distortion score [30], Davies-Bouldin score [31-33], and Silhouette index [34] from SeqProperties [35]; on fluxes from each reaction knockout. Variations within a cluster was calculated as root mean square error (RMSE) between a predicted fluxome and an average fluxome (defined as the set of average fluxes of each reaction within the cluster) of the same cluster which had been used in previous genomic studies [36, 37].

Results and Discussion

GSM for *Synechocystis sp.* PCC 6803, iJN678 [22], was obtained from BiGG database [25]. The model, which is based on Accession number BA000022.2, consists of 622 genes, 795 metabolites, and 863 reactions. Predicted fluxes from individual reaction knockout, defined as deduction of the corresponding reaction flux to zero, were used to analyse reactome-phenome mapping.

Our knockout simulation results show that the knockouts in iJN678 [22] were clustered into 11 clusters using elbow method on distortion score [30] (Figure 1A and Table 1), 29 clusters using Silhouette index [34] (Figure 1B and Table 1), and 35 clusters using Davies-Bouldin score [31-33] (Figure 1B and Table 1). However, 37.3% to 39.7% (322 to 343 reactions) of the knockouts have minimal impact on the phenome as they were clustered together with wildtype phenotype. This may be explained by gene compensation [38-41] and / or gene redundancy [42-44]. Gene redundancy is when there is more than one gene for the same function, which are commonly resulted from gene duplication [45]; whereas gene compensation is when other genes, which may not have the same functions, can compensate for the effects of a gene deletion or mutation. This suggests that there may be substantial compensatory or redundancy in iJN678 [22]. On the other hand, 53.5% (462 reactions) resulted in no fluxes suggesting that knockout of these reactions resulted in lethality; thus, essential reactions. This proportion of essential genes within the genome is within the range of prokaryotes surveyed by Gerdes et al [46]. Collectively, these two clusters account for 90.8% to 93.3% of the knockouts, suggesting that single reaction knockouts is unlikely to result in substantial changes in the fluxome.

Of the 6.72% (58 reactions) to 9.15% (79 reactions) not clustered into wildtype or no flux (essential reaction) cluster, they can be clustered into 9 to 33 phenotypic clusters (Table 1). By combining the remaining 6.72% (58 reactions) to 9.15% (79 reactions) into a single "other phenome" cluster (denoted as "Others" in Table 2), our results show that the mean RMSEs of the three clusters are significant different using all three measures; namely, Elbow [30] ($F = 435$, p-value = 1.5E-131), Silhouette [34] ($F = 427$, p-value = 1.1E-129), and Davies-Bouldin [31-33] ($F = 336$, p-value = 1.4E-103). Pairwise t-test (Table 2) suggests that the mean RMSE of any two clusters are significantly different ($t \geq 3.26$, p-value $\leq$ 1.3E-3). Hence, the mean RMSE of others cluster is significantly larger than the mean RMSE of wildtype cluster ($t \geq 8.08$, p-value $\leq$ 4.1E-11), and the mean RMSE of wildtype cluster is significantly larger than the mean RMSE of no flux (essential reaction) cluster ($t \geq 3.26$, p-value $\leq$ 1.3E-3). This suggests that although 37.3% to 39.7% (322 to 343 knockouts) of the reaction knockouts are clustered together with wildtype, individual reaction knockout may have measurable effects on the fluxes; which may be useful in metabolic engineering [10, 11].

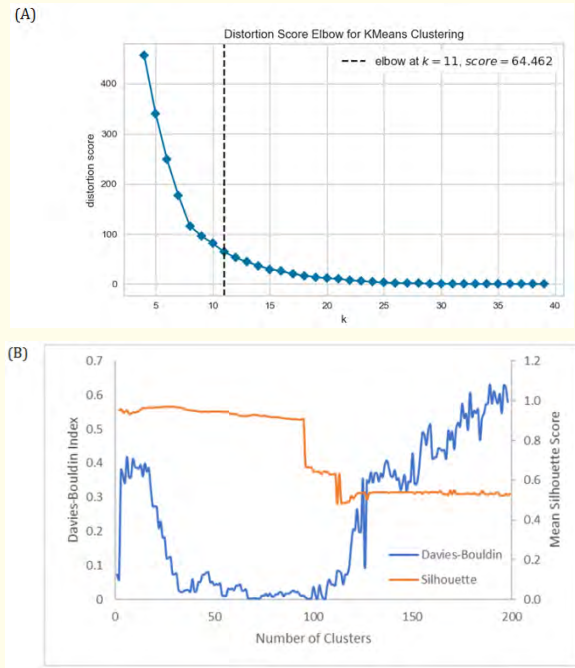

Figure 1: Cluster Analysis of Reaction knockouts in iJN678. Panel A shows results from elbow method using distortion score. Panel B shows the Davis-Bouldin index and mean Silhouette score.

Cluster	***Number of Reactions***		
	Elbow [30]	***Silhouette*** [34]	***Davies-Bouldin*** [31-33]
Wildtype	343	325	322
No Flux	462	462	462
Others-1	2	2	2
Others-2	4	1	1
Others-3	5	5	4
Others-4	3	1	1
Others-5	38	3	3
Others-6	3	3	3
Others-7	1	2	2
Others-8	2	1	1
Others-9	1	1	1
Others-10	0	1	1
Others-11	0	2	2
Others-12	0	29	28
Others-13	0	11	10
Others-14	0	1	1

Others-15	0	1	1
Others-16	0	1	1
Others-17	0	1	1
Others-18	0	2	1
Others-19	0	1	1
Others-20	0	1	1
Others-21	0	1	1
Others-22	0	1	1
Others-23	0	1	1
Others-24	0	1	1
Others-25	0	1	1
Others-26	0	1	1
Others-27	0	1	1
Others-28	0	0	2
Others-29	0	0	1
Others-30	0	0	1
Others-31	0	0	1
Others-32	0	0	1
Others-33	0	0	1

Table 1: Number of Reactions in Each Cluster for iJN678.

Clustering Method	***Comparison***	***df***	***RMSE***		***Statistical Analysis***	
			Mean	***Variance***	***t-Statistic***	***p-value***
Elbow[30]	Wildtype vs No Flux	343	3.3E-3 vs 2.7E-6	8.8E-5 vs 1.5E-12	6.44	4.05E-10*
	Wildtype vs Others	59	3.3E-3 vs 9.2E-2	8.8E-5 vs 7.2E-3	8.08	4.06E-11*
	No Flux vs Others	59	2.7E-6 vs 9.2E-2	1.5E-12 vs 7.2E-3	8.38	1.23E-11*
Silhouette [34]	Wildtype vs No Flux	325	2.5E-4 vs 6.9E-6	1.5E-6 vs 1.5E-12	3.72	2.31E-4*
	Wildtype vs Others	77	2.5E-4 vs 8.1E-2	1.5E-6 vs 6.1E-3	9.13	6.63E-14*
	No Flux vs Others	77	6.9E-6 vs 8.1E-2	1.5E-12 vs 6.1E-3	9.16	5.84E-14*
Davies-Bouldin [31-33]	Wildtype vs No Flux	172	8.6E-5 vs 2.7E-6	1.1E-7 vs 1.5E-12	3.26	1.33E-3*
	Wildtype vs Others	80	8.6E-5 vs 7.9E-2	1.1E-7 vs 6.0E-3	9.21	3.44E-14*
	No Flux vs Others	80	2.7E-6 vs 7.9E-2	1.5E-12 vs 6.0E-3	9.22	3.29E-14*

Table 2: Comparison of Clusters using RootMean Square Error (RMSE).Asterisk in p-value denote significance (p-value < 0.05).

Conclusion

GSM iJN678 based reactome-phenome mapping results suggest that 37.3% to 39.7% (322 to 343 reactions) of the knockouts have minimal impact on the phenome and 53.5% (462 reactions) are essential with the rest of the 58 to 79 reactions clustered into 9 to 33 phenotypic clusters. Fluxome variation within the wildtype cluster suggests that individual reaction knockout may have measurable effects on the fluxes.

Supplementary Materials

Data files for this study can be downloaded from https://bit.ly/Reactome-Phenome-iJN678.

Acknowledgement

This work was conducted as part of Major Project in Temasek Polytechnic School of Applied Science under the Student Project Fund (TP_PR1052).

Conflict of Interest

The authors declare no conflict of interest.

References

1. Ikeuchi M and Tabata S. "Synechocystis sp. PCC 6803 - A Useful Tool in the Study of the Genetics of Cyanobacteria". Photosynthesis Research 70.1 (2001):73-83.
2. Yu Y., et al. "Development of Synechocystis sp. PCC 6803 as a Phototrophic Cell Factory". Marine Drugs 11.8 (2013): 2894-2916.
3. Towijit U., et al. "Co-Over expression of Native Phospholipid-Biosynthetic Genes plsX and plsC Enhances Lipid Production in Synechocystis sp. PCC 6803". Scientific Reports 8.1 (2018): 13510.
4. Veetil VP., et al. "Ethylene Production with Engineered Synechocystis sp PCC 6803 Strains". Microbial Cell Factories 16.1 (2017): 34.
5. Jin X., et al. "Modeling, Design Guidelines, and Detection Limits of Self-Powered Enzymatic Biofuel Cell-Based Sensors". Biosensors & Bioelectronics 168 (2020): 112493.
6. Leow S., et al. "A Unified Modeling Framework to Advance Biofuel Production from Microalgae". Environmental Science & Technology 52.22 (2018): 13591-13599.
7. Ali H., et al. "CFD and Kinetic-Based Modeling to Optimize the Sparger Design of a Large-Scale Photobioreactor for Scaling up of Biofuel Production". Biotechnology and Bioengineering 116.9 (2019): 2200-2211.
8. Pham N., et al. "Genome-Scale Metabolic Modeling Underscores the Potential of Cutaneotrichosporon *oleaginosus* ATCC 20509 as a Cell Factory for Biofuel Production". Biotechnology for Biofuels 14.1 (2021): 2.
9. van Rosmalen RP., et al. "Model Reduction of Genome-Scale Metabolic Models as a Basis for Targeted Kinetic Models". Metabolic Engineering 64 (2021): 74-84.
10. Xu C., et al. "Genome-Scale Metabolic Model in Guiding Metabolic Engineering of Microbial Improvement". Applied Microbiology and Biotechnology 97.2 (2013): 519-539.
11. Kim B., et al. "Applications of Genome-Scale Metabolic Network Model in Metabolic Engineering". Journal of Industrial Microbiology & Biotechnology 42.3 (2015): 339-348.
12. Zhang K., et al. "Genome-Scale Metabolic Model of Caldicellulosiruptor bescii Reveals Optimal Metabolic Engineering Strategies for Bio-based Chemical Production". mSystems 6.3 (2021): e0135120.
13. Nguyen LT and Lee EY. "Biological Conversion of Methane to Putrescine Using Genome-Scale Model-Guided Metabolic Engineering of a Methanotrophic Bacterium Methylomicrobium alcaliphilum 20Z". Biotechnology for Biofuels 12 (2019): 147.
14. Li BT., et al. "Analyzing Transcriptome-Phenotype Correlations". Encyclopedia of Bioinformatics and Computational Biology (Elsevier) (2019): 819-824.
15. Jakutis G and Stainier DYR. "Genotype–Phenotype Relationships in the Context of Transcriptional Adaptation and Genetic Robustness". Annual Review of Genetics 55.1 (2021): 71-91.
16. Kemble H., et al. "Recent Insights into the Genotype-Phenotype Relationship from Massively Parallel Genetic Assays". Evolutionary Applications 12.9 (2019): 1721-1742.
17. Kabimoldayev I., et al. "Basics of Genome-Scale Metabolic Modeling and Applications on C1-Utilization". FEMS Microbiology Letters 365.20 (2018): fny241.
18. Cortassa S., et al. "From Metabolomics to Fluxomics: A Computational Procedure to Translate Metabolite Profiles into Metabolic Fluxes". Biophysical Journal 108.1 (2015): 163-172.

19. Winter G and Krömer JO. "Fluxomics - Connecting Omics Analysis and Phenotypes". Environmental Microbiology 15.7 (2013): 1901-1916.
20. Babele PK and Young JD. "Applications of Stable Isotope-Based Metabolomics and Fluxomics Toward Synthetic Biology of Cyanobacteria". Wiley Interdisciplinary Reviews Systems Biology and Medicine 12.3 (2020): e1472.
21. Castillo S., et al. "Yeast Genome-Scale Metabolic Models for Simulating Genotype-Phenotype Relations". Progress in Molecular and Subcellular Biology 58 (2019): 111-133.
22. Nogales J., et al. "Detailing the Optimality of Photosynthesis in Cyanobacteria Through Systems Biology Analysis". Proceedings of the National Academy of Sciences 109.7 (2012): 2678-2683.
23. Brochado AR., et al. "Impact of Stoichiometry Representation on Simulation of Genotype-Phenotype Relationships in Metabolic Networks". PLoS Computational Biology 8.11 (2012): e1002758.
24. Contador CA., et al. "Use of Genome-Scale Models to Get New Insights into the Marine Actinomycete Genus Salinispora". BMC Systems Biology 13.1 (2019): 11.
25. King ZA., et al. "BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models". Nucleic Acids Research 44.D1 (2016): D515-D522.
26. Feist AM and Palsson BO. "The Biomass Objective Function". Current Opinion in Microbiology 13.3 (2010): 344-349.
27. Orth JD., et al. "What is Flux Balance Analysis?" Nature biotechnology 28.3 (2010): 245-248.
28. Cardoso JGR., et al. "Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories". ACS Synthetic Biology 7.4 (2018): 1163-1166.
29. Ling MH. "AdvanceSyn Toolkit: An Open Source Suite for Model Development and Analysis in Biological Engineering". MOJ Proteomics & Bioinformatics 9.4 (2020): 83-86.
30. Yuan C and Yang H. "Research on K-Value Selection Method of K-Means Clustering Algorithm". J 2.2 (2019): 226-235.
31. Davies DL and Bouldin DW. "A Cluster Separation Measure". IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI 1.2 (1979): 224-227.
32. Mughnyanti M., et al. "Analysis of Determining Centroid Clustering X-Means Algorithm with Davies-Bouldin Index Evaluation". IOP Conference Series: Materials Science and Engineering 725.1 (2020): 012128.
33. Xiao J., et al. "Davies Bouldin Index Based Hierarchical Initialization K-Means". Intelligent Data Analysis 21.6 (2017): 1327-1338.
34. Wang X and Xu Y. "An Improved Index for Clustering Validation Based on Silhouette Index and Calinski-Harabasz Index". IOP Conference Series: Materials Science and Engineering 569.5 (2019): 052024.
35. Ling MHT. "SeqProperties: A Python Command-Line Tool for Basic Sequence Analysis". Acta Scientific Microbiology 3.6 (2020): 103-106.
36. Hadi K., et al. "Distinct Classes of Complex Structural Variation Uncovered across Thousands of Cancer Genome Graphs". Cell 183.1 (2020): 197-210.
37. Park SC., et al. "Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size". Frontiers in Microbiology 10 (2019): 834.
38. Peng J. "Gene Redundancy and Gene Compensation: An Updated View". Journal of Genetics and Genomics 46.7 (2019): 329-333.
39. Buglo E., et al. "Genetic Compensation in a Stable slc25a46 Mutant Zebrafish: A Case for Using F0 CRISPR Mutagenesis to Study Phenotypes Caused by Inherited Disease". PLoS ONE 15.3 (2020): e0230566.
40. Barbaric I., et al. "Appearances Can Be Deceiving: Phenotypes of Knockout Mice". Briefings in Functional Genomics and Proteomics 6.2 (2007): 91-103.
41. El-Brolosy MA and Stainier DYR. "Genetic Compensation: A Phenomenon in Search of Mechanisms". PLoS Genetics 13.7 (2017): e1006780.
42. Nowak MA., et al. "Evolution of genetic redundancy". Nature 388.6638 (1997): 167-171.
43. Láruson ÁJ., et al. "The Importance of Genetic Redundancy in Evolution". Trends in Ecology & Evolution 35.9(2020):809-822.
44. Baker EA., et al. "Extensive Non-Redundancy in a Recently Duplicated Developmental Gene Family". BMC Ecology and Evolution 21.1 (2021): 33.

45. Conrad B and Antonarakis SE. "Gene Duplication: A Drive for Phenotypic Diversity and Cause of Human Disease". Annual Review of Genomics and Human Genetics 8 (2007): 17-35.
46. Gerdes S., et al. "Essential Genes on Metabolic Maps". Current Opinion in Biotechnology 17.5 (2006): 448-456.

Volume 1 Issue 4 May 2022