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

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

[1]Initiative for Biological Systems Engineering, IIT Madras, 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

Abstract

Summary: 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 a variety of 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, which contains synthetic lethals predicted from the metabolic networks of over 130 organisms. These targets include single, double or triple lethal set of genes and reactions, and have been predicted using the Fast-SL algorithm. The workflow used for building CASTLE can be easily applied to other pathogenic models and used to identify novel therapeutic targets.
Availability: CASTLE is available at https://ramanlab.github.io/CASTLE/
Contact: kraman@iitm.ac.in

1 Introduction

The recent years have witnessed the reconstruction of a number of genome-scale metabolic models [1, 2], including pathogenic organisms [3–5], with applications in drug target identification and understanding disease aetiology [6–10]. These genome-scale models of cellular metabolism 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, particularly the removal of one or more genes/reactions [11]. These networks have been studied using Flux Balance Analysis (FBA) and FBA has proven to accurately predict phenotypes following various genetic perturbations [12, 13]. FBA has been employed as a key strategy in antimicrobial discovery pipelines [9, 14]. FBA can also be used to reliably predict synthetic lethal genes in metabolic network which in turn can be used to identify combinatorial targets for various pathogens [15, 16]. These combinatorial targets, which are termed 'Synthetic Lethal Sets' are sets of reactions/genes, in which only the simultaneous deletion of all reactions/genes in the set will abrogate growth and lead to the subsequent death of the organism. (Figure 1).

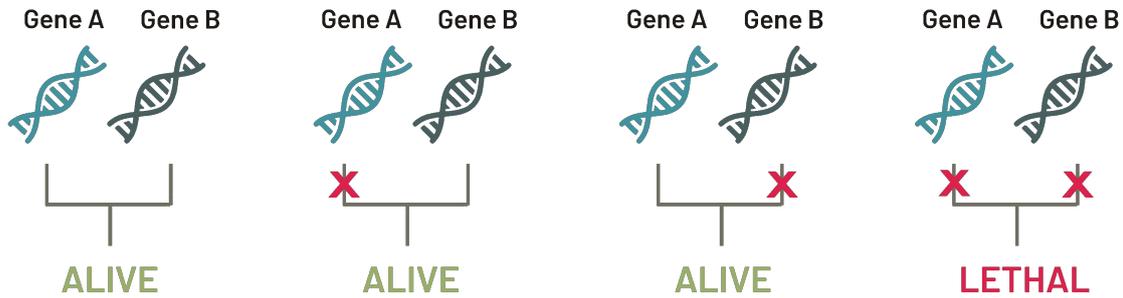[*]Corresponding author. E-mail: kraman@iitm.ac.in

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'

Since the systematic evaluation of such combinatorial deletions in vivo is challenging, computational approaches have been of great interest to overcome this difficulty. A number of algorithms have been developed to predict synthetic lethals in metabolic networks [17–19]. Fast-SL, previously developed in our laboratory, is as yet the most efficient parallel algorithm available to predict synthetic lethal sets in metabolic networks [20]. Fast-SL algorithm circumvents the computational complexity of various other methods through an iterative reduction of the search space for higher order combinatorial targets. Many studies have used Fast-SL to predict synthetic lethals [21–24] 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 pathogenic 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 [11] with the parallel Fast- SL programs and automate the process of identifying single, double, and triple reaction/gene lethal sets for a group of pathogenic 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 [25] and Virtual Metabolic Human (VMH) database [26]. The output results on lethal information were saved as an Excel file in CSV format. These lethal sets were converted into three different formats (JSON, CSV, MAT) to offer users choice in the format they can view the data. This process was repeated for all the 113 organisms and the results were compiled to be published in the database. The names of all organisms along with the information on metabolic models and growth rate are given in Supplementary Information.

3   Results

The synthetic lethal sets (single, double and triple lethal reactions and genes) of 113 pathogenic organisms were identified. The distribution of the number of synthetic lethal sets identified by Fast–SL algorithm is shown in Figure 2. This information was incorporated into CASTLE—a standalone web database. In the home page, the number of each type of lethal set 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.
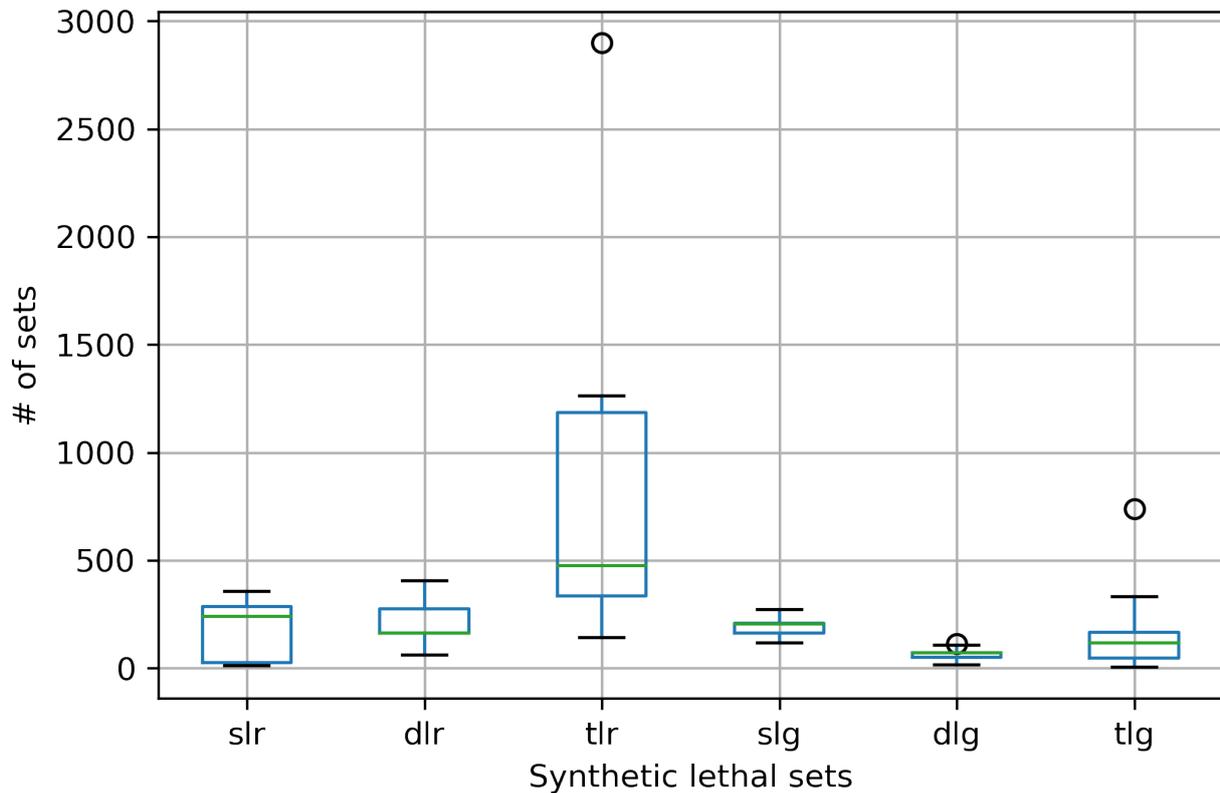

Figure 2: Box plot showing the distribution of the synthetic lethal sets identified in the 113 organisms. The larger box plot of 'tlr' is heavily influenced by the sizable cohorts of triple lethals sets from *Escherichia coli* and *Shigella* organisms. **Abbreviations:** slr – Single Lethal Reactions; dlr = Double Lethal Reactions; tlr = Triple Lethal Reactions; slg = Single Lethal Genes; dlg = Double Lethal Genes; tlg = Triple Lethal Genes

The database also features individual organism page which contains the list of synthetic lethal sets for that particular organism. Equipped with smooth–flowing anchors for better user experience, the individual organism page also provides you with hyperlinks for each synthetic lethal reaction/gene which takes the user to VMH and BiGG databases for further information about the target.

Finally, we provide our synthetic lethal sets as downloadable files available in the organism-specific page. A user can opt for any of the three file formats (JSON, CSV and MAT files) available. The 'About' page provides the user with a brief overview of synthetic lethals and contains instructions on how to use the database effectively.

3

4 Discussion

In this study, synthetic lethal sets (single, double, and triple) were found for 133 pathogenic organisms using the COBRA Toolbox and Parallel Fast-SL for MATLAB. This data was made accessible to global research community through an online database. This database is developed with a motive to fill an existing gap in this area of research, by providing the research community with an easy-to-access, standardized resource unlike any other currently in existence. 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 the models of other pathogenic models, which have not been created yet / have not been analyzed in this study.

Acknowledgments

References

[1] Fang X, Lloyd CJ, and Palsson BO (2020) Reconstructing organisms in silico: Genome-scale models and their emerging applications. *Nature Reviews Microbiology* doi:10.1038/s41579-020-00440-4

[2] O'Brien EJ, Monk JM, and Palsson BO (2015) Using genome-scale models to predict biological capabilities. *Cell* **161**(5):971–987

[3] Raghunathan A, Reed J, Shin S, Palsson B, and Daefler S (2009) Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Systems Biology* **3**(1):38+. doi:10.1186/1752-0509-3-38

[4] Plata G, Hsiao TL, Olszewski KL, LlinasManuel, and Vitkup D (2010) Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. *Molecular Systems Biology* **6**(1). doi:10.1038/msb.2010.60

[5] Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, and Palsson BØ (2016) Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences* **113**(26):E3801–E3809. doi:10.1073/pnas.1523199113

[6] Beste DJ, Hooper T, Stewart G, Bonde B, Avignone-Rossa C, Bushell ME *et al.* (2007) GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosismetabolism. *Genome Biology* **8**(5):R89. doi:10.1186/gb-2007-8-5-r89

[7] Raman K, Yeturu K, and Chandra N (2008) targetTB: a target identification pipeline for Mycobacterium tuberculosis through an interactome, reactome and genome-scale structural analysis. *BMC systems biology* **2**:109. doi:10.1186/1752-0509-2-109

[8] McCloskey D, Palsson B and Feist AM (2013) Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular Systems Biology* **9**:661. doi:10.1038/msb.2013.18

[9] Mienda BS, Salihu R, Adamu A, and Idris S (2018) Genome-scale metabolic models as platforms for identification of novel genes as antimicrobial drug targets. *Future Microbiology* **13**:455–467. doi:10.2217/fmb-2017-0195

[10] Chung WY, Zhu Y, Mahamad Maifiah MH, Shivashekaregowda NKH, Wong EH, and Abdul Rahim N (2020) Novel antimicrobial development using genome-scale metabolic model of Gram-negative pathogens: A review. *The Journal of Antibiotics* doi:10.1038/s41429-020-00366-2

[11] Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols* **6**(9):1290–1307. doi:10.1038/nprot.2011.308

[12] Edwards JS and Palsson BO (2000) Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics* **1**(1):1. doi:10.1186/1471-2105-1-1

[13] Famili I, Forster J, Nielsen J, and Palsson BO (2003) 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* **100**(23):13134–13139. doi:10.1073/pnas.2235812100

[14] Shen Y, Liu J, Estiu G, Isin B, Ahn YY, Lee DS *et al.* (2010) Blueprint for antimicrobial hit discovery targeting metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **107**(3):1082–1087. doi:10.1073/pnas.0909181107

[15] Navid A (2011) Applications of system-level models of metabolism for analysis of bacterial physiology and identification of new drug targets. *Briefings in Functional Genomics* **10**(6):354–364. doi:10.1093/bfgp/elr034

[16] Hartman HB, Fell DA, Rossell S, Jensen PR, Woodward MJ, Thorndahl L *et al.* (2014) Identification of potential drug targets in Salmonella enterica sv. Typhimurium using metabolic modelling and experimental validation. *Microbiology,* **160**(6):1252–1266. doi:10.1099/mic.0.076091-0

[17] Sinha S, Thomas D, Chan S, Gao Y, Brunen D, Torabi D *et al.* (2017) Systematic discovery of mutation-specific synthetic lethals by mining pan-cancer human primary tumor data. *Nature Communications* **8**(1):15580. doi:10.1038/ncomms15580

[18] Huang J, Wu M, Lu F, Ou-Yang L, and Zhu Z (2019) Predicting synthetic lethal interactions in human cancers using graph regularized self-representative matrix factorization. *BMC Bioinformatics* **20**(19):657. doi:10.1186/s12859-019-3197-3

[19] Liu L, Chen X, Hu C, Zhang D, Shao Z, Jin Q *et al.* (2018) Synthetic Lethality-based Identification of Targets for Anticancer Drugs in the Human Signaling Network. *Scientific Reports* **8**(1):8440. doi:10.1038/s41598-018-26783-w. Number: 1 Publisher: Nature Publishing Group

[20] Pratapa A, Balachandran S, and Raman K (2015) Fast-SL: an efficient algorithm to identify synthetic lethal sets in metabolic networks. *Bioinformatics (Oxford, England)* **31**(20):3299–3305. doi:10.1093/bioinformatics/btv352

[21] Stanway RR, Bushell E, Chiappino-Pepe A, Roques M, Sanderson T, Franke-Fayard B *et al.* (2019) Genome-Scale Identification of Essential Metabolic Processes for Targeting the Plasmodium Liver Stage. *Cell* **179**(5):1112–1128.e26. doi:10.1016/j.cell.2019.10.030

[22] Devika NT and Raman K (2019) Deciphering the metabolic capabilities of Bifidobacteria using genome-scale metabolic models. *Scientific Reports* **9**(1):1–9. doi:10.1038/s41598-019-54696-9

[23] Rodenburg SYA, Seidl MF, Judelson HS, Vu AL, Govers F, and Ridder Dd (2019) Metabolic Model of the Phytophthora infestans-Tomato Interaction Reveals Metabolic Switches during Host Colonization. *mBio* **10**(4). doi:10.1128/mBio.00454-19. Publisher: American Society for Microbiology Section: Research Article

[24] Rosario D, Benfeitas R, Bidkhori G, Zhang C, Uhlen M, Shoaie S *et al.* (2018) Understanding the Representative Gut Microbiota Dysbiosis in Metformin-Treated Type 2 Diabetes Patients Using Genome-Scale Metabolic Modeling. *Frontiers in Physiology* **9**:775. doi:10.3389/fphys.2018.00775

[25] King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA *et al.* (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* **44**(D1):D515–522. doi:10.1093/nar/gkv1049

[26] Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G *et al.* (2019) The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research* **47**(D1):D614–D624. doi:10.1093/nar/gky992

Gene A	Gene B		Gene A	Gene B		Gene A	Gene B		Gene A	Gene B
ALIVE			ALIVE			ALIVE			LETHAL	