Journal of Tropical Forest Science 36(2): 165–173 (2024)
https://doi.org/10.26525/jtfs2024.36.2.165
ISSN: 0128-1283, eISSN: 2521-9847

AN OPTIMISED METHOD FOR EXTRACTION OF HIGH-QUALITY RNA FOR RNA-SEQ FROM AMAZONIAN TREES

Camel V[1, *] & Zolla G[2]

[1]*Universidad César Vallejo, Professional School of Environmental Engineering, 13001 Lima, Peru*
[2]*Laboratorio de Fisiologia Molecular de Plantas del PIPS de Cereales y Granos Nativos, Facultad de Agronomía, Universidad Nacional Agraria La Molina, 15024 Lima, Peru*

**vcamelpa@ucv.edu.pe*

Submitted February 2023; accepted November 2023

The study aimed to develop a robust and reliable method to extract ribonucleic acid (RNA) from *Guazuma crinita* and *Calycophyllum spruceanum* woody tissues for RNA sequencing. Thus, 60 plants of each species were obtained, and four RNA extraction methods were evaluated (Direct-zol™ kit, TRIzol, cetyltrimethylammonium bromide (CTAB) and CTAB with Direct-zol™ kit). Subsequently, six samples for each species were selected for sequencing. As a result, the CTAB method combined with TRI-reagent and Direct-zol™ RNA kit obtained RNA of good quality in both species, as the RNA integrity number (RIN) values were higher than 7.9 in all samples. Inhibitors were removed in *G. crinita* with polymerase chain reaction (PCR) inhibitor removal kit. The sequencing process, on average per sample, were 4,559,130 kbp in 45,139,903 reads with a Q20 of 98.49% and Q30 of 95.59% obtained for *G. crinita* and 5,510,902 kbp in 54,563,395 reads with a Q20 of 98.23% and Q30 of 94.94% obtained for *C. spruceanum*. Thus, a robust method was developed, enabling the extraction of RNA in both quality and quantity for RNA-seq experiments, successfully overcoming the challenges posed by inhibitors and high concentrations of exudates, alkaloids, polysaccharides and phenols.

Keywords: CTAB, Direct-zol RNA kit, transcriptomic, TRIzol, RNA-seq, RNA

INTRODUCTION

The Amazon contains approximately 40% of the world's remaining tropical forests (Hubbell et al. 2008, Ter Steege et al. 2016). It is estimated to be home to 50,000 plant species, of which between 6,000 and 16,000 are tree species (Hubbell et al. 2008, Ter Steege et al. 2016). Currently, there is great concern about the rapid deforestation of the Amazon and the possible consequences, such as the extinction of species and genetic erosion, in addition to the reduction of ecosystem services (Degen & Sebbenn 2014, Cardoso et al. 2017, de Lima et al. 2022). On the other hand, due to the long-lived characteristics of trees, genetic diversity is essential for the stability of forest ecosystems (O'Neill et al. 2001). Therefore, proper measures for gene conservation are essential. However, there is currently limited genomic and transcriptomic information on Amazonian trees.

Twenty years ago, the complete genome of a model tree (*Populus trichocarpa*) was published (De Heredia & Vázquez-Poletti 2016). Until now, among the gymnosperms, the most studied species belong to the Pinaceae family. Among the angiosperm families, Salicaceae, Rosaceae, Fabaceae and Fagaceae, there are many species of great interest, such as, *Calycophyllum spruceanum* and *Guazuma crinita* (De Heredia & Vázquez-Poletti 2016). These species greatly interest the forestry industry, agroforestry and ecological restoration in the Amazon region. Biochemical studies consider *C. spruceanum* a medicinal tree of great interest due to the active components of its bark (Odonne et al. 2013, Perin et al. 2020). Pharmacologically, *C. spruceanum* exhibits antioxidant properties that help delay cell aging and photoprotection against UV radiation. It also has immunostimulant, anti-inflammatory, antileukemic, antiviral and antiseptic properties (Santos et al. 2016).

©*Forest Research Institute Malaysia*

On the other hand, wood is an abundant and vital renewable resource on Earth, and its biosynthesis regulates the carbon cycle within ecosystems, reducing CO_2 emissions (Ambavaram et al. 2014, Sundell et al. 2017).

Likewise, biotechnology tools can be used to understand the molecular processes of wood formation and speed up the plant breeding forest in South America. Extracting high-quality DNA and RNA is critical because of the large amounts of resins, polysaccharides, inhibitors, phenols and others that are present in trees (Vennapusa et al. 2020, Saldaña et al. 2022). For instance, phenolic compounds are oxidised to form quinones, binding irreversibly to nucleic acids and proteins. Polysaccharides can coprecipitate and degrade RNA by altering its molecular structure (Sharma et al. 2003). Thus, Ghawana et al. (2011) proposed a phenol-based method to isolate RNA in species with a high presence of secondary metabolites such as *Rheum austral* and *Arnebia euchroma* (Chomczynski 1993). Other protocols using Trizol method are designed to extract RNA from *Tectona grandis* (Galeano et al. 2014). Most protocols are relatively simple modifications of the cetyltrimethylammonium bromide (CTAB) method. However, Ouyang et al. (2014) reported that some kits performed poorly in isolating RNA from *Neolamarckia cadamba*, may be due to the decrease in the efficiency of the spin columns used in the presence of phenolic compounds and alkaloids. Therefore, efficiency improvement is required to combat the presence of inhibitors. Finally, the study aimed to develop a robust and reliable method of RNA isolation from woody tissues that can be used for RNA sequencing.

MATERIALS AND METHODS

Plant material

The evergreen *G. crinita* and the deciduous *C. spruceanum* are two fast-growing pioneer species in the Amazon (Figure 1). They are of great economic, social and environmental importance. Currently, they are used in reforestation and ecological restoration programs. The study worked with 2-year-old trees of *C. spruceanum* (Rubiaceae) and *G. crinita* (Malvaceae), obtained from the Ayahuasca Ecolodge EIRL nursery, Ucayali, Peru, and grown in a net house at the Universidad Nacional Agraria La Molina (12° 05' S, 76° 57' W

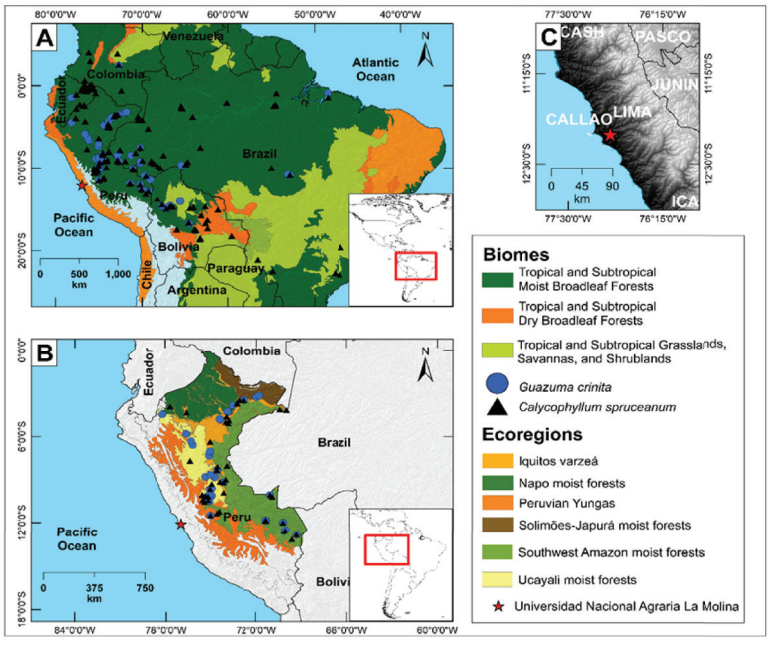


Figure 1 Distribution of *Guazuma crinita* and *Calycophyllum spruceanum* trees through the different biomes of South America, a) distribution of *Calycophyllum spruceanum* and *Guazuma crinita* trees in South America (Brazil, Peru, Colombia, Ecuador and Bolivia), b) Peru coverage of six ecoregions of the jungle, having a more significant presence in the region of Ucayali, Loreto and Madre de Dios, c) location of the Universidad Nacional Agraria La Molina center where the experiment was conducted

and 243.7 m.a.s.l.) (Figure 1). *Guazuma crinita* was cultivated throughout the experiment in a nursery covered by 70% radiation-interfering mesh. According to Paredes et al. (2010), these conditions ensure the optimal temperature, relative humidity and light for adapting bolaina plants. *Calycophyllum spruceanum*, was cultivated in a nursery with 50% light reduction, as in a previous studies by Abanto-Rodriguez et al. (2016) .

In both species, 60 individuals were carefully selected and assigned codes based on their phenotypic characteristics (stem diameter, height, straight stem, healthy leaves, etc.) and good phytosanitary status. Subsequently, they were planted in 12 L pots filled with a mixture of agricultural soil, organic matter and vermiculite in a 3:2:1 ratio. Irrigation was maintained consistently throughout the experiment to prevent water stress.

CTAB solution preparation

One liter of buffer solution containing 20 g of CTAB at 2% (w/v), 12.11 g of Tris-HCl (pH 8.0), 29.224 g of ethylenediaminetetraacetic acid (EDTA), 81.816 g of NaCl and 50.50 ml of beta-mercaptoethanol 5% (v/v) was prepared, and the solution was autoclaved and subsequently stored at 2 ˚C.

RNA extraction

With a scalpel, 1 g of secondary xylem was collected from the basal part of the stem, frozen in liquid nitrogen and stored at -80 ˚C. The frozen tissues were ground with liquid nitrogen in a mortar. Subsequently, the ground tissue was mixed with 20 ml of pre-heated buffer solution (CTAB) for 5 minutes at 65 ˚C, and the samples were homogenised for 1 minute and incubated at 65 ˚C for 30 min in a water bath. Then 20 ml of chloroform/isoamyl alcohol (24:1, v/v) was added and homogenised vigorously for 5 min, and centrifuged at 7800 rpm between 40 and 50 min at 2 ˚C. Since *G. crinita* has many exudates, the chloroform/isoamyl alcohol (24:1, v/v) procedure was repeated.

After centrifugation, 9 ml of supernatant was mixed with 3 ml of LiCl in a new falcon tube, homogenised and incubated at -20 ˚C for 12 hours. Then, it was centrifuged at 7800 rpm for 40 min at 2 ˚C to get a pellet. The pellet was dissolved in 1 ml of TRI-reagent in a 1.5 ml vial tube, mixed with 0.2 ml of chloroform, and centrifuged at 13000 rpm for 20 min at 2 ˚C. The aqueous phase was then transferred to a new tube with 500 µl of isopropanol, vigorously homogenised, and stored at -20 ˚C for 40 min. Then, the mixture was centrifuged at 13,000 rpm for 20 min at 2 ˚C. The supernatant was discarded, and the final pellet was washed with 1 ml of 70% (v/v) ethanol at 13,000 rpm for 3 minutes. This step was repeated twice, and at the end, the pellet was air-dried for 4 min. The RNA was re-suspended with 50 µl of milli-Q water and stored at -80 ˚C. Finally, the integrity (28S/18S) of the isolated RNA was examined by 1.0% agarose gel, and RNA purity (ratio of A260/A280) was evaluated with a nanophotometer for the CTAB method, Trizol and Direct-zol RNA kit.

RNA purification and inhibitor removal

To remove any remaining inhibitors (polyphenols, humic/fulvic acids, tannins, melanin, etc.) that impair the construction of total RNA sequencing libraries, the PCR inhibitor removal kit protocol was used. Finally, the quality and quantity of RNA was analysed, and the samples were stored at -80 ˚C.

RNA integrity number (RIN) analysis and next-generation sequencing (NGS) library quality control (QC)

Six samples of *G. crinita* and *C. spruceanum,* respectively, were tested. The quality and concentration of all extracted RNA samples were assessed with a bioanalyser RNA Pico 6000 chip (Agilent), and the library was prepared using TruSeq Stranded mRNA Library at MACROGEN Inc., Seoul, Korea.

RESULTS AND DISCUSSION

Isolating RNA from lignocellulosic tissues is challenging, as they contain large amounts of polyphenols, polysaccharides and specialised metabolites that precipitate together with

nucleic acids, negatively affecting both yield and purity (Nizam et al. 2023). For this reason, the objective of the study was to develop a robust and reliable method to extract RNA from *G. crinita* and *C. spruceanum* woody tissues for subsequent application in tree genomics (RNA sequencing, gene expression, etc.) (Norwati et al. 2011, Wu et al. 2017).

However, a good quantity and quality of RNA is essential to perform these experiments. Therefore, modifications to the protocols commonly used in trees are necessary to obtain intact RNA without interference from metabolites, as the quality and quantity of RNA vary among different types of tissues and trees (Ma et al. 2015, Chauhan et al. 2018, Yan et al. 2022).

Thus, four protocols were tested (Trizol, Direct-zol RNA kit, CTAB and a combination between CTAB and Direct-zol RNA) in *G. crinita* and *C. spruceanum* (Figure 2).

The results showed that the most widely used kit, the Direct-zol RNA Kit and the TRIzol™ reagent, were not successful in obtaining high-quality RNA due to the presence of specialised metabolites (Figure 2). Nizam et al. (2023) reported similar results for the species *Kandelia candel*, using the RNeasy Plant Mini kit and the TRIzol™ reagent.

Table 1 shows that it was possible to obtain RNA from forest trees with the CTAB method and the combination between CTAB and Direct-zol RNA. The CTAB method is advantageous due to the inclusion of β-mercaptoethanol, a potent reducing agent that prevents the oxidation of phenolic compounds. Consequently, the co-precipitation of oxidised products with RNA during isolation is minimised. Additionally, this method can break disulfide bonds and inactivate ribonucleases (Kanani et al. 2019).

Likewise, the quality of isolated total RNA described in this study was confirmed in several ways. Firstly, Table 1 shows that RNA concentration in *G. crinita* ranged between 2432.7 ± 334.7 ng/µL (CTAB) and 2048.2 ± 281.8 ng/µL (CTAB with Direct-zol RNA kit), and in *C. spruceanum* it ranged between 2248.6 ± 291.9 ng/µL (CTAB) and 1864.2 ± 242.0 ng/µL (CTAB with Direct-zol RNA kit). Moreover, with these two methods, the 260/280 and 260/230 ratios were greater than 1.90, indicating high-quality RNA (Table 1). On the other hand, the RNA concentration was below 40 ng/µl with Trizol and Direct-zol RNA kit (Table 1, Figure 2C, 2D, 2G, 2H) and the 260/280 and 260/230 ratios were below 1.5 and 0.5 respectively, indicating contamination by protein, polysaccharides, salt ion and organic solvents (Jensen et al. 2023, Yan et al. 2022).

Secondly, 1% agarose gels were run to visualise RNA quality. Low-quality RNA was obtained with the Trizol and Direct-zol RNA kit methods (Figure 2). The finding is in agreement with previous studies, since tissue rich in phenolic compounds are oxidised to form quinones that bind RNA, making it difficult to isolate (Ghawana et al. 2011, Ouyang et al. 2014). A good separation of the 28S and 18S ribosomal RNA bands with CTAB method in *G. crinita* samples was not observed, and this pattern was similar in all samples (Figure 2E), although the 260/280 ratio was of 2.08 (Table 1). *Guazuma crinita* presented several exudates, making it challenging to isolate RNA. Thus, an initial pretreatment of CTAB with chloroform/

Table 1 Ribonucleic acid (RNA) concentration and quality from *Guazuma crinita* and *Calycophyllum spruceanum*

Species	Method	Concentration (ng/µl)	A260/A280	A260/A230
Guazuma crinita	Trizol	26.7 ± 4.7	1.43 ± 0.03	0.56 ± 0.03
	Direct-zol RNA kit	35.3 ± 6.3	1.43 ± 0.03	0.58 ± 0.03
	CTAB	2432.7 ± 334.7	2.08 ± 0.03	2.25 ± 0.09
	CTAB & Direct-zol RNA kit	2048.2 ± 281.8	2.1 ± 0.03	2.44 ± 0.10
Calycophyllum spruceanum	Trizol	27.1 ± 24.1	1.36 ± 0.08	0.37 ± 0.03
	Direct-zol RNA kit	35.4 ± 35.9	1.22 ± 0.08	0.54 ± 0.05
	CTAB	2248.6 ± 291.9	2.03 ± 0.10	1.95 ± 0.15
	CTAB & Direct-zol RNA kit	1864.2 ± 242.0	2.11 ± 0.10	2.34 ± 0.17

CTAB = cetyltrimethylammonium bromide, RNA = ribonucleic acid

©*Forest Research Institute Malaysia*

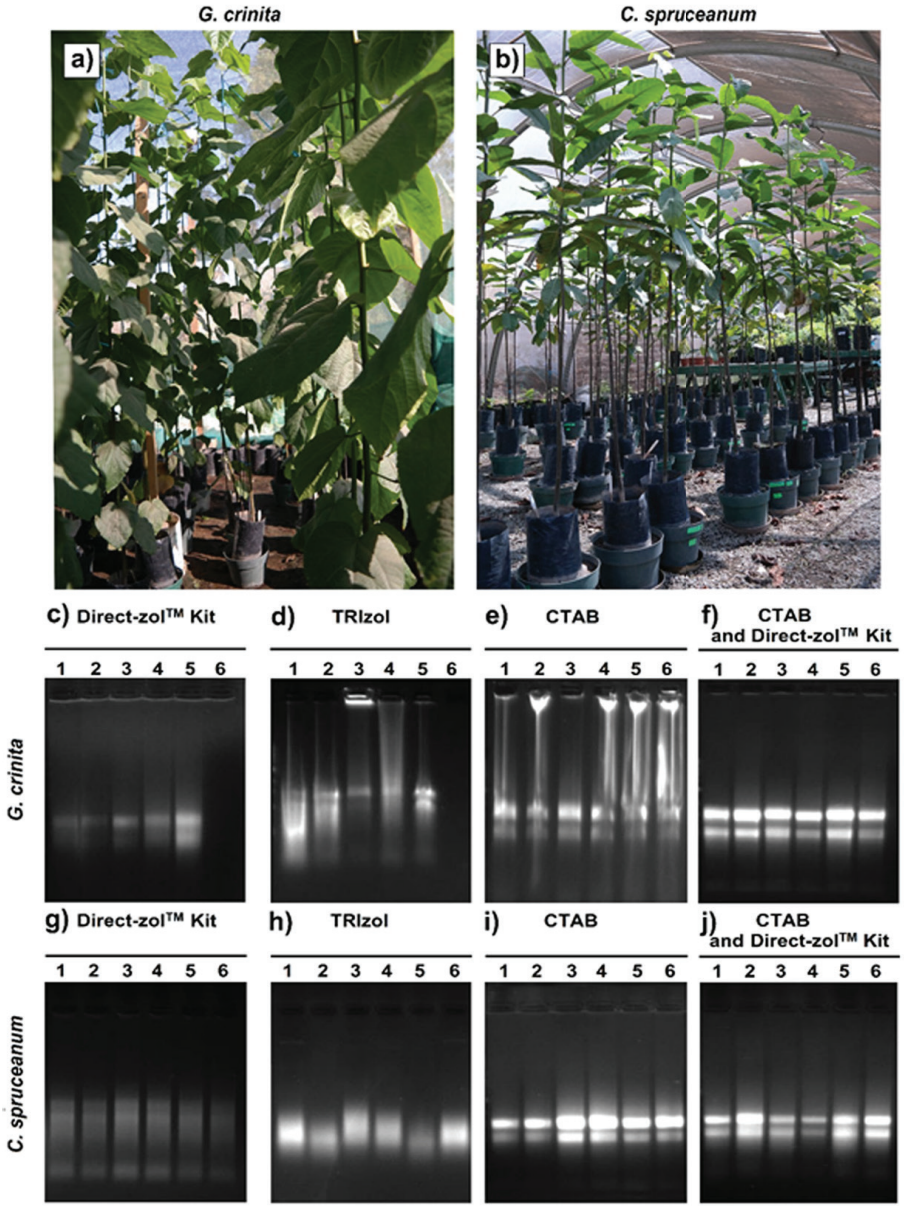


Figure 2 RNA quality in agarose gels from *Guazuma crinita* and *Calycophyllum spruceanum*

isoamyl alcohol (24:1, v/v) was added to improve RNA isolation. Likewise, RNA precipitation with LiCl for 12 hours at -20 °C was performed to help isolate the RNA from contamination, avoiding the oxidation of polyphenols in alkaline conditions. After RNA precipitation, the pellet was dissolved in TRI-reagent, which significantly eliminates contamination and prevents the production of insoluble granules (Ramadoss & Basu 2018). However, it was necessary to improve the RNA isolation from *G. crinita*. Thus, the CTAB method was used in tandem with Direct-zol RNA kit, having RNA of good quality. The 28S and 18S ribosomal RNA bands were well-defined since the kit provided a method of RNA purification by removing polysaccharides and secondary metabolites (Figure 2F). On the other hand, RNA isolation in *C. spruceanum* presented good quality from the secondary xylem tissues with CTAB and CTAB plus Direct-zol RNA kit (Figure 2I, J). However, there are drawbacks to isolating RNA from *C. spruceanum* bark because of the high levels of polyphenols that are easily oxidised to form covalent compounds with RNA, leading to oxidation and degradation (Sharma et al. 2003, Perin et al. 2020).

RNA integrity number (RIN) analysis and next-generation sequencing (NGS) library quality control (QC)

Last but perhaps most importantly, an electropherogram of the total RNA isolated from the secondary xylem of *G. crinita* and *C. spruceanum* was generated. The bioanalyser system was used to validate total RNA high quality, which uses the RIN to assign a numerical value to the RNA quality, including 18S to 28S ribosomal peak ratio, the separation between these peaks, and the presence or absence of degradation (Ma et al. 2015, Agilent Technologies 2016, Jensen et al. 2023). The RIN has values ranging from 10 to 1, where 10 indicates the highest RNA quality and 1 indicates degraded RNA (Agilent Technologies 2016). Thus, the RNA isolated from all the samples of *G. crinita* and *C. spruceanum* showed RIN values greater than 7.9 (Table 2), showing no degradation, and the RNA was of good quality. Generally, an RNA value greater than 7.0 is necessary to yield satisfactory results in next-generation sequencing analysis. The RIN values obtained are significantly higher than those reported in previous studies (Qiu et al. 2013, Ma et al. 2015, Nizam et al. 2023).

The purity of the extracted RNA may depend on the number of secondary metabolites produced by the plants (Sharma et al. 2003). The methods consistently produced RNA with good yield and high quality in *C. spruceanum* and *G. crinita* (Table 2). Therefore, 3 µL were used to test NGS library QC. Only the two-year-old plants of *G. crinita* failed to pass, possibly due to inhibitors affecting Truseq stranded mRNA library (Lamble et al. 2013). A possible explanation is that contaminants could have a more significant effect than the amount of RNA used in the reaction, and only a small proportion of RNA would be accessible to transposase, upsetting the ideal ratio of RNA and enzyme. Thus, to solve this problem in this downstream processing step, total RNA from the two-year-old plants of *G. crinita* was passed through the PCR inhibitor removal kit, allowing the library construction for RNA sequencing (Figure 3). It is worth noting that this step was also employed by Jensen et al. (2023) for extracting RNA from red, green and brown algae, as well as from the cyanobacterium *Arthrospira platensis* and the seagrass *Zostera marina*. These reports reinforce the notion that the modified CTAB method could be applicable to diverse autotrophic organisms (Mitchell et al. 2023).

Finally, total RNA sequencing was performed with a sequencer to obtain paired-end reads. Thus, on average per sample 4,559,130 total kilo base pairs (kbp) in 45,139,903 reads with a Q20 of 98.49% and Q30 of 95.59% were obtained for *G. crinita* and 5,510,902 total kbp in 54,563,395 reads with a Q20 of 98.23% and Q30 of 94.94% were obtained for *C. spruceanum* (Supplementary material S1).

CONCLUSIONS

To have a robust protocol to isolate high-quality RNA from Amazonian trees is truly important, as it will help to understand the molecular mechanisms associated with their development. In this study, a high-throughput method was developed for isolating RNA from the woody

Table 2 Assessment of the completeness of rRNA from *Guazuma crinita* and *Calycophyllum spruceanum*

Species	Method	Concentration (ng/µl)	A260/A280	A260/A230
Guazuma crinita	Trizol	26.7 ± 4.7	1.43 ± 0.03	0.56 ± 0.03
	Direct-zol RNA kit	35.3 ± 6.3	1.43 ± 0.03	0.58 ± 0.03
	CTAB	2432.7 ± 334.7	2.08 ± 0.03	2.25 ± 0.09
	CTAB & Direct-zol RNA kit	2048.2 ± 281.8	2.1 ± 0.03	2.44 ± 0.10
Calycophyllum spruceanum	Trizol	27.1 ± 24.1	1.36 ± 0.08	0.37 ± 0.03
	Direct-zol RNA kit	35.4 ± 35.9	1.22 ± 0.08	0.54 ± 0.05
	CTAB	2248.6 ± 291.9	2.03 ± 0.10	1.95 ± 0.15
	CTAB & Direct-zol RNA kit	1864.2 ± 242.0	2.11 ± 0.10	2.34 ± 0.17

RIN = RNA integrity number, rRNA = ribosomal ribonucleic acid

©*Forest Research Institute Malaysia*

tissues of *G. crinita* and *C. spruceanum*, which can be utilised for RNA sequencing. Compared to the other three methods, the CTAB combined with TRI-reagent and Direct-zolTM RNA kit allowed obtaining good-quality RNA. Furthermore, agarose gel electrophoresis exhibited distinct and intact bands for this method, and the RIN values ranged from 7.9 to 8.3. The results showed that the modified method efficiently obtained high-quality RNA from the secondary xylem, overcoming the problems with inhibitors and high concentrations of exudates, alkaloids, polysaccharides and phenols. It is suitable for downstream experiments such as RNA-seq, qRT-PCR analysis and gene cloning.

ACKNOWLEDGMENTS

The authors would like to thank Jimenez-Davalos J for his critical reading. Funding was obtained by Universidad Nacional Agraria la Molina through PROCIENCIA (grant numbers 177-2015-FONDECYT and 159-2018-FONDECYT-BM-IADT-AV).

REFERENCES

Abanto-Rodriguez C, García-Soria D, Guerra-Árevalo W, Murga-Orrillo H, Saldaña-Ríos G, Vázquez-Reátegui D & Tadashi-Sakazak R. 2016. Organic substrates in *Calycophyllum spruceanum* (Benth.) plants production. *Scientia Agropecuaria* 7: 341–347. https://doi.org/10.17268/sci.agropecu.2016.03.23.

Agilent Technologies. 2016. Comparison of RIN and RINe algorithms for the Agilent 2100 Bioanalyzer and the Agilent 2200 TapeStation systems. Technical Overview 1–4. Agilent, Santa Clara CA.

Ambavaram M, Basu S, Krishnan A, Ramegowda V, Batlang U, Rahman L & Pereira A. 2014. Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nature Communications* 5: 5302 https://doi.org/10.1038/ncomms6302.

Cardoso D, Särkinen T, Alexander S, Amorim A, Bittrich V, Celis M, Forzza R. 2017. Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences of the United States of America* 114: 10695–10700. https://doi.org/10.1073/pnas.1706756114.

Chauhan A, Sharma J, Modgil M & Siddappa S. 2018. Comparison of various RNA extraction methods, cDNA preparation and isolation of calmodulin gene from a highly melanized isolate of apple leaf blotch fungus Marssonina coronaria. *Journal of Microbiological Methods* 151: 7–15. https://doi.org/10.1016/j.mimet.2018.05.023.

Chomczynski P. 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *BioTechniques* 15: 532–537.

De Heredia LU & Vázquez-Poletti J. 2016. RNA-seq analysis in forest tree species: bioinformatic problems and solutions. *Tree Genetics and Genomes* 12: 1–17. https://doi.org/10.1007/s11295-016-0995-x.

De Lima R, Görgens E, Batista A, Da Silva D, De Oliveira C & De Sousa C. 2022. Diversity and big trees patterns in the Brazilian Amazon. *Diversity* 14: 1–10. https://doi.org/10.3390/d14070503.

Degen B & Sebbenn A. 2014. Genetics and tropical forests. Pp 1–30 in Laslo-Pancel and Michael-Köhl (eds) *Tropical Forestry Handbook*. Springer Verlag, Berlin. https://doi.org/10.1007/978-3-642-41554-8.

Galeano E, Vasconcelos T, Ramiro D, De Martin V & Carrer H. 2014. Identification and validation of quantitative real-time reverse transcription PCR reference genes for gene expression analysis in teak (Tectona grandis L.f.). *BMC Research Notes* 7: 464. https://doi.org/10.1186/1756-0500-7-464.

Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj P Kumar S. 2011. An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Research Notes* 4: 85. https://doi.org/10.1186/1756-0500-4-85.

Hubbell S, He F, Condit R, Borda-De-Água L, Kellner J & Ter Steege H. 2008. How many tree species are there in the Amazon and how many of them will go extinct? *Proceedings of the National Academy of Sciences of the United States of America* 105: 11498–11504. https://doi.org/10.1073/pnas.0801915105.

Jensen T, Saleh L, Bents D, Krohn S, Wu Y, Mucke M & Labes A. 2023. Optimised protocols for RNA extraction from a broad taxonomic range of algae. *Journal of Applied Phycology* 35: 1743–1753. https://doi.org/10.1007/s10811-023-02980-7.

Kanani P, Shukla Y, Modi A, Subhash N & Kumar S. 2019. Standardization of an efficient protocol for isolation of RNA from Cuminum cyminum. *Journal of King Saud University - Science* 31: 1202–1207. https://doi.org/10.1016/j.jksus.2018.12.008.

Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G & Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnology* 13. https://doi.org/10.1186/1472-6750-13-104.

Ma Z, Huang B, Xu S, Chen Y, Li S & Lin S. 2015. Isolation of high-quality total RNA from Chinese fir (*Cunninghamia lanceolata* (Lamb.) hook). *PLoS ONE* 10: 1–10. https://doi.org/10.1371/journal.pone.0130234.

Mitchell N, Mcassey E & Hodel R. 2023. Emerging methods in botanical DNA/RNA extraction. *Applications in Plant Sciences* 11: 1–4. https://doi.org/10.1002/aps3.11530.

Nizam A, Kalath H & Kumar A. 2023. A modified method for efficient RNA isolation from mangrove root tissues rich in secondary metabolites. *BioTechniques* 74: 303–317. https://doi.org/10.2144/btn-2022-0078.

Norwati, Abdullah, Norlia, Mohd, Norwati & Suryani. 2011. Expression of BT gene in transgenic teak. *Journal of Tropical Forest Science* 23: 488–491.

O'neill G, Dawson I, Sotelo-Montes C, Guarino L, Guariguata M, Current D & Weber J. 2001. Strategies for genetic conservation of trees in the Peruvian Amazon. *Biodiversity and Conservation* 10: 837–850. https://doi.org/10.1023/A:1016644706237.

Odonne G, Valadeau C, Alban-Castillo J, Stien D, Sauvain M & Bourdy G. 2013. Medical ethnobotany of the Chayahuita of the Paranapura basin (Peruvian Amazon). *Journal of Ethnopharmacology* 146: 127–153. https://doi.org/10.1016/j.jep.2012.12.014.

Ouyang K, Li J, Huang H, Que Q, Li P & Chen X. 2014. A simple method for RNA isolation from various tissues of the tree *Neolamarckia cadamba*. *Biotechnology and Biotechnological Equipment* 28: 1008–1013. https://doi.org/10.1080/13102818.2014.981086.

Paredes O, Soudre-Zambrano M, Chavez J & Guerra W. 2010. Propagación vegetativa de Bolaina Blanca (*Guazuma crinita* Mart.) mediante injerto, bajo condiciones ambientales controladas. *Folia Amazónica* 19: 69. https://doi.org/10.24841/fa.v19i1-2.346.

Perin E, Fontoura B, Lima V & Carpes S. 2020. RGB pattern of images allows rapid and efficient prediction of antioxidant potential in *Calycophyllum spruceanum* barks. *Arabian Journal of Chemistry* 13: 7104–7114. https://doi.org/10.1016/j.arabjc.2020.07.015.

Qiu Z, Wan L, Chen T, Wan Y, He X, Lu S & Lin J. 2013. The regulation of cambial activity in Chinese fir (*Cunninghamia lanceolata*) involves extensive transcriptome remodeling. *New Phytologist* 199: 708–719. https://doi.org/10.1111/nph.12301.

Ramadoss N & Basu C. 2018. Extraction of RNA from recalcitrant tree species *Paulownia elongata*. *Bio-Protocol* 8: 1–8. https://doi.org/10.21769/bioprotoc.2925.

Saldaña C, Rodriguez-Grados P, Chávez-Galarza J, Feijoo S, Guerrero-Abad J, Vásquez H & Arbizu C. 2022. Unlocking the complete chloroplast genome of a native tree species from the Amazon Basin, Capirona (*Calycophyllum spruceanum*, Rubiaceae), and its comparative analysis with other Ixoroideae species. *Genes* 13: 1–113. https://doi.org/10.3390/genes13010113.

Santos A, Ribeiro-Oliveira J & Carvalho C. 2016. Sobre a botânica, a etnofarmacologia e a química de *Calycophyllum spruceanum* (Benth.) Hook. f. ex K. Schum. *Revista Brasileira de Plantas Medicinais* 18: 383–389. https://doi.org/10.1590/1983-084X/15_152.

Sharma A, Gill P & Singh P. 2003. RNA isolation from plant tissues rich in polysaccharides. *Analytical Biochemistry* 314: 319–321. https://doi.org/10.1016/S0003-2697(02)00689-9.

Sundell D, Street N, Kumar M, Mellerowicz E, Kucukoglu M, Johnsson C & Hvidsten T. 2017. Aspwood: high-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in populus tremula. *Plant Cell* 29: 1585–1604. https://doi.org/10.1105/tpc.17.00153.

Ter Steege H, Vaessen R, Cárdenas-López D, Sabatier D, Antonelli A, De Oliveira S & Salomão R. 2016. The discovery of the Amazonian tree flora with an updated checklist of all known tree taxa. *Scientific Reports* 6: 1–15. https://doi.org/10.1038/srep29549.

Vennapusa A, Somayanda I, Doherty C & Jagadish S. 2020. A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fibre. *Scientific Reports* 10: 1–13. https://doi.org/10.1038/s41598-020-73958-5.

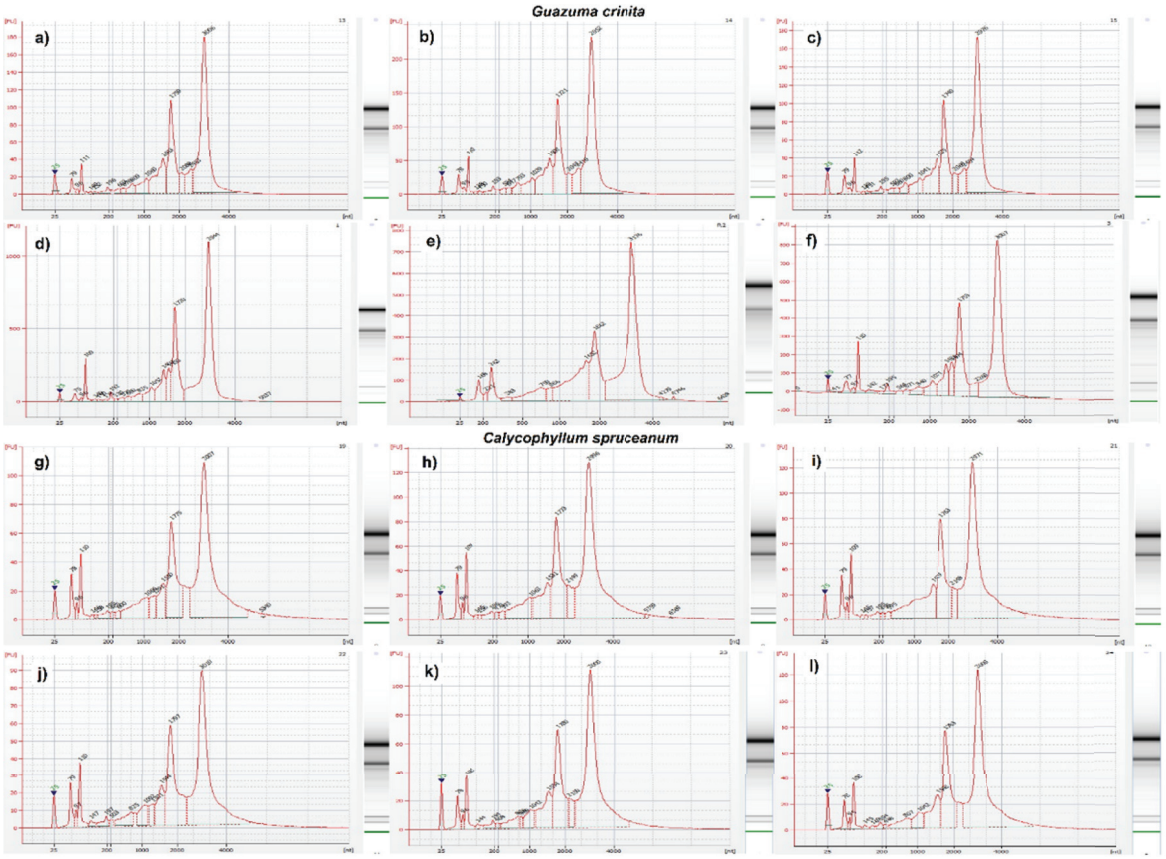
Wu Z, Liu S, Li J, Li M, Du H, Qi L & Lin L. 2017. Analysis of gene expression and quality of agarwood using agar-bit in aquilaria sinensis. *Journal of Tropical Forest Science* 29: 380–388. https://doi.org/10.26525/jtfs2017.29.3.380388.

Yan W, Pendi F & Hussain H. 2022. Improved CTAB method for RNA extraction of thick waxy leaf tissues from sago palm (*Metroxylon sagu* Rottb.). *Chemical and Biological Technologies in Agriculture* 9: 1–14. https://doi.org/10.1186/s40538-022-00329-9.

Appendix 1 S1 Distribution of A, T, G, C base in RNA sequencing

Species	Sample ID	Total read bases (kbp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
Guazuma crinita	BJ1	5,645,955	55,900,552	46.43	53.57	98.54	95.83
	BJ2	5,348,246	52,952,938	46.65	53.35	98.51	95.8
	BJ3	4,189,320	41,478,416	46.47	53.53	98.45	95.69
	BA1	4,072,862	40,325,372	45.4	54.6	98.46	95.36
	BA2	4,048,312	40,082,306	45.3	54.7	98.52	95.48
	BA3	4,050,083	40,099,832	45.11	54.89	98.45	95.35
Calycophyllum spruceanum	CJ1	5,043,953	49,940,132	45.58	54.42	98.42	95.49
	CJ2	5,542,772	54,878,936	45.23	54.77	98.53	95.76
	CJ3	4,968,155	49,189,656	46.07	53.93	98.39	95.44
	CA1	6,468,740	64,046,940	45.55	54.45	97.89	94.02
	CA2	6,533,295	64,686,092	45.67	54.33	97.96	94.17
	CA3	4,508,499	44,638,612	45.22	54.78	98.19	94.76

Appendix 2 S2 RNA integriti number (RIN) analysis from Guazuma crinita and Calycophyllum spruceanum



Appendix 3 S3 Next generation sequencing (NGS) library QC from *Guazuma crinita* and *Calycophyllum spruceanum*