Yan *et al. Chem. Biol. Technol. Agric.* (2022) 9:63
https://doi.org/10.1186/s40538-022-00329-9

Chemical and Biological Technologies in Agriculture

METHODOLOGY

Open Access

Improved CTAB method for RNA extraction of thick waxy leaf tissues from sago palm (*Metroxylon sagu* Rottb.)

Wei-Jie Yan, Fifi Hafizzah Pendi and Hasnain Hussain*

Abstract

Background: There is a growing interest in transcriptomics studies parallel to the advancement of transcriptome databases and bioinformatics, which provided the opportunity to study responses to growths, stimuli and stresses. There is an increase in demand for excellent RNA extraction techniques. General RNA extraction protocols can be used in RNA extraction, but the quality and quantity vary in different types of tissues from different organisms. Hence, a specific RNA extraction method for each organism's tissue type is required to obtain the desired RNA quality and quantity.

Results: The improved CTAB RNA extraction method is superior to the PCI method and MRIP method for thick waxy leaves that were applied for mature sago palm (*Metroxylon sagu* Rottb.) leaf tissue and produce total RNA extract with good purity (OD 260/280 ≥ 1.8, OD 260/230 ≥ 2.0) and integrity (RIN ~ 7). RNA sequencing was conducted with the extracted samples and showed good assembly results (Q20 ≥ 97, Q30 ≥ 91%, assembly mean length ≥ 700 bp).

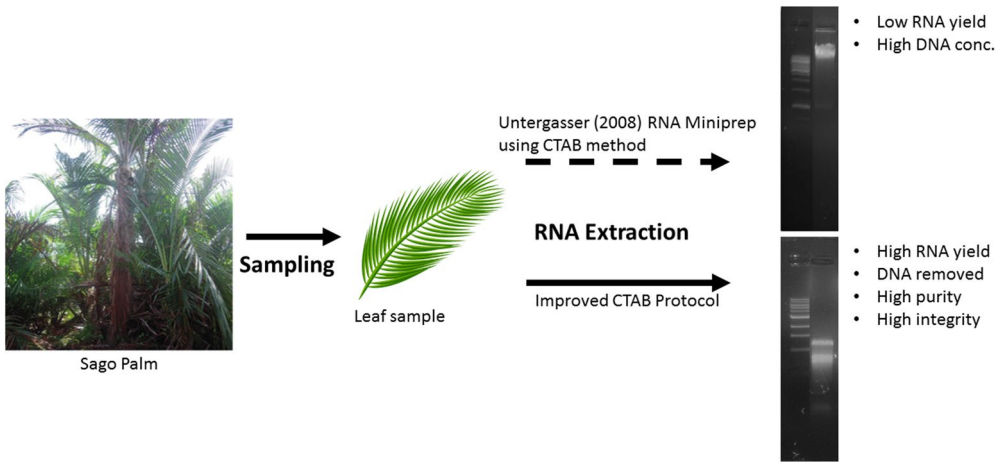
Conclusion: The improved CTAB RNA extraction method enables rapid, cost-effective, and relatively simple RNA extraction from waxy, fibrous and high-in-polyphenol sago palm (*M. sagu* Rottb.) leaf tissue with next-generation RNA sequencing recommended quality.

Keywords: *Metroxylon sagu*, RNA extraction, Transcriptomics, Improved CTAB method

*Correspondence: hhasnain@unimas.my

Centre for Sago Research, Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

Graphical Abstract



Introduction

Sago palm (*Metroxylon sagu*) is one of the starch-producing crops and is the best starch-producing species within the *Metroxylon* spp. which could be found from southern Thailand to the Solomon Islands [1, 2]. Among the *Metroxylon* spp. only *M. sagu* is both hapaxanthic (once-flowering) and soboliferous (produce suckers) [3]. *M. sagu* lifecycle is divided into four stages which is the rosette stage, bole forming stage, inflorescence stage and fruit ripening stage where the starch accumulation in the trunk starts at the bole formation stages and reaches its peak at the beginning of the inflorescence stage which is approximately nine years after plantation [3]. The advantages of sago palm are its abilitiy to adapt to acidic, saline soil with seasonal flooding such as peat swamps which other crops are unable to grow [4]. There is growing interest in this palm as one of the candidates to combat the world food shortage.

Advances in nucleic acid sequencing and bioinformatics technologies have revolutionized our ability to study plant growth biology and its response to stimuli and stresses [5–8]. Several methods (and their adoptees) such as Rochester et al. [9, 10], Chomczynski and Sacchi [11, 12], Chang et al. [13, 14], Schultz et al. [15–17], Salzman et al. [10, 18], Kiefer et al. [19, 20], Hu et al. [21, 22], Hussain [23–29], modified Zeng and Yang [10, 30–32], Wu et al. [10, 33], modified Gasic et al. [20, 22, 34–40], Xiao et al. [41] and several kits such as RNAqueous™ Total RNA Isolation Kit [12, 22], FavorPrep™ Plant Total RNA Mini extraction kit [22], and Qiagen RNeasy Plant Mini Kit [10] were used in sago palm leaf RNA extraction, but the actual quality and quantity were not well defined. Notably, modified Zeng and Yang [30] and modified Gasic et al. [34] methods which utilized CTAB as its extraction buffer are better in young sago palm leaf RNA extraction.

The major differences between the improved CTAB method based on Untergasser [42] with modified Zeng and Yang [30] and modified Gasic et al. [34] are the extraction buffer is not pre-heated to 55 °C–65 °C and the precipitation using ice-cooled isopropanol for 10 min instead of overnight precipitation with lithium chloride. This change minimized the damage to the RNA due to high temperature and the extraction complete

Table 1 General differences between improved CTAB method, PCI method and MRIP method

Features	RNA extraction methods		
	Improved CTAB	PCI	MRIP
Surfactant/chaotropic agent	Cetyltrimethyl ammonium bromide	Sodium dodecyl sulfate	Guanidinium thiocyanate
Phenol content	No	Yes	Yes
Final RNA precipitation	Ice-cooled isopropanol	Absolute ethanol + incubate in -80 °C freezer	Isopropanol + incubate on ice
Estimated extraction time inclusive of DNase treatment	4 h	27 h	2.5 h

Table 2 General differences of improved CTAB method with Untergasser [42]

Features	Improved CTAB method	Untergasser [42] method
RNA pellet re-solubilization	Gentle mixing	65 °C in water for 15 min
DNA removal step	DNase treatment	Did not mention which specifically
RNA precipitation	Ice-cooled isopropanol	Lithium chloride
Estimated extraction time inclusive of DNase treatment	4 h	5 h

more rapidly. The effect of other extraction buffers on mature sago palm leaf; which are waxier, more secondary metabolites, and less viable cells; RNA extraction are still unknown, hence a comparative study of RNA extraction on mature sago palm leaf with different extraction buffers and protocols will demonstrate the potential of the CTAB extraction method.

In this paper, we describe a validated method for the extraction of RNA for sago palm (*Metroxylon sagu* Rottb) leaf tissue for RNA sequencing (RNA-seq) purpose modified from Untergasser [42] RNA Miniprep using the CTAB method [42]. These procedures demonstrated the successful extraction of RNA from sago palm leaf tissue for RNA sequencing. The procedures described here are rapid, cost-effective and relatively simple.

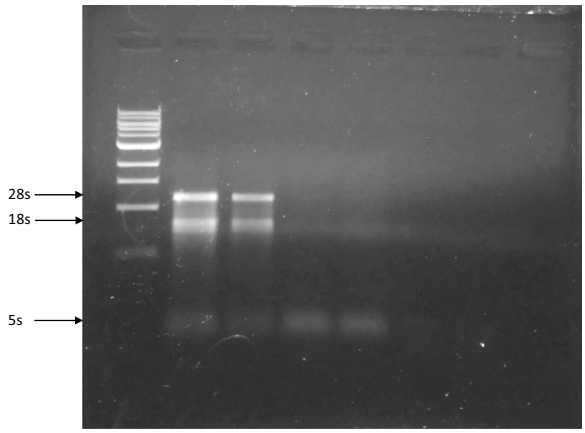


Fig. 1 TBE-AGE assay of sago palm leaf tissue RNA extract between improved CTAB method, PCI method and MRIP methods. Line M = GeneRuler 1 kb DNA Ladder, lane 1 = improved CTAB method replicate 1, lane 2 = improved CTAB method replicate 2, lane 3 = PCI method replicate 1, lane 4 = PCI method replicate 2, lane 5 = MRIP method replicate 1, lane 6 = MRIP method replicate 2. Improved CTAB method shows the best results as shown by the present of intact 28 s and 18 s rRNA band. There is presence of smearing at the 5 s region for CTAB method and PCI method but more significant in the latter method. No band is observed in MRIP method

Materials and methods

Sampling

Sago palm leaf samples were wiped with 70% ethanol to remove debris, stored in labeled containers and snap-frozen in liquid nitrogen on-site to preserve the RNA. The samples need to be kept in liquid nitrogen before long-term storing in a − 80 °C freezer. Selection of the leaf age influences the total RNA yield and purity, in which young leaf gives better RNA yield and purity. In this study, mature sago palm leaf was selected in compliance with the motif of the main study.

Total RNA extraction

Three different extraction methods were compared in this study, i.e., improved CTAB method, PCI method [23] and MRIP method [41], to determine the most suitable method for extraction of total RNA from *Metroxylon sagu* leaf tissue for RNA sequencing (Table 1). The improved CTAB method is a modified method based on Untergasser [42] which focus on the simplicity, rapidity and reproducibility of RNA extraction from mature sago palm leaf

Table 3 Sago palm leaf tissue RNA extract quality and quantity assay between improved CTAB, PCI and MRIP methods by Eppendorf Biophotometer Plus

Method	Replicate	A230	A260	A280	A340	260/280	260/230	RNA weight (µg)
Improved CTAB	**1**	0.316	0.637	0.33	0.001	1.93	2.01	255
	2	0.131	0.25	0.13	0	1.93	1.9	100
PCI	**1**	0.299	0.27	0.161	0.008	1.67	0.9	10.78
	2	0.354	0.289	0.181	0.017	1.6	0.82	11.56
MRIP	**1**	0.941	0.189	0.16	0.044	1.18	0.2	7.56
	2	1.073	0.209	0.18	0.054	1.16	0.2	8.38

Improved CTAB method gives the highest purity and RNA yield in mature sago palm leaf RNA extraction

Table 4 Normality test for 260/280 ratio, 260/230 ratio and RNA weight data. Source: Reprint from IBM SPSS Statistics 28.0.1.1 (15) report

Tests of normality						
	Kolmogorov–Smirnov[a]			**Shapiro–Wilk**		
	Statistic	**df**	**Sig**	**Statistic**	**df**	**Sig**
260/280	0.210	6	0.200*	0.859	6	0.185
260/230	0.219	6	0.200*	0.862	6	0.195
Weight	0.373	6	0.009	0.689	6	0.005

The *p*-values of 260/280 ratio and 260/230 ratio data are higher than α (0.05) value hence they are normally distributed while RNA weight data *p*-value is lower than α (0.05) value hence not normally distributed with 95% confidence

*This is a lower bound of the true significance

[a] Lilliefors significance correction

Table 5 Tukey multiple comparison test of 260/280 ratio and 260/230 ratio data between RNA extraction method. Source: Reprint from IBM SPSS Statistics 28.0.1.1 (15) report

Multiple comparisons							
Tukey HSD							
Dependent variable	**(I) Method**	**(J) Method**	**Mean difference (I-J)**	**Std. Error**	**Sig**	**95% Confidence interval**	
						Lower bound	**Upper bound**
260/280	CTAB	MRIP	0.7600*	0.02972	< .001	0.6358	0.8842
		PCI	0.2950*	0.02972	0.004	0.1708	0.4192
	MRIP	CTAB	− 0.7600*	0.02972	< .001	− 0.8842	− 0.6358
		PCI	− 0.4650*	0.02972	.001	− 0.5892	− 0.3408
	PCI	CTAB	− 0.2950*	0.02972	.004	− 0.4192	− 0.1708
		MRIP	0.4650*	0.02972	.001	0.3408	0.5892
260/230	CTAB	MRIP	1.7550*	0.05553	< .001	1.5230	1.9870
		PCI	1.0950*	0.05553	< .001	0.8630	1.3270
	MRIP	CTAB	− 1.7550*	0.05553	< .001	− 1.9870	− 1.5230
		PCI	− 0.6600*	0.05553	.003	− 0.8920	0.4280
	PCI	CTAB	− 1.0950*	0.05553	< .001	− 1.3270	0.8630
		MRIP	0.6600*	0.05553	.003	0.4280	0.8920

The *P*-values between all the RNA extraction methods are lower than the α (0.05) value hence the 260/280 ratio and 260/230 ratio data between all the RNA extraction methods are significantly with 95% confidence

Based on observed means

The error term is mean square (error) = 0.003

*The mean difference is significant at the 0.05 level

Table 6 Independent-samples Kruskal–Wallis multiple comparison test of RNA weight data between RNA extraction methods. Source: Reprint from IBM SPSS Statistics 28.0.1.1 (15) report

Pairwise comparisons of method					
Sample 1–Sample 2	**Test statistic**	**Std. error**	**Std. test statistic**	**Sig**	**Adj. Sig[a]**
MRIP–PCI	− 2.000	1.871	− 1.069	0.285	0.855
MRIP–CTAB	4.000	1.871	2.138	0.033	0.098
PCI–CTAB	2.000	1.871	1.069	0.285	0.855

There is no significant difference of the extracted RNA weight between the RNA extraction methods as shown by all the adjusted P-value higher than the α (0.05) value with 95% confidence

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same

Asymptotic significances (2-sided tests) are displayed. The significance level is 0.050

[a] Significance values have been adjusted by the Bonferroni correction for multiple tests

Table 7 Mature sago palm leaf tissue (T1, T2, T3, N1, N2, and N3 are biological replicates) RNA extract quality and quantity for RNA sequencing by nanodrop and Agilent 2100 bioanalyzer. Source: Adapted from BGI Pre-sequencing Report

Sample	OD 260/280	OD 260/230	RIN	28 s/18 s	Total mass(μg)
T1	1.91	2.61	6.6	1.4	13.5
T2	1.88	2.23	7.4	2.0	18.7
T3	1.96	2.22	7.5	1.7	37.53
N1	1.99	2.62	7.7	2.0	38.34
N2	1.99	2.38	7.4	1.8	31.45
N3	2.01	2.45	7.5	1.9	35.87

The purity, total RNA amount, RIN and the 28 s/18 s number for all the samples are sufficient for RNA sequencing

sample. PCI method is an RNA extraction method practice on sago palm young leaf sample while MRIP method is a published RNA extraction method for young coconut palm leaf which shows promising results. One gram of mature sago palm leaf samples was grounded finely in prechilled mortar and pestle with liquid nitrogen for all three methods. The ribonucleic acid pellet obtained, which contains both DNA and RNA, from the three protocols were further subjected to DNase treatment. The pellet was resuspended with 0.5 ml of Tris–HCl buffer. Next DNase treatment was conducted with RQ1 RNase-Free DNase (Promega) according to protocol. Then, the mixture was centrifuged at 16,100 *g*, 4 °C for 15 min. After that, the mixture was transferred into a new 1.5-ml microcentrifuge tube. Then, an equal volume of ice-cooled isopropanol was added, mixed and incubated at room temperature for 10 min. After that, the mixture was centrifuged at 16,100 *g*, 4 °C for 15 min. Then, the supernatant was discarded and the pellet was washed with 1 ml of 70% ethanol twice. Lastly, the pellet was resuspended with 100 μl Tris–EDTA buffer (10 mM Tris–HCl, 1 mM disodium EDTA, pH 8.0) for further analysis.

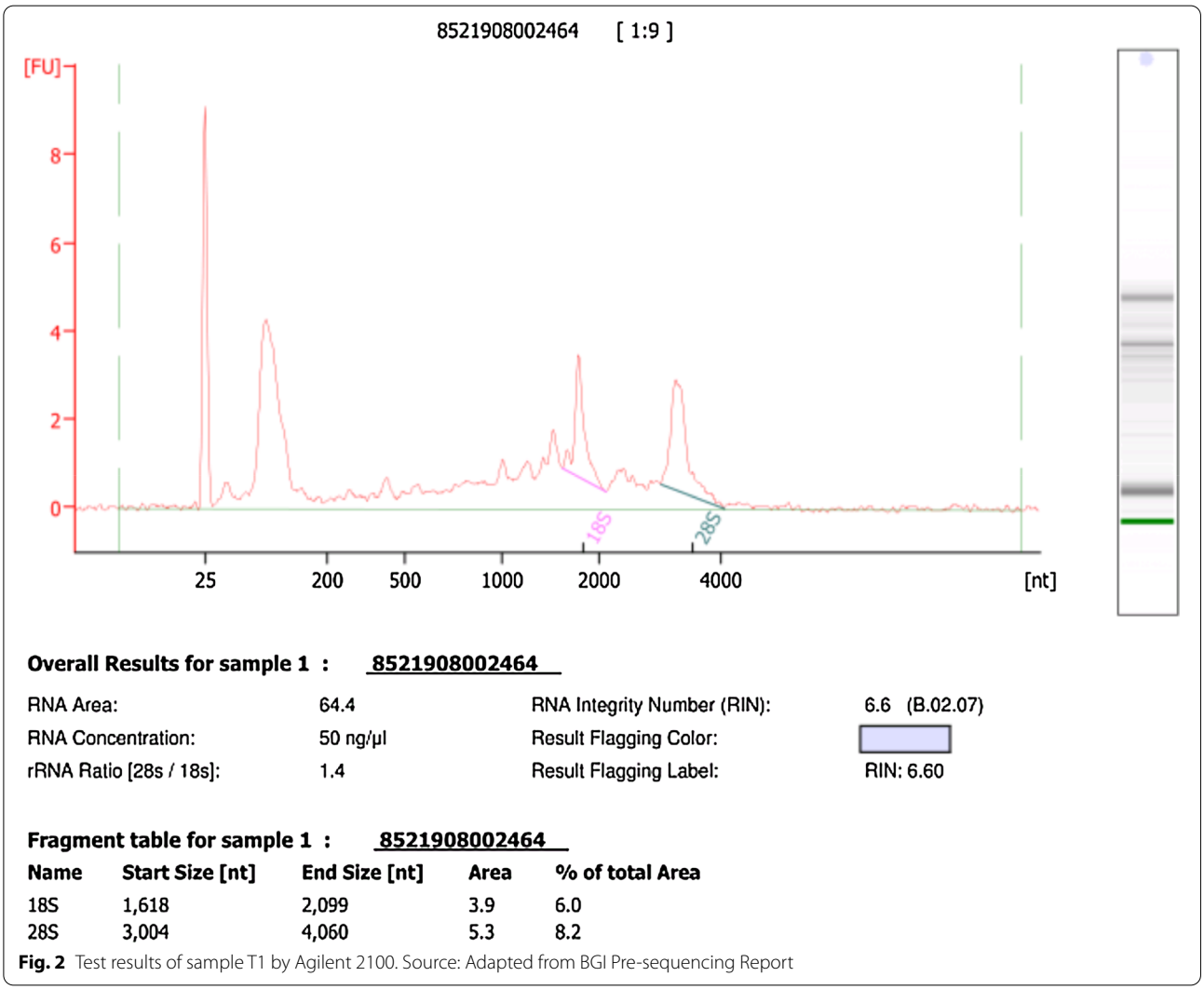
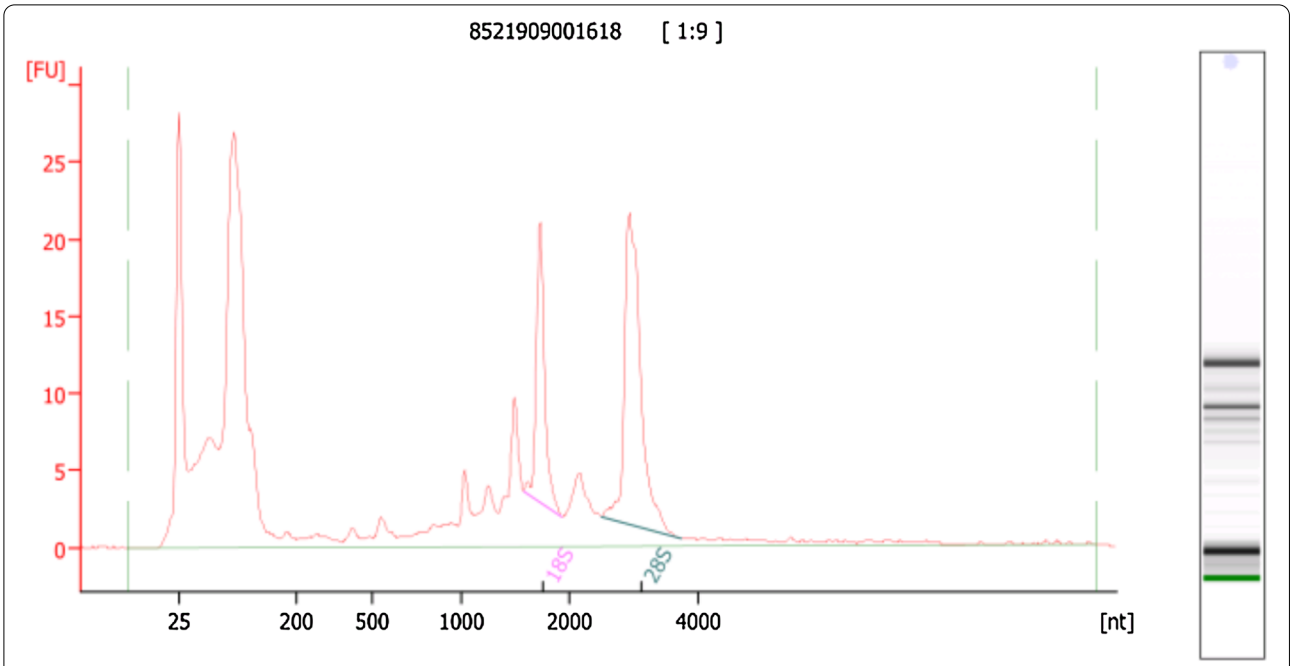


Fig. 2 Test results of sample T1 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report



Overall Results for sample 8 : 8521909001618

RNA Area:	318.7	RNA Integrity Number (RIN):	7.4 (B.02.07, Anomaly Threshold(s) manually adapted)
RNA Concentration:	110 ng/µl		
rRNA Ratio [28s / 18s]:	2.0	Result Flagging Color:	
		Result Flagging Label:	RIN: 7.40

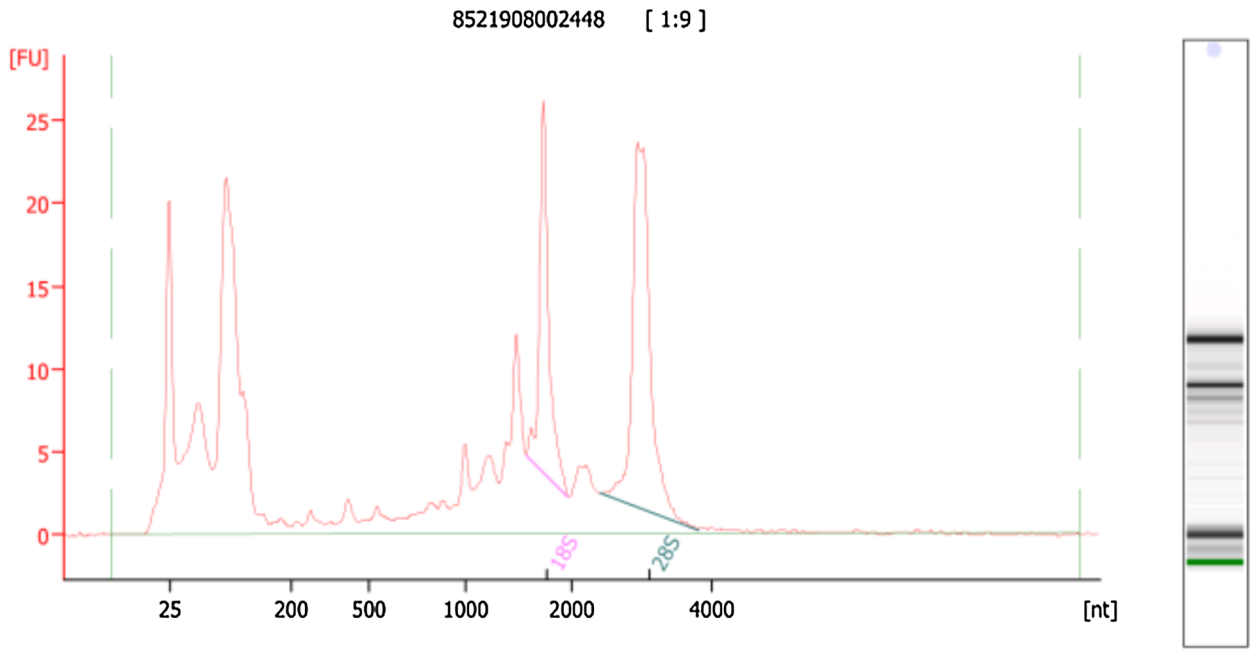
Fragment table for sample 8 : 8521909001618

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,576	1,929	20.9	6.6
28S	2,500	3,722	42.1	13.2

Fig. 3 Test results of sample T2 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report

Improved CTAB method

This method was modified based on RNA miniprep using the CTAB method [42] (Table 2). One gram of the powdered sago palm leaf sample was transferred into a 50-ml polypropylene centrifuge tube. Then, 15 ml of CTAB buffer with 1% (v/v) β-mercaptoethanol was added, mixed and then incubated for 5 min. After that, 15 ml of chloroform was added into the mixture, mixed and then incubated for another 5 min. The mixture was centrifuged at 20,922 *g*, 4 °C for 5 min. Next, the top aqueous layer from the tube was transferred into a new polypropylene centrifuge tube. Then 15 ml of chloroform was added into the aqueous layer tube, mixed and then incubated for 5 min. The mixture was centrifuged at 20,922 *g*, 4 °C for 5 min. Next, the top aqueous layer from tube again was transferred into new 50-ml polypropylene centrifuge tube. Then, equal volume of ice-cooled isopropanol was added, mixed and then incubated in room temperature for 10 min. After that, the mixture was centrifuged at 20,922 *g*, 4 °C for 15 min. Next, the supernatant from the tube was discarded and the pellet was transferred, with 1 ml of 70% ethanol, into a new 1.5-ml microcentrifuge tube. Then, the tube was centrifuged at 16,100 *g*, 4 °C for 2 min. The pellet was washed again with 1 ml of 70% ethanol and then discarded. Lastly, the pellet was resuspended with appropriate buffer for further analysis or storage.



Overall Results for sample 5 : 8521908002448

RNA Area:	337.3	RNA Integrity Number (RIN):	7.5 (B.02.07, Anomaly Threshold(s) manually adapted)
RNA Concentration:	139 ng/µl		
rRNA Ratio [28s / 18s]:	1.7	Result Flagging Color:	
		Result Flagging Label:	RIN: 7.50

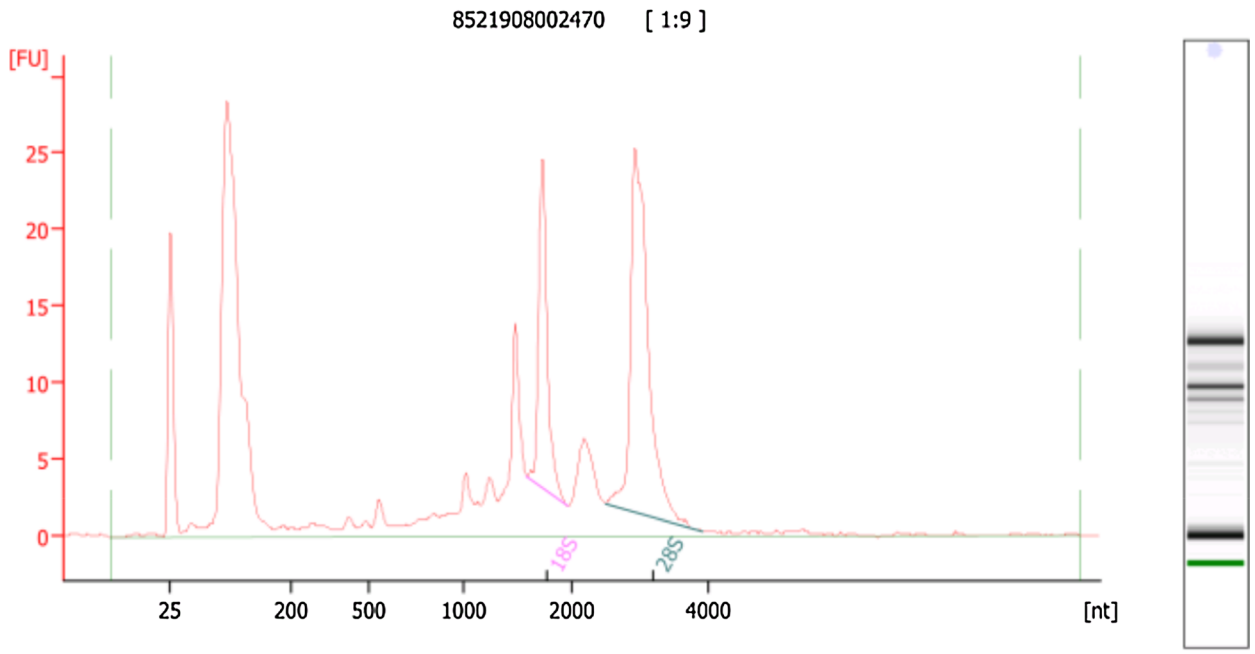
Fragment table for sample 5 : 8521908002448

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,560	1,964	29.8	8.8
28S	2,392	3,803	49.4	14.7

Fig. 4 Test results of sample T3 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report

PCI method

This method was adapted based on the RNA extraction method performed by Hussain [23]. One gram of the powdered sago palm leaf sample was transferred into a 50-ml polypropylene centrifuge tube. Then, 15 ml ice-cooled extraction buffer [150 mM lithium chloride (LiCl), 50 mM Tris–HCl (pH 9.0), 5 mM ethylene diaminetetraacetic acid (EDTA), and 5% w/v sodium dodecyl sulfate (SDS)] was added, mixed and incubated for 5 min. After that, the mixture was centrifuged at 20,922 *g*, 4 °C for 5 min. Next, the top aqueous layer from tube was transferred into a new 50-ml polypropylene centrifuge tube. The following PCI treatment was performed trice where 15 ml of ice-cooled PCI was added, mixed and incubated for 5 min. After that, the mixture was centrifuged at 20,922 *g*, 4 °C for 5 min. Then, the top aqueous layer from the tube was transferred into a new 50-ml polypropylene centrifuge tube. After the PCI treatment, the following chloroform treatment was performed thrice where 15 ml of chloroform were added and mixed for 5 min. Then, the mixture was centrifuged at 20,922 *g*, 4 °C for 5 min. After that, the top aqueous layer from the tube was transferred into a new 50-ml polypropylene centrifuge tube. After the chloroform treatment, 8 M of lithium chloride was added to make the final concentration of 2 M lithium chloride mixture and was incubated for 24 h at 4 °C. Then, the mixture was centrifuged at 20,922 *g*, 4 °C for 30 min. After that, the pellet was transferred into a new 1.5-ml microcentrifuge tube and washed with 70% ethanol. The pellet was then dissolved in 400 µl sterile dH_2O. Then, 1 ml of absolute ethanol and 40 µl 3 M Sodium acetate was added into the mixture and mixed. After that,



Overall Results for sample 10 : 8521908002470

RNA Area:	309.2	RNA Integrity Number (RIN):	7.7 (B.02.07)
RNA Concentration:	142 ng/µl	Result Flagging Color:	
rRNA Ratio [28s / 18s]:	2.0	Result Flagging Label:	RIN: 7.70

Fragment table for sample 10 : 8521908002470

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,581	1,966	26.4	8.5
28S	2,507	3,929	53.1	17.2

Fig. 5 Test results of sample N1 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report

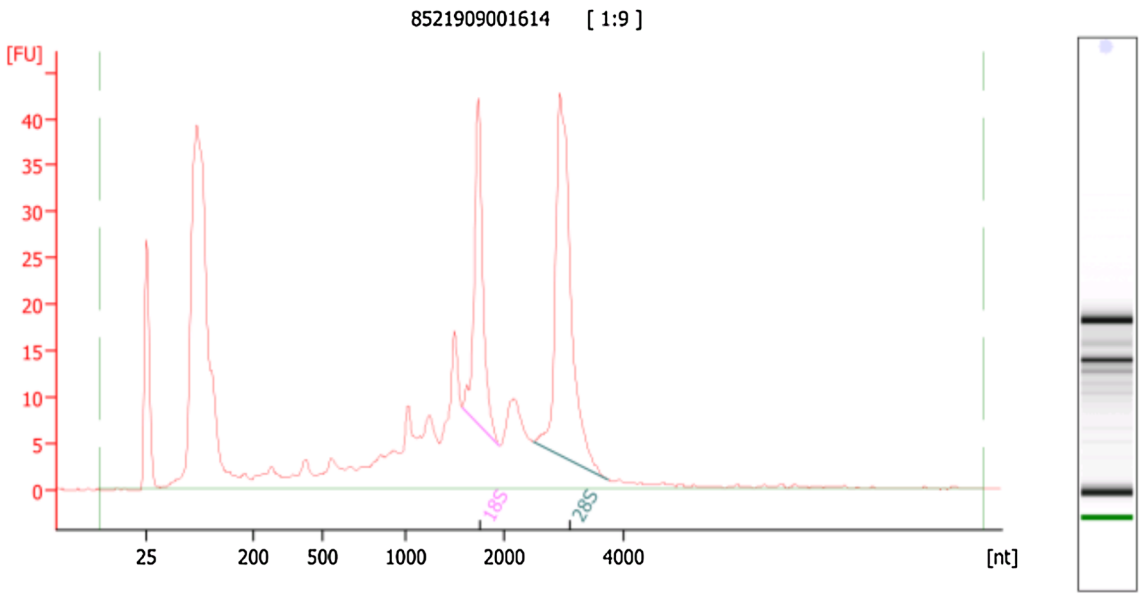
the mixture was incubated at − 80 °C for 20 min. Then, the mixture was centrifuged at 16,100 *g*, 4 °C for 30 min. Next, the supernatant was discarded and the pellet was washed with 70% ethanol twice. Lastly, the pellet was resuspended with appropriate buffer for further analysis or storage.

MRIP method

This method was adapted based on the RNA extraction method performed by Xiao et al. [41]. One gram of the powdered sago palm leaf sample was transferred into a 50-ml polypropylene centrifuge tube. Then, 15 ml MRIP buffer [3.05 g ammonium thiocyanate (% w/v), 9.44 g guanidine thiocyanate (% w/v), 3.33 ml 3 M sodium acetate (pH 5.2) (% v/v) and 38 ml phenol (% v/v), adjust pH to 5.0 with acetic acid] was added, mixed and incubated for 5 min. Then, 5 ml ice-cooled chloroform was added, mixed and further incubated for 5 min. After that, the top aqueous layer from tube was transferred into a new 50-ml polypropylene centrifuge tube. Next, equal volume of ice-cooled isopropanol were added, mixed and incubated for 10 min. Then, the mixture was centrifuged at 20,922 *g*, 4 °C for 15 min. The supernatant was discarded and the pellet was washed with 70% ethanol twice. Lastly, the pellet was resuspended with appropriate buffer for further analysis or storage.

RNA purity test

RNA purity test was performed using Eppendorf Biophotometer Plus for RNA extraction methods comparison and Nanodrop was used to test the RNA quality for RNA sequencing samples. Samples pellet were dissolved in 100 µl Tris–EDTA and transferred into Eppendorf UVette cuvette for the instrument measurement in Eppendorf Biophotometer Plus while Samples pellet were dissolved in nuclease-free water for nanodrop.



Overall Results for sample 4 : 8521909001614

RNA Area:	537.4	RNA Integrity Number (RIN):	7.4 (B.02.07, Anomaly Threshold(s) manually adapted)
RNA Concentration:	185 ng/µl		
rRNA Ratio [28s / 18s]:	1.8	Result Flagging Color:	
		Result Flagging Label:	RIN: 7.40

Fragment table for sample 4 : 8521909001614

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,578	1,957	47.4	8.8
28S	2,485	3,749	85.6	15.9

Fig. 6 Test results of sample N2 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report

The purity of the sample's RNA extract was measured by the OD260/OD230 and OD260/OD280 ratio where the higher the value of the ratio the purer the sample's RNA [43].

RNA quality test

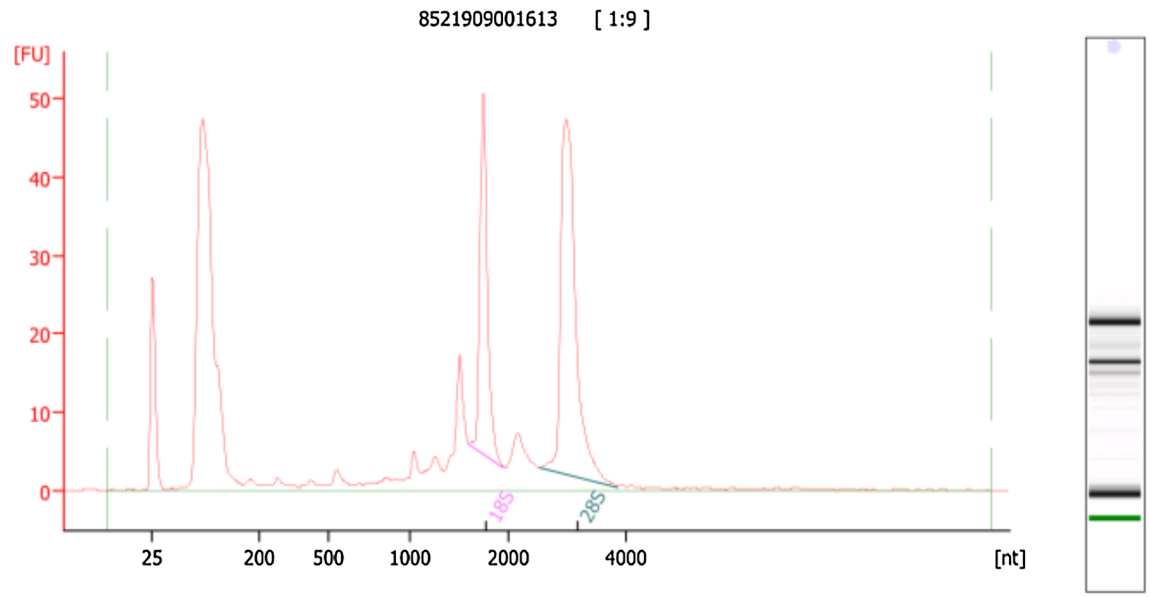
RNA quality test was performed using Tris–borate–EDTA buffered Agarose Gel Electrophoresis (TBE-AGE) for RNA extraction methods comparison and Agilent 2100 Bioanalyzer was used to test the RNA quality for RNA sequencing samples. In TBE-AGE, sample pellets were dissolved in 100 µl Tris–HCl. 5 µl of the RNA samples was mixed with 2 µl loading dye. The samples mixtures and GeneRuler 1-kb DNA ladder were loaded into the gel to perform the electrophoresis. In TBE-AGE, good-quality RNA sample represented by intact 28 s and 18 s rRNA band without smearing. In Agilent 2100 Bioanalyzer, sample pellets were dissolved and analyzed. RNA Integrity Number (RIN) is use to determine the quality of the RNA sample scored from 1 (worst) to 10 (best) [44].

Statistical analysis

IBM SPSS Statistics 28.0.1.1 (15) was used to analyse the 260/280 ratio, 260/230 ratio and RNA weight data obtained for RNA extraction method comparison. All the data were assessed using normality test to determine its parametricity. Parametric data were further subjected to multivariate analysis of variance (MANOVA) coupled with Tukey's test while non-parametric data were subjected to Kruskal–Wallis test.

RNA sequencing

The improved CTAB method was used to extract RNA from sago palm leaf tissue and sent for sequencing. The RNA samples were pelletized in 100% absolute ethanol and shipped for sequencing service. BGISEQ-500 sequencing platform was used in this study. The RNA



Overall Results for sample 3 : 8521909001613

RNA Area:	491.5	RNA Integrity Number (RIN):	8.3 (B.02.07, Anomaly Threshold(s) manually adapted)
RNA Concentration:	169 ng/µl		
rRNA Ratio [28s / 18s]:	1.8	Result Flagging Color:	
		Result Flagging Label:	RIN: 8.30

Fragment table for sample 3 : 8521909001613

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	1,601	1,955	52.2	10.6
28S	2,530	3,845	96.4	19.6

Fig. 7 Test results of sample N3 by Agilent 2100. Source: Adapted from BGI Pre-sequencing Report

sequencing data were measured by Q20 and Q30 values. BGI NGS RNA sequencing guideline for good-quality data are Q20 ≥ 90%, Q30 ≥ 80 [45]. The assembly sequences median length (N50) and the mean length provides information on abundance of the sample RNA length, the closer the N50 value to the mean length the more normalized the distribution of the sample RNA length.

Results and discussion

RNA extraction methods comparison

The improved CTAB method was compared with the PCI method [23] and MRIP methods [41] and showed the best result in RNA quality and quantity (Table 3 and Fig. 1). Due to low RNA yield of MRIP method, there are no visible bands in TBE-AGE. On the other hand, there are presence of visible smear at the 5 s rRNA region, but absent at 28 s and 18 s rRNA band in TBE-AGE of PCI method indicating the considerable degradation of the RNA. CTAB method gives the best RNA quality as shown by the intact 28 s and 18 s rRNA band with slightly visible smear at the 5 s rRNA region in TBE-AGE. This outcome shows that the PCI method is not suitable for extracting RNA from mature sago palm leaf tissue. The lower sample weight used in this test improved the RNA yield and purity of the extracted RNA compared to the reported result. This outcome also shows that MRIP method is not suitable for RNA extraction from mature sago palm leaf tissue although it shows promising result when applied on RNA extraction from young coconut leaf. The purity and the RNA yield are very much lower compared to the reported result on application of MRIP method on young coconut leaf.

The 260/280 ratio, 260/230 ratio and the RNA weight data were subjected to normality test using IBM SPSS Statistics 28.0.1.1 (15). The null hypothesis (H_0) for the normality test assumes the population is normally distributed and the standard significance level (α) is 0.05.

Table 8 RNA sequencing clean reads quality metrics from sequencing report. Source: Reproduced from BGI RNA-Sequencing Report

Sample	Total raw reads (Mb)	Total clean reads (Mb)	Total clean bases (Gb)	Clean reads Q20 (%)	Clean reads Q30 (%)	Clean reads ratio (%)
T1	69.96	66.82	6.68	97.88	91.45	95.51
T2	69.96	66.97	6.7	98.09	92.07	95.72
T3	69.96	66.68	6.67	97.76	91.14	95.31
N1	69.96	66.66	6.67	98	91.78	95.27
N2	69.41	66.62	6.66	97.94	91.59	95.99
N3	69.96	67.33	6.73	97.98	91.63	96.24

All the samples show good sequencing results with above average clean read ratio and clean read Q30 ratio

Sample: T1, T2, T3, N1, N2, and N3 are biological replicates of mature sago palm leaf RNA extract

Total raw reads (Mb): the reads amount before filtering

Total clean reads (Mb): the reads amount after filtering

Total clean bases (Gb): the total base amount after filtering

Clean reads Q20 (%): the rate of bases which quality is greater than 20 value in clean reads

Clean reads Q30 (%): the rate of bases which quality is greater than 30 value in clean reads

Clean reads ratio (%): the ratio of the amount of clean reads

Table 9 Sequenced RNA transcripts quality metrics from sequencing report. Source: Reproduced from BGI RNA-Sequencing Report

Sample	Total number	Total length	Mean length	N50	N70	N90	GC (%)
T1	75,730	56,765,699	749	1334	740	283	45.54
T2	94,989	73,822,623	777	1447	778	285	44.11
T3	84,578	65,560,398	775	1430	776	285	44.67
N1	68,068	55,890,222	821	1472	860	309	45.44
N2	91,100	65,484,623	718	1303	684	270	44.75
N3	97,205	78,047,560	802	1503	821	293	43.65

All the samples show similar assembly results with mean length above 700 bp and N50 above 1300

Sample: T1, T2, T3, N1, N2, and N3 are biological replicates of mature sago palm leaf RNA extract

Total number: the total number of transcripts

Total length: the read length of transcripts

Mean length: the average length of transcripts

N50: the N50 length is used to determine the assembly continuity, the higher the better. N50 is a weighted median statistic that 50% of the total length is contained in Unigenes that are equal to or larger than this value;

N70: similar to N50

N90: similar to N50

GC (%): the percentage of G and C bases in all transcripts

The P-values of 260/280 ratio and 260/230 ratio data are higher than α so we accept H_0, hence they are normally distributed (Table 4). On the other hand, P-value of RNA weight is lower than α so we reject H_0, hence the data are not normally distributed. Since both 260/280 ratio and 260/230 ratio data are normally distributed, they were subjected to MANOVA coupled with Tukey's test while because of the RNA weight data are not normally distributed, they were subjected to independent-samples Kruskal–Wallis test using IBM SPSS Statistics 28.0.1.1 (15). The H_0 for the Tukey's test assumes all means being compared are from the same population and the standard significance level (α) is 0.05. The P-values between all the RNA extraction methods are lower than the α (0.05) value so we reject H_0, hence the 260/280 ratio and 260/230 ratio data between all the RNA extraction methods are not the same (Table 5). The H_0 for the Kruskal–Wallis test assumes the distribution of the weight is the same across the RNA extraction methods and the standard significance level (α) is 0.05. The P-values of all the

RNA extraction methods are higher than the α (0.05) value so we accept H_0, hence distribution of the weight is the same across the RNA extraction methods (Table 6).

The improved CTAB method gives the best RNA yield with an average of 177.5 µg RNA for 1 g mature sago palm leaf tissue while PCI method and MRIP method only gives 11.17 µg and 7.97 µg RNA, respectively, even though the difference is not significant between RNA extraction methods. The RNA purity of the CTAB method for protein contamination as shown by the 260/280 ratio was significantly higher with an average of 1.93 than the PCI method and MRIP method with an average of 1.635 and 1.18, respectively. The RNA purity of the CTAB method for other organic compounds contamination as shown by 260/230 ratio was also significantly higher with an average of 1.955 than the PCI method and MRIP method with an average of 0.86 and 0.2, respectively. Hence, the improved CTAB method is superior for RNA extraction on mature sago palm leaf samples compared with the PCI method and MRIP method.

RNA extraction of mature sago palm leaf tissue by improved CTAB method

Pre-sequencing RNA sample quality and quantity tests were performed on the samples with NanoDrop and Agilent 2100 (refer Table 7 and Figs. 2, 3, 4, 5, 6, 7). Note that the RNA yield by the sequencing service provider is underestimating the actual RNA yield because the sample pellet was not fully dissolved during pellet resuspension by the sequencing service provider (data on actual RNA yield before pelletized and shipped for RNA sequencing are provided as Additional File 1).

There is high intensity of 5 s and some consistent signals pattern at the 18 s and 5 s fast region across the samples which is not ideal compared to the RIN standard but the outcome of the sequencing results showed otherwise. RNA sequencing of the extracted sago palm leaf tissue RNA by improved CTAB method gives a good result as all the samples exceed good-quality data definition for next-generation sequencing where Q20 ≥ 90%, Q30 ≥ 80% (refer Table 8). The *de novo* assembly of the sequenced RNA transcripts shows mean length > 700 bps and N50 > 1300 bps (refer Table 9) can be used as a guideline for future RNA sequencing of sago palm RNA extracts. The high intensity of the 5 s and some consistent signals pattern at the 18 s and 5 s fast region could indicate the RNA profile of mature sago palm leaf tissue and do not represent the degradation of the RNA sample. The below-average sample T1 RIN and 28 s/18 s ratio do not affect the sequencing result much as the clean read ratio, clean read Q30, mean assembly length and N50 values are comparable with other samples.

Conclusion

In this study, three RNA extraction methods, i.e., improved CTAB, PCI and MRIP, were compared and evaluated based on the best RNA purity and quality extracted from sago palm leaf tissue. Overall, the best result of RNA purity, quantity and quality of RNA obtained are by the improved CTAB method. The total RNA extracts using the improved CTAB method were then sequenced and this method generates good sequencing data Q20 ≥ 97, Q30 ≥ 91. The mean length ≥ 700 bp and the N50 value ≥ 1.3 kbp of the assembly can be used as a guideline for future RNA sequencing in sago palm.

Abbreviations

RNA: Ribonucleic acid; CTAB: Cetyltrimethylammonium bromide; spp.: Species (plural); PCI: Phenol–chloroform–isoamyl alcohol; MRIP: Methods for RNA isolation from palms; Tris: (Hydroxymethyl)aminomethane; HCl: Hydrochloric acid; EDTA: Ethylenediaminetetraacetic acid; BGI: Beijing Genome Institute; NGS: Next-Generation Sequencing.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40538-022-00329-9.

Additional file 1. The RNA purity profile prior to shipment for RNA sequencing.

Acknowledgements

This research is supported by Sarawak Research Development Council Grant: (RDCRG/CAT/2019/23). The authors thank Land Custody and Development Authority (LCDA) Sdn. Bhd. for the permission to obtain samples from their plantation and for logistical supports.

Author contributions

WJY planned and performed the experiments, analysis, and is the main author of the manuscript. HH designed the overall study, is the major contributor in data analysis and manuscript preparation. FHP performed the methods verification experiments and contributed to the manuscript drafting. All authors have read and approved the manuscript.

Funding

Open Access funding provided by Universiti Malaysia Sarawak. This study was supported and funded by Sarawak Research Development Council Grant: (RDCRG/CAT/2019/23).

Availability of data and materials

All data are provided in this manuscript. Additional information on the sequencing data can be accessed on:
Repository name: NCBI's Gene Expression Omnibus (GEO).
Data identification number: GSE189085.
Direct URL to data: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE189085.
Repository name: NCBI's Sequence Read Archive (SRA).
Sample ID: GSM5694359 (ST1: Trunking Sample 1).
Data identification number: SRX13165895.
Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165895.
Sample ID: GSM5694360 (ST4: Trunking Sample 4).
Data identification number: SRX13165896.
Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165896.
Sample ID: GSM5694361 (ST5: Trunking Sample 5).
Data identification number: SRX13165897.

Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165897.
Sample ID: GSM5694362 (NT7: Trunking Sample 7).
Data identification number: SRX13165898.
Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165898.
Sample ID: GSM5694363 (NT8: Trunking Sample 8).
Data identification number: SRX13165899.
Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165899.
Sample ID: GSM5694364 (NT9: Trunking Sample 9).
Data identification number: SRX13165900.
Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/SRX13165900.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 13 June 2022 Accepted: 24 August 2022
Published online: 13 September 2022

References
1. Ehara H. Genetic variation and agronomic features of *Metroxylon* palms in Asia and Pacific. In: Ehara H, Toyoda Y, Johnson DV, editors. Sago Palm : multiple contributions to food security and sustainable livelihoods. Singapore: Springer Nature; 2018. p. 45–59.
2. Ellen R. Local Knowledge and Management of Sago Palm (*Metroxylon sagu* Rottboell) Diversity in South Central Seram, Maluku, Eastern Indonesia. J Ethnobiol. 2006;26(2):258–98.
3. Flach M. Sago palm: Metroxylon sagu Rottb. Rome: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute; 1997.
4. Mochamad HB, Muhammad IN, Agief JP, Fendri A, Liska A. Growing area of sago palm and its environment. In: Ehara H, Toyoda Y, Johnson DV, editors. Sago palm: multiple contributions to food security and sustainable livelihoods. Singapore: Springer Nature; 2018. p. 17–29.
5. Abdurakhmonov IY. Bioinformatics: basics, development, and future, bioinformatics—updated features and applications. London: IntechOpen; 2016. https://doi.org/10.5772/63817.
6. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63. https://doi.org/10.1038/nrg2484.
7. Firtzlaff V, Oberländer J, Geiselhardt S, et al. Pre-exposure of Arabidopsis to the abiotic or biotic environmental stimuli "chilling" or "insect eggs" exhibits different transcriptomic responses to herbivory. Sci Rep. 2016;6:28544. https://doi.org/10.1038/srep28544.
8. Liu H, Wang J, Sun H, Han X, Peng Y, Liu J, Liu K, Ding Y, Wang C, Du B. Transcriptome profiles reveal the growth-promoting mechanisms of *Paenibacillus polymyxa* YC0136 on tobacco (*Nicotiana tabacum* L.). Front Microbiol. 2020;11:584174. https://doi.org/10.3389/fmicb.2020.584174.
9. Rochester DE, Winer JA, Shah DM. The structure and expression of maize genes encoding the major heat shock protein, hsp70. Eur Mol Biol Org. 1986;5(3):451–8.
10. Chong IP. Isolation, characterisation of partial small and large subunits of the ADP-glucose pyrophosphorylase gene from sago palm (Metroxylon sagu). (Master dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2008.
11. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. Anal Biochem. 1987;162:156–9.
12. Lau JS. Isolation and characterisation of near complete cDNA and genomic sequences coding for granule-bound starch synthase in sago palm (*Metroxylon sagu*). (Master dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2001.
13. Shu-Jun C, Puryear J, Cairney J. A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep. 1993;11:113–6.
14. Budiani A, Putranto RA, Minarsih H, Riyadi I, Abbas B. Ekspresi dan kloning gen penyandi ADP-Glucose Phyrophosphorylase dari tanaman sagu (*Metroxylon sagu* Rottb.). Menara Perkeb. 2015;83(2):76–85.
15. Schultz DJ, Craig R, Cox-Foster DL, Mumma RO, Medford JI. RNA isolation from recalcitrant plant tissue. Plant Mol Biol Rep. 1994;12(4):310–6.
16. Bong SK. Screening for Flower-Specific cDNA Sequences in Sago Palm (*Metroxylon sagu*) via a Differential Display Technique. (Master dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2004.
17. Jamel B, Hussain MH, Salleh MA, Busri N. Total RNA isolation from sago palm. CRAUN Sago Res J. 2006;2:165–9.
18. Salzman RA, Fujita T, Zhu-Salzman K, Hasegawa PM, Bressan RA. An Improved RNA Isolation Method for Plant Tissues Containing High Levels of Phenolic Compounds or Carbohydrates. Plant Mol Biol Report. 1997;17:11–7.
19. Kiefer E, Heller W, Ernst D. A Simple and Efficient Protocol for Isolation of Functional RNA from Plant Tissues Rich in Secondary Metabolites. Plant Mol Biol Report. 2012;18:33–9.
20. Siti Izyan-Liyana Kamarol (2015). Representational Difference Analysis (RDA) for Identification of Molecular Factors Contributing to Trunking and Non Trunking Sago Palm (*Metroxylon sagu*). [Master dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
21. Chun-Gen Hu, Honda C, Kita M, Zi-Lian Z, Tsuda T, Moriguchi T. A Simple Protocol for RNA Isolation from Fruit Trees Containing High Levels of Polysaccharides and Polyphenol Compounds. Plant Mol Biol Report. 2002;20:69.
22. Anastasia-Shera Edward Atit (2013). Differential Expression Gene Profiling Studies on Trunking and Non-Trunking Sago Palm (*Metroxylon sagu* Rottb.). [Master dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
23. Mohd Hasnain Hussain (2002). Analysis of Debranching Enzymes from Pea and Potato. [Doctor of Philosophy dissertation, University of East Anglia].
24. Baharina Wati Barozah (2005). Synthesis of First Strand cDNA from Total RNA of *Metroxylon Sagu* Leaf. [Bachelor dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
25. Farizan Bin Nan (2006). *Isolation of cDNA Fragment Encoding Starch Synthase Gene from Metroxylon sagu by RT-PCR Method*. [Bachelor dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
26. Tajuddin Sidek Hairaddin (2006). *Isolation of cDNA Fragment Encoding Starch Branching Enzyme (Isoform I) Gene from Metroxylon sagu by RT-PCR method*. [Bachelor dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
27. Rohana Musa (2006). Isolation of cDNA Fragment Encoding Isoamylase Gene from *Metroxylon Sagu* by RT-PCR Method. [Bachelor dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
28. Mohd-Azinuddin Ahmad Mokhtar (2006). Isolation of cDNA fragment encoding starch branching enzyme (isoform II) gene from *Metroxylon sagu* by RT-PCR method. [Bachelor dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
29. Zulaikha Pol Ong (2019). Rapid Amplification of cDNA Ends of Differentially Expressed Genes from Trunking and Non-Trunking Sago Palm (*Metroxylon Sagu* Rottb.). [Master dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
30. Zhen Ying and Yang Tao. RNA Isolation from Highly Viscous Samples Rich in Polyphenols and Polysaccharides. Plant Mol Biol Report. 2002;20:417a–417e.
31. Jerry, anak Gerunsin (2014) Isolation, Characterisation and Expression of Fructose-1,6-bisphophate Aldolase from Sago Palm (*Metroxylon sagu*) Leaves. [Master dissertation, Universiti Malaysia Sarawak]. UNIMAS Institutional Repository
32. Roslan HA, Md.-Anowar Hossain, Jerry Gerunsin,. Molecular and 3D-Structural Characterisation of Fructose-1,6-Bisphosphate Aldolase Derived from *Metroxylon Sagu*. Braz Arch Biol Technol. 2017;60: e17160108. https://doi.org/10.1590/1678-4324-2017160108.

33. Wu Y, Llewellyn DJ, Dennis ES. A quick and easy method for isolating good-quality RNA from cotton (*Gossypium hirsutum* L.) tissues. Plant Mol Biol Rep. 2002;20:213–8.
34. Gasic K, Hernandez A, Korban SS. RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. Plant Mol Biol Rep. 2004;22:437a–437g.
35. Holling N. Screening of leafy-like gene in sago meristem and various tissues. (Bachelor dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2008.
36. Julai N. Rapid amplification of cDNA ends and cDNA screening of alcohol dehyrdrogenase genes from *Metroxylon sagu*. (Bachelor dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2008.
37. Wee CC. Generation of expressed sequence tags (Ests) database and isolation of alcohol dehydrogenase gene from young leaf samples of *Metroxylon sagu*. (Master dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2008.
38. Roslan HA, Anji SB. Characterisation of inflorescence-predominant chitinase gene in *Metroxylon sagu* via differential display. 3 Biotech. 2011;1:27–33. https://doi.org/10.1007/s13205-011-0004-x.
39. Wee CC, Roslan HA. Isolation of alcohol dehydrogenase cDNA and basal regulatory region from *Metroxylon sagu*. Int Sch Res Netw. 2012;2012:839427. https://doi.org/10.5402/2012/839427.
40. Rosli SS. Isolation, heterologous expression and functional characterisation of phenyalanine ammonia lyase gene (msPAL) from sago palm (*Metroxylon sagu* Rottbl.). (Master dissertation, Universiti Malaysia Sarawak). UNIMAS Institutional Repository. 2017.
41. Xiao Y, Yang Y, Cao H, Fan H, Ma Z, Lei X, Mason AS, Xia Z, Huang X. Efficient isolation of high quality RNA from tropical palms for RNA-seq analysis. Plant Omics. 2012;5(6):584–9.
42. Untergasser A. "RNA Miniprep using CTAB" Untergasser's Lab. 2008. http://www.untergasser.de/lab/protocols/miniprep_rna_ctab_v1_0.htm. Accessed 1 Aug 2018.
43. DeNovix Inc., Purity ratios explained. Technical note 130. DeNovix incorporated. 2022. https://www.denovix.com/tn-130-purity-ratios-explained/. Accessed 5 Aug 2022.
44. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006. https://doi.org/10.1186/1471-2199-7-3.
45. BGI. DNBseqTM service overview: RNA-Seq (Transcriptome) sequencing. PN: BGIASO4_061319. BGI Genomics. 2019. https://www.bgi.com/api/assets/01b07d24-e382-4c07-b657-3ee75c9f7c0d. Accessed 5 Aug 2022.

Publisher's Note