

BACKGROUND INFORMATION

Hydrogenases

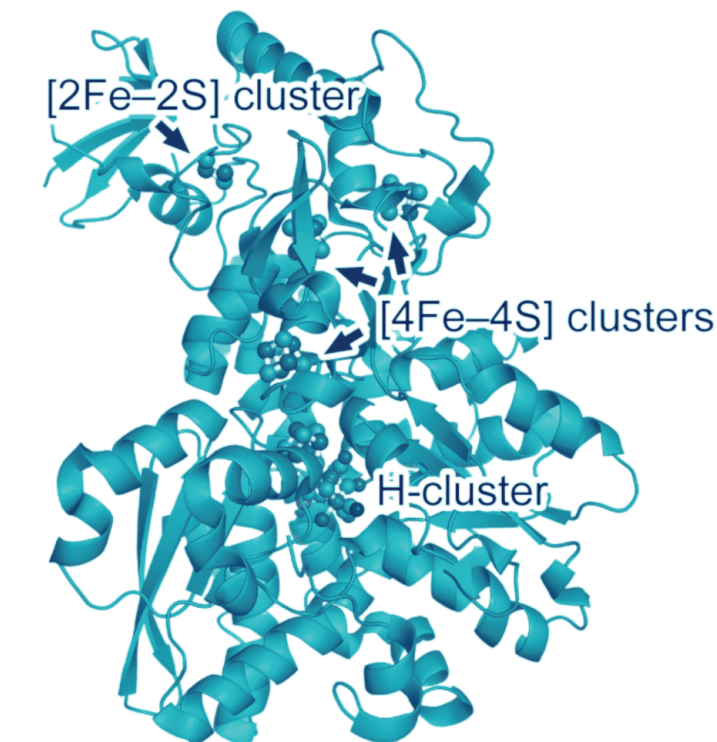


Fig: Fe-Fe hydrogenase protein
Credit: Akin, J., et al (2023): Hydrogenase and Nitrogenase: Key Catalysts in Biohydrogen Production. *Molecules*, 28(3), <https://doi.org/10.3390/molecules28031392>

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 ferredoxin 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] `GTTAACCTGACAAATAGACGGAATTAAAGTATCGT....GCTGTTGTCATACCTATTACCATGCCAAGAATAAAAAATTCTTA` [574 aa] `MVNLTIDGIKVSVPESGTLQAASEVGKIPTLCFHPD....LRKSHDNPEVKTYEELFHEPLGHKSHELLHHTHYHAKNKKFL`

Domain architecture for Swol_1017:



> Swol_1017 "signature"

SP₀-TM₀-PF12510-PF10588-PF12838-PF02906-PF02256-aa574

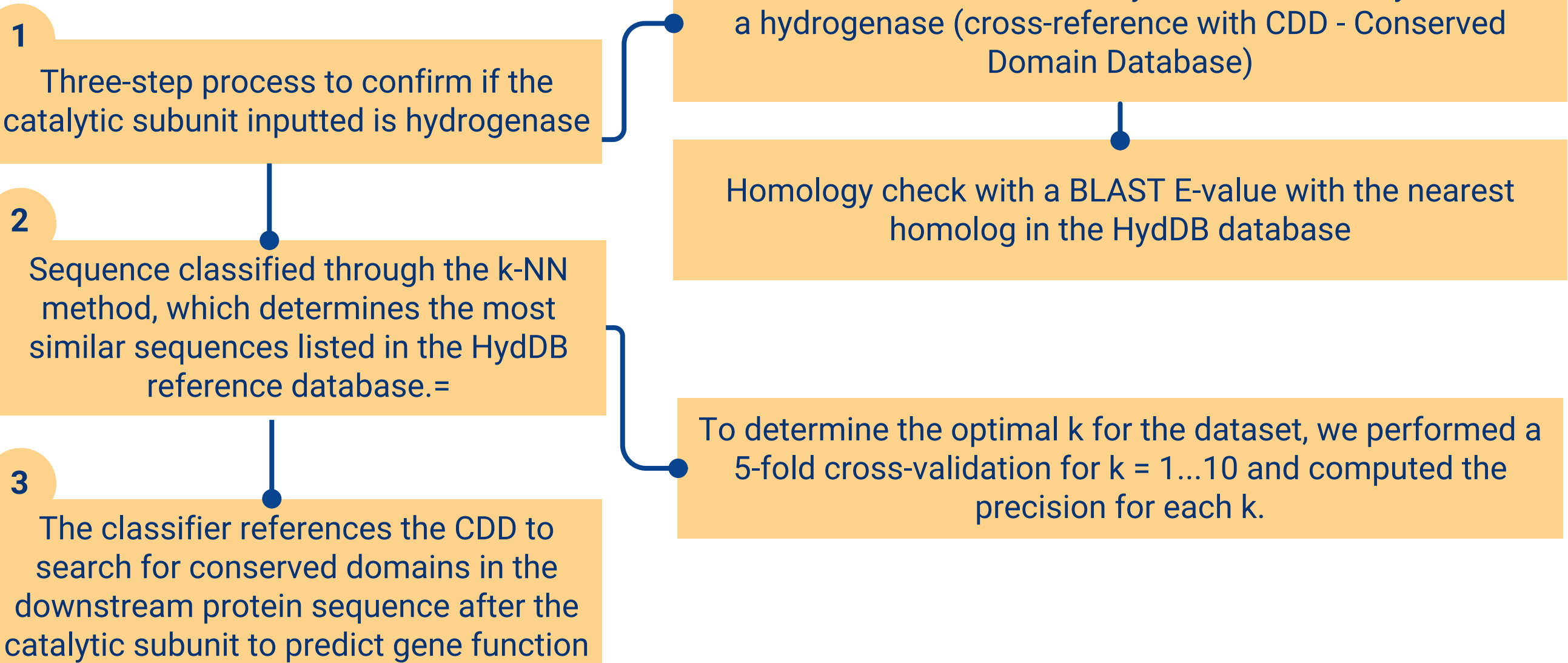
GOALS

- 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.

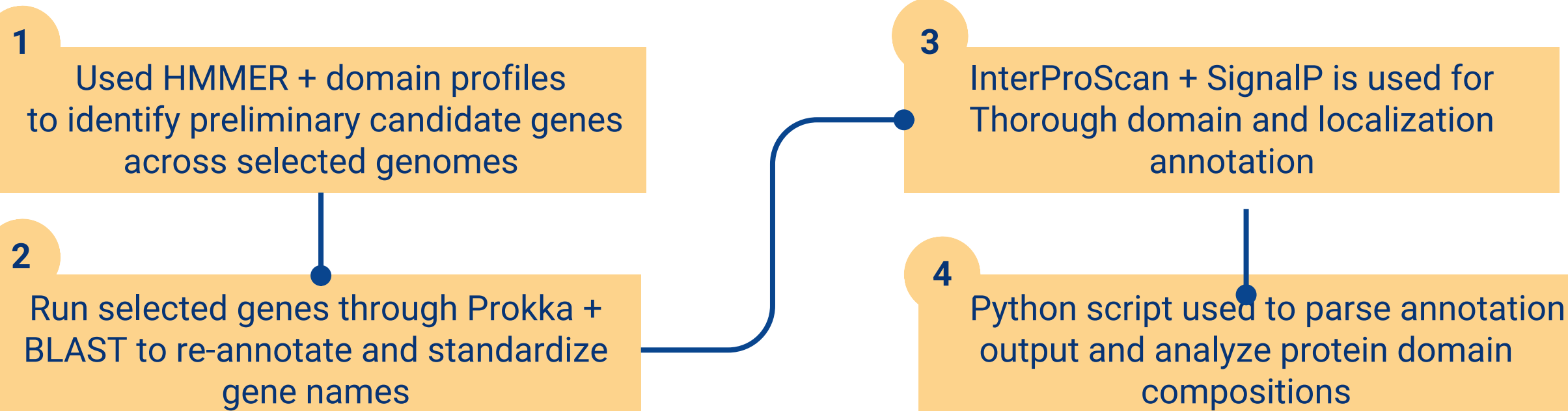
METHODS

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

HydDB



Smp: Domain Search



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.

Phylogenetic Analysis & Characterization of Catalytic Sequences for 4 *Syntrophomonas* Species