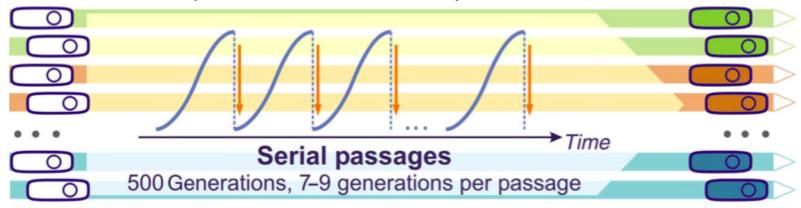
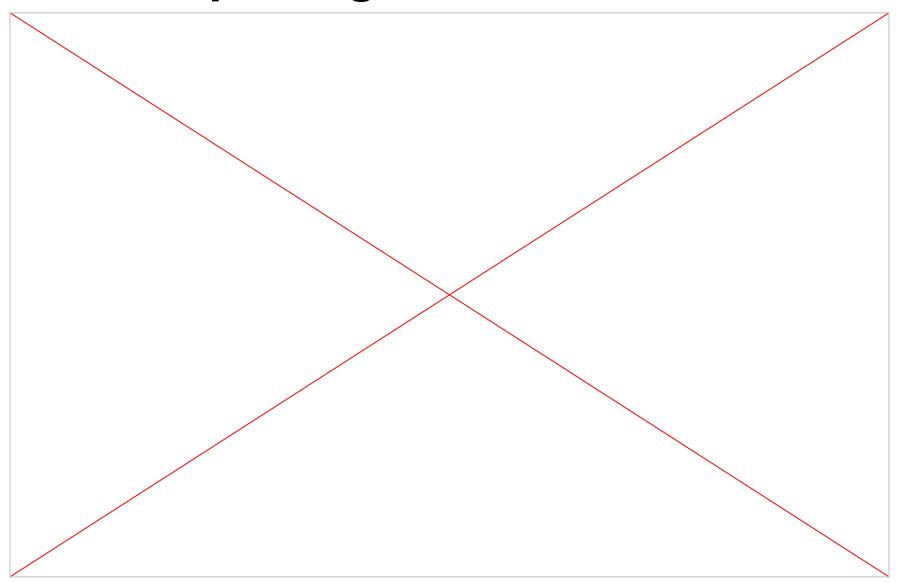


Data set: stress environment evolved strains of *E. coli*

- E. coli K12 MG1655
- Culture Stress Tested:
 - Acidic
 - Oxidative
 - Osmotic
 - N Butanol
 - Control (Minimum Medium)

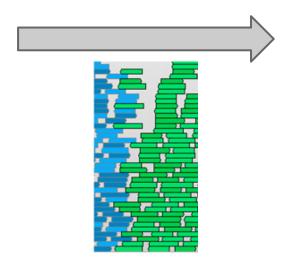


RNA-seq data generation

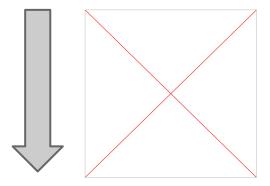


RNA-seq analysis workflow

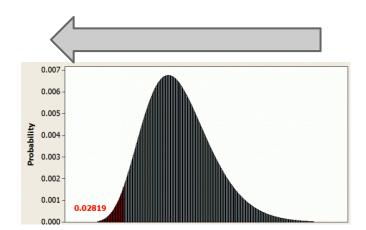
Map reads to reference genome (Bowtie2)



Count reads per gene (HTSeq)

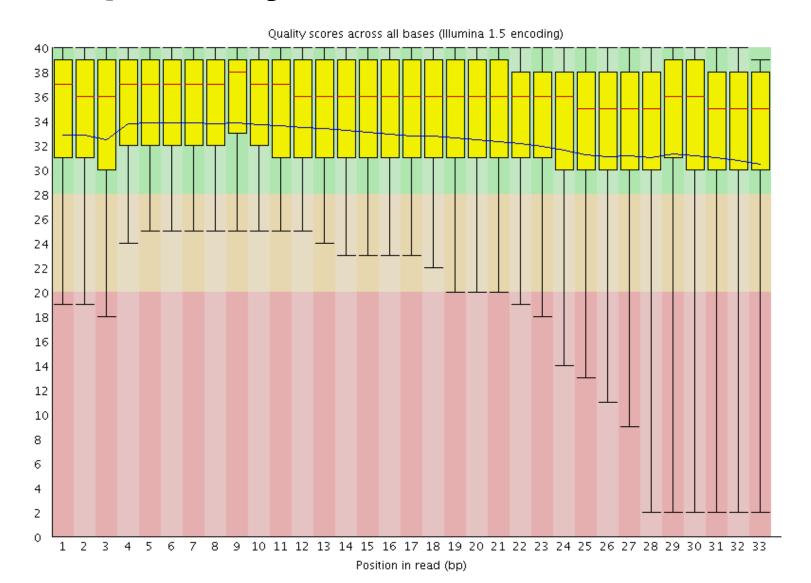


Evaluate probability of gene being differentially expressed



Rescale read counts

Fastq Quality Check

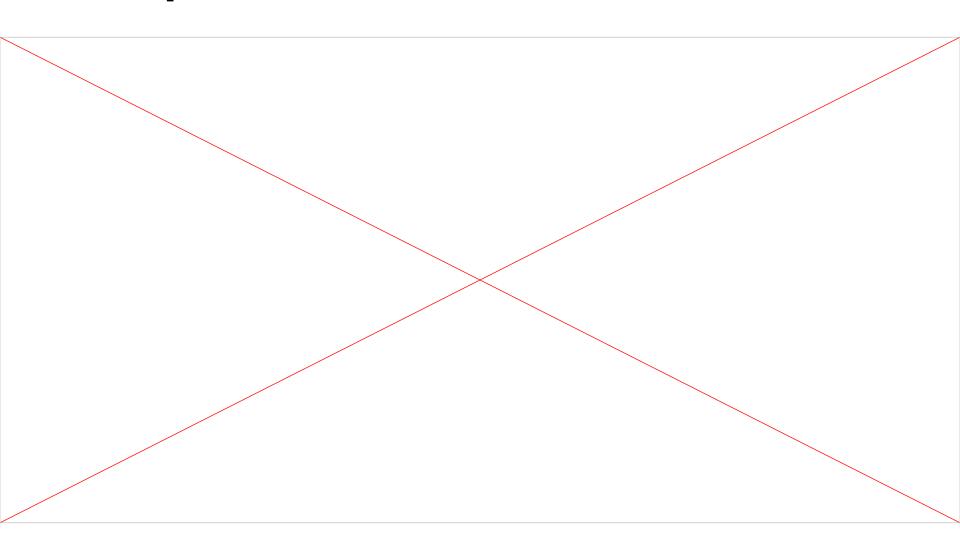


Sequencing Effort/Bowtie Mapping

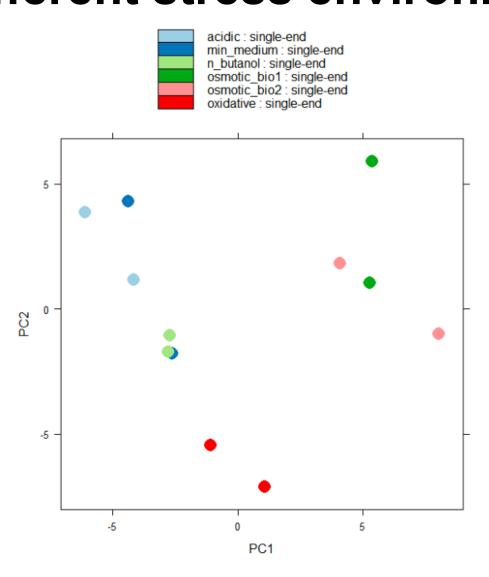
Samples:

- Acidic Conditions = 2 tech replicates
- Oxidative = 2 tech replicates
- Osmotic = 2 tech x 2 bio replicates
- N Butanol = 2 tech replicates
- Control = 2 tech replicates
- Sequencing: ~1-2million single-end reads (33bases/read)
- Mapping: ~70-80% uniquely mapped, ~10-15% aligned >1 times

Heatmap Overview of Pairwise Comparisons of Culture Conditions

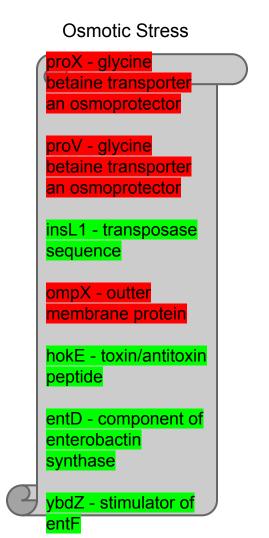


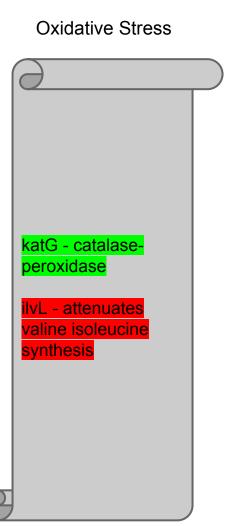
PCA: Qualitative assessment of gene expression similarity between different stress environments



Differentially expressed genes

Acidic Stress flgE - Flagellar regulon flgD - Flagellar regulon flgB - Flagellar regulon

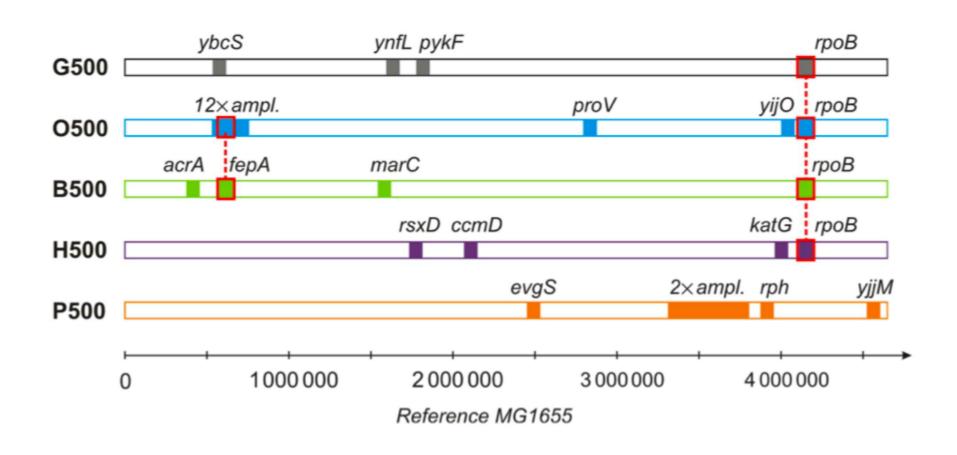




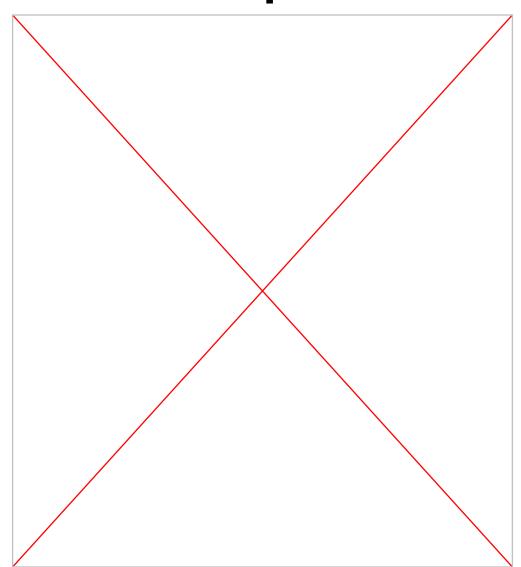
Glycine beatine (osmoprotectant) transporters - like struggling in quicksand



Genetic mutation in addition to gene expression adaptation

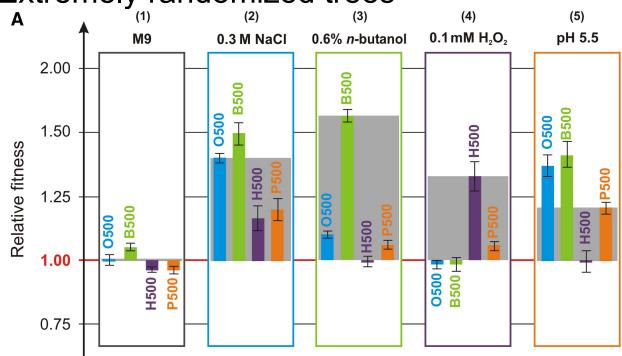


Cross stress protection - potentially a function of Waddington's epigenetic landscape



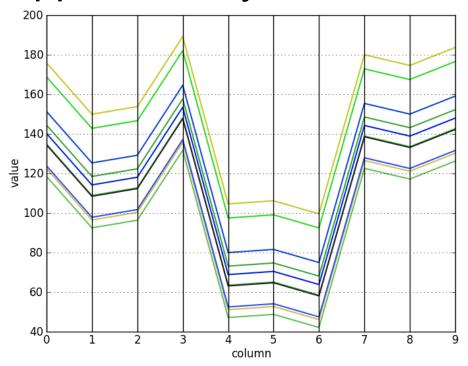
Machine learning approaches

- Regressor to predict fitness
 - LASSO L1 regularized
 - Random forest regressor
 - Extremely randomized trees



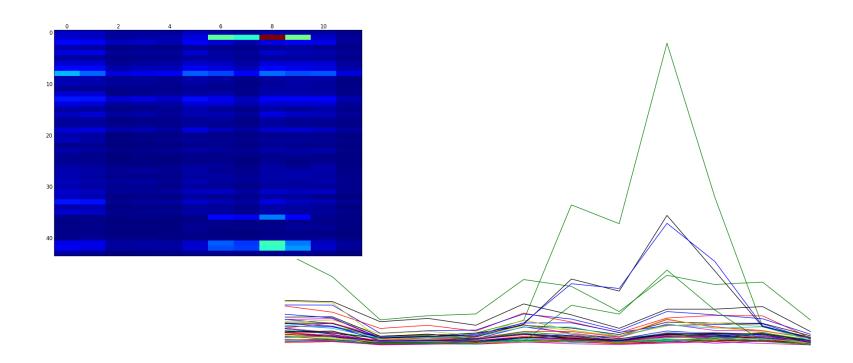
Biclustering of differentially expressed genes

- Algorithm: Cheng and Church
- Objective: low MSR (mean squared residue).
 Rows approximately additive.



Biclustering results

- 100 biclusters, with 3-44 genes
- No enriched GO terms



Questions?