

Mutation bias can obscure signals of selection on codon usage

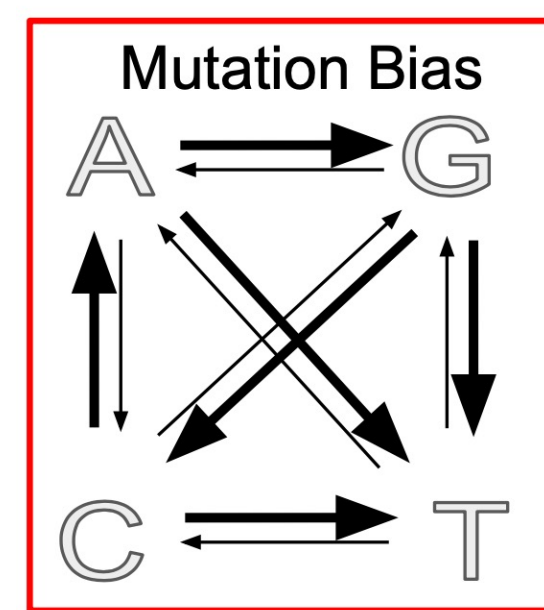
Ahmed Abdellatif; Alexander L. Cope, PhD; Premal Shah, PhD
Department of Genetics, Rutgers University—New Brunswick



INTRODUCTION

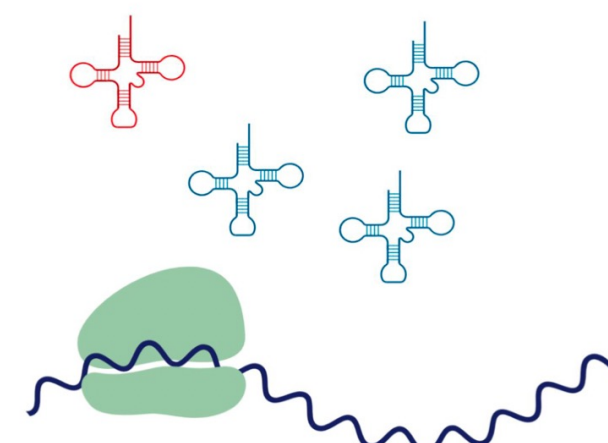
▪ Codon usage bias (CUB): non-uniform usage of synonymous codons

- Histidine: CAT 64%, CAC 36%
- Glutamine: CAA 69%, CAG 31%



▪ Evolutionary forces shape CUB:

- Mutation Bias
- Natural Selection for Translation Efficiency/Accuracy



▪ Key point: Codon usage covaries with gene expression

OBJECTIVE

Quantify across-species variation in codon usage bias

- Separate effects of natural selection and mutation biases
- Use an evolutionary theory-based approach

METHODS

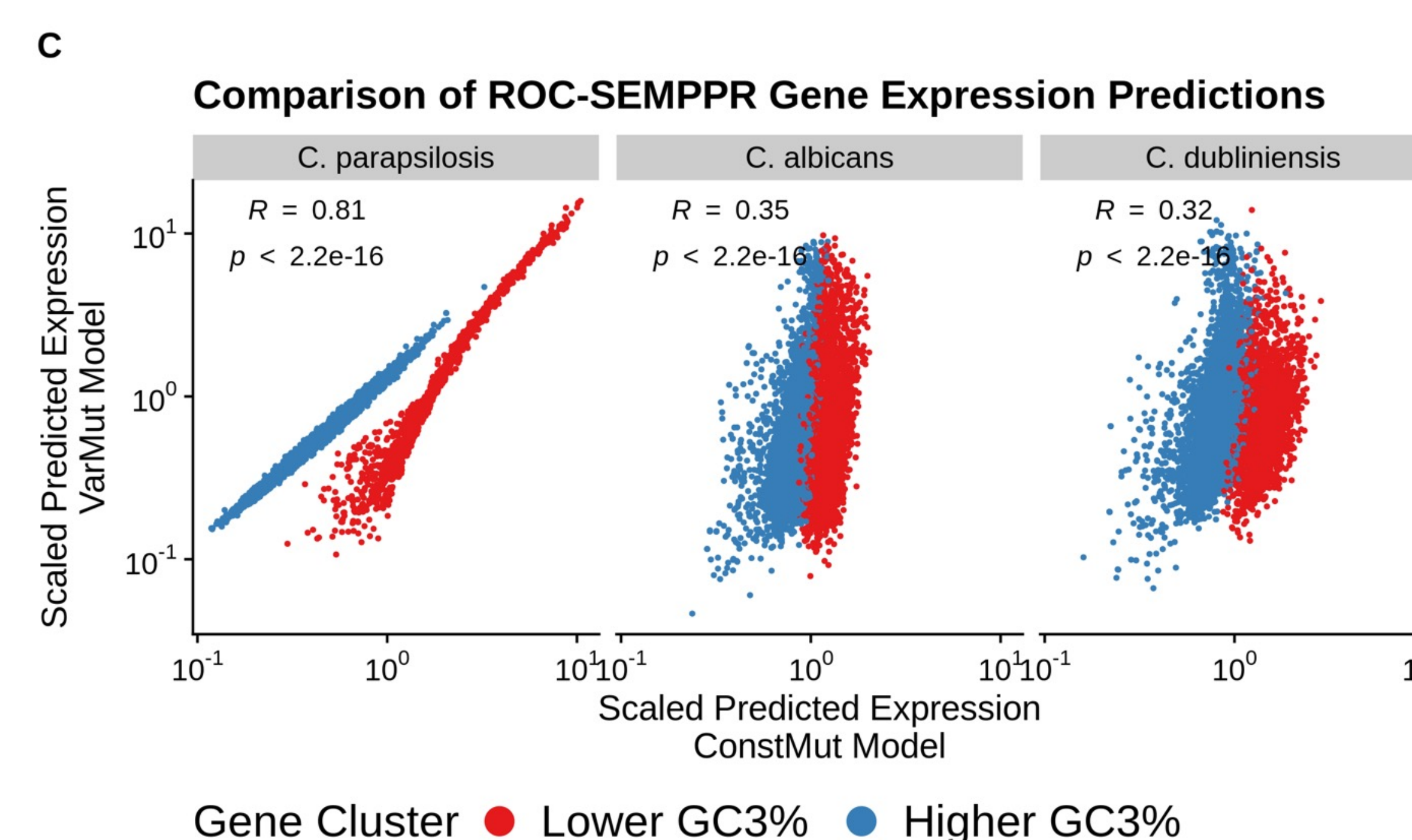
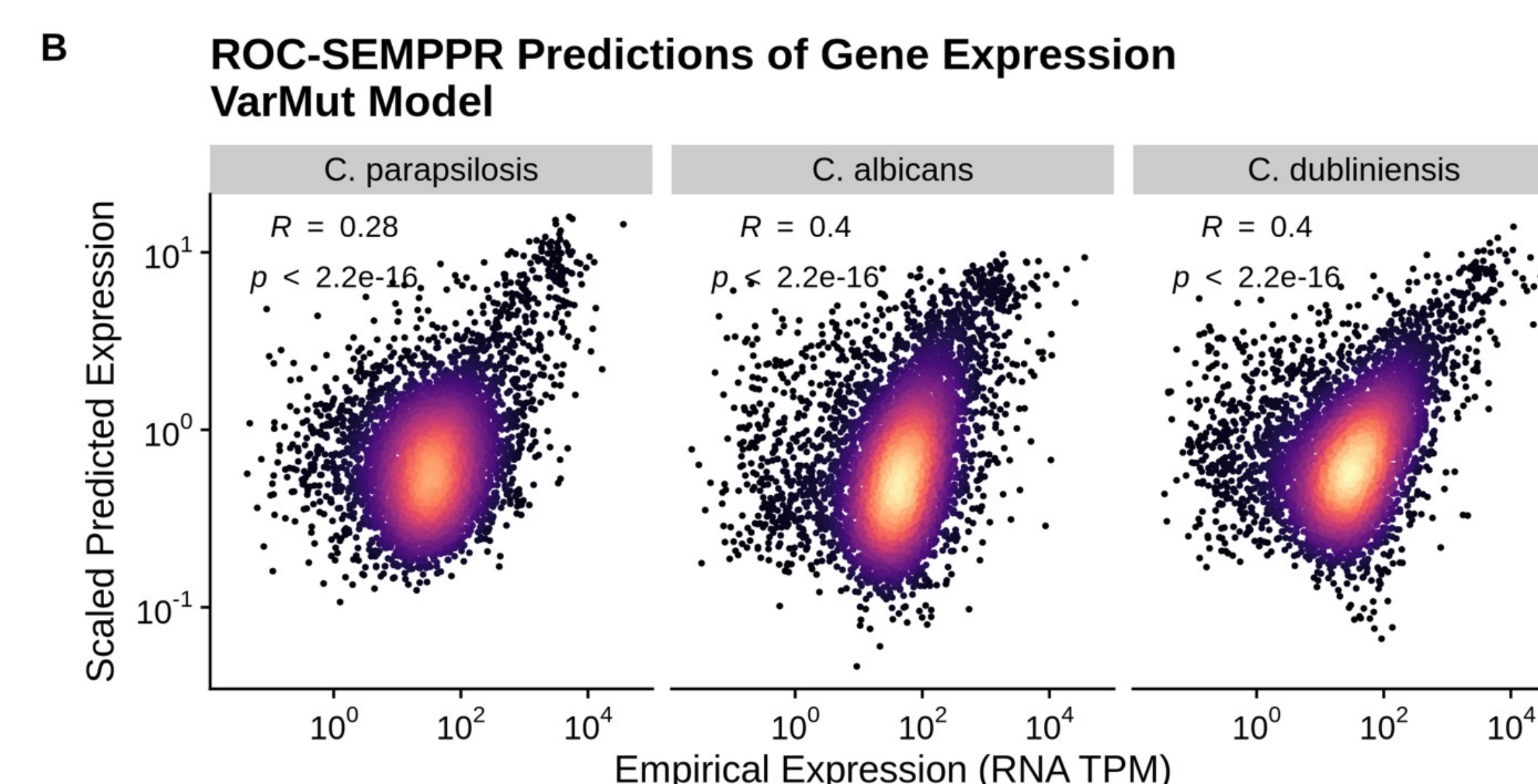
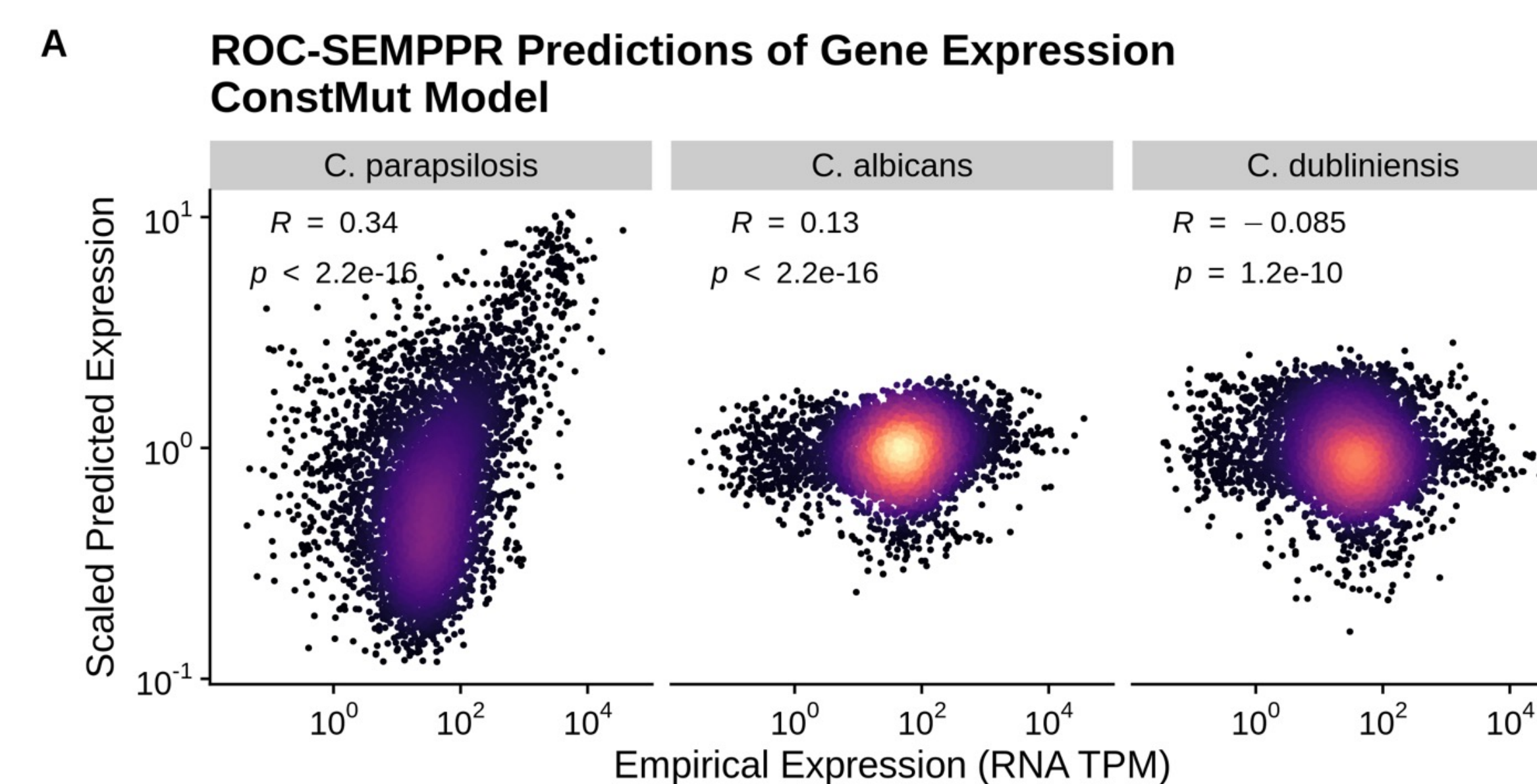
- Examined coding sequences of 49 yeast species
- Predicting gene expression from codon usage
- Expect correlation: no correlation suggests selection is not strong enough to shape codon usage
- Using a clustering approach:
 - Identify genes evolving under different mutational biases
- Clusters: “lower” & “higher” GC3% content
 - Based on avg. freq. of GC nts @ the 3rd codon position
- Evolutionary model: ROC-SEMPPR
 - Allows genes to differ in mutation bias by cluster
- Use the AnaCoDa R package¹

RESULTS

- Estimates of variation in genomic mutation biases are lacking for non-model organisms
- Often do not know which genes are evolving under different mutation biases

➢ **Solution: Unsupervised clustering of genes based on codon usage**

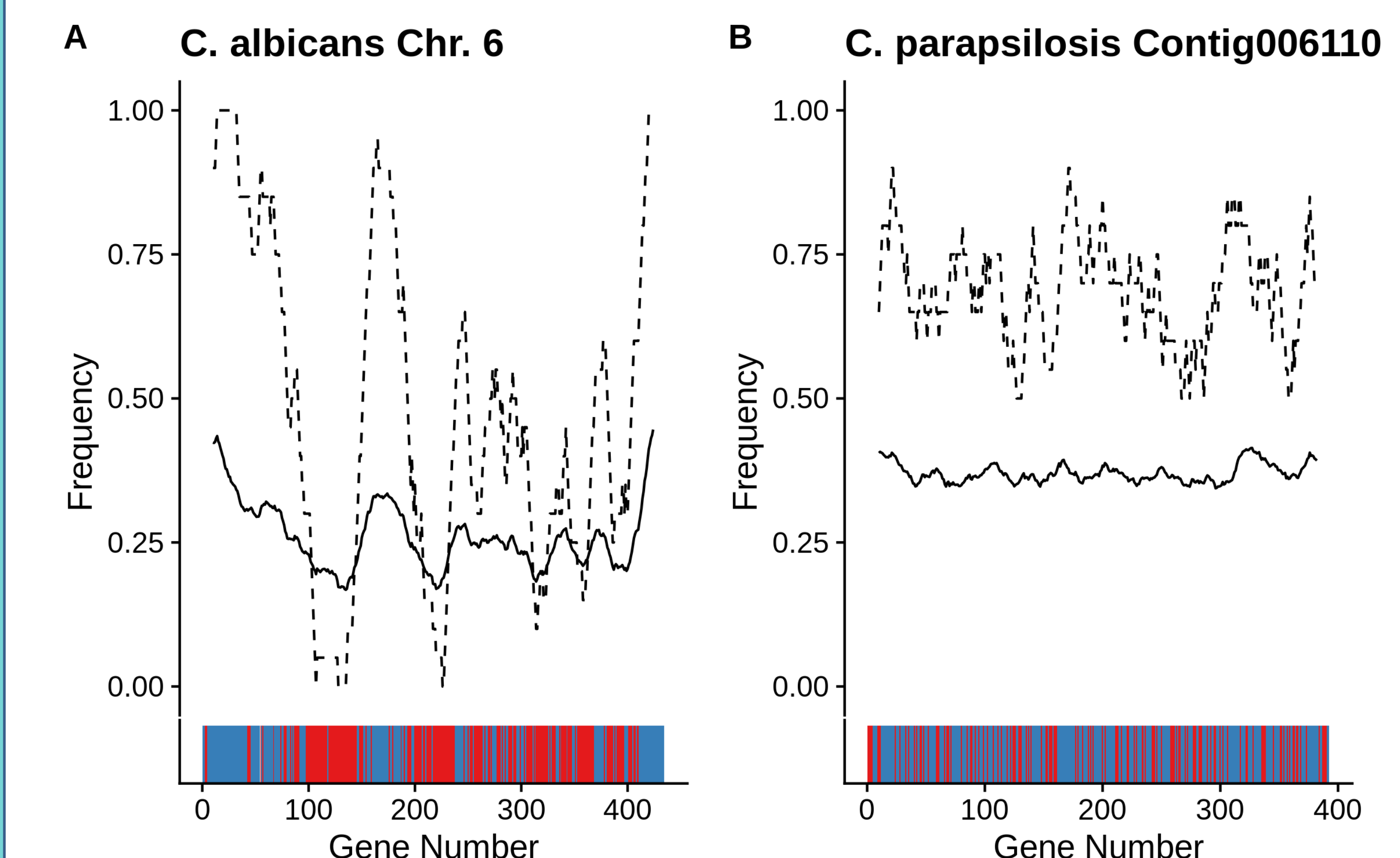
Comparison of predicted and empirical gene expression estimates.



Comparison of predicted and empirical gene expression estimates when (A) assuming mutation bias is the same & (B) allowing mutation bias to vary between the clusters. (C) Comparison of predicted expression estimates between the two models.

RESULTS (cont.)

Variation in GC3% is clustered along chromosomes in some species.



Moving Average

— GC3%
- - Percentage of Genes in Higher GC3% Cluster

Gene Cluster

■ Higher GC3%
■ Lower GC3%

Changes in GC3% content by gene along specified chromosomes in (A) *C. albicans* & (B) *C. parapsilosis*. Solid lines represent moving average of GC3% across 20 gene-long windows, and dashed lines represent number of genes assigned to the higher GC3% cluster in the same window.

CONCLUSIONS

- Failure to account for variation in mutational biases due to processes such as biased gene conversion:
 - Potentially obfuscate signals of natural selection
- Allowing mutation bias to vary can reveal signatures of selection on codon usage

REFERENCES

1. Landerer C, Cope A, Zaretzki R, Gilchrist MA. AnaCoDa: analyzing codon data with Bayesian mixture models. *Bioinformatics* [Internet]. 2018; bty138.