In silico RECONSTRUCTION OF PROTEIN-PROTEIN INTERACTION NETWORKS OF BACTERIA USED IN SYNTHETIC BIOLOGY APPLICATIONS

C.T. Chasapis[1], V. Savvopoulou[1,2], M.I. Klapa[1,*]
[1]Metabolic Engineering & Systems Biology Laboratory, FORTH/ICE-HT, Patras, Greece
[2]Graduate Program in "Chemical Biology", Department of Chemistry, U. of Patras, Patras, Greece
(*mklapa@iceht.forth.gr)

ABSTRACT
The genetic and physiological characterization of bacteria that are metabolic engineering and synthetic biology resources is essential for the identification of appropriate targets for genetic modification towards the development of more efficient strains and the optimization of their fermentation processes. The number of fully sequenced bacteria has dramatically increased in recent years, providing a valuable basis for the design of high-throughput biomolecular (omic) analyses that can help elucidate the regulatory machinery connecting genotype to phenotype. The interpretation of physiological and omic measurement datasets could be greatly assisted by the accurate reconstruction and integrated analysis of the metabolic and protein-protein interaction (PPI) networks of the investigated bacteria. While a lot of effort has been placed in the genome-scale metabolic network reconstruction in bacteria, being a focal area of research in metabolic engineering for many years, this is not the case with the bacterial PPI networks. To-date, experimentally reconstructed PPI networks based on high-throughput experiments exist for very few bacterial systems [1]. Thus, for most of the bacteria, there is a need to *in silico* reconstruct their potential PPI network based on any available experimental data and bioinformatic resources using predictive models.
In light of this limited experimental PPI data availability for microbial systems, we designed and applied a workflow for the *in silico* reconstruction of the PPI network of two bacteria, *Streptomyces lividans* and *Moorella thermoacetica*, which have been used in synthetic biology applications in the recent years. *S. lividans* is a preferred host for the overproduction of secretory proteins and secondary metabolites of industrial and pharmaceutical interest [2], however, it has a very small number of experimentally determined PPIs reported in source databases. Through the applied workflow, we were able to identify more than a thousand experimentally supported PPIs based on comparative genomic analysis with evolutionarily adjacent bacterial systems. *M. thermoacetica* is a thermophilic Gram-positive acetogen that has been used for the non-photosynthetic CO_2 fixation for the production of useful chemicals [3]. While there is a limited number of experimentally determined PPIs for this bacterium too, a larger than for *S. lividans* number of multi-omic datasets are available, supporting a more reliable application of bioinformatics predictive algorithms. Our analysis so far has identified ~240 experimentally supported PPIs, which can be used to enrich our understanding of the molecular physiology of this microorganism. The reconstructed PPI networks of both microbes can form the basis for validation studies of key PPIs and be used in the design of synthetic biology experiments.

INTRODUCTION
The genetic and physiological characterization of bacteria that are metabolic engineering and synthetic biology resources is essential for the identification of appropriate targets for genetic modification towards the development of more efficient strains and the optimization of their cultivation processes. The number of fully sequenced bacteria has dramatically increased in recent years, providing a valuable basis for the design of high-throughput biomolecular (omic) analyses that can help elucidate the regulatory machinery connecting genotype to phenotype. The interpretation of physiological and omic measurement datasets could be greatly assisted by the

accurate reconstruction and integrated analysis of the metabolic and protein-protein interaction (PPI) networks of the investigated bacteria. While a lot of effort has been placed in the genome-scale metabolic network reconstruction in bacteria, being a focal area of research in metabolic engineering for many years, this is not the case with the bacterial PPI networks. To-date, experimentally reconstructed PPI networks based on high-throughput experiments exist for only 9 bacterial systems [1]. Thus, for most bacteria there is a need to *in silico* reconstruct their potential PPI network based on any available experimental data and bioinformatic resources using predictive models.

OBJECTIVES & METHODOLOGY

In light of this limited experimental PPI data availability for microbial systems, we designed and applied a workflow for the *in silico* reconstruction of the PPI network of two bacteria, *Streptomyces lividans* and *Moorella thermoacetica*, which have been used in synthetic biology applications in the recent years. *S. lividans* is a preferred host for the overproduction of secretory proteins and secondary metabolites of industrial and pharmaceutical interest [2], however, there is currently no reported high-throughput experiment supporting the reconstruction of the direct PPI network of *S. lividans* or any other *Steptomyces* species. *Moorella thermoacetica* has been used for the non-photosynthetic CO_2 assimilation into high-value chemicals and biologics [3].

The workflow comprised the mining of any available PPI data for the investigated bacteria from publicly available experimental PPI databases and then the mining of PPIs between *S. lividans* or *M. thermoacetica*, respectively, orthologues in the experimentally reconstructed PPI networks of evolutionary neighboring bacteria. Experimentally predicted PPI networks for all considered bacteria were mined from the STRING database [4]. All reconstructed networks were visualized and analyzed using the Cytoscape software [5].

RESULTS AND DISCUSSION

S. lividans PPI network

To-date, only thirty-nine (39) experimentally determined PPIs have been reported for *Streptomycetes*, among which only one (1) is specifically assigned to *S. lividans* (Table 1). It is also noted that STRING database does not provide a predicted from experimental (mainly transcriptomic) data for the particular organism.

Table 1. *The number of experimentally determined PPIs for Streptomycetes.*

***Streptomyces* species**	**Number of Experimentally Determined PPIs**
Streptomyces lividans	1
Streptomyces coelicolor	18
other Streptomyces species	20
TOTAL	**39**

Following the *in silico* reconstruction workflow presented in this work, combining both experimentally determined and predicted PPI networks for the neighboring bacteria of *S. lividans* augmented by similar information from *Escherichia coli*, we were able to reconstruct a PPI network of 8982 interactions between 1001 UniProt_IDs, including a 1064 experimentally supported PPI network of higher availability. This is indeed a very significant result, because it provides a useful resource to the scientific community for a microorganism of industrial interest, which integrated with information about the gene regulation and metabolic network of this bacterium can form the basis for furthering our understanding of the physiology of this organism and identifying potential targets for useful genetic modifications.

M. thermoacetica PPI network

For many years, metabolic engineering research has focused on the study and evolution of photosynthetic microorganisms and algae for the bioconversion of CO_2 emissions into products of industrial interest. As the yield of these processes and the cost of their maintenance has not proven positive for their use at industrial-scale applications, there has been an increased interest in the recent years in the use of extremophiles, such as, *Moorella thermoacetica* [3] and *Clostridium ljungdahlii* [6], for the non-photosynthetic CO_2 assimilation into the formation of high-value chemicals and biologics. The selection of these two microorganisms was based on their higher metabolic flexibility compared to other extremophiles possessing the non-photosynthetic CO_2 assimilation pathway, and mainly on the availability of reconstructed metabolic models and extensive molecular physiology datasets.

In the context of the BIOMEK (Τ1ΕΔΚ-00279) project, co-financed by Greece and the European Union through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE, we focused our studies on *M. thermoacetica* as it is an obligate anaerobic, fully sequenced, thermophilic and Gram-positive acetogen with a small sized genome, thus flexible for laboratory studies and synthetic biology applications. Its genome-scale metabolic network was reconstructed in 2015 [7], however there has been no study focusing on the reconstruction of its protein-protein interaction (PPI) network, which could enhance our understanding about the physiology of this microorganism and improve the analytical capability of multi-omic studies investigating its physiological boundaries under particular CO_2 availability conditions.

Following the presented in this work in silico PPI network reconstruction workflow, we were able to identify ~240 experimentally supported PPIs including only 9 experimentally determined PPIs. It is of significance to note that this experimentally supported PPI network had substantially augmented the STRING-predicted *M. thermoacetica* network including physiologically significant PPIs that are now absent from the STRING network. The ongoing work includes the analysis of the *in silico* reconstructed PPI network and its integration with the metabolic network to better understand the physiological boundaries of this bacterium.

CONCLUSIONS

The presented workflow for the *in silico* reconstruction of the PPI networks of bacteria of metabolic engineering interest for which no direct experimental information exist can be an important tool in the systems biology analysis of these microorganisms furthering our understanding of their molecular physiology characteristics and susceptibility to any metabolic engineering efforts.

ACKNOWLEDGMENTS

This work was supported by the European Commission 7th Framework Programme “Rewiring the *Streptomyces* cell factory for cost-effective production of biomolecules” (STREPSYNTH) project (n° 613877), a post-doctoral fellowship to Dr. Ch. Chasapis by the Stavros Niarchos Foundation “Advancing Young Researchers’ Human Capital in Cutting Edge Technologies in the Preservation of Cultural Heritage and the Tackling of Societal Challenges” (ARCHERS) program of FORTH/ICE-HT, the projects “ELIXIR-GR: Hellenic Research Infrastructure for the Management and Analysis of Data from the Biological Sciences” (MIS 5002780), and “INSPIRED: The National Research Infrastructures on Integrated Structural Biology, Drug Screening” (MIS 5002550), which are implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Program “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund), and the “Bioconversion of lignite power plant emissions to fuels and fine chemicals” -

BIOMEK (T1EDK-00279) project co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE.

REFERENCES

[1] C.T. Chasapis, C. Rückert, T. Busche, J. Kalinowski, A. Economou, M.I. Klapa. (2019) submitted.
[2] R.P. Mellado. World J. Microbiol. Biotechnol. 27(2011) 2231-2237.
[3] P. Hu, S. Chakraborty, A. Kumar, B. Woolston, H. Liu, D. Emerson, G. Stephanopoulos. PNAS 113 (2016) 3773-3778.
[4] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn et al. Nucleic Acids Res 45(2017) D362-D368.
[5] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker. Genome Res 13(2003) 2498-2504.
[6] S.W. Jones, A.G. Fast, E.D. Carlson, C.A. Wiedel, J. Au, M.R. Antoniewicz, E.T. Papoutsakis, B.P. Tracy. Nat. Commun. 7(2016) 12800.
[7] P. Hu, H. Rismani-Yazdi, G. Stephanopoulos. AICHE 59(2013) 3176–3183.