

# CODON USAGE BIAS IS EVOLUTIONARILY CONSERVED

*Oliver Y. W. Chan<sup>1,#</sup>, Bryan M. H. Keng<sup>1,#</sup>  
and Maurice H. T. Ling<sup>2,3,4,\*</sup>*

<sup>1</sup>Raffles Institution, Singapore

<sup>2</sup>Department of Zoology, The University of Melbourne, Australia

<sup>3</sup>School of Chemical and Biomedical Engineering,  
Nanyang Technological University, Singapore

<sup>4</sup>AdvanceSyn Private Limited, Singapore

## ABSTRACT

Codon usage bias (CUB) reflects the frequency distribution of codons usage in the genome. Several studies suggest that CUB is based on the combinations, which are most chemically efficient and minimise translational error, show that amongst closely related species, CUB is similar. However, previous studies were mainly carried out on a limited number of related species. This study tests the hypothesis that CUB is evolutionarily conserved, and examines CUB over a large set of organisms. Codon usage distributions from 18 organisms across a diversity of classes were examined. The correlations of codon usage frequencies were calculated between and within classes. Our results demonstrated that Pearson's correlation between CUBs of different organisms within the same class is significantly higher than random. The correlation between the CUBs of mammals, birds, insects, yeast, and bacteria also corresponded to evolutionary distance. This suggests that CUB is evolutionarily conserved and the degree of conservation corresponds to evolutionary distance.

**Keywords:** codon usage bias, CUB, evolutionary distance

## INTRODUCTION

The genetic code is the mechanism by which information within genetic material, such as DNA and RNA, is translated into amino acid chains, also known as peptides. Three monomers or nucleotides of DNA or RNA, also known as a codon, represent an amino acid. The genetic code is degenerate as one amino acid may be coded by one of more codons. For

---

<sup>#</sup>Equal contributors.

<sup>\*</sup>Corresponding author: mauriceling@acm.org

example, both CUU and CUG code for the amino acid leucine while AGU and AGC code for serine, the standard translation table. Furthermore, there are also three stop codons UAA, UGA, and UAG, which serve to terminate translation.

Codon Usage Bias (CUB) refers to differences in the frequency of occurrence of synonymous codons in coding DNA (Comeron and Aguadé, 1998). It has been found that different organisms may show particular preferences for one of the several codons that encode a single amino acid (Moriyama and Powell, 1998). Several previous studies have shown that CUB appears to be similar for closely related species (Dass and Sudandiradoss, 2012; Potter et al., 2007). Wang et al. (2013) found that mitochondrial genome of aphids share conserved genomic organization, nucleotide and amino acid composition, and codon usage features. Moura et al. (2011) showed that codon pair usage is conserved in many species while Pal et al. (2011) found that codon usage patterns were uniform across three species of the soil bacterium *Arthrobacter*; namely *A. aurescens* TC1, *A. chlorophenolicus* A6, and *Arthrobacter* sp. FB24.

Several other studies (Iriarte et al., 2013; Rocha, 2004; Higgs and Ran, 2008) suggest that CUB is based on the combinations, which are most chemically efficient and minimise translational error. Porceddu et al. (2013) has shown that in plants, preferred codons are selected to maximize translational accuracy, while Thiele et al. (2012) found that codon usage evolves to maximize growth and adapt to new environmental niches in *Escherichia coli*.

It has also been shown that even amongst species with large evolutionary divergence, there are similarities in equivalent non-coding regions. Siepel et al. (2005) performed compared equivalent genomic regions from different organisms to identify both similarities and differences, and found that there were nearly 1400 highly conserved non-coding sequences between humans and the pufferfish, *Fugu rubripes*. This suggests that these sequences probably serve an important function in all vertebrates, such as developmental regulators. This has further implications for genetic engineering (Gustafsson et al., 2004), where it would be necessary to optimise the proportion of different synonymous codons to be as similar as possible to the host cell, such that the host cell would successfully integrate the foreign DNA into its genome and translate it into protein.

CUB has been suggested as a factor in viral virulence, determining the success of viral exploitation of the host (Brandão, 2013). Previous studies (Zhou et al., 2012; Wong et al., 2010) have found that codon usage amongst viruses matches their target cell's codon usage to a large extent. This suggests that when the proportion of synonymous codons used by a virus is different from the amounts normally present in the host cell, cellular machinery may be able to detect the abnormalities and stop translation of viral genetic material (Coleman et al., 2008).

Taken together, these suggest that various organisms utilise CUB to optimise translational efficiency while maintaining efficient use of resources. In addition, organisms like bacteria are also able to detect when the codons for foreign genetic material do not correspond to their own codon usage biases. However, Yang and Nielsen (2008) found that natural selection may be a driving force in the evolution of synonymous codon usage but their estimation of selection coefficients suggested that CUB is weak and shows mostly neutral mutational bias.

CUB may be important in examining cross species viral infections (Wong et al., 2010). For example, the Spanish flu in 1918 was transmitted from birds to cats to humans (Webster

et al., 1992). This suggests that CUB may have an evolutionary bias. For instance, the CUB of mammalian species would be closer as compared to non-mammalians.

In this study, we analysed codon usage in 18 species across different families to determine their evolutionary relationships. We hypothesize that CUB is evolutionarily conserved. If our results show correlation between classes of organisms and that CUB is correlated to evolutionary time, it will suggest that CUB is evolutionarily conserved.

## MATERIALS AND METHODS

Eighteen sets of publically available codon usage tables from Codon Usage Database ([www.kazusa.or.jp/codon](http://www.kazusa.or.jp/codon)) were selected for use in this study (Table 1). Codon usage tables were generated from protein coding gene sequences released in GenBank version 160. Species were chosen to represent a diversity of classes and datasets with large numbers of codons were used.

**Table 1. Species used**

Species/strain	Common Name	Class	No. of CDS	No. of Codons
<i>Mus musculus</i>	Mouse	Mammal	53036	24533776
<i>Rattus norvegicus</i>	Rat	Mammal	15151	6981149
<i>Homo sapiens</i>	Human	Mammal	93487	40662582
<i>Canis lupus familiaris</i>	Dog	Mammal	1194	559501
<i>Bos taurus</i>	Cow	Mammal	13374	5198458
<i>Gallus gallus</i>	Chicken	Bird	6017	2719057
<i>Anas platyrhynchos</i>	Duck	Bird	151	45582
<i>Drosophila melanogaster</i>	Fruit fly	Insect	42417	21945319
<i>Drosophila simulans</i>	-	Insect	1157	372744
<i>Aedes aegypti</i>	Mosquito	Insect	585	257935
<i>Saccharomyces cerevisiae</i>	Baker's yeast	Yeast	14411	6534504
<i>Schizosaccharomyces pombe</i>	Fission yeast	Yeast	6109	2859399
<i>E. coli</i>	-	Bacteria	8087	2330943
<i>E. coli</i> APEC O1	-	Bacteria	4890	1598551
<i>E. coli</i> UTI89	-	Bacteria	5211	1570760
<i>E. coli</i> 536	-	Bacteria	4629	1445921
<i>E. coli</i> CFT073	-	Bacteria	5379	199310
<i>E. coli</i> O157:H7 EDL933	-	Bacteria	5347	155864

The codon usages for each species, given in percentage, were correlated pairwise, producing 171 correlations. The correlations of codon usage frequencies were calculated between classes and within classes. If codon usages are evolutionarily conserved, we will expect that the correlations between organisms within the same class to be significantly higher than random (no correlation), and higher correlations within classes than between classes.

## RESULTS AND DISCUSSION

It has been shown that conserved nucleic acid sequences and protein sequences may indicate evolutionary relationships between species (Aravind et al., 1999; Vogt, 1990). A particular sequence that has been maintained by evolution despite speciation further implies that a conserved gene is present. This is because such highly conserved sequences normally have an important functional value (Sbisà et al., 1997). Due to the conservation of DNA sequences for protein coding genes (Aravind et al., 1999), it follows that antisense transcripts are also conserved. In addition, recent studies had demonstrated that the expression of protein coding genes (Jordan et al., 2005) and its antisense counterparts are evolutionarily conserved (Ling et al., 2013). Hence, it is highly plausible that codon usages are evolutionarily conserved.

### CUBs Are Significantly Correlated within the Same Class

Our results demonstrated that Pearson's correlation between CUBs of different organisms within the same class ranged from 0.855 to more than 0.999 (Table 2).

**Table 2. Correlations between codon usages of different organisms within the same class**

Comparison	Pearson's Correlation	P-value
Bacteria: <i>E. coli</i> / <i>E. coli</i> 536	0.906	5.72E-13
Bacteria: <i>E. coli</i> / <i>E. coli</i> APEC O1	0.915	9.47E-13
Bacteria: <i>E. coli</i> / <i>E. coli</i> CFT073	0.911	9.98E-14
Bacteria: <i>E. coli</i> / <i>E. coli</i> O157:H7	0.914	3.01E-12
Bacteria: <i>E. coli</i> / <i>E. coli</i> UTI89	0.909	1.09E-13
Bacteria: <i>E. coli</i> 536 / <i>E. coli</i> APEC O1	0.999	3.23E-15
Bacteria: <i>E. coli</i> 536 / <i>E. coli</i> CFT073	>0.999	<7.51E-16
Bacteria: <i>E. coli</i> 536 / <i>E. coli</i> O157:H7	0.999	4.92E-16
Bacteria: <i>E. coli</i> 536 / <i>E. coli</i> UTI89	>0.999	<1.75E-15
Bacteria: <i>E. coli</i> APEC O1 / <i>E. coli</i> CFT073	>0.999	<3.01E-15
Bacteria: <i>E. coli</i> APEC O1 / <i>E. coli</i> O157:H7	>0.999	<1.12E-14
Bacteria: <i>E. coli</i> APEC O1 / <i>E. coli</i> UTI89	>0.999	<8.11E-15
Bacteria: <i>E. coli</i> CFT073 / <i>E. coli</i> O157:H7	0.999	3.09E-14
Bacteria: <i>E. coli</i> CFT073 / <i>E. coli</i> UTI89	>0.999	<1.24E-14
Bacteria: <i>E. coli</i> O157:H7 / <i>E. coli</i> UTI89	0.999	9.84E-15
Bird: <i>A. platyrhynchos</i> / <i>G. gallus</i>	0.978	1.01E-13
Insect: <i>A. aegypti</i> / <i>D. melanogaster</i>	0.859	3.10E-11
Insect: <i>A. aegypti</i> / <i>D. simulans</i>	0.855	5.87E-12
Insect: <i>D. melanogaster</i> / <i>D. simulans</i>	0.975	7.48E-14
Mammal: <i>B. taurus</i> / <i>C. lupus familiaris</i>	0.995	3.77E-15
Mammal: <i>B. taurus</i> / <i>H. sapiens</i>	0.988	6.27E-15
Mammal: <i>B. taurus</i> / <i>M. Musculus</i>	0.985	1.22E-14
Mammal: <i>B. taurus</i> / <i>R. Norvegicus</i>	0.993	5.24E-14
Mammal: <i>C. lupus familiaris</i> / <i>H. sapiens</i>	0.984	2.37E-14
Mammal: <i>C. lupus familiaris</i> / <i>M. musculus</i>	0.984	6.15E-15
Mammal: <i>C. lupus familiaris</i> / <i>R. Norvegicus</i>	0.992	9.30E-15
Mammal: <i>H. sapiens</i> / <i>M. musculus</i>	0.994	9.19E-16
Mammal: <i>H. sapiens</i> / <i>R. norvegicus</i>	0.990	6.73E-15
Mammal: <i>M. musculus</i> / <i>R. norvegicus</i>	0.996	1.19E-15

These correlations are significantly higher than random ( $p\text{-value} < 0.001$ ). This suggests that CUB is an indicator of evolutionary distance. This is consistent with Aravind et al. (1999) whom demonstrated sequence conservation in protein coding genes.

It is inferable that CUB is evolutionarily conserved if protein coding gene sequences are evolutionarily conserved as CUB can be seen as a feature of the sequence. Jordan et al. (2005) had also demonstrated that the expressions of protein coding genes are evolutionarily conserved. Hence, it is plausible to expect a correlation between CUB and the expression of protein coding genes, which had been demonstrated by dos Reis et al. (2003) and Piovesan et al. (2013). This principle had been successfully used in several recent studies to predict expressions of protein coding genes from CUB in *E. coli* (Roymondal et al., 2009), yeast (Das et al., 2009), and cyanobacterium (Das et al., 2012).

### **CUBs Corroborate Evolutionary Time**

An analysis of codon usage datasets shows that all three species' CUB are highly correlated (rat-mouse  $r = 0.996$ ; rat-human  $r = 0.990$ ; mouse-human  $r = 0.994$ ), suggesting a short evolutionary distance. Bejerano et al. (2004) identified ultra-conserved sequences that share complete identity among mouse, rat and human, and appear to be essential for maintaining the ontogeny of mammals. In addition, the correlation of codon usage between rats and mice is higher, indicating a more recent common ancestor (Bourque et al., 2004). A study by Jordan et al. (2005) on the expression of protein coding genes in human, rat, and mouse suggests that the expressions of orthologous genes are evolutionarily conserved.

Further work by Ling et al. (2013) showed that orthologous antisense transcripts are evolutionarily conserved in human, rat, and mouse. Moreover, both groups (Jordan et al., 2005; Ling et al., 2013) suggested that the degree of expression correlations is congruent with evolutionary time as the expressional correlation between human and mouse is lower than that of rat and mouse. Our CUB analysis shows that the correlation between mouse and rat is higher than both human and rat, and human and mouse, which corroborates previous studies (Jordan et al., 2005; Ling et al., 2013).

Analysis of the datasets also revealed that the correlation between the CUBs of mammals, birds, insects, yeast and bacteria also corresponded to evolutionary distance (Figure 1). This is because classes such as animals and birds have a more recent common ancestor, which existed about 65 million years ago (Hedges et al., 1996). The average correlation between mammals and birds was found to be 0.982, while the average correlation between the eukaryotes and bacteria was 0.680. In addition, the correlations between the various groupings were all significant, with  $p\text{-values} < 0.001$ .

Using taxonomic lineage provided by NCBI ([www.ncbi.nlm.nih.gov/taxonomy](http://www.ncbi.nlm.nih.gov/taxonomy)), mammals and birds are amniotic tetrapods, which are classified as vertebrates under the phylum of Chordata, which emerged 525 million years ago (Chen et al., 1999). Insects are classified as invertebrates under the phylum of Arthropoda and share common lineage with the phylum of Chordata as bilateral organisms, under the kingdom of Metazoa, which emerged 635 million years ago (Lake, 1990).

Yeast is classified under the kingdom of Fungi and share common lineage with the kingdom of Metazoa under the superkingdom of Eukaryota, which emerged 2.1 billion years

ago (Philippe et al., 1998). Bacterial organisms are classified under the superkingdom of Eubacteria, which is distinct from the superkingdom of Eukaryota.

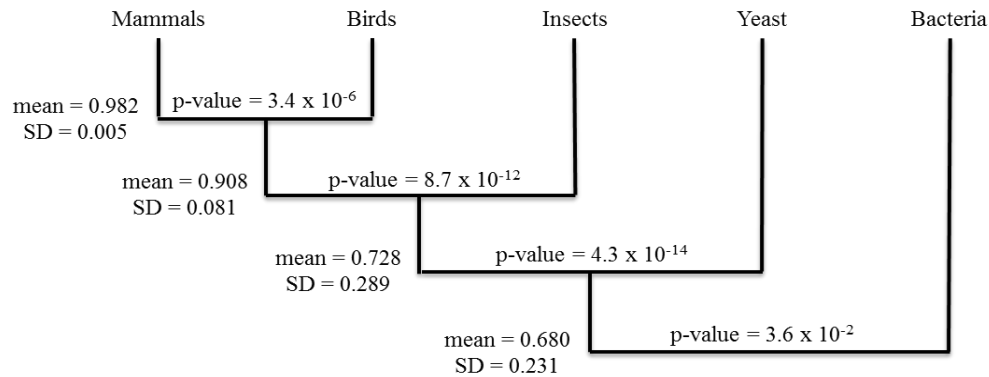


Figure 1. Correlation of codon usages across classes.

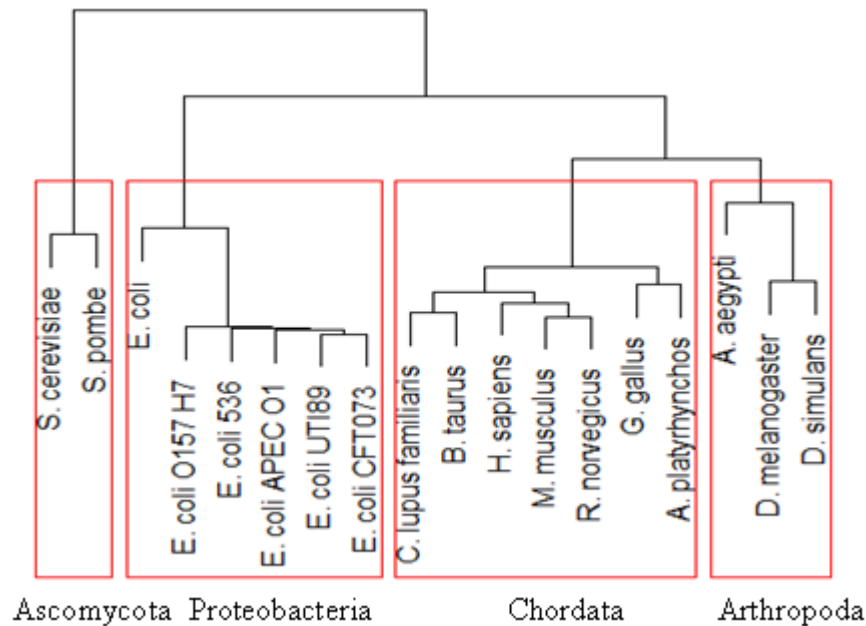


Figure 2. Dendrogram of organisms by codon usages.

A dendrogram (Figure 2) of codon usages shows statistical similarities within phyla of Ascomycota, Proteobacteria, Chordata, and Arthropoda. Within the phylum of Arthropoda, *D. melanogaster* and *D. simulans* have more similar CUBs as compared to *A. aegypti*, due to the fact that these species have the same genus *Drosophila* and are more closely related. Under the phylum of Chordata, birds and mammals have less similar CUBs than CUBs between various species of mammals. Under the phylum of Proteobacteria, the various strains of *E. coli* show much greater CUB similarity as compared to the reference strain. This may be due to the reference strain being representative of all *E. coli* strains; therefore, exhibiting an additive effect of the CUB differences between each strain. However, the dendrogram also

shows that Chordata share greater CUB similarities with Proteobacteria than Ascomycota, despite Chordata and Proteobacteria having a less recent evolutionary ancestor. Several studies had demonstrated that symbiotic prokaryotes evolved similar CUB with their hosts, such as *Verminephrobacter* and earthworms (Kjeldsen et al., 2012), and *Buchnera* and aphids (Moya et al., 2002). As *E. coli* is a symbiotic organism within gastrointestinal tract of mammals (PMID Kaper et al., 2004), it is plausible to expect a degree of CUB similarity between *E. coli* and mammals.

Codon usage bias is used to maximize translational accuracy and efficiency. It has been found that different anticodons that code for the same amino acid are kept to a minimum in rapidly multiplying bacteria species (Rocha, 2004; Ran and Higgs, 2013). This suggests an optimization of the translation machinery to use a small subset of optimal codons and anticodons in fast-growing bacteria and in highly expressed genes.

## CONCLUSION

Our results show correlation between classes of organisms and that CUB is correlated to evolutionary time, suggesting that CUB is evolutionarily conserved.

## REFERENCES

- Aravind, L., Walker, D. R., & Koonin, E. V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic acids research*, 27(5), 1223-1242.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, 304(5675), 132-1325.
- Bourque, G., Pevzner, P. A., & Tesler, G. (2004). Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome research*, 14(4), 507-516.
- Brandão, P. E. (2013). The evolution of codon usage in structural and non-structural viral genes: The case of Avian coronavirus and its natural host *Gallus gallus*. *Virus research*, 178(2), 264-271.
- Chamary, J. V., & Hurst, L. D. (2004). Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage. *Molecular biology and evolution*, 21(6), 1014-1023.
- Chen, J. Y., Huang, D. Y., & Li, C. W. (1999). An early Cambrian craniate-like chordate. *Nature*, 402(6761), 518-522.
- Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science*, 320(5884), 1784-1787.
- Comeron, J. M., & Aguadé, M. (1998). An evaluation of measures of synonymous codon usage bias. *Journal of molecular evolution*, 47(3), 268-274.
- Das, S., Roymondal, U., & Sahoo, S. (2009). Analyzing gene expression from relative codon usage bias in Yeast genome: A statistical significance and biological relevance. *Gene*, 443(1), 121-131.

- Das, S., Roymondal, U., Chottopadhyay, B., & Sahoo, S. (2012). Gene expression profile of the cyanobacterium *Synechocystis* genome. *Gene*, 497(2), 344-352.
- Dass, J., & Sudandiradoss, C. (2012). Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species. *Gene*, 503(1), 92-100.
- De Curtis, I. (2011). Host-Pathogen Interactions: Cheating the Host by Making New Connections. *Current Biology*, 21(5), R192-R194.
- dos Reis, M., Wernisch, L., & Savva, R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic acids research*, 31(23), 6976-6985.
- Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in biotechnology*, 22(7), 346-353.
- Hedges, S. B., Parker, P. H., Sibley, C. G., & Kumar, S. (1996). Continental breakup and the ordinal diversification of birds and mammals. *Nature*, 381(6579), 226-229.
- Higgs, P. G., & Ran, W. (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Molecular biology and evolution*, 25(11), 2279-2291.
- Iriarte, A., Baraibar, J. D., Romero, H., Castro-Sowinski, S., & Musto, H. (2013). Evolution of optimal codon choices in the family Enterobacteriaceae. *Microbiology*, 159(Pt 3), 555-564.
- Jordan, I.K., Marino-Ramirez, L., & Koonin, E.V. (2005). Evolutionary significance of gene expression divergence. *Gene*, 345:119-126.
- Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, 2(2), 123-140.
- Kjeldsen, K. U., Bataillon, T., Pinel, N., De Mita, S., Lund, M. B., Panitz, F., ... & Schramm, A. (2012). Purifying selection and molecular adaptation in the genome of *Verminephrobacter*, the heritable symbiotic bacteria of earthworms. *Genome biology and evolution*, 4(3), 307-315.
- Lake, J. A. (1990). Origin of the Metazoa. *Proceedings of the National Academy of Sciences*, 87(2), 763-766.
- Ling, M.H.T., Ban, Y., Wen, H., Wang, SM, Ge, SX. (2013). Conserved Expression of Natural Antisense Transcripts in Mammals. *BMC Genomics*, 14(1): 243.
- Moriyama, E.N., & Powell, J.R. (1998). Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic acids research*, 26(13), 3188-3193.
- Moura, G.R., Pinheiro, M., Freitas, A., Oliveira, J. L., Frommlet, J. C., Carreto, L., & Santos, M.A. (2011). Species-specific codon context rules unveil non-neutrality effects of synonymous mutations. *PloS one*, 6(10), e26817.
- Moya, A., Latorre, A., Sabater-Muñoz, B., & Silva, F. J. (2002). Comparative molecular evolution of primary (Buchnera) and secondary symbionts of aphids based on two protein-coding genes. *Journal of molecular evolution*, 55(2), 127-137.
- Pal, A., Mondal, U.K., Mukhopadhyay, S., & Bothra, A.K. (2011). Genomic heterogeneity within conserved metabolic pathways of *Arthrobacter* species - a bioinformatic approach. *Bioinformation*, 5(10), 446.



- Philippe, H., Adoutte, A., Coombs, G. H., Vickerman, K., Sleigh, M. A., & Warren, A. (1998). The molecular phylogeny of Eukaryota: solid facts and uncertainties. Evolutionary relationships among protozoa, 25-56.
- Piovesan, A., Vitale, L., Pelleri, M.C., & Strippoli, P. (2013). Universal tight correlation of codon bias and pool of RNA codons (codonome): The genome is optimized to allow any distribution of gene expression values in the transcriptome from bacteria to humans. *Genomics*, 101(5), 282-289.
- Porceddu, A., Zenoni, S., & Camiolo, S. (2013). The signatures of selection for translational accuracy in plant genes. *Genome biology and evolution*, 5(6), 1117-1126.
- Potter, S., Orgeig, S., Donnellan, S., & Daniels, C.B. (2007). Purifying selection drives the evolution of surfactant protein C (SP-C) independently of body temperature regulation in mammals. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 2(2), 165-176.
- Ran, W., & Higgs, P.G. (2012). Contributions of speed and accuracy to translational selection in bacteria. *PloS one*, 7(12), e51652.
- Rocha, E.P. (2004). Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome research*, 14(11), 2279-2286.
- Roymondal, U., Das, S., & Sahoo, S. (2009). Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome. *DNA research*, 16(1), 13-30.
- Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., & Saccone, C. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, 205(1), 125-140.
- Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson H., Spieth J., Hillier L.W., Richards S., Weinstock G.M., Wilson R.K., Gibbs R.A., Kent W. J., Miller W. & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research*, 15(8), 1034-1050.
- Thiele, I., Fleming, R.M., Que, R., Bordbar, A., Diep, D., & Palsson, B.O. (2012). Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. *PloS one*, 7(9), e45635.
- Vogt, P. (1990). Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code". *Human genetics*, 84(4), 301-336.
- Wang, Y., Huang, X.L., & Qiao, G.X. (2013). Comparative Analysis of Mitochondrial Genomes of Five Aphid Species (Hemiptera: Aphididae) and Phylogenetic Implications. *PloS one*, 8(10), e77511.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., & Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiological reviews*, 56(1), 152-179.
- Wong, E.H., Smith, D.K., Rabadan, R., Peiris, M., & Poon, L.L. (2010). Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC evolutionary biology*, 10(1), 253.
- Yang, Z., & Nielsen, R. (2008). Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Molecular biology and evolution*, 25(3), 568-579.
- Zhou, Y., Chen, X., Ushijima, H., & Frey, T.K. (2012). Analysis of base and codon usage by rubella virus. *Archives of virology*, 157(5), 889-899.