



# Functional annotation lesson

Targeted HMM-based annotation

# Profiling datasets using HMMs

The screenshot shows the HMMER web interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and a search icon. On the right of the navigation bar is the EMBL-EBI Hinxton logo. Below the navigation bar, the HMMER logo (a stylized yellow, red, and blue square icon) and the text "HMMER" are displayed, followed by the subtitle "Biosequence analysis using profile hidden Markov Models". A horizontal menu bar below the logo includes links for Home, Search, Results, Software, Help, About, and Contact. Underneath this, a secondary menu bar shows options: phmmmer, **hmmscan**, hmmsearch, and jackhmmer. The main content area features a large teal background with abstract geometric shapes. The title "protein sequence vs profile-HMM database" is centered in a large, light blue font. Below the title are three search options: "Paste a Sequence", "Upload a File", and "Accession Search". A text input field is provided for pasting a sequence, with a placeholder "Paste in your sequence or use the example" and a help icon. Below the input field are two buttons: "Submit" and "Reset". A section titled "▼ HMM Database" is visible, containing a "Protein Families" heading and a selection bar with checkboxes for Pfam, TIGRFAM, Gene3D, Superfamily, PIRSF, and TreeFam. The "Pfam" checkbox is checked. At the bottom of the selection bar are "select all" and "clear all" buttons.

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

**HMMER**  
Biosequence analysis using profile hidden Markov Models

Home Search Results Software Help About Contact

phmmmer hmmscan hmmsearch jackhmmer

protein sequence vs profile-HMM database

Paste a Sequence | Upload a File | Accession Search

Paste in your sequence or use the example ?

Submit Reset

▼ HMM Database ?

Protein Families

Pfam  TIGRFAM  Gene3D  Superfamily  PIRSF  TreeFam

select all | clear all

# Profiling datasets using HMMs

The screenshot shows the HMMER web interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, About us, and a search icon. To the right of the navigation bar is the EMBL-EBI Hinxton logo. Below the navigation bar, the HMMER logo (a stylized yellow, red, and blue square icon) and the text "HMMER" are displayed, followed by the subtitle "Biosequence analysis using profile hidden Markov Models". A horizontal menu bar below the logo includes links for Home, Search, Results, Software, Help, About, and Contact. The "Search" button is highlighted with a black background. Below the menu, there are four tabs: "phmmer", "hmmscan", "hmmsearch" (which is selected and has a black background), and "jackhmmer". The main content area features a large teal gradient background with abstract geometric shapes. The title "protein alignment/profile-HMM vs protein sequence database" is centered in a large, light blue font. Below the title are three links: "Paste an Alignment", "Upload a File", and "Accession Search". A text input field with a placeholder "Paste in your alignment/hmm or use the example" is followed by a "Submit" button and a "Reset" button. A section titled "▼ Sequence Database" contains a list of frequently used databases: Reference Proteomes, UniProtKB, SwissProt, PDB, and Ensembl. A dropdown menu labeled "Current database selection: Reference Proteomes" is shown.

EMBL-EBI Services Research Training About us

EMBL-EBI Hinxton

**HMMER**  
Biosequence analysis using profile hidden Markov Models

Home **Search** Results Software Help About Contact

phmmer hmmscan **hmmsearch** jackhmmer

protein alignment/profile-HMM vs protein sequence database

Paste an Alignment | Upload a File | Accession Search

Paste in your alignment/hmm or use the example ?

Submit Reset

▼ Sequence Database ?

Frequently used databases: [Reference Proteomes](#) [UniProtKB](#) [SwissProt](#) [PDB](#) [Ensembl](#)

Current database selection: [Reference Proteomes](#)

# Getting HMMs from Pfam



HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam  
keyword search Go

**Family: 7tm\_1 (PF00001)**

**Summary**

**Domain organisation**

**Clan**

**Alignments**

**HMM logo**

**Trees**

**Curation & model**

**Species**

**Interactions**

**Structures**

**Jump to...**

enter ID/acc Go

**Summary: 7 transmembrane receptor (rhodopsin family)**

Pfam includes annotations and additional family information from a range of different sources. These sources can be accessed via the tabs below.

Wikipedia: Rhodopsin-like receptors Pfam InterPro

This is the Wikipedia entry entitled "Rhodopsin-like receptors". [More...](#)

Rhodopsin-like receptors [Edit Wikipedia article](#)

Rhodopsin-like receptors are a family of proteins that comprise the largest group of G protein-coupled receptors.<sup>[2]</sup>

Contents [hide]

- 1 Scope
- 2 Function
- 3 Classes
  - 3.1 Subfamily A1
  - 3.2 Subfamily A2
  - 3.3 Subfamily A3
  - 3.4 Subfamily A4
  - 3.5 Subfamily A5
  - 3.6 Subfamily A6
  - 3.7 Subfamily A7
  - 3.8 Subfamily A8
  - 3.9 Subfamily A9
  - 3.10 Subfamily A10
  - 3.11 Subfamily A11
  - 3.12 Subfamily A12
  - 3.13 Subfamily A13
  - 3.14 Subfamily A14
  - 3.15 Subfamily A15
  - 3.16 Subfamily A16

**Rhodopsin-like receptors**

Structure of rhodopsin: A G protein-coupled receptor.<sup>[1]</sup>

Identifiers	
Symbol	7tm_1
Pfam	PF00001
Pfam clan	GPCR_A
InterPro	IPR000276
PROSITE	PDOC00211
SCOPe	1f88 / 1SLDEAM

# Getting HMMs from Pfam

Family: **7tm\_1 (PF00001)**

 653 architectures

 56771 sequences

 12 interactions

 425 species

 344 structures

[Summary](#)

[Domain organisation](#)

[Clan](#)

[Alignments](#)

[HMM logo](#)

[Trees](#)

[Curation & model](#)

[Species](#)

[Interactions](#)

[Structures](#)

[Jump to... !\[\]\(e50091943b385fe16d3277389202856f\_img.jpg\)](#)

[enter ID/acc](#)

[Go !\[\]\(179f167ede0522ebb4ea025b3ad78ca7\_img.jpg\)](#)

## Curation and family details

This section shows the detailed information about the Pfam family. You can see the definitions of many of the terms in this section in the [glossary](#) and a fuller explanation of the scoring system that we use in the [scores](#) section of the help pages.

### Curation

Seed source:	Prosite
Previous IDs:	none
Type:	Family
Sequence Ontology:	<a href="#">SO:0100021</a>
Author:	Sonnhammer ELL 
Number in seed:	64
Number in full:	56771
Average length of the domain:	255.40 aa
Average identity of full alignment:	18 %
Average coverage of the sequence by the domain:	67.02 %

### HMM information

HMM build commands: *build method: hmmbuild -o /dev/null --hand HMM SEED  
search method: hmmsearch -Z 45638612 -E 1000 --cpu 4 HMM pfamseq*

Model details:	Parameter	Sequence	Domain
	Gathering cut-off	30.5	30.5
	Trusted cut-off	30.5	30.5
	Noise cut-off	30.4	30.4

Model length: 263

Family (HMM) version: 21

Download: [download](#) the raw HMM for this family

# Getting HMMs from TIGRFAMS

→ TIGRFAMS Home

→ TIGRFAMS Terms

→ TIGRFAMS Complete Listing

→ TIGRFAMS FTP site

→ TIGRFAMS Resources

## TIGRFAMS COMPLETE LISTING

Accession Name	Function	Isology Type	EC Number
TIGR00001 rpmI_bact	ribosomal protein bl35	equivalog	
TIGR00002 S16	ribosomal protein bS16	equivalog	
TIGR00003 TIGR00003	copper ion binding protein	subfamily_domain	
TIGR00004 TIGR00004	reactive intermediate/imine deaminase	subfamily	
TIGR00005 rluA_subfam	pseudouridine synthase, RluA family	subfamily	5.4.99.-
TIGR00006 TIGR00006	16S rRNA (cytosine(1402)-N(4))-methyltransferase 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase	equivalog	2.1.1.199
TIGR00007 TIGR00007		equivalog	5.3.1.16
TIGR00008 infA	translation initiation factor IF-1	equivalog	
TIGR00009 L28	ribosomal protein bl28	equivalog	
TIGR00010 TIGR00010	hydrolase, TatD family	subfamily	
TIGR00011 YbaK_EbsC	Cys-tRNA(Pro) deacylase	hypoth_equivalog	3.6.1.-
TIGR00012 L29	ribosomal protein ul29	equivalog	
TIGR00013 taut	4-oxalocrotonate tautomerase family enzyme	subfamily	5.3.2.-
TIGR00014 arsC	arsenate reductase (glutaredoxin)	equivalog	1.20.4.1
TIGR00016 ackA	acetate kinase	equivalog	2.7.2.1
TIGR00017 cmk	cytidylate kinase	equivalog	2.7.4.14
TIGR00018 panC	pantoate--beta-alanine ligase	equivalog	6.3.2.1
TIGR00019 prfA	peptide chain release factor 1	equivalog	
TIGR00020 prfB	peptide chain release factor 2	equivalog	
TIGR00021 rplA	ribose 5-phosphate isomerase A	equivalog	5.3.1.6
TIGR00022 TIGR00022	Yhch/YjgK/YiaL family protein	subfamily	
TIGR00023 TIGR00023	acyl-phosphate glycerol 3-phosphate acyltransferase	equivalog	2.3.1.-
TIGR00024 SbcD_rel_arch	putative phosphoesterase	hypoth_equivalog	3.1.-.-
TIGR00025 Mtu_efflux	ABC transporter efflux protein, DrrB family	paralog	
TIGR00026 hi_GC_TIGR00026	deazaflavin-dependent oxidoreductase, nitroreductase family	subfamily	
TIGR00027 methyl_TIGR00027	methyltransferase, TIGR00027 family	subfamily	2.1.1.-
TIGR00028 Mtu_PIN_fam	toxin-antitoxin system PIN domain toxin	subfamily	
TIGR00029 S20	ribosomal protein bs20	equivalog	
TIGR00030 S21p	ribosomal protein bs21	equivalog	
TIGR00031 UDP-GALP_mutase	UDP-galactopyranose mutase	equivalog	5.4.99.9
TIGR00032 argG	argininosuccinate synthase	equivalog	6.3.4.5
TIGR00033 aroC	chorismate synthase	equivalog	4.2.3.5
TIGR00034 aroFGH	3-deoxy-7-phosphoheptulonate synthase	equivalog	2.5.1.54

# Getting HMMs from TIGRFAMS

JCVI Home | TIGRFAMs Home | Genome Properties

## HMM SUMMARY PAGE: TIGR00001

<a href="#">TIGRFAMs Home</a>	Accession	TIGR00001
<a href="#">TIGRFAMs Terms</a>	Name	rpmI_bact
<a href="#">TIGRFAMs Complete Listing</a>	Function	ribosomal protein bL35
<a href="#">TIGRFAMs FTP site</a>	Gene Symbol	rnmT
<a href="#">TIGRFAMs Resources</a>	Trusted Cutoff	51.20
<a href="#">TIGR00001 Seed Alignment</a>	Domain Trusted Cutoff	51.20
	Noise Cutoff	23.65
	Domain Noise Cutoff	23.65
	Isology Type	equivalog
	HMM Length	63
	Mainrole Category	Protein synthesis
	Subrole Category	Ribosomal proteins: synthesis and modification
	Gene Ontology Term	<a href="#">GO:0000315</a> : organellar large ribosomal subunit cellular_component <a href="#">GO:0003735</a> : structural constituent of ribosome molecular_function <a href="#">GO:0006412</a> : translation biological_process <a href="#">GO:0022625</a> : cytosolic large ribosomal subunit cellular_component
	Author	Haft DH, Dodson RJ, Richardson DL
	Entry Date	Apr 20 1999 2:04PM
	Last Modified	Sep 4 2014 11:07AM
	Comment	This ribosomal protein is found in bacteria and organelles only. It is not closely related to any eukaryotic or archaeal ribosomal protein.
	References	SE TIGR DR HAMAP; MF_00514; 522 of 531
	Genome Property	<a href="#">GenProp0174</a> : ribosome, bacterial, large subunit (HMM) <a href="#">GenProp0799</a> : bacterial core gene set, exactly 1 per genome (HMM)

# Getting HMMs from TIGRFAMS

Choo et al., 2004

Choo et al.

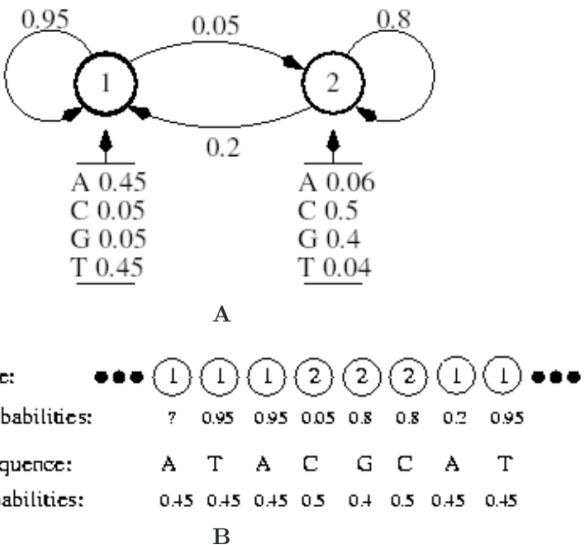


Fig. 1 A. a simple HMM model for generating DNA sequences; B. a generated state sequence and the associated DNA sequence

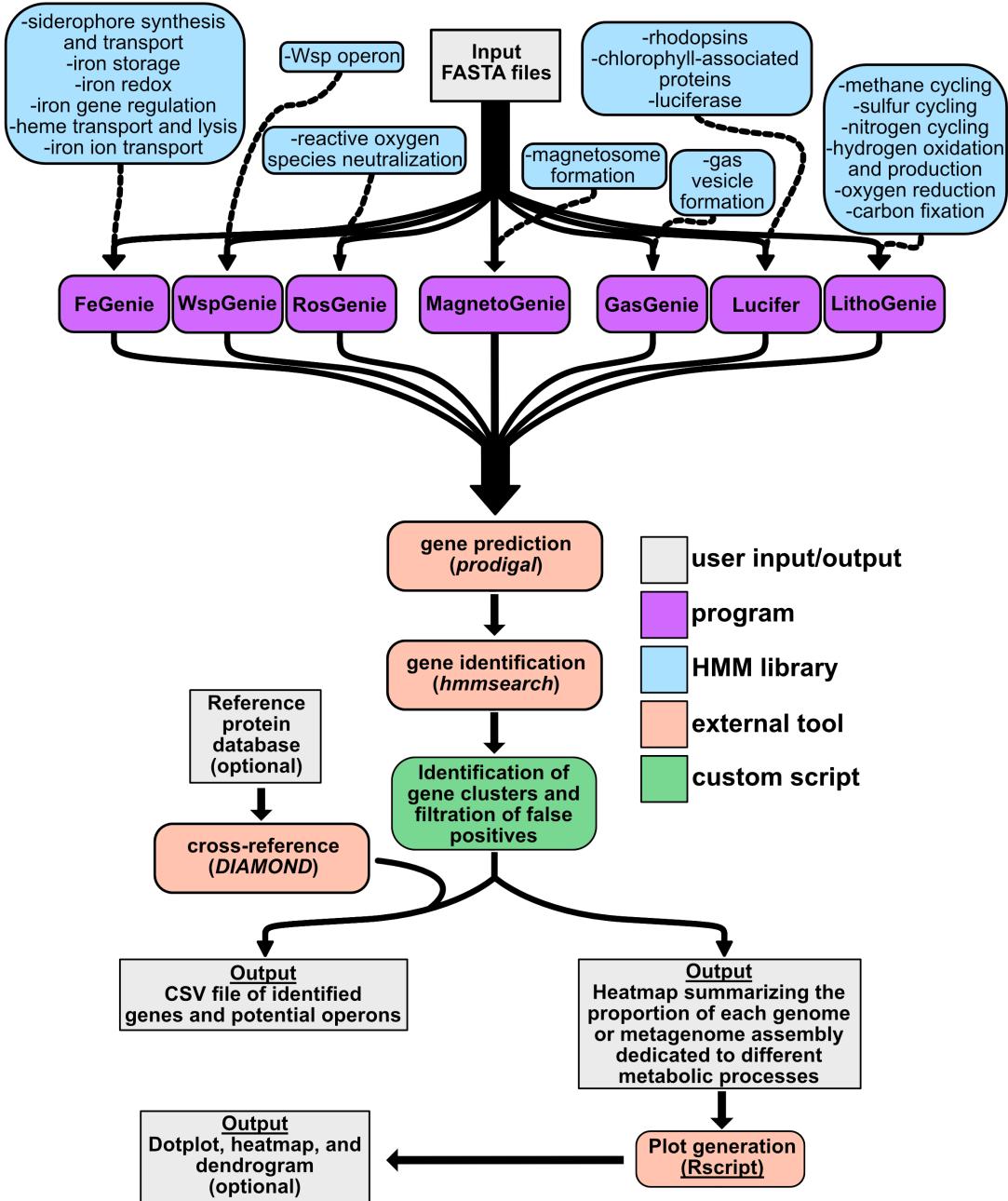
→ **hmmbuild** (HmmGenie tutorial)

- [TIGRFAMs Home](#)
- [TIGRFAMs Terms](#)
- [TIGRFAMs Complete Listing](#)
- [TIGRFAMs FTP site](#)
- [TIGRFAMs Resources](#)

## HMM SEED ALIGNMENT: TIGR00001

[Download Alignment](#)

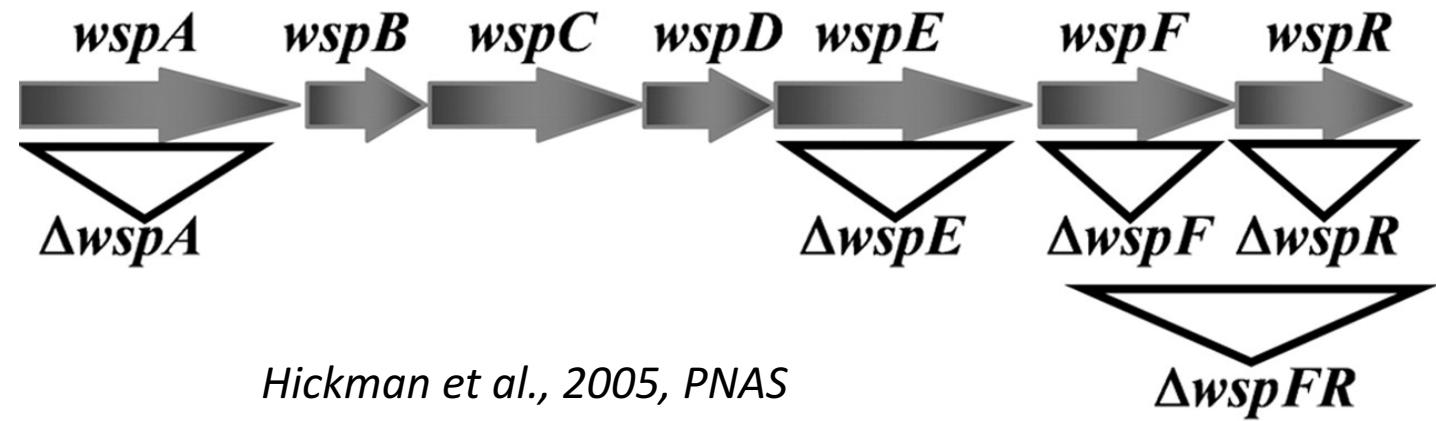
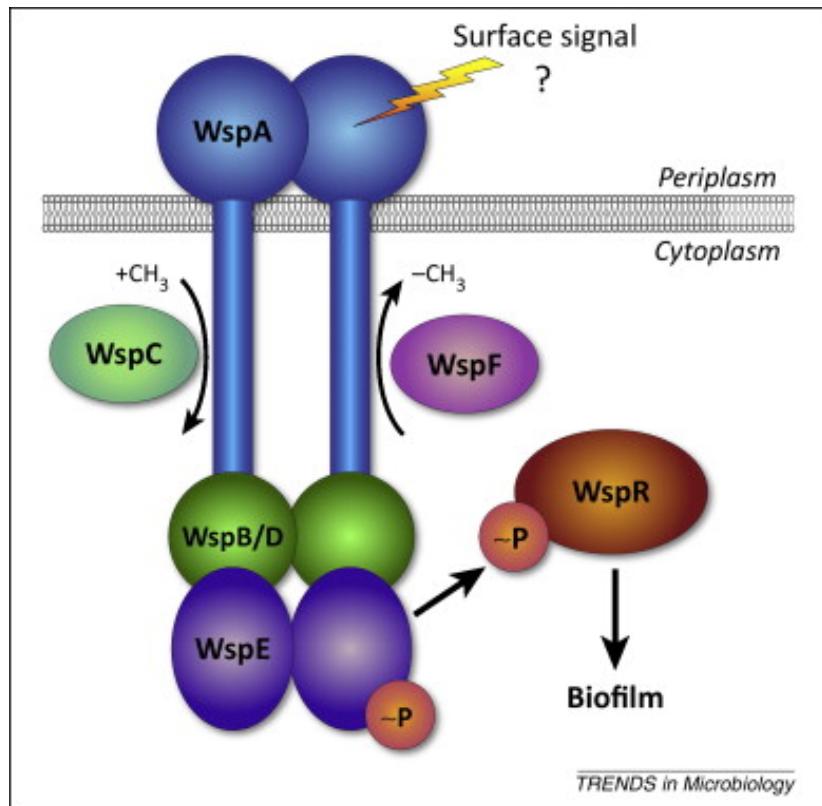
```
>SP|P94976|RL35_MYCTU/2-64
PKAKTHSGASKRFRRTGT.GKIVRQKANRRHLEHKPSTR.T.RRLDGRVVAANDTKRVT.SLLNG
>SP|088059|RL35_STRC0/2-64
PKNKSHSGASKRFRKTGS.GKVLRERAGKRHLLEHKSSRV.T.RRLTGNAEMAPGDAAKIK.KLLGK
>SP|P80341|RL35_THETH/1-63
PKMKTHGAKKRVKITAS.GKVVAMKTGKRHLNWQKSGKEI.RQ.KGRKFVLAKPEAERIKLLLPHY
>SP|Q9RSW6|RL35_DEIR4/2-64
PKMKTHKMAKKRKIITGT.GKVMFKSGKRHQNTGKSGDEI.RG.KGKGFLAKAEWARHKLMLPR
>SP|Q9ZCV1|RL35_RICPR/2-65
PKLTKSAVKKRPFKTAT.GKVIASQAGKHHFMRRTKAQI.RNLRGTTILCPQDGHNKIKYPLPY
>OMNI|NTL01BS2880/2-64
PKMKTHRGSAKRFKKTGS.GKLKRSHAYTSHLFPANKSQQK.RKLRSKAVVSAGDPFKRIK.QMLAN
>SP|P47439|RL35_MYCGE/1-59
..MKTSAAVKRFKLTKS.GQIKRKHAYTSHLAPHKSTKQK.RHLRKQATVSNSELKRIG.ILI...
>GP|150150|gb|AAZ25414.1/2-62
PKMKTRKAAAKRFPTGSGKKIIRRKAFKNHLEHKSEQKHRLSNLALVHEADENVR.LMLPY
>SP|P9X137|RL35_THEME/2-64
PKVKTNRSAAKRFRITKN.GKIMRNHAYRSHKTGKRRNAL.RALRKDKVVSSADKNRVL.RLLGK
>SP|P48959|RL35_SYNYS/2-66
PKLTKRKAACRFPTGSGKKIIRRKAFKNHLEHKSEQKHRLSNLALVHEADENVR.LMLPY
>SP|P14810|RL35_CYAPA/2-64
YKLKTRKAAAKRYKAVGN.KKISRRKAFRSHLQQKSTNRK.RQLSQVVIASPGDTKIIY.LMLPY
>SP|P49567|RL35_ODOSI/2-64
PKLTKRKAALKRYKKTAT.GKFPLRRHAYKGHLLMKKSQTQK.RKLSQIICVSNNNDSKPIK.LMLPY
>SP|P51270|RL35_PORPU/2-64
PKLTKSAKIAKRFKVSS.GKILRHAKSHLQQKSSKHR.RHLSSTCQVDSRDAKNIS.INLPY
>SP|P23326|RL35_SP1OL/90-152
YMKMTHKASAKRFRTVGK.GKIVRRRAKGQHLLAKKNTKRK.NRLSKLIQVDRSDYDNVI.GALPY
>SP|P78496|RL35_GUILTH/2-64
PKLTKRKSARKRFYISAK.GKFVRKRKAFPSHILEKKTSKRK.RNLKRKMIIVFGENLALK.TMLPY
>OMNI|NTL01BC01685/2-64
PKIRTVRGAAXKRFKKTGK.GGFKHKHANLRHILTKKATKRK.RHLRPKAMVSKGDILGLVI.ACIPY
>SP|P52830|RL35_PSESM/2-63
PKMKTKSGAAKRFPLTAN.G.IKHKHFPSHILTKMSTKRK.RQLRGSSLLHPSDVAKVE.RMLRL
>SP|P68845|RL35_VIBCH/2-63
PKMKNNKGAAXKRFKKTAG.G.IKYKHFPSHILTKRTTKNK.RQLRPNAILPKCELAJAVA.RMLPY
>SP|Q9JQN7|RL35_NEIMA/2-64
PKMKTKSSAKKRFKVLGK.GGVKRAHAFKRHILTKTNTKNN.RQLRGTSMVNDRDLASVA.KMLPY
>SP|P083821|RL35_TREPA/2-65
AKMKTKSAAKRFSVTGA.GKVKFKMMNLRHILTKKAPKRK.RKLRHAGFLSKVELKVVKRKLMLPY
>SP|P051207|RL35_BORBU/3-65
HMKMTRKSAAKRFYSPFTVN.GKVKYKKQNLRHILTKTNTKNN.RQLRGTSMVNDRDLASVA.KMLPY
>SP|P066982|RL35_AQUAE/4-66
VNMKTNRSAAKRFVTAK.GKIKRKKSGAHYNTKSSKRK.RHLRKHTYVKDNMLKHVK.ALLKE
>SP|P56057|RL35_HELPY/2-63
PKMKTNRGASKRFRKV.K.NLIKRGSAFKSHILTKSPKRK.ANLNAPKHVHETNAH SVM.SLLCR
>SP|P084841|RL35_CHLTR/2-64
PKMKNSKNSVAAFRFLTGS.GQLKRTRPGKHKLSKRSSQQK.RNLSKQPLVDQGQVGMYK.RMMLV
```



# Which pathways are currently supported:

- Biofilm formation (wsp)
- Reactive oxygen species neutralization
- Magnetosome formation
- Iron
- Photosystems, rhodopsins, and light production (lux)
- Nitrogen-cycling
- Sulfur-cycling
- Hydrogen oxidation/production
- Methanogenesis/methane oxidation
- Oxygen reduction
- Circadian rhythm in cyanobacteria

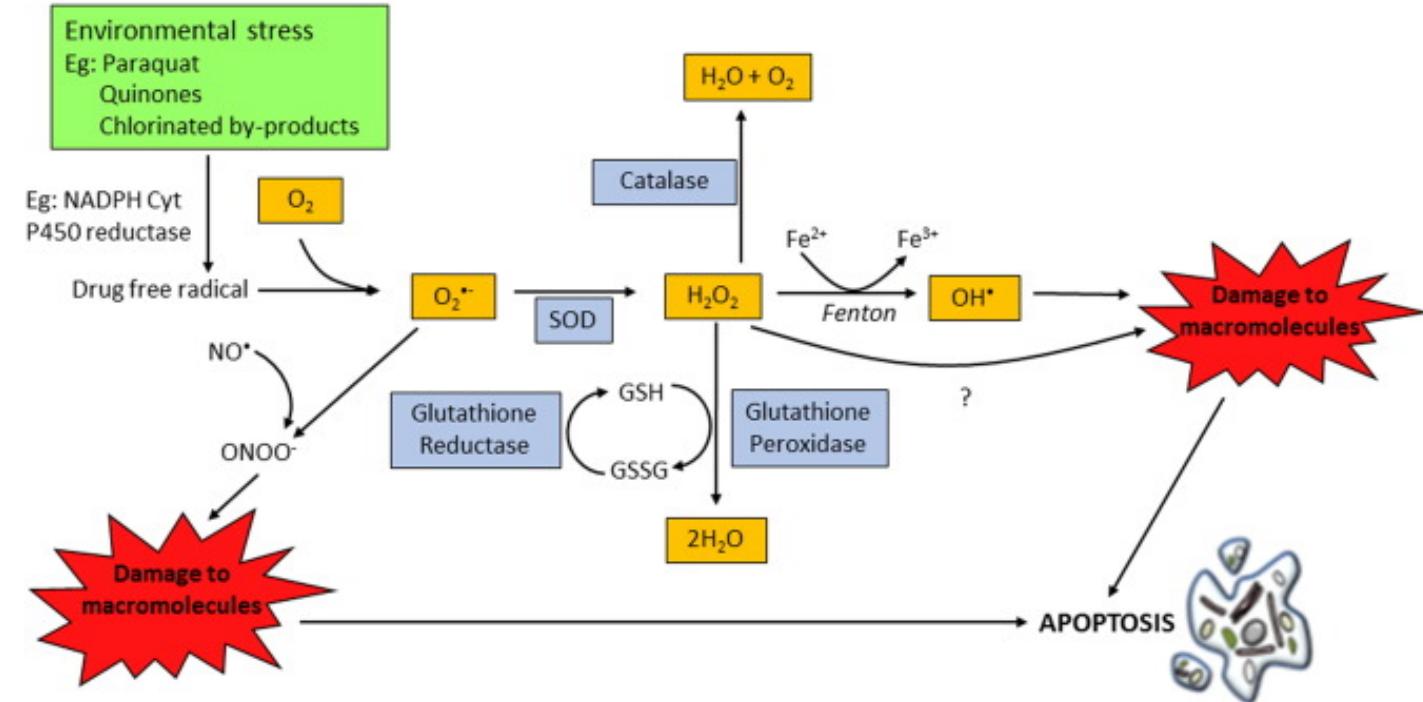
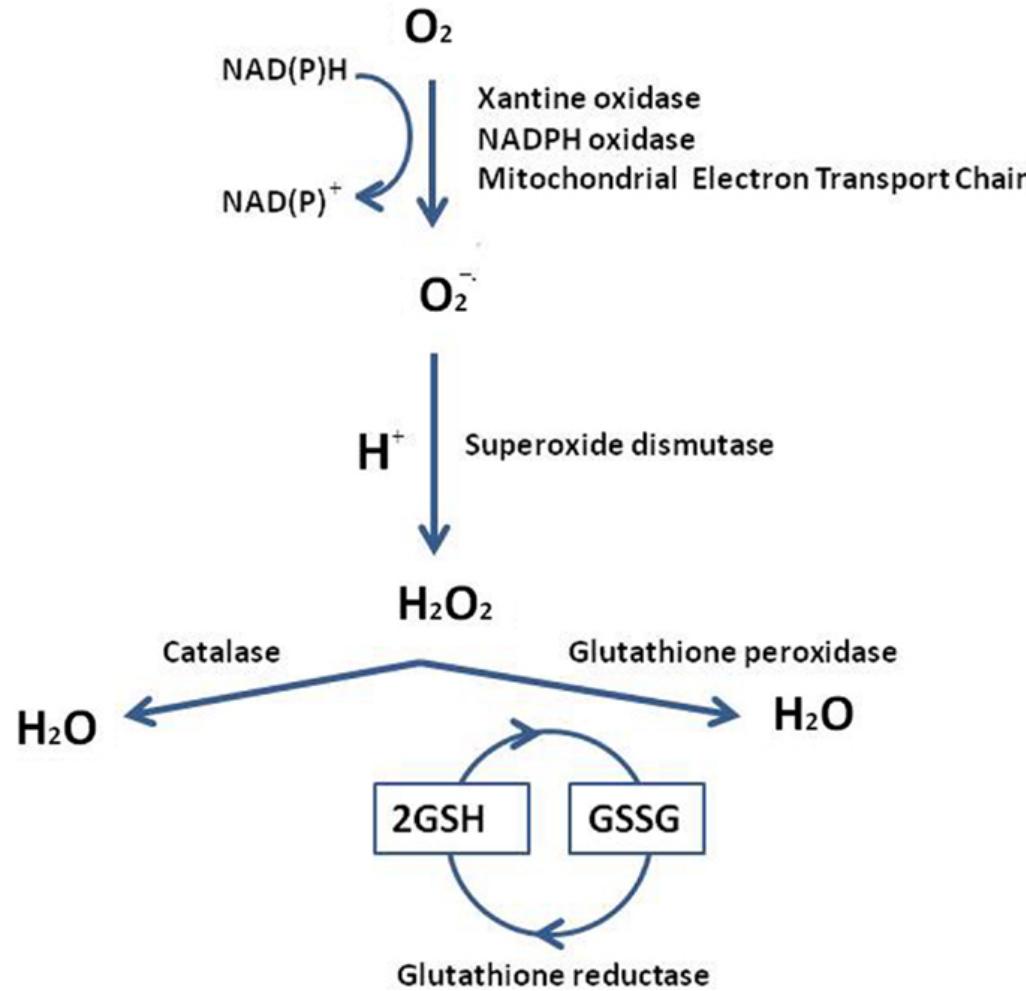
# Biofilm formation - WspGenie



Hickman et al., 2005, PNAS

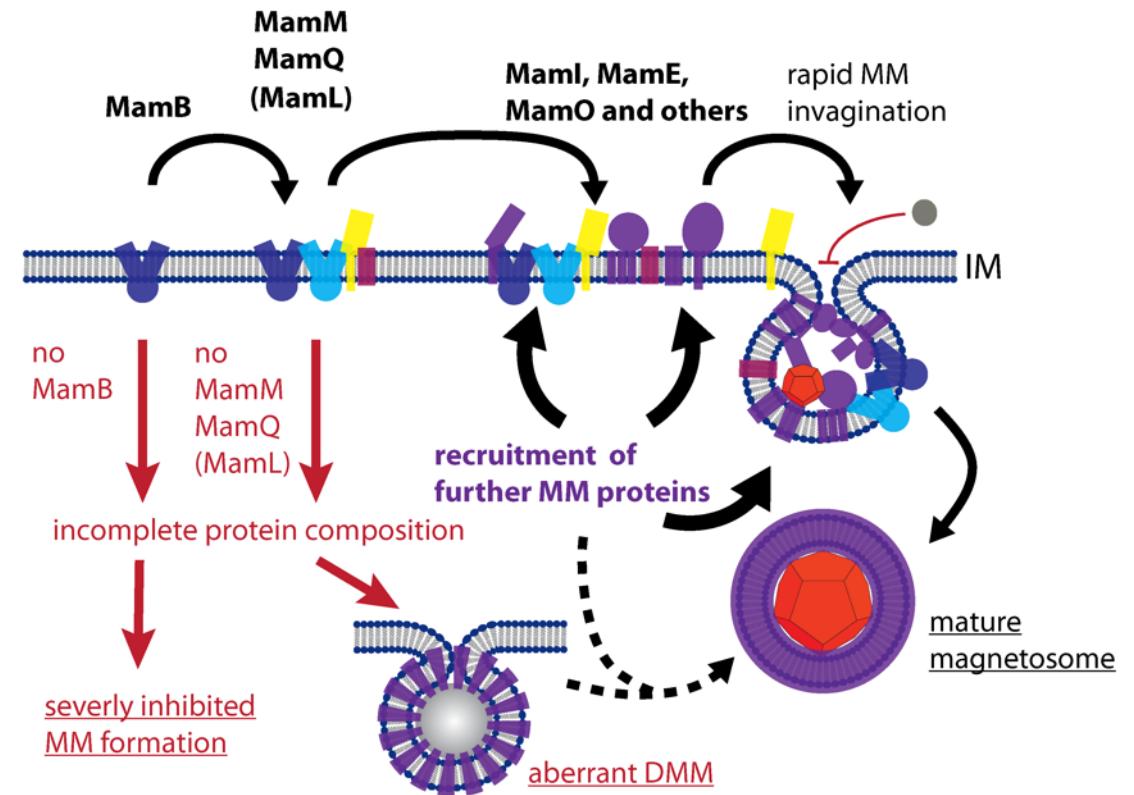
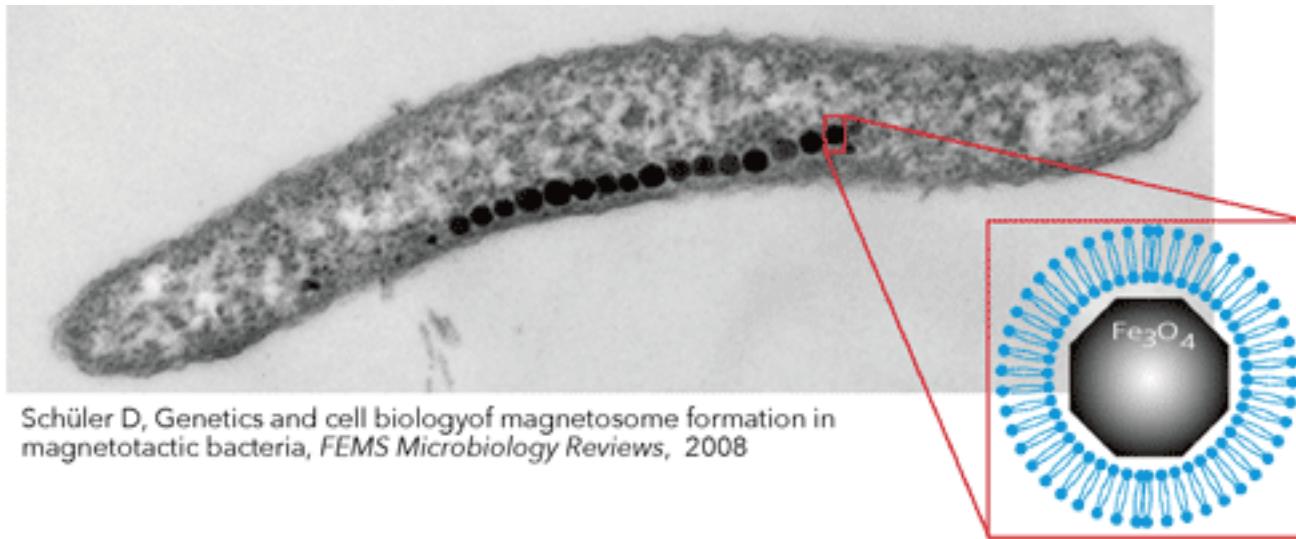
Belas, 2014, Trends in Microbiology

# Reactive oxygen species neutralization - RosGenie



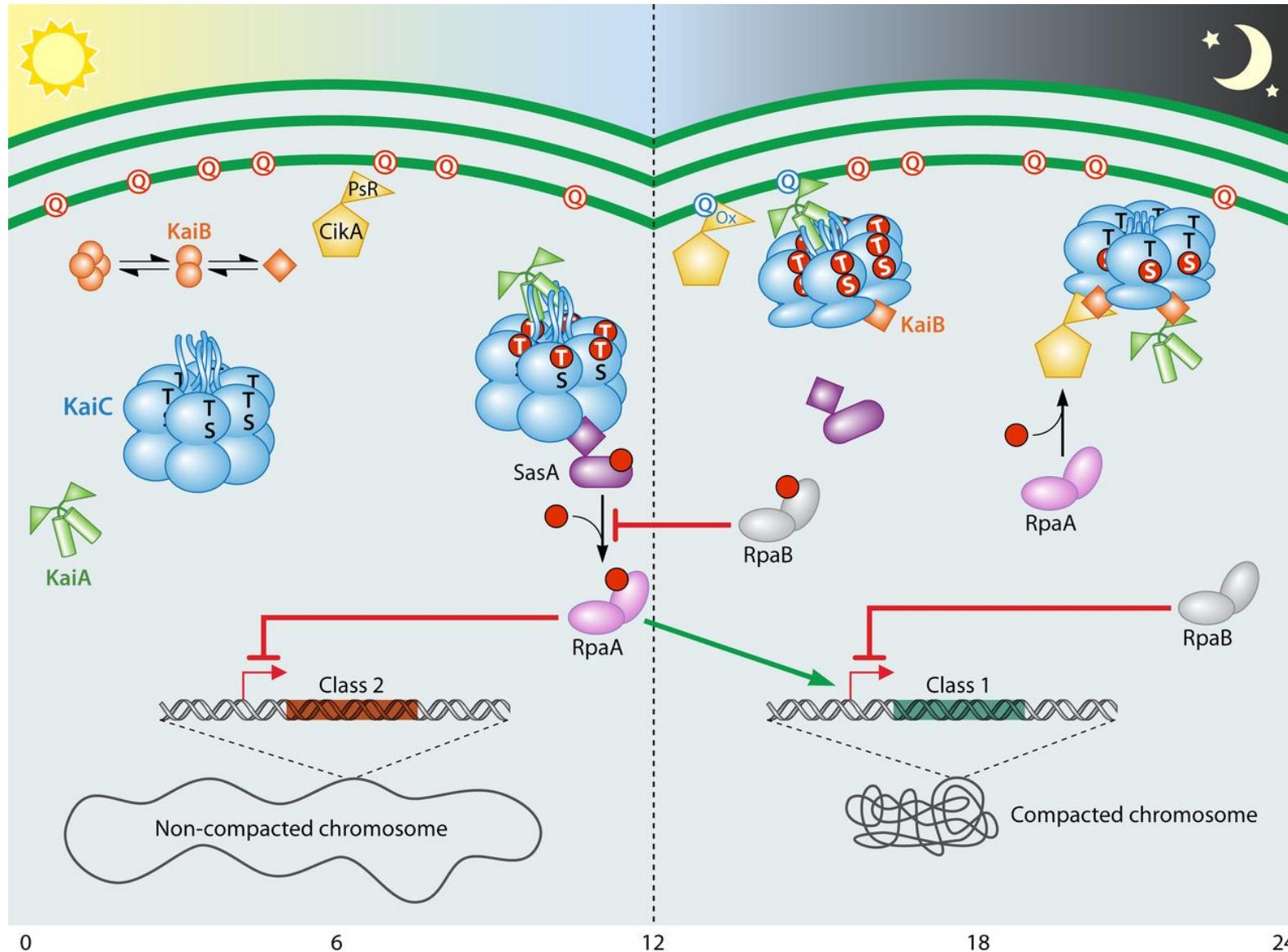
Redza-Dutordoir and Averill-Bates, 2016,  
Biochimica et Biophysica Acta

# Magnetosome formation - MagnetoGenie



Raschdorf et al., 2016, PLOS Genetics

# Circadian rhythm- CircGenie



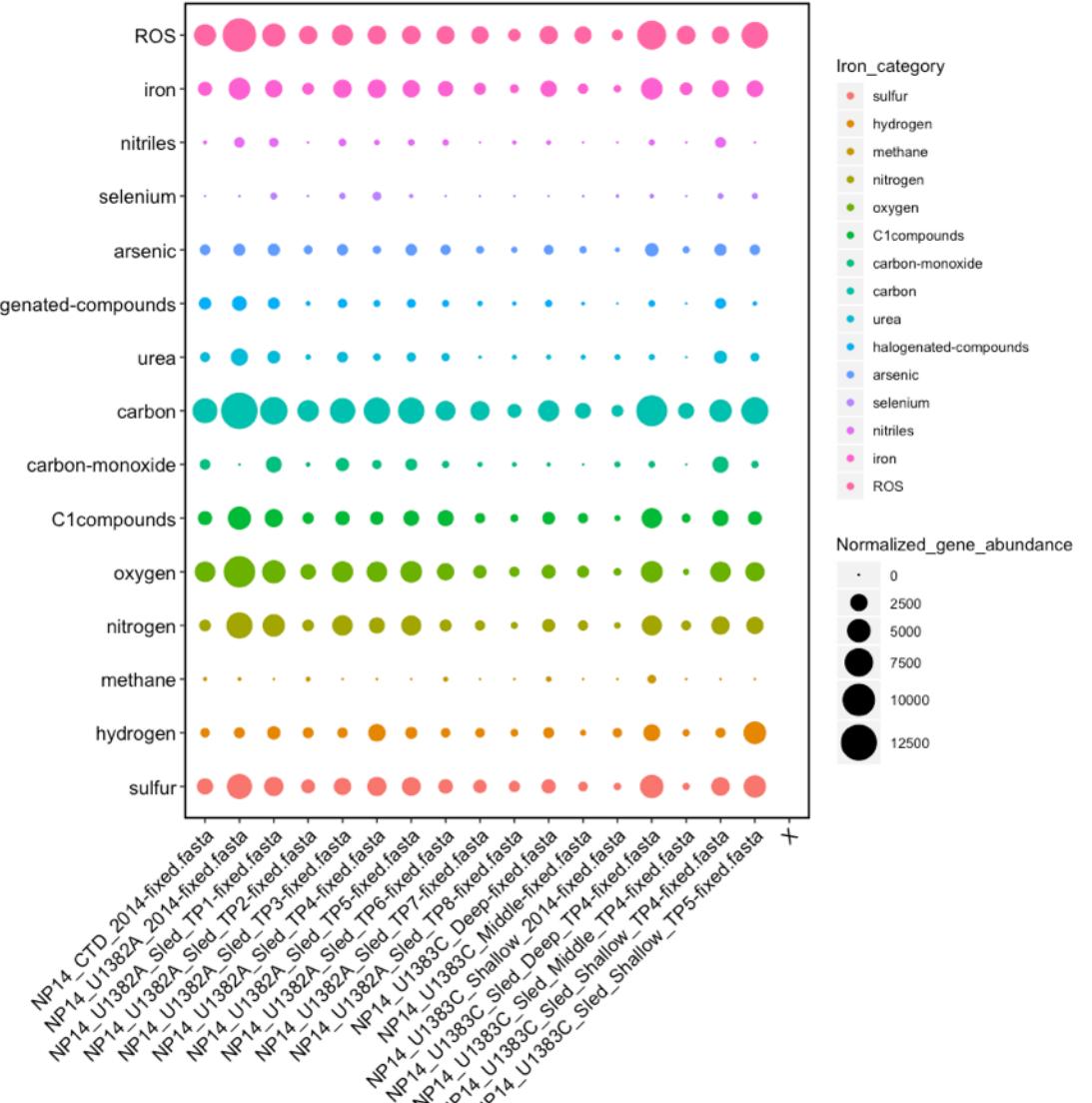
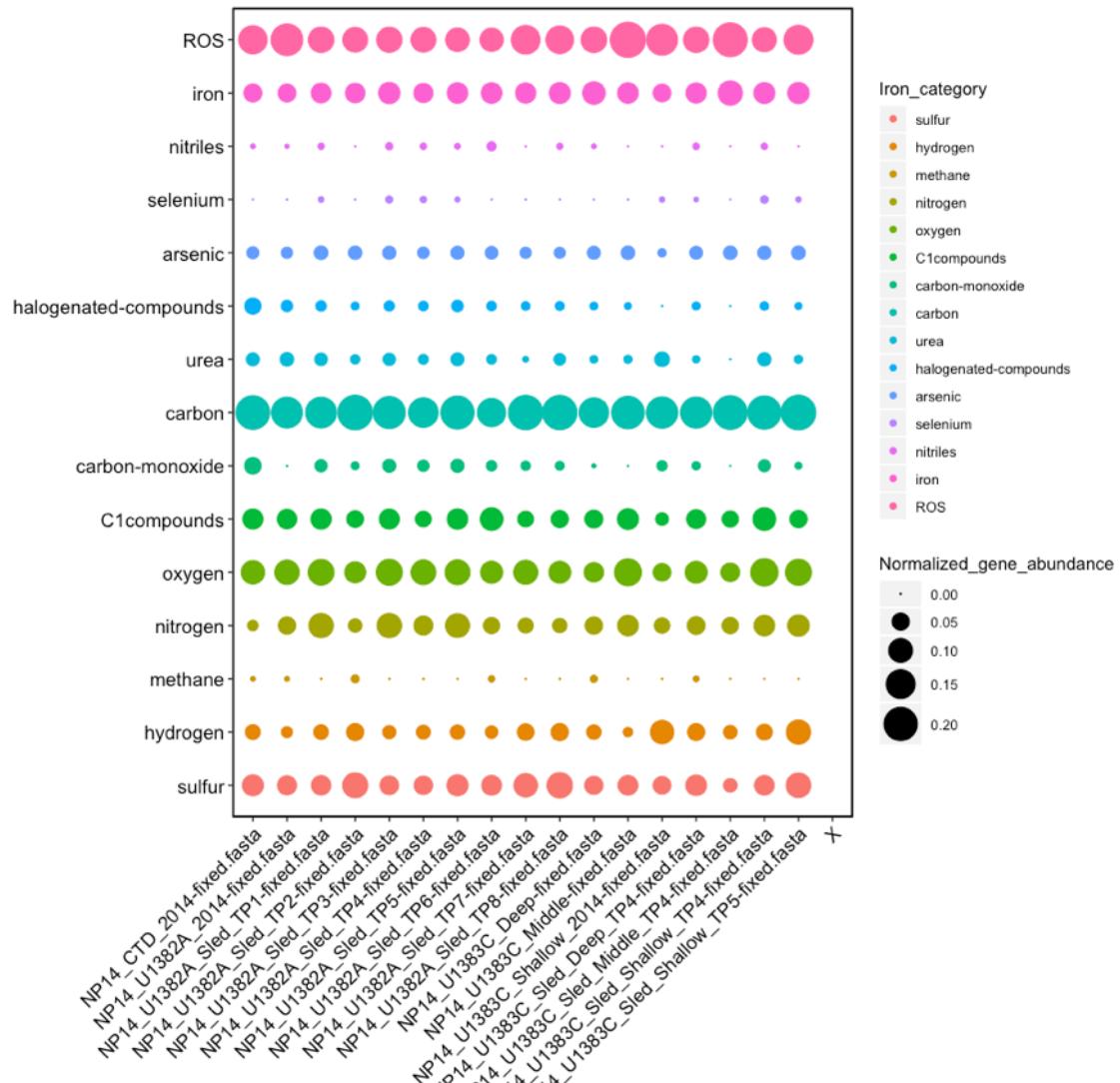
# Lithotrophy - LithoGenie

HMMs compiled by Karthik Anantharaman (UW Madison) – combination of sources (TIGRFAMS, Pfam, and custom HMMs):

[https://github.com/kanant\\_haraman/metabolic-hmms](https://github.com/kanant_haraman/metabolic-hmms)

- Sulfur-cycling
- Iron-cycling
- Nitrogen-cycling
- Methane oxidation/production
- Oxygen reduction
- Hydrogenases
- Arsenic oxidation
- Manganese oxidation (putative)
- Iron-cycling (FeGenie)

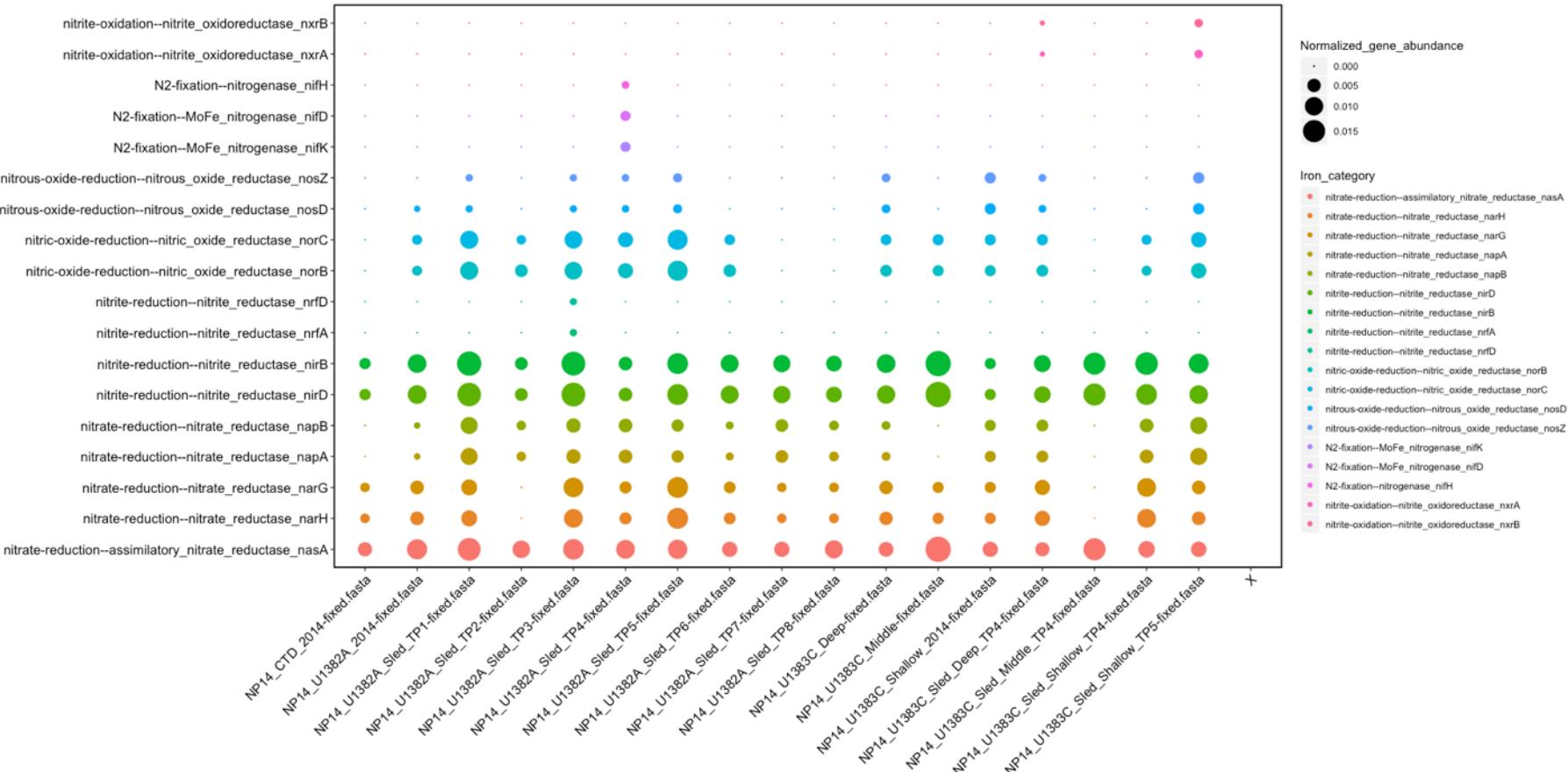
# Once LithoGenie's analyses are finished, it creates a heatmap-style CSV file summarizing the data



*Data from Tully et al., 2017 ISME*

# LithoGenie can be re-run with the ‘–skip’ flag, and a specified substrate category (via -cat flag)

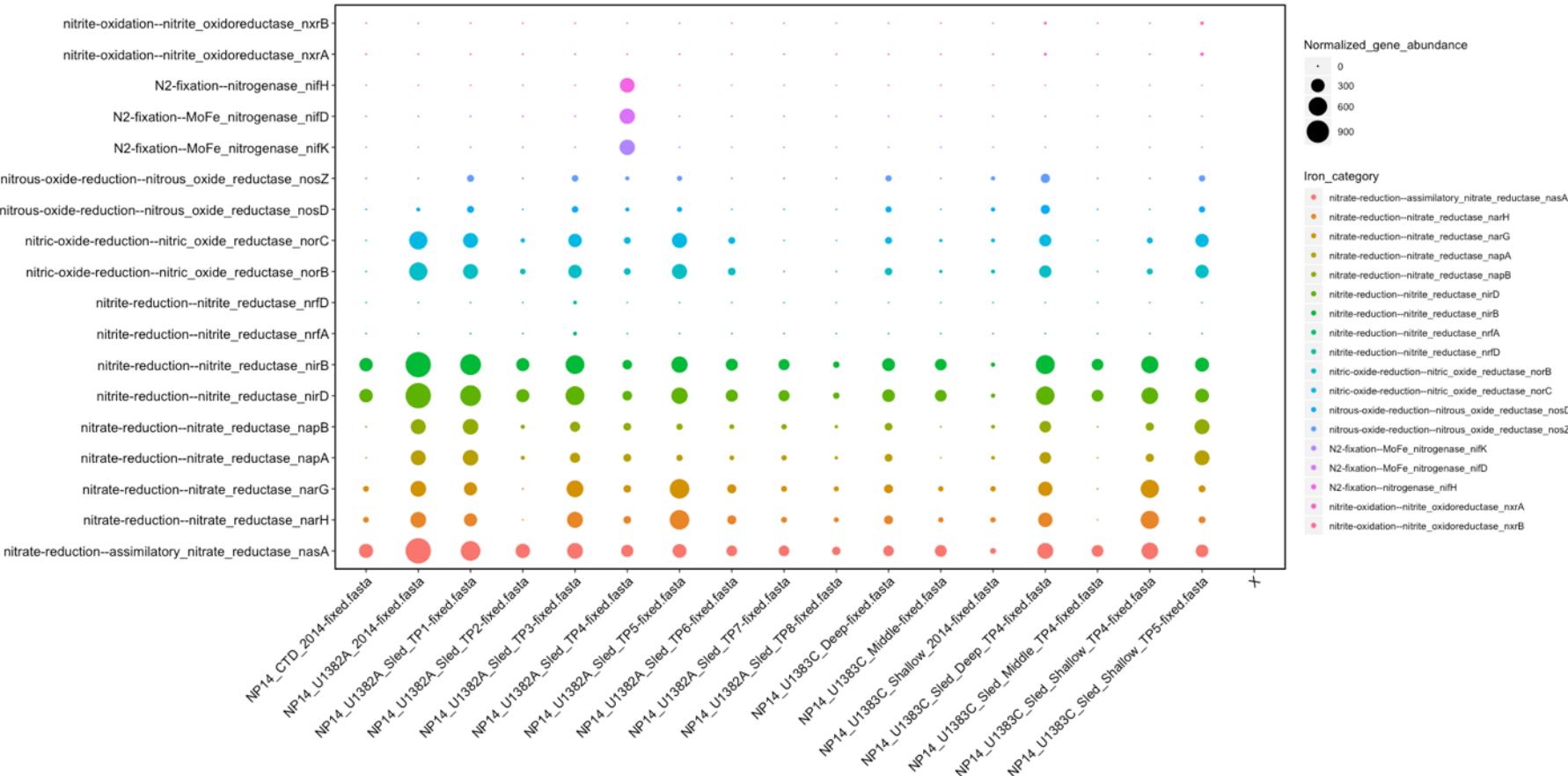
## Nitrogen



Data from Tully et al., 2017 ISME

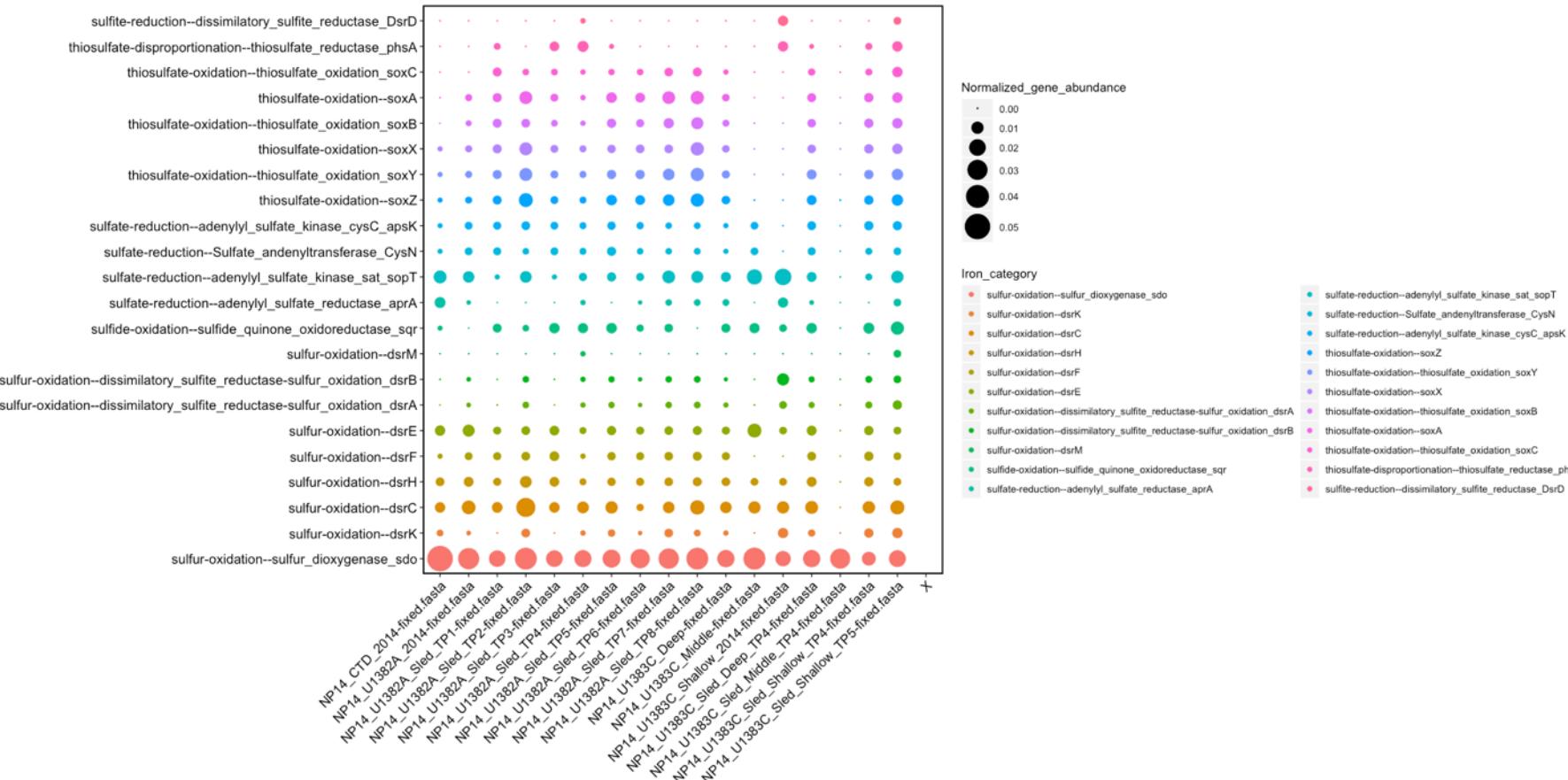
# LithoGenie can be re-run with the ‘–skip’ flag, and a specified substrate category (via -cat flag)

## Nitrogen



# LithoGenie can be re-run with the ‘–skip’ flag, and a specified substrate category (via -cat flag)

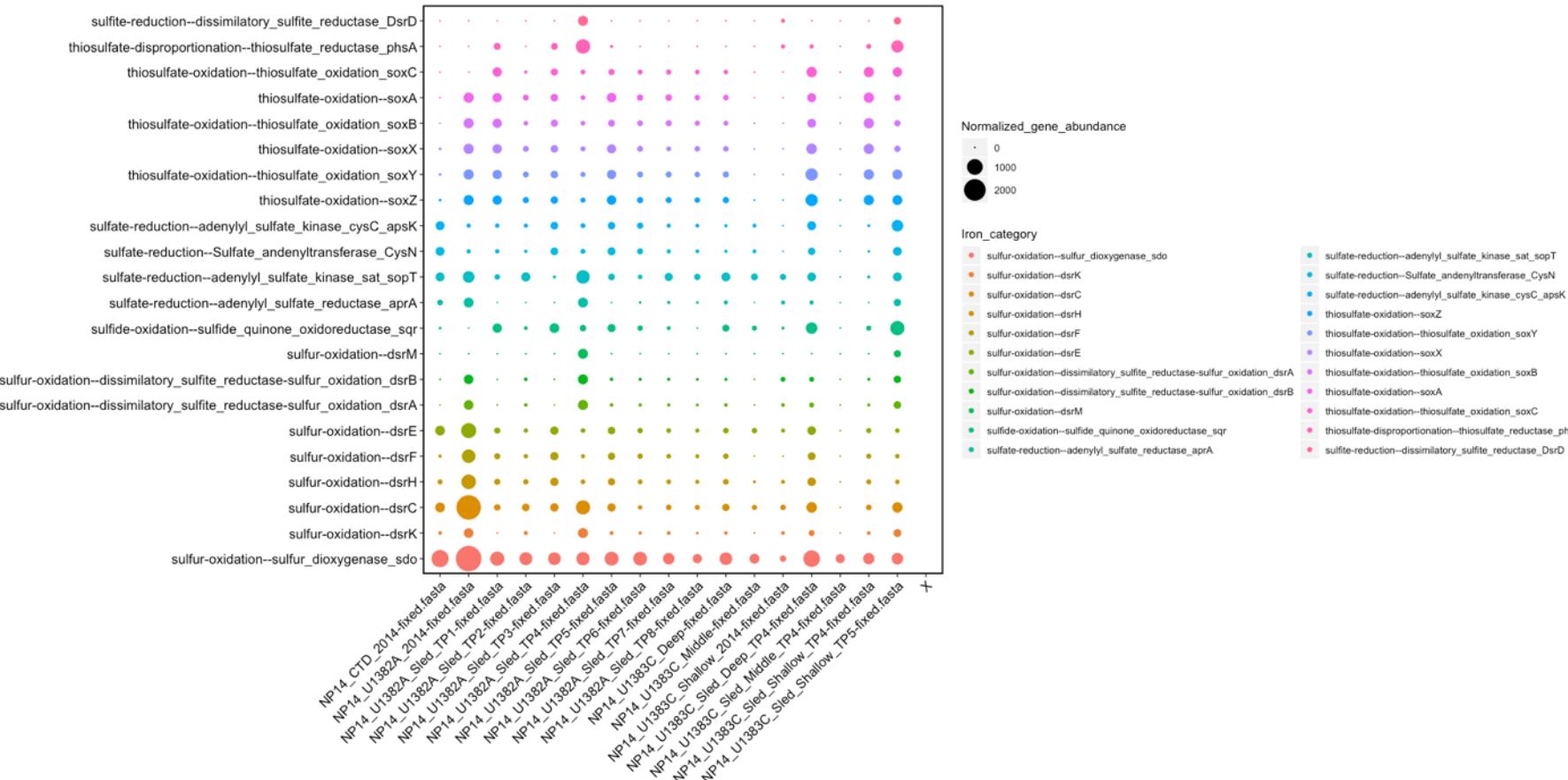
## Sulfur



Data from Tully et al., 2017 ISME

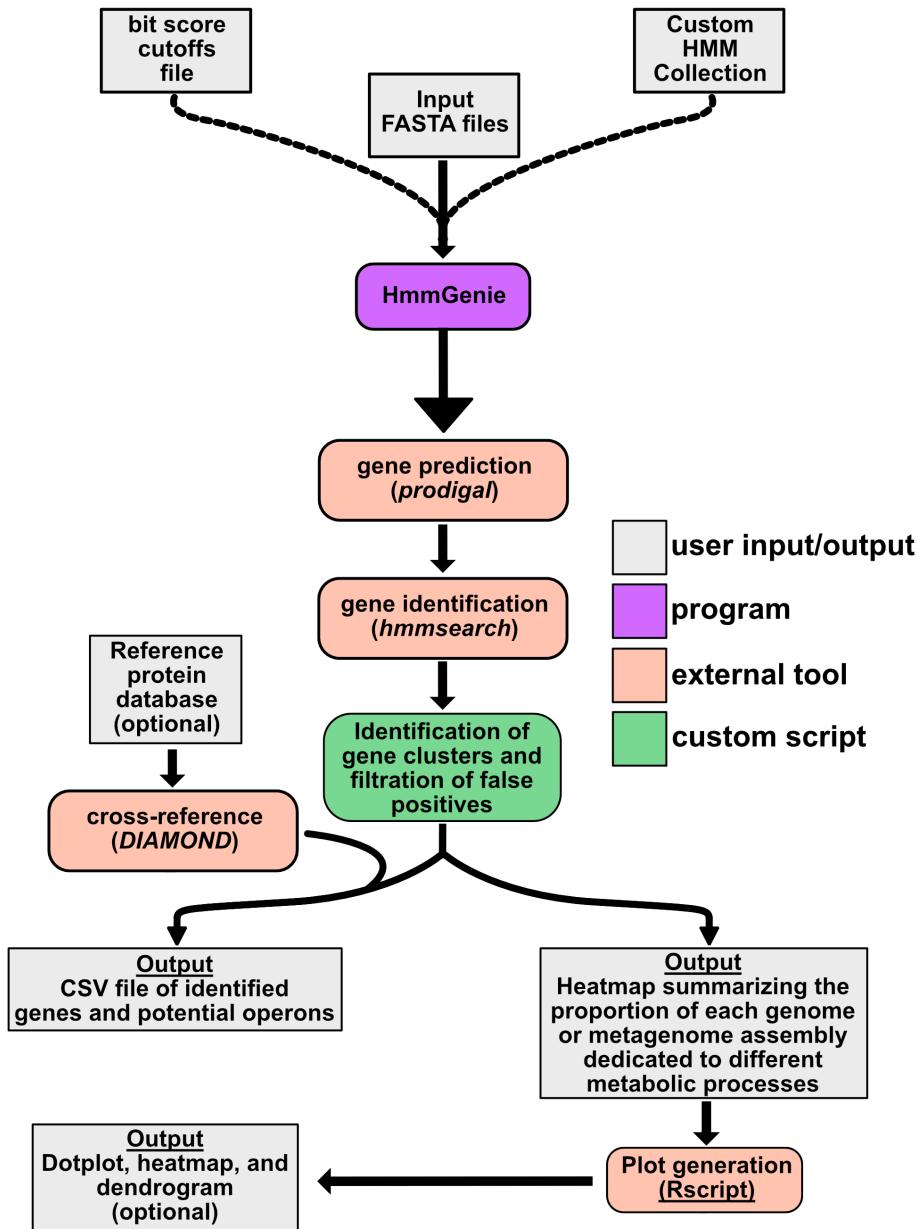
# LithoGenie can be re-run with the ‘–skip’ flag, and a specified substrate category (via -cat flag)

## Sulfur



Data from Tully et al., 2017 ISME

# Making your own genie: HmmGenie





Onto the Jupyter Binder Tutorials:  
MagicLamp and how to make  
your own genie with HmmGenie!

