Ho VT *et al.* (2025)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 53, Issue 1, Article number 14244
DOI:10.15835/nbha53114244
Research Article

# Characterization of the complete chloroplast genome seedless jackfruit (*Artocarpus heterophyllus* var. *seedless*) and its relatives

Viet The HO[1]*, Thi Kim Anh NGO[1], Yohannes Gedamu GEBRE[2], Dang Khoa NGUYEN[3], Sasanti WIDIARSIH[4]

[1]*Ho Chi Minh City University of Industry and Trade, Faculty of Biology and Environment, 140 Le Trong Tan, Tan Phu district, Ho Chi Minh City 700000, Vietnam; thehv@huit.edu.vn (*corresponding author); anhngotk@huit.edu.vn*
[2]*Bio and Emerging Technology Institute (BETin), P.O. Box 5954, Addis Ababa, Ethiopia; yoha.gedamu@gmail.com*
[3]*Technology and IP Commercialization, 22 Le Van Mien, Thao Dien, Thu Duc, Ho Chi Minh City 700000, Vietnam; ndkhoa121098@gmail.com*
[4]*Research Center for Food Crops, Research Organization of Food and Agriculture, National Research and Innovation Agency (BRIN). Jl. Raya Bogor Km. 46, Cibinong 16911, Indonesia; sasa002@brin.go.id*

## Abstract

Jackfruit (*Artocarpus heterophyllus*) is a perennial fruit tree extensively cultivated in Vietnam, where it plays a crucial role in economic development, especially in rural, disadvantaged regions. Recently, a seedless variety has emerged, attracting significant demand due to its superior fruit quality. Remarkably, the edible portion of this variety can comprise more than 90% of the fruit's total weight, increasing its market appeal. As a result, the demand for seedlings of this variety has risen sharply. However, current identification methods for distinguishing seedless jackfruit from other varieties primarily rely on morphological traits, which are often ineffective due to high similarity. To overcome this limitation, the present study aimed to sequence and analyze the complete chloroplast (cp) genome of the seedless jackfruit variety. The cp genome was determined to be 160,385 base pairs in length, consisting of 128 genes, including 84 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. Comparative analysis revealed differences in simple sequence repeat (SSR) patterns between this seedless variety and other jackfruit cp genomes. These findings underscore the value of chloroplast genome characterization as a tool for the precise identification and classification of this novel seedless jackfruit variety.

***Keywords:*** *Artocarpus heterophyllus*; chloroplast genome; next generation sequencing; phylogenetic tree; seedless jackfruit

## Introduction

Jackfruit (*Artocarpus heterophyllus*) is a widely cultivated fruit tree in Vietnam, valued for its versatility and year-round fruit production. The fruit is highly nutritious, rich in proteins, carbohydrates, fats, and essential minerals, making it a vital food source, particularly during periods of scarcity in rural communities. Recently, high-quality seedless jackfruit varieties have entered the market, with over 90% of the fruit's weight consisting of edible flesh. This has led to a surge in demand for seedlings of this variety. However, distinguishing

*Received: 02 Nov 2024. Received in revised form: 15 Dec 2024. Accepted: 13 Feb 2025. Published online: 26 Feb 2025.*
From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

seedless variety from others based solely on morphological traits has proven ineffective due to the high degree of similarity between them. Current jackfruit identification methods primarily rely on observable characteristics such as fruit size, color, flavor, and leaf traits (Suneel *et al*., 2022). While these traits are practical and easy to assess, they are limited in scope, and their allelic expression often follows a dominant-recessive pattern, making it difficult to distinguish between heterozygous and homozygous individuals. Furthermore, many of these traits are polygenic and only expressed during the long juvenile phase (Sampath *et al*., 2020).

Biochemical markers offer a viable alternative for selecting superior jackfruit cultivars (Ibrahim *et al*., 2013; Aseef *et al*., 2017), utilizing parameters such as total soluble solids, titratable acidity, moisture content, reducing and non-reducing sugars, total carbohydrates, crude protein, crude fiber, and crude oil. The reliability of these markers, however, can be influenced by the developmental stage. Additionally, biochemical traits are highly dynamic, varying with the tree's growth stage and environmental conditions. Even within a single jackfruit, these traits can differ depending on geographic location (Azad *et al*., 2007) and between different parts of the fruit (Balamaze *et al*., 2019).

To address the limitations of morphological and biochemical markers in the identification of plant varieties, including jackfruit, numerous studies have employed molecular markers for classification and identification. Molecular markers, which rely on specific DNA sequences, offer precise identification of plant species, unaffected by environmental conditions or the plant's developmental stage. Techniques such as RAPD (Krishnan *et al*., 2015) and SSR (Marjan *et al*., 2019) have been widely applied in jackfruit. However, the advancement of these techniques has faced several challenges: (a) the difficulty of designing universal primers that target homologous markers across all plant species; (b) the ease of amplifying and sequencing proposed DNA markers in some families or genera, while this is problematic in others; and (c) for certain DNA-barcoding markers, genetic gaps between species are pronounced in some plant groups, but absent in others (Das and Dang, 2018). Different studies have proposed the use of DNA barcodes—standardized DNA sequences within the genome—for plant identification (CBOL Plant Working Group, 2009; Kress, 2017). However, our previous study, which utilized three DNA barcode regions (*mat*K, *rbc*L, and ITS), was unable to differentiate seedless jackfruit from other jackfruit cultivars in Vietnam. This limitation arises from the fact that the discriminatory power of these barcodes is generally confined to the genus or species level (Ho *et al.,* 2019).

The chloroplast (cp) genome is significantly smaller than the nuclear and mitochondrial genomes across three plant cells. It provides several advantages for plant taxonomy studies, including haploid inheritance, maternal transmission, a low nucleotide substitution rate, and a highly conserved structure (Feng *et al*., 2023). Furthermore, the cp genome exists in multiple copies per cell, enabling high expression levels of cp-targeted genes. This feature supports the reliable identification of maternal and clonal lines, which is particularly important for crops propagated vegetatively (Gardner *et al*., 2015). Consequently, cp genome sequences have been widely used in phylogenetic analyses for decades (Palmer *et al.,* 1988; Soltis *et al*., 1989). Rapid advancements in sequencing technologies have recently reduced both the cost and time required to sequence complete cp genomes. Studies have sequenced and compared whole cp genomes across different cultivars within the same plant species, uncovering valuable variations that aid in distinguishing plants at subspecies or cultivar levels. Examples include *Oryza sativa* (Tang *et al*., 2004) and *Panax ginseng* (Kim *et al.,* 2015).

The *Artocarpus* genus has also seen the application of next-generation sequencing (NGS) for cp genome analysis. Species such as *A. nanchuanensis* (Li and Song, 2019), *A. hypargyreus* (Li *et al*., 2020), *A. gomezianus* (Lin *et al*., 2021), *A. altilis* (de Souza *et al*., 2021), and *A. tonkinensis* (Tang *et al*., 2021) have had their cp genomes sequenced and studied. To the best of our knowledge, no published study has yet explored the cp genome of seedless jackfruit (*A. heterophyllus).* This research focuses on sequencing and characterizing the complete cp genome of a seedless jackfruit variety collected in Vietnam. The findings provide crucial insights for taxonomy, botanical identification, breeding, and conservation efforts related to this unique jackfruit variety, highlighting its scientific and agricultural value.

## Materials and Methods

*Sample collection and DNA sequencing*

The seedless jackfruit specimen was collected from Ba Lang Ward, Cai Rang District, Can Tho City, Vietnam, and the voucher sample is currently stored at the laboratory of Ho Chi Minh University of Industry and Trade (HUIT) in Vietnam. Total DNA was extracted from fresh leaves using the Isolate II Plant DNA Kit from Bioline (UK). DNA quality was evaluated using 1% agarose gel electrophoresis, while quantity was determined with a Nanodrop spectrophotometer from ThermoScientific (Delaware, USA). A total of 500 ng of DNA was fragmented using the S220 Focused-ultrasonicator from Covaris (USA). Following fragmentation, dA tails were added, adapters were ligated, and the resulting fragments were purified. Library preparation, quality control, cluster generation, and DNA sequencing were performed at Azenta Life Sciences (USA) using the Illumina NovaSeq 6000 sequencer. The quality of the raw reads was assessed using FastQC within the Galaxy portal (*http://usegalaxy.org*) and then submitted to the Sequence Read Archive (SRA) database in NCBI (https://www.ncbi.nlm.nih.gov/) under the SRR28223603 accession number.

*Chloroplast genome assembly and annotation*

The reads from both ends were aligned with the *Artocarpus heterophyllus* reference sequence (NCBI accession number MK303549.1) during the assembly process. Subsequently, the Pilon tool (v. 1.21) was employed to generate a FASTA file, which was then converted to GenBank format using the Geseq program (*https://chlorobox.mpimp-golm.mpg.de/geseq.html*). Any unacceptable characters in the resulting GenBank file were removed using the CleanSeq module from CPGView, available at *http://www.1kmpg.cn/cpgview* (Liu *et al.,* 2023). This software was then utilized for annotating the chloroplast genome, locating genes, and determining the lengths of introns and exons for the genes within the cp genome.

*Comparative analysis and repeat analysis among Artocarpus cp genomes*

The chloroplast genomes of the seedless jackfruit, along with 13 complete cp genomes from 11 species in the *Artocarpus genus* available in the NCBI GenBank were analyzed consisting of *A. heterophyllus* (MK303549.1), *A. heterophyllus* (MG434693), *A. altilis* (NC_059002.1), *A. altilis* (MW367444.1), *A. camansi* (NC_054247.1), *A. champeden* (NC_057308.1), *A. integer* (MT900597.1), *A. petelotii* (NC_056286.1), *A. gomezianus* (NC_080592.1), *A. hypargyreus* (NC_057287.1), *A. excelsus* (OR664167), *A. elasticus* (OR664166), and *A. tonkinensis* (MZ379793).

The MIcroSAttelite (MISA) identification tool was employed to identify SSR motifs, accessible at *https://webblast.ipk-gatersleben.de/misa/*, as described by Beier *et al.* (2017). To identify long repeat regions with a repeat size of ≥ 30 bp and a minimum identity of 90%, the REPuter software (*http://bibiserv.cebitec.uni-bielefeld.de/reputer*) was utilized. This analysis identified four types of repeats: forward (F), reverse (R), complement (C), and retrograde (P), as reported by Kurtz *et al.* (2001).

*Phylogenetic analysis*

A total of 14 chloroplast genomes from the *Artocarpus* genus, along with three cp genomes from different genera in the Moraceae family—namely KM491711.2 (*Morus mongolica*), NC_033979.1 (*Ficus religiosa*), and MH430880.1 (*Broussonetia papyrifera*) were utilized as outgroups for phylogenetic analysis. The alignment of these sequences was performed using the MAFFT program (available at *http://mafft.cbrc.jp/alignment/server/*). The outgroups were selected based on the criteria outlined by Michu (2004), as they represent the sister groups of the *Artocarpus* genus. Phylogenetic trees were constructed using both the Neighbor-Joining method (A) and the Maximum Likelihood method (B), with 500 bootstrap replicates, employing MEGA X software.

## Results

*Genome, sequence assembly, and features of chloroplast*

A total of 115 Gb of paired-end data (150 bp) was generated, yielding 168,251,600 reads, with a Phred score indicating that 97.03% of the reads exceeded Q20. The GC content of the chloroplast genome was approximately 37%. Upon assembly, the cp genome displayed a conserved circular structure, with a total length of 160,385 bp. This finding is consistent with the results reported by Wang *et al*. (2019), who documented a cp genome size of 160,389 bp for jackfruit from China. The genome comprises four distinct regions: a Large Single Copy (LSC), a Small Single Copy (SSC), and two Inverted Repeat (IR) regions, with the IR regions flanking the LSC and SSC regions (Figure 1).

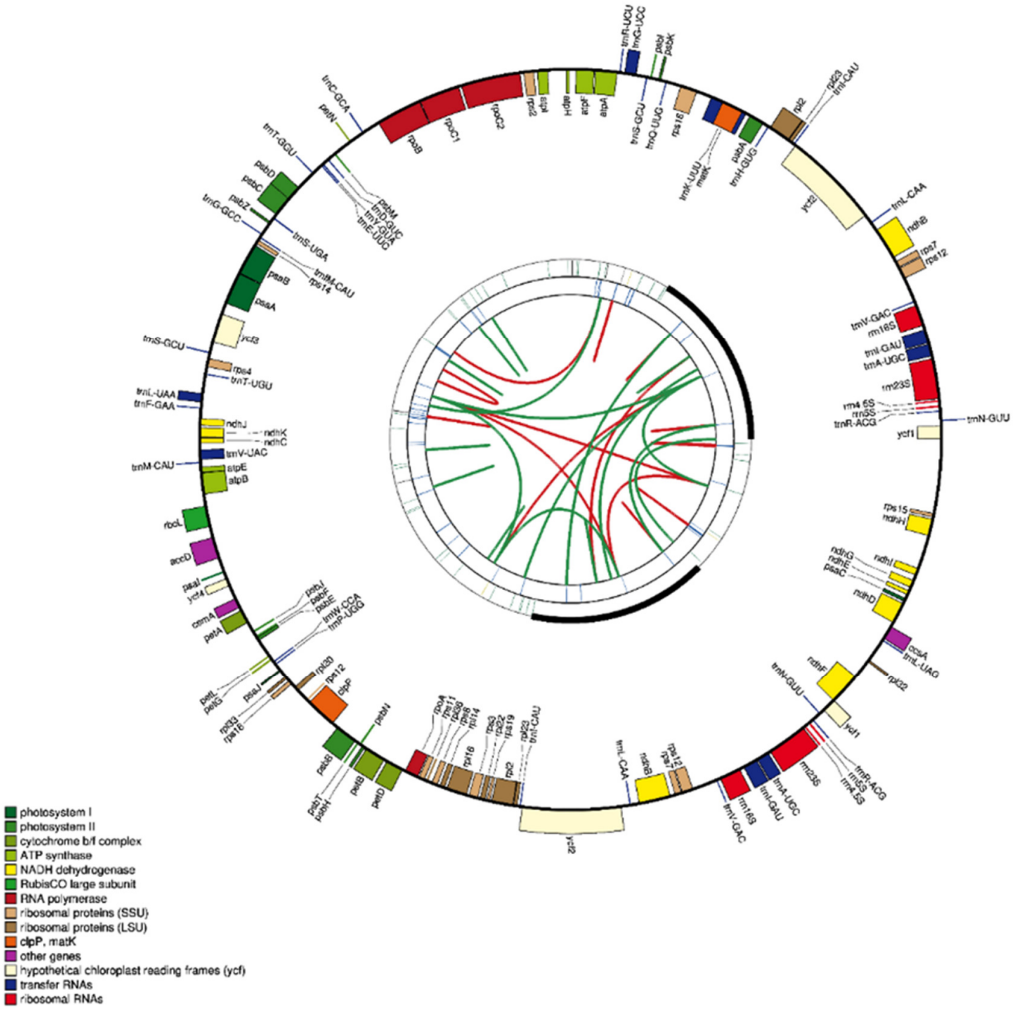


**Figure 1.** The chloroplast genome map of the seedless jackfruit variety from Vietnam is presented. The innermost circle illustrates the forward and reverse repeats, indicated by red and green arcs, respectively. The second circle displays tandem repeats, represented by short bars. The third circle features microsatellite sequences identified using the MISA tool. The outer circle depicts the gene structure of the plastome, with genes color-coded according to their functional categories

The plastid genome of the seedless jackfruit variety is predicted to encode a total of 128 genes, comprising 84 protein-coding genes, 36 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes (Table 1). This finding represents a slight increase compared to previous studies by Liu *et al*. (2018) and de

Souza *et al*. (2021), which identified only 113 genes in jackfruit chloroplast genomes. In contrast, Wang *et al*. (2019) reported a total of 112 genes in the chloroplast of this species. Variations in gene numbers within the chloroplast genome of a single plant species have been documented previously; for instance, in *A. altilis*, de Souza *et al*. (2021) identified 113 genes, whereas Wei *et al*. (2023) reported up to 132 genes. The gene count in jackfruit is comparatively lower than that of other species in the *Artocarpus* genus, such as *A. hypargyreus*, which contains 129 genes (Li *et al*., 2020), and *A. champeden*, which has 131 genes (Niu and Liu, 2021). Although, the variations found in cp genome of terrestrial plants are mostly due to point mutations and deletions/insertions in noncoding regions, the deletions and insertions of genes or large inverted repeat was also reported (Palmer *et al.,* 1988). Soltis and colleagues (1989) suggested that variation of cp genome within species are consequences of cytoplasmic bottleneck during plant domestication. More recently, by applying NGS to characterize whole cp genomes of 11 *Panax ginseng* cultivars, significant variation in gene numbers between these cultivars were suggested causing by the deletion/insertion mutations (Kim *et al.,* 2015).

**Table 1.** Gene composition in this chloroplast genome

| Category of genes | Group of genes | Name of genes |
|---|---|---|
| Genes for photosynthesis | Subunits of ATP synthase I | *atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I |
| | Subunits of photosystem II | *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*I, *psb*J, *psb*K, *psb*M, *psb*N, *psb*T, *psb*Z, *ycf*3 |
| | Subunits of NADH-dehydrogenase | *ndh*B, *ndh*B, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K |
| | Subunits of cytochrome b/f complex | *pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N |
| | Subunits of photosystem I | *psa*A, *psa*B, *psa*C, *psa*I, *psa*J |
| | Subunit of rubisco | *rbc*L |
| Self-replication | Large subunit of ribosome | *rpl*14, *rpl*16, *rpl*2, *rpl*2, *rpl*20, *rpl*22, *rpl*23, *rpl*23, *rpl*32, *rpl*33, *rpl*36 |
| | DNA dependent tRNA polymerase | *rpo*A, *rpo*B, *rpo*C1, *rpo*C2 |
| | Transfer RNAs | *trn*N-GUU, *trn*R-ACG, *trn*A-UGC, *trn*E-UUC, *trn*V-GAC, *trn*L-CAA, *trn*M-CAU, *trn*H-GUG, *trn*K-UUU, *trn*Q-UUG, *trn*S-GCU, *trn*S-CGA, *trn*R-UCU, *trn*D-GUC, *trn*Y-GUA, *trn*E-UUC, *trn*T-GGU, *trn*S-UGA, *trn*G-GCC, *trn*M-CAU, *trn*S-GGA, *trn*T-UGU, *trn*L-UAA, *trn*F-GAA, *trn*M-CAU, *trn*W-CCA, *trn*P-UGG, *trn*M-CAU, *trn*L-CAA, *trn*V-GAC, *trn*E-UUC, *trn*A-UGC, *trn*R-ACG, *trn*N-GUU, *trn*L-UAG |
| | Small subunit of ribosome | *rps*11, *rps*12, *rps*12, *rps*14, *rps*15, *rps*16, *rps*18, *rps*19, *rps*2, *rps*3, *rps*4, *rps*7, *rps*7, *rps*8 |
| Other genes | Subunit of Acetyl-CoA-carboxylase | *acc*D |
| | c-type cytochrome synthesis gene | *ccs*A |
| | Envelop membrane protein | *cem*A |
| | Protease | *clp*P |
| | Maturase | *mat*K |
| | Conserved open reading frames | *ycf*1, *ycf*1, *ycf*1, *ycf*15, *ycf*15, *ycf*2, *ycf*2, *ycf*4 |

In total, 20 genes within the chloroplast genome were found to contain introns, of which 12 were protein-coding genes and 8 encoded transfer RNA (tRNA). Among these, 18 genes contained one intron, while

the genes *ycf*3 and *clp*P were identified with two introns each. The occurrence of two introns in the ycf3 gene has also been documented in *Salix wilsonii* (Chen *et al.,* 2019) and *G. orientalis* (Feng *et al.,* 2023). Notably, *ycf*1 contained the smallest intron, measuring 126 base pairs, while *trn*K-UUU had the largest intron, at 2,582 base pairs.

**Table 2.** The lengths of introns and exons for the splitting genes in cp genome of seedless jackfruit

| Gene | Strand | Start | End | Exon I | Intron I | Exon II | Intron II | Exon III |
|---|---|---|---|---|---|---|---|---|
| *trn*A-UGC | - | 5760 | 6634 | 37 | 802 | 36 | | |
| *trn*E-UUC | - | 6699 | 7714 | 32 | 944 | 40 | | |
| *ndh*B | + | 13140 | 15357 | 775 | 685 | 758 | | |
| *rpl*2 | + | 24082 | 25591 | 399 | 580 | 531 | | |
| *trn*K-UUU | - | 27497 | 30150 | 37 | 2582 | 35 | | |
| *rps*16 | - | 31207 | 32409 | 42 | 934 | 227 | | |
| *trn*S-CGA | + | 35564 | 36372 | 32 | 717 | 60 | | |
| *atp*F | - | 38447 | 39733 | 145 | 732 | 410 | | |
| *rpo*C1 | - | 47918 | 50773 | 430 | 789 | 1637 | | |
| *ycf*3 | - | 71324 | 73435 | 124 | 846 | 230 | 759 | 153 |
| *trn*L-UAA | + | 77036 | 77641 | 35 | 521 | 50 | | |
| *trn*C-ACA | - | 80894 | 81584 | 39 | 596 | 56 | | |
| *clp*P | - | 100335 | 102504 | 71 | 885 | 294 | 694 | 226 |
| *rpl*16 | - | 111585 | 113029 | 9 | 1037 | 399 | | |
| *rpl*2 | - | 114800 | 116309 | 399 | 580 | 531 | | |
| *ndh*B | - | 125034 | 127251 | 775 | 685 | 758 | | |
| *trn*E-UUC | + | 132677 | 133692 | 32 | 944 | 40 | | |
| *trn*A-UGC | + | 133757 | 134631 | 37 | 802 | 36 | | |
| *ndh*A | - | 151416 | 153676 | 553 | 1169 | 539 | | |
| *ycf*1 | - | 155610 | 159692 | 349 | 126 | 3608 | | |

Codon usage was analyzed based on 84 identified protein-coding gene sequences from the plastid genome of seedless jackfruit (Table 3). A total of 49,145 codons, encompassing 64 different codon types that encode all 20 amino acids, were identified. The total number of codons in jackfruit is higher than that found in the chloroplast genome of *Salix wilsonii*, which contained 25,899 codons (Chen *et al*., 2019), but lower than the range observed in various *Ficus* species, which ranged from 53,412 to 53,566 codons (Huang *et al.,* 2022). Leucine was the most abundant amino acid, represented by 5,186 codons, accounting for approximately 10.55% of the total codons, while cysteine was the least abundant, with only 600 codons (about 1.22% of the total). This finding aligns with previous reports by Chen *et al.* (2019) for *S. wilsonii* and Ren *et al.* (2022) for *Coleanthus subtilis*. The most frequently occurring codon was ATT, which encodes isoleucine and comprised approximately 2,009 codons (about 4.01% of the total), while, excluding stop codons, TGC, which encodes cysteine, was the least abundant codon, with 167 occurrences (~0.34% of the total) (Table 3). Analyzing codon usage patterns and nucleotide composition provides a theoretical foundation for genetic modifications of the chloroplast genome, as these preferences are closely related to gene expression and influence protein and mRNA levels within the genome.

**Table 3.** Codon usage in the seedless jackfruit chloroplast genome

| Codon | Amino acid | Number | Frequency (%) | Codon | Amino acid | Number | Frequency (%) |
|---|---|---|---|---|---|---|---|
| GCA | Ala | 708 | 14.41 | CCA | Pro | 570 | 11.60 |
| GCC | Ala | 369 | 7.51 | CCC | Pro | 351 | 7.14 |
| GCG | Ala | 280 | 5.70 | CCG | Pro | 316 | 6.43 |
| GCT | Ala | 1095 | 22.28 | CCT | Pro | 729 | 14.83 |
| TGC | Cys | 167 | 3.40 | CAA | Gln | 1308 | 26.62 |
| TGT | Cys | 433 | 8.81 | CAG | Gln | 410 | 8.34 |
| GAC | Asp | 382 | 7.77 | AGA | Arg | 934 | 19.005 |
| GAT | Asp | 1519 | 30.91 | AGG | Arg | 371 | 7.549 |
| GAA | Glu | 1790 | 36.42 | CGA | Arg | 655 | 13.33 |
| GAG | Glu | 655 | 13.33 | CGC | Arg | 183 | 3.72 |
| TTC | Phe | 1051 | 21.39 | CGG | Arg | 214 | 4.35 |
| TTT | Phe | 1880 | 38.25 | CGT | Arg | 604 | 12.29 |
| GGA | Gly | 1285 | 26.15 | AGC | Ser | 265 | 5.39 |
| GGC | Gly | 353 | 7.18 | AGT | Ser | 747 | 15.20 |
| GGG | Gly | 499 | 10.15 | TCA | Ser | 772 | 15.71 |
| GGT | Gly | 1035 | 21.06 | TCC | Ser | 607 | 12.35 |
| CAC | His | 295 | 6.00 | TCG | Ser | 348 | 7.08 |
| CAT | His | 907 | 18.46 | TCT | Ser | 1111 | 22.61 |
| ATA | Ile | 1340 | 27.27 | ACA | Thr | 749 | 15.24 |
| ATC | Ile | 808 | 16.44 | ACC | Thr | 411 | 8.36 |
| ATT | Ile | 2009 | 40.88 | ACG | Thr | 287 | 5.84 |
| AAA | Lys | 1950 | 39.60 | ACT | Thr | 1011 | 20.57 |
| AAG | Lys | 742 | 15.10 | GTA | Val | 989 | 20.12 |
| CTA | Leu | 681 | 13.86 | GTC | Val | 339 | 6.90 |
| CTC | Leu | 372 | 7.57 | GTG | Val | 343 | 6.98 |
| CTG | Leu | 340 | 6.92 | GTT | Val | 914 | 18.60 |
| CTT | Leu | 1096 | 22.30 | TGG | Trp | 857 | 17.44 |
| TTA | Leu | 1557 | 31.68 | TAC | Tyr | 378 | 7.69 |
| TTG | Leu | 1140 | 23.20 | TAT | Tyr | 1492 | 30.36 |
| ATG | Met | 1168 | 23.77 | TAA | Stop | 294 | 5.98 |
| AAC | Asn | 604 | 12.29 | TAG | Stop | 145 | 2.95 |
| AAT | Asn | 1766 | 35.93 | TGA | Stop | 165 | 3.36 |

*Repeat structure and simple sequence repeats*

A total of 784 simple sequence repeats (SSRs) were identified across 14 chloroplast genomes within the genus *Artocarpus*. The species with the fewest SSRs was *Artocarpus tonkinensis* (MZ379793), which exhibited 51 SSRs, while *Artocarpus altilis* (MW367444.1) had the highest number, with 67 SSRs, resulting in an average of 56 SSRs per chloroplast genome. Nine distinct repeat sequences were detected: A, C, G, T, AT, TA, AAT, ATT, and TAA. Among these, the mononucleotide types T and A were the most prevalent, with frequencies of 458 (58.4%) and 245 (31.3%), respectively. The abundance of these mononucleotide SSR motifs has also been observed in other plant species, such as *S. wilsonii* (Chen *et al*., 2019) and *Dalbergia* (Song *et al.,* 2019).

In contrast, the mononucleotide types C and G exhibited lower repeat frequencies, with only 6 (0.8%) and 21 (2.7%), respectively. The dinucleotide sequences TA and AT were identified with relative frequencies of 1.2% and 5.4%, respectively. Trinucleotide sequences such as AAT were exclusively found in *Artocarpus excelsus* (OR664167), while the sequences ATT and TAA were only present in *A. gomezianus* (NC_080592.1).

The number of simple sequence repeats (SSRs) identified within the genus *Artocarpus* is generally lower than that observed in the two species of the genus *Morus*, specifically *M. atropurpurea* and *M. multicaulis*, which contain 83 and 81 SSRs, respectively (Li, 2016). In comparison, a total of 18 distinct SSR sequences were identified in these two *Morus* species, while only 8 SSR sequences were detected in the *Artocarpus* genus, including 5 SSR sequences specific to the seedless jackfruit variety.

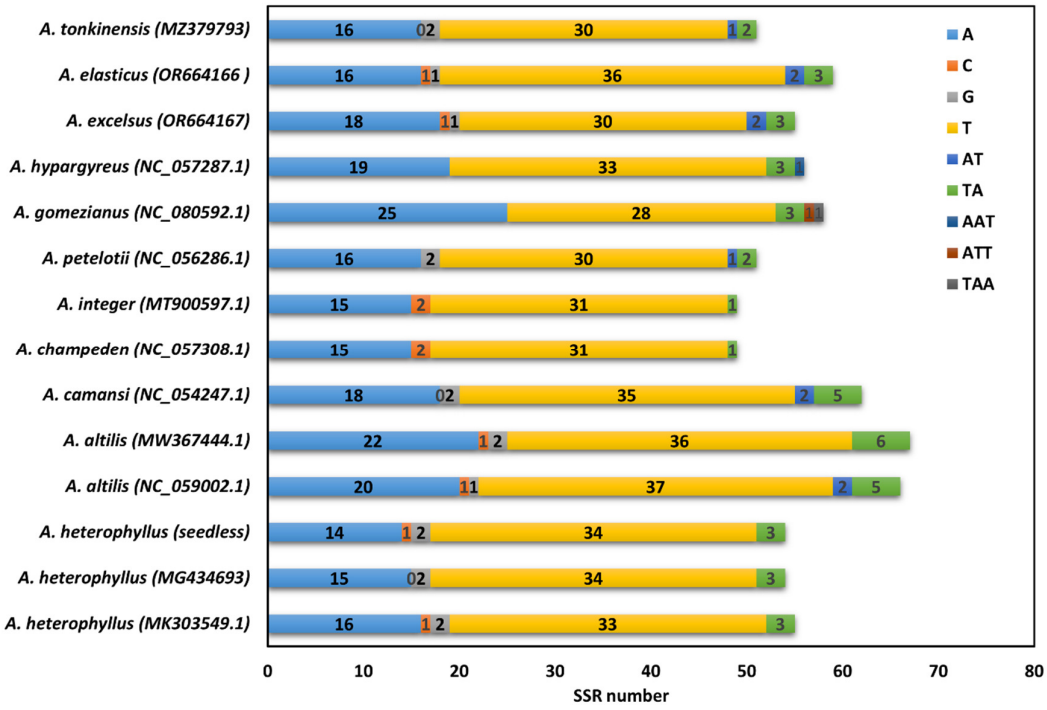


**Figure 2.** The different simple sequences repeat types in the cp genomes of 14 cp sequences from 11 species in *Artocarpus* genus

The REPuter program was utilized to analyze nine chloroplast (cp) sequences and evaluate the abundance of four types of oligonucleotide repeats: forward (F), palindromic (P), reverse (R), and complementary (C). The number and type of repeat elements displayed considerable variation among the nine cp genomes (Table 4), with a range of 38 units identified in *A. chapaensis* to 53 units in *A. zhejiangensis*. Notably, no complementary repeats were detected in the three cp genomes of *A. heterophyllus*.

**Table 4.** Number of repeated sequences in 14 cp genome sequences of 11 species in *Artocarpus* genus

| No. | Accession | F vs. F | F vs. C | F vs. R | F vs. RC* |
|---|---|---|---|---|---|
| 1 | *A. heterophyllus* (MK303549.1) | 21 | 0 | 4 | 24 |
| 2 | *A. heterophyllus* (MG434693) | 22 | 0 | 4 | 24 |
| 3 | *A. heterophyllus* (seedless) | 22 | 0 | 4 | 24 |
| 4 | *A. altilis* (NC_059002.1) | 7 | 5 | 11 | 18 |
| 5 | *A. altilis* (MW367444.1) | 20 | 0 | 4 | 25 |
| 6 | *A. camansi* (NC_054247.1) | 18 | 1 | 5 | 26 |
| 7 | *A. champeden* (NC_057308.1) | 18 | 1 | 5 | 26 |
| 8 | *A. integer* (MT900597.1) | 16 | 0 | 6 | 28 |
| 9 | *A. petelotii* (NC_056286.1) | 19 | 2 | 4 | 24 |
| 10 | *A. gomezianus* (NC_080592.1) | 28 | 0 | 3 | 18 |
| 11 | *A. hypargyreus* (NC_057287.1) | 28 | 0 | 3 | 18 |
| 12 | *A. excelsus* (OR664167) | 14 | 1 | 4 | 24 |
| 13 | *A. elasticus* (OR664166) | 14 | 1 | 4 | 24 |
| 14 | *A. tonkinensis* (MZ379793) | 18 | 3 | 4 | 24 |

(*F: forward; R: reverse; C: complement; RC: reverse complement)

To align the chloroplast genomes of the 11 *Artocarpus* species, the reference sequence MK303549.1 was utilized in conjunction with the mMISTA software for sequence alignment (Figure 2). Overall, the sizes and gene arrangements of the nine analyzed cp genomes exhibited a high degree of conservation.

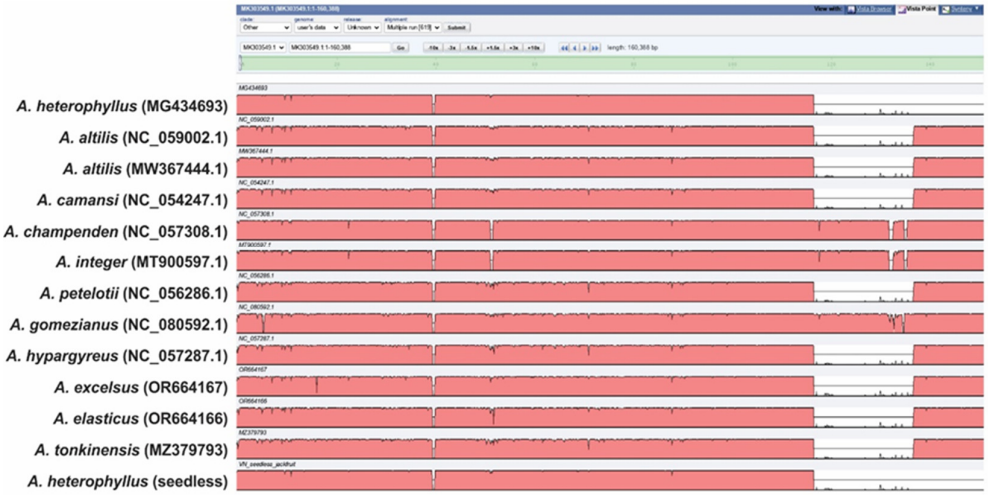


**Figure 2.** A sequence identity plot was generated to compare 13 chloroplast genomes of the *Artocarpus* genus, using *A. heterophyllus* (MK303549.1) as the reference, through the mVISTA tool. A cutoff threshold of 70% identity was applied for the plots, with the Y-axis indicating percent identity ranging from 50% to 100%

The Neighbor-Joining method is a distance-based approach for constructing phylogenetic trees, noted for its speed in tracing evolutionary relationships. In contrast, the Maximum Likelihood method evaluates the likelihood of each potential tree and identifies the one with the highest probability, yielding more accurate and detailed phylogenetic information. Consequently, the phylogenetic tree generated by this method should be employed to validate the results obtained from the Neighbor-Joining approach. Both methods produced

relatively similar phylogenetic outcomes. All 17 chloroplast genome sequences analyzed (comprising 14 *Artocarpus* species and three outgroups) were classified into two major clades. Clade I included four varieties: *A. champeden* (NC_057308.1), *A. integer* (MT900597.1), *A. heterophyllus* (MK303549.1), and *A. gomezianus* (NC_080592.1), while the remaining varieties were grouped within clade II. The chloroplast genomes of the outgroup varieties formed a small clade within clade II. The seedless jackfruit was classified into clade II and exhibited a close relationship with *A. heterophyllus* (MG434693) due to their placement in the same subclade. However, despite both being classified as *A. heterophyllus*, the genome *A. heterophyllus* (MK303549.1) was situated in clade I, indicating significant genetic divergence from the seedless jackfruit. These findings are consistent with the report by de Souza *et al*. (2021), which grouped *A. heterophyllus* with *A. altilis* and *A. camansi*.

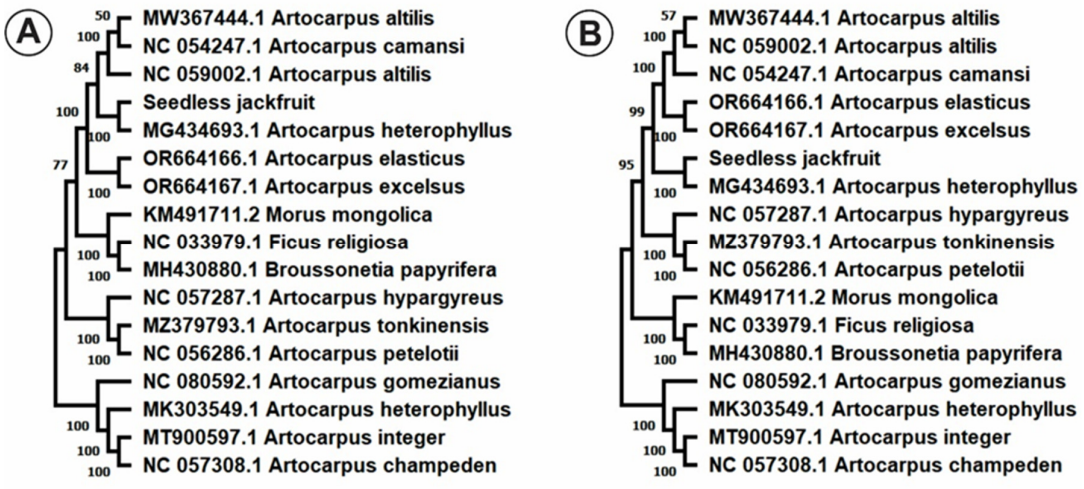


**Figure 4.** The phylogenetic tree of 14 *Artocarpus* chloroplast genomes was constructed using the Neighbor-Joining method (A) and the Maximum Likelihood method (B). The chloroplast genomes of *Morus mongolica*, *Ficus religiosa*, and *Broussonetia papyrifera* served as outgroups. Bootstrap values are indicated by numbers adjacent to the branches

## Discussion

The phylogenetic analysis conducted in this study, utilizing both Neighbor-Joining (NJ) and Maximum Likelihood (ML) methods, provides valuable insights into the evolutionary relationships among Artocarpus species and their outgroups. The congruence observed between the results from both methods reinforces the reliability of the phylogenetic relationships identified in this analysis. Similar patterns have been documented in previous studies comparing distance-based and likelihood-based methods, where ML trees are often employed as a benchmark for validating the results obtained from NJ approaches (Talavera and Castresana, 2007).

Furthermore, the distinct placement of *A. heterophyllus* (MK303549.1) and the seedless jackfruit in separate clades suggests notable genetic divergence within the species. This divergence may be attributed to factors such as geographic isolation, hybridization, or long-term evolutionary processes, all of which have been implicated in contributing to intraspecific genetic variation in other plant taxa (Avise, 2000). However, the unique positioning of *A. heterophyllus* (MK303549.1) in Clade I highlights the need for further investigation into the genetic and evolutionary mechanisms that drive this divergence.

## Conclusions

In this study, we sequenced and characterized the complete chloroplast genome of seedless jackfruit, a distinct variety from Vietnam. Comparative analyses with closely related species revealed notable features, including variations in genome size, gene count, and sequence repeat motifs. The data obtained offer valuable insights into the typical structure and composition of chloroplast genomes in seedless jackfruit from Vietnam. These differences contribute to our understanding of genetic architecture within the *Artocarpus* genus. Furthermore, the identification of unique repeat motifs and highly divergent regions in the chloroplast genome of Vietnamese seedless jackfruit presents opportunities for developing molecular markers. Such markers may be instrumental in future studies focusing on taxonomy and conservation initiatives for this valuable species in Vietnam.

## Authors' Contributions

Conceptualization: Ho, V. T.; Ngo, T. K. A.; Gebre, Y. G.; Nguyen, D. K; Widiarsih, S; Methodology: Ho, V. T.; Ngo, T. K. A.; Gebre, Y. G.; Nguyen, D. K; Widiarsih, S; Funding acquisition: Ngo, T.K.A.; Ho, V.T.; Writing – Original Draft: Ho, V. T.; Gebre, Y. G.; Nguyen, D. K; Widiarsih, S; Writing – Review and Editing: Ho, V. T.; Ngo, T. K. A.; Gebre, Y. G.; Nguyen, D. K; Widiarsih, S; Final approval: Ho, V. T.; Ngo, T. K. A.; Gebre, Y. G.; Nguyen, D. K; Widiarsih, S.

All authors read and approved the final manuscript.

## **Ethical approval** (for researches involving animals or humans)

Not applicable.

## Acknowledgements

This work was supported by the Ho Chi Minh City University of Industry and Trade- Vietnam through the HUIT fund for Science and Technology under the Contract No. 157/HD-DCT.

## Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

## References

Aseef RM, Manikandan K, Kavina M, Vijayakumar RM, Ganesan N M (2017). Biochemical evaluation of local genotypes of jackfruit (*Artocarpus heterophyllus* Lam.) in Pudukkotai District. Journal of Pharmacognosy and Phytochemistry 6(5):2533-2536. *https://www.phytojournal.com/archives/2017/vol6issue5/PartAK/6-5-439-773.pdf*

Avise JC (2000). Phylogeography: the history and formation of species. Harvard University Press.

Azad AK, Jones JG, Haq N (2007). Assessing morphological and isozyme variation of jackfruit (*Artocarpus heterophyllus* Lam.) in Bangladesh. Agroforestry Systems 71:109-125. *https://doi.org/10.1007/s10457-007-9039-8*

Balamaze J, Muyonga JH, Byaruhanga YB (2019). Physico-chemical characteristics of selected jackfruit (*Artocarpus heterophyllus* Lam) Varieties. Journal of Food Research 8(4):11- 22. *https://doi.org/10.5539/jfr.v8n4p11*

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics 33(16):2583-2585. *https://doi.org/10.1093/bioinformatics/btx198*

CBOL Plant Working Group (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106(31): 12794-12797. *https://doi.org/10.1073/pnas.0905845106*

Chen Y, Hu N, Wu H (2019). Analyzing and characterizing the chloroplast genome of *Salix wilsonii*. BioMed Research International 1:5190425. *https://doi.org/10.1155/2019/5190425*

Das S, Dang JC (2018). Applications of DNA barcoding for identification of horticultural plants. Transient 58-66. *https://donboscocollege.ac.in/public/frontend/pdf/Transient-VII-2018.pdf#page=65*

De Souza UJB, dos Santos RN, de Araujo FRN, dos Santos GR, Sarmento RA, de Bellis F, Campos FS (2021). The complete chloroplast genome of *Artocarpus altilis* (Moraceae) and phylogenetic relationships. Mitochondrial DNA part B 6(8):2291-2293. *https://doi.org/10.1080/23802359.2021.1945504*

Feng J, Xion Y, Su X, Liu T, Xiong Y, Zhao J.... Ma X (2023). Analysis of complete chloroplast genome: structure, phylogenetic relationships of *Galega orientalis* and evolutionary inference of Galegeae. Genes 14:176. *https://doi.org/10.3390/genes14010176*

Gardner EM, Laricchia KM, Murphy M, Ragone D, Scheffeler BE, Simpson S, Williams EW, Zerega NJC (2015). Chloroplast microsatellite markers for *Artocarpus* (Moraceae) developed from transcriptome sequences. Applications in Plant Sciences 3(9):1500049 *https://doi.org/10.3732/apps.1500049*

Ho VT, Ngo TKA, Phan HMC, Tran VTT (2019). Initial application of molecular markers in identificating seedless jackfruit. Proceedings National Biotechnology Conference (In Vietnamese) pp 474-477.

Huang Y, Li J, Yang Z, An W, Xie C, Liu S, Zheng X (2022). Comprehensive analysis of complete chloroplast genome and phylogenetic aspects of ten *Ficus* species. BMC Plant Biology 22:253. *https://doi.org/10.1186/s12870-022-03643-4*

Ibrahim M, Islam MS, Helali MOH, Alam AKMS, Shafique MZ (2013). Morphological fruit characters and nutritional food value of different jackfruit (*Artocarpus heterophyllus* Lam.) cultivars in Rajshahi region of Bangladesh. Bangladesh Journal of Scientific and Industrial Research 48(4):287-292. *https://doi.org/10.3329/bjsir.v48i4.18280*

Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, Park HS, Yang TJ (2015). Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. PloS One 10(6):e0117159 *https://doi.ort/10.1371/journal.pone.0117159*

Kress WJ (2017). Plant DNA barcodes: Applications today and in the future. Journal of Systematics and Evolution 55(4): 291-307. *https://doi.org/10.1111/jse.12254*

Krishnan AG, Sabu TS, Sible GV, Xavier L (2015). Genetic diversity analysis in jackfruit selections of Kuttanad region using RAPD technique. International Journal of Scientific and Research Publications 5(4):1-6. *https://www.ijsrp.org/research-paper-0415/ijsrp-p4059.pdf*

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research 29(22):4633-4642. *https://doi.org/10.1093/nar/29.22.4633*

Li Y, Song Y (2019). The chloroplast genome of an endangered tree *Artocarpus nanchuanensis* (Moraceae). Mitochondrial DNA part B 4(1):893-894. *https://doi.org/10.1080/23802359.2019.1574684*

Li DL, Gul J, He JM (2020). The complete chloroplast genome sequence of *Artocarpus hypargyreus*. Mitochondrial DNA part B 5(1):910-911. *https://doi.org/10.1080/23802359.2020.1715293*

Lin X, Lin T, Zhizia S, Priyadi A (2021). The complete chloroplast genome sequence of *Artocarpus gomezianus* (Moraceae) from Xishuangbanna, China. Mitochondrial DNA part B: Resources 6(10):2890-2892. *https://doi.org/10.1080/23802359.2021.1973603*

Liu J, Niu YF, Ni SB, Liu ZY, Zheng C, Shi C (2018). The complete chloroplast genome of *Artocarpus heterophyllus* (Moraceae). Mitochondrial DNA part B: Resources 3(1):13-14. *https://doi.org/10.1080/23802359.2017.1413317*

Liu S, Ni Y, Li J, Zhang Z, Yang Y, Chen H, Liu C (2021). CPGView: A package for visualizing detailed chloroplast genome structures. Molecular Ecology Resources 00:1-11. *https://doi.org/10.1111/1755-0998.13729*

Marjan PS, Krishnan AG, Cyriac A (2019). Molecular characterization and DNA fingerprinting of superior jackfruit genotypes from Kerala using SSR markers. International Journal of Current Microbiology and Applied Sciences 8(1):679-688. *https://doi.org/10.20546/ijcmas.2019.801.076*

Michu E (2007). A short guide to phylogenetic reconstruction. Plant, Soil and Environment 53(10):442-446.

Niu YF, Liu J (2021). The complete chloroplast genome and phylogenetic analysis of *Artocarpus champeden*. Mitochondrial DNA Part B 6(11):3148-3150. *https://doi.org/10.1080/23802359.2021.1987165*

Palmer JD, Jansen RK, Michaels HJ, Chase MW, Manhart JR (1988). Chloroplast DNA variation and plant phylogenny. Annals of the Missouri Botanical Graden 75(4):1180-1206.

Ren J, Tian J, Jian H, Zhu XX, Mutie FM, Wanga VO, ... Hu GW (2022). Comparative and phylogenetic analysis based on the chloroplast genome of *Coleanthus subtilis* (Tratt.) Seidel, a protected rare species of monotypic genus. Frontiers in Plant Science 13. *https://doi.org/10.3389/fpls.2022.828467*

Sampath PM, Shyamalamma S, Swamy GSK, Honnabyraiah MK (2020). A Review: Molecular markers in jackfruit (*Artocarpus heterophyllus* Lam.). Environment and Ecology 38(2):190-197. *https://www.cabidigitallibrary.org/doi/pdf/10.5555/20203173930*

Song Y, Zhang Y, Xu J, Li W, Li M (2019). Characterization of the complete chloroplast genome sequence of *Dalbergia* species and its phylogenetic implications. Scientific Reports 9:4368.

Soltis DE, Soltis PS, Ranker TA, Ness BD (1989). Chloroplast DNA variation in a wild plant, *Tolmiea menziesii*. Genetics 12:819-826.

Suneel RK, Jagadeesh SL, Umesh BC, Netravati, Bharathkumar A (2022). Identification of elite jackfruit (*Artocarpus heterophyllus* Lam.) genotypes for fruit characters, bulb characters and biochemical parameters in Doddaballapur and Tumkur districts of Karnataka, India (in Indian province of Karnataka). The Pharma Innovation SP-11(10):2368-2372. *https://www.thepharmajournal.com/archives/2022/vol11issue10S/PartAB/S-11-10-300-434.pdf*

Talavera G, Castresana J (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56(4):564-577. *https://doi.org/10.1080/10635150701472164*

Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, ... Zhu L (2004). A comparison of rice chloroplast genomes. Plant physiology 135: 412-420. *https://doi.org/10.1104/pp.103.031245*.

Tang LM, Liu XL, Lin JY, Qui B, Chen CZ (2021). The complete chloroplast genome of *Artocarpus tonkinensis*, a tree native to China with diverse beneficial medicinal applications. Mitochondrial DNA part B: Resources 7(1):4-5. *https://doi.org/10.1080/23802359.2021.1997116*

Wang JX, Zhang XF, Wang JH, Zhao KK, Zhu ZX, Wang HF (2019). Complete plastome sequence of *Artocarpus heterophyllus* Lam. (Moraceae): An ediable fruit cultivated throughout the tropics. Mitochondrial DNA part B: Resources 4(1):977-978. *https://doi.org/10.1080/23802359.2019.1581105*

Wei L, Wang HX, Chen DJ, Ke XR, Zhu ZX, Wang HF (2023). The complete plastome sequence of *Artocarpus altilis* (Parkinson ex F.A. Zorn) Fosberg, (Moraceae). Mtochondrial DNA Part B 8(1):154-156. *https://doi.org/10.1080/23802359.2022.2163595*.

**Notes:**

- Material disclaimer: The authors are fully responsible for their work and they hold sole responsibility for the articles published in the journal.
- Maps and affiliations: The publisher stay neutral with regard to jurisdictional claims in published maps and institutional affiliations.
- Responsibilities: The editors, editorial board and publisher do not assume any responsibility for the article's contents and for the authors' views expressed in their contributions. The statements and opinions published represent the views of the authors or persons to whom they are credited. Publication of research information does not constitute a recommendation or endorsement of products involved.