Current Trends in Biotechnology and Pharmacy
Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

Comparative Codon Bias Analysis of WRKY Gene Family in *Oryza sat*iva subsp. *indica* and *japonica*

Bevin Nishanth J, G Atul Babu, and Suji Somasundaram*

Department of Biotechnology, Saveetha School of Engineering, SIMATS, Thandalam,
Chennai - 602 105, Tamil Nadu, India
*Corresponding author: sujis.sse@saveetha.com

Abstract

Codon usage bias (CUB) is an important molecular evolutionary feature shaped by various forces such as mutational pressure and natural selection. In this study, we performed a comprehensive codon usage bias analysis of WRKY transcription factor (TF) genes in *Oryza sativa* subsp. *indica* and *japonica* to explore their underlying evolutionary dynamics. A total of 109 and 128 valid WRKY CDSs were retrieved and analyzed respectively. Several indices including nucleotide composition, GC content, RSCU (Relative Synonymous Codon Usage), ENC (Effective Number of Codons), and neutrality plot analysis were used to assess the CUB patterns. Our results revealed that both subpopulations exhibited moderate codon bias, with the majority of genes displaying ENC values above 40, indicating weak bias. The ENC vs. GC3s plot showed most data points clustered around the expected curve, suggesting mutational pressure as the primary force shaping codon usage. Neutrality plot regression slopes (~0.28) and strong correlations between GC3 and GC12 further supported the dominance of mutational bias. PR2 plots demonstrated slight deviations in A/T and G/C proportions at the third codon position, indicating minor influences of natural selection. Additionally, correspondence analysis (COA) showed that the first axis, largely influenced by GC3s, explained over 90% of codon usage variation, reinforcing the role of base composition. Comparative analysis with existing literature highlighted that our findings align with previous reports on codon usage in stress-responsive gene families in rice. These results collectively suggest that although natural selection may play a minor role, mutational pressure predominantly governs codon usage patterns in WRKY TF genes of *O. sativa*, offering insights for evolutionary genomics and crop improvement.

Keywords: Codon usage bias; WRKY transcription factor; *Oryza sativa*; Mutational pressure; Natural selection; Rice genomics

Introduction

The genetic code is degenerate, meaning that most amino acids can be encoded by multiple synonymous codons. However, organisms often preferentially use certain synonymous codons over others—a phenomenon known as codon usage bias (CUB)(1, 2). This non-random usage is influenced by several factors, including mutational bias, natural selection, translational efficiency, gene expression levels, and tRNA abundance(3). Understanding codon usage bias is vital in molecular biology and biotechnology, as it influences protein translation dynamics, mRNA stability, and the overall cellular economy of gene expression(4). In plants, research on codon usage has provided significant insights into genome evolution, gene function, and adaptation to environmental conditions. Beyond its fundamental evolutionary significance, codon usage patterns are increasingly utilized in synthetic biology to optimize gene constructs for enhanced expression in heterologous systems(5, 6). For example, selecting preferred codons that match host tRNA abundances can improve the efficiency of protein production in genetically modified crops.

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

The WRKY gene family is one of the most functionally diverse and biologically important groups of transcription factors in plants. WRKY transcription factors play key roles in numerous physiological and developmental processes, including seed germination, leaf senescence, secondary metabolism, and, notably, responses to both biotic and abiotic stresses(7). They are defined by a highly conserved WRKYGQK heptapeptide sequence and a zinc finger-like motif, which facilitate specific binding to W-box elements located in the promoter regions of target genes(8). Due to their extensive role in stress-responsive pathways, WRKY genes are often targeted in plant breeding and genetic engineering programs aimed at enhancing crop resilience(9, 10). *Oryza sativa* (rice), a staple food for over half of the global population, provides a compelling system to investigate codon usage bias, particularly because of its two major cultivated subspecies—*indica* and *japonica*. These subspecies are ecologically and genetically distinct. *Indica* varieties are typically grown in tropical and subtropical climates and exhibit greater genetic diversity, whereas *japonica* varieties are adapted to temperate regions and are genetically more uniform. Differences in genome structure, nucleotide composition, and gene expression patterns between these subspecies have been well documented(11, 12). However, the comparative codon usage bias in transcription factor families, such as WRKY, remains underexplored. The genomic divergence between *indica* and *japonica* subspecies not only reflects their evolutionary histories but also subspecies-specific adaptation mechanisms(13). As transcription factors like WRKY genes play central roles in modulating gene expression under environmental stress, analyzing their codon usage patterns can provide deeper insights into the selective pressures shaping these regulatory elements(14). For instance, an elevated GC content at the third codon position (GC3) may reflect mutational bias, while deviations from expected codon usage models may suggest translational selection favoring efficient gene expression(15).

Codon usage bias analysis often utilizes several quantitative metrics. Analyzing GC content at the first (GC1), second (GC2), and third (GC3) codon positions reveals trends in nucleotide distribution. The Effective Number of Codons (ENC) quantifies the degree of codon preference, with lower values *indica*ting a stronger bias. Relative Synonymous Codon Usage (RSCU) values help identify codons that are overrepresented or underrepresented.

Additionally, visual methods like the ENC-GC3s plot, neutrality plot, and PR2 bias plot enable researchers to infer the evolutionary forces, namely mutation pressure versus natural selection, that influence codon usage patterns. This comparative study not only deepens our understanding of the molecular evolution of WRKY genes in *Oryza sativa* but also identifies subspecies-specific codon usage trends that could be leveraged in crop biotechnology. By identifying optimal codons, the findings offer a valuable resource for codon optimization strategies aimed at enhancing transgene expression, particularly under stress conditions where WRKY factors are crucial.

Future research might integrate these results with transcriptomic and proteomic data or extend the analysis to other transcription factor families to develop a comprehensive view of codon usage evolution in rice and other cereals.

Materials and Methodology

Sequence Retrieval of WRKY Gene Family in Rice

The complete coding sequences (CDSs) of WRKY transcription factor genes from *Oryza sativa* subsp. *indica* and *japonica* were obtained from the Plant Transcription Factor Database v5.0 (PlantTFDB). Each sequence underwent quality control to ensure accuracy in codon usage analysis.

The filtering criteria were as follows(16) Each CDS had to start with the ATG start codon and end with a valid stop codon (TAA, TAG, or TGA). Sequences containing internal stop codons were excluded. Only sequences with lengths that are multiples of three and exceed 300 base

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

pairs (bp) were retained. Redundant or incomplete gene models were discarded to maintain dataset consistency.

Analysis of Nucleotide Composition and GC Content

The overall GC content, as well as the position-specific GC content at the first (GC1), second (GC2), and third (GC3) codon positions, was calculated for each gene. In addition, the nucleotide composition at the third codon position, namely, the frequencies of A3, T3, G3, and C3, was analyzed to evaluate codon base bias. These values were computed using Python scripts.

Calculation of Codon Usage Indices

The degree of codon usage bias was evaluated using the following indices:

***Relative Synonymous Codon Usage* (RSCU)**

The RSCU value is defined as the ratio of the observed frequency of a codon to its expected frequency assuming equal usage of all synonymous codons for a particular amino acid. RSCU values were calculated using custom Python scripts, and codons were classified based on their usage bias:

RSCU = 1: Codon usage is unbiased (random selection)(17).
RSCU > 1: Codon is used frequently than expected and considered preferred(18).
RSCU < 1: Codon is used less frequently than expected and considered non-preferred(17).

Codons with an RSCU value greater than 1.6 were classified as over-represented, whereas those with an RSCU value below 0.6 were considered under-represented(19). Non-degenerate codons, including AUG (Met), UGG (Trp), and the three stop codons (UAG, UAA, UGA), were excluded from the RSCU-based bias analysis because they do not have synonymous counterparts.

***Effective Number of Codons* (ENC)**

ENC values were used to quantify the overall codon usage bias. Values range from 20 (maximum bias) to 61 (no bias). Genes with ENC values <50 were interpreted as having moderate to strong codon bias. The top 10% of genes with the lowest ENC values (*indica*ting high codon bias) were considered to represent highly expressed genes. The top 10% of genes with the highest ENC values (*indica*ting low codon bias) were treated as low-expression genes(20).

High frequency and Optimal codons

Codons with RSCU values greater than 1 are considered high-frequency codons(21). To explore codon usage preferences, the top 10% of genes with the highest and lowest effective number of codons (ENC) values were selected to create datasets representing high- and low-expression genes, respectively. RSCU values for codons in both datasets were calculated and compared using ΔRSCU. Codons with RSCU > 1 in the high-expression dataset and RSCU < 1 in the low-expression dataset were identified. A ΔRSCU ≥ 0.08 and RSCU > 1 between the two datasets was used as the criterion to designate a codon as putatively optimal(22).

Multivariate Analyses of Codon Bias
ENC-GC3s Plot

To investigate the association between codon usage bias and mutational pressure, an ENC–GC3s plot (commonly referred to as an ENC plot) was generated. In this plot, the x-axis denotes the GC3s value, representing the GC content at synonymous third codon positions, while the y-axis displays the observed ENC value for each gene. The expected ENC values were determined using the following formula(20):

$$ENC = 2 + GC3s + \frac{29}{GC3s^2 + (1 - GC3s)^2}$$

If mutation pressure alone governs codon bias, the gene data points are expected to fall on or near the standard curve. Conversely, if natural selection or other factors also influence codon usage, the data points will fall below the expected ENC curve(23). To measure each gene's deviation

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

from neutrality, the relative deviation was calculated using the formula:

$$\frac{\mathrm{ENC}_{expected} - \mathrm{ENC}_{observed}}{\mathrm{ENC}_{expected}}$$

A higher deviation value *indica*tes a stronger influence of selection or additional evolutionary constraints on codon usage.

Neutrality Plot

To assess the primary evolutionary forces—natural selection or mutational pressure—shaping codon usage patterns in WRKY genes, a neutrality plot was constructed. In this plot, GC3 (GC content at the third codon position) is plotted along the x-axis, while GC12, the mean GC content of the first and second codon positions (GC1 and GC2), is plotted along the y-axis. The relationship between GC3 and GC12 was analyzed using a linear regression model. The slope of the regression line provides insights into the balance of evolutionary forces. A slope close to 1.0 suggests that mutation pressure primarily determines codon usage, whereas a slope near 0.0 indicates a significant role of natural selection in shaping codon preference(24). The statistical significance of the regression coefficient was also evaluated.

PR2 Bias Plot

To further explore the impact of strand-specific mutational asymmetry and natural selection, PR2-bias plots were created. In this analysis, the x-axis represents the ratio. G3 / (G3 + C3), while the y-axis denotes the ratio A3 / (A3 + T3)(25). According to Parity Rule 2 (PR2) conditions, where A equals T and G equals C, the data points are expected to cluster around the central point (0.5, 0.5). Deviations from this central point *indica*te asymmetric base usage at the third codon position, which may be due to directional selection, mutational bias, or both. A uniform distribution of data points around the center suggests mutation-driven usage, whereas skewed clustering indicates the presence of selective constraints(11).

***Correspondence analysis* (COA)**

COA is a multivariate statistical technique commonly employed to examine codon usage patterns. In this study, COA was performed using the RSCU values of the codons. Each coding sequence (CDS) was encoded as a 59-dimensional vector, omitting codons corresponding to methionine (Met), tryptophan (Trp), and termination signals. Within this space, each point corresponded to a synonymous codon (26).

Results and Discussions

Nucleotide Composition and GC Content of WRKY Genes

The analysis of nucleotide composition and GC content in WRKY transcription factor (TF) genes of *Oryza sativa* subspecies has uncovered significant patterns. Both *indica* (109 CDS) and *japonica* (128 CDS) WRKY genes exhibited high GC content, with overall values of 61.37% and 61.03%, respectively (Fig 1). These represent the highest GC levels reported for WRKY genes across plant species to date(27, 28). This substantial GC richness stands in stark contrast to dicotyledonous species such as *Helianthus annuus* (43.42%)(27) and *Glycine max* (45.15%)(29). Even among monocots, rice surpasses *Musa acuminata*, which exhibits a GC content of 55.49%(30). These findings establish *O. sativa* as possessing one of the most GC-enriched WRKY gene families characterized thus far. The third codon position (GC3) content is particularly elevated in rice WRKY genes, with *indica* and *japonica* reaching 73.93% and 73.22%, respectively. These values exceed those observed in *Musa acuminata* (62.13%)(30) by more than 10 percentage points and far exceed typical dicot values, such as *H. annuus* (39.60%)(27) and *G. max* (43.37%)(29). This strong GC3 bias underscores the well-established monocot-dicot dichotomy in codon usage, wherein monocots preferentially employ G/C-ending codons, while dicots exhibit a tendency toward A/T-ending codons(31, 32). Additionally, the analysis of nucleotide composition revealed a pronounced bias

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

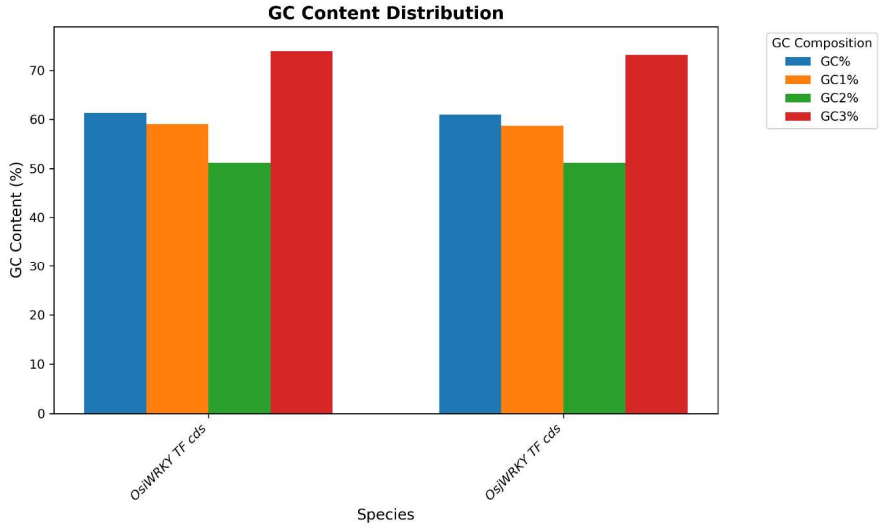

Fig 1: GC content distribution across codon positions in *Oryza sativa* WRKY genes

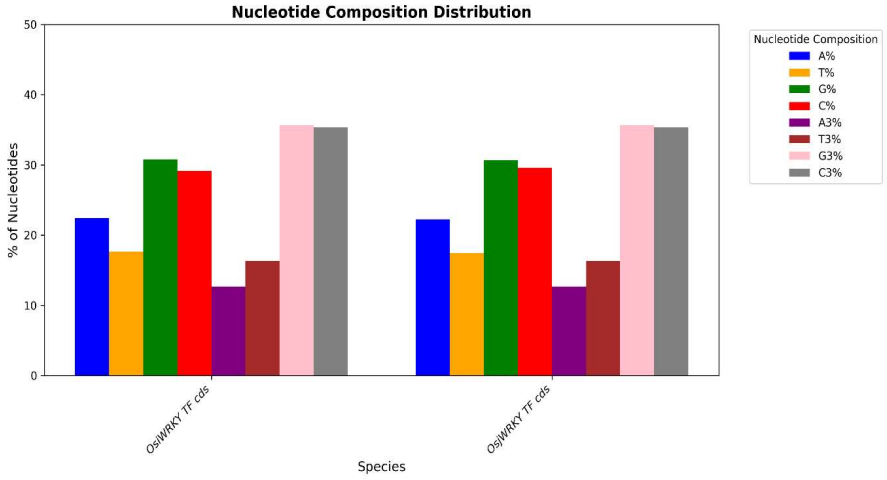

Fig 2: Base composition (A, T, C, G) at the third codon position

towards C and G nucleotides at the third codon position (Fig 2). In *indica*, C3% = 35.32% and G3% = 35.69%, significantly surpassing A3% = 12.69% and T3% = 16.31%. A similar trend was observed in *japonica*, with C3% = 35.35%, G3% = 35.66%, A3% = 12.64%, and T3% = 16.35%. These findings support the notion that codon usage in rice WRKY genes is highly GC-biased, particularly at synonymous sites, potentially influencing mRNA stability, translation efficiency, or gene expression.

Codon Usage Indices in *Oryza sativa* WRKY Genes

***Effective Number of Codons* (ENC)**

Codon usage bias serves as a crucial *indica*tor in genetic studies. Lower Effective Number of Codons (ENC) values suggest a stronger bias, while values nearing 61

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

*indica*te a uniform use of synonymous codons. In this study, the WRKY genes of *Oryza sativa indica* showed an average ENC of 49.92, whereas those of *Oryza sativa japonica* had a similar ENC of 49.68. These moderate ENC values suggest that WRKY genes in both subspecies exhibit a weak to moderate codon usage bias, likely shaped by a combination of natural selection and mutation pressure.

Relative Synonymous Codon Usage (RSCU) Patterns

The RSCU analysis revealed a preference for codons ending in guanine or cytosine, consistent with the GC-rich composition of rice WRKY genes. In *indica*, codons with high preference (RSCU > 1.6) include GGC (1.98), AGG (1.93), CCG (1.87), CTC (1.84), and CTG (1.76), while in *japonica*, the preferred codons are GGC (1.96), AGG (1.87), CCG (1.86), CTC (1.86), and CTG (1.82). Conversely, codons with low representation (RSCU < 0.6), such as AGA, TTA, ATA, CGA, GTT, and GGT, predominantly end in adenine or thymine. This biased codon usage suggests that selection pressures related to translation may favor certain codons to enhance translational efficiency or tRNA availability, particularly in highly expressed WRKY genes.

Optimal Codon Usage and Expression Implications

The identification of 25 and 26 optimal codons in *indica* and *japonica*, respectively, highlights a substantial degree of codon optimization in rice WRKY genes. This level of optimization notably surpasses that of *M. acuminata*, which utilizes only 14 optimal codons(30), and its higher than that of *H. annuus*, where only 3 optimal codons were identified(27). The high number of optimal codons in rice suggests extensive selection pressure favoring efficient translation, likely driven by the functional importance of WRKY transcription factors in abiotic and biotic stress responses (Table 1).

Table 1: Optimal Codon Identification

Species	Num Genes	High Freq Codons (n)	High Freq Codons	Optimal Codons (n)	Optimal Codons
OsiWRKY_TF	109	28	AAC, AAG, ACC, ACG, AGC, AGG, ATC, ATG, CAC, CAG, CCG, CGC, CGG, CTC, CTG, GAC, GAG, GCC, GCG, GGC, GTC, GTG, TAC, TCC, TCG, TGC, TGG, TTC	25	AAC, AAG, ACC, ACG, AGC, ATC, CAC, CAG, CCG, CGC, CGG, CTC, CTG, GAC, GAG, GCC, GCG, GGC, GTC, GTG, TAC, TCC, TCG, TGC, TTC
OsjWRKY_TF	128	28	AAC, AAG, ACC, ACG, AGC, AGG, ATC, ATG, CAC, CAG, CCG, CGC, CGG, CTC, CTG, GAC, GAG, GCC, GCG, GGC, GTC, GTG, TAC, TCC, TCG, TGC, TGG, TTC	26	AAC, AAG, ACC, ACG, AGC, ATC, CAC, CAG, CCG, CGC, CGG, CTC, CTG, GAC, GAG, GCC, GCG, GGC, GGG, GTC, GTG, TAC, TCC, TCG, TGC, TTC

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

The most frequently preferred codons in rice, including GGC, AGG, CCG, CTC, and CTG (each with RSCU > 1.6), consistent with the overall genomic bias toward these nucleotides in monocots(31, 32). The consistency of these codons across both rice subspecies further supports the hypothesis that translational selection has acted conservatively and intensively on these genes.

ENC-Plot Analysis

The connection between codon usage bias and GC3s content in the WRKY transcription factor (TF) genes of *Oryza sativa* subsp. *Indica* and *japonica* were assessed using ENC-GC3s plots (Fig. 3A). In both subspecies, most data points were located on or very near the expected curve, *indica*ting that mutational pressure, particularly the GC

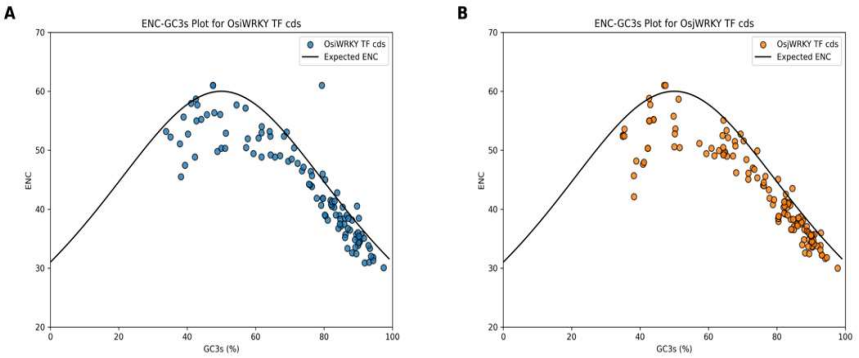

Fig 3: A) ENC-GC3s plot for WRKY genes in *O. sativa* subsp. *indica*. Most genes fall below the expected ENC curve, *indica*ting the role of selection. B) ENC-GC3s plot for *O. sativa* subsp. *japonica*. Similar to *indica*, most WRKY genes deviate from the expected curve.

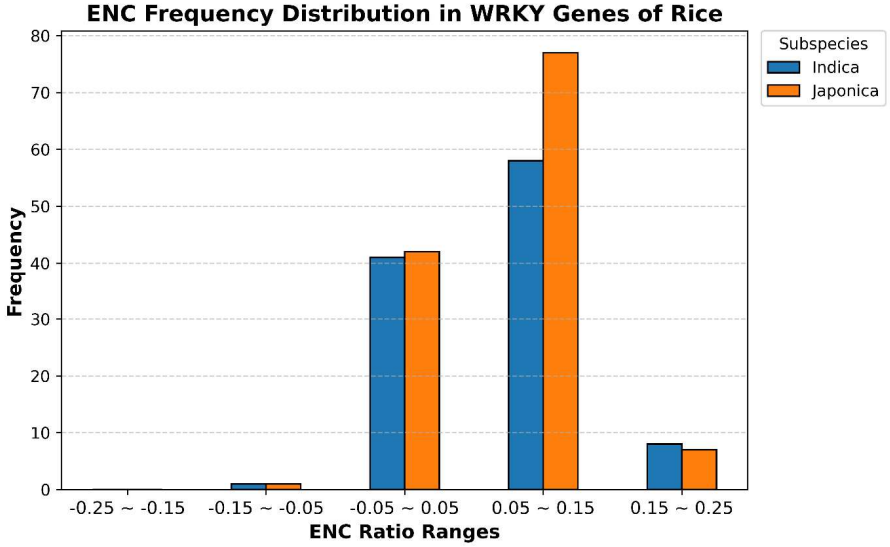

Fig 3: B) ENC frequency distribution of WRKY genes in *O. sativa indica* and *japonica*.

Comparative Codon Bias Analysis of WRKY Gene Family

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

composition at the third codon position, predominantly shapes codon usage patterns. With only a minority of genes deviating below the curve—these exceptions likely reflecting selection for translation or expression efficiency. This pattern is nearly identical to findings in *Musa acuminata* (banana), where WRKY genes display high GC3, weak but directed bias, and almost all genes conform closely to the mutation-driven expectation(30). Similarly, in sunflower (*Helianthus annuus*), WRKY genes also cluster along the neutral expectation, further supporting mutation as the chief force(27). In soybean (*Glycine max*), a positive correlation between GC12 and GC3 confirms a dominant mutational bias, though optimal codons may still favor C/G in the third position(33, 29). In contrast, analyses of *Arabidopsis thaliana* and *Brassica rapa* reveal higher GC in *Arabidopsis*, but more genes in both species deviate from the neutral curve, *indica*ting a stronger role for natural selection in certain dicot lineages(34). This diversity is echoed by neutrality plot and hierarchical clustering studies that suggest a spectrum of mutational versus selective pressure across dicots, with some species more selection-driven and others more mutation-driven. Supporting this, ENC ratio frequency distribution analysis showed that over 90% of WRKY genes had ENCobs values closely aligning with ENCexp, with most ratios ranging between 0.05 and 0.15 (Fig. 3B). This constraint is even more pronounced than in other monocots, *indica*ting possibly stronger genome-wide mutational processes or consistent translational machinery affecting codon choice.

PR2-Plot Analysis

To explore the effects of mutational bias and natural selection on nucleotide composition at the third codon position, PR2-bias plots were generated for WRKY transcription factor genes in *Oryza sativa* subsp. *indica* and *japonica* (Fig. 4). These plots display the ratio of A3 / (A3 + T3) on the Y-axis and G3 / (G3 + C3) on the X-axis, allowing for the visualization of deviations from the expected parity value of 0.5. PR2-bias plots in O. sativa WRKY TFs reflect a clear base composition asymmetry: substantial T > A and G > C preference at the third codon position. This is a common but not universal feature among plant WRKY genes. For instance, in *Helianthus annuus*, the PR2 plot *indica*tes a comparable AT > GC preference at the third codon, again dominated by mutation pressure, but some genes show evidence for selective effects (e.g., HaWRKY51, HaWRKY91, HaWRKY109)(27). In banana and other monocots, similar GC3-driven asymmetries predominate, often linked to mutational spectra, but with limited, locus-specific effects

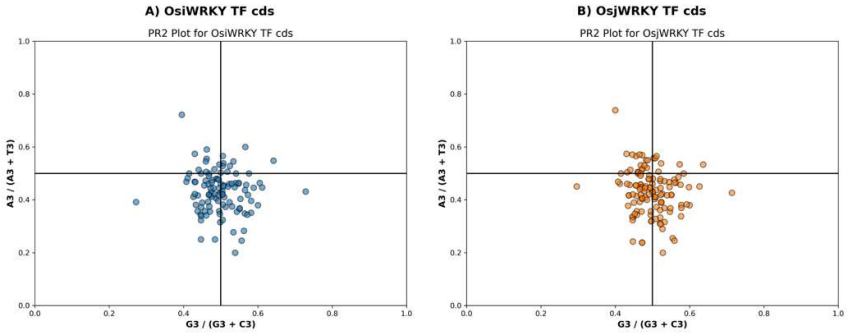

Fig 4: A) PR2 bias plot for *O. sativa* subsp. *indica*. Most genes fall below the 0.5 A/T axis and slightly left of the 0.5 G/C axis, *indica*ting T and C preference. B) PR2 bias plot for *O. sativa* subsp. *japonica*. A similar pattern of T3 and C3 dominance is observed, suggesting conserved strand bias.

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

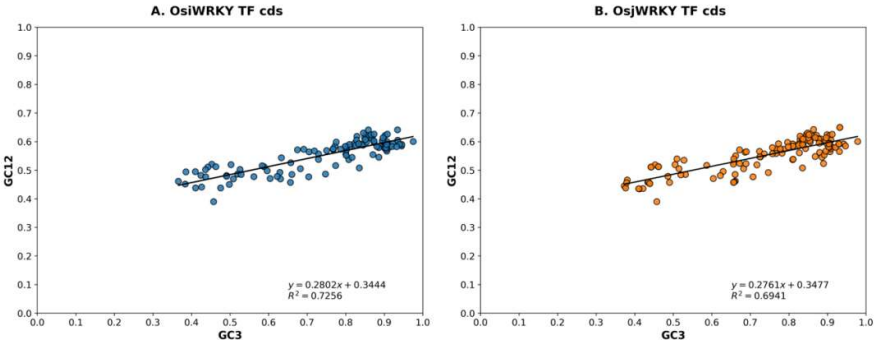

Fig 5: A) Neutrality plot for *O. sativa* subsp. *indica*. The slope of 0.2802 and strong correlation (r = 0.85) suggest codon usage is shaped by both selection and mutational bias. B) Neutrality plot for *O. sativa* subsp. *japonica*. Similar to *indica*, a slope of 0.2761 and r = 0.83 *indica*tes partial neutrality with significant selection influence.

of natural selection(30). However, in *Arabidopsis* and *Brassica*, PR2 analyses reveal that both natural selection and mutation contribute, with selection being more influential than in rice or banana(34).

Neutrality Plot Analysis

Neutrality plots were employed to investigate the relative contributions of mutational pressure and natural selection to codon usage bias in WRKY transcription factor genes of *Oryza sativa* subsp. *indica* and *japonica* (Fig. 5). In both subspecies, a moderate positive correlation was found between GC3 and GC12, with R^2 values of 0.7256 for *indica* and 0.6941 for *japonica*. The corresponding low regression slopes were 0.2802 and 0.2761, *indica*ting that natural selection plays a dominant role in shaping codon usage bias in rice WRKY genes. These findings are consistent with previous studies in monocot species such as *M. acuminata(30)* and dicots like *Arabidopsis thaliana* and *Brassica napus(34)*, where selection rather than mutational pressure governs codon preference, but in contrast with species like sunflower, where mutational pressure plays a larger role(27). The predominance of selection-driven codon usage in rice WRKY genes underscores their functional significance as key regulators of stress signaling pathways. Efficient translation, particularly under rapidly changing environmental conditions, may be essential for the prompt activation of stress-responsive transcriptional programs(35, 36).

***Correspondence Analysis* (COA)**

It was performed on the RSCU values of WRKY transcription factor genes in *Oryza sativa* subsp. *indica* and *japonica* to explore the main trends in codon usage variation (Fig. 6). In both subspecies, the first principal axis (Axis 1) accounted for a significant portion of the total variance, explaining 92.3% in *indica* and 91.6% in *japonica* that are among the highest reported for plant WRKY genes. These results significantly exceed variance explanations observed in *M. acuminata* (36.87–38.67%)(30) and *Phoenix dactylifera* (23.28%)(32), suggesting that compositional constraints, particularly GC3 content, overwhelmingly influence codon usage in rice. The second axis (Axis 2) captured 4.1% and 5.0% of the variation, respectively. The dominance of Axis 1 in explaining codon variation suggests that GC3 composition is the principal determinant of codon usage, with minimal contributions from secondary factors such as gene expression or protein function. These observations reinforce the interpretation that long-term

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

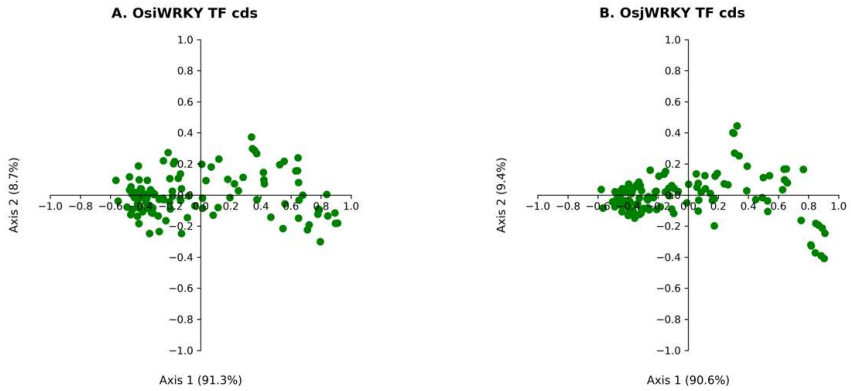

Fig 6: COA plots of WRKY codon usage in (A) *indica* and (B) *japonica*.

evolutionary pressures have strongly shaped the compositional landscape of rice WRKY genes, driving the selection of codons optimal for both translational efficiency and regulatory flexibility.

Conclusion

Analysis of WRKY transcription factors in rice *indica*tes that both mutational pressure and natural selection shape the codon usage of these factors. Mutational bias stands out as the primary influence, while natural selection plays a moderate role, particularly in refining codon preferences in stress-responsive genes. These insights deepen our understanding of the evolutionary constraints on key regulatory gene families and set the stage for future endeavors in gene expression optimization and synthetic biology within rice improvement programs.

Future Applications

The characterization of codon usage bias in rice WRKY genes has important implications for plant biotechnology and genetic engineering. The findings from this study may be leveraged in the following applications:

- Transgene Optimization: Identified optimal codons can be used to redesign WRKY gene constructs for improved expression in rice transformation systems.
- Synthetic Gene Design: Codon usage information can guide the synthesis of custom WRKY genes tailored for enhanced translation efficiency and expression stability.
- Cross-Species Engineering: Understanding species-specific codon usage preferences can inform gene transfer strategies, allowing adaptation of WRKY sequences for optimal expression in heterologous hosts.

These applications hold promise for enhancing stress tolerance in crops through precise modulation of transcription factor activity.

References

1. Parvathy, S.T., V. Udayasuriyan, and V. Bhadana, (2022). *Codon usage bias.* Molecular biology reports. 49(1): p. 539-565.
2. Supek, F., (2016). *The code of silence: widespread associations between synonymous codon biases and gene function.* Journal of molecular evolution. 82(1): p. 65-73.
3. Quax, T.E., N.J. Claassens, D. Söll, and J. van der Oost, (2015). *Codon bias as a means to fine-tune gene expression.* Molecular cell. 59(2): p. 149-161.
4. Liu, Y., Q. Yang, and F. Zhao, (2021). *Synonymous but not silent: the codon usage code for gene expression and protein folding.* Annual review of biochemistry. 90(1): p. 375-401.
5. Al-Hawash, A.B., X. Zhang, and F. Ma, (2017). *Strategies of codon optimization for high-level heterologous protein expression in*

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

microbial expression systems. Gene Reports. 9: p. 46-53.
6. Webster, G.R., A.Y.H. Teh, and J.K.C. Ma, (2017). *Synthetic gene design—The rationale for codon optimization and implications for molecular pharming in plants.* Biotechnology and bioengineering. 114(3): p. 492-502.
7. Wani, S.H., S. Anand, B. Singh, A. Bohra, and R. Joshi, (2021). *WRKY transcription factors and plant defense responses: latest discoveries and future prospects.* Plant Cell Reports. 40(7): p. 1071-1085.
8. Rinerson, C.I., R.C. Rabara, P. Tripathi, Q.J. Shen, and P.J. Rushton, (2015). *The evolution of WRKY transcription factors.* BMC plant biology. 15(1): p. 66.
9. Tripathi, P., R.C. Rabara, and P.J. Rushton, (2014). *A systems biology perspective on the role of WRKY transcription factors in drought responses in plants.* Planta. 239(2): p. 255-266.
10. Yang, L., S. Fang, L. Liu, L. Zhao, W. Chen, X. Li, Z. Xu, S. Chen, H. Wang, and D. Yu, (2025). *WRKY transcription factors: Hubs for regulating plant growth and stress responses.* Journal of Integrative Plant Biology. 67(3): p. 488-509.
11. Chakraborty, S., S. Yengkhom, and A. Uddin, (2020). *Analysis of codon usage bias of chloroplast genes in Oryza species: Codon usage of chloroplast genes in Oryza species.* Planta. 252(4): p. 67.
12. Wang, M., L. Huang, Y. Kou, D. Li, W. Hu, D. Fan, S. Cheng, Y. Yang, and Z. Zhang, (2023). *Differentiation of morphological traits and genome-wide expression patterns between rice subspecies indica and japonica.* Genes. 14(10): p. 1971.
13. Chin, J.H., Y.-J. Lee, W. Jiang, H.-J. Koh, and M.J. Thomson, (2017). *Characterization of indica–japonica subspecies-specific InDel loci in wild relatives of rice (Oryza sativa L. subsp. indica Kato and subsp. japonica Kato).* Genetic Resources and Crop Evolution. 64(2): p. 405-418.
14. Li, S., M.A. Khoso, H. Xu, C. Zhang, Z. Liu, S. Wagan, K. Dinislam, and L. Liu, (2024). *WRKY Transcription Factors (TFs) as Key Regulators of Plant Resilience to Environmental Stresses: Current Perspective.* Agronomy. 14(10): p. 2421.
15. De Oliveira, J.L., A.C. Morales, L.D. Hurst, A.O. Urrutia, C.R. Thompson, and J.B. Wolf, (2021). *Inferring adaptive codon preference to understand sources of selection shaping codon usage bias.* Molecular Biology and Evolution. 38(8): p. 3247-3266.
16. Sharp, P.M. and E. Cowe, (1991). *Synonymous codon usage in Saccharomyces cerevisiae.* Yeast. 7(7): p. 657-678.
17. Song, Y.-F., Q.-H. Yang, X.-G. Yi, Z.-Q. Zhu, X.-R. Wang, and M. Li, (2022). *Comparative analysis of codon usage patterns in chloroplast genomes of cherries.* Forests. 13(11): p. 1891.
18. Sharp, P.M. and W.-H. Li, (1987). *The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications.* Nucleic acids research. 15(3): p. 1281-1295.
19. Yengkhom, S., A. Uddin, and S. Chakraborty, (2019). *Deciphering codon usage patterns and evolutionary forces in chloroplast genes of Camellia sinensis var. assamica and Camellia sinensis var. sinensis in comparison to Camellia pubicosta.* Journal of Integrative Agriculture. 18(12): p. 2771-2785.
20. Wright, F., (1990). *The 'effective number of codons' used in a gene.* Gene. 87(1): p. 23-29.
21. Liu, X.-Y., Y. Li, K.-K. Ji, J. Zhu, P. Ling, T. Zhou, L.-Y. Fan, and S.-Q. Xie, (2020). *Genome-wide codon usage pattern analysis reveals the correlation between codon usage bias and gene expression in Cuscuta australis.* Genomics. 112(4): p. 2695-2702.
22. Liu, H., Y. Lu, B. Lan, and J. Xu, (2020). *Codon usage by chloroplast gene is bias in Hemiptelea davidii.* Journal of Genetics. 99(1): p. 8.
23. Wang, H.-C. and D.A. Hickey, (2007). *Rapid divergence of codon usage patterns within the rice genome.* BMC evolutionary biology. 7(Suppl 1): p. S6.
24. He, B., H. Dong, C. Jiang, F. Cao, S. Tao, and L.-a. Xu, (2016). *Analysis of codon*

Vol. 20 (Supplementry Issue 1A), March 2026, ISSN 0973-8916 (Print)., 2230-7303 (Online)
10.5530/ctbp.2026.1s.8

usage patterns in Ginkgo biloba reveals codon usage tendency from A/U-ending to G/C-ending. Scientific Reports. 6(1): p. 35927.
25. Sueoka, N., (1999). *Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+ C content of third codon position.* Gene. 238(1): p. 53-58.
26. Wang, Z., Q. Cai, Y. Wang, M. Li, C. Wang, Z. Wang, C. Jiao, C. Xu, H. Wang, and Z. Zhang, (2022). *Comparative analysis of codon bias in the chloroplast genomes of theaceae species.* Frontiers in Genetics. 13: p. 824610.
27. Gao, Y., Y. Lu, Y. Song, and L. Jing, (2022). *Analysis of codon usage bias of WRKY transcription factors in Helianthus annuus.* BMC Genomic Data. 23(1): p. 46.
28. Xie, Z., Z.-L. Zhang, X. Zou, J. Huang, P. Ruas, D. Thompson, and Q.J. Shen, (2005). *Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells.* Plant physiology. 137(1): p. 176-189.
29. Song, H., P. Wang, L. Hou, S. Zhao, C. Zhao, H. Xia, P. Li, Y. Zhang, X. Bian, and X. Wang, (2016). *Global analysis of WRKY genes and their response to dehydration and salt stress in soybean.* Frontiers in Plant Science. 7: p. 9.
30. Sun, J., J. Zhang, J. Zhang, B.J. Ferguson, and A. Chen, (2025). *Codon Usage Bias Analysis of the WRKY Gene Family in Musa acuminata.* bioRxiv: p. 2025.01. 08.631839.
31. Kawabe, A. and N.T. Miyashita, (2003). *Patterns of codon usage bias in three dicot and four monocot plant species.* Genes & genetic systems. 78(5): p. 343-352.
32. Mazumdar, P., R. Binti Othman, K. Mebus, N. Ramakrishnan, and J. Ann Harikrishna, (2017). *Codon usage and codon pair patterns in non-grass monocot genomes.* Annals of Botany. 120(6): p. 893-909.
33. Sinha, K., S. Jana, P. Pramanik, and B. Bera, (2024). *Selection on synonymous codon usage in soybean (Glycine max) WRKY genes.* Scientific reports. 14(1): p. 26530.
34. Srivastava, S., S. Chanyal, A. Dubey, A. Tewari, and G. Taj, (2019). *Patterns of codon usage bias in WRKY genes of Brassica rapa and Arabidopsis thaliana.* J Agric Sci. 11(4): p. 76.
35. Ramamoorthy, R., S.-Y. Jiang, N. Kumar, P.N. Venkatesh, and S. Ramachandran, (2008). *A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments.* Plant and cell physiology. 49(6): p. 865-879.
36. Wei, X., H. Zhou, D. Xie, J. Li, M. Yang, T. Chang, D. Wang, L. Hu, G. Xie, and J. Wang, (2021). *Genome-wide association study in rice revealed a novel gene in determining plant height and stem development, by encoding a WRKY transcription factor.* International Journal of Molecular Sciences. 22(15): p. 8192.