**Introduction**

Over 3500 microbial genomes have been sequenced and this number rises rapidly, leading to new insight in:

- pathogenicity and drug resistance,
- host-microbiome interactions
- survival in extreme environments etc.

Microbes may **adapt to diverse environments by changes in gene expression**, encompassing:

- changes in basal expression levels, and
- changes in transcriptional regulation.

Predicting such changes will prove useful in elucidating various aspects of microbial physiology and ecology.

**Motivation, goals**

It is known that the expression level of a gene influences the rate of evolutionary change in the corresponding protein: **highly expressed proteins evolve slower**.

We aim to systematically investigate whether this **correlation is strong enough to be predictive** of gene / protein levels in practical terms.

Furthermore, a tremendous amount of **high-throughput phylogenetics data is available** e.g. in PhyloMeDB, presenting a unique opportunity for unbiased large-scale screens.

We aim to further develop a data mining methodology for **systematic screens for evolutionary signatures in phylogenomics data**, in theory applicable to any gene functional property.

**Methods**

Reconstructing phylogenetic trees for all proteins across 19 diverse bacterial genomes (phylogenomes) allowed us to compare the **terminal branch lengths** in the trees to experimentally measured mRNA levels.

In addition, we have systematically examined a number of **other features extracted from the phylogenetic trees**.

**Results**

The **signatures of evolutionary history** at the protein sequence level captured by our phylogenetic tree descriptors predict gene expression equally well as the codon biases at the DNA level.

The two sources of expression-related evolutionary signal complement each other.

**Example regression tree for E. coli K12**

![Example regression tree for E. coli K12](image)

The predictive power of each set of features was evaluated using regression trees.

The predictive ability of the tree features varies greatly between the 19 bacterial genomes. Additionally, the correlation coefficients between the codon bias and the tree features are themselves correlated.