Biological Forum – An International Journal **15(5): 886-892(2023)**

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Nuclear Ribosomal DNA (nrDNA) Sequence based Molecular Markers for Plant Phylogeny: Potential and Pitfalls

Inamul Haque*
Department of Botany, Derozio Memorial College,
Rajarhat Road, PO R-Gopalpur, Kolkata (West Bengal), India.

(Corresponding author: Inamul Haque)*
(Received: 09 March 2023; Revised: 10 April 2023; Accepted: 19 April 2023; Published: 20 May 2023)
(Published by Research Trend)

ABSTRACT: Phylogenies play a valuable role in our understanding of biological diversity. They help to structure classifications and provide us with insights into the events that took place during the course of evolution. For the assessment of phylogeny, historically used morphological data is now being replaced by the more advantageous molecular data, specifically DNA sequence data. One of the most important region used for phylogenetic inference is the nuclear ribosomal DNA (nrDNA) region. The ITS1-5.8S-ITS2, ITS2, intergenic spacer and external transcribed spacer regions have been routinely used for phylogenetic analysis. Evaluating accurate phylogenetic inference with the advent of molecular data has its own set of challenges. If the background of the genes or sequences included in the analysis are not considered and understood, it can lead to major pitfalls resulting in inaccurate phylogenetic reconstructions. A better understanding of the different molecular processes which operate in these regions is of paramount importance since they can directly affect the phylogenetic analysis. This review summarizes the different problems that can arise when using nrDNA sequences for phylogenetic analysis, and how to overcome them.

Keywords: Phylogeny, molecular markers, internal transcribed spacer, external transcribed spacer, nuclear ribosomal DNA.

INTRODUCTION

Plant phylogeny is the study of the evolutionary relationships among plants (Schenck & Busta 2022). It is a complex and constantly evolving field that draws on multiple lines of evidence, including morphology, anatomy, embryology, paleobotany, and molecular biology. While morphological and other characters have been traditionally used in plant phylogenetics, they have several drawbacks.

One of the main issues with using morphological characters is that they can be highly variable and often influenced by environmental factors. This can make it difficult to accurately compare and analyze morphological characters across different species. Another issue is that some morphological characters can exhibit homoplasy (Al Sayad & Yassin 2019), meaning that they have evolved independently in different lineages and do not reflect true evolutionary relationships and this can lead to incorrect phylogenetic inferences. Furthermore, morphological characters may not be informative enough to distinguish between closely related species or resolve deep evolutionary relationships. This is especially true for groups of plants that have undergone rapid radiation and speciation, where morphological characters may not have had enough time to evolve and differentiate. In addition, morphological characters can be difficult to quantify and measure objectively, leading to subjectivity and potential errors in data analysis. This is particularly problematic in cases where morphological characters are used in combination with molecular data, as the different types of data may have different weights and uncertainties. The use of morphological characters can also be influenced by taxonomic expertise and bias, with different taxonomists potentially interpreting characters differently and leading to inconsistent results. Morphological characteristics have been the mainstay of phylogenetic analysis for many years (Durán-Castillo *et al.,* 2022). However, with the advent of molecular markers, the landscape has changed significantly (Hua *et al.,* 2022).

DNA sequence based molecular markers are DNA fragments that are used to identify genetic variation among individuals and populations. These markers are derived from DNA sequences and are used in various fields of study, including evolutionary biology, conservation biology, and forensic science. Molecular markers, in contrast to morphological ones, are DNA sequences that have evolved independently of other traits and can provide a more accurate picture of evolutionary relationships. Molecular markers and morphological characteristics are both important tools in plant systematics and phylogeny (Adhikari *et al.,* 2017). While both can be used to infer evolutionary relationships among plants, molecular markers are

generally considered to be superior to morphological characteristics in certain respects (Franco *et al.,* 2022).

Molecular markers have several advantages over morphological ones. They are less prone to homoplasy, evolve at a constant rate, and the changes are not affected by the environment, which makes them more reliable. They are more informative and can provide a higher degree of resolution at the species level and can help to clarify the relationships among closely related taxa. Molecular markers are easier to compare across different taxa, which are largely conserved across species and can be easily compared (Mahima *et al*., 2020). Analyzing the DNA sequences of different taxa, scientists can determine the patterns of evolution that have occurred over time. This information can then be used to infer the evolutionary relationships among different species and to reconstruct the evolutionary history of a group of organisms.

Nuclear ribosomal DNA (nrDNA) markers are one of the most commonly used DNA sequence-based molecular markers in plant phylogenetic studies and these markers have been used in multiple studies (Yu *et al.,* 2022). The widespread use of the nrDNA can be attributed to its universal role in all free-living organisms. Across a wide range of taxa, the nrDNA locus has an identical or nearly identical structure. This locus offers substantial insights in phylogenetic research, as specific regions of the nrDNA loci are differentially conserved, enabling its usage at different taxonomic levels (Acharya *et al.,* 2022). These regions can also be used for the development of DNA barcodes (Reddy *et al*., 2022).

NUCLEAR RIBOSOMAL DNA (nrDNA) SEQUENCE BASED MARKERS

In plants, the 18S, 5.8S and 25S (in mammals 28S) rRNAs are encoded by a single transcription unit which is known as 45S nrDNA as represented in Fig. 1. This single transcription unit in mammals is 47S nrDNA, and 35S nrDNA in yeast (Sáez-Vásquez & Delseny, 2019). The genomes of plants harbour numerous 45S rRNA genes, ranging from hundreds to thousands in number, which are typically organized in tandem arrays located in the nucleolus organizer regions (NOR). The transcribed sequence of 45S nrDNA includes the 18S, 5.8S, and 25S rRNA sequences, which are separated by internal transcribed spacers (ITS1 and ITS2) and are flanked by external transcribed spacers (ETS). Intervening between each 45S nrDNA coding sequence are intergenic spacer (IGS) regions, housing A motifs, two spacer promoters (SP1 and SP2), and the gene promoter (GP). Across plant species, the sequence encompassing the transcription initiation site exhibits a high degree of similarity. Mutations in this consensus region can result in the abolishment or inhibition of rRNA transcription, as well as disruption of the position of transcription initiation (Sáez-Vásquez & Delseny 2019). The major regions of the nrDNA used for phylogenetics are, the entire internal transcribed spacer (ITS), only the ITS2, intergenic spacer (IGS), and external transcribed spacer (ETS).

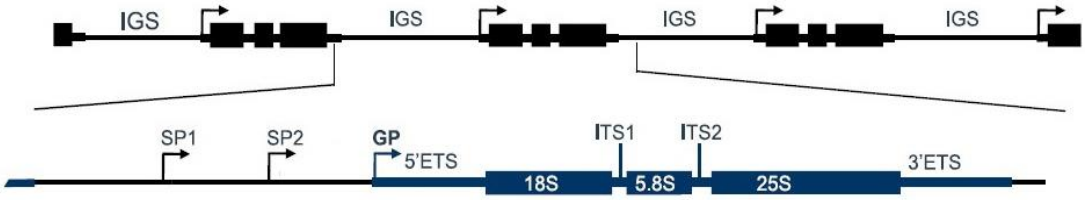

Fig. 1. Organization of 45S rDNA tandem repeat in plants (modified from Sáez-Vásquez & Delseny 2019).

A. ITS1-5.8S-ITS2 nrDNA sequence based markers

nrDNA markers are among the most commonly used molecular markers in plant phylogenetic studies due to their high degree of conservation and high copy number within the genome. The nrDNA genes consist of repeated sequences, with the major classes of nrDNA genes including the 18S-5.8S-25S genes and 5S genes and intergenic spacers (IGS), which are arranged in tandem repeats/arrays on different loci (Baldwin *et al.,* 1995; Goffová & Fajkus 2021; Soltis *et al.,* 1997). The internal transcribed spacers (ITS1, 5.8S and ITS2) of the 18S-5.8S-25S nrDNA has been used extensively for phylogenetic studies of plants. The 5.8S subunit, which is highly conserved, is flanked on either side by the more variable ITS regions. The ITS region is transcribed but are not incorporated into mature ribosomes and, are of functional importance as deletions in certain regions of the ITS results in inhibition of production and maturation of large and small subunit rRNAs (Musters *et al*., 1990; van Nues *et al*., 1994).

The ITS regions have immense potential for phylogenetic and evolutionary studies, which is reflected in the scientific literature (Banerjee *et al.,* 2018), but the benefits of using these regions may be offset by certain phenomena that can introduce homoplasy and confound phylogenetic analysis. However, a comprehensive understanding of these underlying processes have the potential to provide deeper insights into evolutionary history.

(i) Orthologous and paralogous problem. The importance of 18S-25S nrDNA arrays and their RNA products as a crucial element of eukaryotic NORs has been recognized for a considerable period of time. These regions play a crucial role in the formation and maintenance of the nucleolus, which is responsible for the production of ribosomes, essential for protein synthesis. In eukaryotic genomes the NOR loci can vary in number and distribution, and both major and minor arrays or nrDNA are present which can exhibit locational variability among species (Hemleben *et al.,* 2021). The formation of major arrays in some cases is believed to have occurred through the amplification of minor arrays, while in other instances, major arrays

have reduced in size and complexity to become minor arrays through the loss of repeats (Alvarez & Wendel 2003).

In order to accurately infer the evolutionary history from DNA sequence, it is crucial to ensure that the gene(s) that are being analyzed should be orthologous. Orthologous genes share a common ancestor due to organismal cladogenesis and are appropriate for phylogenetic analysis as they have the ability to uncover events which can be responsible for divergence among species. However, if gene duplication events have occurred in the past, the duplicated sequences that are present in the lineages are considered to be paralogous. Using paralogous genes instead of orthologous genes in phylogenetic analysis can result in a confounding of organismal divergence events, leading to error prone assessment of orthology and paralogy and ultimately, phylogenetic in-congruence. Therefore, it is imperative to ensure proper sampling of orthologous and paralogous sequences in a study to avoid inaccurate conclusions about evolutionary history (Alvarez & Wendel 2003). Also, it should not be assumed that all nrDNA sequences are truly orthologous, rather they may be a combination of orthologous and paralogous sequences (Buckler *et al.,* 1997; Hartmann *et al.,* 2001; Mayol & Rosselló 2001).

(ii) The array problems. The collective evolution of nrDNA genes is a fascinating phenomenon that involves all copies appearing to evolve in synchrony. Rather than each gene copy accumulating unique mutations, all the copies of repeats present in an array or the entire genome may have the same mutations due to intergenic sequence homogenization processes. This uniformity is referred to as concerted evolution, and is a complex mechanism (like high-frequency gene conversion and unequal crossing over) that allows for the efficient maintenance of ITS regions into identical copies (Elder & Turner 1995). The mechanisms of concerted evolution play a crucial role in removing paralogous sequences, making it easier to determine true orthology among taxa and improve the accuracy of phylogenetic reconstruction. But, concerted evolutionary mechanisms that lead to sequence homogenization may not always be able to keep up with the variation-generating processes at the organism and genomic levels. Therefore, to conclude that only a single form of ITS sequence exists in a given taxa is erroneous. Furthermore, it can also be argued that chimeric ITS sequences may arise as a result of maintenance of two or more repeat types following a hybridization event. A phylogenetic reconstruction from these chimeric sequences will tend to occupy basal phylogenetic positions with either parental lineage (Rosazlina *et al.,* 2021).

Another possibility of hybridization event is that only one repeat type sequence may tend to dominate the entire population of arrays in the genome, this repeat type can be from either of the parent. The possibility that only a single repeat type remains in each descendant presents a challenge for phylogenetic analysis of ITS sequences, as it cannot provide a clear understanding of the history of genomic merger.

The ITS1-5.8S-ITS2 repeats can be found in hundreds to thousands of copies at one or more chromosomal locations, but due to the evolutionary instability of nrDNA arrays, not all repeats remain functional over time. Some copies may degrade into pseudogenes unless they are immediately deleted or rescued by concerted evolution, as a result, genomes may accumulate a variety of dead or dying repeats of different ages. These non-functional pseudogenes pose a serious problem for phylogenetic analysis as they may evolve independently of the functional ITS1-5.8S-ITS2 repeats (Hartmann *et al.,* 2001; Muir *et al.,* 2001; Yang *et al*., 1999).

(iii) The alignment issues. The absence of protein-encoding function in ITS sequences presents challenges in achieving natural alignment guides for sequence alignment. Moreover, the issue of sequence alignment is further compounded by the tendency of ITS sequences to accumulate indels and, their typically high GC content. As sequence accuracy, alignment, and gap treatment are critical factors influencing phylogenetic outcomes, these issues merit careful consideration (Hsieh *et al.,* 2022).

The challenges associated with alignment and sequencing, compensatory base changes (discussed below), paralogy, pseudogenes, and incomplete concerted evolution, can ironically result in increased homoplasy in phylogenetic datasets (Cao *et al.,* 2022). The accuracy of a phylogeny is not impacted much by homoplasy if the said characters are uninformative. However, in cases where homoplasy is distributed in a manner that appears as synapomorphy, it can potentially lead to misleading phylogenetic conclusions.

B. The interesting ITS2

The ITS2 region of the nrDNA arrays have opened up an interesting approach of using secondary structure prediction based alignment for phylogenetic analysis. The cleavage and processing of ITS2 region is not only complicated and fascinating (Fromm *et al.,* 2017), it is also important for bio-genesis of functional ribosomes (Côté & Peculis 2001). Probably, because of the functional aspect of the ITS2 region, it appears that the secondary structure formed by the single stranded rRNA (consisting of base paired stems and unpaired loops) is conserved across eukaryote groups (Joseph *et al.,* 1999; Schultz *et al.,* 2005). The ITS2 region consists of conserved and variable sequences, this gives ITS2 region the ability to be comparable over a large number of taxa.

The ITS2 secondary structure generally consists of four helices (I-IV), the sequence variability of helices I and IV can be used for identification upto subspecies level, whereas the primary sequence of helices II and III along with their adjacent single stranded regions are highly conserved. However, there is variation in the number of helices across different eukaryotes and only helices II and III are universally recognizable and common (Coleman, 2007). The base pair interaction in the stem regions of ITS2 are usually the four canonical Watson-Crick base pair, but non-canonical base pairs do occur occasionally (Antczak *et al*., 2019).

The secondary structure of RNA exhibits a higher degree of conservation than its sequence when under functional constraint. Sequences that are homologous and capable of performing functions like, cleavage, catalysis, or binding may contain different nucleotide, which poses a challenge to the conventional approach of sequence-based alignment. However, by considering the conserved common secondary structures of these homologous RNA sequences during folding, the accuracy and reliability of the aligned sequences can be significantly enhanced.

Even though the ITS2 region is fast evolving, it can maintain its secondary structure with the help of compensatory base changes (CBCs). A compensatory base change occurs when both nucleotides that pair with each other in a double-stranded helix (the stem) are altered, in the case of hemi-CBC, only one of the nucleotide change happens while maintaining the pairing (Zhang *et al.,* 2020). But nucleotide sequences and DNA evolutionary models used for phylogenetic analyses assume that evolution of each site in a sequence occurs independently. To compensate for the site dependency of CBCs, it was earlier recommended to assign half weight to all (Wheeler & Honeycutt 1988) but, recent workers suggest assigning half weight only to double CBCs rather than all CBCs (Zhang *et al.,* 2020).

Apart from the use of ITS2 CBCs for phylogenetic analysis, the region has the potential for another application, which is directly utilizing CBC numbers for species delimitation. But the prevalence of ITS2 CBCs primarily relates to broader taxonomic groups above the species level, rather than distinguishing between individual species (Li *et al*., 2019) and should be used with caution.

*C. The intergenic spacer (IGS***)**

The IGS region, which acts as a separator between the tandem arrays of nrDNA, is a complex region encompassing regulatory elements such as promoters, enhancers and terminators. Various repeating elements or sub-repeats, multiple types of enhancers, and promoter regions are present in this region which play a critical role in the regulation of rRNA transcription and transcript processing. It can be concluded that the IGS region serves as a vital functional region containing nucleotide sequences that trigger and/or terminate transcription, making it an essential component in the intricate control of rRNA synthesis. Even though IGS region has a functional role, the IGS sub-repeats of different species do not share high sequence similarity. This can be attributed to the fast indel (insertion-deletion) rates among the short mono-nucleotide repeats which are observed rather abundantly in the IGS. The short mono-nucleotide repeats tend to be common among closely related species and may indicate a species specific pattern, indicating a common evolutionary history.

Due to the variability observed in certain regions of the IGS, some of which may even surpass the widely utilized ITS in terms of variability, these regions have been explored as potential phylogenetic markers (Hu *et al.,* 2019). However, the use of IGS region has been criticized for its high sequence variability, presence of sub-repeats and, difficulty in sequence alignment due to the length variability of the sub-repeats. This region of nrDNA is known for its rapid evolution, with multiple internal sub-repeats that exhibit dynamic changes in both size and structure over time. This rapid evolution poses challenges in conducting comparative analyses and primer designing, as the IGS region can exhibit significant variability among different species, making it difficult to draw direct comparisons in some cases.

Even though, the presence of numerous reiterated sub-repeats within the IGS sequence pose a problem, it has been effectively employed in some studies to infer phylogeny between closely related species (Krawczyk *et al.,* 2017). To circumvent the problem of sub-repeats, a method called dropout alignment can also be used (Ryu *et al.,* 2008).

D. The ETS

There are two ETS, 5' and 3' (sometimes called ETS1 and ETS2), which are separated by the IGS. The major focus of phylogenetic analysis is on the 5' end of ETS, the 3' end of ETS have occasionally been used (Wu et al., 2020). The major reason for the prevalence of 5' end of ETS over other regions of the IGS is due to higher levels of conservation at the sequence and structural level. The nrDNA intergenic spacer (IGS) is characterized by a progressive reduction in sequence conservation from the 18S gene towards the central region of the nrDNA IGS, which is composed of repetitive elements (Fehrer *et al.,* 2021). These characteristics of the nrDNA IGS region present significant challenges for primer development and sequence alignment, particularly outside of the ETS region, even at lower taxonomic levels. Although precise substitution rates for 5' ETS may not be directly comparable to those of ITS, given the unlikelihood of a shared universal substitution rate, evidence indicates that relative substitution rates in 5' ETS are notably higher, ranging from 1.3 to 7 times that of ITS (Linder et al., 2000). The high rate of evolution helps in resolving phylogenetic trees at lower taxonomic levels, providing much better topological partitioning than the ITS. The 5' ETS generally exhibits superiority over ITS in terms of sequence divergence levels, numbers of parsimony informative sites, and resolving power.

Even though 5' ETS has demonstrated clear advantages in phylogenetic studies, it is often challenging to assign unambiguous homology to sequences at higher taxonomic levels. Also, the procedure for designing primers for amplification of 5' ETS is time consuming and technically challenging. Concerted evolution also seems to be operational at a higher level as only marginal levels of polymorphism were observed within 5' ETS sequence of individuals. But, the occurrence of hybridization was observed to lead to a notable presence of ambiguous sites (Noyes, 2006).

CONCLUSIONS

The wealth of knowledge on the phylogeny of life is largely attributed to morphological data. The

hierarchical classification systems that have been developed using morphological datasets do share similar phylogenetic nodes with those predicted by sequence based markers. Despite their limited resolution, these classifications establish a foundation of diagnostic anchor points (Caddah *et al.*, 2022). DNA sequence analyses can then serve to validate, clarify, reinforce, and enhance accuracy for phylogenetic areas that lack sufficient morphological data, using these anchor points as a framework. The reason behind DNA data overtaking morphology in phylogenetic studies is that a substantial amount of the valuable morphological diversity has already undergone meticulous examination.

The nrDNA region of eukaryotes do serve as a valuable marker for phylogenetic analyses due to their high sequence level variability, conserved flanking regions, rapid concerted evolution under similar functional constraints, and their small size. There are multiple challenges associated with the usage of these sequence. These challenges can be overcome with a proper understanding of the sequences and considerable precautions have to be undertaken for phylogenetic analyses to be accurate. Studies have consistently demonstrated that integrating both ITS and 5' ETS datasets in phylogenetic analyses tends to yield higher support and resolution of trees. Therefore, incorporating a 5' ETS dataset into an existing ITS-based phylogeny appears to be a promising approach for enhancing phylogenetic accuracy (Chen *et al.*, 2022). Rather than relying solely on nrDNA sequences for phylogenetic analysis, it is beneficial to consider using single-copy nuclear genes as an alternative. These genes, which are inherited from both parents, are becoming more prevalent in phylogenetic analysis. They typically avoid concerted evolution and feature codons that limit alignment ambiguity, making it easier to conduct homologous comparisons.

FUTURE SCOPE

The advances in nucleic acid sequencing technology has opened a new dimension to marker technology by not only providing a large amount of sequencing data at a low economic cost (Danilevicz *et al*., 2021), but also numerous bioinformatics software for a multifaceted and analytical approach (Draper *et al.*, 2022). This has induced a paradigm shift in phylogenetics resulting in more studies in the direction of phylogenomics, which is an expansion of phylogenetics that takes into account not only the evolution of nucleotides but also broader phenomena that influence entire genomes (Boutte *et al.*, 2022).

Acknowledgement. I extend my sincere thanks to the Teacher Incharge, Derozio Memorial College, for providing all facilities. My sincere thanks to Dr. M. Sarkar, head of the Department of Chemistry, Derozio Memorial College for helping with the drafting of this article. I also extend my sincere thanks to Mr. Joy Chakraborty for helping me with graphics. I also thank the anonymous referees for their useful suggestions.

Conflict of Interest. None.

REFERENCES

Acharya, G. C., Mohanty, S., Dasgupta, M., Sahu, S., Singh, S., Koundinya, A. V. V., Kumari, M., Naresh, P. & Sahoo, M. R. (2022). Molecular phylogeny, DNA barcoding, and ITS2 secondary structure predictions in the medicinally important *Eryngium* genotypes of east coast region of India. *Genes*, *13*(9), 1678.

Adhikari, S., Saha, S., Biswas, A., Rana, T. S., Bandyopadhyay, T. K. & Ghosh, P. (2017). Application of molecular markers in plant genome analysis: A review. *The Nucleus*, *60*(3), 283–297.

Al Sayad, S. & Yassin, A. (2019). Quantifying the extent of morphological homoplasy: A phylogenetic analysis of 490 characters in *Drosophila*. *Evolution Letters*, *3*(3), 286–298.

Alvarez, I. & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, *29*(3), 417–434.

Antczak, M., Zablocki, M., Zok, T., Rybarczyk, A., Blazewicz, J. & Szachniuk, M. (2019). RNAvista: A webserver to assess RNA secondary structures with non-canonical base pairs. *Bioinformatics (Oxford, England)*, *35*(1), 152–155.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S. & Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, *82*(2), 247–277.

Banerjee, S., Bharati, P., Gangwar, S. & Rai, D. V. (2018). Phylogenetic studies of Rudraksha; *Elaeocarpus* spp. *International Journal of Theoretical & Applied Sciences*, *10*(1), 15–20.

Boutte, J., Fishbein, M. & Straub, S. C. K. (2022). NGS-Indel Coder v2.0: A streamlined pipeline to code indel characters in phylogenomic data. *Methods in Molecular Biology (Clifton, N.J.)*, *2512*, 61–72.

Buckler, E. S., Ippolito, A. & Holtsford, T. P. (1997). The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. *Genetics*, *145*(3), 821–832.

Caddah, M. K., Meirelles, J., Nery, E. K., Lima, D. F., Nicolas, A. N., Michelangeli, F. A. & Goldenberg, R. (2022). Beneath a hairy problem: Phylogeny, morphology, and biogeography circumscribe the new *Miconia* supersection *Discolores* (Melastomataceae: Miconieae). *Molecular Phylogenetics and Evolution*, *171*, 107461.

Cao, R., Tong, S., Luan, T., Zheng, H. & Zhang, W. (2022). Compensatory base changes and varying phylogenetic effects on angiosperm ITS2 genetic distances. *Plants (Basel, Switzerland)*, *11*(7), 929.

Chen, Y.-P., Turdimatovich, T. O., Nuraliev, M. S., Lazarević, P., Drew, B. T. & Xiang, C. L. (2022). Phylogeny and biogeography of the northern temperate genus *Dracocephalum* s.l. (Lamiaceae). *Cladistics: The International Journal of the Willi Hennig Society*, *38*(4), 429–451.

Coleman, A. W. (2007). Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research*, *35*(10), 3322–3329.

Côté, C. A. & Peculis, B. A. (2001). Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. *Nucleic Acids Research*, *29*(10), 2106–2116.

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E. & Edwards, D. (2021). High-throughput genotyping technologies in plant taxonomy. *Methods*

in Molecular Biology (*Clifton, N.J.*), *2222*, 149–166.
Draper, I., Villaverde, T., Garilleti, R., Burleigh, J. G., McDaniel, S. F., Mazimpaka, V., Calleja, J. A. & Lara, F. (2022). An NGS-Based phylogeny of orthotricheae (Orthotrichaceae, Bryophyta) With the proposal of the new genus *Rehubryum* From Zealandia. *Frontiers in Plant Science*, *13*, 882960.
Durán-Castillo, M., Hudson, A., Wilson, Y., Field, D. L. & Twyford, A. D. (2022). A phylogeny of Antirrhinum reveals parallel evolution of alpine morphology. *The New Phytologist*, *233*(3), 1426–1439.
Elder, J. F. & Turner, B. J. (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology*, *70*(3), 297–320.
Fehrer, J., Slavíková, R., Paštová, L., Josefiová, J., Mráz, P., Chrtek, J. & Bertrand, Y. J. K. (2021). Molecular evolution and organization of ribosomal DNA in the hawkweed tribe *Hieraciinae* (Cichorieae, Asteraceae). *Frontiers in Plant Science*, *12*, 647375.
Franco, F. F., Amaral, D. T., Bonatelli, I. A. S., Romeiro-Brito, M., Telhe, M. C. & Moraes, E. M. (2022). Evolutionary genetics of cacti: Research biases, advances and prospects. *Genes*, *13*(3), 452.
Fromm, L., Falk, S., Flemming, D., Schuller, J. M., Thoms, M., Conti, E. & Hurt, E. (2017). Reconstitution of the complete pathway of ITS2 processing at the pre-ribosome. *Nature Communications*, *8*(1), 1787.
Goffová, I., & Fajkus, J. (2021). The rDNA Loci-Intersections of Replication, Transcription, and Repair Pathways. *International Journal of Molecular Sciences*, *22*(3), 1302.
Hartmann, S., Nason, J. D., & Bhattacharya, D. (2001). Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. *Journal of Molecular Evolution*, *53*(2), 124–134.
Hemleben, V., Grierson, D., Borisjuk, N., Volkov, R. A. & Kovarik, A. (2021). Personal perspectives on plant ribosomal RNA genes research: From precursor-rRNA to molecular evolution. *Frontiers in Plant Science*, *12*, 797348.
Hsieh, C. L., Yu, C. C., Huang, Y. L. & Chung, K. F. (2022). *Mahonia* vs. *Berberis* unloaded: Generic delimitation and infrafamilial classification of berberidaceae based on plastid phylogenomics. *Frontiers in Plant Science*, *12*.
Hu, X., Yu, F., Huang, Y., Sun, L., Li, X., Yang, S., Chen, K., Huang, F., Zeng, K., Zhang, M. & Deng, Z. (2019). Characterization analysis of the 35S rDNA intergenic spacers in *Erianthus arundinaceus*. *Gene*, *694*, 63–70.
Hua, Z., Jiang, C., Song, S., Tian, D., Chen, Z., Jin, Y., Zhao, Y., Zhou, J., Zhang, Z., Huang, L. & Yuan, Y. (2023). Accurate identification of taxon-specific molecular markers in plants based on DNA signature sequence. *Molecular Ecology Resources*, *23*(1), 106–117.
Joseph, N., Krauskopf, E., Vera, M. I. & Michot, B. (1999). Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Research*, *27*(23), 4533–4540.
Krawczyk, K., Nobis, M., Nowak, A., Szczecińska, M. & Sawicki, J. (2017). Phylogenetic implications of nuclear rRNA IGS variation in Stipa L. (Poaceae). *Scientific Reports*, *7*(1), Article 1.
Li, M., Zhao, H., Zhao, F., Jiang, L., Peng, H., Zhang, W. & Simmons, M. P. (2019). Alternative analyses of compensatory base changes in an ITS2 phylogeny of *Corydalis* (Papaveraceae). *Annals of Botany*, *124*(2), 233–243.
Linder, C. R., Goertzen, L. R., Heuvel, B. V., Francisco-Ortega, J. & Jansen, R. K. (2000). The complete external transcribed spacer of 18s-26s rDNA: amplification and phylogenetic utility at low taxonomic levels in asteraceae and closely allied families. *Molecular Phylogenetics and Evolution*, *14*(2), 285–303.
Mahima, K., Sudhakar, J. V. & Sathishkumar, R. (2020). Molecular phylogeny of the *Ficus virens* complex (Moraceae). *Genome*, *63*(12), 597–606.
Mayol, M., & Rosselló, J. A. (2001). Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. *Molecular Phylogenetics and Evolution*, *19*(2), 167–176.
Muir, G., Fleming, C. C. & Schlötterer, C. (2001). Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. And *Quercus robur* L. *Molecular Biology and Evolution*, *18*(2), 112–119.
Musters, W., Boon, K., van der Sande, C. A., van Heerikhuizen, H. & Planta, R. J. (1990). Functional analysis of transcribed spacers of yeast ribosomal DNA. *The EMBO Journal*, *9*(12), 3989–3996.
Noyes, R. D. (2006). Intraspecific nuclear ribosomal DNA divergence and reticulation in sexual diploid *Erigeron strigosus* (Asteraceae). *American Journal of Botany*, *93*(3), 470–479.
Reddy, V. D., Mehandi, S., Janeja, H. S., Saxena, K. & Prakash, S. (2022). Concept on plant DNA barcodes and their application in identification of plants. *Biological Forum – An International Journal*, *14*(2a), 360–368.
Rosazlina, R., Jacobsen, N., Ørgaard, M. & Othman, A. S. (2021). Molecular evidence of the hybrid origin of *Cryptocoryne* × *purpurea* Ridl. Nothovar. Purpurea (Araceae). *PloS One*, *16*(1), e0239499.
Ryu, S., Do, Y., Fitch, D. H. A., Kim, W. & Mishra, B. (2008). Dropout alignment allows homology recognition and evolutionary analysis of rDNA intergenic spacers. *Journal of Molecular Evolution*, *66*(4), 368–383.
Sáez-Vásquez, J. & Delseny, M. (2019). Ribosome biogenesis in plants: from functional 45s ribosomal DNA organization to ribosome assembly factors. *The Plant Cell*, *31*(9), 1945–1967.
Schenck, C. A. & Busta, L. (2022). Using interdisciplinary, phylogeny-guided approaches to understand the evolution of plant metabolism. *Plant Molecular Biology*, *109*(4–5), 355–367.
Schultz, J., Maisel, S., Gerlach, D., Müller, T. & Wolf, M. (2005). A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA (New York, N.Y.)*, *11*(4), 361–364.
Soltis, D. E., Soltis, P. S., Nickrent, D. L., Johnson, L. A., Hahn, W. J., Hoot, S. B., Sweere, J. A., Kuzoff, R. K., Kron, K. A., Chase, M. W., Swensen, S. M., Zimmer, E. A., Chaw, S.-M., Gillespie, L. J., Kress, W. J. & Sytsma, K. J. (1997). Angiosperm phylogeny inferred from 18s ribosomal DNA sequences. *Annals of the Missouri Botanical Garden*, *84*(1), 1–49.
van Nues, R. W., Rientjes, J. M. J., van der Sande, C. A. F. M., Zerp, S. F., Sluiter, C., Venema, J., PIanta, R. J. & Raué, H. A. (1994). Separate structural elements within internal transcribed spacer 1 of *Saccharomyces cerevisiae* precursor ribosomal RNA direct the formation of 17S and 26S rRNA. *Nucleic Acids Research*, *22*(6), 912–919.
Wheeler, W. C. & Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Molecular Biology and*

Evolution, *5*(1), 90–96.

Wu, C. S., Sudianto, E., Hung, Y. M., Wang, B. C., Huang, C. J., Chen, C. T. & Chaw, S. M. (2020). Genome skimming and exploration of DNA barcodes for Taiwan endemic cypresses. *Scientific Reports*, *10*(1), 20650.

Yang, Y. W., Lai, K. N., Tai, P. Y., Ma, D. P. & Li, W. H. (1999). Molecular phylogenetic studies of *Brassica, rorippa*, *Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S-25S rDNA. *Molecular Phylogenetics and Evolution*, *13*(3), 455–462.

Yu, X. Q., Jiang, Y. Z., Folk, R. A., Zhao, J. L., Fu, C. N., Fang, L., Peng, H., Yang, J. B., & Yang, S. X. (2022). Species discrimination in *Schima* (Theaceae): Next-generation super-barcodes meet evolutionary complexity. *Molecular Ecology Resources*, *22*(8), 3161–3175.

Zhang, W., Tian, W., Gao, Z., Wang, G. & Zhao, H. (2020). Phylogenetic utility of rRNA ITS2 sequence-structure under functional constraint. *International Journal of Molecular Sciences*, *21*(17), 6395.

Zhang, X., Cao, Y., Zhang, W. & Simmons, M. P. (2020). Adenine·cytosine substitutions are an alternative pathway of compensatory mutation in angiosperm ITS2. *RNA*, *26*(2), 209–217.

How to cite this article: Inamul Haque (2023). Nuclear Ribosomal DNA (nrDNA) Sequence based Molecular Markers for Plant Phylogeny: Potential and Pitfalls. *Biological Forum – An International Journal*, *15*(5): 886-892.