Project properties	
Title	In silico medium prediction for culturable and unculturable microorganisms.
Group	Systems and Synthetic Biology
Project type	Thesis
Credits	36
Supervisor(s)	Dr. Jasper Koehorst, Dr. Peter Schaap
Examiner(s)	Dr. Peter Schaap, Dr. Jasper Koehorst
Contact info	peter.schaap@wur.nl, jasper.koehorst@wur.nl, robert1.smith@wur.nl
Begin date	2022/09
End date	
Description	Type of thesis: Computational work. Efficient culturing of "novel' microorganisms can be very challenging and requires a great deal of experience and is often a process of trial and error. As a result, by far most microorganisms in nature still cannot be cultured (~99% by some estimates). Recent developments in sequencing, in particular in the field of metagenomics, allows us to extract genomic data from currently nonculturable organisms. This opens the way to move away from the trial-and-error black box approaches by including rational data-driven method such as genome and pathway analysis to determine microbial nutrient needs. Integrating these new developments will require new approaches to rapidly bring new organisms into culture. Aim: to build a tool that links ecological closeness and phylogenetic similarity of culturable and unculturable species with its genome based functional potential. Three main sources of data will be used: 1) Experimental data from BacDive and https://komodo.modelseed.org. Here we can find a large catalogue of lab media that have been manually developed to date, and tools to explore what insight these known media can give into predicting new organism-media pairings: 2) A similar catalogue from MediaDB https://mediadb.systemsbiology.net/defined_media/ and 3) a large in-house collection of 400.000 FAIR *de novo* functionally and taxonomically annotated microbial genomes. References: Harnessing the landscape of microbial culture media to predict new organism–media pairings. Nat Commun. 2015; 6: 8493. MediaDB: A Database of Microbial Growth Conditions in Defined Media. PLoS One. 2014; 9(8) Conventional culture methods with commercially available media unveil the presence of novel culturable bacteria. Gut Microbes. 2019; 10(1): 77–91.
Used skills	- Programming (Python or Java) - SPARQL: a semantic query language for databases— to manipulate functional genome data stored in Resource Description Framework (RDF) format. - Pattern recognition - Supervised Machine Learning: Frequent Pattern Mining, Random Forest, PCoA, among others
Requirements	- An understanding of the principles of microbial systems - Experience with at least one programming language and some familiarity with the Linux environment.