Outline for MM model Paper

* Introduction/Background
  + Methane is an important chemical in the global carbon cycle, it’s a greenhouse gas, and a fuel source. Methanogens make about 1 GT of it per year, so they’re hugely important microorganisms
  + Methanococcus maripaludis S2 is a model methanogen; it possesses a fast doubling time, grows readily in a chemostat, and has a well-developed set of genetic tools.
  + Metabolic models are useful because they serve as organism knowledge bases and because they can be simulated to predict growth phenotypes for potential wet lab experiments. They have promise for guiding metabolic engineering efforts such as harnessing the unique energy metabolism of our hydrogenotrophic methanogen.
  + We have already constructed the most current metabolic models for two Methanosarcina and have also developed a likelihood-based gap filling method for building new metabolic models with increased gene homology. We have combined our expertise in modeling methanogens with our new gapfilling tool to produce the first manually-curated genome scale metabolic model constructed using likelihood-based gapfilling
* Methods
  + Standard model-building methodology
  + Code generating method (codes I want to distribute with the model)
    - maxGrowthOn\_\_ codes
    - simulateKOPanel code
    - switchToFormate code
    - switchToSpecificFerredoxins code
  + Chemostat culture growth method
  + Metabolomics data method
  + Dry cell weight measurement method
* Results
  + Basic model data
    - # rxns, mets, genes (currently 662, 687, 494)
    - % genome covered
    - # gene-associated rxns, transports, exchanges
  + Results of likelihood-based gapfill
    - We have # gapfilled reactions in our model, and unlike other gapfilling, our reactions come equipped with “likelihood” scores. These scores give us and other users insight into the reaction’s place in the model and go beyond the standard method of evaluating a reaction. Now we can easily see the least likely reactions and point ourselves at these entry points as a means of improving the model.
    - We can actually look at all non-exchanges in our model and see that nearly 90% of our reactions are gene-associated. This is a direct result of using our gap filling approach; making the model with traditional gap-filling misses 66 genes that we automatically catch.
  + Comparison with growth yield data
  + Comparison with knockouts
    - Compared with data from Leigh lab across 6 papers, we match up very well with KO data.
  + Comparison with metabolomics data
    - TBD
* Discussion
  + Another group published a model of M. maripaludis in 2014; this work marked the first effort to represent M. maripaludis metabolism *in silico*, though the model deviated from published literature in several notable ways:
    - Missed electron bifurcation pathway, the essential connection that completes the Wolfe cycle
    - Includes sulfate transport, which M. maripaludis is known not to do
    - Uses methanophenazine, a cofactor we know to be missing from M. maripaludis
  + The other group’s model relied primarily on the KEGG and MetaCyc databases, making use of only 16 other literature sources for their reactions. Though we relied heavily on the DOE Kbase, we also used many literature sources. Furthermore, due to our use of reaction likelihoods, we are able to score each reaction accordingly, as opposed to simply attributing each reaction to the database it was taken from. This gives us a much better way of evaluating how we did in our model.
  + This model represents the first manually curated model that was constructed with likelihood-based gapfilling, at least to our knowledge. The likelihood scores lend an element of accountability to our gapfilling, but we’ve also strived for accountability making our decisions explicit throughout the curation process (this is a tie-in with Ben’s paper).

Figure 1. Comparison of predicted and experimental growth yields. At the moment, our growth predictions don’t really match, but we think that’s more a function of poor growth measurements than of our model’s prediction capabilities. We’re still waiting on the controllers to show up, but I anticipate being able to re-measure these numbers next month using the a microfiltering system to measure optical density vs. dry cell weight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| KO Genes | H2 | Formate | H2 + Formate | Formate + CO | |  |
| ∆hmd | N | N | N | N | |  |
| ∆mtd | N | N | N | N | |  |
| ∆frcA | N | N | N | N | |  |
| ∆fruA | N | N | N | N | |  |
| ∆frcA∆fruA | N | N | N | N | |  |
| ∆vhcAU∆vhuA | N | N | N | N | |  |
| ∆hdrB2 | N | N | N | N | |  |
| ∆fdhA1 | N | N | N | N | |  |
| ∆fdhA2 | N | N | N | N | |  |
| ∆fdhA1∆fdhA2 | N | L | N | L | |  |
| ∆fdhA2∆fdhB2 | N | N | N | N | |  |
| ∆ehbF | N | N | N | N | |  |
| ∆3H2ase | N | N | N | N | |  |
| ∆5H2ase | L | N | N | N | |  |
| ∆6H2ase | L | N | N | N | |  |
| ∆6H2ase∆cdh | L | N | N | N | |  |
| ∆6H2asesupp | L | N | N | N | |  |
| ∆7H2asesupp | L | N | N | N | | **TOTAL** |
| **Total Correct:** | **10 of 10** | **14 of 16** | **2 of 2** | **1 of 2** | | **27 of 30** |
|  |  |  |  |  |  | |
| Figure 2. Knockout lethality predictions from FBA and agreement with experimental results. Our model achieves 90% agreement with experimental results for central catabolic knockouts, corresponding to a Matthew’s Correlation Coefficient of 0.67. | | | | | | |

|  |  |  |
| --- | --- | --- |
| ***Methanococcus maripaludis S2* model comparison** | | |
| Model | iMR533 | iMM518 |
| Protein Coding Genes | 533 | 518 |
| % ORF Coverage | 31 | 30 |
| Intra/Extracellular Metabolites | 656/52 | 556/49 |
| Dead End Metabolites | 265 | 163 |
| Internal Reactions | 574 | 570 |
| Exchange Reactions | 57 | 49 |
| Gene-Associated Reactions | 565 | 464 |
| % Reactions Associated with Genes (non-exchange) | 89 | 75 |
| Table 1A. A comparison between iMR533 and iMM518 indicates that our model covers slightly more of the genome, including over 100 more gene-associated reactions. Both models include approximately the same number of reactions, but our model has roughly 100 more internal metabolites and dead end metabolites. Though this represent the portion of metabolism that cannot carry flux, all of our model's dead end metabolites are part of gene-associated reactions and thus represent promising avenues for future model expansion. | | |
|
|
|
|
|