
Towards the Complete Functional Characterization of a Minimal Bacterial Proteome

David M. Bianchi¹, James F. Pelletier², Clyde A. Hutchison III³, John I. Glass³, Zaida Luthey-Schulten^{1,*}

1 Department of Chemistry, University of Illinois Urbana-Champaign, 600 S Mathews Ave. Urbana, IL 61801

2 Centro Nacional de Biotecnología, Calle Darwin no. 3, 28049 Madrid, ESP

3 J. Craig Venter Institute, 4120 Capricorn Ln. La Jolla, CA 92037

* zan@illinois.edu

Supporting Information

Section S1 Additional Genes Characterized

Table S1. Table of General Genes Characterized Computationally. JCVI-syn3A genes involved in more general cellular functions in ascending order of gene locus tag (NCBI Entry - NCBI GenBank: CP002027.1). The general cellular function and the name of the most similar analogous gene discovered from the computational workflow for a related species is given.

Gene	General Function
<i>azoR/0029</i>	NADH-azoreductase
<i>ftsH/0039</i>	Membrane Protein Protease
<i>ywnG-like/0146</i>	Unknown - <i>B. subtilis</i> ortholog
<i>ytoP-pepA/0154</i>	Aminopeptidase
<i>prkC/0264</i>	Serine-threonine protein kinase
<i>ylxR/0298</i>	Transcription Termination Nucleic Acid Binding
<i>ylxQ/0299</i>	RNA Binding Ribosomal Protein L7A
<i>ynzC-like/0315</i>	DNA Damage Control
<i>dnaD-like/0352</i>	DNA Replication Machinery
<i>thiJ-like/0400</i>	Protein deglycase
<i>yqfO-like/0409</i>	GTP cyclohydrolase
<i>yqgC-like/0424</i>	Unknown - <i>B. subtilis</i> ortholog
<i>ymdB/0431</i>	Damage Control phosphoesterase
<i>cutC-like/0433</i>	Copper Homeostasis Regulator
<i>prp-like/0500</i>	Ribosomal protein processing protease
<i>pepF-like/0693</i>	Growth Regulation Peptidase
<i>fic/0805</i>	Filamentation/Protein Adenylation
<i>whiA/0817</i>	Cell Division/Chromosome Segregation Regulation
<i>yqxA-like/0830</i>	Growth factor
<i>ychF/0872</i>	Ribosome Binding Factor ATPase

Table S2. Table of General Genes Characterized Computationally. JCVI-syn3A genes involved in more general cellular functions in ascending order of gene locus tag (NCBI Entry - NCBI GenBank: CP002027.1). The p-value for significant structural similarity generated from FATCAT structural alignment¹ and the closest PDB Structure to which the AlphaFold prediction is aligned are given for each gene (AF2 signifies an AlphaFold predicted structure where experiment is unavailable).

Gene	Structure p-value (RMSD Å)	Closest PDB
<i>azoR/0029</i>	0.00 (2.35)	3W77
<i>ftsH/0039</i>	0.00 (2.25)	2DHR
<i>ywnG-like/0146</i>	3.28×10^{-4} (2.10)	AF2 <i>B. subtilis</i> ²
<i>ytoP-pepA/0154</i>	0.00 (2.00)	3JRU
<i>prkC/0264</i>	0.00 (2.18)	4EQM
<i>ylxQ/0298</i>	0.00 (1.30)	3V7Q
<i>ylxR/0299</i>	7.43×10^{-14} (1.69)	1G2R
<i>ynzC-like/0315</i>	8.94×10^{-12} (0.58)	3BHP
<i>dnaD-like/0352</i>	6.94×10^{-9} (2.08)	2V79
<i>thiJ-like/0400</i>	2.12×10^{-13} (2.79)	3EFE
<i>yqfO-like/0409</i>	4.44×10^{-16} (2.51)	AF2 <i>B. subtilis</i> ³
<i>yqgC-like/0424</i>	3.13×10^{-4} (2.63)	AF2 <i>B. subtilis</i> ⁴
<i>ymdB/0431</i>	0.00 (1.02)	4B2O
<i>cutC-like/0433</i>	0.00 (1.82)	1TWD
<i>prp-like/0500</i>	4.29×10^{-14} (1.36)	2P92
<i>pepF-like/0693</i>	9.95×10^{-4} (4.35)	2QR4
<i>fic/0805</i>	4.05×10^{-11} (3.17)	3EQX
<i>whiA/0817</i>	1.11×10^{-16} (2.23)	3HYI
<i>yqxA-like/0830</i>	3.81×10^{-6} (1.86)	AF2 <i>B. subtilis</i> ⁵
<i>ychF/0872</i>	0.00 (1.87)	1JAL

Section S2 JCVI-syn3A AlphaFold2 Predictions Database

We also provide a database of AlphaFold2 predictions for all of the approximately 450 protein-coding genes in JCVI-syn3A (generated using the same program parameters given in Section Methods). Along with this database we provide a Jupyter Notebook for visualization and analysis of AlphaFold2 predicted structures (with associated prediction confidence coloring *etc.*) and relevant associated data.

The database and Jupyter Notebook can be found at [Database Link](#).

Section S3 Summary of Updated JCVI-syn3A Gene Product Annotations

Spreadsheet_S11.xlsx gives the updated gene product annotations generated from this study along with the previous gene product annotations and associated Transposon Mutagenesis gene essentiality data that was published by Breuer *et al.*⁶ for all 450+ protein-coding genes in JCVI-syn3A. This data is available at: [SI Table Link](#), under the Supplementary Table link.

References

1. Li, Z.; Jaroszewski, L.; Iyer, M.; Sedova, M.; Godzik, A. *Nucleic Acids Research* **2020**, *48*, W60–W64.
2. Uncharacterized protein YwnG - AlphaFold structure prediction.
<https://alphafold.ebi.ac.uk/entry/P71042>, Date Accessed: 2022-04-03.
3. GTP cyclohydrolase 1 type 2 homolog - AlphaFold structure prediction.
<https://alphafold.ebi.ac.uk/entry/P54472>, Date Accessed: 2022-04-03.
4. Uncharacterized protein YqgC - AlphaFold structure prediction.
<https://alphafold.ebi.ac.uk/entry/P54486>, Date Accessed: 2022-04-03.
5. Uncharacterized protein YqxA - AlphaFold structure prediction.
<https://alphafold.ebi.ac.uk/entry/P38425>, Date Accessed: 2022-04-03.
6. Breuer, M. et al. *eLife* **2019**, *8*.