



University *of* Delaware

Validation of the Short Time-series Expression Miner (STEM)
to Iron Cycling in a Shallow Alluvial Aquifer

Student: Paul Soper

Advisors: Michela Taufer, Clara Chan, and Stephen Herbein



Motivation

- We (Chan) are studying Zetaproteobacteria, an iron-oxidizing bacteria commonly found at deep-sea hydrothermal vents
- Researchers at Carnegie Mellon developed software (STEM) specifically for short time series experiments like those we have performed
- Our research questions are similar to those addressed by the paper *Metatranscriptomic evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling in a shallow alluvial aquifer* by Jewell *et al.*
- We want to use this better understood system to see whether STEM correctly clusters genes involved in iron oxidation



Contributions

- STEM is described in *STEM: a tool for the analysis of short time series gene expression data* by Ernst and Bar-Joseph
- We studied the its application to a relatively well understood data set similar to the one we have for Zetaproteobacteria



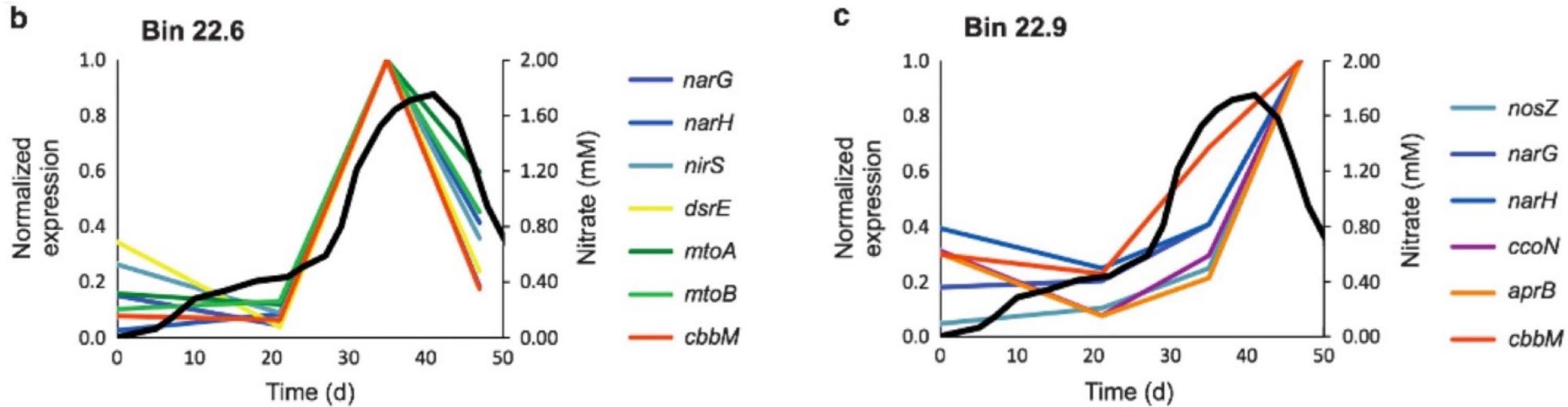
Background

- Short time series of gene expression include a very large number of genes, and a very small number of time points (four in this study)
- Jewell *et al.* studied microbial iron oxidation in a shallow alluvial aquifer in Rifle, Colorado. [ISME Journal (2016) 10, 2106-2117]
- We believe that understanding this system will help us better understand the system we are studying



Background

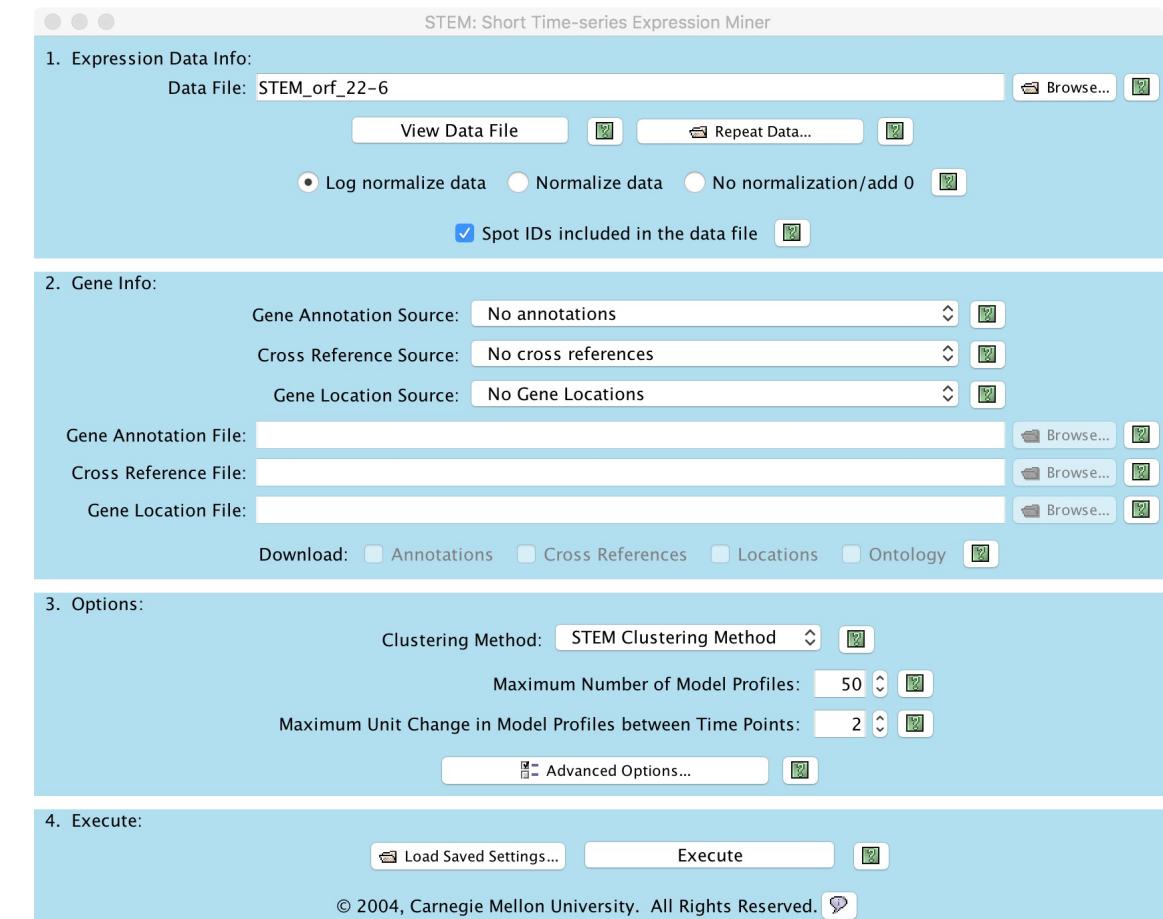
- The data we want to reproduce are presented in two graphs from Jewell *et al.*
- One refers to bin 22.6 and the other to bin 22.9





Methodology – Tools

- We downloaded STEM from Carnegie-Mellon University
- STEM generates an exhaustive set of profiles, then selects the ones that match the data
- Values are the log of the ratio of a gene's expression at time $t > 0$ to its expression at $t = 0$





Methodology – Data

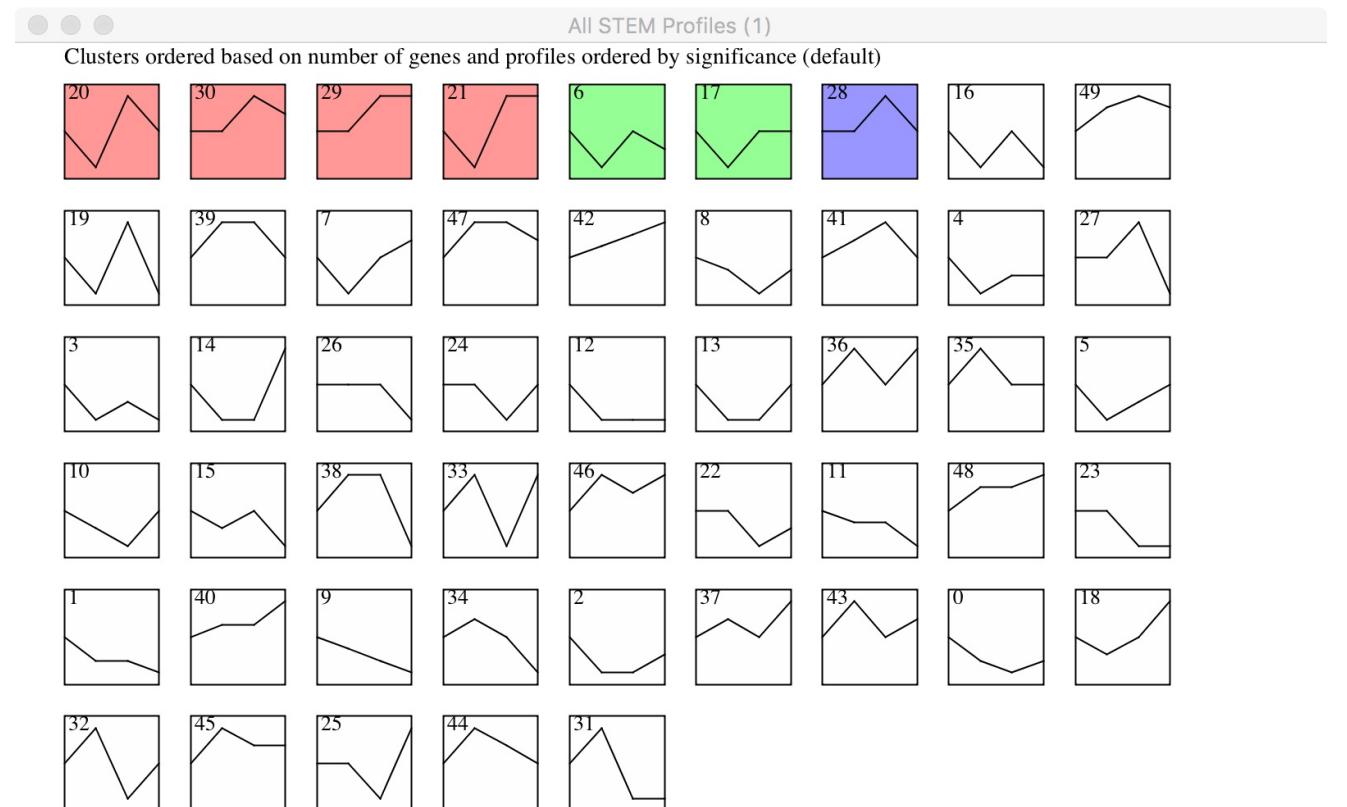
- Data were downloaded from the Supplemental Information provided for the Jewell paper
- Some open reading frames (contiguous stretches of RNA) were labelled with the corresponding gene
 - cyc2
 - mtoA
 - mtoB
 - drsL
 - drsO
- We filtered the data to include only bins 22.6 and 22.9, which were processed separately



STEM Clusters the Bin 22.6 Genes of Interest

- Profiles with colored backgrounds are statistically significant

gene	profile	cluster
dsrL	20	1
cyc2	29	1
mtoA	29	1
dsrO	29	1
mtoB	30	1

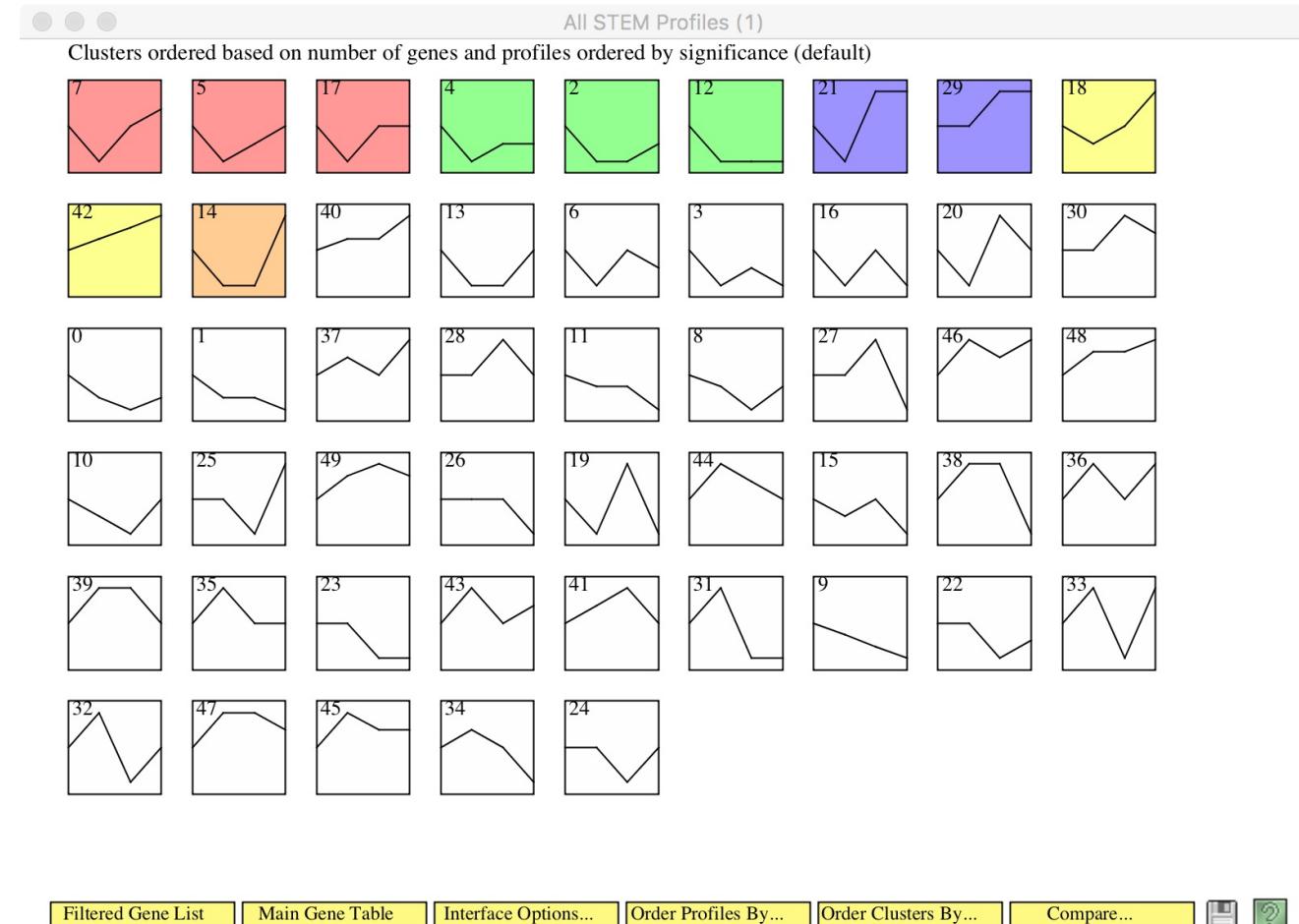




STEM Clusters the Bin 22.9 Genes of Interest

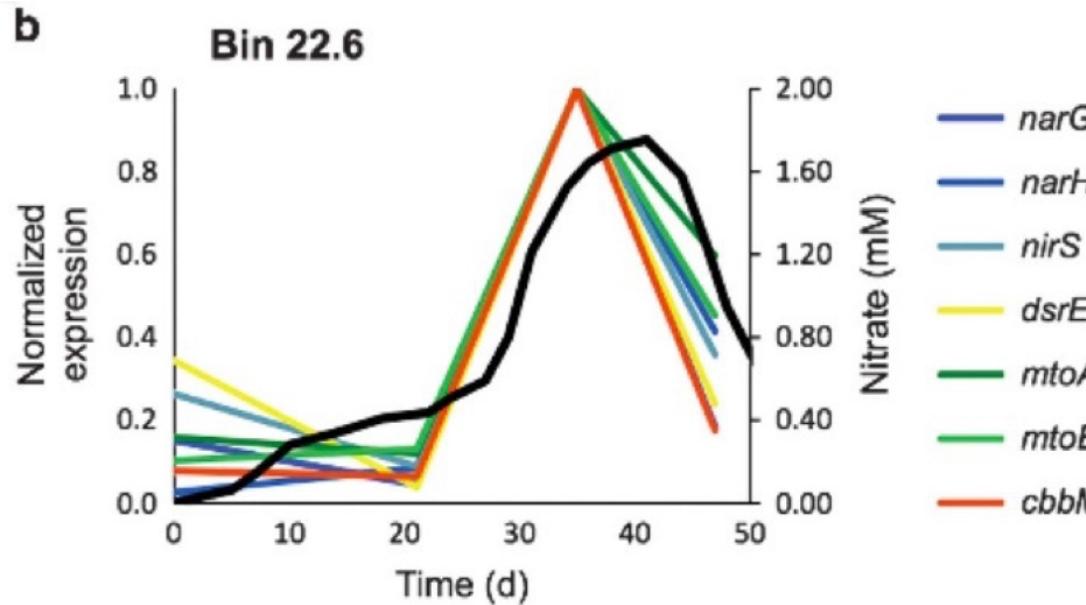
- Profiles with colored backgrounds are statistically significant

gene	profile	cluster
dsrO	5	1
dsrK	5	1
dsrP	5	1
dsrL	7	1
mtoB	21	3
dsrF	21	3
cyc2	42	4

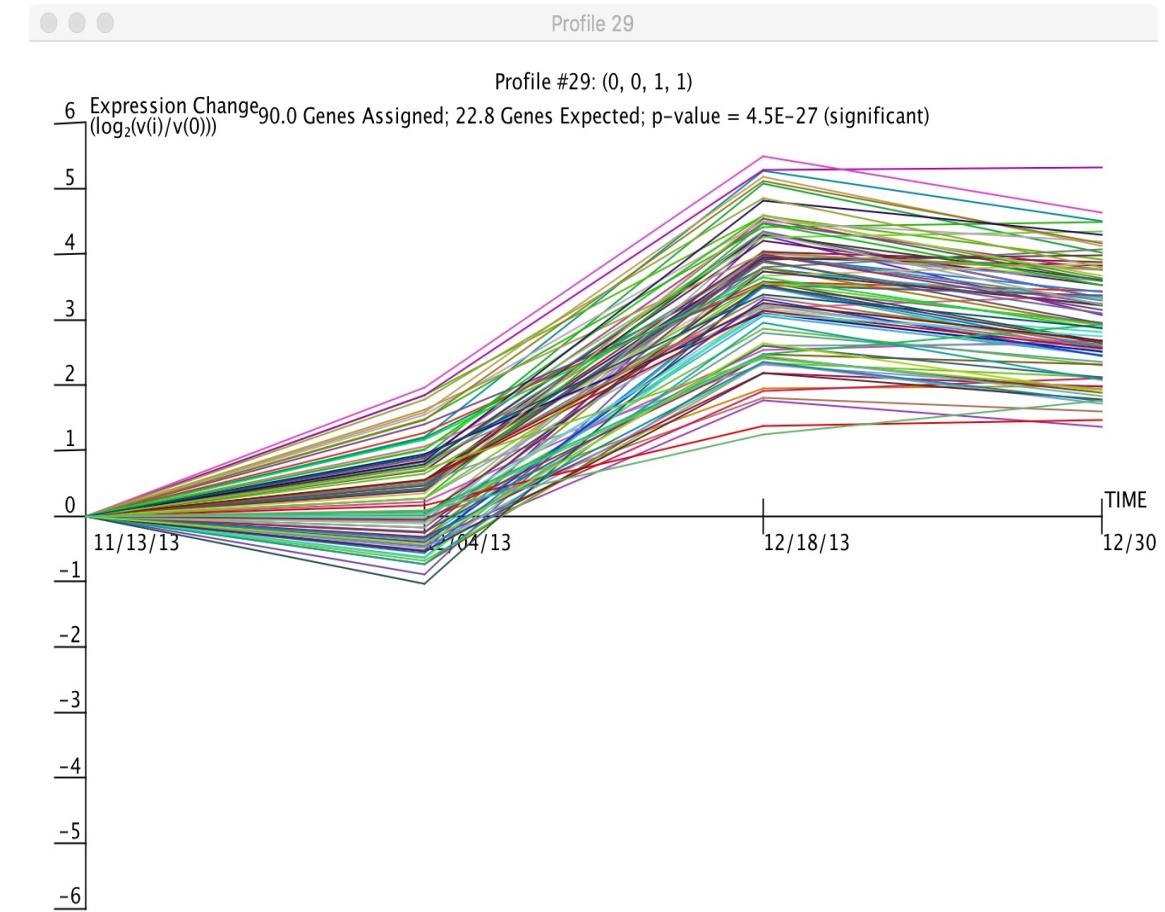




Bin 22.6 Jewell vs. STEM profile 29



Jewell et al.	STEM
Normalized (RPKM)	Ratio to value at t=0



Profile Gene Table

Profile GO Table

Cluster Gene Table

Cluster GO Table



Lessons Learned

- Our work shows that STEM clusters similar time expression profiles of important genes in iron oxidation
- The selected profiles are similar to those reported in Jewell *et al.*