The custom HMM library includes **205** HMMs that are subdivided into **42** metabolic pathways; these are summarized in **Table 1**. Each category is further sub-divided into functional class, following the standard established by Anantharaman *et al*. (2016), in which 28 of the HMMs used by LithoGenie were developed. The aim of this program is to allow users to infer genomic potential for lithotrophic metabolisms in ‘omics datasets, leading to hypothesis experiments.

The HMMs used by LithoGenie were sourced from 3 different repositories:

1. Metabolic HMMs from Anantharaman *et al*., 2016, which are freely available in the following GitHub repository: <https://github.com/kanantharaman/metabolic-hmms>
2. TIGRFAMS
3. Pfam

In addition, genes for which HMMs were not publically-available, we designed HMMs based on multiple sequence alignments of a curated set of protein sequences collected from UniProtKB. After representative sequences for each gene were downloaded from UniProt multiple, we generated multiple sequence alignments (MSAs), which were then hand-curated, and used as seeds to build the HMMs using HMMER’s *hmmbuild* program. Bit score cutoffs were chosen by querying each HMM against NCBI’s nonredundant (nr) protein database using HMMER’s *hmmsearch* program; output was evaluated manually, and a bit score cutoff that best delineated between true and false positives was chosen for each HMM.

Table 1: Breakdown of the HMMs used by LithoGenie, and each HMM’s associated pathway. Superscripts indicate the source of HMM; lack of superscript indicates that the HMM was designed here, specifically for LithoGenie.

|  |  |  |
| --- | --- | --- |
| **Chemical Species** | **Pathway** | **HMM** |
| **Sulfur** | sulfide-oxidation | FCC† SQR\* |
|  | sulfur-oxidation | SOR†, SDO\*, DsrCA‡B‡EFHMK |
|  | thiosulfate-oxidation | SoxAB‡C‡XY‡Z |
|  | sulfite-reduction | DsrD†, AsrA‡B‡C‡ |
|  | sulfate-reduction | AprA‡, SopT‡, CysC‡N‡ |
|  | thiosulfate-disproportionation | PhsA\* |
| **Hydrogen** | hydrogen-oxidation | FeFe Hydrogenase Group B1B3\*, FeFe Hydrogenase Group A\*, NiFe Hydrogenase Groups 1\*, 2a\*, 2b\*, 3a\*, 3b\*, 3c\*, 3d\*, 4\* |
| **Methane** | methane/ammonia-oxidation | PmoA‡B‡C‡ |
|  | methane-oxidation | MmoB†D‡ |
|  | methane-production | McrA‡B‡G‡ |
| **Nitrogen** | N2-fixation | Fe nitrogenase ‡ (alpha/beta/gamma)  V-containing nitrogenase (alpha‡/beta‡/delta‡)  NifKDH‡ |
|  | nitrite-oxidation | NxrA\*B\* |
|  | nitrate-reduction | NapA‡B†, NarG‡H‡, NasA |
|  | nitrite-reduction | NrfH‡A†D‡, NirB‡D‡K‡S\* |
|  | nitric-oxide-reduction | NorB\*C\* |
|  | nitrous-oxide-reduction | NosD†Z‡ |
|  | anammox | HzoA\*, HzsA\* |
| **Oxygen** | oxygen-reduction | CoxA‡B‡, CcoN‡O‡P‡, CyoA‡D‡E‡, CydA†B‡, QoxA‡B‡ |
| **carbon-monoxide** | CO-oxidation | CoxS\*M\*L‡ |
| **C1compounds** | methanol-oxidation | MxaF\*, MDO\* |
|  | methylamine-oxidation | MauA‡B‡ |
|  | formaldehyde-oxidation | ESD‡, FhcD‡, FdhA‡L‡, MCH‡, FAE‡, S-hydroxymethyl-glutathione dehydrogenase class III alcohol dehydrogenase‡, S-hydroxymethyl-mycothiol dehydrogenase‡ |
|  | formate-oxidation | Formate dehydrogenase‡ (alpha/beta/gamma) |
| **Carbon** | carbon-fixation | Rubisco‡ (forms I, II, III, II-III, and IV), ChdhB‡C‡D‡, AclA\*B\*, propionyl-CoA synthase\*, malonyl-CoA synthase\*, 4-hydroxybutyryl-CoA synthetase\*, 4-hydroxybutyryl-CoA dehydratase\* |
|  | organic-carbon-oxidation | AckA, G6PD, Gdh2, IDH1, PdhA, SdhA, XDH, GltA |
|  | mixed-acid-fermentation | PfhB, TdcE, LdhA, Mdh, Ppc, PykFAa |
| **Urea** | urea-hydrolysis | UreC‡B‡A‡ |
| **Halogenated-compounds** | halogenated-compounds-breakdown | DehII‡, reductive dehalogenase‡ |
|  | perchlorate-reduction | PcrA‡B‡ |
|  | chlorite-reduction | Cld† |
| **Arsenic** | arsenide-oxidation | AoxA‡B‡ |
|  | arsenate-reduction | ArsC‡ (three families) |
| **Selenium** | selenate-reduction | YgfM‡K‡, Mo-binding subunit‡ |
| **Nitriles** | nitrile-hydratase | NthA‡B‡ |
| **Iron** | iron-reduction | MtrC‡AB‡, OmcS, OmcZ, OmcF, CymA |
|  | iron-oxidation | MtoAB, Cyc2 (3 families) FoxABC, FoxEYZ, Cyc1, sulfocyanin, [FmnAB, DmkAB, EetAB, Ndh2, PplA]\*\* |

\*HMM for this gene designed by Karthik Anantharaman.

† HMM for this gene retrieved from Pfam.

‡ HMM for this gene retrieved from TIGRFAMS.

\*\* Genes within brackets are all part of an eight-gene EET operon from *Listeria monocytogenes* (REF).