Sago Palm Genome Size Estimation via Real-Time Quantitative PCR

Hairul Azman Roslan[1]*, Md Anowar Hossain[2], Ngieng Ngui Sing[1] and Ahmad Husaini[1]

[1]Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia
[2]Department of Biochemistry and Molecular Biology, University of Rajshahi, Rajshahi, Bangladesh

Received: 9 December 2019, Revised: 23 February 2020, Accepted: 27 February 2020

Abstract

Sago palm, *Metroxylon sagu* Rottb., is an underutilized indigenous food crop that can be found mainly in the South East Asia and Pacific regions. It is a main starch producer and socioeconomically important crop in the South East Asia region including Malaysia. The sago starch provides for considerable potential to food security in the places where it is grown. However, not many molecular works have been reported thus far. In the post genomic era, sago plant genome sequencing is very important for sustainable starch development in these regions. Therefore, determination of the genome size is prerequisite to full genome sequencing and assembly. Here we report on the use of real-time quantitative polymerase chain reaction (qPCR) in determining the genome size. For this work, we calculated the genome size, Γ (bp) of *M. sagu* based on qPCR-derived copy number of two single copy genes. *Pichia pastoris,* with a known genome size, was used as a control to estimate sago palm genome size. With this technique, the genome size of *M. sagu* was calculated to be 1.87 Gbp. This genome size information would be beneficial for subsequent molecular work including genome sequencing and analysis on this economically important crop plant.

Keywords: Genome size, *Metroxylon sagu*, real-time PCR, copy number, *Pichia pastoris*
DOI 10.14456/cast.2020.10

1. Introduction

Sago palm (*Metroxylon sagu* Rottb.) is a palm that is widely distributed in the South East Asia region. It is a starch-producing crop and economically important to the state of Sarawak in Malaysia. Sarawak has the largest acreage planted with sago palm and exports sago starch to countries such as Taiwan, Japan, Singapore and others, generating incomes of up to US$10.8 million/year [1]. Despite its importance, little research in molecular biology has been completed on *M. sagu* molecular biology, and it has not been sequenced, because genome size is not known for this emerging crop. The genome size is a prerequisite in genome sequencing project where it is needed to calculate the number of clones needed to be generated in shotgun sequencing and library

*Corresponding author: Tel.: +60 82 58 3038 Fax: +60 82 58 3160
E-mail: hairulroslan@hotmail.com

screening. The information would also be beneficial to estimate read coverage in next generation sequencing of the genome and its subsequent analyses [2].

An organism genome size is defined as the amount of its DNA and commonly denoted as "C" value or "Γ". The content represents the amount of DNA per haploid genome and is calculated in picograms. Several techniques can be used to estimate genome size such as pulse field gel electrophoresis (PFGE) where the genome was cut into smaller pieces and directly measured [3]. PFGE have been successfully used to determine smaller genomes typically prokaryotes and unicellular eukaryotes [3]. Earlier methods used to estimate genome size are based on either determining the phosphate contents in DNA strands of a defined number of cells or on re-association kinetics of high molecular weight of genomic DNA (C_ot assay) [3]. More recently, D'Hondt and co-workers reported the use of flow and absorption cytometry that determined genome size characterization via DNA-tagged fluorescent dyes [4].

In this work, to estimate the genome size of sago palm, we followed a real-time quantitative polymerase chain reaction (qPCR) technique described by Mounsey and co-workers in 2012 [5] where the amount of genetic materials can be determined based on a known amount of DNA. The real time qPCR combines the nucleic acid amplification and identification in one step and finally quantify the product in a real time by fluorescent detection. This method uses a known DNA mass that was determined by UV spectrophotometry and relatively pure DNA content, that is then used to calculate the C-value by dividing the mass of the DNA sample by the copy number of single copy genes [5]. Using this simple strategy, the genome sizes of various eukaryotic organisms have been estimated such as fungus [6], lichen [7], house-fly [8], fruit fly [9], mites [10], wasp [11], human liver fluke [12], Venus flytrap [13] and three different eukaryotes (yeast, swordtail fish and human) [14]. Here we report on the work to determine the genome size of sago palm using the real-time quantitative PCR technique.

2. Material and Methods

2.1 Collection of plant materials

Fresh young leaf samples (5 months old) of *M. sagu* were collected from the Faculty of Resource Science and Technology garden located at GPS coordinates of 1.469852°N, 110.428175°E. The midribs and stem were removed and then washed with distilled water to remove any debris. The leaves were then surface sterilized using 70% ethanol, blotted dry and cut into small pieces. The processed samples were either used immediately or stored at -80ºC.

2.2 *Metroxylon sagu* DNA isolation and purification

Total genomic DNA isolation was carried out using the CTAB method according to Wee and Roslan [1]. Approximately 0.1 g of the *M. sagu* leaves were ground to powder in liquid nitrogen and added into 1 ml of cetyl trimethylammonium bromide (CTAB) buffer (2% CTAB; 1.4 M NaCl; 20 mM EDTA; 100 mM Tris-HCl, pH 8.0) preheated together with 2 µl of β-mercaptoethanol at 65°C. The mixture was incubated for 1 h at 65°C before gently mixed with 400 µl of 24:1 ratio of chloroform: isoamyl alcohol and centrifuged at 13,000 rpm for 5 min.

The upper aqueous layer containing DNA was transferred to a new tube and treated with 0.2 µl of RNase A (10 µg/µl) (Qiagen, USA) at 37ºC for 1 h, to digest any possible RNA. Genomic DNA was then precipitated overnight at -20°C with equal volume of ice-cold propan-2-ol. The mixture was then centrifuged at 13,000 rpm for 2 min and the resulting pellet was washed

with 1 ml of 70% ethanol, air-dried and re-suspended in 50 µl of 15 mM TE buffer. The DNA samples were stored at -80°C for use in quantification and quality checking.

2.3 DNA isolation from *Pichia pastoris*

Pichia pastoris (strain GS115) was used as the positive control for the method. Three milliliter of *P. pastoris* was cultured overnight in yeast peptone dextrose (YPD) medium. Yeast genomic DNA was isolated using the method described by Mounsey and co-workers [5]. Genomic DNA was extracted by homogenizing *P. pastoris* with a conical grinder in a microcentrifuge frozen with liquid nitrogen. Two milliliter of digestion buffer (800 mM guanidine HCl; 30 mM Tris-Cl, pH 8.0; 30 mM EDTA, pH 8.0; 5% Tween-20; 0.5% Triton X-100) was added to the homogenate containing 2 µl RNase A (10 µg/µl) (Fermentas, Lithuania) and the mixture was incubated at 37°C for 30 min. To further digest all proteins, 100 µl of proteinase K (Fermentas, Lithuania) was added and incubated at 50°C for 1 h. The lysate was then centrifuged at 4000 x g for 10 min at 4°C. The resulting supernatant was transferred into a Genomic-tip 20/G (Qiagen, USA) and DNA extraction was performed according to the manufacturer's instructions. The DNA pellet was then re-suspended in 50 µl of 15 mM Tris-EDTA buffer and stored at -80°C prior to further quantification and quality checking.

2.4 DNA quality control

The concentration and purity of all extracted DNA samples were determined by measuring absorbance at 260 nm using U2900 UV-VIS spectrophotometer (Hitachi, Japan). The DNA purity with A260:A280 ratio of between 1.8 and 1.9 was regarded as having good quality, where one A260 unit correspond to 50 µg/ml DNA.

2.5 PCR primer design and preparation of qPCR standards

Sago palm primers were designed based on sago palm EST sequence available in the National Center for Biotechnology Information database (GenBank). Two sago palm single-copy nuclear genes were selected for this study, the beta-actin (GI: JK731311.1) and elongation factor (EF3) (GI: JK731265.1) genes. For the positive control using *P. pastoris* (genome size 9.43 Mbp), the primers selected were the genes actin [AF216956] and elongation factor 3 (EF3) [FN392322] [5]. Detailed information for the primers used is shown in Table 1. The PCR conditions were optimized to determine that the optimum annealing temperatures were between 55-65°C. The PCR reactions contained 0.2 mM dNTPs, 0.4 µM of each primer, 2.5 mM $MgCl_2$, 0.2 U *Taq* Polymerase, and 1 µl template gDNA. PCR cycle conditions include an initial denaturation at 95°C for 2 min, followed by 35 cycles of a denaturation-step of 94°C for 30 s, annealing-step of 55-65°C for 30 s, and elongation-step of 72°C for 1 min, and a final extension for 10 min at 72°C. The PCR products were visualized on 1.5% agarose gels, purified and ligated into pGEM-T vector (Promega, USA). The PCR products were then sequenced to confirm the genes had been amplified as intended. The plasmid was linearized with *NcoI* restriction enzyme and purified. The DNA was then quantified using NanoDrop 2000/2000c spectrophotometer (Thermo Scientific, USA).

Table 1. Primers used for qPCR in this study.

Gene	Primers	Sequence (5'→3')	PCR Size	Tm
Beta-actin (*P. pastoris*)	PpAct-F	GAT GCG ATC TCC GTT TGA TT	246 bp	60.04
	PpAct-R	GGA AGC GTA CAG GGA CAA AA		60.11
Elongation Factor 3 (*P. pastoris*)	PpEF3-F	CCA TTT GGA CAC TGT CAA CG	246 bp	60.0
	PpEF3-R	CCT GGT TCT GGG AAC TTG AA		60.08
Beta-actin (*M. sagu*)	MsAct-F	GCA CGA TTG AAG GAC CAC TT	185 bp	60.12
	MsAct-R	TGC TGA TCG TAT GAG CAA GG		59.97
Elongation Factor (*M. sagu*)	MsEF-F	CTC TCA CAG CAA AAC GAC CA	216 bp	60.02
	MsEF-R	GTT ATG CCC CTG TGC TTG AT		59.96

2.6 Real-time quantitative PCR

Real-time qPCR was utilized to quantify the amount of a targeted genomic DNA according to modified method described by Mounsey and co-workers [5]. The quantity of amplified product monitored in qPCR was achieved via calculating the fluorescence signal intensities upon integration into DNA strands and are proportional to the amount of amplified PCR. The signal curves that are produced by the standards and target samples, in the same run, were then used to quantify the target DNA.

Real-time qPCR experiment was conducted using Rotor-Gene Q Cycler (Qiagen, USA). A PCR total volume of 25 µl reaction mixture was prepared containing 12.5 µl of 2X SYBR green master mix (Applied Biosystems, USA), 1.0 µl 10 µM of each primer, and 1.0 µl template DNA. The PCR cycle conditions include an initial denaturation-step at 95°C for 3 min followed by 40 cycles of denaturation-step at 95°C for 15 s, annealing-step at 60°C for 20 s and extension-step at 72°C for 30 s. Two reaction runs were conducted with each run consisted of a series of linearized plasmid standards (six, 10-fold dilutions of standard DNA as template), genomic DNA and a no-template control. After the PCR extension step, SYBR Green I fluorescence was calculated and a melting curve analysis was measured. The melting curve confirms the amplification of specific product and was achieved with a temperature gradient of 0.2 Ks^{-1} from 62°C to 95°C.

A DNA weight-to-moles calculator was used to calculate the copy number (Practical Molecular Biology, http://molbiol.edu.ru/eng/scripts/h01_07.html). This calculation determined the number of template concentration and size. To calculate the copy number of targeted gDNA, a standard curve was generated using the C_T values for plasmid amplifications. In order to estimate the genome size, only qPCR runs with standard curve efficiency of more than 90% and duplicates of less than 0.5 C_T standard deviation were used. The C value of a single copy gene of a nuclear genome (picograms) was calculated by dividing the input of the template concentration by the qPCR-derived copy number [5]. The genome size of the unknown gDNA was estimated using the genome size formula described by Mounsey *et al*. [5] and Dolezel *et al*. [15]:

$$\text{Genome size (bp)}, \Gamma = (0.978 \times 10^9) \times C \text{ (pg)}$$

Where, the mean weight of one nucleotide base pair is 1.023×10^{-9} pg.

3. Results and Discussion

The technique used in this work requires the calculation of the absolute amount of single-copy genes present in a DNA sample that is then used to estimate the C values and Γ. To achieve this aim, DNA dilution with a known concentration of a standard (single-copy and gene specific) was required. Two pairs of primers for *M. sagu* single copy genes of beta-actin and translation elongation factor (EF3) were designed and used to amplify the single-copy genes. The PCR products were cloned into a cloning vector, products were confirmed via sequencing and the plasmids were linearized by restriction enzyme *NcoI*. The linearized DNAs were used as standard DNAs in real-time PCR to determine the copy number of genomic DNA present in genomic DNA samples. To validate the protocol, positive control real-time PCR using the beta-actin and elongation factor (EF3) primers was carried out to estimate *P. pastoris* genome size.

An amplification curve generated for elongation factor gene sequence quantification of *P. pastoris* is shown in Figure 1, where fluorescence signals from SYBR Green I were measured at 520 nm. The different color curves are the standard PCR amplified from 10^6 to 10^1 copies of the genes, while the curves for the targeted genomic sample, measured six times, are indicated by arrows. The red curve represents the amplification profile of the no-template control (NTC). A C_T value calibration curve of the standard versus the concentrations in copies per microliter, was plotted using Q-Rex software and used to determine the target copy numbers based on their fractional cycle number or C_T values (Figure 2), where the values are proportional to the log of initial target concentrations. To determine the concentrations of targeted gDNA, a calibration curve of the C_T values of the standard dilution series versus the concentrations was then calculated [5, 14].

Melting curve analysis was done by slowly increasing the PCR amplification temperature from 72° to 90°C. The signal was continuously recorded to ensure that non-specific products were not amplified before the signal curves reached the threshold. The cooperative melting process of the dsDNA causes an abrupt decrease in fluorescence signal that can be seen as a clear peak in the negative derivative (±dF/dT) of the melting curves. The melting curve analysis of the specific PCR product showed that the standard melted at approximate 85°C (Figure 3). The same melting temperature of 85°C was also observed for the PCR product of the targeted genomic DNA, therefore confirming the specificity of the amplification that gave rise to the same PCR product.

The methanotrophic yeast *P. pastoris* strain GS115 was chosen as the positive control for the method due to the availability of its complete genome sequence which was reported to be 9.43 Mbp [16] and 8.7 Mbp using similar technique [5]. Comparatively via real-time qPCR, the genome size of *P. pastoris* was estimated to be 8.52 Mbp (Table 2) which was within 10% of the size reported by De Schutter *et al.* [16] and closer to the size by Mounsey *et al.* [5]. Quantitative real-time PCR of two single copy genes of *M. sagu* gave a mean genome estimate of 1.87 Gbp for *M. sagu* (Table 2) with no large differences in genome size estimated using either of the two specific primers. Assuming the 10% upper error limit for the genome size of *P. pastoris*, the actual genome size of *M. sagu* is therefore estimated to be 2.06 Gbp.

This estimated genome size of *M. sagu* is very much larger than that of any other economically important plant of known genome size such as agarwood (*Aquilaria malaccensis*) which range from 894.65 to 938.88 Mbp [17]. The genome size for the Para rubber tree (*Hevea brasiliensis*) has been estimated to range from 1.37 to 1.47 Gbp [18] and the genome size for oil palm (*Elaeis guineensis*) closely related to *M. sagu* and from the same family Arecaceae was estimated at the 1.8 Gbp [19]. However, genome size across angiosperms has been revealed to be

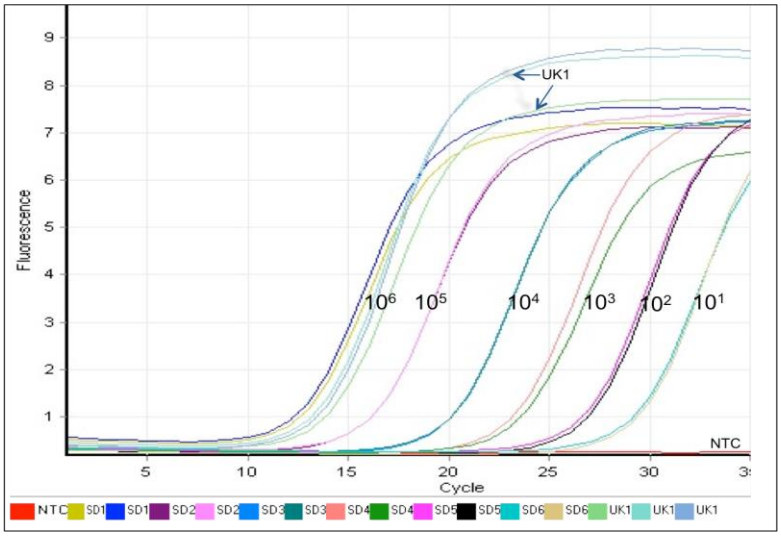

Figure 1. Signal curves obtained from real-time PCR. An amplification curve used for quantification of elongation factor gene from *P. pastoris*. Fluorescence signals from SYBR Green 1 were measured at 520 nm. Curves obtained for the standards with 10^6, 10^5, 10^4, 10^3, 10^2 copies of the standards PCR products as templates are shown in different colors (SD1 to SD6, each duplicate samples). Samples curves are shown by black arrows (UK1, triplicate samples); NTC indicates the no-template control.

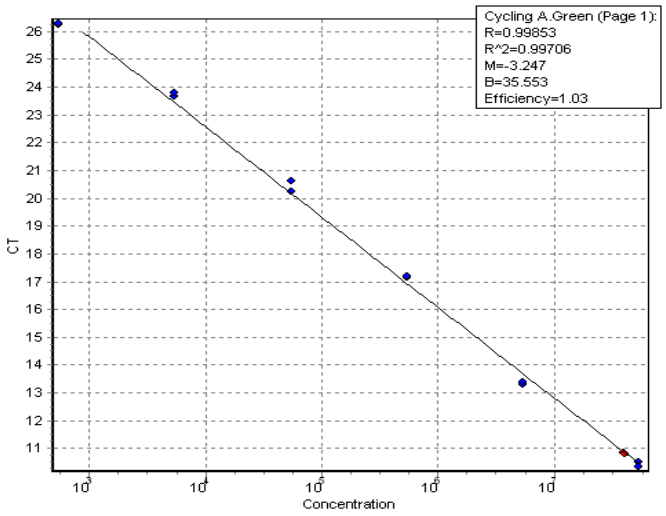

Figure 2. C_T values of serially 10-fold diluted standard of known concentrations (copy number) of elongation factors (EF3) gene sequence and unknown genomic samples of *P. pastoris.* Standard and unknown samples C_T values are indicated by blue and red colors, respectively.

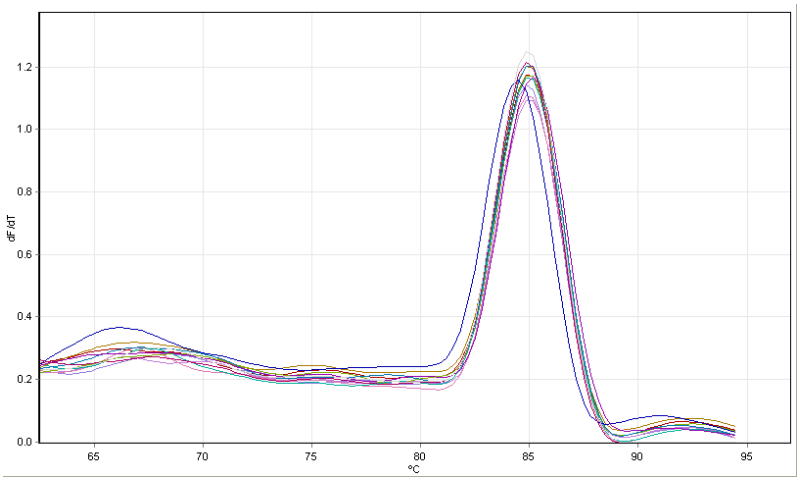

Figure 3. Melting curve analysis of standard and samples after amplification with the PpEF3 specific primers (Amplification curve is shown in Figure 1). Melting curve analysis was carried directly after the PCR by slowly increasing the temperature at 0.2 Ks^{-1} from 72 to 90°C while the signal was taken continuously. Cooperative melting process of the dsDNA causes a steep decrease in the fluorescence signal around the melting temperature of the PCR products. The temperature where the top value is reached is considered as the melting temperature: the specific PCR product obtained from the amplification of the standard PCR product melts at 85°C. Same melting temperature was also determined for the PCR product of the genomic DNA sample.

Table 2. Quantitative real time PCR based genome size estimates of *P. pastoris* and *M. sagu*

Species	**Sample conc.** (ng/µl)	**Gene**	**qPCR products size** (bp)	**Target copies/µl**	**C** (Mean, ±SD, pg)	**Genome Size, Γ** (Mean, ±SD)	**Average Genome size** (Γ)
P. pastoris	134	Beta-actin	246	$15.5x10^6$	0.0086 ± 0.0006	8.45 (± 0.5) Mbp	8.52 (± 0.42) Mbp
	254	Elongation factor 3	246	$28.81x10^6$	0.0088 ± 0.0004	8.62 (± 0.4) Mbp	
M. sagu	151	Beta-actin	185	$7.95x10^4$	1.89 ± 0.11	1.85 ± 0.09 Gbp	1.87 (± 0.08) Gbp
	160	Elongation factor	209	$8.20x10^4$	1.95 ± 0.12	1.90 ± 0.08 Gbp	

diverse with differences in C value ranging up to 2000-fold [20]. The genome size estimated here for *M. sagu* of 1.87 Gbp, is therefore still considered small compared to the monocot *Trillium hagae* (*Melanthiaceae*) that is known to have the largest genome size (129.54 Gbp) among all angiosperms analyzed to date [21]. The relatively small size of *M. sagu* genome nevertheless

allows a fair challenge to its genome sequence project. Although flow cytometry remains the most ideal method for determination of genome size, this method is not available to many investigators, and the PCR method we use here is demonstrated to be quite reliable. Up until now, there has been no reported analyses of *M. sagu* genome size. Therefore, this is the first report for genome size estimation for *M. sagu*, which will be helpful for future *M. sagu* genome sequencing and subsequent molecular analyses projects.

4. Conclusions

Quantitative real time PCR is an invaluable tool that could be used in many applications including in evaluating the genome size of an underutilized crop plant, *Metroxylon sagu*. The qPCR technique indicated that the genome size of *M. sagu* is determined to be 1.87 Gbp. With the determination of genome size of *M. sagu*, this would enable other molecular research of the species, such as genome sequencing, and further drive molecular analysis of this important but underutilized crop plant.

5. Acknowledgements

This study was funded by grants derived from UNIMAS (F07/DPD01/1119/2014(01)) and partially funded by the Malaysia Ministry of Higher Education (FRGS/2/2013/ST03/UNIMAS/02/2). The authors would also like to thank the Faculty of Resource Science and Technology (UNIMAS) for providing the research facilities.

References

[1] Wee, C.C. and Roslan, H.A., 2012. Expressed sequence tags (ESTs) from young leaves of *Metroxylon sagu*. *3 Biotech*, 2, 211-218.

[2] Hare, E.E. and Johnston, J.S., 2011. Genome size determination using flow cytometry of propidium iodide-stained nuclei. *Methods in Molecular Biology*, 772, 3-12.

[3] Zhang, J.Z. and Fan M.Y., 2002. Determination of genome size and restriction fragment length polymorphism of four Chinese rickettsial isolates by pulsed-field gel electrophoresis. *Acta Virologica*, 46, 25-30.

[4] D'Hondt, L., Hofte, M., Van Bockstaele, E. and Leus, L., 2011. Applications of flow cytometry in plant pathology for genome size determination, detection and physiological status. *Molecular Plant Pathology*, 12, 815-828.

[5] Mounsey, K.E., Willis, C., Burgess, S.T.G., Holt, D.C., McCarthy, J. and Fischer, K., 2012. Quantitative PCR-based genome size estimation of the astigmatid mites *Sarcoptes scabiei, Psoroptes ovis* and *Dermatophagoides pteronyssinus. Parasites & Vectors*, 5, 3. https://doi.org/10.1186/1756-3305-5-3.

[6] Henk, D.A. and Fisher, M.C., 2012. The gut fungus *Basidiobolus ranarum* has a large genome and different copy numbers of putatively functionally redundant elongation factor genes. *Plos One*; 7 (2), e31268. https://doi.org/10.1371/journal.pone.0031268

[7] Armaleo, D. and May, S., 2009. Sizing the fungal and algal genomes of the lichen *Cladonia grayi* through quantitative PCR. *Symbiosis*, 49, 43, https://doi.org/10.1007/s13199-009-0012-3

[8] Gao, J. and Scott, J.G., 2006. Use of quantitative real-time polymerase chain reaction to estimate the size of the house-fly *Musca domestica* genome. *Insect Molecular Biology*, 15, 835-837.
[9] Tsoumani, K.T. and Mathiopoulos, K.D., 2012. Genome size estimation with quantitative real-time PCR in two Tephritida species: *Ceratitis capitata* and *Bactrocera oleae*. *Journal of Applied Entomology*, 136, 626-631.
[10] Kim, J.H., Roh, J.Y., Kwon, D.H., Kim, Y.H., Yoon, K.A., Yoo, S., Noh, S.J., Park, J.H., Shin, E.H., Park, M.Y. and Lee, S.H., 2014. Estimation of the genome sizes of the chigger mites *Leptotrombidium pallidum* and *Leptotrombidium scutellare* based on quantitative PCR and k-mer analysis. *Parasites & Vectors*, 7, 279. https://doi.org/10.1186/1756-3305-7-279.
[11] Park, B. and Kim, Y., 2012. Genome size estimation of an endoparasitoid wasp, *Cotesia plutellae*, using quantitative real-time polymerase chain reaction. *Journal of Asia-Pacific Entomology*, 15, 349-353.
[12] Kaewkong, W., Imtawil, K., Maleewong, W., Intapan, P.M., Sri-Aroon, P., Wongkham, S. and Wongkham, C., 2012. Genome size estimation of liver fluke *Opisthorchis viverrini* by real-time polymerase chain reaction based method. *Parasitology International*, 61, 77-80.
[13] Jensen, M.K., Vogt, J.K., Bressendorff, S., Seguin-Orlando, A., Petersen, M., Sicheritz-Pontén, T. and Mundy, J., 2015. Transcriptome and genome size analysis of the Venus flytrap. *Plos One*, 10: e0123887. https://doi.org/10.1371/journal.pone.0123887
[14] Wilhelm, J., Pingoud, A. and Hahn, M., 2003. Real-time PCR-based method for the estimation of genome sizes. *Nucleic Acids Research*, 31: e56. https://doi.org/10.1093/ nar/gng056
[15] Dolezel, J., Bartos, J., Voglmayr, H. and Greilhuber, J., 2003. Nuclear DNA content and genome size of trout and human. *Cytometry. Part A ,* 51, 127-128.
[16] De Schutter, K., Lin, Y-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouzé, P., Van de Peer, Y. and Callewaert, N., 2009. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology*, 27, 561-566.
[17] Siti Suhaila, A.R., Mohd Saleh, N., Norwati, M., Mahani, M.C., Namasivayam, P. and Kandasamy, K.I., 2018. DNA content and genome size of highly valued malaysian agarwood, *Aquilaria malaccensis* LAMK. *Malaysian Applied Biology*, 47, 13-21.
[18] Tang, C., Yang, M., Fang, Y., Luo, Y., Gao, S., Xiao, X., An, Z., Zhou, B., Zhang, B., Tan, X., Yeang, H.Y., Qin, Y., Yang, J., Lin, Q., Mei, H., Montoro, P., Long, X., Qi, J., Hua, Y., He, Z., Sun, M., Li, W., Zeng, X., Cheng, H., Liu, Y., Yang, J., Tian, W., Zhuang, N., Zeng, R., Li, D., He, P., Li, Z., Zou, Z., Li, S., Li, C., Wang, J., Wei, D., Lai, Z.Q., Luo, W., Yu, J., Hu, S. and Huang, H., 2016. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature plants*, 2, 16073 https://doi: 10.1038/nplants.2016.73.
[19] Singh, R., Ong-Abdullah, M., Low, E-L., Manaf, M.A., Rosli, R., Nookiah, R., Ooi, L.C., Ooi, S.E., Chan, K.L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A., Sambanthamurthi, R., 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, 500, 335-339.
[20] Pellicer, J., Fay, M.F. and Leitch, I.J., 2010. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* 164, 10-15.
[21] Zonneveld, B.J.M., 2010. New record holders for maximum genome sizes for monocots and eudicots. *Journal of Botany*, 2010, article ID 527357. https://doi.org/ 10.1155/2010/527357.