

1 **Title page**

2 **Codon Usage Bias Analysis of the *WRKY* Gene Family in *Musa acuminata***

3 **Jiaman Sun^{1,2†}, Ji Zhang^{2†}, Jinzhong Zhang¹, Brett J. Ferguson³, Andrew Chen^{4*}**

4 1. Guangdong Provincial Key Laboratory of Conservation and Precision Utilization of Characteristic
5 Agricultural Resources in Mountainous Areas, School of Life Science, Jiaying University, Meizhou,
6 Guangdong, 514015, China.

7 2. Guangxi Academy of Agricultural Sciences, Nanning 530007, China.

8 3. Integrative Legume Research Group, School of Agriculture and Food Sustainability, The University
9 of Queensland, St. Lucia, QLD 4072, Australia

10 4. School of Agriculture and Food Sustainability, The University of Queensland, St Lucia, Brisbane,
11 QLD 4072, Australia.

12 * Corresponding author, a.chen2@uq.edu.au

13 †These authors contributed equally to the work.

14

15

16

17

18

19

20

21 Abstract

22 Codon usage bias (CUB) is a universal phenomenon across all organisms and species, providing valuable
 23 insights into codon usage patterns to understand evolutionary mechanisms. The *WRKY* transcription factor
 24 family plays a pivotal role in regulating plant responses to various stresses at both physiological and
 25 biochemical levels. Bananas (*Musa* spp.), widely consumed as a fruit or staple food, are an important
 26 horticulture crop in the tropical and subtropical regions around the world. A total of 151 *WRKY* transcription
 27 factors were identified across the genome of the *Musa acuminata* banana cultivar 'Guijiao 9', which has been
 28 shown to be resistant to the Tropical Race 4 of the *Fusarium* wilt of banana. The codons of these
 29 transcription factors, selected based on their expression from RNA-Seq data, were analysed to investigate
 30 the evolutionary patterns of *WRKY* genes in *Musa acuminata*. The average GC content of *MaWRKY* genes
 31 was 55.49%, with a GC3 content of 62.13%, indicating a preference for G/C-ending codons. Among the
 32 codons, 27 were identified as high frequency, with 22 ending in G or C. The high effective number of
 33 codons (ENC) values (34.53–60.79) and low codon adaptation index (CAI) values (0.143–0.3) suggested
 34 weak CUB and that was associated with low gene expression levels. ENC-plot, PR2-plot, and neutrality
 35 analysis revealed that both natural selection and mutation pressure contributed to the observed CUB, with
 36 natural selection being the dominant factor influencing the codon usage of *MaWRKY* genes in *M. acuminata*
 37 'Guijiao 9'. Fourteen optimal codons, all ending in G or C, were identified. This analysis provides a
 38 theoretical foundation for further understanding the evolutionary mechanisms of *WRKY* genes and
 39 identifying key genes involved in resistance to biotic stress in *Musa*. Analysing the evolution of *WRKY*
 40 family genes can offer valuable genetic insights into banana growth and development, supporting the
 41 phenotypic selection of banana hybrids in breeding programs.

42

43 **Keywords:** *Musa acuminata*, 'Guijiao 9', *WRKY* transcription factors, codon usage bias, natural selection,
 44 mutation pressure, *Fusarium* wilt of banana, *Fusarium oxysporum* f. sp. *cubense*, Tropical Race 4.

45 Introduction

46 A codon is a triad nucleotide sequence that encodes amino acids on messenger RNA. Codons that encode the
 47 same amino acid are referred to as synonymous codons. Over the course of evolution, certain species and
 48 specific genes tend to favour particular codons for encoding amino acids, leading to synonymous codon
 49 usage bias (Kumar et al., 2016). This bias influences gene expression levels, with stronger codon usage bias
 50 generally correlating with higher gene expression (Jiang et al., 2008). Low-frequency codons help maintain
 51 codon diversity and contribute to gene expression regulation, playing a critical role in species stability and
 52 evolution of each species (Liu et al., 2010). Three primary forces shaping codon usage bias have been
 53 identified: mutation pressure, natural selection and random genetic drift (Behura and Severson, 2013).
 54 Mutation pressure causes deviations in nucleotide composition, while natural selection restricts codon usage
 55 bias to optimise protein production efficiency in highly expressed genes (Guan et al., 2018). Additional
 56 factors influencing codon usage bias include nucleic acid composition (Mondal et al., 2016), gene
 57 expression level (Hambuch and Parsch, 2005), gene length (Wei et al., 2014), and tRNA abundance (Deng et
 58 al., 2020). Analysis of codon usage bias is crucial for understanding the evolutionary mechanisms and codon
 59 usage patterns of specific plants. Synonymous codon usage bias differs between monocotyledons and
 60 dicotyledons. Monocotyledons tend to favour codons ending in G/C, while dicotyledons prefer A/T endings
 61 (Murray et al., 1989). GC content also serves as a distinguishing factor between these two groups. Related
 62 species often exhibit similar codon usage biases, making it an important tool for inferring evolutionary
 63 relationships among different plant species (Ma et al., 2015).

64
 65 Codon usage bias reflects the origins, evolution, and mutations of species or genes, providing essential
 66 insights for gene function analysis, protein expression and protein structure research (Wu et al., 2015; He et
 67 al., 2016). In recent years, studies on codon preference in gene families and individual genes across plants
 68 (Liu et al., 2020), animals (Uddin et al., 2020), viruses (Deb et al., 2020; Xin et al., 2020) and other
 69 organisms has been increasingly widespread. WRKY transcription factor family is one of the largest
 70 transcription factor families in plants, playing a key role in physiological and biochemical processes such as
 71 hormone synthesis, signal transduction, plant growth and development, metabolism regulation and
 72 biological responses to both biotic and abiotic stresses (Govardhana and Satyan, 2020; Li et al., 2020). The
 73 widespread use of next-generation sequencing technologies for whole genome and transcriptome sequencing,
 74 along with the availability of bioinformatics tools have greatly advanced in-depth studies of WRKY
 75 transcription factor family genes and their functional roles in plant's ability to adapt to environmental
 76 challenges (Wani et al., 2021; Goyal et al., 2023; Liu et al., 2023; Deng et al., 2024). Banana (*Musa*. spp) is
 77 an important fruit crop, which supplies essential nutrients to millions of people around the world (Drenth
 78 and Kema, 2021). Cultivated bananas originated from natural intra- and interspecific hybridisation between
 79 two diploid species, *Musa acuminata* (AA genome) and *Musa balbisiana* (BB genome) (Drenth and Kema,
 80 2021). However, due to various biotic and abiotic stresses, global banana yields have significantly declined
 81 in recent years (Siamak and Zheng, 2018; Patel et al., 2019).

82
 83 One of the major diseases threatening the banana production dominated by 'Cavendish' from the AAA
 84 subgroup is the Fusarium wilt of banana, caused by the *Fusarium oxysporum* f. sp. *cubense* tropical race 4
 85 (Foc TR4) (Zorrilla-Fontanesi et al., 2020; Roberts et al., 2024). Current efforts have been focused on
 86 containment and deterrence through biosecurity measures; biological control agents; integrated disease
 87 management strategies; and the development of resistant banana varieties (Swarupa et al., 2014; Ploetz,

2015; Dita et al., 2018; Pegg et al., 2019). Resistant (R) genes, such as Resistance Gene Analog 2 (RGA2), which was isolated from *Foc TR4*-resistant *Musa acuminata* ssp. *malaccensis*, demonstrated strong resistance in the field when over-expressed in 'Cavendish' transgenic lines (Dale et al., 2017). Forward genetic studies have also identified QTLs conferring resistance to *Foc TR4* and subtropical race 4 (STR4) (Ahmad et al., 2020; Chen et al., 2023a, 2023b). Candidate R genes, specifically pattern recognition receptors, were identified.

Several *WRKY* genes, such as *MaWRKY28*, *MaWRKY71*, *MaWRKY40* and *MaWRKY22*, have been associated with banana's response to biotic stress, including *Foc TR4* (Sun et al., 2019; Li et al., 2012). However, research on the codon usage pattern of *WRKY* genes in banana has not been reported. In our previous study, *Musa. acuminata* banana 'Guijiao 9' exhibited strong resistance to *Foc-TR4*, with lower disease incidence and severity compared to susceptible banana cultivars (Sun et al., 2019). In this study, 151 members of *WRKY* gene family were identified from *M. acuminata* 'Guijiao 9' transcriptome data. Factors influencing codon preference of *MaWRKY* and potential evolutionary models were determined by analysing the CUB of *MaWRKY* genes in 'Guijiao 9'. These findings contribute to understanding the function of the *WRKY* gene family and provide insights for codon optimisation of *MaWRKY* in the process of biotic stress.

Results

Codon composition of *MaWRKY* genes reveals a G/C bias

The codon composition and ENC values of 151 banana *MaWRKY* family members were statistically analysed, with detailed codon bias parameters provided in Additional File 2. The ENC values ranged from 34.53 to 60.79, indicating varying levels of codon bias, although the overall CUB was weak. CAI values for *MaWRKY* genes ranged from 0.143 to 0.3, suggesting low gene expression levels.

The average content of A3s and T3s was not significantly different from G3s and C3s ($P > 0.05$) (Fig. 1). However, the combined average content of G3s+C3s differed significantly from that of A3s+T3s. The average GC content was 55.49%, ranging from 51.05% to 58.95%, indicating a relatively high GC content and a preference for G/C-ending codons in *MaWRKY* genes.

Analysis of G/C content at different codon positions showed that GC3s (62.13% on average) were much higher than GC1s (54.65%) and GC2s (48.24%), suggesting preference for G or C at the third codon position. GC3s exhibited greater variability (ranging from 54.20% to 69.40%) compared to GC1s (ranging from 52.24% to 57.93%) and GC2s (ranging from 45.65% to 50.75%). This variability indicates that the first and second codon positions are relatively stable, while the third position shows greater fluctuation, potentially contributing to the observed codon usage bias.

Codon Usage Parameter Correlation Analysis

The Pearson Correlation coefficients for various codon usage indices were calculated and are presented in Table 1 to assess the relationships between factors influencing codon usage. ENC value showed a significant positive correlation with A3s ($r = 0.726$, $p < 0.05$) and T3s ($r = 0.690$, $p < 0.05$) but was negatively correlated with G3s ($r = -0.616$, $p < 0.05$), C3s ($r = -0.703$, $p < 0.05$), GC ($r = -0.704$, $p < 0.05$) and GC3s (r

125 = -0.750, $p < 0.05$). These findings indicate that the nucleotide composition at the third position of
126 synonymous codons strongly influences codon usage bias.

127 Additionally, G3s was positively correlated with C3s ($r = 0.615$, $p < 0.05$), GC3s ($r = 0.844$, $p < 0.01$) and
128 GC ($r = 0.695$, $p < 0.01$). Conversely, A3s exhibited a negative correlation with G3s ($r = -0.801$, $p < 0.01$),
129 C3s ($r = -0.842$, $p < 0.01$), GC3s ($r = -0.938$, $p < 0.01$) and GC ($r = -0.938$, $p < 0.01$). Similarly, T3s showed
130 a negative correlation with G3s ($r = -0.776$, $p < 0.05$), C3s ($r = -0.913$, $p < 0.01$), GC3s ($r = -0.959$, $p < 0.01$),
131 and GC ($r = -0.881$, $p < 0.01$). These negative correlations suggest that the frequency of A or T at the third
132 codon position has little influence on the codon usage bias of *MaWRKY* genes.

133 **Principal Component Analysis**

134 Principal Component Analysis (PCA) was performed on the RSCU values of *MaWRKY* genes sequences.
135 The contribution of the first 20 factors to the variance in codon usage bias is shown in Fig. 2A. The first four
136 factors accounted for 58.47% of the total variance, capturing the majority of the differences in codon usage
137 patterns.

138 PCA analysis of the 59 codons in *MaWRKY* family members, categorised by their endings (A, T, C or G),
139 revealed distinct distribution patterns (Fig. 2B). Codons ending in A and T were clustered on the left side of
140 axis 1, and those ending in G and C were primarily located on the right side. This suggests similar RSCU
141 usage patterns for codons ending in A or T and for those ending in G or C. The distribution of codons ending
142 in A and T was denser and showed partial overlap, whereas codons ending in G and C were more dispersed,
143 with those ending in C being spread across all four quadrants. These findings indicate a natural
144 evolutionary or long-time domestication bias in banana codon usage toward G and C endings.

145 **Factors affecting codon usage bias in *MaWRKY* genes of *Musa acuminata***

146 ENC-plot analysis commonly used to assess the effect of mutation pressure on codon usage bias. The ENC
147 values are plotted against GC3 values (Fig. 3A). The standard curve shows that the relationship between
148 ENC and GC3s is shaped primarily by mutation pressure rather than selection. The ENC values for
149 *MaWRKY* family members generally align with the standard curve in the ENC-plot, indicating that mutation
150 pressure is a major factor influencing codon usage bias for these genes (Fig. 3A). However, the ENC values
151 of many *MaWRKY* genes deviate significantly from the standard curve (Fig. 3A), suggesting that mutation
152 pressure is not the sole factor driving codon bias. Other factors, such as natural selection and gene
153 expression, may also contribute to the observed codon bias.

154
155 PR2 analysis was performed to assess the impact of mutation and selection pressure on codon usage by
156 examining whether there was a mutation imbalance between A/T (U) and C/G. In the PR2 plot, $A3 / (A3 +$
157 $T3)$ is plotted as the ordinate and $G3 / (G3 + C3)$ as the abscissa for *MaWRKY* family members to explore
158 the influence of evolutionary factors (Fig. 3B). The $A3 / (A3 + T3)$ or $G3 / (G3 + C3)$ values for most
159 *MaWRKY* genes deviated from 0.5, suggesting that codon bias was additionally influenced by factors other
160 than base mutation of encoding gene amino acids (Fig. 3B). This indicates that additional pressures, such as
161 natural selection, likely played a role in the evolution of *MaWRKY* genes.

162
163 To further identify the factors influencing codon usage bias, a neutrality plot analysis was performed by
164 comparing GC3 (abscissa) and GC12 (ordinate) to examine the role of mutation-selection equilibrium in

165 codon usage variation. A linear regression line was plotted with GC3 (abscissa) and GC12 (ordinate) values
166 (Fig. 3C). A regression coefficient (RC) value of less than 0.5 suggests a greater influence of natural
167 selection, while an RC greater than 0.5 indicates a greater impact of mutation pressure. The RC value of
168 0.1718 indicates that the natural selection had a more significant effect on the codon usage bias of *MaWRKY*
169 gene family in banana than mutation pressure (Fig. 3C).
170

171 **Relative synonymous codon usage (RSCU) in *MaWRKY* gene family**

172 To investigate the patterns of synonymous codon usage bias in *MaWRKY* gene family, the RSCU of 59
173 synonymous codons (excluding the initiation codon, termination codon and tryptophan) was analysed. The
174 results revealed 27 high-frequency codons (RSCU > 1) in *MaWRKY* family members, including AGG, AGA,
175 CUC, CUG, UUG, AGC, UCC, UGG, GUG, GUC, ACC, ACG, GGC, GCC, GCU, CCG, CCA, AUC, UGC,
176 AAG, UUC, AAC, GAG, UAC, CAG, CAC, and GAU, respectively (Fig. 4). Among these, 22 codons
177 ended in G/C, indicating a preference for G or C at the end of the codons in *MaWRKY* members. The codon
178 AGG of Arginine (RSCU = 1.800) exhibited the strongest preference, accounting for 30% of the
179 synonymous codons, while the codon AUC for Isoleucine accounted for 59.0%. These high-frequency
180 codons show preferential usage, which contributed to the deviation of the ENC value of *MaWRKY* genes
181 from an ENC value of 61.

182 **The optimal codons of *MaWRKY* Genes**

183 The optimal codons in *MaWRKY* genes of *M. acuminata* 'Guijiao 9' were identified by combining
184 high-frequency codons (RSCU > 1) and highly expressed codon ($\Delta\text{RSCU} \geq 0.08$). The analysis revealed 26
185 highly expressed codons with $\Delta\text{RSCU} \geq 0.08$ (Table 2) and 27 high-frequency codons with RSCU > 1 (Fig.
186 4). In total, 14 optimal codons were determined for *MaWRKY* genes in *M. acuminata* 'Guijiao 9', including
187 CUC, CUG, AUC, GUG, AGC, CCG, ACC, UAC, CAC, CAG, AAC, GAG, UUC and GGC. Notably, all of
188 these optimal codons end with G or C (Table 2).

189 **Codon usage patterns of WRKY Family Gene members across different plant species**

190 Using the Euclidean distance, the RSCU values of 25 WRKY gene family members across different plant
191 species were clustered with SPSS software v22.0 (Fig. 5). The clustering results of WRKY genes in these
192 plant species were consistent with their phylogenetic tree, as shown in Phytozome v12.1
193 (<https://phytozome.jgi.doe.gov/>). This finding suggests a correlation between species classification and
194 codon usage patterns. It also indicates that the codon usage patterns of WRKY genes in different species align
195 with the evolutionary processes of their respective genomes.
196

197 A negative correlation was observed between GC3s content and ENC value when comparing the GC3s and
198 ENC values of WRKY gene family members from different plant species (Fig. 5). The variation in GC3s
199 content and ENC values differed among different plants categories. For example, in monocotyledons, the
200 GC3s content ranged from 60.03% to 75.88%, and the ENC value ranged from 42.67 to 52.20, while in
201 dicotyledons, the GC3s content varied between 37.77% and 52.98%, and the ENC value ranged from 51.85
202 to 54.26. Plant species that are closely related species showed similar GC3s content and ENC value. 'Guijiao
203 9' and another *Musa acuminata* representative were clustered adjacently due to their similar ENC and GC3

values, as both belong to the monocotyledon group. The ENC value of 'Guijiao 9' and *Musa acuminata* is most similar to that of tree species like *Populus trichocarpa*, *Manihot esculenta*, and *Citrus sinensis*, but significantly different from gramineous species such as *Oryza sativa*, *Zea mays* PH207, and *Setaria viridis*.

A heat map analysis of the 59 RSCU values (excluding the initiation codon, termination codon and tryptophan) for 25 plant species revealed that the codons were divided into two groups: those ending with G/C and those ending with A/T. Among the 25 plant species, Algae, Pteridophytes and monocotyledons favoured codons ending in G/C, while dicotyledons preferred those ending in A/T (Fig. 6).

Discussion

During the process of gene expression, the selective use of codons varies from species to species. Different organisms exhibit different preference for synonymous codons encoding the same amino acid, closely linked to their genetic characteristics. CUB can influence mRNA stability, transcription, protein translation accuracy, and protein folding, thereby fine-tuning gene expression.

In this study, 151 *MaWRKY* genes were identified in *M. acuminata* using bioinformatics tools for codon analysis. GC composition has been shown to influence codon and amino acid usage, with the third base of a codon could directly reflecting codon usage patterns (Chen et al., 2013; Gao et al., 2022). The average GC content and GC3 content of *MaWRKY* codons were 55.49% and 62.13% respectively, indicating a preference for G/C-ended codons over A/U-ended codons in *M. acuminata*. This preference was further supported by RSCU analysis, which confirmed *MaWRKY* genes favour G/C-ending codons. These results were consistent with the codon usage patterns observed in other monocotyledons, such as in *Musa basjoo*, *Zea mays* and *oryza sativa* (Ma et al., 2015).

The ENC value and CAI value provide insights into CUB and gene expression levels. Higher ENC values suggest lower CUB and gene expression, while higher CAI values indicate better codon adaptiveness and potentially higher gene expression levels. In this study, the *MaWRKY* genes exhibited high ENC values (34.53-60.79) and low CAI values (0.143-0.3), indicating weak CUB and low gene expression levels of these genes in 'Guijiao 9'. This was consistent with studies showing that most *WRKY* family genes display stress-induced expression patterns (Dong et al., 2003; Kayum et al., 2015). Furthermore, the GC3 content showed greater variability than GC1 and GC2 with codons ending in G/C exhibiting a broad distribution. These findings suggest that nucleotide composition plays a significant role in influencing CUB in *MaWRKY* genes. Mutation pressure and natural selection are recognised as important factors influencing codon usage bias (Sharp et al., 2010; He et al., 2016). In this study, ENC-Plot analysis showed that most ENC values for *MaWRKY* codons did not align with the expected curve, with a few codons exhibiting strong bias near an ENC value of 35. This finding indicates that mutation pressure is not the primary factor driving codon preference. Instead, other factors, such as natural selection and gene expression, likely play significant roles.

PR2 plot analysis further supported this conclusion, as the $A3 / (A3 + T3)$ and $G3 / (G3 + C3)$ values for most *MaWRKY* members deviated from 0.5. This deviation suggests that codon bias in *MaWRKY* genes is influenced more by pressures like natural selection than by mutation. Neutrality plot analysis provided additional evidence, with a regression coefficient of 0.17, highlighting that natural selection exerts a stronger influence on codon bias than mutation pressure.

246
247 Base on ENC, PR2 and neutrality plot analyses, the CUB of *MaWRKY* genes in *M. acuminata* is influenced
248 by multiple factors, with natural selection being the dominant force. These findings align with the study by
249 He et al. (2016), which reported that Ginkgo biloba unique genes involved in environment adaptation
250 preferentially using G/C-ending codons, with natural selection as the primary driver of codon usage bias.
251 Interestingly, while monocotyledons like *M. acuminata* favour G/C-ending codons, dicotyledons such as
252 *Helianthus annuus* (Gao et al., 2022) and *Brassica napus* (Li et al., 2013) exhibit a preference for A/T
253 (U)-ending codons. In these species, mutation pressure appears to play a larger role in shaping codon usage
254 bias (Gao et al., 2022; Li et al., 2013). Thus, the codon usage preferences of *WRKY* gene families vary
255 significantly among different species, reflecting distinct evolutionary pressures and adaptations. An optimal
256 codon is defined as one that is over-represented in highly expressed genes and is associated with higher gene
257 expression levels and efficient translation (Zhou et al., 2009). It was generally believed that GC-rich
258 organisms favour codons ending with G or C, while AT-rich organisms prefer codons ending with A or T
259 (Hershberg et al., 2009; Palidwor et al., 2015). In our study, all 14 identified optimal codons of *MaWRKY*
260 genes ended with G or C (Table 2), aligning with the GC-rich composition of *MaWRKY* genes in *M.*
261 *acuminata*. Similar patterns have been observed in other monocotyledon, such as *Oryza sativa* (Liu et al.,
262 2004) and *Zea mays* (Liu et al., 2010), whose genomes are GC-rich, and whose optimal codons also end
263 with G or C. This finding has practical implications for introducing point mutations and codon modifications
264 to enhance the production of specific proteins in host cells. Codon optimisation based on these preferences
265 can improve gene expression efficiency in biotechnological and agriculture applications.

266
267 RSCU values serve as a significant indicator of evolutionary relationships among different organisms and are
268 widely used in studies of biological evolution in plants. Cluster analysis of RSCU values for *WRKY* genes
269 across various plant species revealed that the codon usage patterns of *WRKY* gene families align with the
270 codon usage preferences of their respective genomes during evolution (Fig. 5). *MaWRKY* genes clustered
271 with other monocotyledon species in a single branch, exhibiting a preference for codons ending with G/C
272 (Fig. 6). This suggests that the codon usage patterns of *WRKY* family genes are consistent among related
273 species, reflecting evolutionary.

274 **Conclusions**

275 In the present study, 151 coding sequences of *WRKY* genes from *M. acuminata* 'Guijiao 9' were analysed to
276 investigate codon usage bias and its potential influencing factors. The results revealed that *MaWRKY* genes
277 in *M. acuminata* 'Guijiao 9' exhibit a preference for codons ending in G/C. Overall, the codon usage bias
278 was weak, and the expression levels of these genes were low, indicating high variability in synonymous
279 codon usage among *MaWRKY* genes. While mutation pressure played a role, natural selection emerged as
280 the primary factor in shaping the codon usage patterns of *MaWRKY* genes during evolution. The codon
281 usage patterns of *MaWRKY* genes provide valuable insights into their evolutionary dynamics and offer a
282 foundation for identifying and developing suitable gene targets with resistance to biotic stress.

283 **Materials and Methods**

284 **Screening, Identification and Characterization of Sequence**

285 The complete genome of *M. acuminata* double-haploid 'Pahang' was downloaded in FASTA format from the
 286 National Centre for Biotechnology Information (NCBI) banana genome database
 287 (<https://www.ncbi.nlm.nih.gov/datasets/taxonomy/214687/>, GenBank assembly accession:
 288 GCA_904845865.1), and was used as the reference genome. The Coding Domain Sequences (CDS) of the
 289 *WRKY* gene family were retrieved from the transcriptome data of *Musa acuminata* banana 'Guijiao 9'
 290 (Additional File 1). Corresponding amino acid and nucleotide sequences were obtained through local
 291 BLAST. HMMER 3.3 (<http://hmmer.org/download.html>) was used to identify sequences containing
 292 complete *WRKY* domain using the reference sequence (PF03106) obtained from PFAM
 293 (<http://pfam.sanger.ac.uk/>). The presence of the *WRKY* domain in these sequences was further confirmed
 294 using NCBI's CDD tool (<https://www.ncbi.nlm.nih.gov/cdd>). ORFfinder was then used to identify CDS from
 295 the retained sequences (<https://www.ncbi.nlm.nih.gov/orffinder/>). Ultimately, a total of 151 *M. acuminata*
 296 *WRKY* genes (*MaWRKY*) were identified and included in the subsequent analysis (Additional File 2).
 297 The CDS sequences of the *WRKY* gene family from 24 different species, including Bryophyte, Algae,
 298 Pteridophyte, monocotyledon and dicotyledon, were downloaded from the JGI Phytozome12.1 database
 299 (<https://phytozome.jgi.doe.gov/pz/portal.html>) and utilised for Cluster analysis.

300 **Indices of Codon Usage Bias**

301 Following the screening of *MaWRKY* gene CDS sequences, the key indices of codon usage bias were
 302 calculated, including the effective number of codons (ENC), codon adaptation index (CAI), relative
 303 synonymous codon usage (RSCU), total GC content of each CDS (GC), and GC content at the first, second
 304 and third codon positions (GC1, GC2, GC3). Additionally, the nucleotide composition at synonymous third
 305 codon positions (A3s, T3s, G3s, C3s) was analysed. These calculations were performed using the CodonW
 306 program (version 1.4.2). Correlations between nucleotide contents were assessed using SPSS statistical
 307 software (version 23.0), with values exceeding 0.5 and below -0.5 indicating strong positive and strong
 308 negative correlations, respectively.

309
 310 The Effective Number of Codons (ENC) is a key metric for evaluating the degree of codon usage bias. ENC
 311 values range from 20, representing extreme bias where only one codon is used for each amino acid, to 61,
 312 indicating no bias where all synonymous codons are equally utilised. Higher ENC values suggest weaker
 313 codon usage bias and lower expression levels (Sharp and Li, 1986; Wright, 1990; Zhao et al., 2016).
 314 Typically, an ENC value above 40 indicates weak CUB (Wright, 1990). The Codon Adaptation Index (CAI)
 315 ranges from 0 to 1, reflecting how well a gene's codon usage aligns with those of highly expressed genes
 316 (Sharp and Li, 1987). Higher CAI values signify better adaptiveness and higher potential expression
 317 efficiency. Relative Synonymous Codon Usage (RSCU) measures the observed frequency of a codon
 318 relative to its expected frequency, independent of the amino acid composition of gene product (Wang et al.,
 319 2016). An RSCU value of 1 signifies no bias, meaning the codon is used as frequently as expected. Values
 320 greater than 1 denote a preference for the codon (positive bias), while values below 1 indicate under
 321 utilisation (negative bias) (Zhao et al., 2016).

322

323 ***ENC-plot Analysis***

324 ENC-plot analysis is commonly used to identify factors influencing codon usage bias by plotting ENC
325 values against GC3s values. The standard curve represents the relationship between ENC and GC3s, helping
326 to distinguish whether mutation pressure or natural selection is the dominant factor (Jiang et al., 2008; Liu et
327 al., 2019). When codon usage bias is primarily driven by mutation pressure, the points tend to lie on or just
328 below the standard curve. In contrast, if natural selection and other factors play a larger role, the points will
329 fall below the standard curve (Wright, 1990).

330 ***Parity Rule 2 (PR2) Bias Plot Analysis***

331 A PR2 bias plot is created by plotting the AT bias ($A3 / (A3 + T3)$) on the y-axis and the GC bias ($G3 / (G3$
332 $+ C3)$) on the x-axis. The midpoint at 0.5 indicates an equal balance between $G = C$ and $A = T$, suggesting
333 no significant bias from either mutation or selection pressure (Deb et al., 2020). If the genes are clustered
334 near the centre, it suggests that the base frequencies are relatively balanced, and the codon bias is mainly
335 affected by mutation pressure. Conversely, If the genes are far from the centre, other factors may be
336 affecting the codon usage bias.

337 ***Neutrality Plot Analysis***

338 A neutrality plot is used to assess the extent to which codon usage bias is influenced by mutation versus
339 selection in organisms, by comparing GC3 (x-axis) and GC12 (y-axis). In this plot, each gene is represented
340 by a point. If the regression coefficient of the plot approaches 1, the points will show a clear pattern,
341 indicating selection plays a significant role in codon usage bias. If the regression coefficient deviates from 1,
342 it suggests that factors other than mutation pressure are also influencing the codon usage bias (Sueoka,
343 1988).

345 ***Relative Synonymous Codon Usage (RSCU) Cluster Analysis***

346 RSCU cluster analysis was employed to group similar elements. Euclidean distance, a common method for
347 calculating clustering distance, measures the linear correlation between two variables. A smaller clustering
348 distance indicates higher similarity between the clustered genes. The RSCU values of the *MaWRKY* gene
349 family members in 'Guijiao 9' were compared with those from 24 other plant species to reveal the
350 evolutionary relationships and characteristics among them.

351 ***Principal Component Analysis (PCA) of RSCU***

352 Principal Component Analysis (PCA) is a multivariate statistical method to examine relationships among
353 multiple variables, and is often applied to analyse trends in synonymous codon usage patterns (Kanaya et al.,
354 1996). In this study, the RSCU values for 59 synonymous codons from *MaWRKY* genes was reduced from
355 151 dimensions (representing 151 *MaWRKY* genes) to two principal components through dimensionality
356 reduction. Fifty-nine synonymous codons did not include the initiation codon AUG, tryptophan codon UGG,
357 and the termination codons UAA, UAG, and UGA (Zhang et al., 2018).

358 ***Optimal Codon Analysis***

359 The optimal codon plays a significant role in improving the speed and accuracy of translation (Duret et al.,
360 1999). To predict the optimal codon for *MaWRKY*, the Δ RSCU method was applied (Qiu et al., 2011). The
361 *MaWRKY* gene sequences were sorted based on their ENC values. The top 10% and the bottom 10% of the

genes, based on ENC value, were selected to create two new databases (high expression and low expression, respectively), and the RSCU values for the codons were calculated. A codon was considered a high-frequency codon if its RSCU value was greater than 1, while a codon was regarded as a high-expression codon if the Δ RSCU value was equal to or greater than 0.08. A codon is defined as optimal if it met both criteria: $RSCU > 1$ and $\Delta RSCU \geq 0.08$.

Statistical analysis

The CDS sequences of the *MaWRKY* were analysed using CodonW (<http://codonw.sourceforge.net/>) and EMBOSS online software (<http://www.bioinformatics.nl/emboss-explorer/>). Codon-related parameters were statistically analysed using Microsoft Office Excel 2016 and MeV software version 4.9.0 (www.tm4.org/mev.html) was used to construct the heat map. ENC-plot analysis, Neutrality plot, and Box diagram were generated using Origin 8.0 and HemI 1.0 (<http://hemi.biocuckoo.org/down.php>) software. Clustering analysis and principal component analysis (PCA) of RSCU values for *WRKY* genes in different plant species were performed by Statistical Product and Service Solutions (SPSS) version 22.0.

Abbreviations

CUB	Codon usage bias
ENC	Effective number of codons
CAI	Codon adaptation index
RSCU	Relative synonymous codon usage
PCA	Principal component analysis
PR2	Parity Rule 2
GC1, GC2, GC3	The GC content at the first, second and third position, respectively
A3s, T3s, G3s, C3s	The content of each nucleotide of the codon at synonymous third positions

Supplementary Information

Additional file 1. The RNA-Seq data of *Musa acuminata* banana 'Guijiao 9'

Additional file 2. The composition indices values of codon usage in *MaWRKY* genes

Authors' contributions

Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - original draft, and Writing - review & editing: J.S.; Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing - review & editing: J.Z.; Formal Analysis, Investigation, Supervision, and Writing - review & editing: Jinzhong Z; Supervision, and Writing - review & editing: B.J.F.; Formal Analysis, Investigation, Supervision, Writing - original draft, and Writing - review & editing: A.C. All authors have read and approved the final manuscript.

Funding

This work was supported by Guangxi Natural Science Foundation (2021GXNSFAA196014) and Guangxi Science and Technology Planning Project (Guike AA21196005). A.C. was supported by The Bill and Melinda Gates Foundation (Project Grant ID: OPP1093845) through its grant to the International Institute of Tropical Agriculture (IITA) under the project Accelerated Breeding of Better Bananas, grant number IITA

20600.15/0008-8—Phase II as well as Hort Innovation Australia through grant ‘BA21000’, using the banana research and development levy and contributions from the Australian Government. Hort Innovation is the grower-owned, not-for-profit research and development corporation for Australian horticulture.

Availability of data and materials

The genome datasets generated and analysed in the study are available in NCBI, <https://www.ncbi.nlm.nih.gov/datasets/taxonomy/214687/>, and supplementary files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Declaration of Interest Statement

The authors declare no competing interests.

Table 1. Pearson correlation coefficients among parameters influencing codon bias in *MaWRKY* genes of *Musa acuminata* banana 'Guijiao 9'. A3s, T3s, G3s and C3s denote the nucleotide content at the third codon position. GC represents the overall GC content of each coding domain sequence, while GC1, GC2, and GC3 refer to the GC content at the first, second, and third codon positions, respectively. ENC indicates Effective Number of Codons.

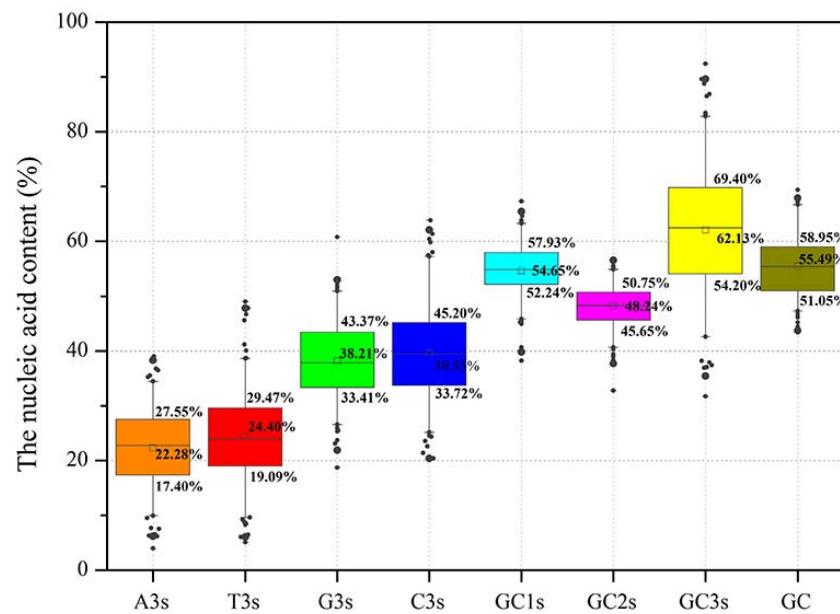
Indices	A3s	T3s	G3s	C3s	GC	GC1	GC2	GC3s	ENC
A3s	1.000								
T3s	.809**	1.000							
G3s	-0.801**	-0.776*	1.000						
C3s	-0.842**	-0.913**	0.615*	1.000					
GC	-0.896**	-0.881**	0.695*	0.840**	1.000				
GC1s	-0.479	-0.373	0.246	0.356	0.705*	1.000			
GC2s	-0.443	-0.441	0.154	0.370	0.681*	0.572	1.000		
GC3s	-0.938**	-0.959**	0.844**	0.929**	0.916**	0.417	0.417	1.000	
ENC	0.726*	0.690*	-0.616*	-0.703*	-0.704*	-0.357	-0.317	-0.750*	1.000

* and ** indicate p values < 0.05 and 0.01 , respectively.

Table 2. Identification of optimal codons in MaWRKY Genes of *M. acuminata* 'Guijiao 9'.

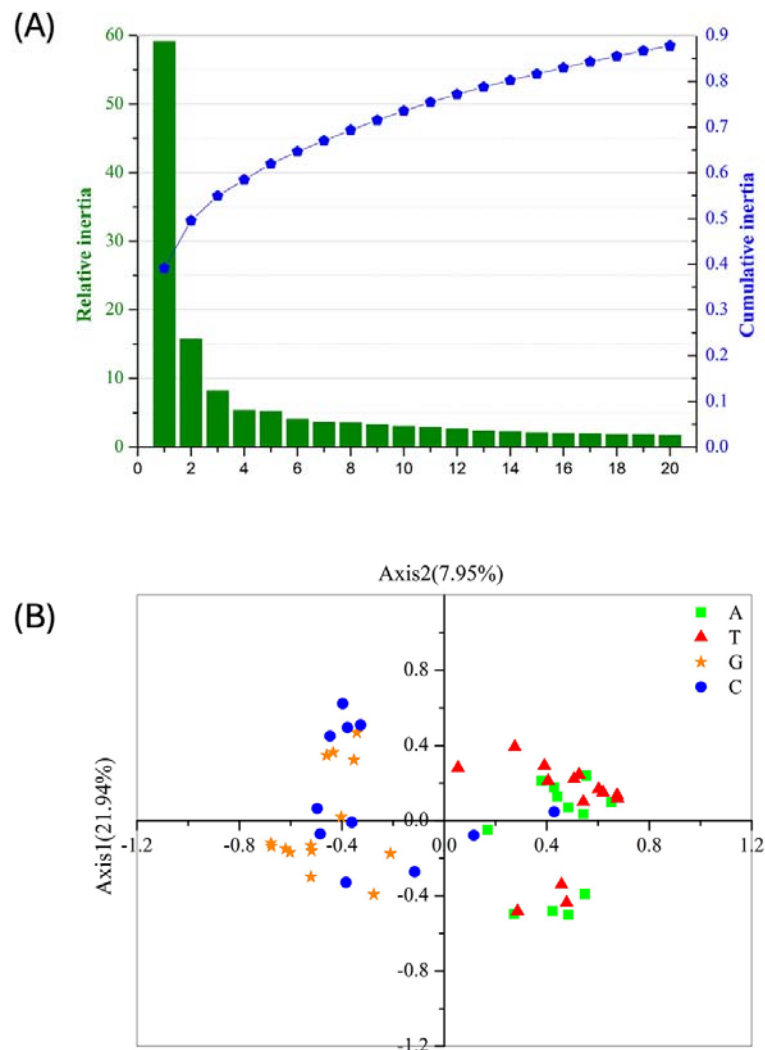
AA	Codon	RSCU		Δ RSCU	AA	Codon	RSCU		Δ RSCU
		High expression	Low expression				High expression	Low expression	
Leu (L)	UUA	0.37	0.40	-0.03	Tyr (Y)	UAU	0.48	0.73	-0.26
	<u>UUG</u>	0.48	1.63	-1.14		UAC *	1.52	1.27	0.26
	CUU*	0.73	0.69	0.04	TER	UAA*	0.38	0.00	0.38
	CUC *	2.13	1.10	1.03		UAG	0.75	2.25	-1.50
	CUA	0.37	0.53	-0.16		UGA*	2.25	0.75	1.50
	CUG *	1.92	1.66	0.26	His (H)	CAU	0.77	1.27	-0.50
Ile (I)	AUU	0.36	0.64	-0.27		CAC *	1.23	0.73	0.50
	AUC *	2.43	1.95	0.48	Gln (Q)	CAA	0.69	1.21	-0.52
	AUA	0.21	0.42	-0.21		CAG *	1.31	0.79	0.52
Met (M)	AUG	1.00	1.00	0.00	Asn (N)	AAU	0.40	0.99	-0.59
Val (V)	GUU	0.56	0.69	-0.14	Lys (K)	AAC *	1.60	1.02	0.59
	<u>GUC</u>	1.00	1.24	-0.24		AAA	0.49	0.50	-0.01
	GUA	0.29	0.47	-0.18		<u>AAG</u>	1.51	1.50	0.01
	GUG *	2.16	1.59	0.57	Asp (D)	<u>GAU</u>	0.79	1.23	-0.44
Ser (S)	<u>UCU</u>	0.67	0.89	-0.22		GAC*	1.21	0.77	0.44
	UCC	1.43	0.64	0.80	Glu (E)	GAA	0.40	0.78	-0.38
	UCA	0.37	0.95	-0.58		GAG *	1.60	1.22	0.38
	AGU*	0.27	0.40	-0.13	Cys (C)	UGU	0.41	0.61	-0.20
	AGC *	2.12	1.90	0.22		<u>UGC</u>	1.44	1.39	0.05
Pro (P)	<u>UCG</u>	1.14	1.22	-0.08	Phe (F)	UUU	0.31	0.68	-0.37
	CCU	1.07	1.36	-0.29		UUC *	1.69	1.32	0.37
	CCC	1.02	0.70	0.32	Trp (W)	UGG	1.00	1.00	0.00
	<u>CCA</u>	0.60	1.34	-0.74		CGU*	0.71	0.12	0.59
	CCG *	1.31	0.60	0.71	Arg (R)	CGC*	1.29	0.30	0.98
Thr (T)	ACU	0.74	0.81	-0.08		CGA*	0.49	0.41	0.08
	ACC *	1.69	0.51	1.18		CGG*	1.11	0.68	0.42
	ACA	0.43	1.53	-1.10		<u>AGA</u>	1.23	2.12	-0.90
	<u>ACG</u>	1.14	1.15	-0.01		AGG	1.19	2.36	-1.18
Ala (A)	<u>GCU</u>	1.05	1.12	-0.07	Gly (G)	GGU*	0.72	0.38	0.34
	GCC *	0.99	0.62	0.37		GGC *	2.05	1.63	0.42
	GCA*	1.00	0.89	0.12		GGA	0.75	1.08	-0.33
	GCG	0.96	1.38	-0.42		GGG	0.63	0.90	-0.28

* indicates Δ RSCU ≥ 0.08 , the underlined text represents high-frequency codons (RSCU > 1), and the bold text highlights the optimal codon in MaWRKY genes of *M. acuminata* 'Guijiao 9'.



423

424 **Fig. 1** Nucleic acid composition of *MaWRKY* codons. The Y-axis represents the nucleic acid content, while
 425 the X-axis corresponds to parameters related to *MaWRKY* codons. These parameters included the total GC
 426 content of each coding domain sequence (CDS), the nucleotide composition at the third codon position (A3s,
 427 T3s, G3s, C3s), the GC content at each codon position-GC1, GC2, and GC3 (representing the first, second,
 428 and third positions, respectively) for each CDS.



429

430 **Fig. 2** Principal component analysis (PCA) of *MaWRKY* codon usage frequency. **(A)** The relative and
 431 cumulative contributions of the first 20 factors to the total variance in PCA; **(B)** PCA plot of codon RSCU
 432 values on codons ending with A, T, G, or C. Green squares represent codons ending with A; red triangles
 433 indicate those ending with T, orange stars indicate codons ending with G, and blue circles indicate
 434 codons ending with C.

435

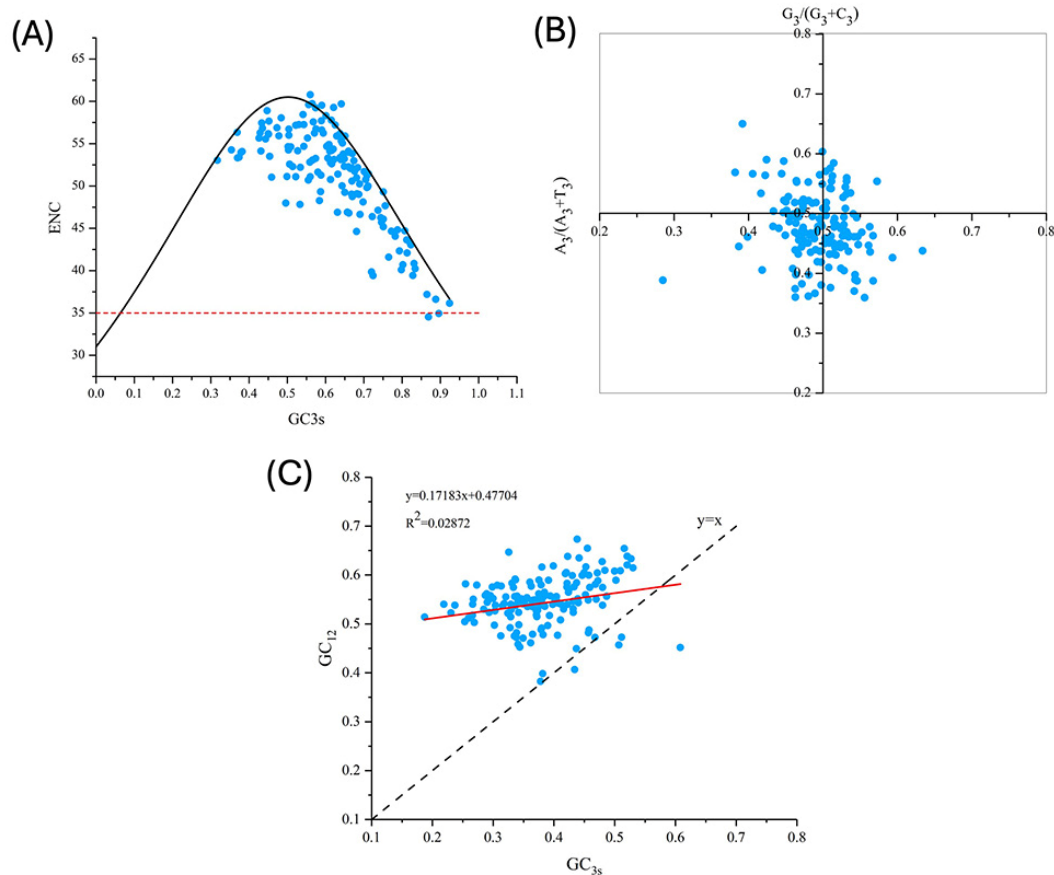


Fig. 3 (A) ENC-plot analysis of *MaWRKY* against GC3s. Points on or near the curve indicate bias caused by mutation pressure, while points away from the curve suggest influence from natural selection or other factors. The red-dotted line means the lowest ENC value of *MaWRKY*. **(B)** PR2 plot analysis. The midpoint at 0.5 represents an equal balance between $G = C$ and $A = T$, indicating no bias between mutation and selection pressure. Deviations from 0.5 suggest that codon bias is primarily influenced by factors other than base mutation of gene-encoding amino acids. **(C)** Neutrality plot analysis of *MaWRKY*. The plot compares GC3 (x-axis) and GC12 (y-axis). An RC value less than 0.5 indicates a stronger influence of natural selection, whereas an RC greater than 0.5 suggests a greater impact of mutation pressure. The red straight line represents the fitted curve.

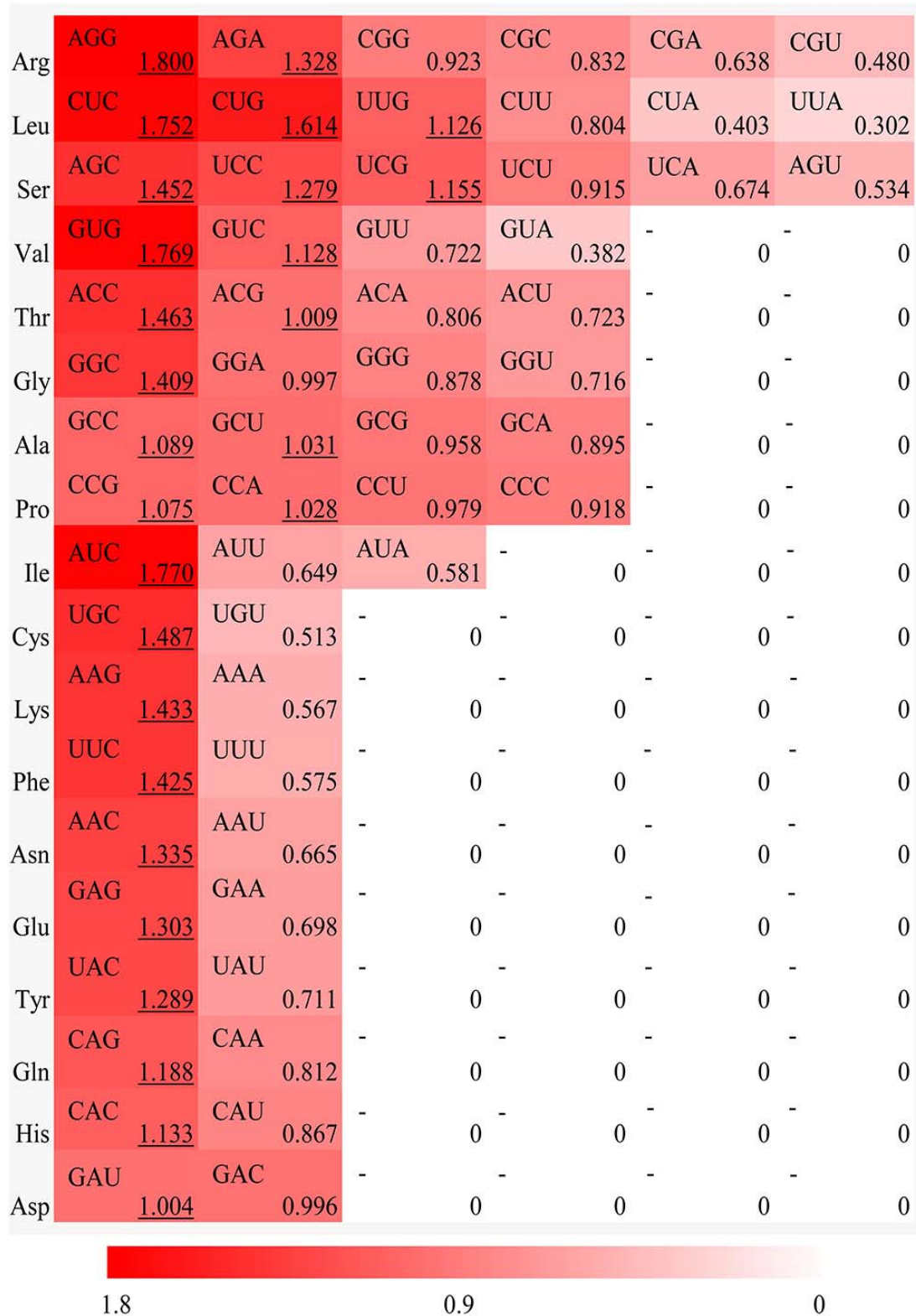
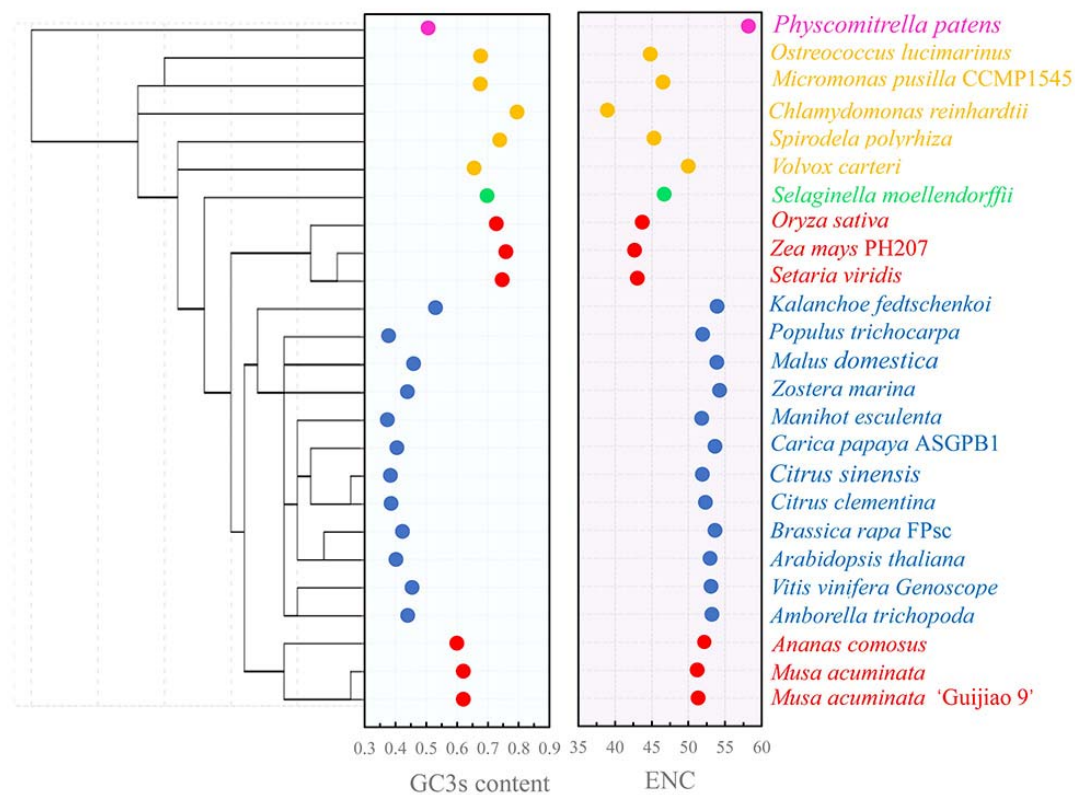


Fig. 4 Relative synonymous codon usage (RSCU) analysis for each amino acid in the *WRKY* gene family of *Musa acuminata* banana 'Guijiao 9'. The red color bar represents the RSCU value, with deeper shades indicating higher RSCU values



450

451 **Fig. 5** Cluster Analysis of RSCU values of 25 *WRKY* family members across different plant species, along
 452 with the mean GC3 and ENC values for each corresponding species. Purple represents Bryophytes, orange
 453 represents Algae, green represents Pteridophytes, red represents monocotyledons, and blue represents
 454 dicotyledons.

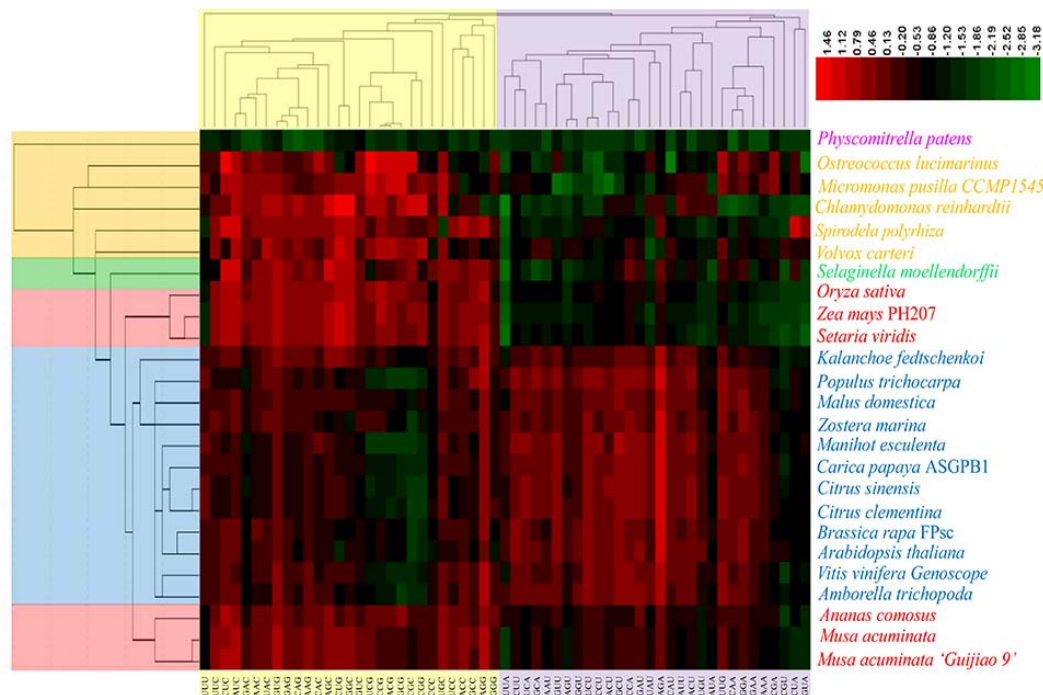


Fig. 6 Heatmap analysis of the RSCU values for 25 WRKY family members across different plant species. The color bars represent the RSCU values, with a colour gradient from green to red indicating low to high RSCU values, respectively. The redder the color, the higher the RSCU value, while the greener the color, the lower the RSCU value.

References

- Ahmad, F., Martawi, N.M., Poerba, Y.S., Hans, S., Henk, K., Gert, H.J., 2020. Genetic mapping of Fusarium wilt resistance in a wild banana *Musa acuminata* ssp. *malaccensis* accession. *Theor. Appl. Genet.* 133(12): 3409-3418.
- Behura, S.K., Severson, D.W., 2013. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol Rev.* 88: 49-61.
- Chen, A., Sun, J.M., Martin, G., Gray, L.A., Hribová, E., Christelová, P., Yahiaoui, N., Rounsley, S., Lyons, R., Batley, J., Chen, N., Hamill, S., Rai, S.K., Coin, L., Uwimana, B., D'Hont, A., Dolezel, J., Edwards, D., Swennen, R., Aitken, E.A.B., 2023a. Identification of a Major QTL-Controlling Resistance to the Subtropical Race 4 of *Fusarium oxysporum* f. sp. *cubense* in *Musa acuminata* ssp. *malaccensis*. *Pathogens* 12, 2: 289.
- Chen, A., Sun, J.M., Viljoen, A., Mostert, D., Xie, Y.C., Mangila, L., Bothma, S., Lyons, R., Hřibová, E., Christelová, P., 2023b. Genetic Mapping, Candidate Gene Identification and Marker Validation for Host Plant Resistance to the Race 4 of *Fusarium oxysporum* f. sp. *cubense* using *Musa acuminata* ssp. *malaccensis*. *Pathogens* 12, 6: 820.
- Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., 2013. Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem Biophys Res Commun.* 430:1344-1348.
- Dale, J., James, A., Paul, J.Y., Khanna, H., Smith, M., Peraza-Echeverria, S., Garcia- Bastidas, F., Kema, G., Waterhouse, P., Mengersen, K., Harding, R., 2017. Transgenic Cavendish bananas with resistance to Fusarium wilt tropical race 4. *Nat. Commun.* 8(1): 1496.
- Deb, B., Uddin, A., Chakraborty, S., 2020. Codon usage pattern and its influencing factors in different

481 genomes of hepadnaviruses. Arch Virol. 165: 557-570.

482 Deng, Y., Hedayioglu, F., Kalfon, J., Chu, D., Haar, T., 2020. Hidden patterns of codon usage bias across
483 kingdoms. J R Soc Interface 17: 20190819.

484 Deng, P.J., Wang, Z., Li, W.Y., Chen X.H., Liu, D.Q., 2024. WRKY11 up-regulated dirigent expression to
485 enhance lignin/lignans accumulation in *Lilium regale* Wilson during response to *Fusarium* wilt. Journal
486 of Integrative Agriculture, 23(8): 2703-2722.

487 Dita, M., Barquero, M., Heck, D., Mizubuti, E.S.G., Staver, C.P., 2018. Fusarium Wilt of Banana: Current
488 Knowledge on Epidemiology and Research Needs Toward Sustainable Disease Management. Front.
489 Plant Sci. 9:1468.

490 Dong, J., Chen, C., Chen, Z., 2003. Expression profiles of the Arabidopsis WRKY gene superfamily during
491 plant defense response. Plant Mol Biol. 51: 21-37.

492 Drenth, A., Kema, G., 2021. The Vulnerability of Bananas to Globally Emerging Disease Threats.
493 Phytopathology 111: 2146-2161.

494 Duret, L., Mouchiroud, D., 1999. Expression pattern and, surprisingly, gene length shape codon usage in
495 *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. PANS. 96: 4482-4487.

496 Gao, Y., Lu, Y., Song, Y., Jing, L., 2022. Analysis of codon usage bias of WRKY transcription factors in
497 *Helianthus annuus*. BMC Genomic Data 23:46.

498 Govardhana, M., Satyan, K.B., 2020. In-silico Analysis of Cucumber (*Cucumis sativus* L.) Genome for
499 WRKY Transcription Factors and cis-acting Elements. Comput Bio Chem. 85: 107212.

500 Goyal, P., Devi, R., Verma, B., Hussain, S., Arora, P., Tabassum, R., Gupta, S., 2023.
501 WRKY transcription factors: evolution, regulation, and functional diversity in plants. Protoplasma
502 260: 331–348.

503 Guan, D.L., Ma, L.B., Salabat, K.M., Zhan, X.X., Xu, S.Q., Xie, J.Y., 2018. Analysis of codon usage
504 patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant
505 selection constraints. BMC Genomics19: 542.

506 Hambuch, T.M., Parsch, J., 2005. Patterns of synonymous codon usage in *Drosophila melanogaster* genes
507 with sex-biased expression. Genetics 170: 1691-1700.

508 He, B., Dong, H., Jiang, C., Cao, F., Xu, L.A., 2016. Analysis of codon usage patterns in *Ginkgo biloba*
509 reveals codon usage tendency from A/U-ending to G/C-ending. Sci Rep. 6: 35927.

510 He, B., Dong, H., Jiang, C., Cao, F.L., Tao, S.T., Xu, L.A., 2016. Analysis of codon usage patterns in
511 *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C ending. Sci. Rep. 6: 35927.

512 Hershberg, R., Petrov, D.A., 2009. General rules for optimal codon choice. Plos genet. 5: e1000556.

513 Jiang, Y., Deng, F., Wang, H., Hu, Z., 2008. An extensive analysis on the global codon usage pattern of
514 baculoviruses. Arch Virol. 153: 2273.

515 Kanaya, S., Kudo, Y., Nakamura, Y., Ikemura, T., 1996. Detection of genes in *Escherichia coli* sequences
516 determined by genome projects and prediction of protein production levels, based on multivariate
517 diversity in codon usage. Bioinformatics. 12: 213-225.

518 Kayum, M.A., Jung, H.J., Park, J.I., Ahmed, N.U., Saha, G., Yang, T.J., 2015. Identification and expression
519 analysis of WRKY family genes under biotic and abiotic stresses in *Brassica rapa*. Mol Genet
520 Genomics. 290: 79-95.

521 Kumar, N, Bera, B.C., Greenbaum, B.D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B.N.,
522 Virmani, N., 2016. Revelation of influencing factors in overall codon usage bias of equine influenza
523 viruses. Plos one 11(4): e0154376.

524 Li, C.Y., Deng, G.M., Yang, J., Viljoen, A., Jin, Y., Kuang, R.B., Zuo, C.W., Lv, Z.C., Yang, Q.S., Sheng,

O., Wei, Y.R., Hu, C.H., Dong, T., Yi, G.J., 2012. Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *cubense* tropical race 4. BMC Genom. 13: 374.

Li, G.Y., Wang, Z., Zhang, Z.Y., Fang, H.D., Tan, X.L., 2013. The base composition and codon use of the WRKY gene family of the *Brassica napus*. J Biol. 30: 42-45.

Li, X., Tang, Y., Zhou, C., Zhang, L., Lv, J., 2020. A Wheat WRKY Transcription Factor TaWRKY46 Enhances Tolerance to Osmotic Stress in transgenic Arabidopsis Plants. Int J Mol Sci. 21: 1321.

Liu, H., He, R., Zhang, H., Huang, Y., Tian, M., Zhang, J., 2010. Analysis of synonymous codon usage in *Zea mays*. Mol Biol Rep. 37: 677.

Liu, H., Lu, Y., Lan, B., Xu, J., 2020. Codon usage by chloroplast gene is bias in *Hemiptelea davidii*. J Genet. 99: 8.

Liu, J.N., Fang, H.C., Liang, Q., Dong, Y.H., Wang, C.X., Yan, L.P., Ma, X.M., Zhou, R., Lang, X.Y., Gai, S.S., Wang, L.C., Xu, S.Q., Yang, K.Q., Wu, D.J., 2023. Genomic analyses provide insights into the evolution and salinity adaptation of halophyte *Tamarix chinensis*. GigaScience 12: 1–17.

Liu, Q., Feng, Y., Dong, H., Xue, Q., 2004. Synonymous codon usage bias in *Oryza sativa*. Plant Sci. 167: 101-105.

Liu, S., Qiao, Z., Wang, X., Zeng, H., Li, Y., Cai, N., Chen, Y., 2019. Analysis of codon usage patterns in “*Lonicerae Flos*” (*Lonicera macranthoides* Hand. -Mazz.) based on transcriptome data. Gene. 705: 127-132.

Ma, Q., Li, C., Wang, J., Wang, Y., Ding, Z., 2015. Analysis of synonymous codon usage in FAD7 genes from different plant species. Genet. Mol Res. 14: 1414-1422.

Mondal, S.K., Kundu, S., Das, R., Roy, S., 2016. Analysis of phylogeny and codon usage bias and relationship of GC content, amino acid composition with expression of the structural nif genes. J Biomol Struct Dyn. 34: 1649-1666.

Murray, E.E., Lotzer, J., Eberle, M., 1989. Codon usage in plant genes. Nucleic Acids Res. 17: 477-498.

Palidwor, G.A., Perkins, T.J., Xia, X.A., 2010. General model of codon bias due to GC mutational bias. PLoS One. 5: e13431.

Patel, P., Yadav, K., Srivastava, A.K., Suprasanna, P., Ganapathi, T.R., 2019. Overexpression of native Musa-miR397 enhances plant biomass without compromising abiotic stress tolerance in banana. Sci Rep. 9: 1-15.

Pegg, K.G., Coates, L.M., O'Neill, W.T., Turner, D.W., 2019. The Epidemiology of Fusarium Wilt of Banana. Front Plant Sci. 10, doi:10.3389/fpls.2019.01395.

Ploetz, R.C., Randy, C., 2015. Fusarium Wilt of Banana. *Phytopathology* 105, 12: 1512-21.

Qiu, S., Bergero, R., Zeng, K., Charlesworth, D., 2011. Patterns of codon usage bias in *Silene latifolia*. Mol Biol Evol. 28: 771-780.

Roberts, J.M., Carvalhais, L.C., O'Dwyer, C., Rincón-Flórez, V.A., Drenth, A., 2024. Diagnostics of Fusarium wilt in banana: Current status and challenges. Plant Pathol 73(4): 760-776.

Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that influence the evolution of codon bias. Phil Trans R Soc B. 365: 1203-1212.

Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons. Nucleic acids Res. 14: 7737-7749.

Sharp, P.M., Li, W.H., 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15: 1281-95.

Siamak, S.B., Zheng, S., 2018. Banana Fusarium wilt (*Fusarium oxysporum* f. sp. *cubense*) control and

569 resistance, in the context of developing wilt-resistant bananas within sustainable production systems.
570 Horti Plant J. 4: 208-218.

571 Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. PANS. 85: 2653-2657.

572 Sun, J., Zhang, J., Fang, H., Peng, L., Wei, S., Li, C., Zheng, S.J., Lu, J., 2019. Comparative transcriptome
573 analysis reveals resistance-related genes and pathways in *Musa acuminata* banana 'Guijiao 9' in
574 response to *Fusarium* wilt. Plant Physiol Bioch. 141: 83-94.

575 Swarupa, V., Ravishankar, K.V., Rekha, A., 2014. Plant defense response against *Fusarium oxysporum* and
576 strategies to develop tolerant genotypes in banana. Planta 239, 735–751.

577 Uddin, A., Mazumder, T.H., Barbhuiya, P.A., Chakraborty, S., 2020. Similarities and dissimilarities of
578 codon usage in mitochondrial ATP genes among fishes, aves, and mammals. IUBMB Life72: 899-914.

579 Wang, S.F., Su, M.W., Tseng, S.P., Li, M.C., Tsao, C.H., Huang, S.W., 2016. Analysis of codon usage
580 preference in hemagglutinin genes of the swine-origin influenza A (H1N1) virus. J Microbio Immuno.
581 49: 477-86.

582 Wani, S.H., Anand, S., Singh, B., Bohra, A., Joshi, R., 2021. WRKY transcription factors and plant defense
583 responses: latest discoveries and future prospects. Plant Cell Rep 40(7):1071-1085

584 Wei, L., He, J., Jia, X., Qi, Q., Liang, Z., Zheng, H., Ping, Y., Liu, S.Y., Sun, J.C., 2014. Analysis of codon
585 usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. BMC Evol Biol. 14:
586 262.

587 Wright, F., 1990. The effective number of codons used in a gene. Gene. 87: 23-29.

588 Wu, Y., Zhao, D., Tao, J., 2015. Analysis of codon usage patterns in herbaceous Peony (*Paeonia lactiflora*
589 Pall.) based on transcriptome data. Genes 6: 1125.

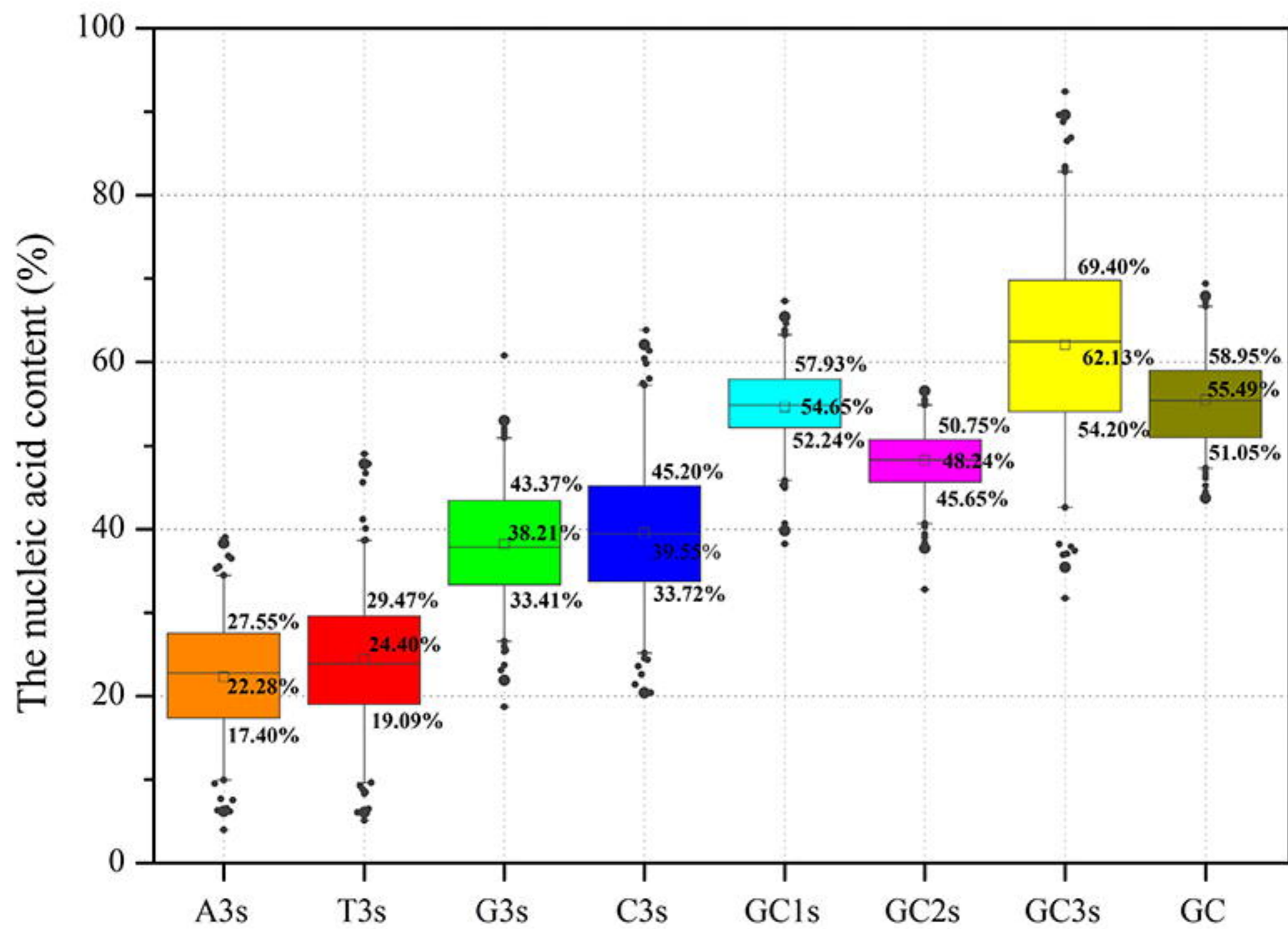
590 Xin, W., Liu, Y., Yang, Y., Sun, T., Niu, L., Ge, J., 2020. Detection, genetic, and codon usage bias analyses
591 of the VP2 gene of mink bocavirus. Virus Genes 56: 306-315.

592 Zhang, X., Cai, Y., Zhai, X., Liu, J., Zhao, W., Ji, S., Su, S., Zhou, J.Y., 2018. Comprehensive analysis of
593 codon usage on rabies virus and other lyssaviruses. Int J Mol Sci. 19: 2397.

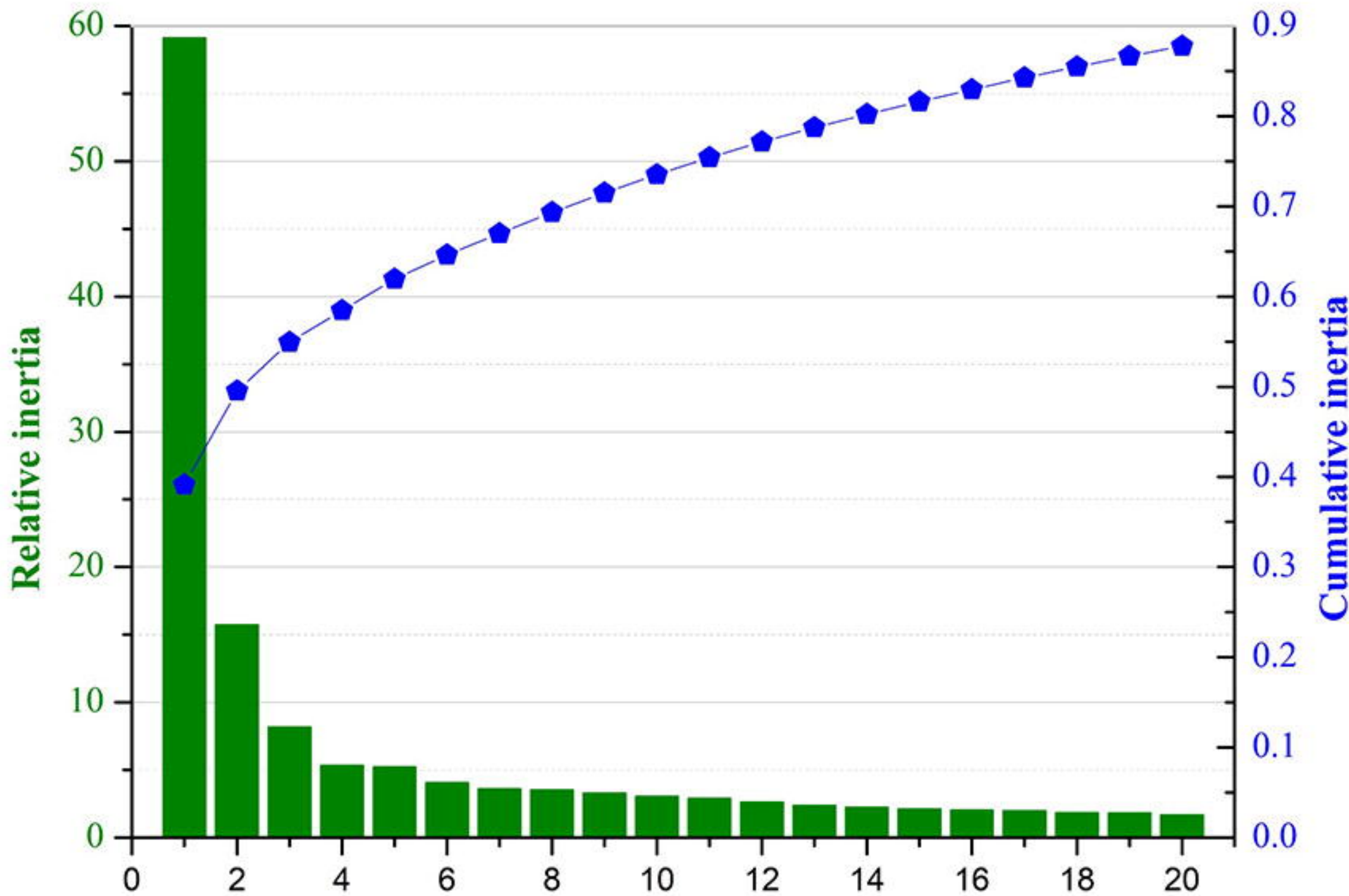
594 Zhao, Y., Zheng, H., Xu, A., Yan, D., Jiang, Z., Qi, Q., Sun, J.C., 2016. Analysis of codon usage bias of
595 envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution. BMC
596 Genom. 17: 677.

597 Zhou, T., Weems, M., Wilke, C.O., 2009. Translationally optimal codons associate with structurally
598 sensitive sites in proteins. Mol Biol Evol. 26: 1571-1580.

599 Zorrilla-Fontanesi, Y., Pauwels, L., Panis, B., Signorelli, S., Vanderschuren, H., Swennen,
600 R., 2020. Strategies to revise agrosystems and breeding to control *Fusarium* wilt of banana. Nat Food
601 599–604.

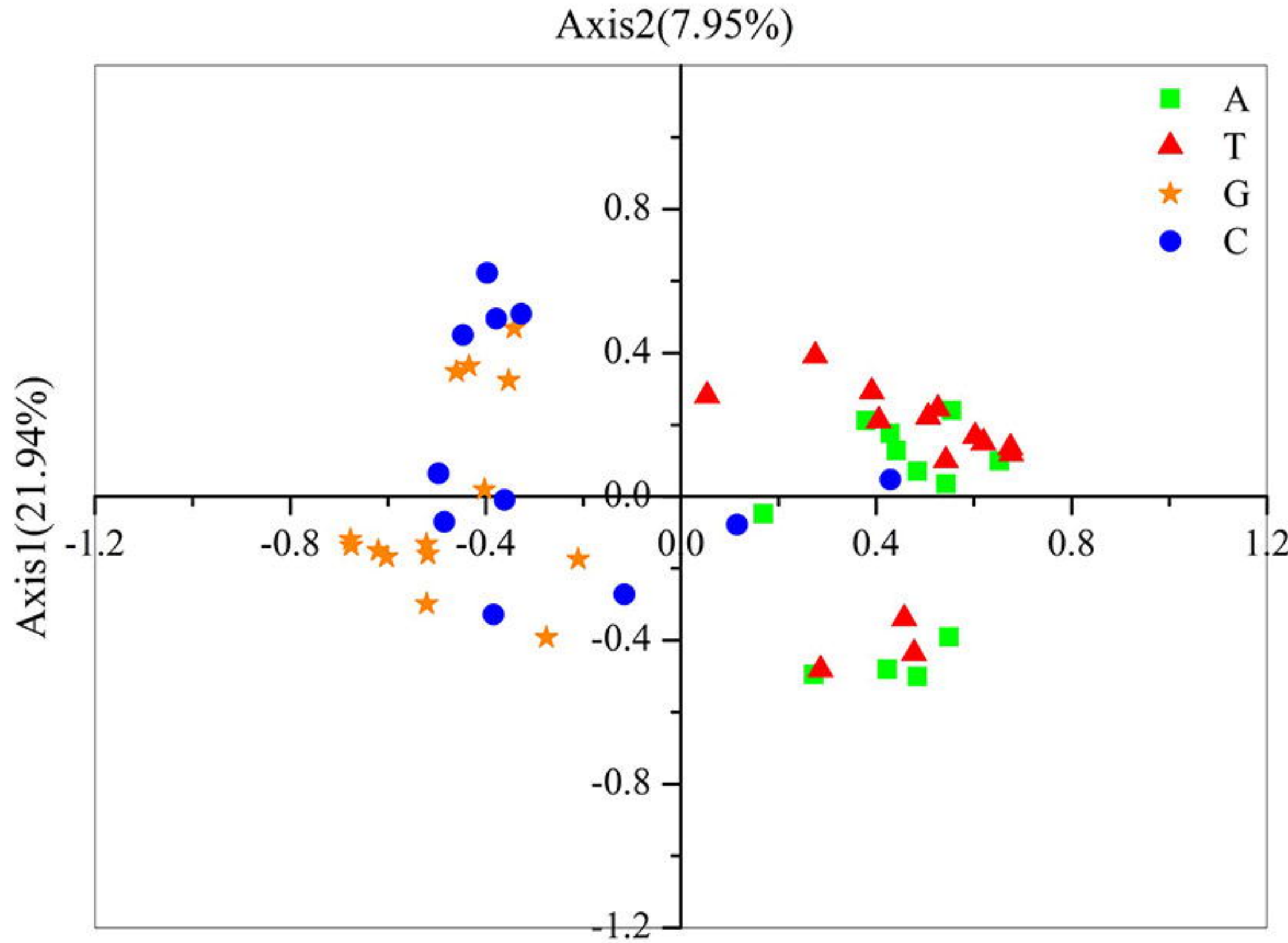


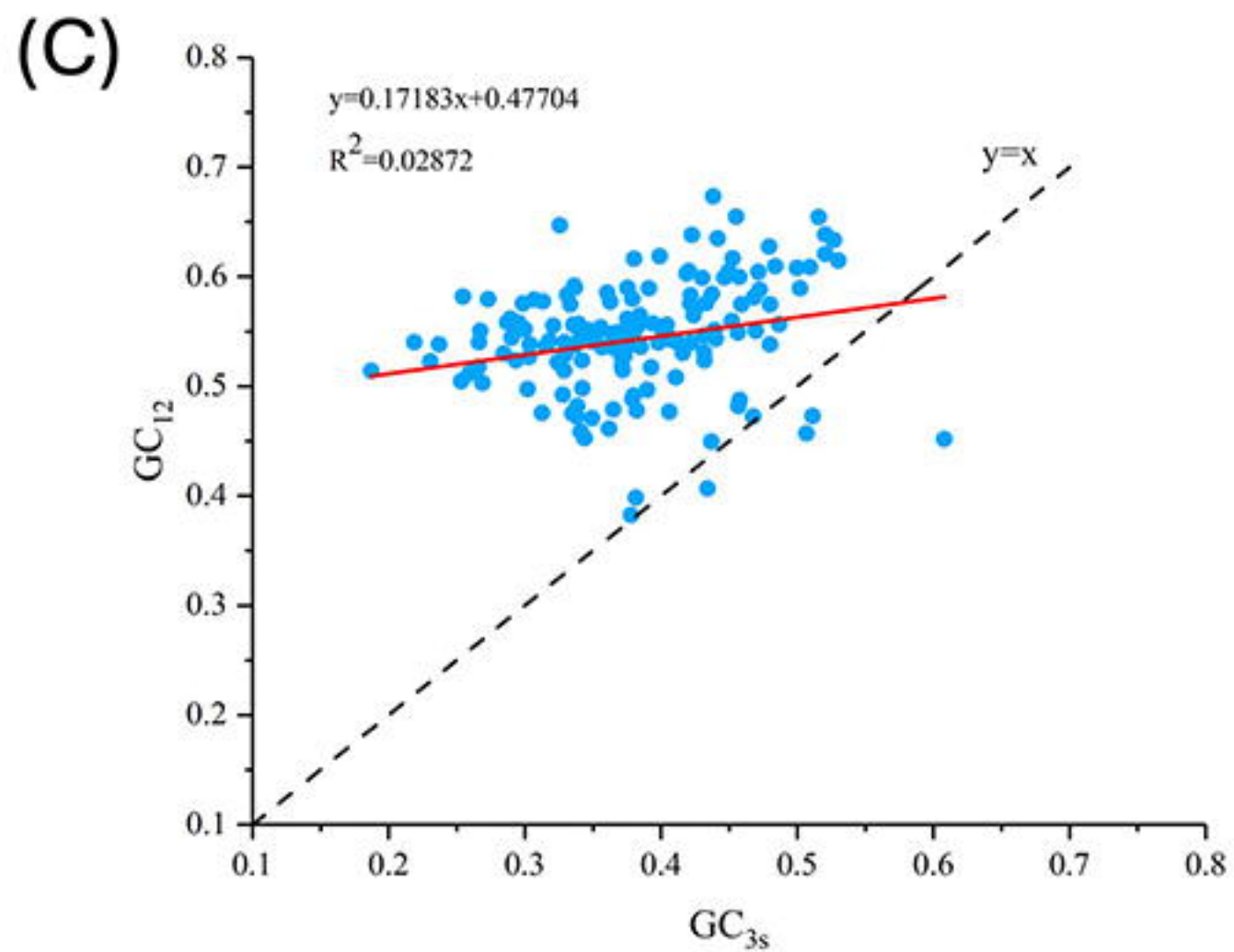
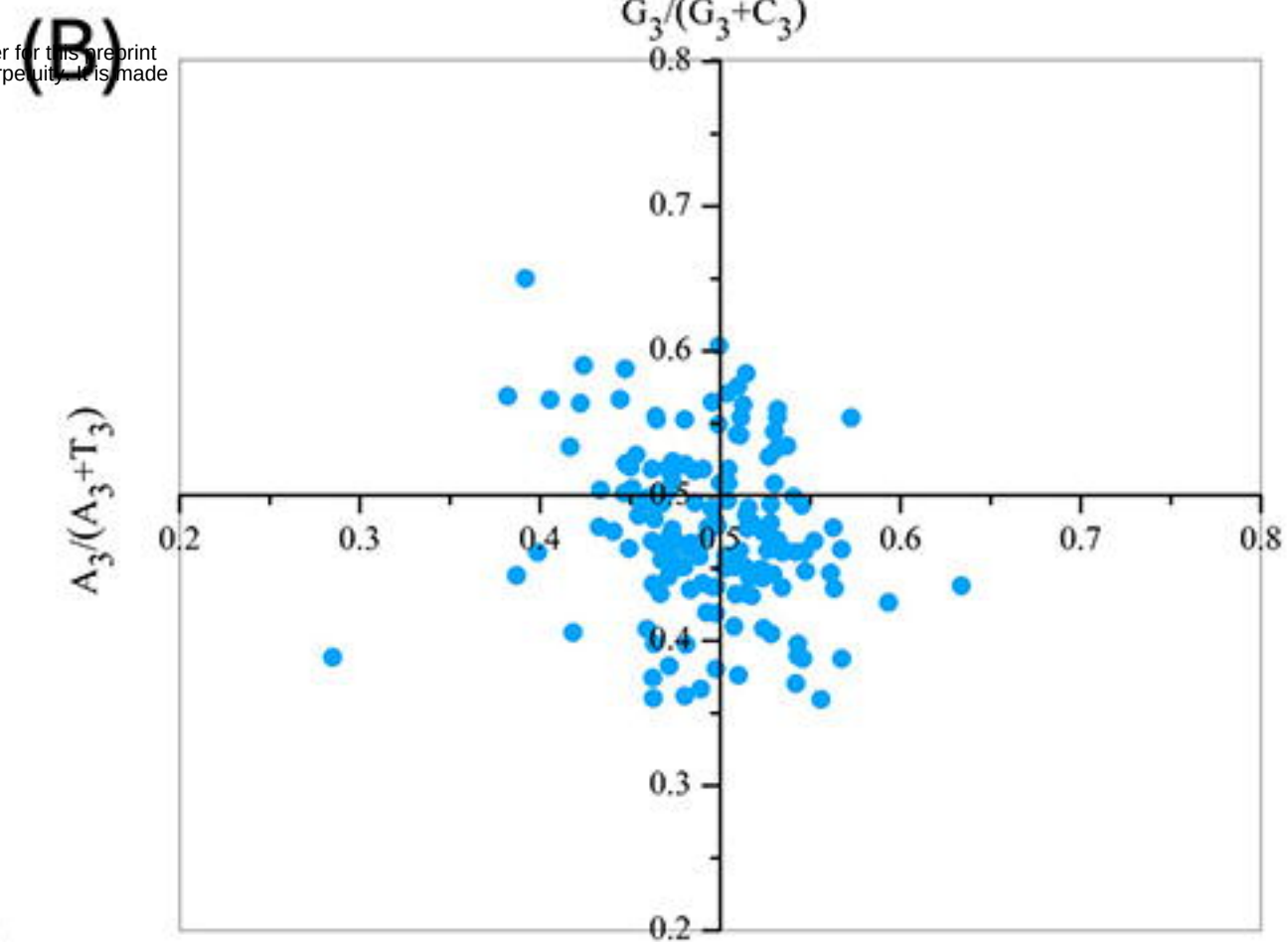
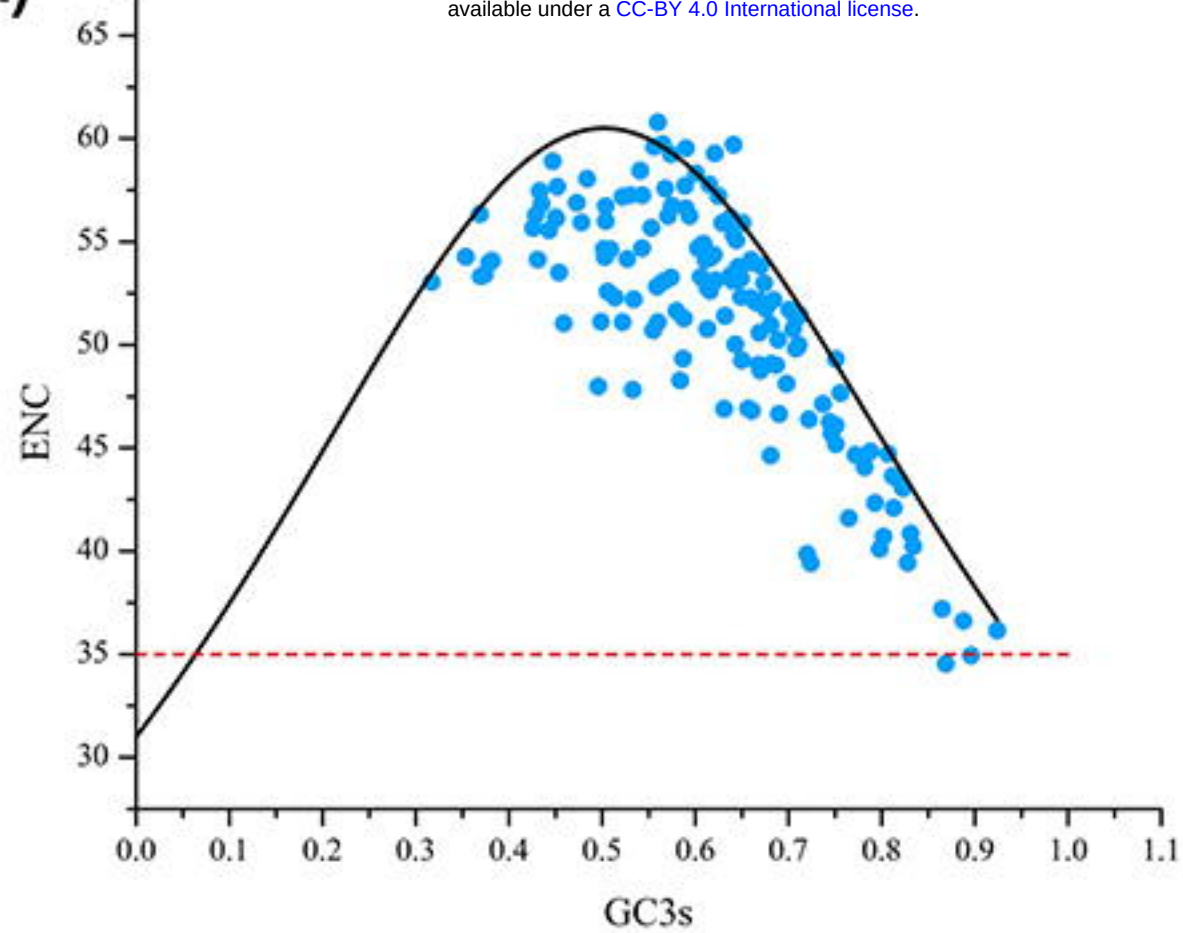
(A)



bioRxiv preprint doi: <https://doi.org/10.1101/2025.01.08.631839>; this version posted January 10, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](#).

(B)





bioRxiv preprint doi: <https://doi.org/10.1101/2025.01.08.631839>; this version posted January 10, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

