

2 **Differences in Codon Usage Bias between genomic**  
3 **regions in the yeast *Lachancea kluyveri*.**

4 CEDRIC LANDERER<sup>1,2,\*</sup>, RUSSELL ZARETZKI<sup>3</sup>, AND MICHAEL  
5 A. GILCHRIST<sup>1,2</sup>

6 <sup>1</sup>Department of Ecology & Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-  
7 1610

8 <sup>2</sup>National Institute for Mathematical and Biological Synthesis, Knoxville, TN 37996-3410

9 <sup>3</sup>Department of Business Analytics & Statistics, Knoxville, TN 37996-0532

10 \*Corresponding author. E-mail: cedric.landerer@gmail.com

Version dated: April 16, 2018

## Abstract

Large efforts have been made to develop and explore models to understand intra-genomic variation in codon usage bias (CUB) and the contributions of mutation and selection to its evolution. Comparative studies have been undertaken to further our understanding of variation in codon usage between species. However, limited efforts have been made to understand how CUB is affected, and in return effects hybridization or introgression events between species with potentially large differences in CUB. In this study, we explore the CUB of *Lachancea kluyveri* which has experienced a large introgression covering the whole left arm of chromosome C, affecting about 10% of all genes. The *L. kluyveri* genome provides insights about the adaptation of introgressed regions to the novel genomic environment, with potentially large differences in selection for translation efficiency due to factors like tRNA availability, effective population size, or differences in mutation environment.

We analyzed the CUB of the endogenous *L. kluyveri* genome and compared it to the CUB of the exogenous, introgressed region while separating the effects of mutation bias and selection for translation efficiency on CUB. Our results show distinct CUB between the endogenous and exogenous regions of the *L. kluyveri* genome. We show that this differences can be mostly attributed to differences in mutation bias.

The introgression into the *L. kluyveri* genome is of additional interest as the source has not yet been identified. Given our ability to clearly distinguish CUB between the exogenous and the endogenous region we explored if CUB can identify possible candidates for the origin of the introgression. The estimation of CUB and its separation into contributions of mutation and selection across a variety of yeasts allowed us to identify two candidates for the origin of the exogenous genes. We used orthogonal information about synteny to validate candidates obtained by matching CUB.

## Outline

- CUB changes due to differences in mutation, selection, and drift.

- most studies assume only one environment for mutation, selection and drift and therefore only one codon usage.

  - This assumptions can be violated for multiple reasons, like introgression/horizontal gene transfer (HGT), population bottlenecks, etc.
- Variation in CUB has previously only been studied in bacteria where HGT is common.

  - HGT only transfers small amount of genes, probably with little to no impact on overall CUB.
  - However, exogenous material can accumulate if HGT is frequent [2].
  - Previous studies have shown that genes with similar CUB are more likely to be transferred, potentially mitigating effects of accumulation [4].
  - Hybridization/Introgression should have a larger impact on CUB due to the amount of material transferred, possibly affecting the outcome of a study if ignored.
- In this study, we look at *L. kluyveri*.

  - *L. kluyveri* has experienced a recent ( $55.5e10$  generations) large scale introgression [1], clearly marked by elevated GC-content [3].
  - We expect that CUB differs between the introgressed exogenous region and the endogenous region due to the great (13%) difference in GC-content between the two regions.
    - \* We find differences in CUB between the two regions.
    - \* Taking this difference into account, we can increase our ability to extract biological information (predicting gene expression).
    - \* Thanks to our ability to distinguish between effects of mutation and selection on CUB, we are able to attribute most of the difference in CUB to mutation bias.

\* Figure 2 shows the CUB if we ignore the introgression (dotted), and for the endogenous (solid) and exogenous (dashed) respectively.

– At this point, the source of the introgression has not been identified.

\* Since we can clearly distinguish between the endogenous and exogenous CUB, can we use this information to find possible donor organisms?

\* We analyzed CUB for several yeasts and found several species with similar selection for translation efficiency, and a few with similar mutation bias, but only two with high agreement in both (*gossypii* and *dubliensis*, Figure 3).

\* We validated our findings with orthogonal information from synteny where analyzed a subset of our initial yeast set.

\* We found several closely related species with syntenious regions, but only one species that also showed agreement in CUB allowing us to exclude *dubliensis* (Figure 4)

## Introduction

## Results

## Discussion

## References

[1] A Friedrich, C Reiser, G Fischer, and J Schacherer. Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Molecular Biology and Evolution*, 32(1):184 – 192, 2015.

[2] JG Lawrence and H Ochman. Amelioration of bacterial genomes: Rates of change and exchange. *Journal of Molecular Miology*, 44:383–397, 1997.

- [3] Clia Payen, Gilles Fischer, Christian Marck, Caroline Proux, David James Sherman, Jean-Yves Coppe, Mark Johnston, Bernard Dujon, and Ccile Neuvglise. Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Research*, 19(10):1710–1721, 2009.
- [4] T Tuller, Y Girshovich, Y Sella, A Kreimer, S Freilich, M Kupiec, U Gophna, and E Rupp. Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Research*, 39(11):4743–4755, 2011.

## Figures and Tables

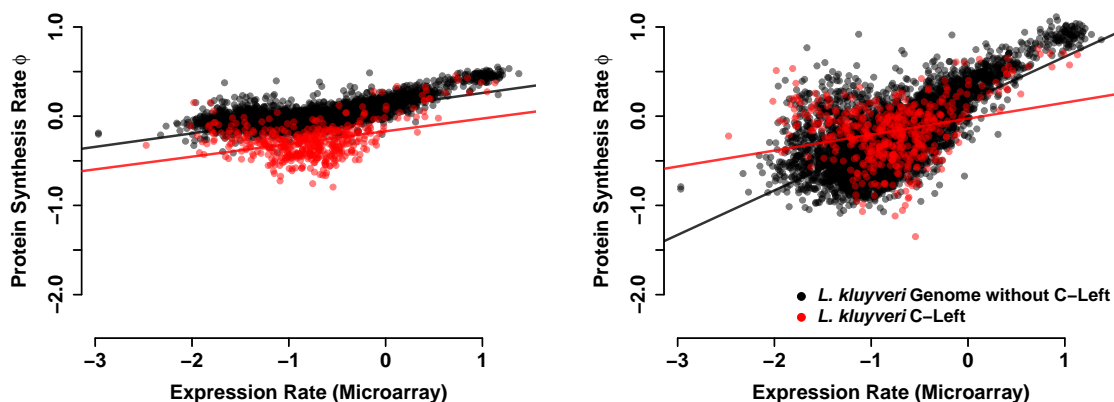


Figure 1: Person correlation of predicted protein synthesis rate  $\phi$  with observed expression rate. )

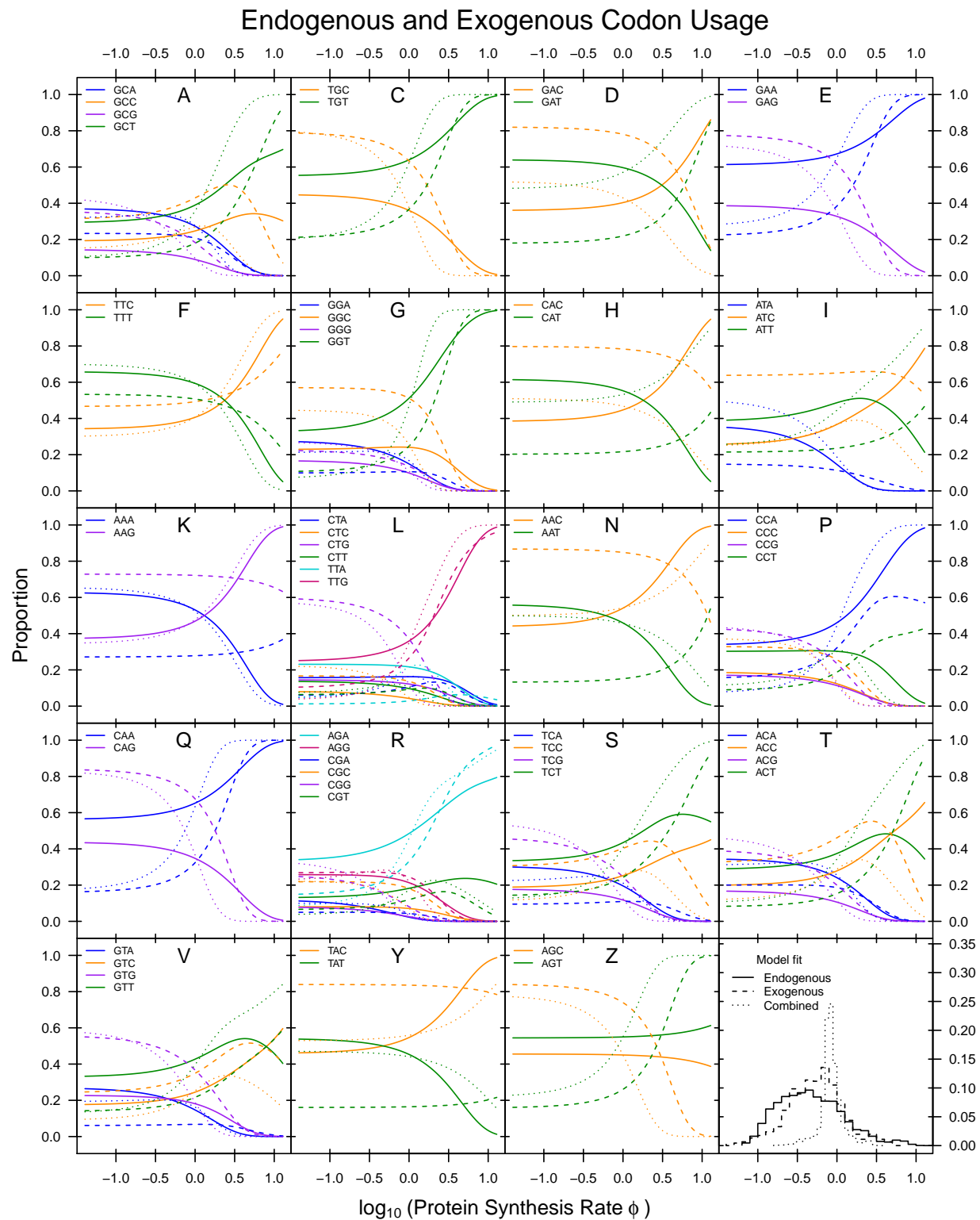


Figure 2: Codon Usage

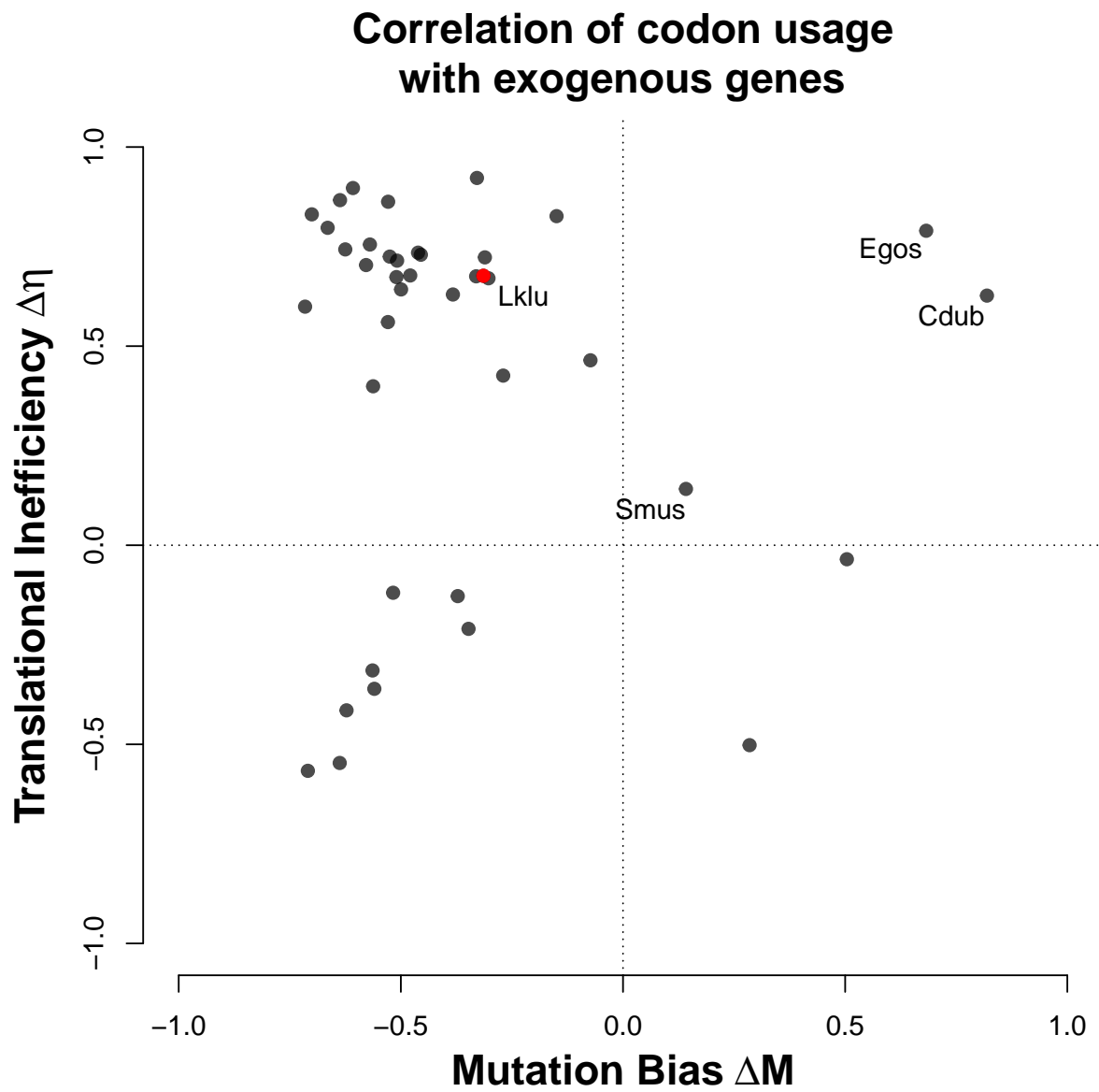


Figure 3: Codon Usage

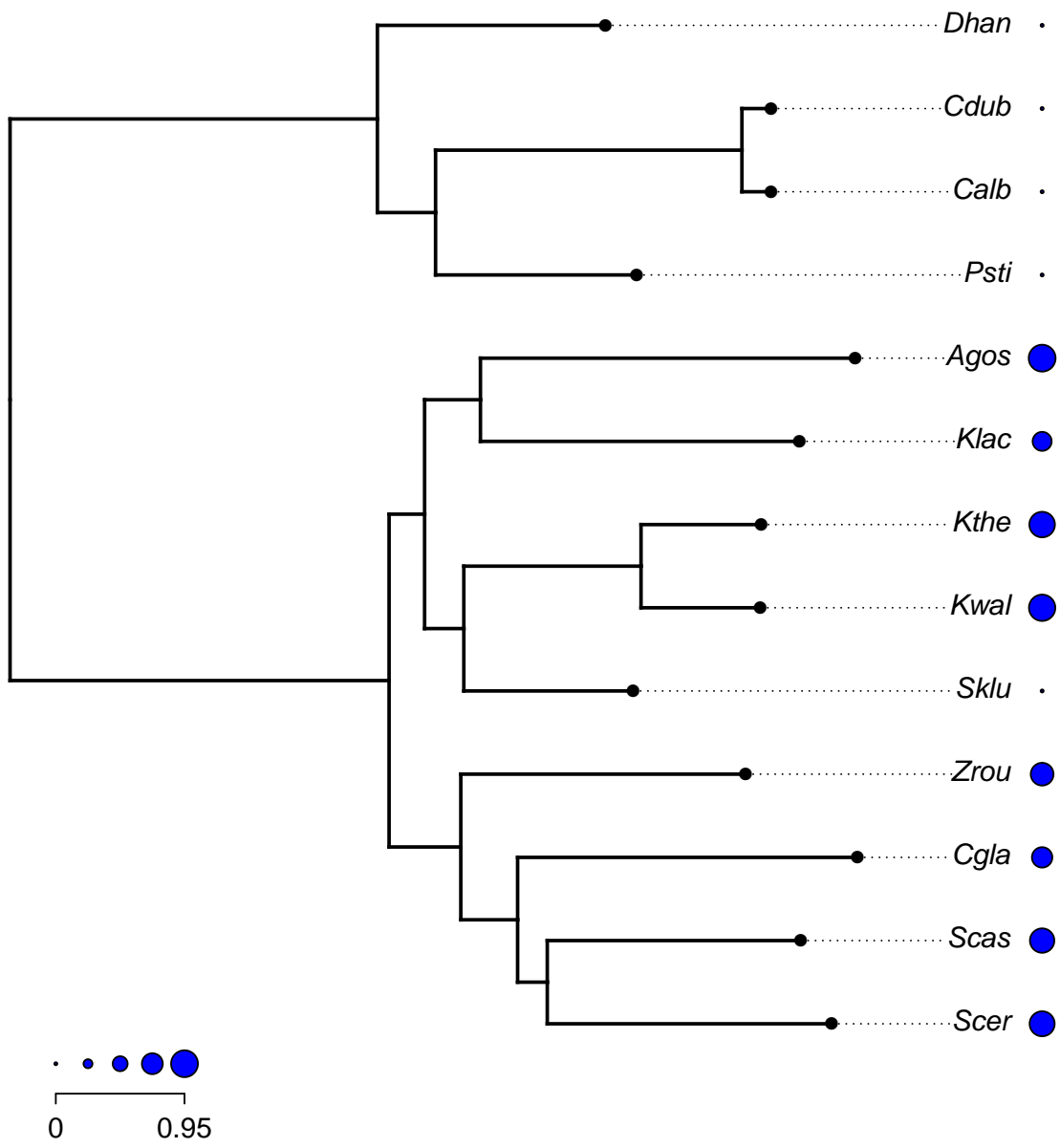


Figure 4: Codon Usage