MDPI

Article

CASTLE: A database of synthetic lethal sets predicted from genome-scale metabolic networks

Vimaladhasan Senthamizhan [1,2], Sunanda Subramaniam [3], Arjun Raghavan [3,4], Lakshmi Chanemougam [5] and Karthik Raman [1,2,3,*]

[1] Centre for Integrative Biology and Systems Medicine, Chennai - 600 036, India
[2] Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), IIT Madras, Chennai - 600 036, India
[3] Bhupat and Jyoti Mehta School of Biosciences, Department of Biotechnology, Indian Institute of Technology (IIT) Madras, Chennai - 600 036, India
[4] Department of Biomedical Engineering, McMaster University, Hamilton, ON, Canada. Present address: University of Manitoba, Winnipeg, Canada
[5] School of Chemical and Biotechnology, Sastra University, Thanjavur - 613401, India
* Correspondence: kraman@iitm.ac.in

Abstract: Genome-scale metabolic networks have been reconstructed for hundreds of organisms over the last two decades, with wide-ranging applications, including the identification of drug targets. Constraint-based approaches such as flux balance analysis have been effectively used to predict single and combinatorial drug targets in various metabolic networks. We have previously developed Fast-SL, an efficient algorithm to rapidly enumerate all possible synthetic lethals from metabolic networks. Here, we introduce CASTLE, an online standalone database containing synthetic lethals predicted from the metabolic networks of over 110 organisms. These targets include single, double or triple lethal sets of genes and reactions, and have been predicted using the Fast-SL algorithm. These synthetic lethals also capture novel functional associations between genes. The workflow used for building CASTLE can be easily applied to other genome-scale models to identify novel functional associations/combinatorial targets. CASTLE database can be accessed using the following link: https://ramanlab.github.io/CASTLE/

Keywords: synthetic lethals; metabolic networks; flux balance analysis; lethal genes; lethal reactions

Citation: Senthamizhan. V; Subramaniam. S; Raghavan. A; Chanemougam. L; Raman. K CASTLE. *Metabolites* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

1. Introduction

The recent years have witnessed the reconstruction of genome-scale metabolic models for a wide variety of organisms [1,2], including pathogenic organisms, with applications in drug target identification and understanding disease aetiology [3–5]. These models capture all known reactions in an organism's metabolic network, and are particularly useful to predict the growth rate of an organism, and its phenotype upon various perturbations, especially the removal of one or more genes/reactions [6]. These networks have been studied using Flux Balance Analysis (FBA), which has been proven to accurately predict phenotypes following various genetic perturbations [7,8]. FBA can also be used to reliably predict synthetic lethal genes in metabolic networks, which in turn can be used to identify combinatorial targets for various pathogens [9,10]. These 'synthetic lethal sets' are sets of reactions/genes, in which only the simultaneous deletion of all reactions/genes in the set will abrogate growth of the organism (Figure 1).

Many algorithms have been developed to predict synthetic lethals in metabolic networks [11–13]. Fast-SL, previously developed in our laboratory, is as yet the most efficient parallel algorithm available to predict synthetic lethal sets in metabolic networks [14]. Fast-SL circumvents the computational complexity of various other methods through an iterative reduction of the search space for higher-order combinatorial targets.

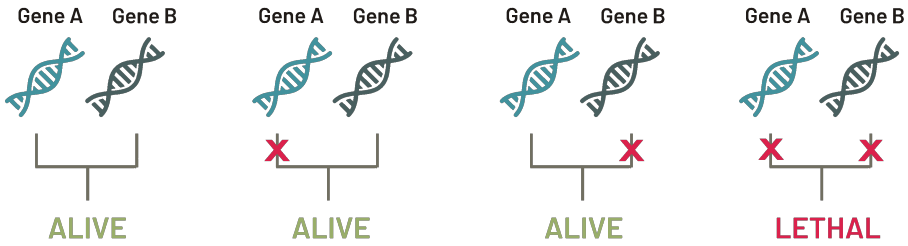

Figure 1. Schematic illustrating the logic of synthetic lethality. If there are two targets A and B, deleting any one of them does not affect the survival of the organism. Only the simultaneous deletion of both targets has a lethal effect on the organism. In this case, targets A and B are known as 'double lethals'. This logic can be extended to 'single lethals' and 'triple lethals'.

Many studies have used Fast-SL to predict synthetic lethals [15–17] and explore them as possible drug targets. To further enable the analysis of such combinatorial deletions across organisms, we used Fast-SL to computationally identify synthetic lethal sets for a variety of organisms. The results of the analyses have been compiled and published in CASTLE (Computational Analysis of SynThetic LEthals)–a standalone web database that we report here.

2. Methods

In order to identify synthetic lethal sets of genome-scale metabolic models, we used the parallel version of the Fast-SL program. A MATLAB script was written to interface the COBRA Toolbox [6] with the parallel Fast-SL programs and automate the process of identifying single, double, and triple reaction/gene lethal sets for a group of organisms. The metabolic model was first read using the COBRA toolbox; the parallel Fast-SL script was then employed to identify the lethals, which were written out as MAT files.

The genome-scale metabolic models themselves were selected and downloaded from the BiGG database [18] and Virtual Metabolic Human (VMH) database [19]. The output results on lethality information were saved in CSV format. These lethal sets were converted into three different formats (JSON, CSV, MAT) to offer users choice for the format in which they can view the data. This process was repeated for all the 113 organisms and the results were published in the database.

Interactive widgets have been created for each organism taken from the VMH database. The genes in each network are queried against the VMH API to provide further information. The networks are created using the NetworkX Python package [20], and the interactive widget is built by embedding the network in Bokeh 1.0.0 [21] environment. The HTML snippets created by the Bokeh environment are embedded into the individual pages of the organisms.

STRING-DB v11.5 [22] offers API functionalities to get interaction score for a set of genes/proteins in an organism. By the nature of gene IDs used in the metabolic models, BiGG database genes are readily identifiable by STRING-DB APIs. Double and Triple synthetic lethal genes were posted to the STRING-DB APIs to retrieve the interaction scores of the gene pairs. The API offers multiple scores like gene neighbourhood score, gene fusion score, experimental score and phylogenetic profile score along with an overall score value, which is obtained by adding the probabilities of all the scores. This overall score is retrieved from the API response for each gene pair and the distribution of the scores were plotted to visualise the interaction strength among the lethal genes. For triple lethal genes, average of the interaction scores between the gene pairs formed from the triplet was calculated and considered as the final interaction score of the triplet.

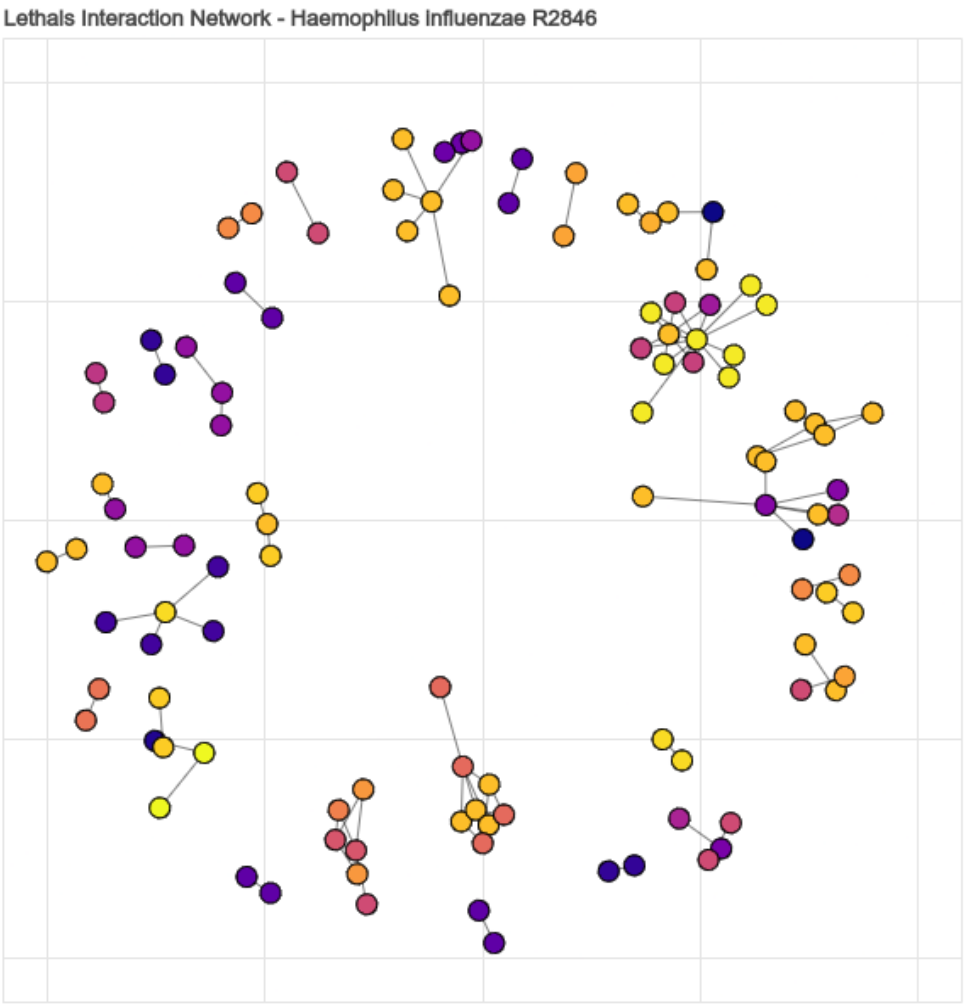

Figure 2. Interactive network of *Haemophilus influenzae* R2846. Each node in the network corresponds to a gene, and the edges connect other genes, which form a lethal set together. Hovering on a lethal gene node, the user can find information about the gene such as the gene name, function, KEGG ID and lethal type. The color of the node denotes the functional group to which the gene belongs. It can be noted from the figure that similar colored nodes form a cluster with each other - meaning that genes in a synthetic lethal set belong to one or closely related functional groups. Such interactive networks have been generated for all of the organisms taken from the VMH database.

3. Results

3.1. Database

Synthetic lethal sets (single, double and triple lethal reactions and genes) of 113 organisms were identified. This information was incorporated into the CASTLE database. On the home page, the number of each type of lethal set is listed, which can be selected to open further information about the reactions/genes. Further, a download icon is available, which, when selected, will automatically download all available results as a ZIP file.

The synthetic lethal sets for an organism contained some genes being repeated across multiple sets, suggesting that some genes are more prevalent in the lethality sets than others. To visualise this, an interactive network connecting each duplet or triplet of synthetic lethals was generated using NetworkX and Bokeh. The interactive widget shows the lethal sets in the form of a network where the lethal genes are depicted as nodes and the edges are connected to other genes that form a lethal set together.

The widget displays various information regarding the lethal gene, such as gene name, function, KEGG ID and lethal type, when hovered over the node.

A sample network of triple lethal genes belonging to *Haemophilus influenzae* R2846 is shown in Figure 2. These networks also represent novel functional linkages between the genes, many of which are hitherto unidentified in databases such as the STRING.

3.2. Functional redundancy in synthetic lethals

Metabolic networks have evolved to reduce the disruption of key metabolic pathways by the establishment of redundant genes/reactions [23]. These network visualisations substantiate that statement evidently. The genes in a lethal set usually belong to a single subsystem, denoted by the same coloured nodes in a cluster. Such a phenomenon suggests that lethal genes adopt functional redundancy to compensate for the disruption in key metabolic pathways in the event of external perturbations.

Further inspection of these networks revealed that some lethal sets contain genes involved in closely related functionalities. To give an example, most of the genes involved in purine (or) pyrimidine catabolism (e.g. KEGG IDs R02297, R01227, R01876) form lethal sets with genes involved in nucleotide salvage pathways (e.g. KEGG IDs - R00966, R00190, R00185). Such functional relations in lethal sets also indicate that synthetic lethality is able to unravel the inherent redundancy in metabolic networks.

3.3. Gene-gene interaction among the lethal sets

In order to further substantiate the functional associations in synthetic lethal genes, the gene-gene interaction scores among the gene pairs of double and triple lethal genes were retrieved using STRING-DB API. Fig 3 shows the distribution of overall interaction scores of the doublets and triplets.

The interaction scores of the gene pairs in synthetic lethal sets are skewed towards the right end of the distribution. STRING-DB's API normalizes the score to the range of 0 to 1. The distribution plots show that most of the gene pairs have scores close to 1.0, denoting there is substantial evidence that the lethal gene sets are functionally associated and interacting with each other. The scores of triple lethal genes are less skewed than its double lethal counterparts. One possible reason for this behaviour might the averaging of the gene pairs inside a triplet which might bring down the overall interaction score.

Only around half of the gene pairs in our synthetic lethal sets were identified by STRING-DB and the interaction scores were retrieved. But these scores suggest that the unidentified gene pairs might very well have functional associations that is yet to be discovered. Through these analyses, we would like to project the unidentified interactions as predicted functional associations and argue that the interactions between these gene pairs can be revealed through targeted experimental and computational studies.

4. Discussion

Synthetic lethality provides a conceptual framework for the development of cancer-specific cytotoxic agents. This paradigm has not been exploited in the past because there were no robust methods for systematically identifying synthetic lethal genes. CASTLE is built to address this issue where we identified synthetic lethal sets for 113 organisms using Fast-SL algorithm.

In addition to the identification of lethal sets, our analysis also sheds light into how metabolic networks espouse functional redundancy to cope up for any perturbations in the system. There seems to be substantial evidence that the synthetic lethals exhibit some form of functional association. Hence, we publish these synthetic lethals as gene-gene interaction predictions that can be confirmed in the future studies.

This database is envisaged to help researchers tackle the twin problem of drug resistance and drug side effects mentioned previously, as well as discover new therapeutic targets. Furthermore, the standardized procedure used can be easily extended to

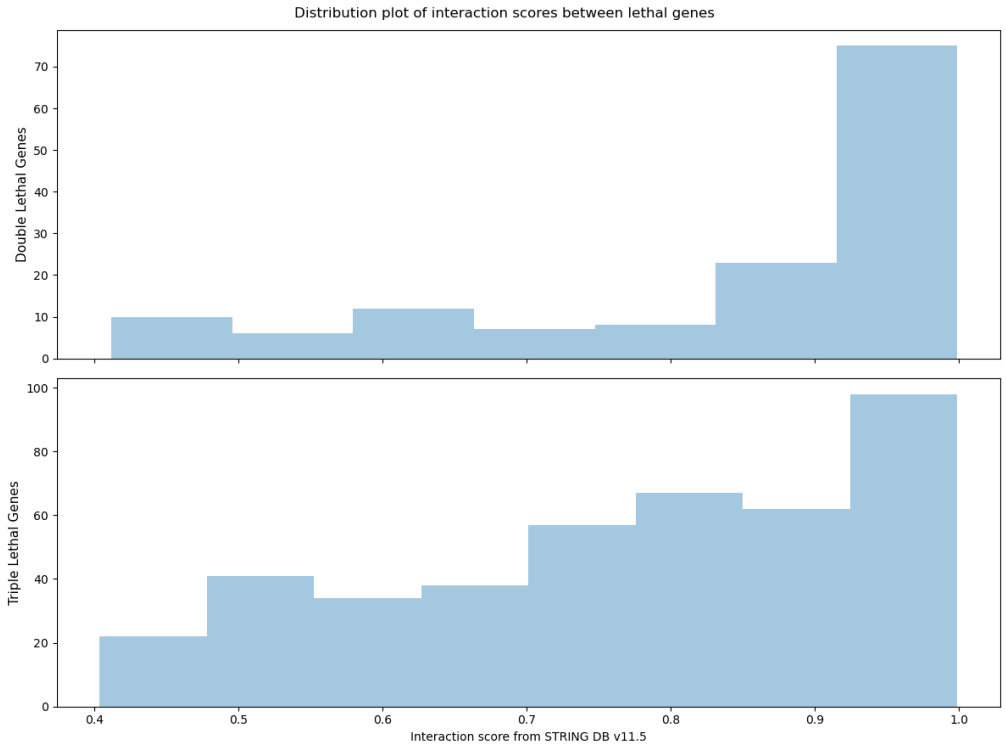

Figure 3. Distribution plots showing the distribution of interaction scores of double and triple lethal genes. The interaction scores range from 0 to 1 and it is a sum of probabilities of multiple scores like gene neighborhood, gene fusion, experimental score and phylogenetic profiles. The plot shows that most of the gene pairs has interaction scores close to 1.0 which means there is substantial evidence for the interactions among synthetic lethal gene pairs. It is to be noted that no threshold has been imposed on the scores and all the values retrieved from STRING-DB were above 0.4.

the models of other genome-scale models, which have not been created yet / have not been analyzed in this study. In the future versions of CASTLE, it is planned to identify synthetic lethals for all the organisms from VMH database and publish the results in our database.

Author Contributions: Conceptualization, K.R; methodology, S.S and A.R.; software, V.S.; validation, S.S. and A.R.; writing—original draft preparation, V.S.; writing—review and editing, V.S. and K.R. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: CASTLE is available at https://ramanlab.github.io/CASTLE/

Acknowledgments: VS acknowledges Centre for Integrative Biology and Systems Medicine (IBSE) for the post-baccalaureate fellowship. AR acknowledges funding from Mathematics of Information Technology and Complex Systems (MITACS) and support from Mr. Ryan Caldwell and others at MITACS.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fang, X.; Lloyd, C.J.; Palsson, B.O. Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications. *Nature Reviews. Microbiology* **2020**, *18*, 731. doi: 10.1038/s41579-020-00440-4.
2. O'Brien, E.J.; Monk, J.M.; Palsson, B.O. Using Genome-Scale Models to Predict Biological Capabilities. *Cell* **2015**, *161*, 971–987.
3. Beste, D.J.; Hooper, T.; Stewart, G.; Bonde, B.; Avignone-Rossa, C.; Bushell, M.E.; Wheeler, P.; Klamt, S.; Kierzek, A.M.; McFadden, J. GSMN-TB: a web-based genome-scale network model

of Mycobacterium tuberculosismetabolism. *Genome Biology* **2007**, *8*, R89. doi:10.1186/gb-2007-8-5-r89.
4. Raman, K.; Yeturu, K.; Chandra, N. targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. *BMC Systems Biology* **2008**, *2*, 109. doi:10.1186/1752-0509-2-109.
5. McCloskey, D.; Palsson, B.; Feist, A.M. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular Systems Biology* **2013**, *9*, 661. doi:10.1038/msb.2013.18.
6. Schellenberger, J.; Que, R.; Fleming, R.M.T.; Thiele, I.; Orth, J.D.; Feist, A.M.; Zielinski, D.C.; Bordbar, A.; Lewis, N.E.; Rahmanian, S.; Kang, J.; Hyduke, D.R.; Palsson, B. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols* **2011**, *6*, 1290–1307. doi:10.1038/nprot.2011.308.
7. Edwards, J.S.; Palsson, B.O. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics* **2000**, *1*, 1. doi:10.1186/1471-2105-1-1.
8. Famili, I.; Forster, J.; Nielsen, J.; Palsson, B.O. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 13134–13139. doi:10.1073/pnas.2235812100.
9. Navid, A. Applications of system-level models of metabolism for analysis of bacterial physiology and identification of new drug targets. *Briefings in Functional Genomics* **2011**, *10*, 354–364. doi:10.1093/bfgp/elr034.
10. Hartman, H.B.; Fell, D.A.; Rossell, S.; Jensen, P.R.; Woodward, M.J.; Thorndahl, L.; Jelsbak, L.; Olsen, J.E.; Raghunathan, A.; Daefler, S.; Poolman, M.G. Identification of potential drug targets in Salmonella enterica sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology* **2014**, *160*, 1252–1266. doi:10.1099/mic.0.076091-0.
11. Sinha, S.; Thomas, D.; Chan, S.; Gao, Y.; Brunen, D.; Torabi, D.; Reinisch, A.; Hernandez, D.; Chan, A.; Rankin, E.B.; Bernards, R.; Majeti, R.; Dill, D.L. Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. *Nature Communications* **2017**, *8*, 15580. doi:10.1038/ncomms15580.
12. Huang, J.; Wu, M.; Lu, F.; Ou-Yang, L.; Zhu, Z. Predicting synthetic lethal interactions in human cancers using graph regularized self-representative matrix factorization. *BMC Bioinformatics* **2019**, *20*, 657. doi:10.1186/s12859-019-3197-3.
13. Liu, L.; Chen, X.; Hu, C.; Zhang, D.; Shao, Z.; Jin, Q.; Yang, J.; Xie, H.; Liu, B.; Hu, M.; Ke, K. Synthetic Lethality-based Identification of Targets for Anticancer Drugs in the Human Signaling Network. *Scientific Reports* **2018**, *8*, 8440. Number: 1 Publisher: Nature Publishing Group, doi:10.1038/s41598-018-26783-w.
14. Pratapa, A.; Balachandran, S.; Raman, K. Fast-SL: an efficient algorithm to identify synthetic lethal sets in metabolic networks. *Bioinformatics* **2015**, *31*, 3299–3305. doi:10.1093/bioinformatics/btv352.
15. Stanway, R.R.; Bushell, E.; Chiappino-Pepe, A.; Roques, M.; Sanderson, T.; Franke-Fayard, B.; Caldelari, R.; Golomingi, M.; Nyonda, M.; Pandey, V.; Schwach, F.; Chevalley, S.; Ramesar, J.; Metcalf, T.; Herd, C.; Burda, P.C.; Rayner, J.C.; Soldati-Favre, D.; Janse, C.J.; Hatzimanikatis, V.; Billker, O.; Heussler, V.T. Genome-Scale Identification of Essential Metabolic Processes for Targeting the Plasmodium Liver Stage. *Cell* **2019**, *179*, 1112–1128.e26. doi:10.1016/j.cell.2019.10.030.
16. Devika, N.T.; Raman, K. Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models. *Scientific Reports* **2019**, *9*, 1–9. doi:10.1038/s41598-019-54696-9.
17. Rodenburg, S.Y.A.; Seidl, M.F.; Judelson, H.S.; Vu, A.L.; Govers, F.; Ridder, D.d. Metabolic Model of the Phytophthora infestans-Tomato Interaction Reveals Metabolic Switches during Host Colonization. *mBio* **2019**, *10*. Publisher: American Society for Microbiology Section: Research Article, doi:10.1128/mBio.00454-19.
18. King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **2016**, *44*, D515–522. doi:10.1093/nar/gkv1049.
19. Noronha, A.; Modamio, J.; Jarosz, Y.; Guerard, E.; Sompairac, N.; Preciat, G.; Daníelsdóttir, A.D.; Krecke, M.; Merten, D.; Haraldsdóttir, H.S.; Heinken, A.; Heirendt, L.; Magnúsdóttir, S.; Ravcheev, D.A.; Sahoo, S.; Gawron, P.; Friscioni, L.; Garcia, B.; Prendergast, M.; Puente,

A.; Rodrigues, M.; Roy, A.; Rouquaya, M.; Wiltgen, L.; Žagare, A.; John, E.; Krueger, M.; Kuperstein, I.; Zinovyev, A.; Schneider, R.; Fleming, R.M.; Thiele, I. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research* **2019**, *47*, D614–D624. doi:10.1093/nar/gky992.

20. Hagberg, A.A.; Schult, D.A.; Swart, P.J. Exploring Network Structure, Dynamics, and Function using NetworkX. Proceedings of the 7th Python in Science Conference; Varoquaux, G.; Vaught, T.; Millman, J., Eds.; , 2008; pp. 11 – 15.
21. Bokeh Development Team. *Bokeh: Python library for interactive visualization*, 2018.
22. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al.. String V11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* **2018**, *47*. doi:10.1093/nar/gky1131.
23. Sambamoorthy, G.; Raman, K. Understanding the evolution of functional redundancy in metabolic networks. *Bioinformatics* **2018**, *34*, i981–i987. doi:10.1093/bioinformatics/bty604.