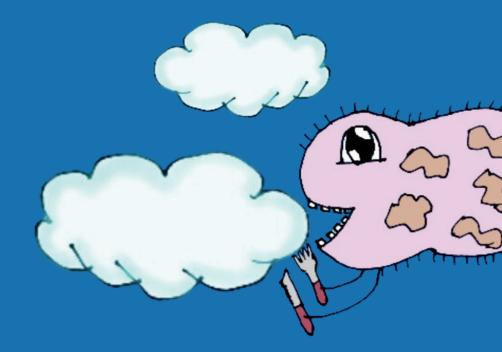


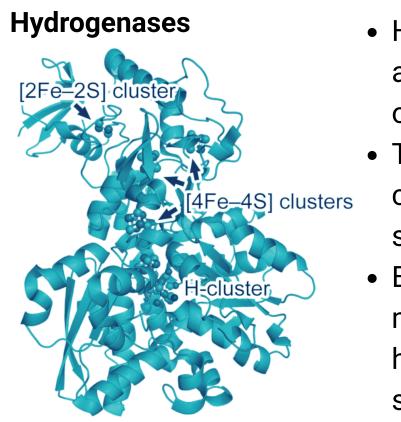
Identifying Hydrogenases In Bacteria Using An Exploratory Computational Tool

AVANTIKA MOHAN^{1,4}, Robert Gunsalus¹, Thomas Holton², Daniel Ha³

¹Department of Microbiology, Immunology, and Molecular Genetics, UCLA | ²Institute for Genomics and Proteomics, UCLA | ³Institute for Quantitative and Computational Biosciences, UCLA | ⁴BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA



BACKGROUND INFORMATION



 Hydrogenases are a subclass of enzymes found in prokaryotes, archaea, and eukaryotes, involved in the production and consumption of hydrogen

- They contain iron-sulfur (FeS) proteins that have active sites consisting of inorganic sulfide and iron atoms bound by cysteinyl sulfur atoms to the polypeptide chain
- Based on their metallocenter composition, they are divided into three main groups: [NiFe] hydrogenases, [FeFe] hydrogenases, and [Fe] hydrogenases. This makes them extremely diverse; in fact the Fe-Fe subunit, found in S.wolfei, isn't even homogeneous across organisms

Protein Families (PF) & Domain Architectures

- Protein families are groups of proteins that share similar sequences, or structure and function as a result of a common evolutionary ancestor
- Eg: PF13510: Fer 2_4 is a type of 2Fe-2S ferrodoxin family
- Existing search tools, like HydDB, rely on DNA and/or amino-acid primary sequence data, and have been of limited value in predicting gene and enzymatic function for a diverse group of organisms
- When analyzing a gene, we can consider its domain architecture, which refers to the functional units, 'domains,' that make up its protein sequence
- Here, the genes are broken down into the protein families they are comprised of. Along with information about their amino acid length, signal proteins, and transmembrane, we create a unique gene 'signature'

Hyd1abc gene neighborhood in S. wolfei

> Swol_1017: NADP-reducing hydrogenase subunit HndD [Syntrophomonas wolfei Goettingen: NC_008346] (-)str. [1725 bp] GTTAACCTGACAATAGACGGAATTAAAGTATCCGT.....GCTGTTGCATACTCATTACCATGCCAAGAATAAAAAATTCTTA [574 aa] MVNLTIDGIKVSVPEGSTILQAASEVGIKIPTLCFHPD.....LRKSHDNPEVKTLYEEFLHEPLGHKSHELLHTHYHAKNKKFL Domain architecture for Swol 1017:

PF13510 PF10588 PF12838 PF02906 PF02256
> Swol_1017 "signature"

SP_o -TM_o - PF12510 - PF10588 - PF12838 - PF02906 - PF02256 - aa574

GOALS

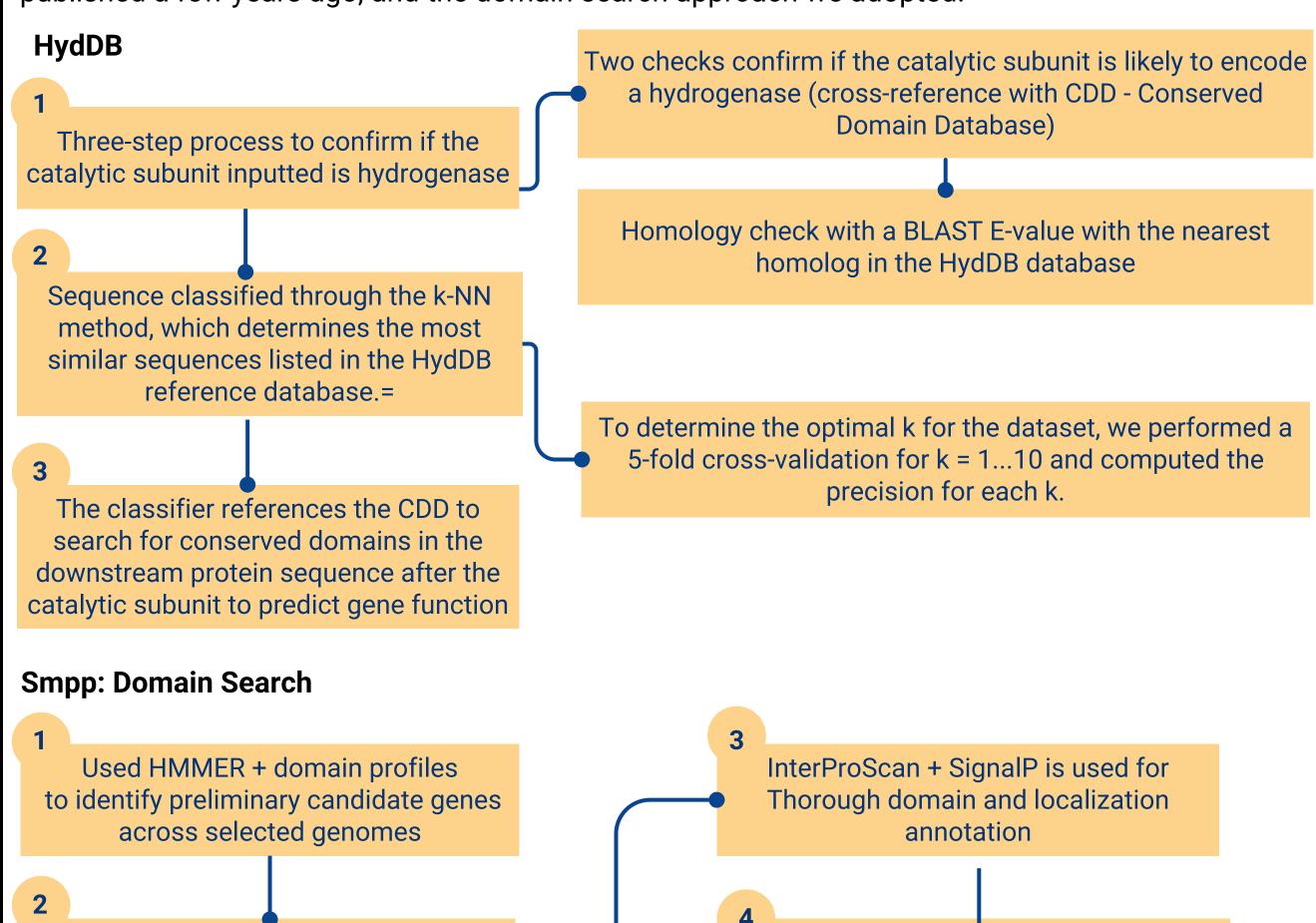


Building a computational tool that recognizes protein families (pfams) in hydrogenase genes to better predict enzyme function across a subset of sequenced genomes.

Conduct an analysis using *Syntrophomonas wolfeii* as a model organism to select a 'signature associated with catalytically active hydrogenase subunits, and compare against three other species in the *Syntrophomonas* genus.

METHODS

A comparative workflow diagram between HydDB, a hydrogenase identifying and classifying tool published a few years ago, and the domain search approach we adopted.



In the subsequent analysis of the 4 different *Syntrophomonas* genomes, I used HMMER, which uses hidden Markov models, to analyze protein sequences and convert them into domain architectures for each sequence. After which, Python analysis was used to parse the annotations further.

Python script used to parse annotation

output and analyze protein domain

compositions

Run selected genes through Prokka +

BLAST to re-annotate and standardize

gene names

are displayed as they appear and are aligned to show similarity across enzyme-coding sequences. Protein Family Distributions across Syntrophomonas Genomes Hierarchical Clustering of all 'Hydrogenase' Sequences Fig: Heatmaps showing the distribution of protein families across 4 Syntrophomonas genomes (wolfei, palmitatica, methylbutyratica, and zehnderi). [Left] Heatmap recording the abundance of 2275 protein families. [Bottom] Heatmap depicting the most (30) commonly occurring protein families and their values per genome. Table shows total PF hits per genome. **UMAP** projection of sequences Fig: After combining protein sequences Top 30 Pfam Domain Hits per Genome Novel sequences found in (A)'hydrogenase' keyword search ('unknown') and (B) sequence hits based on a hydrogenase catalytic protein families ('known'), hierarchical clustering was plotted with a dendogram (above) and UMAP projection (below). The goal was to see where the unknown sequences that were annotated as hydrogenases would cluster with known catalytic hydrogenase sequences. The table The distance cutoff was set as shows 'unknown' suspected 0.3 after cluster vs cutoff analysis hydrogenases per genome.

CONCLUSION & DISCUSSION

- The 'domain search' tool displays genes, their locus tags, and the positions of the PFs within the genes identical to the PFs inputted. These can be used to filter candidates for hydrogenase, formate dehydrogenases, or dehydrogenase (amongst many other) genes.
- Knowledge about PFs also reveals insights into the chemistry and functionality of genes, which can be predicted based on 'signatures.' These genes can then be tested, creating a methodology that can be applied to identifying and classifying novel genes.
- Analysis of the *Syntrophomonas* species reveals a rich diversity in hydrogenase sequences (including small, large, and accessory subunits) as seen in the hierarchical clustering. Analysis of the known catalytic sequences shows an insight into the conserved and additional PFs that comprise the active subunits. More species can be studied to reveal further conservation.

REFERENCES & ACKNOWLEDGEMENTS

- 1. Minor CM, Takayesu A, Ha SM, Salwinski L, Sawaya MR, Pellegrini M, Clubb RT. A genomic analysis reveals the diversity of cellulosome displaying bacteria. Front Microbiol. 2024 Oct 30;15:1473396. doi: 10.3389/fmicb.2024.1473396. PMID: 39539715; PMCID:
- PMC11557425.

 2. Xuan, J., He, L., Wen, W., & Feng, Y. (2023). Hydrogenase and Nitrogenase: Key Catalysts in Biohydrogen Production. *Molecules*, 28(3), 1392. https://doi.org/10.3390/molecules28031392
- 1392. https://doi.org/10.3390/molecules28031392
 3. Søndergaard D, Pedersen CN, Greening C. HydDB: A web tool for hydrogenase classification and analysis. Sci Rep. 2016 Sep 27;6:34212.
- doi: 10.1038/srep34212. PMID: 27670643; PMCID: PMC5037454.

 4. Losey NA, Mus F, Peters JW, Le HM, McInerney MJ. Syntrophomonas wolfei Uses an NADH-Dependent, Ferredoxin-Independent [FeFe]-
- Hydrogenase To Reoxidize NADH. Appl Environ Microbiol. 2017 Sep 29;83(20):e01335-17. doi: 10.1128/AEM.01335-17. PMID: 28802265; PMCID: PMC5626996.

 5. Birrell, J. A., Rodríguez-Maciá, P., Reijerse, E. J., Martini, M. A., & Lubitz, W. (2021). The catalytic cycle of [FeFe] hydrogenase: A tale of
- two sites. Coordination Chemistry Reviews, 449, 214191. https://doi.org/10.1016/j.ccr.2021.214191

 I am grateful to the B.I.G Summer program for providing a platform to learn so many new cool techniques, my PI, Dr. Robert Gunsalus, for working with me, encouraging me, and answering my questions, Thomas Holton, for making bash seem approachable, and the new friends I made during this time, thank you for laughing at all my jokes.