PROTEIN-PROTEIN INTERACTION NETWORKS: CO_EVOLUTION OF DEGREE CONNECTIVITY AND CODON BIAS IN BACTERIALGENOMES

Andrea Giansanti
Dipartimento di Fisica, Sapienza Università di Roma
andrea.giansanti@uniroma1.it

NETWORK DAYS PADOVA, 24-25 Oct. 2024

DIPARTIMENTO DI FISICA

SAPIENZA
UNIVERSITÀ DI ROMA

PROTEIN-PROTEIN INTERACTION NETWORKS AND CODON BIAS: A THEME

Dilucca M, Cimini G, Semmoloni A, Deiana A, Giansanti A. ***Codon Bias Patterns of E. coli's Interacting Proteins***. PLoS One. 2015 Nov 13;10(11):e0142127.
doi: 10.1371/journal.pone.0142127.

Dilucca M, Cimini G, Giansanti A. ***Essentiality, conservation, evolutionary pressure and codon bias in bacterial genomes***. Gene. 2018 Jul 15;663:178-188.
doi: 10.1016/j.gene.2018.04.017.

Dilucca M, Cimini G, Giansanti A. ***Bacterial Protein Interaction Networks: Connectivity is Ruled by Gene Conservation, Essentiality and Function***. Curr Genomics. 2021 Feb;22(2):111-121.
doi: 10.2174/1389202922666210219110831.

Dilucca M, Cimini G, Forcelloni S, Giansanti A. ***Co-evolution between codon usage and protein-protein interaction in bacteria***. Gene. 2021 Apr 30;778:145475.
doi: 10.1016/j.gene.2021.145475.

Codon bias: how the sequence of a gene can tune its translation ”using” a restricted repertoire of synonymous codons

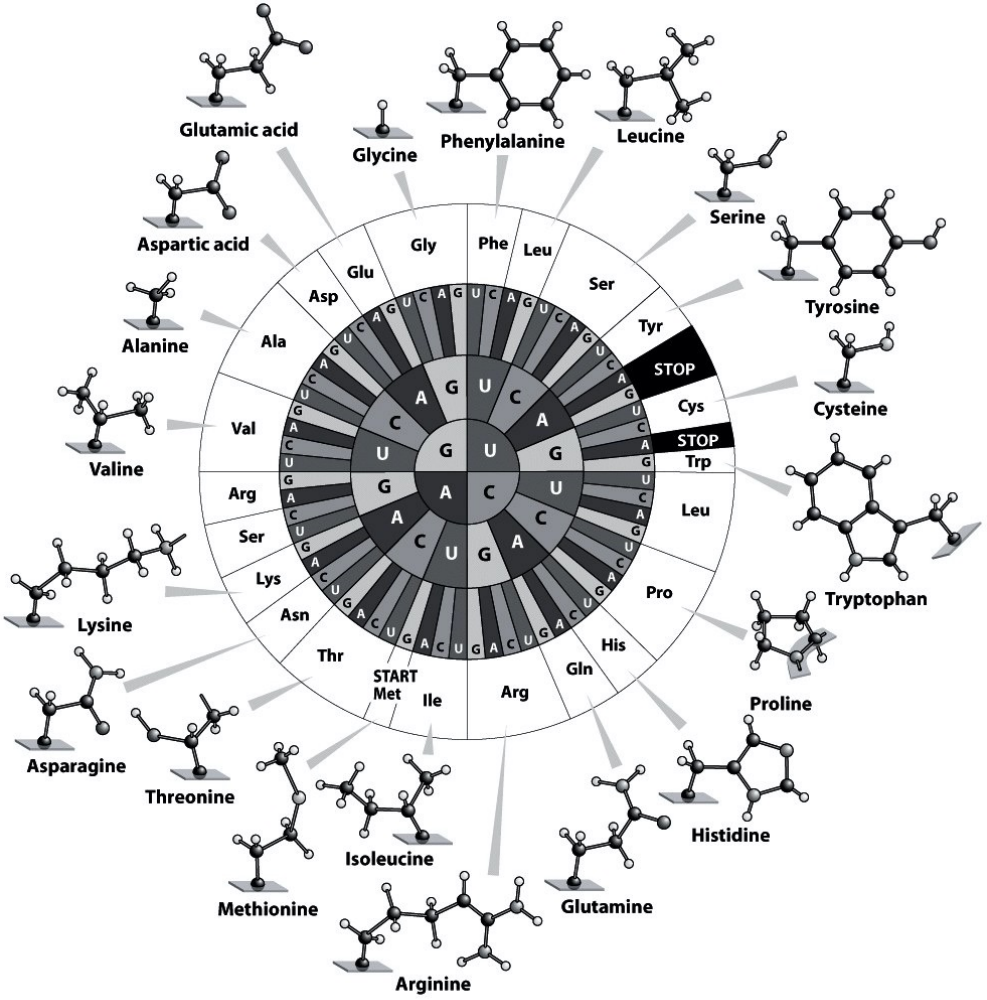


Figure 1.4 Physical Biology of the Cell (© Garland Science 2009)

The genetic code is degenerate

CODON BIAS MEASURES: RSCU

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}}$$

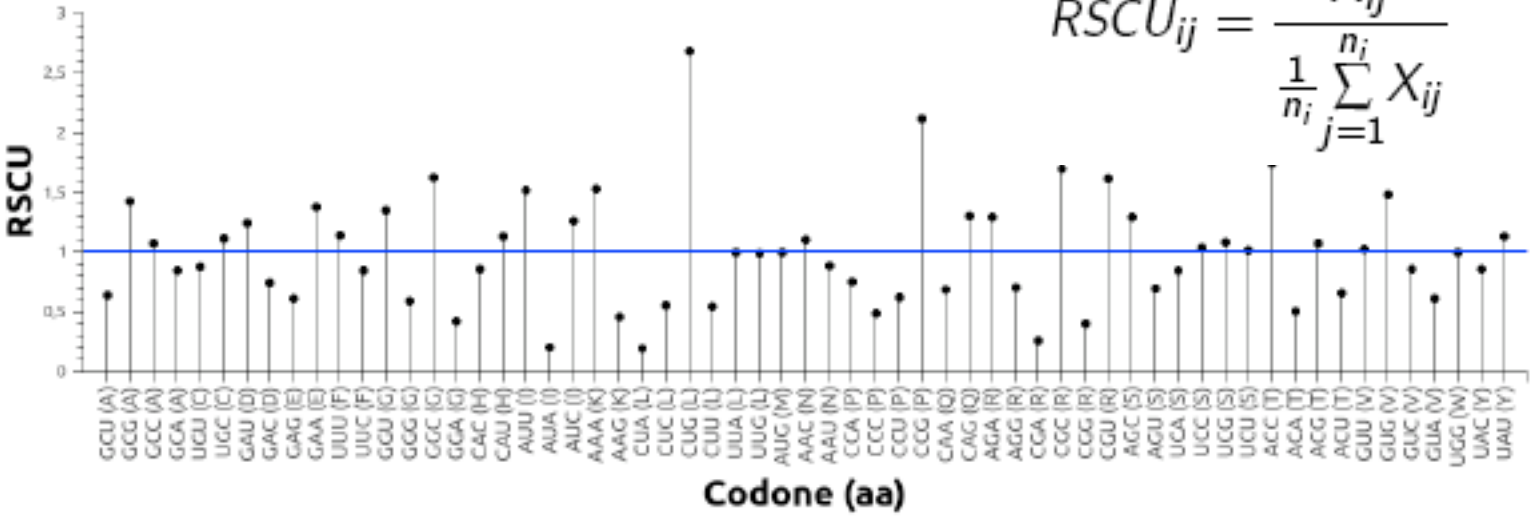
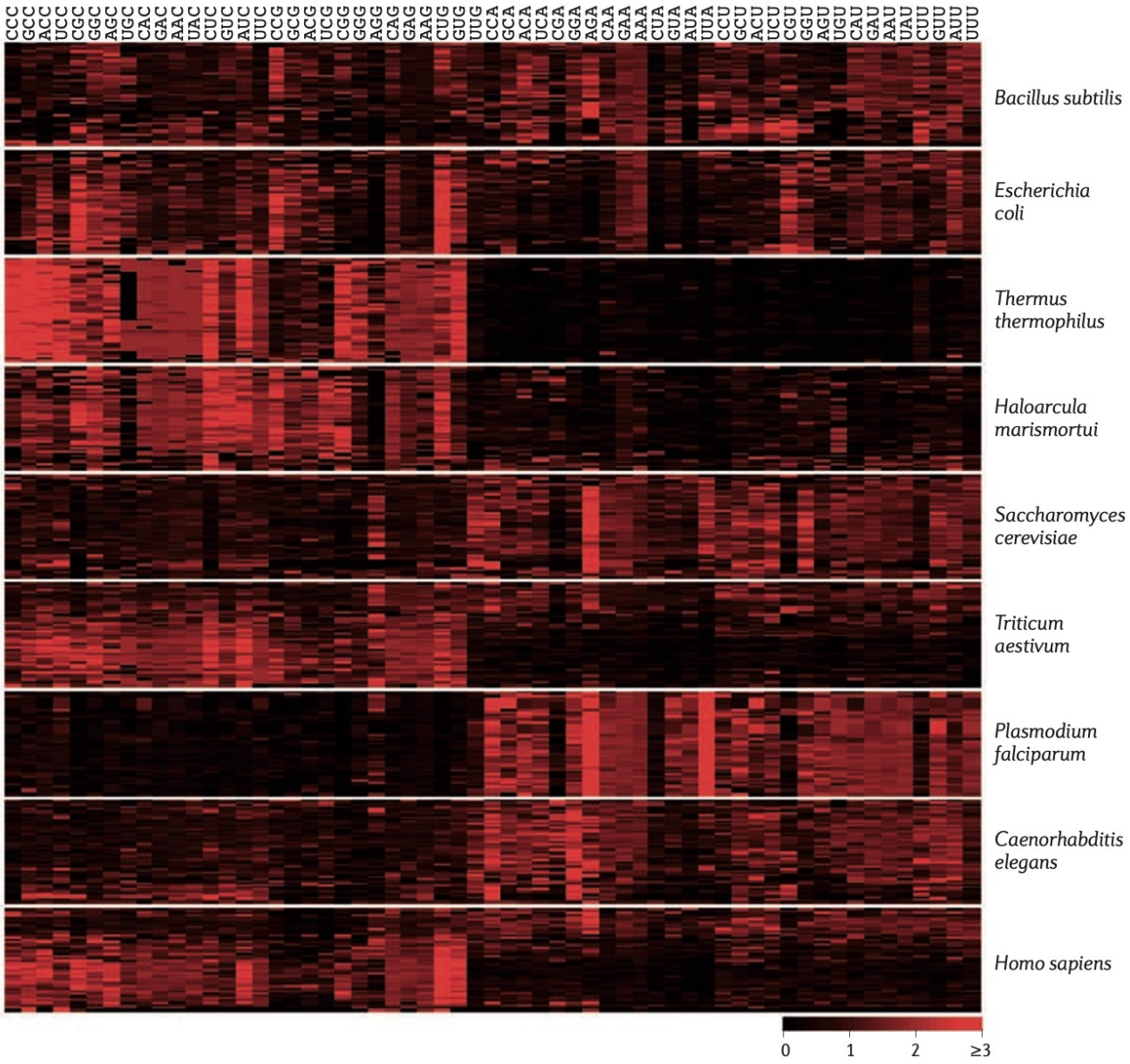


Figure 1 | **Codon bias within and between genomes.** The relative synonymous codon usage (RSCU)[127] is plotted for 50 randomly selected genes from each of nine species. RSCU ranges from 0 (when the codon is absent), through 1 (when there is no bias) to 6 (when a single codon is used in a six-codon family). Methionine, tryptophan and stop codons are omitted. Genes are in rows and codons are in columns, with C- and G-ending codons on the left side of each panel. Note the extensive heterogeneity of codon usage among human genes. Other measures of a gene's codon bias include the codon adaptation index (CAI; the similarity of codon usage to a reference set of highly expressed genes)[35], the frequency of 'optimal' codons (FOP)[28] and the tRNA adaptation index (tAI; the similarity of codon usage to the relative copy numbers of tRNA genes)[128].

CUB whithin and between genomes

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



BASIC BIOLOGY OF CODON BIAS

- Codon usage varies widely between species, between genes in a genome and between sites in a gene.
- Explanations for natural variation in codon usage fall into two categories:
  MUTATIONAL (e.g. viral genes) and
  SELECTIVE (e.g. essential genes in bacteria)

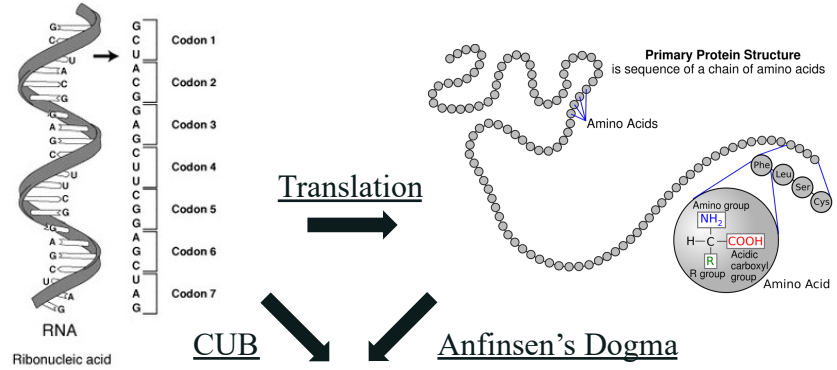
- **Natural selection** → favors optimal (fast) codon
- **Mutational bias** → allows the persistence of non-optimal (slow) codons

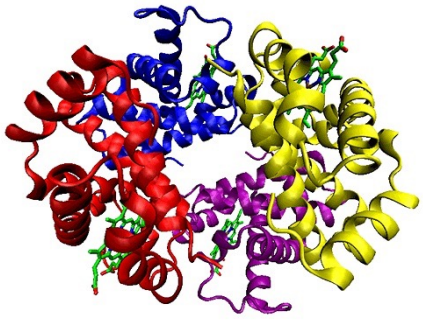
CODON USAGE BIAS MIGHT BE RELATED TO:

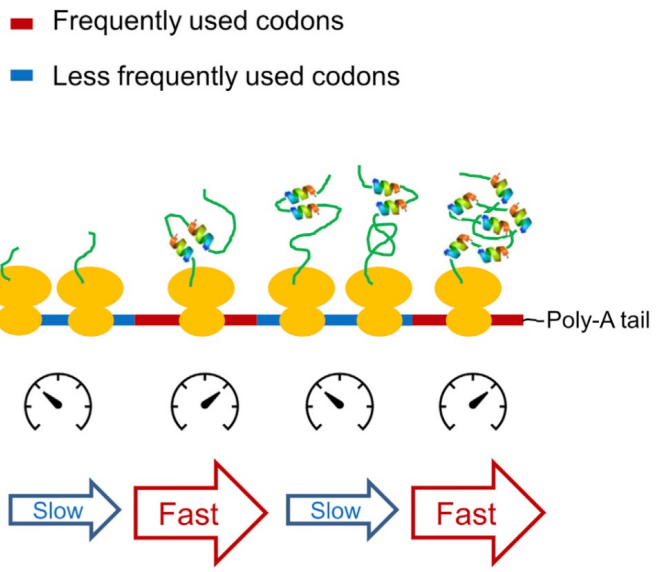
1. selection against translational errors (accuracy)
2. speed of translation elongation (efficiency)
3. effect on protein folding
4. stability of mRNA secondary structures
5. co-adaptation with tRNA abundance
6. control of differential gene expression based on tRNA levels in the cell
7. Environmental (ecological, e.g. microbiomes) adaptation of a species, etc.
8. networks of proteins that fold and assemble co-translationally

CUB and cotranslational folding

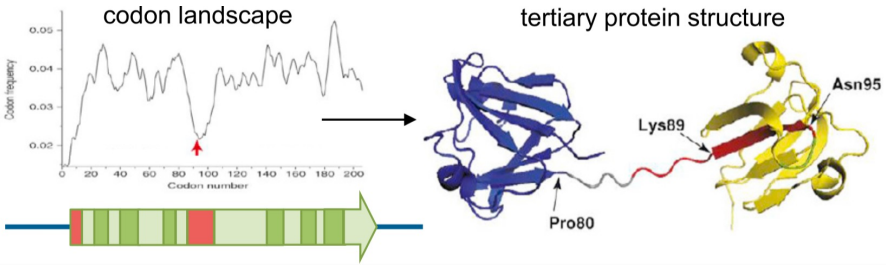
- There is a shared idea that codon usage can modulate translational rates (relevant for biotechnologies!)
- Replacing a preferred codon with a rare codon can cause a slowdown of translation







- Rare codons permit the correct folding towards the native state



CUB AS AN ECOLOGICAL FINGERPRINT OF A SPECIES

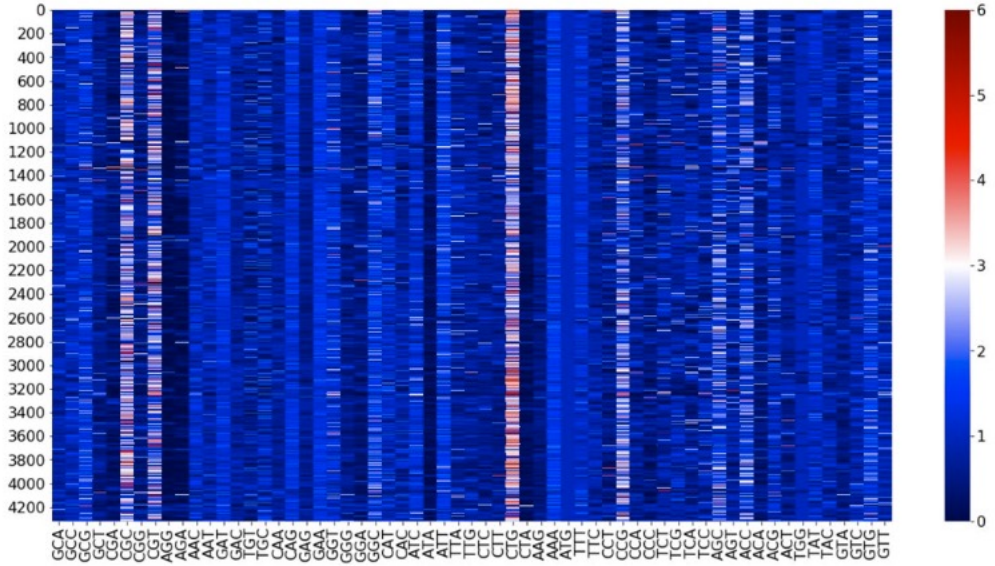


Fig. 1 Heat map of RSCU values for each gene of *E. coli* strain K12 substrain MG1655. The 4319 CDSs are given in the rows and the 61 codons are in columns. Codons are shown in the alphabetical order of the amino acids they code for, i.e. from the four synonymous codons encoding Ala to the four ones encoding Val. We note that RSCU vectors of different genes are very similar to each other

From:Arella D, Dilucca M, Giansanti A, ***Codon usage bias and environmental adaptation in microbial organisms*** **Molecular Genetics and Genomics (2021) 296:751–762**

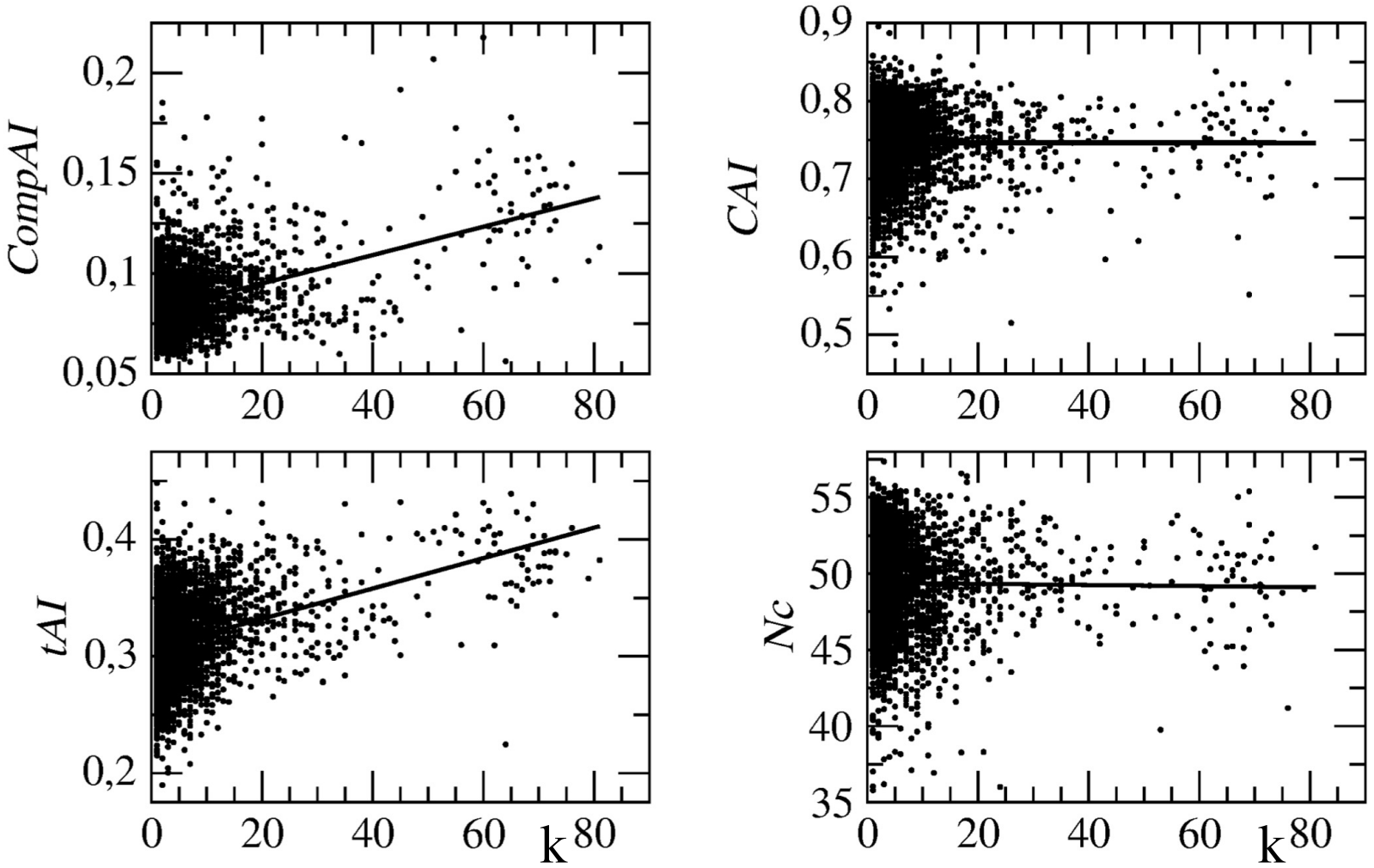


Fig 6. Relation between the various codon bias indices of genes and the degree *k* of the corresponding proteins in the PIN of E.coli. Solid lines are linear fits. *CompAI* and *tAI* of a gene definitely increase with the connectivity of the corresponding protein in the PIN, whereas the other two indices are less sensitive to this parameter.

doi:10.1371/journal.pone.0142127.g006

Current Genomics, 2021, *22*, 111-121 111

RESEARCH ARTICLE

Bacterial Protein Interaction Networks: Connectivity is Ruled by Gene Conservation, Essentiality and Function

Maddalena Dilucca[1,*], Giulio Cimini[2] and Andrea Giansanti[3]

[1]*Dipartimento di Fisica, Sapienza University of Rome, 00185, Rome, Italy;* [2]*Dipartimento di Fisica, Tor Vergata University of Rome, 00133, Rome, Italy Istituto dei Sistemi Complessi CNR UoS, Rome, Italy;* [3]*Dipartimento di Fisica, Sapienza University of Rome, 00185, Rome, Italy INFN Roma1 Unit, Rome, Italy*

42 bacterial species

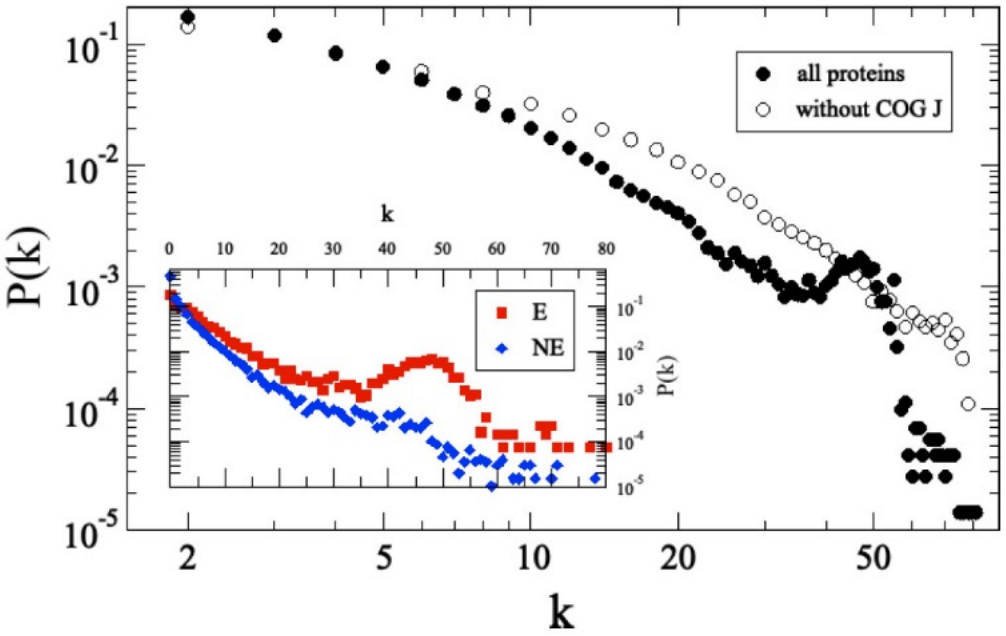


Fig. (1). Probability distribution $P(k)$ for the number of connections k of each protein averaged over the bacterial species considered in Table **1** (full dots), compared with the degree distribution after removal of the proteins corresponding to genes in COG J, related to translational processes (empty dots). Inset: $P(k)$ for essential (E) and nonessential (NE) genes, averaged over DEG-annotated genomes. Note that the average degree is higher for essential genes than for nonessential ones, and the two probability distributions are quite distinct. The region of the curve for low k can be well approximated by a power law [38]. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

COG ONTOLOGY

COG ID	Functional classification
	INFORMATION STORAGE AND PROCESSING
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
	CELLULAR PROCESSES AND SIGNALING
D	Cell cycle control, cell division, chromosome partitioning
T	Signal transduction mechanisms
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, chaperones
	METABOLISM
C	Energy production and conversion
G	Carbohydrate transport and metabolism
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism

Table 2. Functional classification of COG clusters.

Gene Conservation and degree connectivity of the PPI

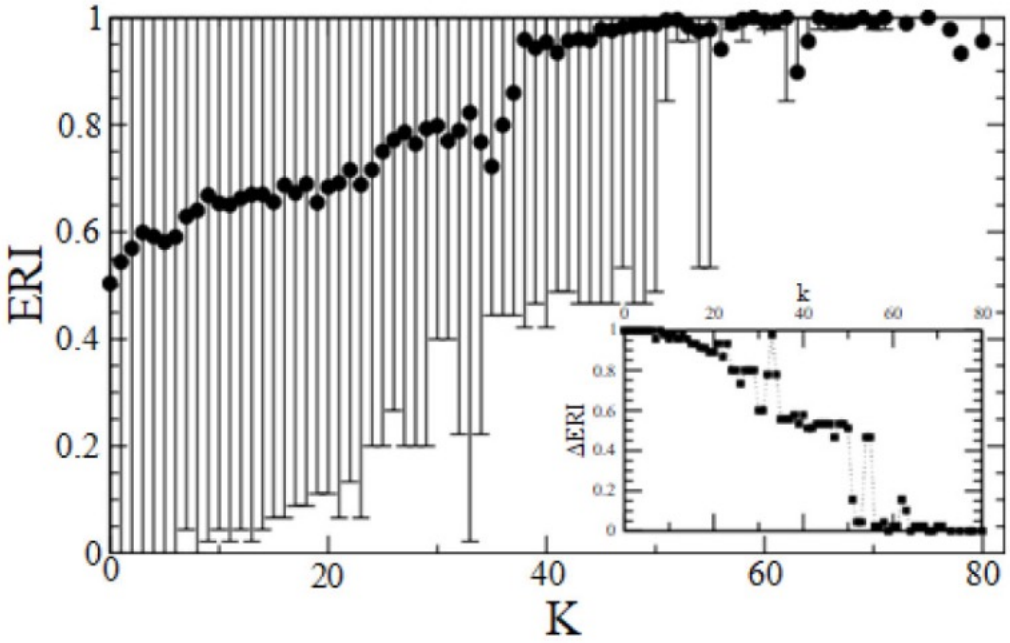


Fig. (2). Average ERI values of bacterial genes as a function of the degrees k of the corresponding proteins, for all the considered genomes. Error bars are standard deviations of ERI values associated to a given k value. Inset: amplitude of the error bar (ΔERI) as a function of k.

“topological” transition in the ontology driven by k

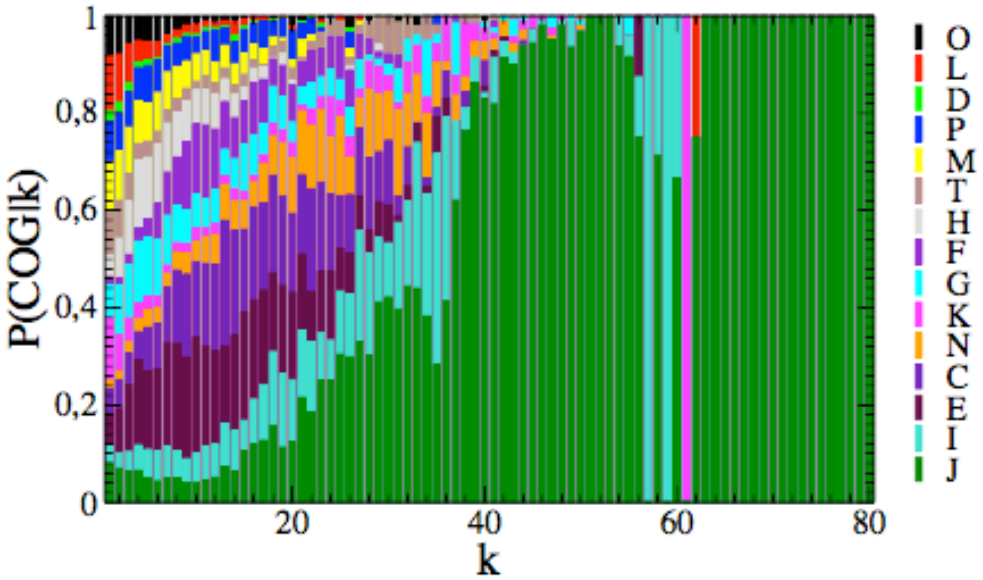


Fig 4. Probability distribution $P(\mathrm{COG}|k)$ of belonging to a given COG for proteins with degree k, over all considered genomes. Proteins with low connectivity have a very heterogeneous COG composition, whereas, those with high k basically belong only to COG J.

Evolutionary pressure on the gene vs connectivity of the proteinin the PPI

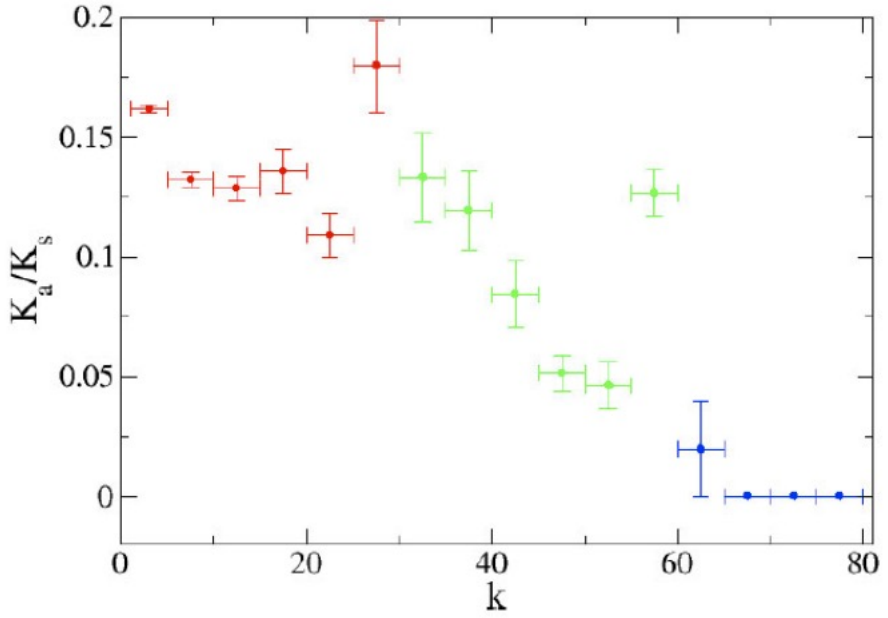


Fig. (3). (K_a/K_a) of groups of genes corresponding to proteins with different connectivity degrees *k*. As in the following Fig. (**4**) and in Fig. **S5**, low connectivities are shown in red, intermediate in green and high connectivities in blue. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

K_a/K_s is the ratio of nonsynonymous substitutions per nonsynonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s) [19]. This parameter is widely accepted as a straightforward and effective way of separating genes subject to purifying evolutionary selection ($K_a/K_s < 1$) from genes subject to positive selective Darwinian evolution ($K_a/K_s > 1$). There are different methods to

TAKE HOME MESSAGE:
CODON BIAS IS FELT BY THE TOPOLOGY OF PPI NETWORKS

LOOKING AHEAD:
-MULTIPLEXES

-ARE THERE NETWORKS OF CO-TRANSLATIONALLY ASSEMBLED PROTEINS ?

- COMMUNITIES OF INTERACTING PROTEINS
THAT DRIVE SPECIATION (FORSDYKE PLOTS)

(see: S. Forcelloni and A. Giansanti. *Mutations in disordered proteins as early indicators of nucleic acid changes triggering speciation.* *Sci Rep* **10**, 4467 (2020). https://doi.org/10.1038/s41598-020-61466-5)

-tRNA BIOLOGY:causal disease networks, precision medicine

Forsdyke Plots

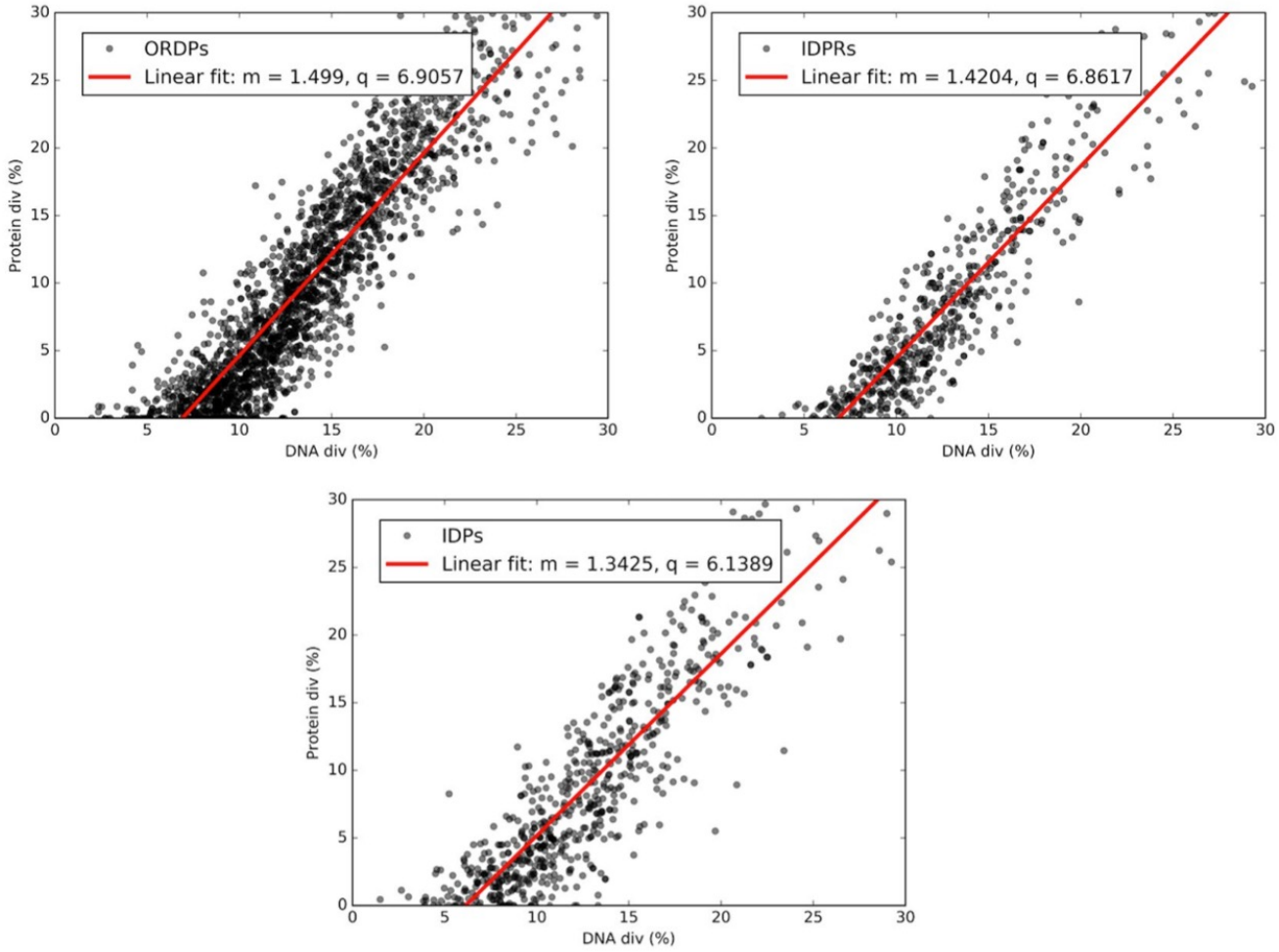


Figure 1. Forsdyke plot. Phenotype (Protein div) vs. nucleotide (DNA div) sequence divergence between human and homologous mouse genes. Each point corresponds to an individual gene. In each panel, we report the best-fit line in red, together with the associated values of the slope (m) and the intercept (q) in the legend.