

Fundamentals of Codon Bias

Willow Kion-Crosby¹, Michael Manhart², Unab Javed¹, Manindra Rachakonda¹ and Alex Morozov¹

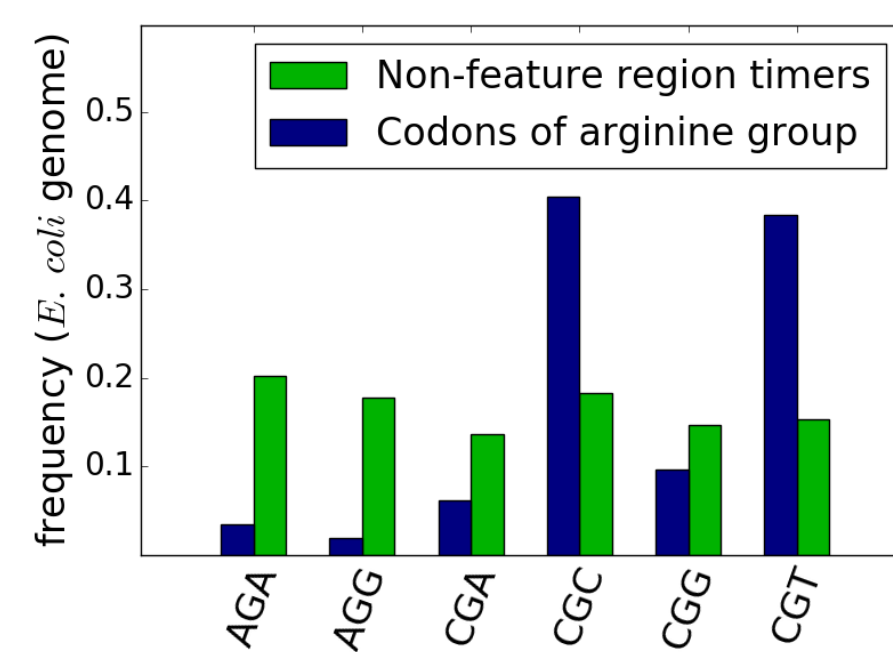
¹Department of Physics & Astronomy, Rutgers, The State University of New Jersey

²Department of Chemistry & Chemical Biology, Harvard University

Introduction

The presence of codon usage bias has been a long-standing mystery [1]. Potential reasons for this bias include,

- Speed of translation.
- Protein synthesis accuracy.



Predictive model

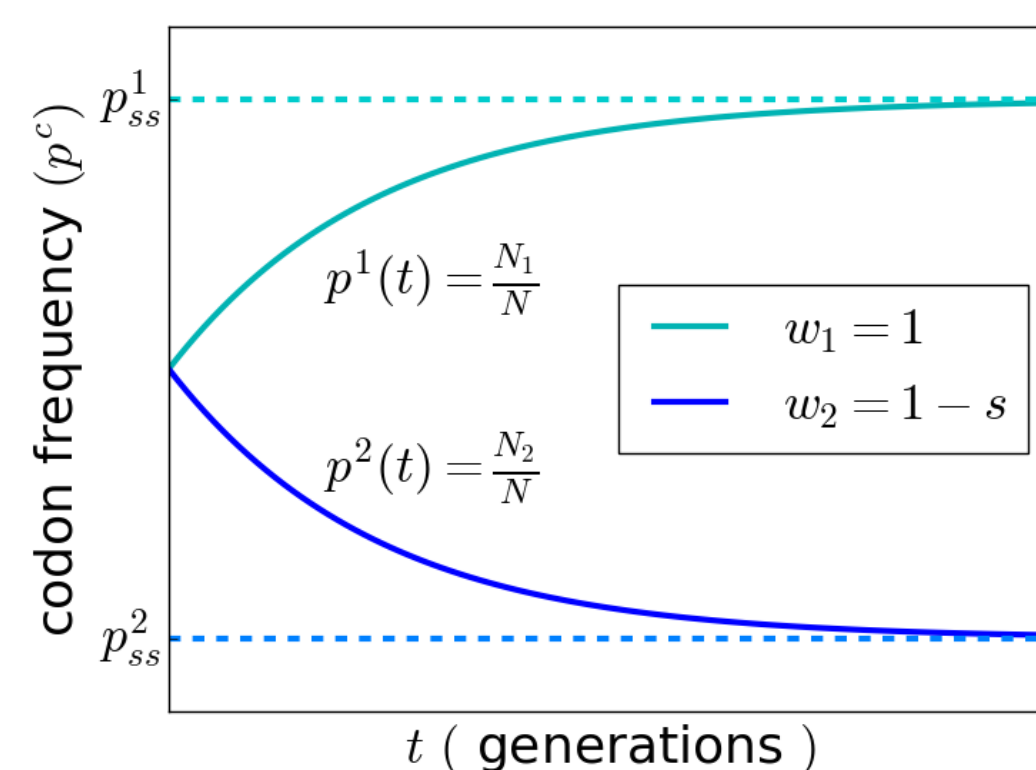
Population genetics model with single-point mutation and absolute fitness. Parameters based on biophysical arguments.

$$|N(t+1)\rangle = (\mathbf{I} + \mathbf{M})\mathbf{W}|N(t)\rangle$$

$$\Rightarrow \langle 1|\mathbf{W}|p_{ss}^c\rangle = (\mathbf{I} + \mathbf{M})\mathbf{W}|p_{ss}^c\rangle.$$

- \mathbf{M} , mutational matrix
- \mathbf{W} , absolute fitness
- $N(t)$, cell number
- $p^c(t)$, codon frequency

After many generations, $t \rightarrow \infty$, frequencies reach steady-state p_{ss}^c .



[1] J. B. Plotkin and G. Kudla. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12(1):32–42, Jan 2011.

[2] M. Bulmer. The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 129(3):897–907, Nov 1991.

[3] Ziheng Yang and Rasmus Nielsen. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 17(1):32, 2000.

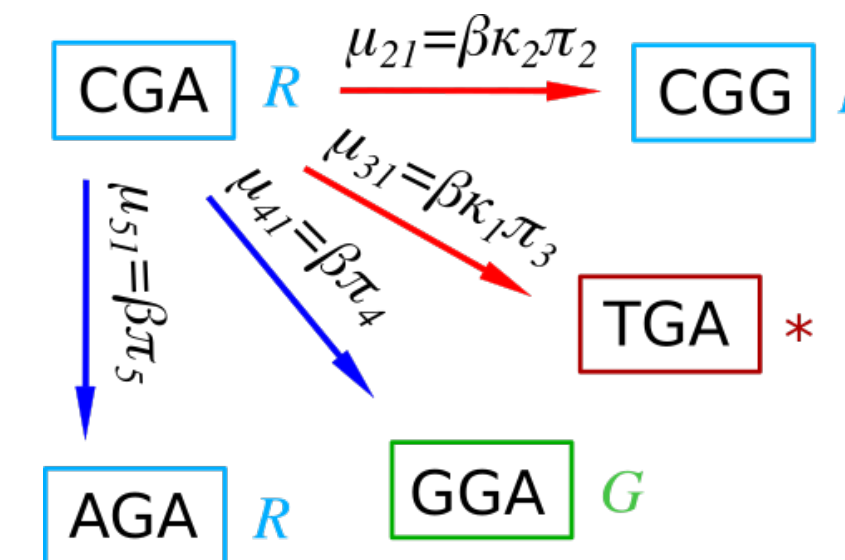
[4] F. V. Murphy and V. Ramakrishnan. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.*, 11(12):1251–1252, Dec 2004.

Inference of mutational rates

Obtained through detailed-balance from trimer frequencies.

$$\pi_1\mu_{21} = \pi_2\mu_{12} \Rightarrow \begin{cases} \mu_{21} = \beta\kappa_2\pi_2 \\ \mu_{12} = \beta\kappa_1\pi_1 \end{cases}$$

Introduces model parameters β , κ_1 and κ_2 corresponding to mutational scale, and transition / transversion rate ratios [3].



Protein production rate

Rate of protein production from two mechanisms with effects on fitness captured by a single parameter, T^0 :

- Single mRNA translation rate affects total protein production due to a finite ribosomal reservoir [2].
- tRNA with high cellular concentrations result in faster translation. [1]

Amino acid fidelity & wobble hypothesis

$1 - s$ penalty for missense mutations and noncognate tRNA binding due to the wobble hypothesis [4].

Wobble rel. rates (U at 5' anticodon)	
U/A	$r_{U/A} = 1$
U/G	$r_{U/G} = r_1$
U/U	$r_{U/U} = r_2$
U/C	$r_{U/C} = r_3$

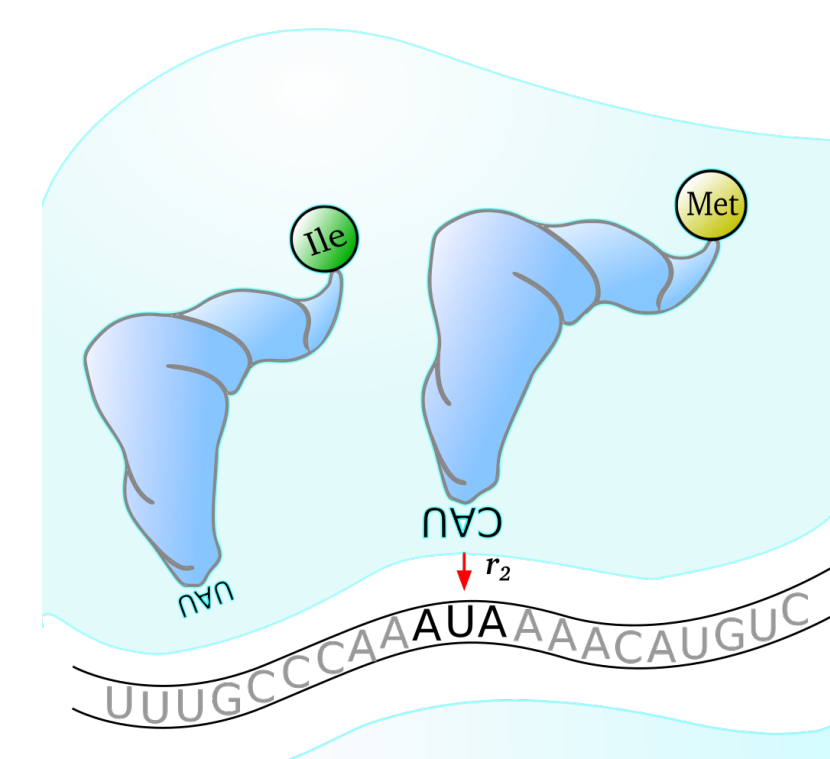


Figure 1: Methionine taking the place of an isoleucine on the protein **thrA** due to wobble.

Results

This model has been fit to a 10% training portion of the genome of *Escherichia coli* K-12 MG1655 and tested on the remaining 90%, and the full *Saccharomyces cerevisiae* S288c genome.

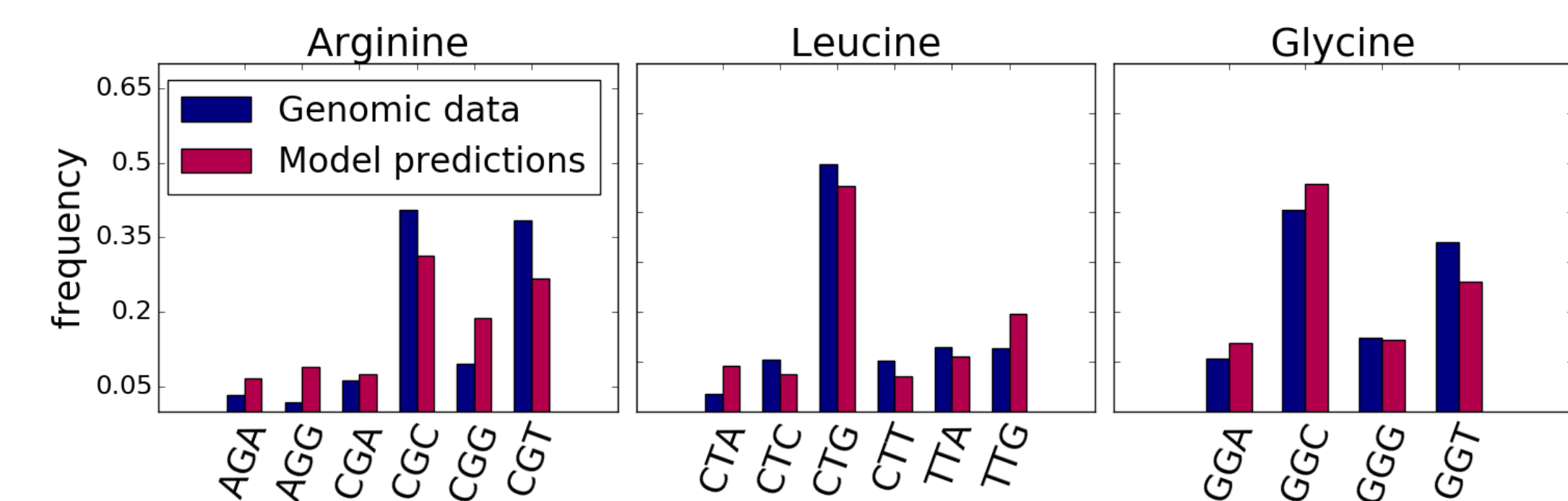


Figure 2: Model prediction of codon frequencies in *E. coli*

Observation: frequencies appear to depend significantly on mutational "closeness" to deleterious codon states (e.g. stop codons.)

Parameter fit values (10% of <i>E. coli</i> genome)	
β	$34.5 \pm 0.2 (\times 10^{-9})$
κ_1	2.24 ± 0.01
κ_2	0.730 ± 0.004
s	$13.2 \pm 0.2 (\times 10^{-9})$
T^0	$7.00 \pm 0.04 (\times 10^{-9})$
r_1	0.715 ± 0.004
r_2	0.153 ± 0.001
r_3	0.242 ± 0.001

