Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Global Genetic Diversity, Phylogenetic Relationships and Codon Usage Bias of Stagonosporopsis Species Causing Gummy Stem Blight

B. P. Maruthi Prasad

maruthiprasad1996@gmail.com

University of Agricultural Sciences, Bangalore

J. Harish

University of Agricultural Sciences, Bangalore

M. Pallavi

University of Agricultural Sciences, Bangalore

Research Article

Keywords: RPB2 gene, Haplotype diversity, Population structure, Relative synonymous codon usage (RSCU), Neutrality plot, Nucleotide composition

Posted Date: April 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9440555/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Global Genetic Diversity, Phylogenetic Relationships and Codon Usage Bias of Stagonosporopsis Species Causing Gummy Stem Blight

B. P. Maruthi Prasad[1]*, J. Harish[2] and M. Pallavi[1]
[1]Department of Genetics and Plant Breeding, University of Agricultural Sciences, GKVK, Bangalore, 560065, Karnataka, India
[2]Department of Plant Pathology, University of Agricultural Sciences, GKVK, Bengaluru, Karnataka, 560065, India
*Corresponding author's email: maruthiprasad1996@gmail.com

ABSTRACT

Gummy stem blight (GSB) is a devastating fungal disease of cucurbit crops worldwide, caused by species of the genus *Stagonosporopsis*. Despite its global economic importance, comprehensive studies integrating phylogenetic relationships, population genetics, and codon usage bias across international isolates remain limited. This study aimed to investigate the global genetic diversity, evolutionary relationships, and codon usage patterns of Stagonosporopsis species using RNA polymerase II second largest subunit (RPB2) gene as a molecular marker. A total of 234 sequences from diverse geographical regions were retrieved from GenBank and analyzed using phylogenetic reconstruction, haplotype network analysis, and comprehensive codon usage bias indices including relative synonymous codon usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), neutrality plot, and parity rule 2 (PR2) analysis. Phylogenetic analysis revealed distinct species-level clades with high bootstrap support, indicating that *RPB2* is a reliable marker for resolving evolutionary relationships within the genus. Haplotype analysis identified 43 distinct haplotypes with high haplotype diversity (Hd = 0.861), while analysis of molecular variance (AMOVA) indicated that most genetic variation (92.06%) occurs within populations, suggesting historical gene flow and global dispersal. Codon usage analysis revealed moderate bias shaped primarily by mutational pressure, with additional influences from natural selection, as evidenced by ENC-GC3 plots, neutrality plots, and PR2 bias. Optimal codons were identified, and correspondence analysis revealed distinct clustering patterns associated with nucleotide composition. These findings provide the integrated perspective on the global genetic structure and molecular evolution of *Stagonosporopsis* species. The observed genetic diversity and codon usage patterns have important implications for understanding pathogen adaptation, population dynamics, and evolutionary mechanisms, ultimately contributing to improved disease management strategies and resistance breeding programs in cucurbit crops.

KEYWORDS: RPB2 gene, Haplotype diversity, Population structure, Relative synonymous codon usage (RSCU), Neutrality plot, Nucleotide composition

INTRODUCTION

Cucurbit crops such as watermelon (*Citrullus lanatus*), melon (*Cucumis melo*), cucumber (*Cucumis sativus*), and squash represent economically important horticultural commodities cultivated worldwide for their nutritional value and economic significance. These crops contribute substantially to global vegetable production and provide essential vitamins, minerals, and dietary fibre for human consumption. However, the productivity and quality of cucurbit crops are frequently threatened by several biotic stresses, among which fungal diseases represent a major constraint to sustainable production. One of the most destructive fungal diseases affecting cucurbit crops is gummy stem blight (GSB), which occurs widely in both tropical and temperate regions (Seblani et al., 2023). The disease causes severe yield losses through symptoms such as stem cankers, leaf lesions, fruit rot, and seedling damping-off, leading to significant economic losses in cucurbit production systems worldwide (Keinath, 2011).

Gummy stem blight is caused by a complex of closely related fungal species belonging to the genus *Stagonosporopsis*, including *Stagonosporopsis cucurbitacearum*, *Stagonosporopsis citrulli*, and *Stagonosporopsis caricae*. These pathogens were previously classified within the genus Didymella (as *D. bryoniae*) but were later reassigned to the genus Stagonosporopsis based on molecular phylogenetic and taxonomic revisions. Species within this genus are capable of infecting a wide range of cucurbit hosts and can persist in infected plant debris and seeds, thereby facilitating rapid dissemination across production regions and continents (Stewart et al., 2015). The global distribution and adaptability of these pathogens highlight the importance of understanding their genetic diversity, evolutionary relationships, and population structure in order to develop effective disease management strategies and support breeding programs aimed at improving resistance in cucurbit crops.

Advances in molecular biology have significantly improved the ability to investigate evolutionary relationships and taxonomic classification of fungal pathogens (Niranjan et al., 2025). Molecular phylogenetic approaches based on conserved gene regions have become essential tools for resolving species boundaries and understanding evolutionary divergence among fungal taxa. Gene markers such as the RNA polymerase II second largest subunit (*RPB2*), the internal transcribed spacer (ITS) region, and translation elongation factor 1-alpha (*TEF1-α*) are widely used in fungal systematics because of their conserved structure and informative sequence variation (Schoch et al., 2012; Větrovský et al., 2016). The *RPB2* locus was first demonstrated to resolve phylogenetic relationships among ascomycetes by Liu et al. (1999), and it has since become a standard marker for fungal taxonomy. Phylogenetic analyses based on these molecular markers provide valuable insights into species relationships, evolutionary history, and the genetic structure of pathogen populations. Recent studies have demonstrated that multilocus phylogenetic approaches can effectively resolve species complexes and uncover cryptic diversity in fungal pathogens associated with plant diseases (Marin-Felix et al., 2019; Wittstein et al., 2020; Cafà et al., 2020).

In addition to phylogenetic analysis, investigations of codon usage bias have become an important component of molecular evolutionary studies. Codon usage bias refers to the unequal usage of synonymous codons encoding the same amino acid in protein-coding genes. Codon

usage bias is a universal phenomenon observed across all domains of life and arises from the combined effects of mutation, natural selection, and genetic drift (Parvathy et al., 2021). Codon usage bias has been linked to multiple biological and environmental factors, including gene expression levels, protein folding, and environmental adaptation across diverse organisms (Baeza et al., 2024). This phenomenon is influenced by several evolutionary forces, including mutation pressure, natural selection, translational efficiency, and genetic drift. Analyses of codon usage patterns can therefore provide insights into genome evolution, gene expression regulation, and adaptive processes occurring in microbial genomes (Hershberg and Petrov 2008; Plotkin & Kudla 2011). Various statistical indices have been developed to quantify codon usage bias, including relative synonymous codon usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), and parity rule 2 (PR2) bias. These indices are widely applied in evolutionary genomics to evaluate codon preference and infer the underlying evolutionary mechanisms shaping synonymous codon usage patterns.

Several graphical and statistical approaches are also commonly used to examine the factors influencing codon usage bias. For example, ENC–GC3 plots, neutrality plots, and correspondence analysis are widely applied to evaluate the relative contributions of mutation pressure and natural selection to codon usage patterns in coding sequences. These analytical methods enable researchers to determine whether codon preference primarily arises from mutational constraints or from selective pressures associated with translational efficiency and protein expression (Wright, 1990). Recent studies on fungal genomes have demonstrated that codon usage patterns are often shaped by a complex interplay of mutational bias, natural selection, and genomic compositional constraints (Han et al., 2022; Wu et al., 2023).

To our knowledge, limited studies have integrated phylogenetic, population genetic, and codon usage analyses for globally distributed isolates of *Stagonosporopsis* species. This study addresses this gap by simultaneously examining these features across a global dataset. Integrating these approaches can provide a comprehensive understanding of pathogen evolution, population genetic structure, and molecular adaptation. Such information is particularly valuable for plant pathology and breeding programs because genetic variation within pathogen populations can influence virulence evolution, host adaptation, and the durability of disease resistance.

Therefore, the present study aimed to investigate the global genetic diversity and evolutionary characteristics of *Stagonosporopsis* species associated with gummy stem blight using the RNA polymerase II second largest subunit (*RPB2*) gene as a molecular marker. Specifically, we performed phylogenetic analyses to elucidate evolutionary relationships among isolates from diverse geographical regions. In addition, haplotype diversity and haplotype network analyses were conducted to examine population genetic structure. Furthermore, comprehensive codon usage bias analyses were performed using multiple statistical indices and graphical approaches to evaluate the evolutionary forces shaping codon usage patterns in the *RPB2* gene. The findings of this study provide new insights into the molecular evolution and global diversity of *Stagonosporopsis* species and may contribute to improved disease management strategies and resistance breeding programs in cucurbit crops.

MATERIALS AND METHODS

Sequence retrieval

Nucleotide sequences of the RNA polymerase II second largest subunit (*RPB2*) gene of *Stagonosporopsis* species were retrieved from the GenBank database available at the National Center for Biotechnology Information (NCBI). A total of 234 sequences representing isolates from different geographical regions worldwide were obtained and used for subsequent analyses to assess global genetic diversity within the genus (Supplementary Table S1). The sequences were downloaded from the GenBank repository maintained by NCBI (Benson et al., 2018).

Recombination analysis

Recombination events among the aligned nucleotide sequences were analyzed using Recombination Detection Program version 4 (RDP4). The dataset was examined using multiple detection methods implemented in the software, including RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, 3Seq, LARD, and PhylPro. Default settings were used, and recombination events supported by more than one method with significant p-values were considered reliable (Martin et al., 2015).

Phylogenetic relationships among global *Stagonosporopsis* isolates

Multiple sequence alignment of the 234 RPB2 gene sequences was performed using the MUSCLE algorithm (Edgar, 2004) implemented in MEGA (version 11) (Tamura et al., 2021). The aligned dataset was subsequently used for phylogenetic reconstruction using the maximum likelihood (ML) method implemented in IQ-TREE (Nguyen et al., 2015; Trifinopoulos et al., 2016; Minh et al., 2020). Model selection was conducted using ModelFinder, which automatically identifies the best-fit nucleotide substitution model based on statistical criteria such as the Bayesian Information Criterion (BIC) (Kalyaanamoorthy et al., 2017). Phylogenetic tree construction was performed using the selected substitution model, and branch support was evaluated using ultrafast bootstrap analysis with 1000 replicates to assess the robustness of the inferred phylogenetic relationships (Hoang et al., 2018).

Haplotype analysis

Genetic diversity and haplotype structure among the sequences were assessed using DnaSP version 6 (Rozas et al., 2017). Several population genetic parameters, including the number of haplotypes (H), haplotype diversity (Hd), and nucleotide diversity (π), were calculated to evaluate the extent of genetic variation within the dataset. These indices provide important measures of sequence polymorphism and allow inference of the genetic structure and evolutionary relationships among isolates from different geographical regions (Rozas et al., 2017).

Haplotype network construction

A haplotype network was constructed to visualize the genealogical relationships among haplotypes and to infer potential evolutionary connections among isolates from different geographical regions. The network was generated using the median-joining algorithm implemented in PopART version 1.7 (Leigh and Bryant, 2015), which is widely used for reconstructing haplotype relationships in population genetic studies (Bandelt et al., 1999; Leigh and Bryant, 2015). This method combines features of minimum-spanning trees and parsimony analysis to identify possible ancestral haplotypes and mutation events among sequences.

Comprehensive analysis of codon usage pattern

RPB2 gene sequences of Stagonosporopsis retrieved from the NCBI database with complete open reading frames and lengths that were multiples of three nucleotides were retained for analysis to avoid frame-shift errors. RPB2 gene sequences with complete open reading frames and lengths that were multiples of three nucleotides were retained to avoid frame-shift errors. Sequences containing internal stop codons, ambiguous bases (Ns), or frameshifts were excluded to ensure high-quality data for codon usage analysis (Wright, 1990; Fu et al., 2023).

All statistical analyses and graphical visualizations were conducted using Python scripts implemented in Jupyter Notebook. Data manipulation and statistical calculations were performed using the Python libraries Pandas, NumPy, and SciPy (Harris et al., 2020; Virtanen et al., 2020), while graphical outputs such as RSCU heatmaps, ENC plots, neutrality plots, PR2 plots, and correspondence analysis plots were generated using Matplotlib and Seaborn (Hunter, 2007; Waskom, 2021).

Nucleotide composition analysis

Nucleotide composition was calculated to determine the base distribution within the coding sequences. The overall frequencies of adenine (A), thymine (T), guanine (G), and cytosine (C) were determined. In addition, nucleotide frequencies at the three codon positions (A1, T1, G1, C1; A2, T2, G2, C2; A3, T3, G3, C3) were calculated. The GC content and GC content at the first, second, and third codon positions (GC1, GC2, GC3) were estimated to evaluate compositional constraints influencing codon usage patterns (Fu et al., 2023; Wu et al., 2023).

$$GC = \frac{G + C}{A + T + G + C}$$

$$GC_{12} = \frac{GC_1 + GC_2}{2}$$

$$GC_3 = \frac{G_3 + C_3}{A_3 + T_3 + G_3 + C_3}$$

Relative synonymous codon usage (RSCU)

The Relative Synonymous Codon Usage (RSCU) value was calculated to evaluate codon usage preference independent of amino acid composition (Sharp and Li, 1986). The RSCU value was computed using the following formula:

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

Where X_{ij} is the observed number of occurrences of the j^{th} codon for the i^{th} amino acid and n_i is the observed number of occurrences of the i^{th} amino acid.

An RSCU value > 1 indicates that a codon is used more frequently than expected, whereas RSCU < 1 indicates under-representation (Sharp and Li, 1986; Fu et al., 2023).

Effective number of codons (ENC)

The Effective Number of Codons (ENC) was calculated to quantify the overall magnitude of codon usage bias. ENC values range from 20 (maximum bias) to 61 (no bias). The ENC value was calculated according to the method proposed by Wright (1990):

$$ENC = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

Where F_k represents the average homozygosity for codon families with k synonymous codons (Wright, 1990; Fu et al., 2023). Lower ENC values indicate stronger codon usage bias.

Codon adaptation index (CAI)

The Codon Adaptation Index (CAI) was used to evaluate codon adaptation and potential translational efficiency rather than direct gene expression levels (Sharp and Li, 1987). The CAI value was calculated as:

$$CAI = \left(\prod_{i=1}^{L} w_i \right)^{\frac{1}{L}}$$

Where L is the total number of codons in the gene abd w_i is the relative adaptiveness of the i^{th} codon. CAI values range between 0 and 1, where higher values indicate stronger codon preference and potentially higher translational efficiency (Sharp and Li, 1987; Fu et al., 2023).

Parity Rule 2 (PR2) bias analysis

The PR2 bias plot was constructed to examine the relative influence of mutation pressure and selection on codon usage bias (Sueoka, 1995), following the methodology recommended by Li et al. (2023) for evaluating these forces in fungal coding sequences. The following ratios were calculated:

$$PR2_x = \frac{A_3}{A_3 + T_3}$$

$$PR2_y = \frac{G_3}{G_3 + C_3}$$

Deviation from the center of the plot (0.5, 0.5) indicates unequal usage of complementary nucleotides at the third codon position (Sueoka, 1995; Li et al., 2023).

Neutrality plot analysis

A neutrality plot was constructed by plotting GC12 against GC3 to estimate the relative contributions of mutation pressure and natural selection in shaping codon usage bias (Sueoka, 1988). The regression slope between GC12 and GC3 was used to estimate the relative contributions of mutation pressure and natural selection. The regression slope of the neutrality plot indicates the relative contribution of mutation pressure versus natural selection (Sueoka, 1988; Li et al., 2023).

$$GC12 = \frac{GC1 + GC2}{2}$$

ENC–GC3 plot analysis

The ENC–GC3 plot was generated to evaluate whether codon usage bias was influenced primarily by mutational pressure. The expected ENC value under mutation-only conditions was calculated using the following equation (Wright, 1990):

$$ENC_{expected} = 2 + GC_3 + \frac{29}{GC_3^2 + (1 - GC_3)^2}$$

Genes falling close to the expected curve are mainly influenced by mutation pressure, whereas genes below the curve suggest additional effects of natural selection (Wright, 1990; Wu et al., 2023).

Correspondence analysis (COA)

Correspondence analysis (COA) was performed on the RSCU matrix to identify the major trends in codon usage variation among genes. COA reduces multidimensional codon usage data into principal axes representing the major sources of variation in codon usage patterns (Greenacre, 1984; Wu et al., 2023).

RESULTS AND DISCUSSION

Recombination analysis

Recombination analysis of the aligned *Stagonosporopsis* RNA polymerase II second largest subunit (RPB2) gene sequences was performed using RDP4 to investigate potential recombination signals within the dataset. RDP4 integrates several recombination detection algorithms, including RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan, 3Seq, LARD, and PhylPro, which apply different statistical and phylogenetic approaches to increase the reliability of recombination detection (Martin et al., 2015). The use of multiple algorithms is recommended because recombination signals supported by several independent methods are generally considered more reliable (Posada, 2002).

The analysis revealed limited evidence of recombination among the analyzed RPB2 sequences of *Stagonosporopsis*. Among the detection methods implemented in RDP4, GENECONV detected recombination signals in two sequences with an average p-value of 1.108×10^{-2}, BootScan detected one sequence with an average p-value of 1.168×10^{-3}, and the 3Seq method detected one sequence with a statistically significant signal (3.516×10^{-4}). In contrast, other algorithms including RDP, MaxChi, Chimaera, SiScan, LARD, and PhylPro did not detect recombination signals within the dataset (Supplementary Table S2). The detection of recombination signals by only a subset of the available algorithms suggests that recombination events within the RPB2 gene region are relatively rare or weakly supported in the analyzed population.

A potential recombination event (Event 3) was detected in the sequence MW853797, with predicted recombination breakpoints located between nucleotide positions 242 and 1046 of the alignment (Table 1, Supplementary Figure S1). The analysis suggested OQ271774 as the major parental sequence, showing approximately 99.8% sequence similarity, whereas the minor parent was inferred to be an unknown sequence related to LR983949. However, the RDP4 output indicated that the recombination signal may be attributable to processes other than true recombination and that only one parental sequence was confidently identified. Such warnings

may arise when recombination signals are weak or when high sequence similarity among isolates leads to ambiguous parental assignments.

The overall results therefore indicate limited recombination activity within the RPB2 gene region of Stagonosporopsis. The detection of recombination signals by only a subset of algorithms suggests that recombination events in the RPB2 gene are infrequent or weakly supported in the analyzed dataset. This pattern is consistent with the expectation for conserved housekeeping genes, where recombination is often limited compared to rapidly evolving genomic regions (Taylor & Fisher, 2003). The low recombination signal also implies that the observed genetic variation is more likely driven by mutation and lineage divergence rather than recombination-mediated genetic exchange (Gladieux et al., 2014).

The absence of strong recombination signals also supports the reliability of the phylogenetic analyses performed in the present study. Recombination can obscure phylogenetic relationships by generating mosaic sequences that violate the assumption of tree-like evolution. However, the limited recombination detected here suggests that the RPB2 gene provides a stable phylogenetic signal for reconstructing evolutionary relationships among *Stagonosporopsis* species. Consequently, the sequences used in this study are suitable for investigating the global genetic diversity and evolutionary dynamics of this important group of fungal plant pathogens.

Table 1. Putative recombination events detected in RPB2 gene sequences of *Stagonosporopsis* isolates using RDP4, including recombinant sequences, parental sequences, breakpoint positions, detection methods, and statistical support.

Event	Recombinant Sequence	Major Parent	Minor Parent	Breakpoint Positions (nt)	Detection Methods	Average P-value
3	MW853797	OQ271774	LR983949 (Unknown)	242–1046	GENECONV, BootScan, 3Seq	3.516×10^{-4}

Phylogenetic relationships among global *Stagonosporopsis* isolates

Phylogenetic relationships among global *Stagonosporopsis* isolates were inferred using the maximum likelihood (ML) method based on the RPB2 gene sequences. The ML phylogeny revealed a well-resolved tree topology with several distinct clades corresponding to recognized species within the genus. The clustering of isolates primarily according to species identity rather than geographic origin indicates that evolutionary divergence within *Stagonosporopsis* is largely driven by species-level differentiation rather than regional isolation. This suggests that these pathogens have undergone global dispersal, likely facilitated by seed transmission and international trade, resulting in genetically similar populations across different geographic regions (Fig. 1).

The phylogenetic tree clearly separated major species complexes, including *Stagonosporopsis cucurbitacearum*, *S. citrulli*, and *S. caricae*, which are known causal agents of gummy stem blight in cucurbit crops. Isolates belonging to *S. cucurbitacearum* formed a large monophyletic clade with strong bootstrap support, indicating a high level of genetic similarity among isolates collected from different geographic regions. The formation of well-supported monophyletic clades with limited intra-species divergence suggests that *S. cucurbitacearum* is globally

distributed yet genetically conserved at the RPB2 locus. Such conservation is characteristic of essential genes under strong functional constraints, where sequence variation is restricted despite broad geographic distribution (Stewart et al., 2015; Marin-Felix et al., 2019).

In contrast, other species such as *Stagonosporopsis pogostemonis*, *S. vannaccii*, and *S. papillata* formed separate lineages within the phylogeny, reflecting their evolutionary divergence within the genus. Several isolates identified only as *Stagonosporopsis* sp. also clustered near related species groups, suggesting possible taxonomic proximity or the presence of yet-uncharacterized lineages. These findings are consistent with previous taxonomic revisions that have expanded the diversity of the genus *Stagonosporopsis* through molecular phylogenetic approaches (Aveskamp et al., 2008; Marin-Felix et al., 2019).

The presence of distinct clades corresponding to different host-associated species further highlights the evolutionary specialization within the genus. For example, species such as *S. helianthi*, *S. chrysanthemi*, and *S. trachelii* grouped together with isolates associated with their respective host plants, suggesting potential host adaptation and ecological differentiation. Such host-associated diversification has been documented in several fungal plant pathogens, where evolutionary divergence is influenced by host range and ecological factors.

Interestingly, isolates representing *Stagonosporopsis citrulli* and *S. cucurbitacearum* were placed in closely related but distinct clades, supporting previous reports that these species represent separate evolutionary lineages responsible for gummy stem blight disease. Earlier phylogenetic studies using multilocus sequence analysis (MLSA) have similarly demonstrated that the gummy stem blight pathogen complex consists of multiple cryptic species within the genus *Stagonosporopsis* (Stewart et al., 2015).

The clustering of isolates into well-supported species clades confirms the usefulness of the RPB2 locus for phylogenetic and taxonomic studies in this group. Moreover, the observed genetic structure indicates that evolutionary divergence within *Stagonosporopsis* is largely driven by species-level differentiation rather than geographic separation, highlighting the global distribution and evolutionary stability of several important phytopathogenic lineages.

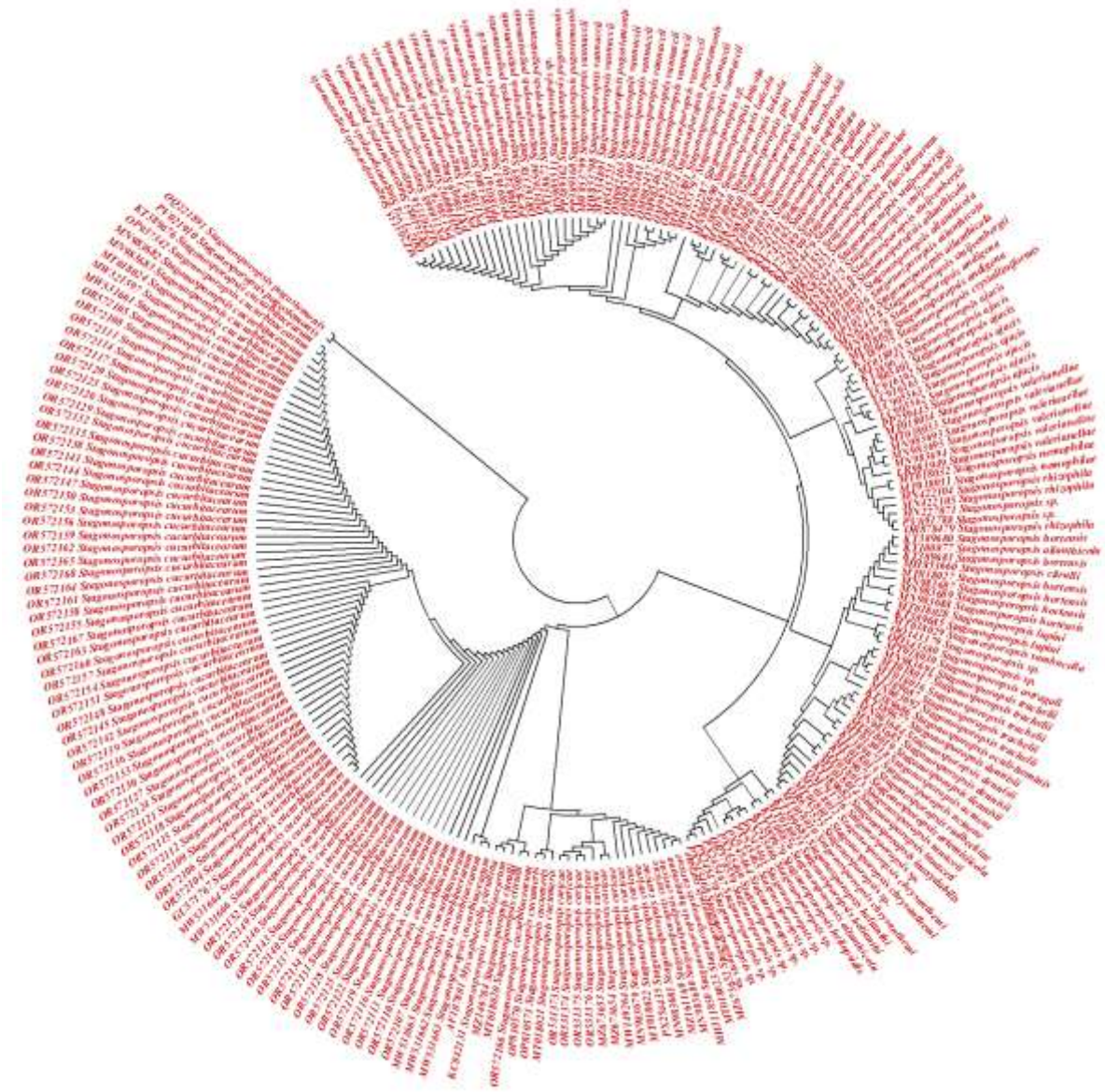


Figure 1. Maximum likelihood phylogenetic tree based on RPB2 gene sequences showing evolutionary relationships among global *Stagonosporopsis* isolates, with bootstrap support values (1000 replicates) indicated at the nodes.

Haplotype diversity and global population structure

Haplotype analysis of the RPB2 gene sequences of *Stagonosporopsis* isolates was conducted using DnaSP v6 to assess genetic diversity and population structure. The dataset comprised 234 sequences, with 171 sites analyzed after excluding gaps and missing data. A total of 126 polymorphic (segregating) sites and 164 mutations were detected, indicating considerable sequence variability within the dataset (Table 2).

The analysis identified 43 distinct haplotypes, with a high haplotype diversity (Hd = 0.861 ± 0.018), suggesting the presence of substantial genetic variation among global *Stagonosporopsis* isolates. The high haplotype diversity observed in this study reflects the presence of multiple distinct genetic lineages within global *Stagonosporopsis* populations, suggesting a long evolutionary history combined with ongoing mutation-driven diversification. The coexistence

of numerous rare haplotypes alongside a few dominant ones is indicative of population expansion following diversification, a pattern commonly observed in fungal pathogens with widespread dispersal (Nei, 1987; Rozas et al., 2017). The nucleotide diversity (π) was estimated at 0.10580, with an average number of nucleotide differences (k = 18.092) among sequences, further supporting the presence of moderate to high genetic variation within the studied populations (Table 2).

Population genetic parameters such as Watterson's theta (θW = 0.12219) and theta estimated from total mutations (θ = 0.15904) were also relatively high, indicating an appreciable level of sequence polymorphism across the RPB2 locus (Table 2). These diversity indices suggest that mutation-driven diversification may play an important role in shaping the genetic structure of *Stagonosporopsis* populations. Similar levels of genetic diversity have been reported in other fungal plant pathogens where mutation and evolutionary divergence contribute significantly to pathogen adaptation and diversification (McDonald and Linde, 2002).

Tajima's D value was negative (D = −1.04156), although statistically non-significant (P > 0.10) (Table 3). The negative Tajima's D value suggests an excess of low-frequency polymorphisms, which may reflect recent population expansion or purifying selection. However, the lack of statistical significance indicates that these deviations are not strong enough to reject the null hypothesis of neutrality. This pattern is consistent with the behavior of conserved genes such as RPB2, where evolutionary changes are constrained and largely governed by neutral processes rather than strong directional selection. Similarly, Fu and Li's neutrality tests produced non-significant values, further supporting the hypothesis that the observed genetic variation in the dataset may not be strongly influenced by recent selective sweeps or demographic bottlenecks. Neutral evolution has been commonly reported in housekeeping genes such as RPB2 that are frequently used for phylogenetic inference in fungal taxa (Liu et al., 1999; Schoch et al., 2012).

Analysis of insertion–deletion (InDel) polymorphisms revealed eight InDel events, with relatively low InDel haplotype diversity (0.017), indicating that insertion–deletion variation contributed minimally to overall genetic diversity in the dataset (Table 4). Nevertheless, the presence of multiple polymorphic sites and conserved regions suggests that the RPB2 locus retains both variable and conserved domains, making it suitable for phylogenetic and evolutionary analyses in *Stagonosporopsis*.

The high haplotype diversity coupled with moderate nucleotide diversity suggests that *Stagonosporopsis* populations possess a wide range of genetic variants distributed across different geographical regions. Such genetic diversity may facilitate adaptation to diverse environmental conditions and host plants. Similar patterns of global genetic variation have been documented in several plant pathogenic fungi, where gene flow and mutation contribute to the emergence of diverse pathogen lineages (McDonald and Linde, 2002).

Table 2. Genetic diversity parameters of global *Stagonosporopsis* isolates based on RPB2 gene sequences, including haplotype diversity and nucleotide diversity indices

Parameter	Value
Number of sequences (N)	234
Number of analyzed sites	171
Polymorphic sites (S)	126

Total mutations (Eta)	164
Number of haplotypes (H)	43
Haplotype diversity (Hd)	0.861 ± 0.018
Nucleotide diversity (π)	0.10580
Average nucleotide differences (k)	18.092
Watterson's theta (θW)	0.12219
Theta from mutations (θ)	0.15904

Table 3. Neutrality test statistics (Tajima's D, Fu's Fs, and Fu & Li's tests) for RPB2 gene sequences of *Stagonosporopsis* isolates.

Test	Value	Significance
Tajima's D	-1.04156	Not significant
Fu's Fs	1.696	Not significant
Fu & Li's D*	-0.402	Not significant
Fu & Li's F*	-0.853	Not significant

Table 4. Summary of insertion–deletion (InDel) polymorphism statistics in RPB2 gene sequences of *Stagonosporopsis* isolates.

Parameter	Value
Total InDel events	8
Average InDel length	10.419
InDel haplotypes	2
InDel haplotype diversity	0.017
InDel diversity per site	0.00055

Haplotype network analysis

Haplotype analysis of the RPB2 gene sequences of *Stagonosporopsis* isolates revealed substantial genetic variation across the global dataset. A total of 43 haplotypes were identified among the 234 sequences analyzed, indicating the presence of multiple genetically distinct lineages within the species complex. The haplotype frequency analysis showed that a few haplotypes were dominant within the dataset, whereas many haplotypes occurred at low frequencies, suggesting the coexistence of common ancestral haplotypes and several rare or recently derived variants. In particular, one major haplotype was represented by 77 isolates, followed by another common haplotype with 34 isolates, while several haplotypes were represented by only one or two sequences (Fig. 2). The detailed distribution of Stagonosporopsis isolates across haplotypes is provided in Supplementary Table S3. The observed haplotype distribution, characterized by a few dominant haplotypes and numerous low-frequency variants, reflects a typical star-like network structure associated with population expansion. The presence of central, high-frequency haplotypes likely represents ancestral lineages, whereas peripheral haplotypes indicate recent mutational events. This pattern supports the hypothesis of ongoing diversification within a globally connected population (Nei, 1987).

The median-joining haplotype network further illustrated the genealogical relationships among the detected haplotypes. The network displayed a complex topology with multiple mutational connections among haplotypes and the presence of several peripheral haplotypes derived from central nodes (**Fig. 3**). The central haplotypes, which were represented by the highest frequencies, likely correspond to ancestral or widely distributed lineages of *Stagonosporopsis*. In contrast, the numerous low-frequency haplotypes connected through few mutational steps suggest the occurrence of recent diversification events within the pathogen population. The absence of clear geographic clustering in the haplotype network suggests extensive gene flow among global populations of *Stagonosporopsis*. Such genetic homogenization is likely driven by anthropogenic factors, including the international movement of infected seeds and plant materials, which facilitate long-distance dispersal of pathogen lineages.

Analysis of molecular variance (AMOVA) provided further insights into the population genetic structure of the pathogen. The results indicated that the majority of genetic variation (92.06%) occurred within populations, whereas only 7.94% of the variation was attributed to differences among populations (**Table 5**). Using microsatellite markers, Brewer et al. (2015) similarly reported high genetic diversity within populations of *Stagonosporopsis* species, supporting our finding that most variation occurs within rather than among geographic populations. The fixation index ($\Phi ST = 0.079$) suggested relatively low genetic differentiation among global populations. Moreover, the permutation test showed that the ΦST value was not statistically significant ($P = 0.081$), indicating that population subdivision was weak. These findings suggest that *Stagonosporopsis* populations from different geographical regions are genetically similar and may share a common evolutionary history.

The predominance of within-population variation is consistent with patterns reported in many plant pathogenic fungi, where mutation, recombination, and dispersal contribute to high levels of genetic diversity within populations while maintaining relatively low differentiation among geographic regions (McDonald and Linde, 2002). The observed high haplotype diversity and low population differentiation are consistent with the presence of both mating types in global populations, which can facilitate gene flow and reduce genetic drift (Li et al., 2017). At a finer spatial scale, Li et al. (2016) found significant population structuring among watermelon fields in the southeastern United States, suggesting that while global gene flow is extensive, local selection and limited dispersal can generate differentiation at the field level. Similarly, a population study of *Stagonosporopsis* species in Korea using β-tubulin and microsatellite markers revealed high genetic diversity within populations and identified a new lineage of *S. cucurbitacearum*, suggesting regional diversification (Jeong et al., 2022). Such patterns may also reflect the movement of infected plant materials, seed transmission, or global agricultural trade, which facilitate the dissemination of pathogen lineages across continents.

The high genetic diversity observed in global Stagonosporopsis populations has direct implications for disease management, as it facilitates the rapid evolution of resistance to systemic fungicides. Indeed, Li et al. (2019) documented differences in fungicide resistance profiles between S. citrulli and S. cucurbitacearum, including multiple resistance to QoIs (azoxystrobin), SDHIs (boscalid, fluopyram), and DMIs (tebuconazole), with resistance levels varying both within and between species. These findings underscore the need for integrated management strategies that rotate fungicides with different modes of action and incorporate resistance monitoring.

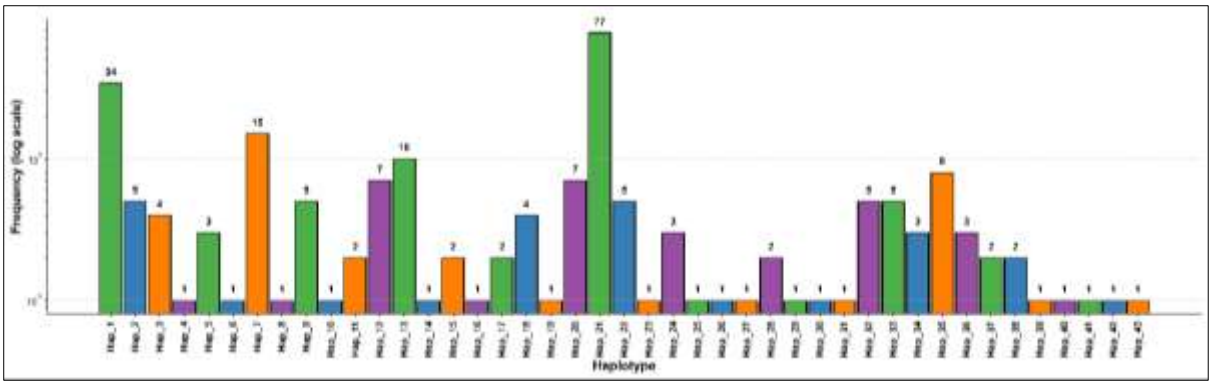


Figure 2. Frequency distribution of haplotypes identified from RPB2 gene sequences of global *Stagonosporopsis* isolates, showing the relative abundance of each haplotype.

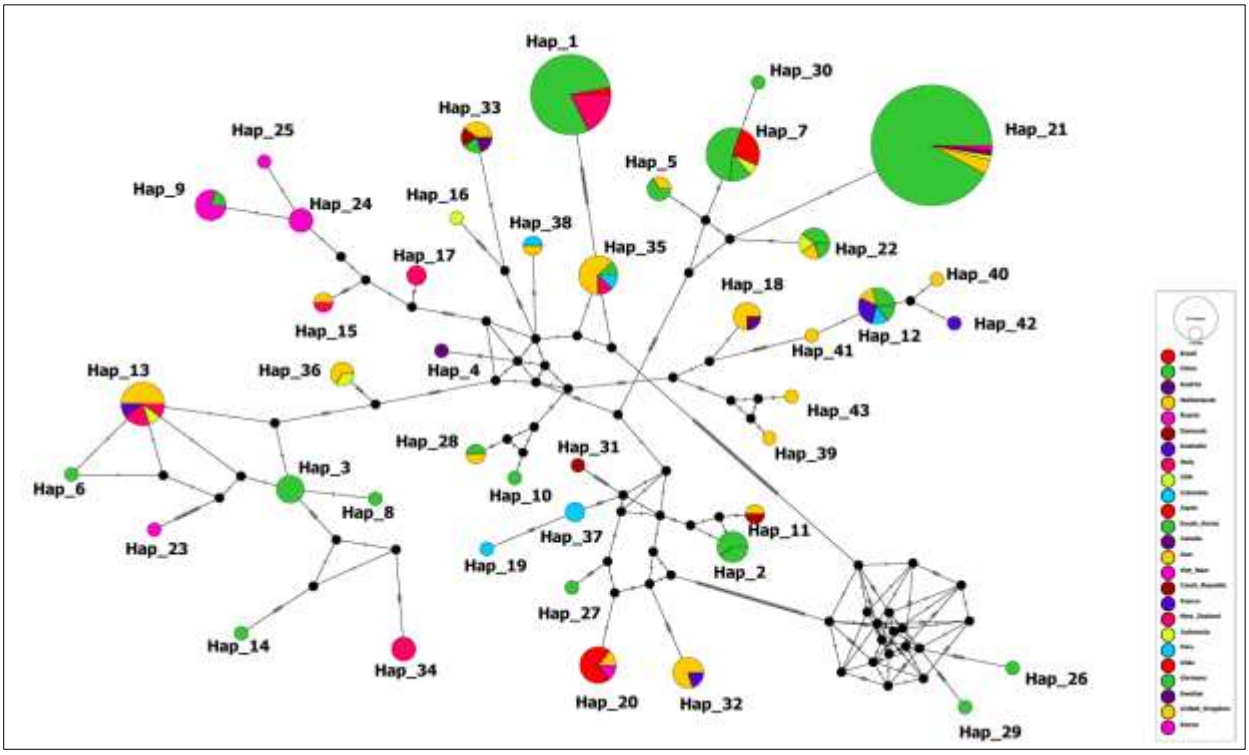


Figure 3. Median-joining haplotype network based on RPB2 gene sequences of global *Stagonosporopsis* isolates, illustrating genealogical relationships among haplotypes; circle size is proportional to haplotype frequency, and connecting lines represent mutational steps.

Table 5. Analysis of molecular variance (AMOVA) of global *Stagonosporopsis* populations based on RPB2 gene sequences, showing partitioning of genetic variation within and among populations.

Source of variation	df	Sum of squares	Variance components (σ^2)	Percentage of variation (%)
Among populations	24	10,068.052	19.652	7.94

Within populations	209	47,629.713	227.893	92.06
Total	233	57,697.765	247.546	100

Comprehensive analysis of codon usage pattern

Nucleotide Composition and Base Usage Bias

The nucleotide composition of the coding sequences of the RNA polymerase II second largest subunit (RPB2) gene from *Stagonosporopsis* isolates was analyzed to investigate the underlying genomic features influencing codon usage patterns. The overall nucleotide composition, including GC content and codon position-specific nucleotide frequencies, is summarized in **Supplementary Table S4**. The analysis revealed moderate GC content across the RPB2 sequences, with noticeable variation among the three codon positions.

In particular, the third codon position (GC3) showed greater variability compared with the first (GC1) and second (GC2) positions, indicating that synonymous sites are more susceptible to nucleotide substitution events. Such variation at the third codon position is commonly observed in fungal genomes and is considered a major determinant of codon usage bias (Sueoka, 1988; Hershberg and Petrov, 2008). Codon usage patterns are also influenced by environmental conditions, genome size, and gene expression dynamics, reflecting both adaptive and neutral evolutionary processes (Ling et al., 2024). Moreover, elevated GC3 variability is often associated with mutational bias rather than strong translational selection, reinforcing the role of neutral evolutionary processes in determining codon preference. Similar nucleotide composition patterns have been reported in fungal pathogens where base composition plays a key role in shaping codon preference (Sharp et al., 2010).

Codon Usage Bias Indices

Codon usage bias in the RPB2 gene sequences was evaluated using several commonly used indices, including the effective number of codons (ENC), codon adaptation index (CAI), codon bias index (CBI), and frequency of optimal codons (FOP). The calculated values for these indices are summarized in Supplementary Table S5.

The ENC values obtained in this study ranged within the typical range associated with moderate codon usage bias, suggesting that RPB2 sequences do not exhibit extremely strong codon preference. ENC values theoretically range from 20 (extreme bias) to 61 (no bias), and values observed in this dataset indicate that codon usage bias is present but not highly constrained.

Correlation analysis among nucleotide composition and codon bias indices further revealed significant relationships among several parameters (Fig. 4). For example, strong positive correlations were observed between CAI, CBI, and FOP, indicating that these metrics capture related aspects of codon usage bias. Conversely, negative correlations were observed between GC1 and codon bias indices, suggesting differential evolutionary constraints across codon positions. Similar correlations between codon bias indices and nucleotide composition have been widely reported in microbial genomes.

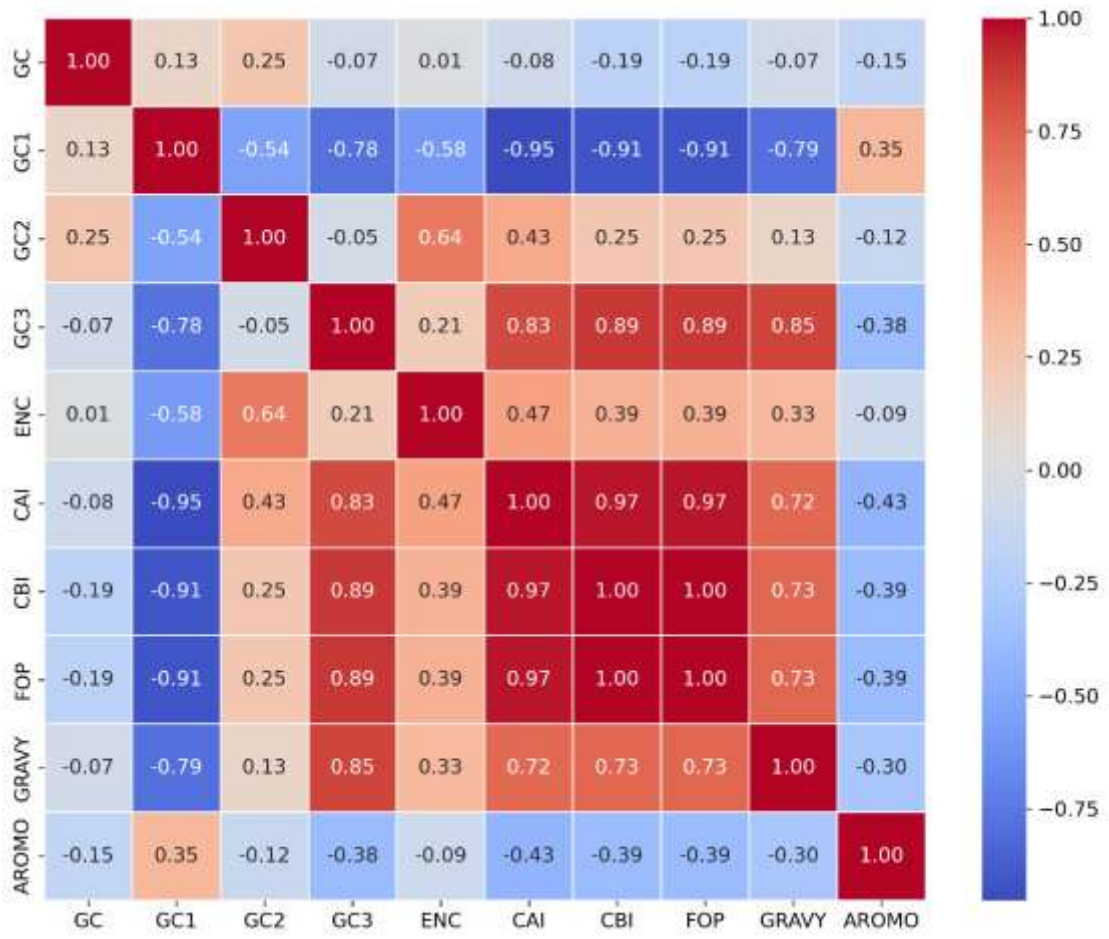


Figure 4. Correlation matrix showing relationships among nucleotide composition parameters and codon usage bias indices (ENC, CAI, CBI, and FOP) in RPB2 gene sequences of *Stagonosporopsis* isolates.

Relative Synonymous Codon Usage (RSCU) Patterns

To further investigate codon preference patterns, Relative Synonymous Codon Usage (RSCU) values were calculated for all synonymous codons in the RPB2 gene sequences. Codons with RSCU values greater than 1.0 were considered preferentially used, whereas those with values less than 1.0 were underrepresented. The distribution of RSCU values across codons is visualized in the heatmap (Fig. 5), which highlights the codon preference landscape of the RPB2 gene.

The heatmap demonstrates that several synonymous codons show clear preferential usage, suggesting that codon selection is not entirely random. Codon preference is often associated with translational efficiency and tRNA abundance, influencing gene expression levels as well as the efficiency and accuracy of protein synthesis (Sharp and Li, 1987; dos Reis et al., 2004).

The variability of codon usage among codons is further illustrated in the variability plot (Fig. 6), which shows differences in standard deviation across codons. Codons exhibiting high variability may reflect differential selective pressures across isolates or adaptive divergence within the species

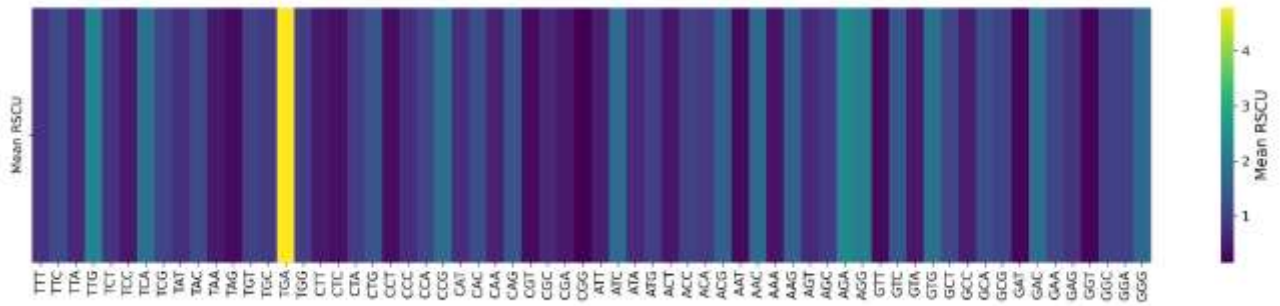


Figure 5. Heatmap of relative synonymous codon usage (RSCU) values for RPB2 gene sequences of *Stagonosporopsis* isolates, illustrating codon preference patterns.

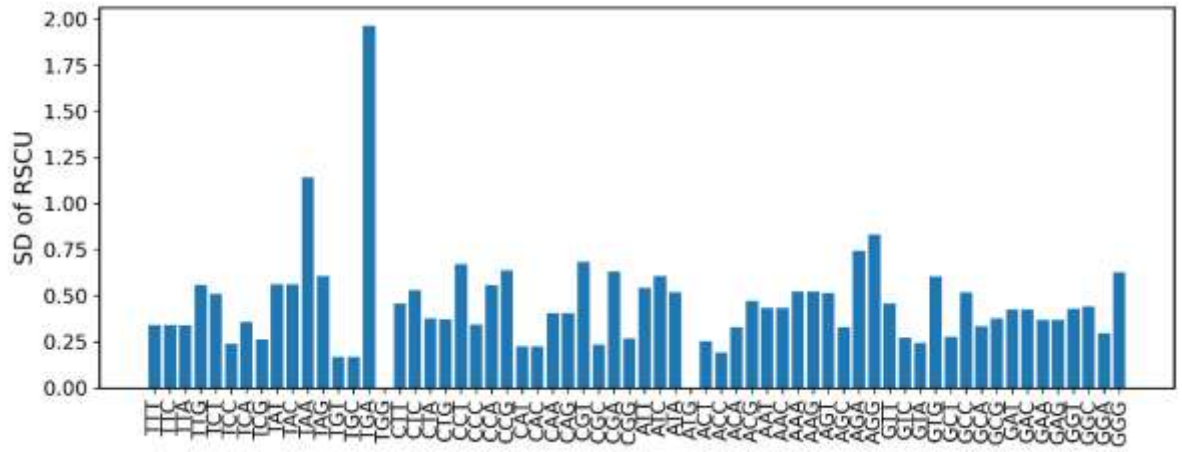


Figure 6. Variability in codon usage among synonymous codons in RPB2 gene sequences of *Stagonosporopsis* isolates, represented by standard deviation of RSCU values.

Optimal Codon Identification

Optimal codons were identified by comparing codon usage patterns between sequences with relatively high and low codon bias. Codons exhibiting significantly higher usage frequencies in highly biased genes were considered optimal codons. Several codons exceeded the ΔRSCU threshold, indicating strong preferential usage. These optimal codons may correspond to abundant tRNAs and could potentially enhance translational efficiency.

Optimal codon identification is widely used to infer translational selection in microbial genomes, as highly expressed genes often preferentially use codons corresponding to abundant tRNA species (Sharp et al., 2010).

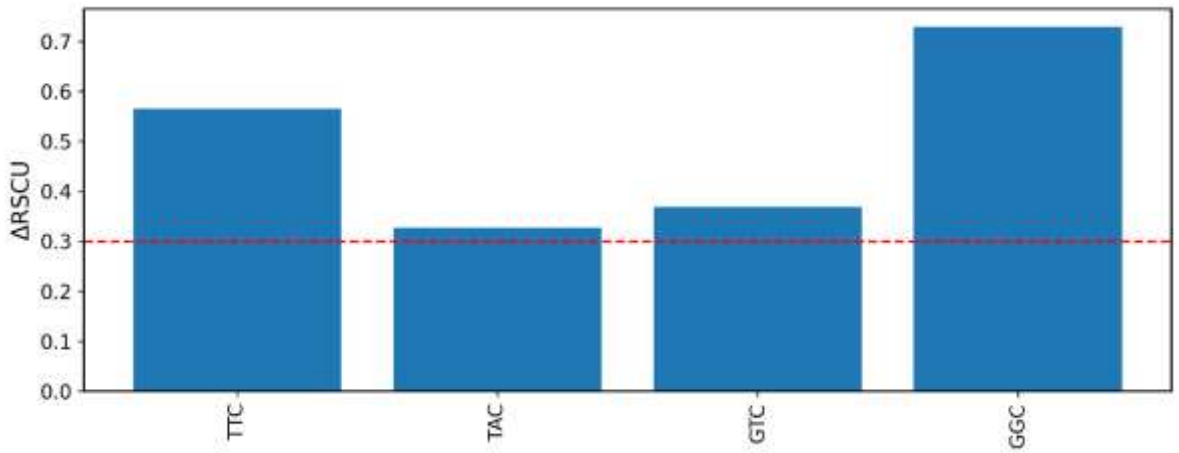


Figure 7. Identification of optimal codons in RPB2 gene sequences of *Stagonosporopsis* isolates based on differences in relative synonymous codon usage (ΔRSCU) between high- and low-bias gene groups.

Influence of Mutational Pressure: ENC–GC3 Analysis

To investigate the influence of mutational pressure on codon usage bias, an ENC–GC3 plot was constructed (Fig. 8). Most data points are located below the expected curve, suggesting that factors beyond simple nucleotide composition influence codon usage patterns in the RPB2 gene. Deviation from the expected curve typically indicates the involvement of natural selection or other evolutionary forces in shaping codon usage (Wright, 1990).

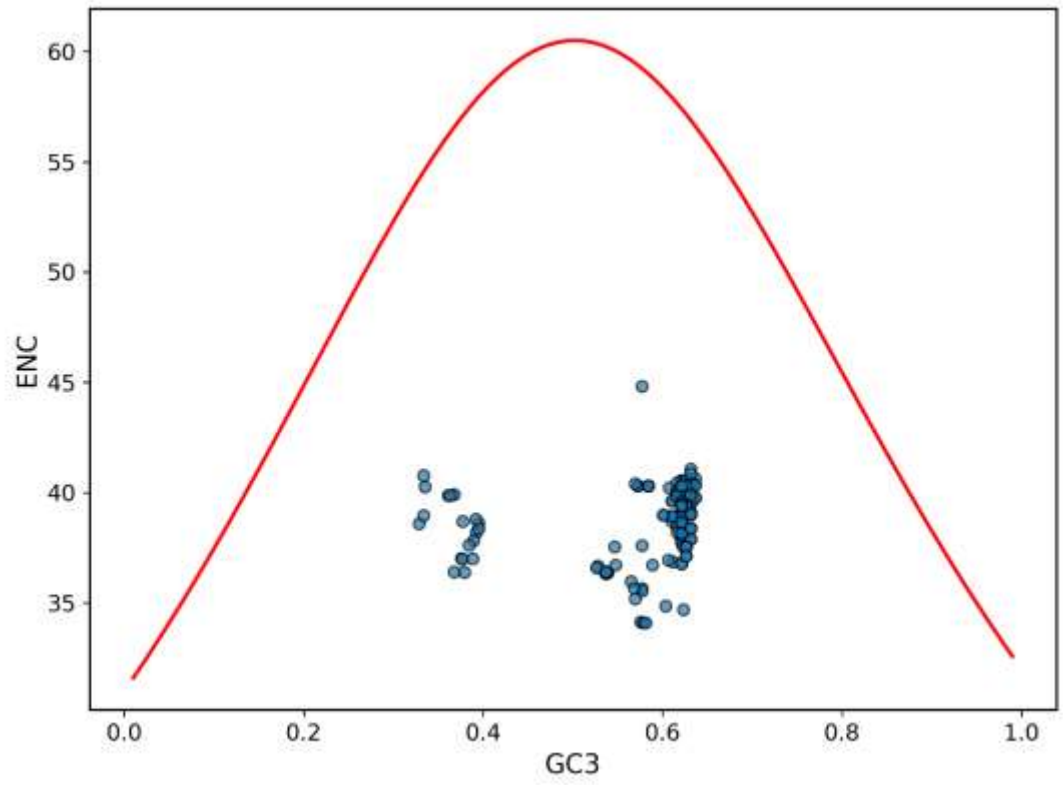


Figure 8. ENC–GC3 plot showing the relationship between effective number of codons (ENC) and GC content at the third codon position (GC3) for RPB2 gene sequences of *Stagonosporopsis* isolates; the expected curve represents codon usage under mutational pressure alone.

Neutrality Plot Analysis

Neutrality plot analysis was performed to further examine the relative contributions of mutation pressure and natural selection to codon usage evolution (Fig. 9). A strong negative correlation was observed between GC12 and GC3, with a regression equation of GC12 = −0.499 GC3 + 76.50 and a high coefficient of determination ($R^2 = 0.918$). This strong relationship indicates that mutation pressure contributes significantly to codon usage patterns in the RPB2 gene. However, the slope of the regression line suggests that selective constraints may also influence nucleotide composition at the first and second codon positions (Sueoka, 1988).

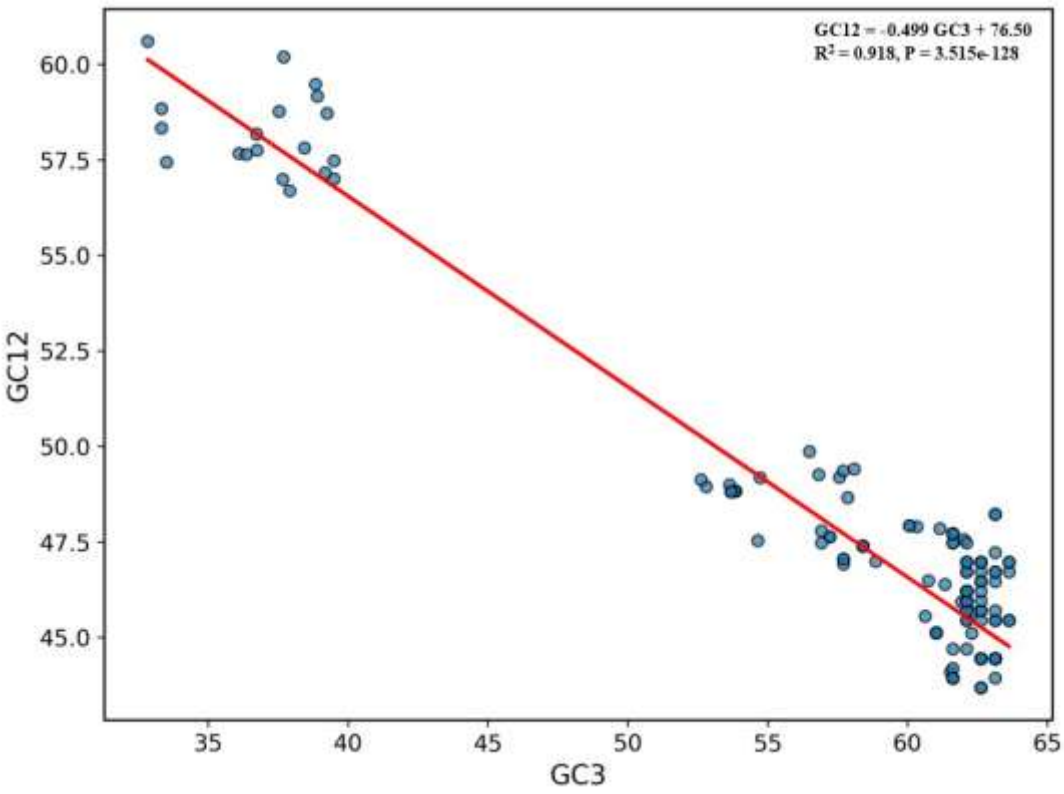


Figure 9. Neutrality plot showing the relationship between GC content at the first and second codon positions (GC12) and the third codon position (GC3) in RPB2 gene sequences of *Stagonosporopsis* isolates, used to assess the relative influence of mutation pressure and natural selection.

Parity Rule 2 (PR2) Bias Analysis

Parity rule 2 (PR2) analysis was conducted to evaluate nucleotide bias at synonymous sites (Fig. 10). Under PR2 expectations, complementary nucleotides (A=T and G=C) should occur at equal frequencies in the absence of selection or mutational bias. Several sequences deviate from the central equilibrium point (0.5, 0.5), indicating unequal usage of complementary nucleotides at the third codon position. Such deviations suggest that codon usage patterns are influenced by a combination of mutational pressure and selective constraints.

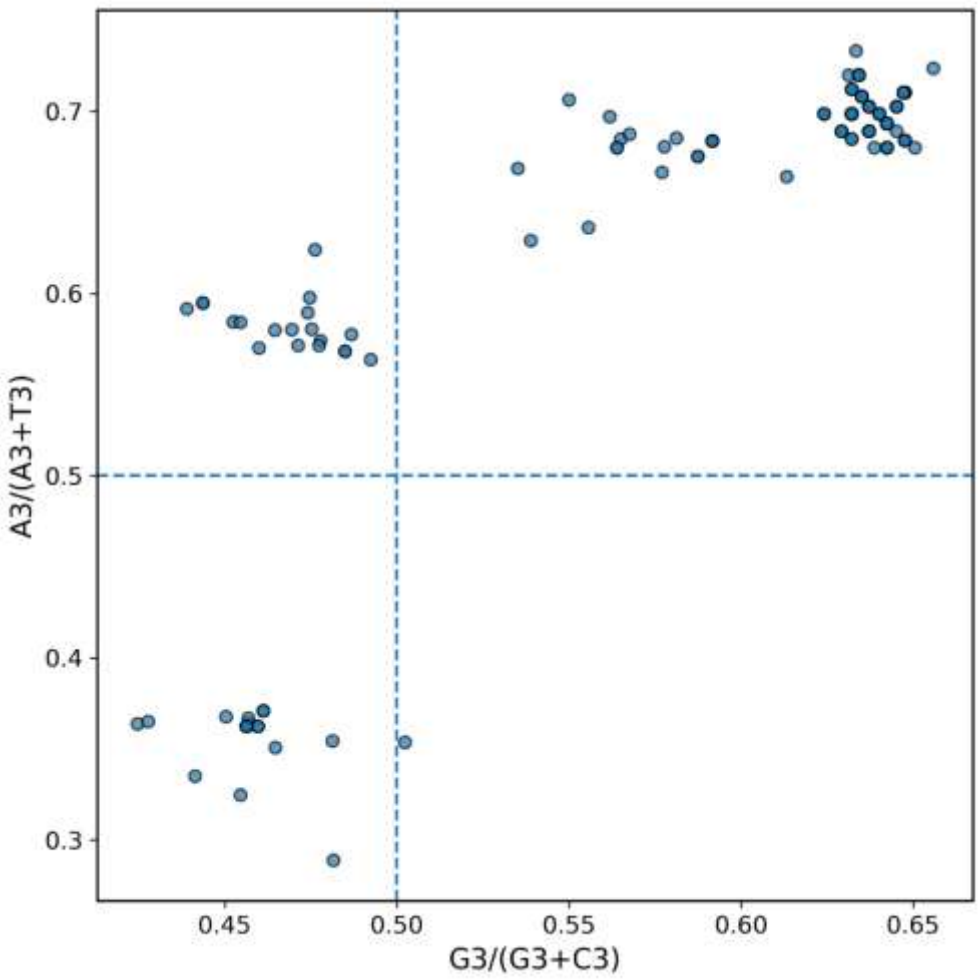


Figure 10. Parity rule 2 (PR2) bias plot showing the distribution of A3/(A3+T3) versus G3/(G3+C3) in RPB2 gene sequences of *Stagonosporopsis* isolates; deviation from the center (0.5, 0.5) indicates nucleotide usage bias at synonymous sites.

Correspondence Analysis of Codon Usage

Correspondence analysis (COA) was performed to explore major trends in codon usage variation among RPB2 sequences (Fig. 11), where sequences are distributed along principal axes representing major sources of variation in codon usage patterns. The clustering pattern observed in the COA plot indicates the presence of distinct codon usage patterns among groups of sequences. Codon usage patterns can vary not only between species but also among genes within the same genome, reflecting functional specialization and evolutionary constraints (Baeza et al., 2024). Color coding based on GC3 values further suggests that nucleotide composition contributes significantly to the observed variation in codon usage (Sharp et al., 2010).

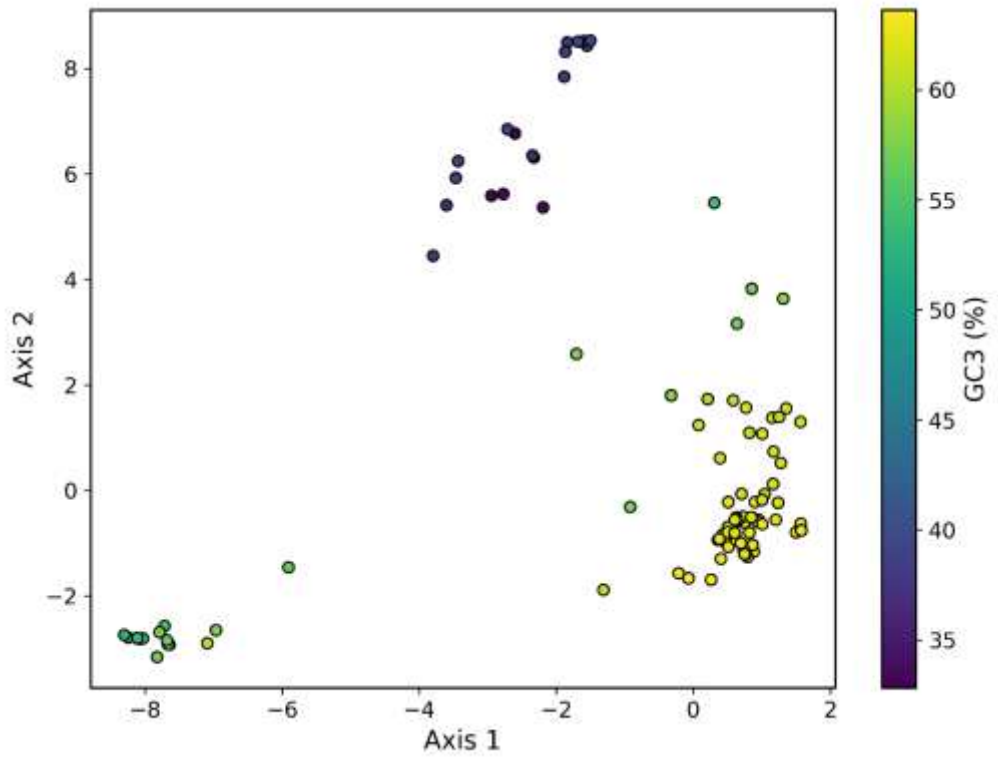


Figure 11. Correspondence analysis (COA) plot based on RSCU values showing major trends in codon usage variation among RPB2 gene sequences of *Stagonosporopsis* isolates, with axes representing principal components of variation.

Relationship Between Codon Bias and Gene Expression

To assess the relationship between codon bias and gene expression potential, the association between ENC and CAI values and GC3 and CAI were analyzed (Fig. 12 and Fig. 13). These analyses suggest that sequences with higher CAI values tend to exhibit slightly lower ENC values, indicating increased codon bias. This trend suggests that translational selection may influence codon preference in the RPB2 gene.

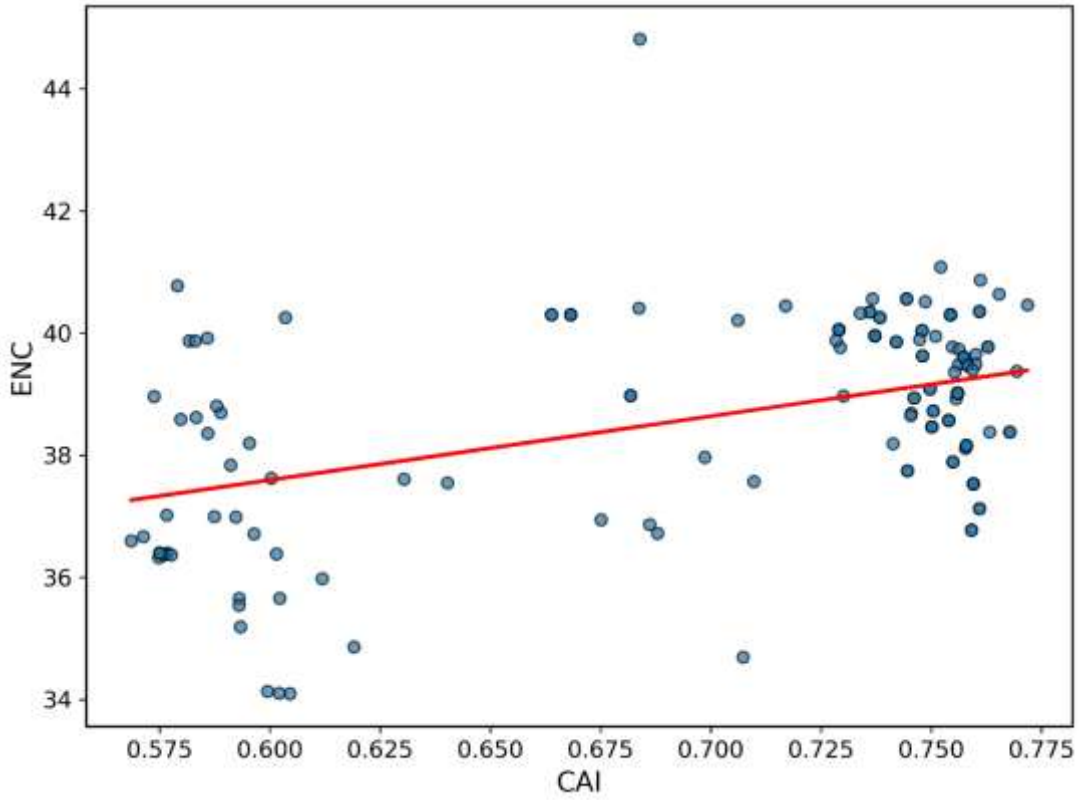


Figure 12. Scatter plot showing the relationship between effective number of codons (ENC) and codon adaptation index (CAI) for RPB2 gene sequences of *Stagonosporopsis* isolates.

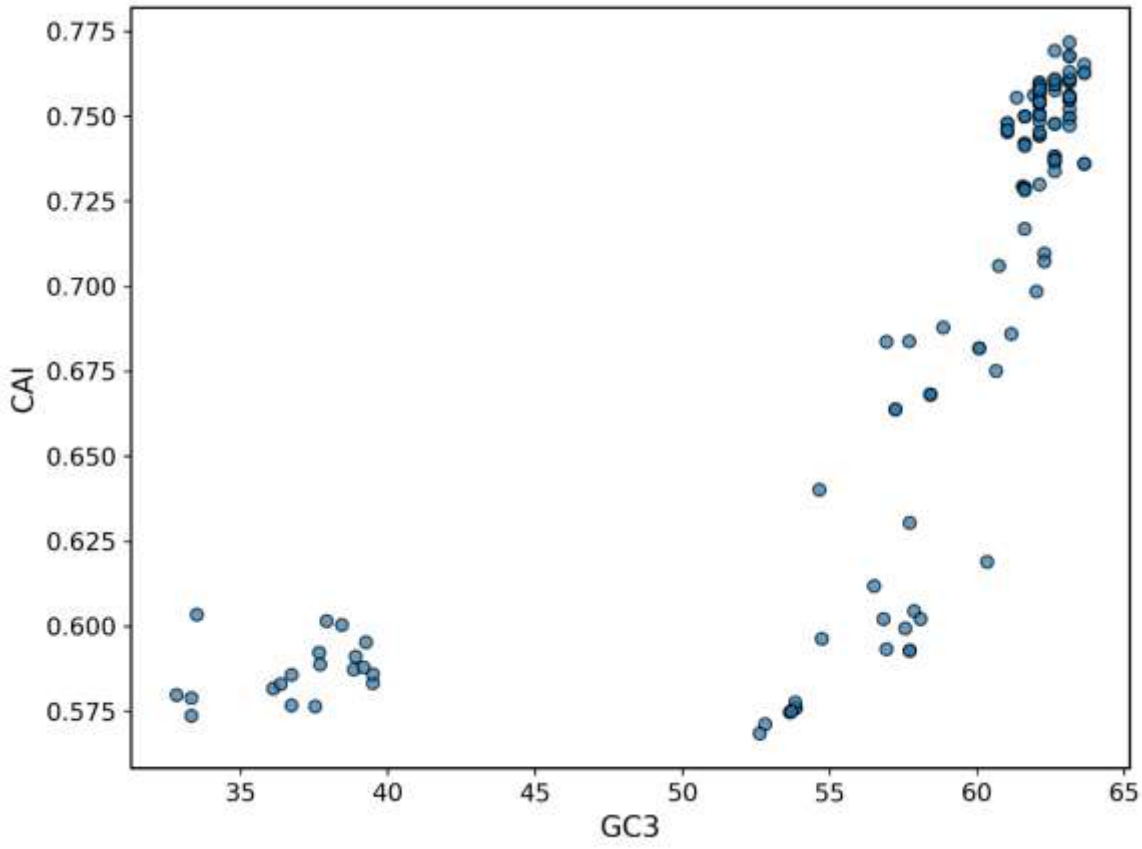


Figure 13. Scatter plot showing the relationship between GC content at the third codon position (GC3) and codon adaptation index (CAI) for RPB2 gene sequences of *Stagonosporopsis* isolates.

Dinucleotide Bias Analysis

Dinucleotide relative abundance analysis was conducted to investigate potential biases in dinucleotide usage within the RPB2 gene sequences (Fig. 14). The heatmap indicates uneven distribution of dinucleotide frequencies, suggesting that certain dinucleotide combinations are either overrepresented or underrepresented in the RPB2 gene sequences. Such biases can arise from mutation patterns, genome composition, or selective pressures acting on coding sequences (Kariin and Burge, 1995).

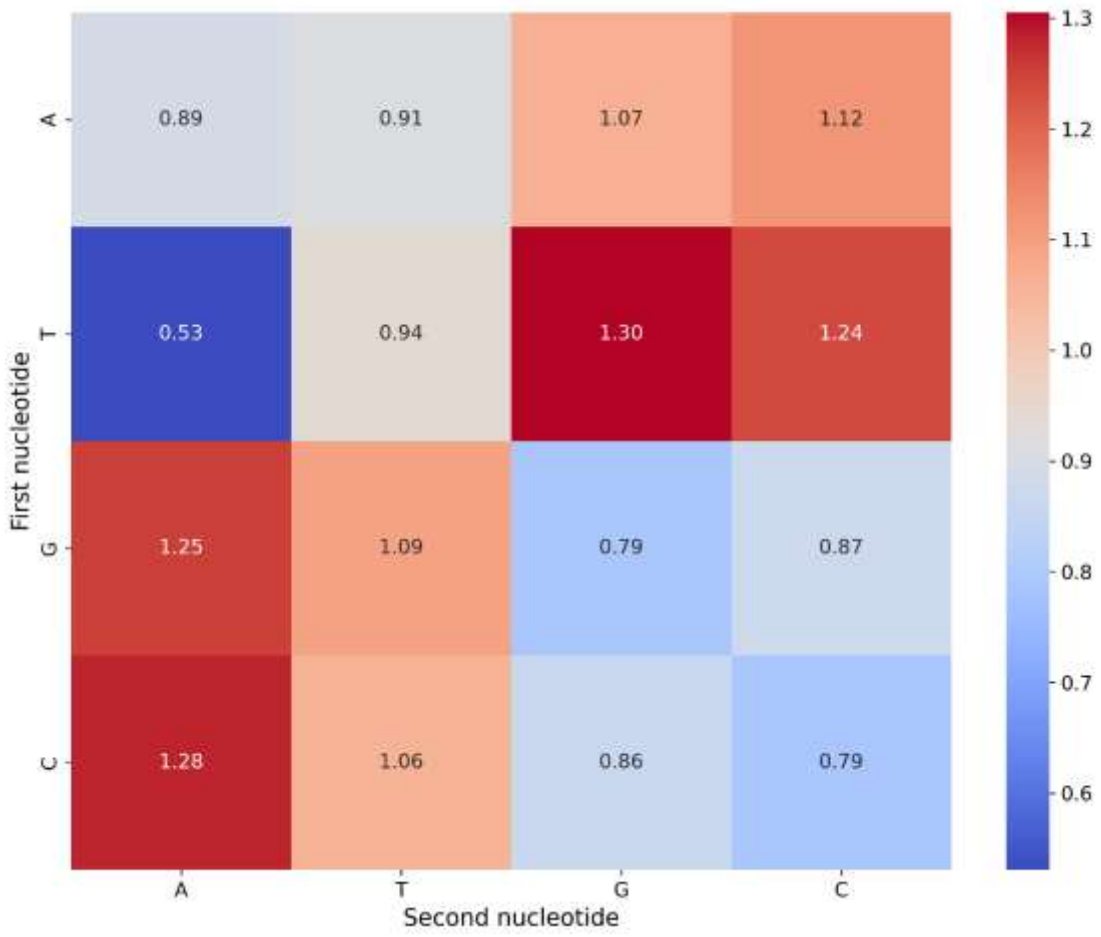


Figure 14. Heatmap of dinucleotide relative abundance in RPB2 gene sequences of *Stagonosporopsis* isolates, highlighting overrepresented and underrepresented dinucleotide combinations.

Collectively, the results demonstrate that codon usage patterns in the RPB2 gene of *Stagonosporopsis* are primarily shaped by mutational pressure, with additional contributions from natural selection and translational constraints. The interplay between these forces reflects an evolutionary balance between maintaining essential gene function and accommodating nucleotide composition bias. This pattern is consistent with observations in fungal genomes, where codon usage bias is influenced by a combination of mutational bias, translational selection, tRNA availability, and GC content (Sharp et al., 2010; Hershberg and Petrov, 2008; Zhao et al., 2020; Wint et al., 2022; Zavala et al., 2024). Codon usage bias has also been shown to reflect adaptive evolutionary processes, including responses to environmental and cellular constraints (Yang et al., 2023). These findings highlight the importance of integrating codon usage analysis with phylogenetic and population genetic approaches to better understand pathogen evolution and adaptation.

CONCLUSION

This study provides a comprehensive analysis integrating phylogenetic relationships, population genetics, and codon usage bias of global Stagonosporopsis species associated with

gummy stem blight. The findings demonstrate that the *RPB2* locus serves as a reliable phylogenetic marker, clearly resolving major species complexes with high bootstrap support. The high haplotype diversity coupled with low population differentiation suggests historical gene flow and global dispersal of these pathogens, likely facilitated by seed transmission and agricultural trade. Codon usage analysis revealed that mutational pressure is the primary evolutionary force shaping codon usage patterns in the *RPB2* gene, with secondary influences from natural selection. The identification of optimal codons and the observed deviations in ENC-GC3, neutrality, and PR2 plots collectively indicate that both compositional constraints and selective forces contribute to the molecular evolution of these fungi. These findings enhance our understanding of the global genetic structure and evolutionary dynamics of *Stagonosporopsis* species, providing valuable insights for disease management strategies, resistance breeding programs, and future epidemiological studies of this economically important pathogen complex.

Author contributions

All authors contributed to the study conception and design. B.P. Maruthi Prasad and J. Harish conceived and designed the study. B.P. Maruthi Prasad performed the data analysis. The manuscript was drafted by B.P. Maruthi Prasad and M. Pallavi. All authors read and approved the final manuscript.

Data availability

The sequence datasets analyzed in this study were retrieved from the NCBI GenBank database, and accession numbers are provided in Supplementary Table S1.

Conflicts of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding

No funding received.

Acknowledgment

The first author gratefully acknowledges the Council of Scientific and Industrial Research (CSIR), New Delhi, India, for awarding a Senior Research Fellowship (SRF) under grant number 09/0271(17512)/2024-EMR-I to pursue a Ph.D. program at the University of Agricultural Sciences, Bangalore.

REFERENCES

Aveskamp MM, De Gruyter J, Crous PW (2008) Biology and recent developments in the systematics of *Phoma*, a complex genus of major quarantine significance. **Fungal Divers** 31:1–18. http://www.fungaldiversity.org/fdp/sfdp/31-1.pdf

Baeza M, Sepulveda D, Cifuentes V, Alcaíno J (2024) Codon usage bias in yeasts and its correlation with gene expression, growth temperature, and protein structure. **Front Microbiol** 15:1414422. https://doi.org/10.3389/fmicb.2024.1414422

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. **Mol Biol Evol** 16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Benson DA, Cavanaugh M, Clark K et al (2018) GenBank. **Nucleic Acids Res** 46:D41–D47. https://doi.org/10.1093/nar/gkx1094

Brewer MT, Rath M, Li HX (2015) Genetic diversity and population structure of cucurbit gummy stem blight fungi based on microsatellite markers. **Phytopathology** 105:815–824. https://doi.org/10.1094/PHYTO-10-14-0282-R

Cafà G, Boufleur TR, Linhares de Castro RR et al (2020) Genome sequence data of the soybean pathogen *Stagonosporopsis vannaccii*: a resource for studies on Didymellaceae evolution. **Mol Plant Microbe Interact** 33:1022–1024. https://doi.org/10.1094/MPMI-01-20-0016-A

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Res** 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics** 131:479–491.

Fu Y, Liang F, Li C et al (2023) Codon usage bias analysis in macronuclear genomes of ciliated protozoa. **Microorganisms** 11:1833. https://doi.org/10.3390/microorganisms11071833

Gladieux P, Ropars J, Badouin H et al (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. **Mol Ecol** 23:753–773. https://doi.org/10.1111/mec.12631

Han L, Jia X, Chen R, Liu C (2022) Analysis of codon usage bias of the genome of medicinal fungus *Wolfiporia cocos*. **J Agric Sci Technol** 24:83. https://doi.org/10.13304/j.nykjdb.2020.0827

Harris CR, Millman KJ, Van Der Walt SJ et al (2020) Array programming with NumPy. **Nature** 585:357–362. https://doi.org/10.1038/s41586-020-2649-2

Hershberg R, Petrov DA (2008) Selection on codon bias. **Annu Rev Genet** 42:287–299. https://doi.org/10.1146/annurev.genet.42.110807.091442

Hoang DT, Chernomor O, Von Haeseler A et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. **Mol Biol Evol** 35:518–522. https://doi.org/10.1093/molbev/msx281

Hunter JD (2007) Matplotlib: A 2D graphics environment. **Comput Sci Eng** 9:90–95. https://doi.org/10.1109/MCSE.2007.55

Jeong YJ, Kwon OK, Jeong AR et al (2022) Population structure of *Stagonosporopsis* species associated with cucurbit gummy stem blight in Korea. **Plant Pathol J** 38:522. https://doi.org/10.5423/PPJ.OA.07.2022.0102

Kalyaanamoorthy S, Minh BQ, Wong TK et al (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. **Nat Methods** 14:587–589. https://doi.org/10.1038/nmeth.4285

Karlin S, Burge C (1995) Dinucleotide relative abundance extremes: a genomic signature. **Trends Genet** 11:283–290. https://doi.org/10.1016/S0168-9525(00)89076-9

Keinath AP (2011) From native plants in central Europe to cultivated crops worldwide: the emergence of *Didymella bryoniae* as a cucurbit pathogen. **HortScience** 46:532–535.

Leigh JW, Bryant D, Nakagawa S (2015) POPART: full-feature software for haplotype network construction. **Methods Ecol Evol** 6. https://doi.org/10.1111/2041-210X.12410

Li HX, Brewer MT (2016) Spatial genetic structure and population dynamics of gummy stem blight fungi within and among watermelon fields in the southeastern United States. **Phytopathology** 106:900–908. https://doi.org/10.1094/PHYTO-01-16-0006-R

Li HX, Gottilla TM, Brewer MT (2017) Organization and evolution of mating-type genes in three *Stagonosporopsis* species causing gummy stem blight of cucurbits and leaf spot and dry rot of papaya. **Fungal Biol** 121:849–857. https://doi.org/10.1016/j.funbio.2017.06.002

Li HX, Nuckols TA, Harris D et al (2019) Differences in fungicide resistance profiles and multiple resistance to a quinone-outside inhibitor (QoI), two succinate dehydrogenase inhibitors (SDHI), and a demethylation inhibitor (DMI) for two *Stagonosporopsis* species causing gummy stem blight of cucurbits. **Pest Manag Sci** 75:3093–3101. https://doi.org/10.1002/ps.5426

Li Q, Luo Y, Sha A et al (2023) Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. **Front Microbiol** 14:1134228. https://doi.org/10.3389/fmicb.2023.1134228

Ling L, Zhang S, Yang T (2024) Analysis of codon usage bias in chloroplast genomes of *Dryas octopetala* var. *asiatica* (Rosaceae). **Genes** 15:899. https://doi.org/10.3390/genes15070899

Liu YJ, Whelen S, Hall BD (1999) Phylogenetic relationships among ascomycetes: evidence from an RNA polymerase II subunit. **Mol Biol Evol** 16:1799–1808. https://doi.org/10.1093/oxfordjournals.molbev.a026092

Marin-Felix Y, Hernández-Restrepo M, Iturrieta-González I et al (2019) Genera of phytopathogenic fungi: GOPHY 3. **Stud Mycol** 94:1–124. https://doi.org/10.1016/j.simyco.2019.05.001

Martin DP, Murrell B, Golden M et al (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. **Virus Evol** 1:vev003. https://doi.org/10.1093/ve/vev003

McDonald BA, Linde C (2002) The population genetics of plant pathogens and breeding strategies for durable resistance. **Euphytica** 124:163–180.

Minh BQ, Schmidt HA, Chernomor O et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. **Mol Biol Evol** 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Naqvi SAH, Abbas A, Hasnain A et al (2025) Advancing fungal phylogenetics: integrating modern sequencing, dark taxa discovery, and machine learning. **Arch Microbiol** 207:192. https://doi.org/10.1007/s00203-025-04392-2

Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press. https://doi.org/10.7312/nei-92038

Nguyen LT, Schmidt HA, Von Haeseler A et al (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. **Mol Biol Evol** 32:268–274. https://doi.org/10.1093/molbev/msu300

Parvathy ST, Udayasuriyan V, Bhadana V (2022) Codon usage bias. **Mol Biol Rep** 49:539–565. https://doi.org/10.1007/s11033-021-06749-4

Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bias. **Nat Rev Genet** 12:32–42. https://doi.org/10.1038/nrg2899

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. **Mol Biol Evol** 19:708–717.

Reis MD, Savva R, Wernisch L (2004) Solving the riddle of codon usage preferences: a test for translational selection. **Nucleic Acids Res** 32:5036–5044. https://doi.org/10.1093/nar/gkh834

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. **Mol Biol Evol** 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Schoch CL, Seifert KA, Huhndorf S et al (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. **Proc Natl Acad Sci USA** 109:6241–6246. https://doi.org/10.1073/pnas.1117018109

Seblani R, Keinath AP, Munkvold G (2023) Gummy stem blight: one disease, three pathogens. **Mol Plant Pathol** 24:825–837. https://doi.org/10.1111/mpp.13339

Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. **Philos Trans R Soc B** 365:1203–1212. https://doi.org/10.1098/rstb.2009.0305

Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. **J Mol Evol** 24:28–38. https://doi.org/10.1007/BF02099948

Sharp PM, Li WH (1987) The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic Acids Res** 15:1281–1295. https://doi.org/10.1093/nar/15.3.1281

Stewart JE, Turner AN, Brewer MT (2015) Evolutionary history and variation in host range of three *Stagonosporopsis* species causing gummy stem blight of cucurbits. **Fungal Biol** 119:370–382. https://doi.org/10.1016/j.funbio.2014.12.008

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. **Proc Natl Acad Sci USA** 85:2653–2657. https://doi.org/10.1073/pnas.85.8.2653

Sueoka N (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. **J Mol Evol** 40:318–325. https://doi.org/10.1007/BF00163236

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. **Mol Biol Evol** 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Taylor JW, Fisher MC (2003) Fungal multilocus sequence typing—it's not just for bacteria. **Curr Opin Microbiol** 6:351–356. https://doi.org/10.1016/S1369-5274(03)00088-2

Trifinopoulos J, Nguyen LT, von Haeseler A et al (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. **Nucleic Acids Res** 44:W232–W235. https://doi.org/10.1093/nar/gkw256

Větrovský T, Kolařík M, Žifčáková L et al (2016) The rpb2 gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. **Mol Ecol Resour** 16:388–401. https://doi.org/10.1111/1755-0998.12456

Virtanen P, Gommers R, Oliphant TE et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nat Methods** 17:261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom ML (2021) Seaborn: statistical data visualization. **J Open Source Softw** 6:3021. https://doi.org/10.21105/joss.03021

Wint R, Salamov A, Grigoriev IV (2022) Kingdom-wide analysis of fungal protein-coding and tRNA genes reveals conserved patterns of adaptive evolution. **Mol Biol Evol** 39:msab372. https://doi.org/10.1093/molbev/msab372

Wittstein K, Cordsmeier A, Lambert C et al (2020) Identification of *Rosellinia* species as producers of cyclodepsipeptide PF1022 A and resurrection of the genus *Dematophora* as inferred from polythetic taxonomy. **Stud Mycol** 96:1–16. https://doi.org/10.1016/j.simyco.2020.01.001

Wright F (1990) The effective number of codons used in a gene. **Gene** 87:23–29.

Wu P, Xiao W, Luo Y et al (2023) Comprehensive analysis of codon bias in 13 *Ganoderma* mitochondrial genomes. **Front Microbiol** 14:1170790. https://doi.org/10.3389/fmicb.2023.1170790

Yang Q, Xin C, Xiao QS et al (2023) Codon usage bias in chloroplast genes implicates adaptive evolution of four ginger species. **Front Plant Sci** 14:1304264. https://doi.org/10.3389/fpls.2023.1304264

Zavala B, Dineen L, Fisher KJ et al (2024) Genomic factors shaping codon usage across the Saccharomycotina subphylum. **G3 (Bethesda)** 14:jkae207. https://doi.org/10.1093/g3journal/jkae207

Zhao Q, Wu J, Zhang L et al (2020) Genome-scale analyses and characteristics of putative pathogenicity genes of *Stagonosporopsis cucurbitacearum*, a pumpkin gummy stem blight fungus. **Sci Rep** 10:18065. https://doi.org/10.1038/s41598-020-75235-x

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplymentary.docx