Jurnal Teknologi

Full paper

Phylogenetic Analysis of Eight Malaysian Pineapple Cultivars using a Chloroplastic Marker (rbc*L* gene)

Norfadilah Hamdan[a], Azman Abd Samad[a], Topik Hidayat[b*], Faezah Mohd Salleh[a]

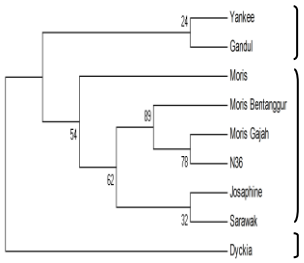
[a]*Department of Biotechnology and Medical Engineering, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*
[b]*Department of Biology Education, Indonesia University of Education (UPI), Jalan Dr Setiabudi, 229 Bandung, 40154, Indonesia*

*Corresponding author: topik28@yahoo.com

Article history

Received :29 July 2013
Received in revised form :
23 September 2013
Accepted :29 September 2013

Graphical abstract



Abstract

To date, Malaysian pineapple cultivars has only been characterized morphologically. A more consistent and accurate method such as biomarker is highly crucial to distinguish and establish the genetic relationship between different cultivars. In this work, we conducted a phylogenetic analysis of eight Malaysian pineapple cultivars using a chloroplastic DNA biomarker, ribulose-bisphosphate carboxylase (rbc*L*) gene. The rbc*L* gene was isolated from genomic DNA, amplified and sequenced. The rbc*L* gene of *Ananas comosus* is approximately 1100 bp. From the multialignment of eight cultivars, the percentage of sequence similarity ranged from 71.1% to 94.98% and is highly conserved throughout the sequences. Phylogenetic analysis which is carried out using maximum parsimony method revealed that the eight Malaysian pineapple cultivars can be classified into two groups. The first group consist of Yankee and Gandul cultivars while Moris, Moris Bentanggur, Moris Gajah, N36, Josaphine and Sarawak falls under the second group. Bootstrap values in some branches are low which reflect the small number of informative characters (981 are conserved, 12 are potentially informative). Formation of several group or subclades is due to its similar genetic pattern, thus supporting this classification. This study confirmed that rbc*L* gene is a good indicator to determine the phylogenetic relationship distinguishing the Malaysian pineapple cultivars.

Keywords: *A. comosus;* Malaysian pineapple; rbc*L*; phylogenetic; maximum parsimony

Abstrak

Sehingga kini, kultivar nanas Malaysia dikelaskan berdasarkan morfologi sahaja. Kaedah yang lebih konsisten dan lebih tepat seperti penanda biologi sangat penting untuk membezakan dan menentukan hubungan genetik antara kultivar yang berbeza. Dalam kertas kerja ini, kami telah menjalankan analisis filogenetik lapan kultivar nanas Malaysia menggunakan DNA penanda biologi, gen kloroplastik *ribulose-bisphosphate carboxylase* (rbc*L*). Gen rbc*L* telah diekstrak daripada DNA genomik, diamplifikasi dan dijujukan. Gen rbc*L* daripada *Ananas comosus* adalah kira-kira 1100 bp. Daripada jujukan terjajar lapan kultivar, peratus persamaan antara jujukan adalah dari 71.1% hingga 94.98% dan sangat abadi *(conserved)* di sepanjang jujukan. Analisis filogenetik yang dijalankan menggunakan kaedah *Maximum parsimony* menunjukkan bahawa lapan kultivar nanas Malaysia boleh dikelaskan kepada dua kumpulan. Kumpulan pertama terdiri daripada kultivar Yankee dan Gandul manakala kultivar Moris, Moris Bentanggur, Moris Gajah, N36, Josaphine dan Sarawak termasuk di dalam kumpulan kedua. Nilai *Bootstrap* di beberapa cabang yang rendah menunjukan nilai sifat yang berinformasi adalah kecil (981 abadi, 12 berinformatif). Pembentukan beberapa kumpulan atau subkumpulan adalah disebabkan oleh komposisi genetik yang sama, sekali gus menyokong klasifikasi ini. Kajian ini mengesahkan bahawa gen rbc*L* adalah penunjuk yang baik untuk menentukan hubungan filogenetik untuk membezakan kultivar nanas Malaysia.

Kata kunci: *A. comosus;* nanas Malaysia; rbc*L*; filogenetik; parsimony maksimum

1.0 INTRODUCTION

Pineapple (*Ananas comosus*) is a type of tropical plant which is believed to originate from East Area of South America (Malaysian Pineapple Industry Board, MPIB, 2012). It was introduced in Malaya in the 16th century by the Portuguese. In 1921, pineapple started to be planted in Singapore and several states in Malaya as cash crop. Currently, pineapple plantation has been expanded in peat soil area especially in Johor, the southeast state of Peninsular Malaysia (Malaysian Pineapple Industry Board, 2012). In Malaysia, the three most common cultivars widely planted are Spanish (also known as Maspine, Josapine and Hybrid pineapple), Smooth Cayenne (Sarawak pineapple) and Queen (Morris pineapple) cultivars (Malaysian Pineapple Board, 2012).

In the past, most of the classification of pineapple cultivars is based on morphological characteristics, including the Malaysian cultivars. The parameters that researcher usually used are fruit size, fruit shape, fruit colour, leaf shape, leaf colour and size of the crown. The drawback of classification based on morphological characters is inconsistency that arose due to disagreements among morphologist who applied different classification for interpretation of the characteristics. Thus, a more accurate approach by using molecular data has gained considerable interest among the morphologist.

Molecular approach has offered effective method in addressing many phylogenetic questions which cannot be solved using morphology characters. Previous studies (Schulte *et al.*, 2008; Sheng-Guo *et al.*, 2008; Spreitzer *et al.*, 2002; Osaloo and Kawano, 1999; Chase *et al.*, 1993) claimed that large subunit of the ribulose- bisphosphate carboxylase (rbc*L*) gene is suitable for inference phylogenetic relationship at higher taxonomic levels. The rbc*L* gene is usually up to 1200 bp in size and the use of this gene in phylogenetic analysis has been reviewed in many studies (Spreitzer *et al.*, 2002; Chase *et al.*, 1993; Clegg, 1993). This is due to its advantages where the gene is highly conserved and thus have slow evolution rate in comparison with nuclear gene (Wolfe *et al.*, 1987). An important consequence of a conservative rate of nucleotide substitution is that the synthetic oligonucleotide primers can be designed for the purposes of PCR amplification which are nearly universally applicable to all angiosperm plant taxa. This is the huge advantage for evolutionary research because the molecular data from the study of one species (for example maize) can be applied to other distantly related species (like spinach) which make it suitable to study plant phylogeny (Clegg, 1993).

To date, Malaysian pineapple cultivars classification is limited to morphological characteristics, and information for molecular classification is scarce. Thus, in this work, we constructed a systematic phylogenetic tree of eight Malaysian pineapple cultivars *(Ananas comosus*) using the chloroplastic rbc*L* gene as the biomarker. The analysis of the sequences highlights the genetic variability between these cultivars and promotes further novel information regarding its evolutionary relationships.

2.0 EXPERIMENTAL

2.1 Cultivar selection and Sampling

Leaf samples from eight Malaysian pineapple cultivars; Gandul, N36, Moris, Moris Gajah, Moris Bentanggur, Josaphine, Sarawak Green Local, Masphine and Yan Kee were obtained from Malaysian Pineapple Board Industry (MPIB), Johor. The cultivars were selected based on their economically important role in Malaysia (Malaysian Pineapple Board Industry, 2012).

2.2 DNA Analysis

Young leaves were ground using mortar and pestle with the aid of liquid nitrogen. Genomic DNA was isolated from leaf samples using QIAGEN Dneasy Mini Plant Kit as described in Hidayat *et al.*, (2011). The polymerase chain reaction (PCR) was carried out by using rbc*L*-F (CTT GGC AGC ATT CCG AGT A) primer and rbc*L*-R (TCA CAA GCA GCA GCC AGT TC) primer. PCR reaction was performed in a 50 µL reaction volume using MyCycler Thermal Cycler System (Bio- Rad). The reaction component were prepared on ice and each of the PCR tubes contain 10x Standard *Taq* Reaction buffer (10 mM Tris- HCl, 50 mM KCl, 1.5 mM $MgCl_2$), DNA template, 10 µM forward primer (rbc*L*- F), 10 µM reverse primer (rbc*L*- R), 10 µM dNTPs mix, enzyme Taq DNA polymerase (New England BioLabs) and distilled water. PCR reaction is as follows: 1 cycle at 94°C (predenaturation) for 2 minutes; 30 cycles at 94°C (denaturation) for 1 minute, 49°C (annealing) for 1 minute, and 72°C (extension) for 2 minutes; and ended with 1 cycle at 72°C (final extension) for 5 minutes. All amplified products were cloned into pGEM-T Easy (Promega) and transformed into NEB5α competent cell (New England BioLabs). Plasmids were extracted using QIAprep® Spin Miniprep Kit (Qiagen) according to QIAprep® Miniprep Handbook with slight modification. The presence of the insert was confirmed with plasmid digestion using *Eco*R1 (Promega), before sending for sequencing (1st Base Laboratories, Malaysia).

2.3 Bioinformatics Analysis

All the DNA sequences obtained from rbc*L* gene were aligned using ClustalX software and the sequences were saved as FASTA format. Both head and tail of the sequences were edited using BIOEDIT software. The aligned sequences which comprised of eight samples with an outgroup were subjected to phylogenetic analysis. Phylogenetic tree was constructed using MEGA 5 software. Phylogenetic analyses were done via Maximum Parsimony approach. A species called *Dyckia* from subfamily Pitcairnioideae was designed as the outgroup as it has been recognized as the closest relatives of the Bromeliaceae, the *Ananas comosus* family (Duval *et al.*, 2003).

3.0 RESULTS AND DISCUSSION

In this study, young leaves were used as DNA source in order to minimize contamination that can inhibit amplification. The amplified rbc*L* gene was approximately 1.1 kb and confirmed via sequencing. Multiple sequence alignment analysis revealed that the percentage of sequence similarity ranged from 71.1% to 94.98% and is highly conserved throughout the sequences (Figure 1). Phylogenetic tree was constructed using Maximum Parsimony (MP) method and bootstraps reliability tests with 1000 replicates. Insertion and deletions were treated as missing data. Based on the analysis using MEGA 5, it was found that the aligned rbc*L* comprised of about 1094 characters. Out of these characters, 981 characters were found to be conserved and only 12 characters were potentially informative. A site is parsimony-informative if it contains at least two types of nucleotides, and at least two of them occur with a minimum frequency of two.

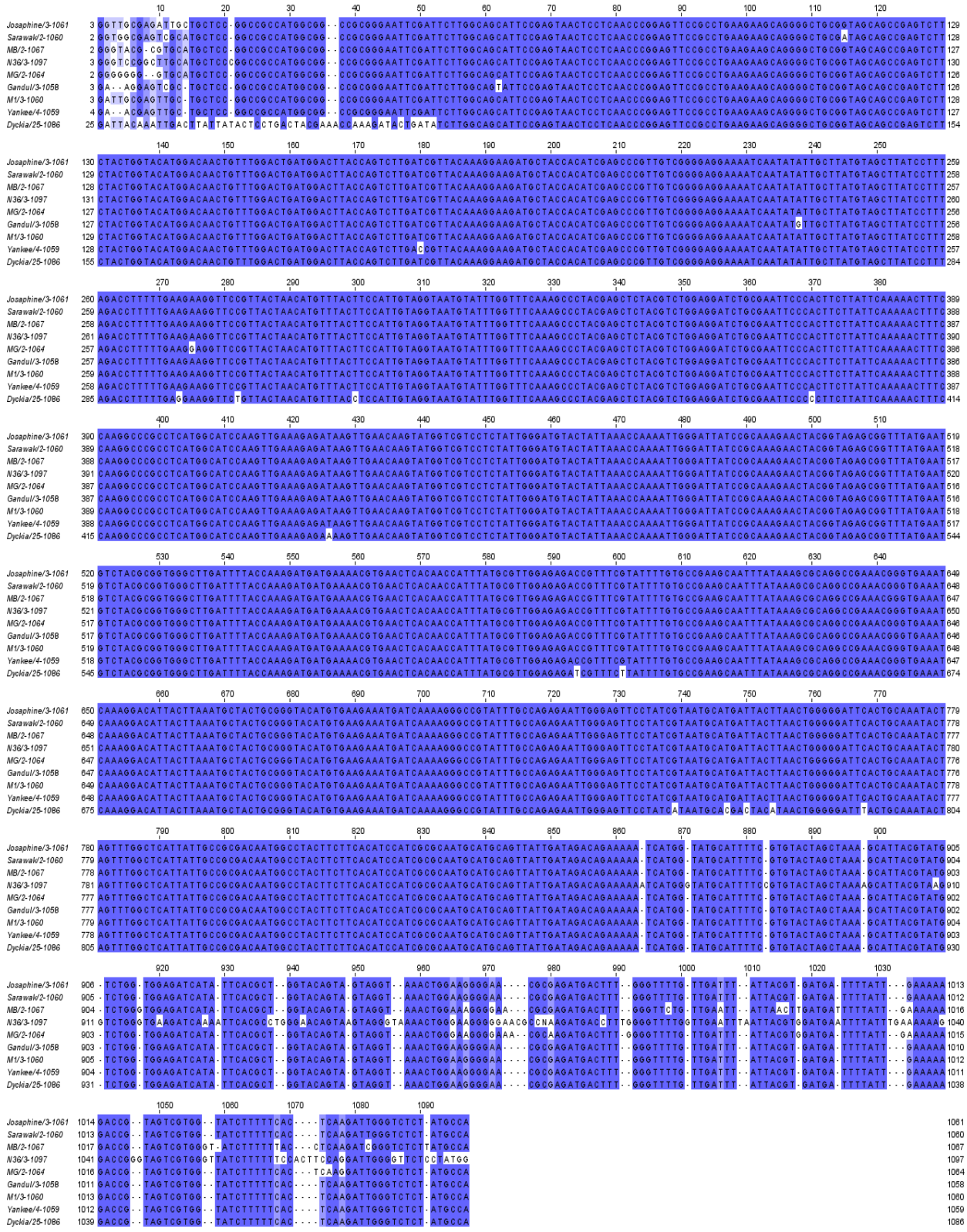


Figure 1 Multiple sequence alignment of the eight sequences of malaysian pineapple cultivars. conserved region are highlighted in dark blue

The consensus tree based on these nine constructed tree is shown in Figure 2. The consensus tree was constructed with a consistency index (CI) of 0.948 and a retention index (RI) of 0.789. The CI serves to measure the relative amount of homoplasy on the tree. Homoplasy occurs when the characters are similar but they are not derived from common ancestor. The CI values will equivalent to 1 if the characters reveal no homoplasy. On the other hand, RI measures the proportion of

synapomorphy on the tree which means it measures the proportion of similarities on the tree. The RI of 1 indicates that it shows no homoplasy which means the characters are completely consistent with phylogeny. Meanwhile, RI of 0 indicates proable maximum amount of homoplasy on the tree (Nei and Kumar, 2000).

Based on the constructed rbc*L* phylogeny tree, eight cultivars were clearly separated into two major Clades (Figure 2). Yan Kee and Gandul were separated into first clades with 24% bootstrap support (BS) while Moris, Moris Gajah, Moris Bentanggur, N36, Josaphine and Sarawak were separated to form the second clades with 54% bootstrap support (BS). It also revealed that Josaphine is a sister to Sarawak cultivar although both are derived from different group where Josapahine is from Spanish cultivar and Sarawak is from Cayenne cultivar. The presence of these two cultivars in one clade might be due to the close relationship between them because the Josaphine cultivar is a hybrid selected from a cross between Johor (Spanish) and Sarawak (Smooth Cayenne) (Chan *et al.*, 2003). On the other hand, it can be seen that the Moris Bentanggur (MR Bentanggur) and Moris Gajah (MR Gajah) cultivars are grouped together in subclade of second clades. This result supports the hypothesis that both of them are derived from the same group (Malaysian Pineapple Industry Board, 2012). The BS values for these two cultivars were relatively high in the phylogenetic tree shown in Figure 2 which is 89. Since both of them are derived from the same group, they shared many similarities in their morphologies such as dark green skin when the fruits ripen, long and closely arranged spine, acuminate leaf shape and rough leaf surface and tapered shape fruit with projected eyes (Malaysian Pineapple Industry Board, 2011)

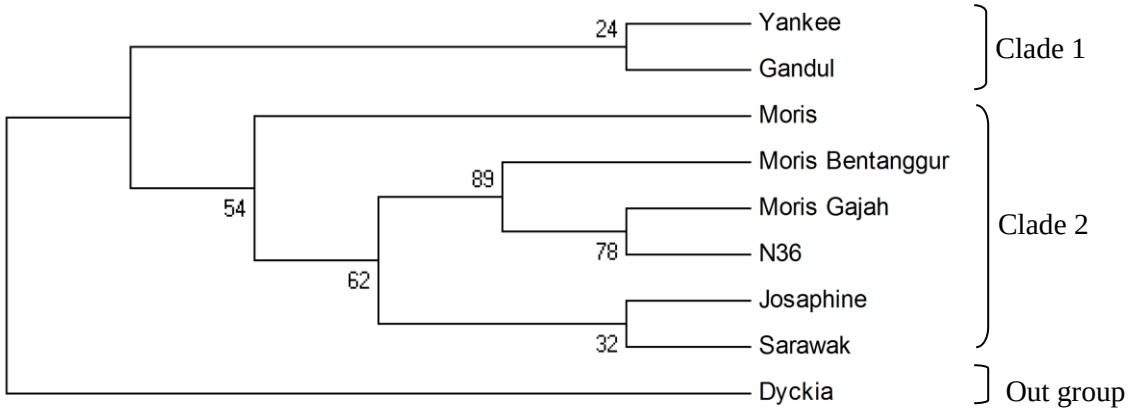


Figure 2 One of the nine Maximum Parsimony tree(s) of the Malaysian pineapple based on rbc*L* gene (Length = 78 steps, CI = 0.948, RI = 0.789). The value in the branch shows the bootstraps support (BS)

On the other hand, the phylogenetic tree also shows that N36 cultivar were separated from Josapahine and Sarawak cultivars although both Josaphine and N36 are hybrid from Sarawak cultivar. Josaphine and N36 were closely related to Spanish group than Cayenne group. As mentioned earlier, Josaphine (Spanish) cultivar is hybrid selected from a cross between Johor (Spanish) and Sarawak (Smooth Cayenne) while N36 cultivar is derived from a hybrid selected from a cross between Gandul (Spanish) and Sarawak (Smooth Cayenne) cultivar (Chan *et al.*, 2003; Hamid, 2005). However, very little morphological similarities are shared by these cultivars which support the different clades grouping in Figure 2. The only similarity shared between N36 and Sarawak cultivar is both have pale yellow pulp colour with acidity content around 0.5 to 0.9% and sugar content about 12 to 16 ° Brix (Chan *et al.*, 2003).

Bootstraps (BS) value is a measure on how consistently the data support the given taxon. Generally, the BS value will reflect the significant of the topology of the branch. If the BS value for a certain clades is close to 100%, then nearly all of the informative characters that build the topology of the branch are considered statistically significant (Neil and Kumar, 2000; Berry and Gascuel, 1996). In this study, sequences from rbc*L* gene were used to construct the tree. As mentioned in the previous study (Chase, 1993; Clegg, 1993), chloroplast gene is highly conserved and the evolutionary rate in the gene is very slow. Out of 1216 characters, 981 were conserved and 12 were potentially informative. It was also shown in the Figure 2 that BS value in some branches was less than 50. This is reflected by the small number of informative characters (only 12 characters were potentially informative) to build the tree. The number of informative characters sites of rbc*L* is low in this study and this is due to the high CI, RI and substitution rates at which the variable sites in the rbc*L* region evolved, indicated a lower level of homoplasy (Hidayat *et al.*, 2005). BS value may increase with increasing taxon sample size. There were evidences from previous studies which claimed that phylogenetic accuracy can be increased with increased taxon sampling (Soltis *et al.*, 1998; Graybeal, 1998; Hillis, 1996).

4.0 CONCLUSION

The study showed that the eight Malaysian pineapple cultivars (Gandul, N36, Moris, Moris Gajah, Moris Bentanggur, Josaphine, Sarawak Green Local, Masphine and Yan Kee) can be classified into two main clades. The result indicates similar genetic pattern between the cultivars and therefore might have a tendency to follow the same evolutionary path. Thus, this work suggested that rbc*L* gene could be used as a good biomarker to distinguish the Malaysian pineapple cultivars and determine their phylogenetic relationship. Although this study is considered to be preliminary, it gives additional information on phylogenetic analysis of Malaysian pineapple cultivars using sequences of rbc*L* as the available data on Malaysian pineapple cultivars is still insufficient. Classification and identification of

pineapple cultivar are important for record documentation purposes and possibly a systematic database development. Additional information on the relationship pattern among the cultivars is important as it can be used as source of knowledge and information for successful interbreeding on creating new cultivars in the future.

Acknowledgement

The author would like to express sincere thanks to the Malaysian Pineapple Industry Board for providing pineapple samples and Faculty of Biosciences and Medical Engineering, UTM Skudai, Malaysia for funding.

References

[1] Botella, J. R., and Smith, M. 2008. Genomics of Pineapple, Crowning The King of Tropical Fruits. In P. H. Moore, and Ming, R. (Ed.), *Genomics of Tropical Crop Plants*. 1: 441–451. New York: Springer.

[2] Berry, V., and O. Gascuel. 1996. On the Interpretation of Bootstrap Trees: Appropriate Threshold of Clade Selection and Induced Gain. *Mol. Biol. Evol.* 13: 999–1011.

[3] Chan, Y. K., Coppens d'eeckenbrugge, G., and Sanewski, G. M. 2003. Chapter 3: Breeding and Variety Improvement In D. P. Bartholomew, Paull, R. E., and Rohrbach, K. G. (Ed.), *The Pineapple: Botany, Production and Uses*. 33–51. Wallingford, UK: CABI Publishing.

[4] Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R., Hills, H. G., Qui, Y. L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K. J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q. Y., Plunkett, G. M., Soltis, P. S., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H., Graham, S., Barrett, S. C. H., Dayanandan, S., and Albert, V. A. 1993. Phylogenetics of Seed Plants: An Analysis of Nucleotide Sequences from the Plastid GenerbcL. *Ann. Mo. Bot. Gard.* 80: 528–580.

[5] Clegg, M. T. 1993. Review: Chloroplast Gene Sequences and the Study of Plant Evolution. *Proc. Natl. Acad. Sci. USA.* 90: 363–367.

[6] Duval, M. F., Buso, G. S. C., Ferreira, F. R., Noyer, J. L., Coppens d'Eeckenbrugge, G., Hamon, P., and Ferreira, M. E. 2003. Relationships in Ananas and Other Related Genera Using Chloroplast DNA Restriction Site Variation. *Genome NRC Research Press*. 46: 990–1004.

[7] Graybeal, A. 1998. Is It Better to Add Taxa or Characters to a Difficult Phylogenetic Problem? *Systematic Biology*. 47: 9–17.

[8] Hamid, M. J. A., and Ali, A. K. 2005. An Assessment of the Impact of Technology on Josaphine Pineapple Grown in Malaysia. Economics and Technology Management Research Centre, MARDI.

[9] Hidayat, T., Yukawa, T., and Ito, M. 2005. Molecular Phylogenetics of Subtribe Aeridine (Orchidaceae): Insights from Plastid matK and Nuclear Ribosomal ITS Sequences. *J. Plant Res*. 118: 271–284.

[10] Hillis, D. M. 1996. Inferring Complex Phylogenetic. *Nature.* 383: 130–131.

[11] Malaysian Pineapple Industry Board. Kultivar Nanas.

[12] Nei, M., and Kumar, S. 2000. Chapter 7: Phylogenetic Inference-Maximum Parsimony Method. In M. Nei, and Kumar, S. (Ed.), *Molecular Evolution and Phylogenetics.* New York: University Press Incorporation. 115–143.

[13] Osaloo, S. K., and Kawano, S. 1999. Molecular systematics of Trilliaceae II. Phylogenetic Analyses of Trillium and its Allies Using Sequences of rbcL and matK Genes of cpDNA and Internal Transcribed Spacers of 18S–26S nrDNA. *Plant Species Biology*. 14: 75–94.

[14] Schulte, K., Barfuss, M. H. J., and Zizka, G. 2008. Phylogeny of Bromelioideae (Bromeliaceae) Inferred from Nuclear and Plastid DNA Loci Reveals the Evolution of the Tank Habit Within the Subfamily. *Molecular Phylogenetics and Evolution.* 51: 327–339.

[15] Sheng-Guo, J., Ke-Ke, H., Jun, W., and Sheng-Li, P. 2008. A Molecular Phylogenetic Study of Huperziaceae Based on Chloroplast rbcL and psbA-trnH Sequences. *Journal of Systematics and Evolution.* 46(2): 213–219.

[16] Soltis, D. E., Soltis, P. S., Mort, M. E., Chase, M. W., Savolainen, V., Hoot, S. B., and Morton, C. M. 1998. Inferring Complex Phylogenies Using Parsimony: An Empirical Approach Using Three Large DNA Data Sets for Angiosperms. *Syst. Biol.* 47: 32–42.