

# Cloning and Sequence Analysis of 5S Ribosomal RNA Gene(s) and Associated Intergenic Spacer Regions in *Carica* Species

Mahipal Singh<sup>1</sup>, Pushpa Yadav<sup>2</sup> & Anand K. Yadav<sup>1</sup>

<sup>1</sup> Agricultural Research Station, Fort Valley State University, USA

<sup>2</sup> Department of Biology, Middle Georgia State University, Macon, USA

Correspondence: Mahipal Singh, Agricultural Research Station, 1005 State University Drive, Fort Valley State University, Fort Valley, GA 31030, USA. Fax: 478-825-6376. E-mail: singhm@fvsu.edu

Received: September 5, 2019

Accepted: October 1, 2020

Online Published: March 30, 2021

doi:10.5539/ijb.v13n1p1

URL: doi: 10.5539/ijb.v13n1p1

## Abstract

The 5S ribosomal RNA gene(s) and their associated intergenic spacer regions were amplified from *Carica papaya* and *Carica quercifolia* by polymerase chain reaction. Both *Carica* species exhibited differently sized amplification products. Sequence analysis of these PCR products revealed that the 5S rRNA genes are arranged as tandem repeats in these regions. Sequence data revealed that the 5S rRNA gene from *Carica quercifolia* was 119 bp in length. Sequence variation was observed in various 5S rRNA gene copies cloned from *Carica quercifolia*. Only truncated 5S rRNA gene but with its full spacer region was recovered from *Carica papaya*. Interestingly, intergenic spacer sequence cloned from *Carica papaya* contained two specific domains, a 30bp “CT” rich domain exhibiting 95-100% homology to several human chromosomes and a domain matching with mitrocomin precursor, a photo-protein from *Mitrocoma cellularia*. The role of 5S rRNA gene and their spacer regions in discerning the germplasm and in adaptation of the species is discussed.

**Keywords:** Carica Papaya, Carica Quercifolia, Genetic Variation, Ribosomal RNA Genes, Spacer Region, Tandem Repeats, 5S rRNA Genes

## 1. Introduction

Ribosomal RNA (rRNA) genes are present in multiple copies per genome in all the eukaryotes studied. Nuclear encoded 5S rRNA genes (5S rDNA) occur in high copy number as tandemly arranged repeats with a highly conserved coding region (Long and David, 1980; Specht *et al.*, 1990) and a divergent non-transcribed spacer region in eukaryotes (Flavell, 1986; Rogers *et al.*, 1986; Reddy and Appels, 1989; Trontin *et al.*, 1999; Singh and Singh, 2001). The number of 5S rRNA genes in plants is often much higher than those of 18S, 5.8S and 25S rRNA genes and except in *Marchantia polymorpha* (Sone *et al.*, 2000), the 5S rRNA genes are usually not linked to these genes. They provide valuable genetic markers for the analysis of genome relationships and phylogenetic reconstruction (Terauchi *et al.*, 1992), in differentiating plant varieties (Nybom *et al.*, 1992), and in identifying plants in admixtures (Ko and Henry; 1996, Dhiman and Singh, 2003).

*Carica* is the largest genus of the Caricaceae family with about 48 species. *Carica papaya* L. commonly known as “papaya” is a commercial fruit crop species that provides highly nutritious fruits. *Carica quercifolia* is a mountainous species. Published reports on ribosomal RNA gene (s) in papaya are lacking except a sequence of 18S-26S rRNA in GeneBank. The objective of this study, therefore, was to clone 5S ribosomal RNA gene(s) from *Carica* species and to understand their organization in the genome, in order to provide tools to study role of 5S rRNA genes and associated spacers. This will be useful in characterization of huge *Carica* germplasm as well as in understanding adaptation of these species to temperatures.

## 2. Material and Methods

### 2.1 Plant Material and DNA Isolation

In this study two *Carica* species, *Carica papaya* (HCAR-80) and *Carica quercifolia* (HCAR-226) were used. Young plants (2-3 month old), obtained by germinating seeds in the greenhouse, were used for DNA isolation. Total genomic DNA was extracted from young fresh leaves using the molecular reagents and procedure from Genelute Plant Genomic DNA Kit (Sigma Chemicals, St. Louis, MO). In brief, fresh leaf tissue (200 mg) was powdered in a pestle and mortar using liquid nitrogen. To this powdered material, 350 µl of lysis buffer was added

and kept at 65°C for 10 min. To the lysate, 130 µl precipitation buffer was added and kept in ice for 5 min. The solution was passed through a filtration column. To this eluted solution 700 µl of binding buffer was added and then passed through a binding column. The column was washed twice with wash buffer. The DNA was eluted in 100 µl of TE buffer and kept frozen at -20°C until used.

### 2.2 Polymerase Chain Reaction and Cloning of the PCR Products

The 5S rRNA gene and their associated spacer regions were amplified by using consensus primers complementary to and based on the sequence of 3' [M27 5'-TTTAGTGCTGGTATGATCGC-3'] and 5' [M28 5'-TGGGAAGTCCTCGTGTGCA-3'] ends of 5S rRNA gene coding regions (5S rDNA) from plants as described (Singh and Singh, 2001). The primers were custom synthesized (Genemed Synthesis Inc., San Francisco, CA). Advantage Taq PCR kit (Clontech Laboratories Inc, Palo Alto, CA) was used to amplify the genomic DNA. The 25 µl PCR reaction mix contained 2.5 µl 10X PCR buffer, 1 µl of dNTPs (10µM each), 1 µl (200 µM) each of the forward and reverse primers, 1.5 mM MgCl<sub>2</sub>, 10 ng of template DNA and 0.5 units of *Taq* DNA polymerase. The PCR reaction was performed in an Icyler machine (BioRad) using hot start set at 94°C for 3 min and then 35 cycles of 94°C, 1 min; 68°C, 1 min; 72°C, 1 min and a final 72°C expansion cycle for 7 min. The PCR products were fractionated in 1.4% agarose gel. The individual bands were excised with a razor blade and purified using the QIAquick Gel Extraction Kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions. The individual fragments were cloned in pCR4-TOPO vector (Invitrogen, Inc.) as per manufacturer's instructions. The clones were selected on Ampicillin (50 µg/mL) LB plates. The individual colonies were picked up and analyzed by colony PCR using T3 and T7 primers as forward and reverse primers, respectively. The DNA of the selected clones was isolated from each colony using a DNA isolation kit (Qiagen Inc., Valencia, CA).

### 2.3 Sequencing and Sequence Analysis

Nucleotide sequences of the cloned PCR DNA fragments were obtained on an ABI prism automatic sequencer (Laragen Inc., Los Angeles, CA) using T3 and/or T7 as sequencing primers. Individual fragment sequences were analyzed using BLAST (Altschul et al., 1997).

## 3. Results and Discussion

### 3.1 Gene Cloning Approach

The 5S rRNA genes are arranged as tandem repeats in several plant species (Szymanski et al., 2000). Assuming it to be true for *Carica* species, we followed a cloning approach as shown diagrammatically in Figure 1. In this approach, the consensus primers complementary to and based on the sequence of the 3' and 5' ends of the 5S rRNA gene coding regions are designed as forward and reverse primers, respectively. These primers are then used to amplify the 5S ribosomal RNA genes and their associated spacer regions. In this approach, the larger amplification products theoretically should contain the intact 5S rRNA gene or even multiple copies of the gene, if these genes are arranged as tandem repeats in the genome. For example, the second predicted PCR product (fig. 1) should contain one copy of the 5S rRNA gene and the third product should contain two copies.

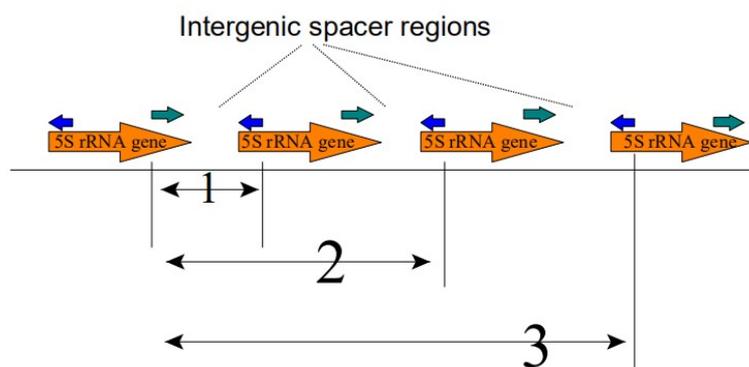


Figure 1. Diagrammatic sketch to elucidate the 5S rRNA gene cloning approach: Numbers 1, 2 and 3 are the possible amplified PCR products containing putative 5S rRNA gene copies

### 3.2 Amplification of 5S Ribosomal RNA Gene Containing DNA Fragments Support Their Tandem Arrangement in Genome

The 5S rRNA gene repeat units were amplified from *Carica quercifolia* and *Carica papaya* genome using the consensus 5S rRNA gene based PCR primers M27 and M28. Several PCR-amplified products ranging from 300 bp to 2000 bp could be seen in the agarose gel (Figure 2). The intensity of PCR products on agarose gel seem to decrease from lower to the higher molecular size (bottom to top of the gel). It supports the tandem arrangement of the 5S rRNA gene repeat units, since it takes longer time for DNA polymerase to travel through the larger DNA sequence as compared to a shorter sequence and, therefore, the turnover rate of PCR products is higher for smaller sequences. In addition, the size of the amplified products appears to be multiples of the lowest molecular weight PCR products which also support their tandem arrangement. As seen in figure 2, PCR-amplification of each *Carica* species results in different sizes of the amplified products and thus they seem to contain different repeat units. The sequence analysis of these clones supports this observation.

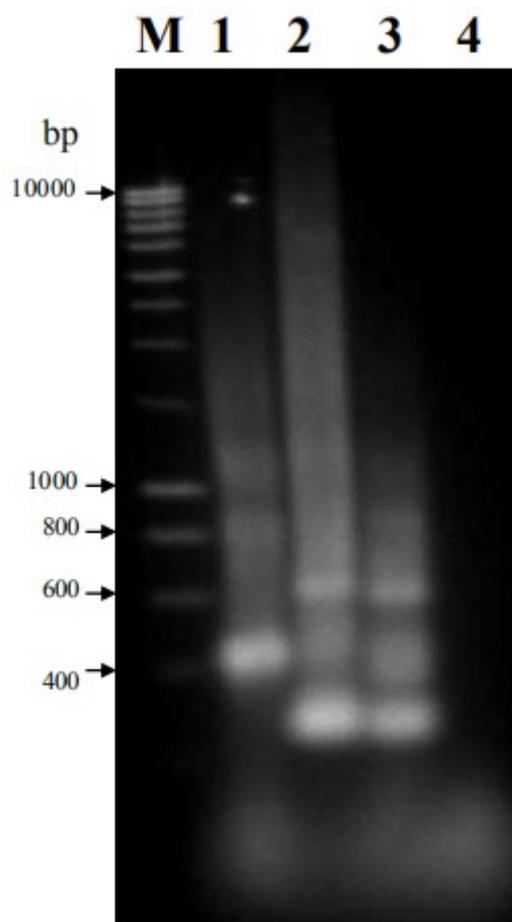


Figure 2. PCR amplified products obtained by using primers M27 and M28 and separated in a 1.4% agarose gel. Lane 1 represent *Carica quercifolia*, lane 2 and 3 represent *C. papaya*. Lane 4 is control reaction without template DNA. M is mol. wt. marker (DNA ladder)

### 3.3 Cloning and Analysis of Amplified Fragments in pCR4-TOPO Vector

A total of nine amplified products of different sizes (four from *Carica quercifolia* and five from *Carica papaya*) were randomly selected. They were cloned in pCR4-TOPO vector and tested for the right inserts by colony PCR, using T3 and T7 as forward and reverse primers, respectively. A representative sample of colony PCR is shown in Figure 3. Selected clones were also tested for the right insert by *Eco* R1 digestion (data not shown). A total of eight clones representing four differently sized PCR products of *Carica quercifolia* and two clones of *Carica papaya* (which represented the best-characterized inserts, by both colony PCR and *Eco* R1 digestion) were selected and their inserts sequenced from both directions using T3 and T7 primers.

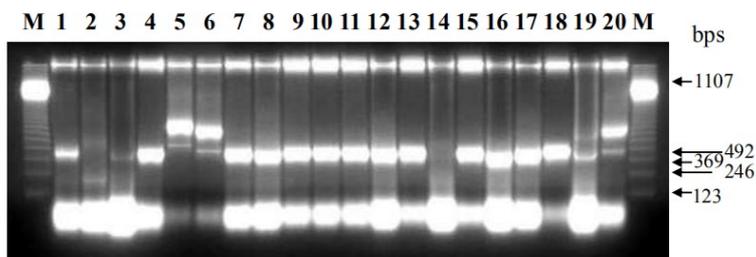


Figure 3. Agarose gel electrophoresis of PCR products from 20 different clones using primers T3 and T7. Lanes 1-20 show 20 different clones. M is mol. wt. marker (a 123 bp ladder)

### 3.4 Sequence Analysis of the 5S Ribosomal RNA Gene and Their Associated Spacers in *Carica Quercifolia*

Two of the eight clones of *C. quercifolia* (Cq) contained only truncated parts of the 5S rRNA gene with full intergenic spacer regions. Since these two clones represented the lowest-sized PCR bands, this result was in full agreement with our cloning strategy. The other six clones contained at least one full copy of the 5S rRNA gene. The spacer regions of the clones differed in their sequence and size. Here, we show sequence analysis of clone Cq2 which contained 586 bp of the insert sequence (Figure 4). BLAST analysis of the complete 586-base sequence indicated that there are two regions of homology. Region one contained sequence from 1-34 bases and region two from 357–475 bases. It was observed that these two regions contain 84-97% homology with 5S ribosomal RNA genes of other plant species. None of the sequences present in the available databases exhibited 100% homology to Cq2 sequence including the two above mentioned 5S rRNA gene homology regions and thus 5S rRNA gene present in *C. quercifolia* seems to be unique to this species.

```

1  TTTAGTGTCTG GTATGATCGC ACCCGTCATG TTTCGTGCTG CATATTCCTA TACCTCTTCC
   Truncated 5S rRNA gene
61 TCTCGAACGC GATCCGGTGC AACCGTCGTT CGCAATAGCC CCCTTCGAGT ACTATAAACG
121 TAATGCATT CTGAATGCAT CTCCTCGGGG TCGTTCCTCC TAAAAACATT GCTTGCCCCA
181 TGATTACTTT CTCITTAATAT AACTTTCACG GCTGTGTCAC TCTTTCTCCC AAAAAAAAAA
241 CAAGTATCTT TTTTAACCAA ATGACCACCG TCGTTGCCAA TTGCCACTAC CAGCAGCATC
301 GCCAAGCCTC TAGTGACAAC TGTACCGCAT CGTCCCCGG ATCCAAAAAA ACGGGCGGGT
361 GCAACACGAG GACTTCCAG GAGGTCACCC ATCTAGTAC TACTCTCGCC CAAGCACGCT
421 TCACTGCGGA GTTGTGATGG GATCCGGTGC ATTANTGGCT GGTATGATCG CACCCGTCAT
   Full length 5S rRNA gene sequence
481 GGATCAGCAT CGCCAAGTCT CTAACGATAA TTGTACAGGC ATCGAATCCC CGGAATCCTT
541 CCCTGGGGAC CGAAAAAAAAA CGGGGCGGGT GGCAACACCA AGGACAA

```

Figure 4. Complete sequence of the 0.6 kb insert of the clone Cq2-T3 containing a full copy of 5S rRNA gene (Bases 357-475 as shown by underline) as well as truncated gene (Bases 1-24) cloned from *Carica quercifolia* plants. Arrows indicate 5'-3' orientation of the sequence

### 3.5 Variation in the 5S rRNA Gene Sequences in *C. Quercifolia*

Upon aligning the 5S rRNA gene from clone Cq2 with eight other clones from *Carica quercifolia* that were sequenced in this study, surprisingly, only 85-90% bases were identical and thus variation in the 5S rRNA gene sequence, exist within *C. quercifolia* species. Earlier sequence differences in 5S rRNA genes have been reported in plants. Although 5S rRNA gene sequences are highly conserved, it is intriguing to note that the sequence differences exist in the copies of a gene within the same genome of a species. Further study is needed to determine the specific role played by these differences in the RNA gene sequences in plants. It seems that along with the multiple copies of the 5S rRNA gene sequence, variations in 5S rRNA gene copies within a single species and probably in the individual copies of a gene in the *C. quercifolia* genome is an important adaptation for survival of these plants in colder climates. It could be a mechanism to create a population of heterogeneous transcripts, which could lead to the stability of the ribosomes, the protein synthesizing machinery of the plants at colder climates. The heterogeneity may enable plants to enhanced adaptation or survival in adverse weather conditions, for example, heat, cold or chemical stress.

### 3.6 Sequence Analysis of the 5S rRNA Gene and Their Associated Intergenic Spacers in *Carica Papaya*

Sequence analysis of the *Carica papaya* clones revealed only a truncated 5S rRNA gene but with a full non-transcribed intergenic spacer region of 213 bp (Figure 5). It was surrounded by the 5S rRNA conserved regions used as primers at both ends. When this 213 base long intergenic spacer sequence was analyzed using BLAST, 23 hits (as shown in Figure 6) were obtained showing homology exclusively at two regions in the sequence (domain 'A' and 'B'). Domain 'A' is 24bp [CTCCTTCTCCTTTCTTCCCCCTCC] long and exclusively comprises of "CT" bases, which match to various human genomic clones including chromosome 1, 6, 8, 11, 15, 16, 17, 18, and 19. It is important to note that this sequence did not show homology to any of the plant sequences except a 17 bp exact match to *Arabidopsis thaliana* putative peroxidase ATP8a mRNA (Yamada et al., 2002, Acc No AY063051). This sequence is usually found around a restriction endonuclease (*Not I*) recognition site. However, in this 213 base long sequence no *Not I* restriction site was seen. When analyzed for sequence tag sites (STS), none was found in either strand. Similarly, none of these two domains matched the binding sites for plant transcription factors. Moreover, it was observed that the 213 bp intergenic spacer sequence contained several sites for binding, predominantly the heat shock protein factors of *Drosophila* and other transcription factors from humans. These sites were predominantly present in the domains observed. It is intriguing to have observed the *Drosophila* heat shock factors in plants. Whether they are the plant's version of heat shock protein factors (HSP) remain to be seen. The 213 bp DNA fragment when translated into protein and analyzed using protein BLAST, gave one hit in frame II. It exhibited overall 72% homology to a 28 amino acid stretch of mitrocomin precursor, a photo protein from *Mitrocoma cellularia*. The homology was specifically at Ca<sup>2+</sup> binding domain (Fagan et al., 1993). This stretch of 84 nucleotides extends from almost the middle of the intergenic spacer region up to the 5' end (M28 complementary primer side) of the 5S rRNA gene (figure 5). This region of the 5S rRNA gene is most likely involved in regulation of transcription. It would be interesting to study the role of this region in transcript regulation of 5S rRNA gene in order to explore the dynamics and importance of this specific region. Taken together, it seems that spacer region of 5S rRNA genes are not lying in genome for only stabilizing but may be actively involved in regulation of these genes.

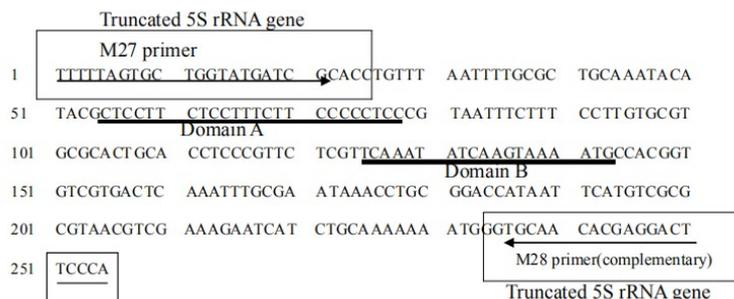


Figure 5. Complete nucleotide sequence of the 0.3 kb insert in clone Cp15-T3 containing truncated gene(s) cloned from *Carica papaya* plants. Note the two domains in intergenic spacer sequence showing homology to various other eukaryotic genes present in nucleotide sequence databases

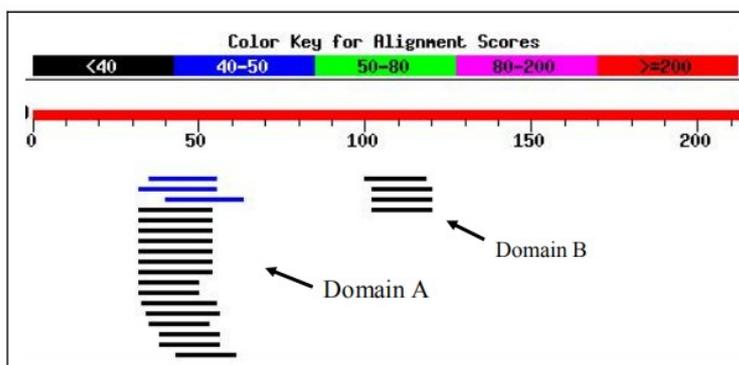


Figure 6. BLAST analysis of the 213 base long intergenic spacer region of *Carica papaya*. Note the two patches of nucleotide sequence (domains A and B) showing homology to several human chromosomes and other eukaryotic sequences

#### 4. Conclusion

In conclusion, we have isolated at least one full-length 5S ribosomal RNA gene from *Carica quercifolia* along with its spacer, while only the spacer sequence from *Carica papaya* was recovered. The 5S rRNA gene repeat units are of different sizes in both the species and are arranged in tandem. Sequence structure and spacers of both species are different.

#### Acknowledgements

We would like to thank Dr. Francis Zee of USDA, ARS, Hilo Hawaii for providing seeds of *Carica* species used in this study. The nucleotide sequence of 5S rRNA gene of *C. quercifolia* was deposited in GeneBank and bears the Accession Number AF506758.

#### Conflict of interests

Authors declare that there is no conflict of interests regarding the publication of this manuscript.

#### References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402. <https://doi.org/10.1093/nar/25.17.3389>
- Dhiman, B., & Singh, M. (2003). Molecular Detection of Cashew Husk (*Anacardium occidentale*) Adulteration in Market Samples of Dry Tea (*Camellia sinensis*). *Planta Medica*, 69(9), 882-884. <https://doi.org/10.1055/s-2003-43211>
- Fagan, T. F., Ohmiya, Y., Blinski, J. R., Inouye, S., & Tsuji, F. I. (1993). Cloning, expression and sequence analysis of cDNA for the Ca (2+) binding photo protein, mitrocomin. *FEBS Letters*, 333, 301-305. [https://doi.org/10.1016/0014-5793\(93\)80675-K](https://doi.org/10.1016/0014-5793(93)80675-K)
- Flavell, R. B. (1986). Repetitive DNA and chromosomes. *Philosophical Transactions of Royal Society Lond. SerB*, 312, 227-242. <https://doi.org/10.1098/rstb.1986.0004>
- Ko, H. L., & Henry, R. J. (1996). Specific 5S ribosomal RNA primers for plant species identification in admixtures. *Plant Molecular Biology Reports*, 14, 33-43. <https://doi.org/10.1007/BF02671900>
- Long, E. O., & David, I. B. (1980). Repetitive genes in eukaryotes. *Annual Review of Biochemistry*, 49, 727-764. <https://doi.org/10.1146/annurev.bi.49.070180.003455>
- Nybohm, H., Gardiner, S., & Simon, C. J. (1992). RFLPs obtained from an rDNA probe and detected with enhanced chemiluminescence in Apples. *Horticultural Sciences*, 27, 355-356. <https://doi.org/10.21273/HORTSCI.27.4.355>
- Reddy, P., & Appels, R. (1989). Organization and evolution of the 5S rRNA gene family in wheat and related species. *Genome*, 32, 1003-1005. <https://doi.org/10.1139/g89-545>
- Rogers, S. O., Honda, S., & Bendich, A. J. (1986). Variation in the ribosomal RNA genes among individuals of *Vicia faba*. *Plant Molecular Biology*, 6, 339-345. <https://doi.org/10.1007/BF00034941>
- Singh, D., & Singh, M. (2001). Organization of 5S ribosomal RNA genes in tea (*Camellia sinensis*). *Genome*, 44, 143-146. <https://doi.org/10.1139/g00-095>
- Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, M., Yamato, K., Ohmido, N., Fukui, K., Fukujawa, H., & Ohyama, K. (2000). Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. *Plant Molecular Biology*, 41, 676-685. <https://doi.org/10.1023/A:1006398419556>
- Specht, T., Wolters, J., & Erdmann, V. A. (1990). Compilation of 5S rRNA and 5S rRNA gene sequences. *Nucleic Acids Research*, 18, 2215-2230. <https://doi.org/10.1093/nar/18.suppl.2215>
- Szymanski, M., Barciszewska, M. Z., Barciszewski, J., & Erdmann, V. A. (2000). 5S ribosomal RNA database Y2K. *Nucleic Acids Research*, 28, 166-167. <https://doi.org/10.1093/nar/28.1.166>
- Terauchi, R., Chikaleke, V. A., Thottopilly, G., & Hahn, S. K. (1992). Origin and phylogeny of Guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theoretical Applied Genetics*, 83, 743-751. <https://doi.org/10.1007/BF00226693>

Trontin, J., Grandemange, C., & Favre, J. (1999). Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix decidua* Mill.) and Japanese larch (*Larix kaempferi* (Lamb.) Carr.). *Genome*, 42, 837-848. <https://doi.org/10.1139/g99-015>

Yamada et al. (2002). *Arabidopsis thaliana* putative peroxidase ATP8a (*At4g30170*) mRNA complete cds. Genebank Acc No: AY063051 (unpublished).

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).