Review

Integrating Molecular Evolution and Computational Biology: Bridging Disciplines for Future Research

Pratyush Kumar Beura [1,]*

[1] Department of Molecular Biology and Biotechnology, Tezpur University
* Correspondence: kumar.pratyush09@gmail.com

Scopus Author ID 57476210500

Received: 13 March 2025; Accepted: 11 July 2025; Published: 31 July 2025

Abstract: Molecular evolution, conventionally rooted in classical evolutionary theory and comparative biology, has entered a transformative era driven by advances in genomics, bioinformatics, and computational modeling. This review traces the conceptual foundations of molecular evolution, beginning with the central dogma and codon degeneracy, and explores how variations such as single nucleotide variants (SNVs) shape protein structure and function. It highlights the evolutionary implications of codon usage bias, substitution models, and the mutation and selection balance in across genomes. Recent advances in artificial intelligence (AI), machine learning, biostatistics, and mathematical modeling have revolutionized our understanding of molecular evolution. AI-driven approaches and mathematical algorithms enhance analyses of genetic variation, protein evolution, and evolutionary dynamics. Updated computational platforms such as IQ-TREE 2, RAxML-NG, BEAST 2, PAML, and HyPhy, along with R and Python-based pipelines, have revolutionized evolutionary studies by enabling accurate modeling of mutation dynamics, phylogenetic reconstructions, and selection analyses.Additionally, the chemistry of amino acid exchangeability introduces new perspectives in evolutionary studies. This convergence of computational biology with mathematics, chemistry, and data science has transformed evolutionary biology into a multidisciplinary and collaborative research area to solve long standing biological queries. This opens up opportunities for a successful career in multidisciplinary research in evolutionary biology.

Keywords: Molecular evolution; Single Nuelcotide Variations; Selection; Codon Usage Bias; Computational Biology

1. Introduction

The central dogma of molecular biology itself integrates biology of gene expression, likelihood of mutation dynamics, and the chemistry of protein folding. With only a few exceptions found in extranuclear DNA, nearly all life forms adhere to the standard genetic code [1] (Figure 1). Gene expression begins with a non-reactive DNA and results into the formation of a functional polypeptide, outlining the fundamental process described by the central dogma of molecular biology [2-4]. Proteins are synthesized from DNA via the genetic code, making it one of the cell's most complex molecular processes. This fundamental process of gene-to-protein translation laid the molecular foundation for understanding how genetic variation drives evolutionary change. In the mid-1800s and early 1900s, as biologists sought to understand the factors influencing phenotypes, Darwin's theory of evolution by natural selection became a foundation of evolutionary biology [5,6]. Ronald Fisher later integrated Mendelian inheritance with natural selection, advancing Neo-Darwinism [7-10]. By the 1990s, evolutionary biologists and population geneticists explored allele frequencies in populations [11,12] speciation [13,14] micro and macroevolution [15,16] fossil evidence of homologous and analogous structures [17,18] and genetic drift [19]. The discovery of DNA's double-helix structure [20] revolutionized molecular genetics, by shifting focus to molecular evolution as scientists recognized nucleotide sequences (A, T, C, G) as key determinants of phenotypes.

In the present day, biological sciences welcome approaches from diverse domains like mathematical and data sciences to solve long standing biological questions. The integrative approach of researchers from non-biological background is also vital for an in-depth understanding of molecular evolution. Beyond the traditional doctrine of evolutionary biologists on fossil evidence and carbon dating, modern research highlights the capabilities of retrieval of molecular data in uncovering evolutionary signatures in inter and intra species studies. Molecular evolution, as a multidisciplinary field, deciphers genetic information to reconstruct evolutionary history through various cladistics tools. Furthermore, cutting-edge high throughput sequencing technologies of biomolecules (nucleic acid and amino acid) have proven as a gateway of new era of multidisciplinary biological research.

	U		C		A		G		
U	UUU	F	UCU	S	UAU	Y	UGU	C	U
	UUC		UCC		UAC		UGC		C
	UUA	L	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	W	G
C	CUU		CCU	P	CAU	H	CGU	R	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Q	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	I	ACU	T	AAU	N	AGU	S	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	K	AGA	R	A
	AUG	M	ACG		AAG		AGG		G
G	GUU	V	GCU	A	GAU	D	GGU	G	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	E	GGA		A
	GUG		GCG		GAG		GGG		G

Figure 1. A standard genetic code table displays all 64 codons, including 61 sense and three stop codons, and their assigned amino acids. It consists of 16 boxes, split into eight split and eight family boxes, illustrating codon degeneracy. For instance, UUU and UUC code for phenylalanine (Phe, F), GUN codons code for valine (Val). Except for minor variations in the start codon, E. coli follows the standard genetic code.

This article highlights the significance of an interdisciplinary approach in understanding molecular evolution. Recent advancements such as AlphaFold [21] that optimized artificial intelligence (AI) to solve the long-standing mystery of protein folding, exemplify the power of computational methods in biological data interpretation and prediction. The availability of extensive genomic data across species in various databases has further transformed molecular evolution research, allowing for deeper insights into evolutionary patterns. Multidisciplinary approaches, integrating AI, machine learning (ML), and mathematical modelling, have enabled researchers to tackle complex biological problems more effectively. Computational inference has become a crucial tool for studying molecular evolution, offering new perspectives and greater precision in analyzing genetic variation, evolutionary forces, and functional genomics. This integration of computational techniques can be further expanded to explore molecular evolution in even greater detail.

2. The importance of codon assignment and usage bias

Since amino acids are the building blocks of proteins, their codon assignments emphasizes their significance since prebiotic Earth. The assignment of 61 sense codons to 20 amino acids, known as codon degeneracy, is a fundamental aspect of the genetic code [22]. In the standard genetic code, degeneracy ranges from zero to six, with some amino acids encoded by a single codon while others have multiple synonymous codons (e.g., Phe: UUU, UUC). However, certain synonymous codons are preferentially used, a phenomenon called codon usage bias (CUB), varies by species [23]. CUB is influenced by factors such as GC content, tRNA abundance, gene expression, and growth temperature [24-26]. Studies in E. coli also reveal distinct CUB patterns in high (HEG) and low (LEG) expressed genes [27]. CUB is analzsed through base substitutions rather than insertions/deletions, with point mutations classified as transitions (ti) or transversions (tv), based on purine (R) and pyrimidine (Y) interchanges [28,29]. Despite more possible tv pathways, ti occurs more frequently—about four times higher in E. coli neutral regions [30,31], shaping DNA sequence evolution. Additionally, codon reassignment challenges the conventional genetic code, impacting protein synthesis and function [32].

3. Single nuelcotide variations (SNVs) and their discrepancies among codons

Point mutations, whether transitions (ti) or transversions (tv), can have either synonymous or non-synonymous consequences in genes. Synonymous mutations do not alter the encoded amino acid, whereas non-synonymous mutations lead to changes in the amino acid sequence of the resulting polypeptide [33-36]. The impact of single-nucleotide variants (SNVs) can be analyzed across the 61 sense codons through theoretical calculations and observations. Since each nucleotide within a codon can be substituted by three alternative nucleotides, every triplet codon has the potential to generate nine different codon combinations due to SNV [37,38]. These nine codon variants can be classified based on their synonymous or non-synonymous effects. For example, in the case of the codon UUU, substituting the third nucleotide results in the codons UUC, UUA, and UUG. Among these, UUU→UUC is a synonymous transition (*Sti*), but UUU→UUA and UUU→UUG are non-synonymous transversions (*Ntv*). By extending this analysis to the remaining two positions of the codon, the total number of possible synonymous transitions (*Sti*), synonymous transversions (*Stv*), non-synonymous transitions (*Nti*), and non-synonymous transversions (*Ntv*) can be determined. For UUU, these values are calculated as 1 *Sti*, 0 *Stv*, 2 *Nti*, and 6 *Ntv*. However, these values are not uniform across all degenerate codons. Two-fold degenerate (TFD) and four-fold degenerate

(FFD) codons exhibit different substitution patterns (Figure 2). Notably, TFD codons have no possibilities for *Stv*, which differs from FFD codons [97]. A comprehensive table summarizing the number of possible SNVs and their corresponding effects is presented in the genetic code table (Table 1) [97].

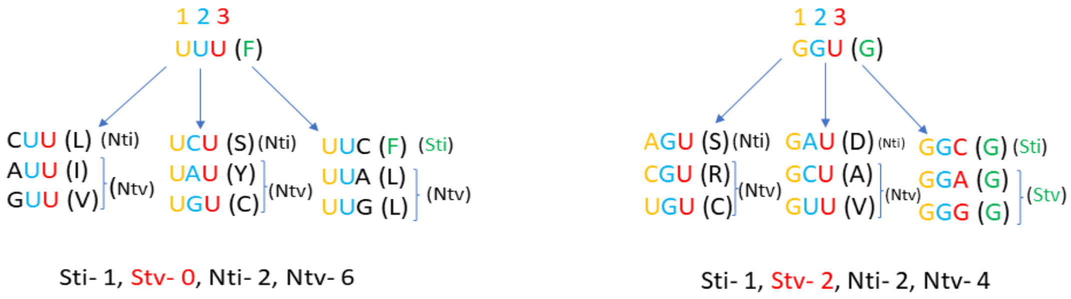

Figure 2. The expected values of *Sti*, *Stv*, *Nti*, and *Ntv* are calculated for two different types of degenerate codons. For instance, UUU, a TFD codon, has expected values of one *Sti*, zero *Stv*, two *Nti*, and six *Ntv*. Contrarily, GGU (FFD codon) has expected values of 1 *Sti*, 2 *Stv*, 2 *Nti*, and 4 *Ntv*. Notably, TFD codons do not allow for synonymous changes through tv. As a result, their distribution in coding sequences plays a crucial role in estimating transition bias.

Table 1. Estimated *Sti*, *Stv*, *Nti* and *Ntv* for all the codon

Codon	Sti	Stv	Nti	Ntv	Codon	Sti	Stv	Nti	Ntv	Codon	Sti	Stv	Nti	Ntv	Codon	Sti	Stv	Nti	Ntv
UUU	1	0	2	6	UCU	1	2	2	4	UAU	1	0	2	4	UGU	1	0	2	5
UUC	1	0	2	6	UCC	1	2	2	4	UAC	1	0	2	4	UGC	1	0	2	5
UUA	2	0	1	4	UCA	1	2	2	2	UAA	X	X	X	X	UGA	X	X	X	X
UUG	2	0	1	5	UCG	1	2	2	3	UAG	X	X	X	X	UGG	0	0	1	6
CUU	1	2	2	4	CCU	1	2	2	4	CAU	1	0	2	6	CGU	1	2	2	4
CUC	1	2	2	4	CCC	1	2	2	4	CAC	1	0	2	6	CGC	1	2	2	4
CUA	2	2	1	4	CCA	1	2	2	4	CAA	1	0	1	6	CGA	1	3	1	3
CUG	2	2	1	4	CCG	1	2	2	4	CAG	1	0	1	6	CGG	1	3	2	3
AUU	1	1	2	5	ACU	1	2	2	4	AAU	1	0	2	6	AGU	1	0	2	6
AUC	1	1	2	5	ACC	1	2	2	4	AAC	1	0	2	6	AGC	1	0	2	6
AUA	0	2	3	4	ACA	1	2	2	4	AAA	1	0	2	5	AGA	1	1	2	4
AUG	0	0	3	6	ACG	1	2	2	4	AAG	1	0	2	5	AGG	1	1	2	5
GUU	1	2	2	4	GCU	1	2	2	4	GAU	1	0	2	6	GGU	1	2	2	4
GUC	1	2	2	4	GCC	1	2	2	4	GAC	1	0	2	6	GGC	1	2	2	4
GUA	1	2	2	4	GCA	1	2	2	4	GAA	1	0	2	5	GGA	1	2	2	3
GUG	1	2	2	4	GCG	1	2	2	4	GAG	1	0	2	5	GGG	1	2	2	4

4. Emergence of substitution models in molecular evolution

The structural similarity between nucleotides facilitates the preferential selection or retention of transitions (ti) over transversions (tv) during the proofreading stage immediately following DNA replication. Additionally, the pairing of purine:purine (R:R) or pyrimidine:pyrimidine (Y:Y) is typically disallowed due to its potential to distort the DNA backbone structure [39]. It is well established that transitions and transversions do not occur at equal frequencies, despite transversions having more possible mutation pathways than transitions. In neutral regions of E. coli, transitions have been observed to occur approximately four times more frequently than transversions [30,31]. To quantify this transition bias, several substitution models have been proposed over time. The first such model, introduced by Jukes and Cantor in 1969, assumed an equal probability of all four nucleotides undergoing substitutions, with uniform rates of substitution between any two nucleotides [40]. Subsequently, the widely recognized Kimura two-parameter model (K80) was introduced, highlighting the unequal substitution rates between transitions and transversions in genomes [41]. The most widely accepted codon substitution model (CSM) emerged in 1994, proposed by Muse and Gaut. This model distinguishes between synonymous and non-synonymous substitution rates and accounts for the effects of purifying selection on non-synonymous substitutions [42].

5. Mutation and selection act as central forces in evolutionary dynamics

Since all population contain wild types as the most common form of allele, they also do contain variations in them, which are commonly known as variants or alternatives of the most common types [43]. Variants are the consequences of genetic mutations. Among different types of genetic mutations or chromosomal aberrations, SNVs are the most prevalent form of base substitutions observed across all populations [44]. Mutations are recognized as the primary instigator of variations in the DNA base sequence, playing a pivotal role not only in the evolutionary process [45] and in the development of complications such as cancer [46]. Nevertheless, mutation alone is inconsiderable, the selection of mutations is an essential step, particularly concerning the fitness of organisms, making it a pivotal driving force in the process of molecular evolution [47]. Most of the variations are unnoticeable to us as the variations first undergo the process of selection. The variations having an enhancing in fitness of the organism are usually selected and others are purged out of the population. Between the two types of selections (positive and negative), positive selection refers to the variant in a population providing higher fitness and reproductive success than the individuals carrying the non-variants [48]. Negative selection refers to the selective removal of the harmful or deleterious genetic variants from the population [49]. However, the concept of mutation-selection balance explains about the introduction of new deleterious mutations and purging out of the harmful mutations through purifying section [50,51]. But such fundamental understanding of mutation and selection did not contemplate to the neutral theory by Motto Kimura [52]. As Kimura explained the majority of the variations in the populations are nearly neutral, meaning they do not confer to the fitness of the organism to a large extent; hence the variations are due to genetic drift in smaller populations where chance events play a pivotal role in shaping the fate of the variants [52,53]. Non-synonymous substitutions are generally considered more deleterious to an organism's fitness than synonymous ones [54,55]. However, they can also enhance fitness by facilitating beneficial amino acid exchanges [56], with advantageous mutations shaping selection forces. Interestingly, recent studies suggest that even synonymous mutations can alter protein folding [57,58].

6. Strand-asymmetry, mutation bias, and GC content variability in prokaryotic genomes

Chargaff's first parity rule was crucial in supporting Watson and Crick's proposed DNA double-helix model [59,60]. Unlike the first rule, the second parity rule has known violations but generally applies to double-stranded sequences with similar substitution and selection patterns [61,62]. Local deviations arise due to replication and transcription pressures [63]. During DNA replication, the leading strand (LeS) and lagging strand (LaS) are synthesized differently [64], with the LeS experiencing greater single-stranded exposure [65]. Cytosine deamination in ssDNA occurs with a half-life of ~200 years [66], contributing to strand-dependent mutations, known as asymmetric directional mutation pressure [67]. Cytosine deamination and guanine oxidation are key sources of base substitutions, particularly G→T mutations [68,69]. A similar process occurs during transcription, where the non-template strand is exposed, leading to C→T/G→A mutations [70,71]. These transitions significantly contribute to polymorphism in both coding and non-coding regions. Since mutations tend to be AT-biased (Hershberg & Petrov, 2010), GC content variability in bacteria has been widely debated. GC% varies across prokaryotic genomes, from 13% in Zinderia insecticola to 75% in Aneromyxobacter dehalogenans [72]. While selectionist views have largely been rejected, the mutationist perspective, which attributes GC content to mutational pressure [73], is widely accepted. In prokaryotic genomes, where protein-coding genes dominate, GC mutational pressure is influenced by selective constraints. Weaker selection in neutral regions amplifies the impact of GC mutational pressure on genome composition [74].

7. Multidisciplinary integrative approach into evolutionary biology

Prior to the emergence of molecular evolution, evolutionary biologists primarily studied evolutionary history through inter-species comparisons and fossil records. However, statistics/mathemetical frameworks provided stornger backbone to both molecular as well as evolutionary genetics. Eminent researchers like Ronald Fisher's contribution in statistics provided a strong foundation for the theoretical biological research. It bridged the gap among different fields and fine-tuned population genetics and studies. However, the nascent phase of molecular evolution was traditionally viewed as an inter-species process, with researchers using phylogenetic methods like maximum likelihood, maximum parsimony, and Bayesian inference to trace evolutionary history and common ancestry [75]. Some methods rely on prior data to generate posterior data, each with their own advantages and limitations. The incorporation of algorithms like Markov chain Monte Carlo (MCMC) significantly enhanced Bayesian phylogenetics [76]. These methods have also influenced modern taxonomy [77]. Since mutation drives evolution [78], many unnoticed mutations shape genetic diversity. The rise of bioinformatics/computational biology and databases like DDBJ,NCBI, EMBL, and UniProt [79-84] have enabled large-scale genetic analyses. Phylogenetic analysis now plays a key role in studying strain-level evolution and genetic diversity. Eventually, tThe development of protein folding pattern prediction algorithms and programs has significantly enhanced researchers' understanding of these functional biomolecules in recent times.

8. Emerging computational opportunities and challenges in molecular evolution

Molecular evolution is increasingly driven by computational tools that enable large-scale analysis, modeling, and interpretation of genetic data. Several updated tools and platforms have recently emerged, significantly improving the accuracy and speed of evolutionary studies. For example, IQ-TREE 2 and RAxML-NG allow efficient construction of phylogenetic trees from genome-scale datasets using maximum likelihood approaches. BEAST 2 supports Bayesian inference for evolutionary time estimation and population dynamics, while HyPhy is extensively used for detecting selection pressure at the molecular level. Tools like PAML continue to offer robust methods for estimating substitution rates and evolutionary parameters, and MEGA12 remains a widely used, user-friendly platform for sequence alignment, model testing, and tree reconstruction among researchers [85-90]. Several R packages are widely used for molecular evolution, covering tasks such as phylogenetic analysis, sequence alignment, selection detection, evolutionary rate estimation, and comparative genomics, Model testing & ancestral reconstruction and data visualization. Simialrly, core python libraries like Biopython, ETE Toolkit, DendroPy, tree construction and visualization have been found to be helpful in evolutionary studies [91,92]. However, several custom pipelines can be made using different scripting tools for different purposes like sequence alignments, phylogeny study and to perform various evolutionary analyses.

Despite these advances, several computational challenges remain and offer exciting opportunities for multidisciplinary exploration. One essential requirement is the development of scalable algorithms for analyzing high-throughput phylogenomic data. There is also growing interest in integrating AI-ML with evolutionary models to predict functional outcomes of mutations or to classify gene families more accurately. Another emerging frontier is the simulation of protein structural evolution, which remains a computationally intensive task. Lastly, the integration of multi-omics data (genomics, transcriptomics, proteomics, and metabolomics) poses both a challenge and an opportunity. Therefore, construction of advanced pipelines could better offer exciting findings and help us better understand how organisms adapt and evolve at the molecular level.

9. Conclusions

Robustness is a ubiquitous property among biological systems (Kitano, 2004). The genetic code is considered robust because it exhibits redundancy and is tolerant to most harmful mutations [93]. This ensures genetic fidelity and preserves both biological information and function. However, some codons are susceptible to nonsense mutations, leading to truncated proteins. To maintain functional integrity, purifying selection plays a crucial role in eliminating deleterious mutations. The assignment of amino acids to codons remains debated, particularly regarding the impact of non-synonymous changes through ti or tv. In split codon boxes, ti is facilitated by the adjacent placement of purines (U←→C) and pyrimidines (A←→G). However, the rationale for the amino acid placement in neighbouring boxes is unclear. The amino acid exchangeability patterns across a wide range of organisms can be interesting to study, as all the organisms share the same genetic code table to a great extent. Furthermore, factors such as gene localization, strand bias, and gene expression influence mutational spectra [94], while context-dependent mutations remain a growing research focus [95,96]. Despite the near-universal applicability of the standard genetic code, studying homologous genes in prokaryotes and eukaryotes may reveal deeper insights into the genetic code's evolution and protein essentiality.

Therefore, modern-day molecular evolution research offers ample multidisciplinary opportunities to unravel the mysteries of evolutionary forces underlying our existence. This can be utilized in personalized medicine and biomedical research to study the evolutionary prospects of different diseases, ultimately contributing to more effective prevention and treatment strategies. In the Indian context, integrating molecular evolution into undergraduate education would inspire younger generations to explore multidisciplinary research, fostering future scientific advancements.

Multidisciplinary Domains

This research covers the domains: (a) Biology, (b) Computer Science

Funding

This research received no external funding.

Acknowledgments

I sincerely thank the anonymous reviewer(s) for their valuable suggestions, which have helped improve the quality of this manuscript. I also extend my heartfelt gratitude to Prof. Suvendra Kumar Ray and Prof. Siddhartha Sankar Satapathy for their insightful input, continuous support, and guidance throughout the journey.

Conflicts of Interest

The author(s) declare no conflict of interest.

Declaration on AI Usage

Artificial Intelligence (AI) tools were used in the preparation of this manuscript as follows: ChatGPT, developed by OpenAI was utilized for language editing, drafting of the abstract, with all outputs reviewed and edited by the author. The author remains responsible for the content's integrity and originality

References

[1] Nirenberg, M. W.; The genetic code; *Scientific American* **1963**, 208(3), 80-95.

[2] Crick, F.; Francis Crick. The Double Helix; *acta crystallographia* **1951**, 2-3.

[3] Watson, J. D.; Crick, F. H. C.; On protein synthesis; *The Symposia of the Society for Experimental Biology* **1958**, 12, 138-163.

[4] Crick, F.; Central dogma of molecular biology; *Nature* **1970**, 227(5258), 561-563; https://doi.org/10.1038/227561a0

[5] Gadgil, M.; Bossert, W. H.; Life historical consequences of natural selection; *The American Naturalist* **1970**, 104(935), 1-24; https://doi.org/10.1086/282638

[6] Gibbs, H. L.; Grant, P. R.; Oscillating selection on Darwin's finches; *Nature* **1987**, 327(6122), 511-513; https://doi.org/10.1038/327511a0

[7] Fisher, R. A.; XV. —The correlation between relatives on the supposition of Mendelian inheritance; *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* **1919**, 52(2), 399-433; https://doi.org/10.1017/S0080456800012163

[8] Berry, A.; Browne, J.; Mendel and Darwin; *Proceedings of the National Academy of Sciences* **2022**, 119(30), e2122144119; https://doi.org/10.1073/pnas.2122144119

[9] Dawkins, R.; *The Blind Watchmaker* Norton & Company: New York, NY, USA, **1986** ISBN: 978-0393351491.

[10] Esposito, M.; From human science to biology: The second synthesis of Ronald Fisher; *History of the Human Sciences* 2016, 29(3), 44-62; https://doi.org/10.1177/0952695116645958

[11] Hardy, G. H.; Mendelian proportions in a mixed population; *Science*; **1908** 28(706), 49-50; https://doi.org/10.1126/science.28.706.49

[12] Stern, C.; The Hardy-Weinberg law; *Science* 1943, 97(2510), 137-138; https://doi.org/10.1126/science.97.2510.137

[13] Mayr, E.; *Animal Species and Evolution*; Harvard University Press: Cambridge, MA, USA, **1963**

[14] Chandler, A. C.; The effect of extent of distribution on speciation; *The American Naturalist* **1914**, 48(567), 129-160; https://doi.org/10.1086/279445

[15] Simpson, G. G.; *The Major Features of Evolution*; Columbia University Press: New York, NY, USA, **1953.**

[16] Stanley, S. M.; A theory of evolution above the species level; *Proceedings of the National Academy of Sciences* **1975**, 72(2), 646-650; https://doi.org/10.1073/pnas.72.2.646

[17] Boyden, A.; Homology and analogy: A critical review of the meanings and implications of these concepts in biology; *American Midland Naturalist* **1947**, 648-669; https://doi.org/10.2307/2421727

[18] Zangerl, R.; The methods of comparative anatomy and its contribution to the study of evolution; *Evolution* **1948**, 351-374; https://doi.org/10.2307/2405743

[19] Goodhart, C. B.; The Sewall Wright Effect; *The American Naturalist* **1963**, 97(897), 407-409; https://doi.org/10.1086/282295

[20] Watson, J. D.; Crick, F. H. C.; The structure of DNA; *Cold Spring Harbor Symposia on Quantitative Biology* **1953**, 18, 123-131; https://doi.org/10.1101/SQB.1953.018.01.020

[21] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Hassabis, D.; Highly accurate protein structure prediction with AlphaFold; *Nature* **2021**, 596(7873), 583-589; https://doi.org/10.1038/s41586-021-03819-2

[22] Crick, F.; Chapter 8: The genetic code; In *What Mad Pursuit: A Personal View of Scientific Discovery* Basic Books: New York, NY, USA, **1988**; pp. 89-101.

[23] Ikemura, T.; Codon usage and tRNA content in unicellular and multicellular organisms; *Molecular Biology and Evolution* **1985**, 2(1), 13-34; https://doi.org/10.1093/oxfordjournals.molbev.a040343

[24] Gouy, M.; Gautier, C.; Codon usage in bacteria: Correlation with gene expressivity; *Nucleic Acids Research* **1982**, 10(22), 7055-7074; https://doi.org/10.1093/nar/10.22.7055

[25] McInerney, J. O.; Replicational and transcriptional selection on codon usage in Borrelia burgdorferi; *Proceedings of the National Academy of Sciences* **1998**, 95(18), 10698-10703; https://doi.org/10.1073/pnas.95.18.10698

[26] Seward, E. A.; Kelly, S.; Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms; *Genome Biology* **2016**, 17, 1-15; https://doi.org/10.1186/s13059-016-0930-7

[27] Sen, P.; Kurmi, A.; Ray, S. K.; Satapathy, S. S.; Machine learning approach identifies prominent codons from different degenerate groups influencing gene expression in bacteria; *Genes to Cells* **2022**, 27(10), 591-601; https://doi.org/10.1111/gtc.12985

[28] Collins, D. W.; Jukes, T. H.; Rates of transition and transversion in coding sequences since the human-rodent divergence; *Genomics* **1994**, 20(3), 386-396; https://doi.org/10.1006/geno.1994.1188

[29] Beura, P. K.; Sen, P.; Aziz, R.; Satapathy, S. S.; Ray, S. K.; Transcribed intergenic regions exhibit a lower frequency of nucleotide polymorphism than the untranscribed intergenic regions in the genomes of Escherichia coli and Salmonella enterica; *Journal of Genetics* **2023**, 102(1), 22; https://doi.org/10.1007/s12041-023-01397-7

[30] Seplyarskiy, V. B.; Kharchenko, P.; Kondrashov, A. S.; Bazykin, G. A.; Heterogeneity of the transition/transversion ratio in Drosophila and Hominidae genomes; *Molecular Biology and Evolution* 2012, 29(8), **1943-1955**; https://doi.org/10.1093/molbev/mss066

[31] Sen, P.; Aziz, R.; Deka, R. C.; Feil, E. J.; Ray, S. K.; Satapathy, S. S.; Stem region of tRNA genes favors transition substitution towards keto bases in bacteria; *Journal of Molecular Evolution* **2022**, 90(1), 114-123; https://doi.org/10.1007/s00239-022-10055-0.

[32] Osawa, S.; Jukes, T. H.; Codon reassignment (codon capture) in evolution; *Journal of Molecular Evolution* **1989**, 28(4), 271-278; https://doi.org/10.1007/BF02103414.

[33] Holmes, E. C.; Patterns of intra-and interhost nonsynonymous variation reveal strong purifying selection in dengue virus; *Journal of Virology* **2003**, 77(20), 11296-11298; https://doi.org/10.1128/JVI.77.20.11296-11298.2003.

[34] Nei, M.; Gojobori, T.; Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions; *Molecular Biology and Evolution* **1986,** 3(5), 418-426; https://doi.org/10.1093/oxfordjournals.molbev.a040410.

[35] Subramanian, S.; Significance of population size on the fixation of nonsynonymous mutations in genes under varying levels of selection pressure; *Genetics* **2013**, 193(3), 995-1002; https://doi.org/10.1534/genetics.112.147074.

[36] Teng, S.; Michonova-Alexova, E.; Alexov, E.; Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions; *Current Pharmaceutical Biotechnology* **2008**, 9(2), 123-133; https://doi.org/10.2174/138920108783955173.

[37] Gojobori, T.; Li, W. H.; Graur, D.; Patterns of nucleotide substitution in pseudogenes and functional genes; *Journal of Molecular Evolution* **1982**, 18(5), 360-369; https://doi.org/10.1007/BF02101694.

[38] Graur, D.; Li, W. H.; *Molecular Evolution*; Sinauer Associates: Sunderland, MA, USA, 2000.

[39] Wahl, M. C.; Sundaralingam, M.; Crystal structures of A-DNA duplexes; *Biopolymers: Original Research on Biomolecules* **1997**, 44(1), 45-63

[40] Jukes, T. H.; Cantor, C. R.; Evolution of protein molecules; In *Mammalian Protein Metabolism*; Munro, H. N., Ed.; Academic Press: New York, NY, USA, **1969**, Volume 3, 21-132. [No DOI available]

[41] Kimura, M.; A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences; *Journal of Molecular Evolution* **1980**, 16(2), 111-120; https://doi.org/10.1007/BF01731581.

[42] Muse, S. V.; Gaut, B. S.; A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome; *Molecular Biology and Evolution* **1994**, 11(5), 715-724; https://doi.org/10.1093/oxfordjournals.molbev.a040152 .

[43] Griffiths, A. J. F.; *An Introduction to Genetic Analysis*, 7th ed.; W. H. Freeman: New York, NY, USA, **2000.**

[44] Shendure, J.; Aiden, E. L.; The expanding scope of DNA sequencing; *Nature Biotechnology* **2012,** 30(11), 1084-1094; https://doi.org/10.1038/nbt.2421.

[45] Hershberg, R.; Mutation—the engine of evolution: studying mutation and its role in the evolution of bacteria; *Cold Spring Harbor Perspectives in Biology* **2015**, 7(9), a018077; https://doi.org/10.1101/cshperspect.a018077.

[46] Prehn, R. T.; The role of mutation in the new cancer paradigm; *Cancer Cell International* 2005, 5(1), 9; https://doi.org/10.1186/1475-2867-5-9.

[47] Sukhodolets, V. V.; The role of natural selection in evolution; *Genetika* **1986**, 22(2), 181-193.

[48] Gregory, T. R.; Understanding natural selection: Essential concepts and common misconceptions; *Evolution: Education and Outreach* **2009**, 2, 156-175; https://doi.org/10.1007/s12052-009-0114-3.

[49] Gulisija, D.; Crow, J. F.; Inferring purging from pedigree data; *Evolution* **2007**, 61(5), 1043-1051; https://doi.org/10.1111/j.1558-5646.2007.00088.x.

[50] Sarkar, S.; Haldane as biochemist: The Cambridge decade, **1923–1932**; In *The Founders of Evolutionary Genetics* Boston Studies in the Philosophy of Science; Springer: Dordrecht, Netherlands, **1992**; Volume 142, 53-81; https://doi.org/10.1007/978-94-011-2864-0_4.

[51] Nachman, M. W.; Haldane and the first estimates of the human mutation rate; *Journal of Genetics* **2004,** 83(3), 231-233; https://doi.org/10.1007/BF02717955.

[52] Kimura, M.; *The Neutral Theory of Molecular Evolution* Cambridge University Press: Cambridge, UK, **1983.** [No DOI available]

[53] Hurst, L. D.; Genetics and the understanding of selection; *Nature Reviews Genetics* **2009**, 10, 83-93; https://doi.org/10.1038/nrg2506 .

[54] Eyre-Walker, A.; Keightley, P. D.; The distribution of fitness effects of new mutations; *Nature Reviews Genetics* **2007**, 8(8), 610-618; https://doi.org/10.1038/nrg2146 .

[55] Schmidt, S.; Gerasimova, A.; Kondrashov, F. A.; Adzhubei, I. A.; Kondrashov, A. S.; Sunyaev, S.; Hypermutable non-synonymous sites are under stronger negative selection; *PLoS Genetics* **2008**, 4(11), e1000281; https://doi.org/10.1371/journal.pgen.1000281 .

[56] Rodrigue, N.; Philippe, H.; Lartillot, N.; Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles; *Proceedings of the National Academy of Sciences* **2010,** 107(10), 4629-4634; https://doi.org/10.1073/pnas.0908092107.

[57] Shabalina, S. A.; Spiridonov, N. A.; Kashina, A.; Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity; *Nucleic Acids Research* **2013**, 41(4), 2073-2094; https://doi.org/10.1093/nar/gks1205.

[58] Liu, Y.; Yang, Q.; Zhao, F.; Synonymous but not silent: The codon usage code for gene expression and protein folding; *Annual Review of Biochemistry* **2021**, 90, 375-401; https://doi.org/10.1146/annurev-biochem-052820-105656.

[59] Elson, D.; Chargaff, E.; On the deoxyribonucleic acid content of sea urchin gametes; *Experientia* **1952,** 8(4), 143-145; https://doi.org/10.1007/BF02154632.

[60] Forsdyke, D. R.; Mortimer, J. R.; Chargaff's legacy; *Gene* **2000,** 261(1), 127-137; https://doi.org/10.1016/S0378-1119(00)00479-2.

[61] Lobry, J. R.; Properties of a general model of DNA evolution under no-strand-bias conditions; *Journal of Molecular Evolution*; **1995,** 40(3), 326-330; https://doi.org/10.1007/BF00163239.

[62] Sueoka, N.; Intrastrand parity rules of DNA base composition and usage biases of synonymous codons; *Journal of Molecular Evolution* **1995**, 40(3), 318-325; https://doi.org/10.1007/BF00163238.

[63] Francino, M. P.; Ochman, H.; Strand asymmetries in DNA evolution; *Trends in Genetics* **1997**, 13(6), 240-245; https://doi.org/10.1016/S0168-9525(97)01119-4.

[64] Kornberg, A.; DNA replication; *Trends in Biochemical Sciences*; **1984,** 9(4), 122-124; https://doi.org/10.1016/0968-0004(84)90110-5.

[65] Powdel, B. R.; Satapathy, S. S.; Kumar, A.; Jha, P. K.; Buragohain, A. K.; Borah, M.; Ray, S. K.; A study in entire chromosomes of violations of the intra-strand parity of complementary nucleotides (Chargaff's second parity rule); *DNA Research* **2009**, 16(6), 325-343; https://doi.org/10.1093/dnares/dsp019.

[66] Lindahl, T.; Nyberg, B.; Heat-induced deamination of cytosine residues in deoxyribonucleic acid; *Biochemistry* **1974,** 13(16), 3405-3410; https://doi.org/10.1021/bi00713a035.

[67] Lobry, J. R.; Sueoka, N.; Asymmetric directional mutation pressures in bacteria; *Genome Biology*; **2002,** 3(10), 1-14; https://doi.org/10.1186/gb-2002-3-10-research0058.

[68] Kino, K.; Sugiyama, H.; Possible cause of G·C→C·G transversion mutation by guanine oxidation product, imidazolone; *Chemistry & Biology* **2001**, 8(4), 369-378; https://doi.org/10.1016/S1074-5521(01)00020-6.

[69] Bhagwat, A. S.; Hao, W.; Townes, J. P.; Lee, H.; Tang, H.; Foster, P. L.; Strand-biased cytosine deamination at the replication fork causes cytosine to thymine mutations in Escherichia coli; *Proceedings of the National Academy of Sciences* **2016**, 113(8), 2176-2181; https://doi.org/10.1073/pnas.1525329113.

[70] Francino, M. P.; Ochman, H.; Strand asymmetries in DNA evolution; *Trends in Genetics* **1997**, 13(6), 240-245; https://doi.org/10.1016/S0168-9525(97)01119-4.

[71] Mugal, C. F.; von Grünberg, H. H.; Peifer, M.; Transcription-induced mutational strand bias and its effect on substitution rates in human genes; *Molecular Biology and Evolution* **2009**, 26(1), 131-142; https://doi.org/10.1093/molbev/msn241.

[72] Zhao, X.; Zhang, Z.; Yan, J.; Yu, J.; GC content variability of eubacteria is governed by the pol III α subunit; *Biochemical and Biophysical Research Communications* **2007**, 356(1), 20-25; https://doi.org/10.1016/j.bbrc.2007.02.086.

[73] Muto, A.; Osawa, S.; The guanine and cytosine content of genomic DNA and bacterial evolution; *Proceedings of the National Academy of Sciences* **1987**, 84(1), 166-169; https://doi.org/10.1073/pnas.84.1.166.

[74] Brocchieri, L.; The GC content of bacterial genomes; *Journal of Phylogenetics & Evolutionary Biology* **2014**, 2, 1000133; https://doi.org/10.4172/2329-9002.1000133.

[75] Revell, L. J.; Mahler, D. L.; Peres-Neto, P. R.; Redelings, B. D.; A new phylogenetic method for identifying exceptional phenotypic diversification; *Evolution* **2012,** 66(1), 135-146; https://doi.org/10.1111/j.1558-5646.2011.01435.x.

[76] Brooks, S.; Markov chain Monte Carlo method and its application; *Journal of the Royal Statistical Society: Series D (The Statistician)* **1998,** 47(1), 69-100; https://doi.org/10.1111/1467-9884.00114.

[77] Heath, T. A.; Hedtke, S. M.; Hillis, D. M.; Taxon sampling and the accuracy of phylogenetic analyses; *Journal of Systematics and Evolution* **2008,** 46(3), 239-257; https://doi.org/10.3724/SP.J.1002.2008.08016.

[78] Hershberg, R.; Mutation the engine of evolution: Studying mutation and its role in the evolution of bacteria; *Cold Spring Harbor Perspectives in Biology* **2015,** 7(9), a018077; https://doi.org/10.1101/cshperspect.a018077.

[79] Mashima, J.; Kodama, Y.; Fujisawa, T.; Katayama, T.; Okuda, Y.; Kaminuma, E.; Takagi, T.; DNA data bank of Japan; *Nucleic Acids Research* **2017**, 45(D1), D25-D31; https://doi.org/10.1093/nar/gkw1001 .

[80] NCBI Resource Coordinators; Database resources of the National Center for Biotechnology Information; *Nucleic Acids Research* **2016**, 44(D1), D7-D19; https://doi.org/10.1093/nar/gkv1290.

[81] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E.; The protein data bank; *Nucleic acids research* **2000**, *28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

[82] Zardecki, C., Dutta, S., Goodsell, D. S., Lowe, R., Voigt, M., & Burley, S. K.; Educational resources supporting molecular explorations through biology and medicine. *Protein Science* **2022**, *31*(1), 129-140. https://doi.org/10.1002/pro.4200

[83] The UniProt Consortium , UniProt: the Universal Protein Knowledgebase in 2025, *Nucleic Acids Research*, Volume 53, Issue D1, 6 January **2025,** Pages D609–D617,. https://doi.org/10.1093/nar/gkae1010

[84] Benson, D. A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Ostell, J.; Pruitt, K. D.; Sayers, E. W.; GenBank. *Nucleic Acids Research* **2018,** 46(D1), D41–D47; https://doi.org/10.1093/nar/gkx1094

[85] Steane, D. A. et al.; Phylogenomic insights using IQ-TREE 2 and BEAST 2; *Molecular Phylogenetics and Evolution* **2023**, 184, 107781; https://doi.org/10.1016/j.ympev.2023.107781

[86] Kozlov, A. M. et al.; RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference; *Bioinformatics* **2020**, 36(5), 2672–2674; https://doi.org/10.1093/bioinformatics/btaa093

[87] Suchard, M. A. et al.; Bayesian phylogenetic and phylodynamic data integration using BEAST 2.6; *Virus Evolution* **2020**, 6(1), veaa042; https://doi.org/10.1093/ve/veaa042

[88] Kosakovsky Pond, S. L. et al.; HyPhy 2.5—A customizable platform for evolutionary hypothesis testing using phylogenies; *Molecular Biology and Evolution* **2020**, 37(1), 295–299; https://doi.org/10.1093/molbev/msz197

[89] Kumar, S.; Stecher, G., Suleski.; M., Sanderford.; M., Sharma, S.; & Tamura, K.; MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution* **2024**, *41*(12), msae263. https://doi.org/10.1093/molbev/msae263

[90] Yang, Z.; PAML 4: Phylogenetic analysis by maximum likelihood; *Molecular Biology and Evolution*; **2007,** 24(8), 1586–1591; https://doi.org/10.1093/molbev/msm088

[91] R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2024; https://www.R-project.org/

[92] Langa, Ł. Python 3.9.13 is now available. *Python Insider*, **2022**, May 17. Archived from the original on May 17, 2022. Retrieved May 21, 2022. Available at: https://blog.python.org/2022/05/python-3913-is-now-available.html

[93] Maeshiro, T.; Kimura, M.; The role of robustness and changeability on the origin and evolution of genetic codes; *Proceedings of the National Academy of Sciences* **1998**, 95(9), 5088-5093; https://doi.org/10.1073/pnas.95.9.5088.

[94] Franklin, I.; Lewontin, R. C.; Is the gene the unit of selection?; *Genetics* **1970**, 65(4), 707-734. [No DOI available]

[95] Sung, W.; Ackerman, M. S.; Gout, J. F.; Miller, S. F.; Williams, E.; Foster, P. L.; Lynch, M.; Asymmetric context-dependent mutation patterns revealed through mutation–accumulation experiments; *Molecular Biology and Evolution* **2015**, 32(7), 1672-1683; https://doi.org/10.1093/molbev/msv055.

[96] Zhu, Y.; Neeman, T.; Yap, V. B.; Huttley, G. A.; Statistical methods for identifying sequence motifs affecting point mutations; *Genetics*; **2017**, 205(2), 843-856; https://doi.org/10.1534/genetics.116.193029.

[97] Beura, P. K..; *A Study on Single Nucleotide Variations in Different Regions of Escherichia coli Genome Sequences*, **2024,(**Doctoral dissertation, Tezpur University).