A Study on Nipah (Nypa Fruticans) Genetic Diversity in Malaysia Based on Amplified trnL – trnF of cpDNA

Abdul Muin-Mohammad, Sulong Ibrahim, Jasnizat Saidin, Haramaini Ariffin, and Mohd Faiz Mahamud

***Abstract*—**Nipah (*Nypa fruticans*) is widely use in human everyday life, unfortunately, genetic study of nipah is poorly documented. This study was carried out to evaluate nipah genetic variability in Malaysia using phylogenetic approach. Ninety nipah leaves samples were collected throughout Malaysia. Nipah isolated DNA was amplified using *trn*L – *trn*F (*trn*LF) primers. Nipah phylogenetic tree was constructed using Maximum-Likelihood (ML) and Bayesian method. Both methods indicated that nipah samples divided into three major clade; Clade A, Clade B and Clade A1. Clade A and Clade B was highly supported by bootstrap and credibility value (with 99% and 98% in ML method, 100% and 100% in Bayesian method, respectively). Nipah morphology may evolved from being small in size and length but produced more seed cluster to being big in size and length but produced less seed cluster. Further study should be conducted to verify nipah evolution pattern.

***Keywords*—**Bayesian, Maximum-Likelihood, *Nypa fruticans*, Phylogenetic, *trn*L-*trn*F

I. Introduction

NIPAH (*Nypa fruticans*) is widely known for its versatility in terms of its usage. Nipah widely use in everyday human life including thatching, wall-dwelling, alcohol and vinegar production. Nipah able to produce sap up to 18,165 L per hector per year [1], [2]. Nipah is known to be mangrove palm due to its habitat that strives in mangrove areas [3]. This study was conducted using non-coding amplified *trn*L-*trn*F (*trn*LF) chloroplast genome. *trn*LF contain 390-615 bp intron within *trn*L and about 160-440 bp spacer. *trn*LF had been used to study phylogenetic construction at taxonomic level making this genome an ideal genome for this study [4]. Phylogenetic tree was constructed using Maximum-Likelihood (ML) method [5] via Mega 5.10 and Bayesian method via MrBayes [6]. This study was conducted with the aim to study nipah genetic diversity with attention given to nipah in Malaysia. It was expected that we can attained nipah phylogeographical distribution pattern in Malaysia using genetic information obtained from phylogenetic analysis.

Abdul Muin-Mohammad is with the Institute Oceanography and Environment (INOS), Universiti Malaysia Terengganu (UMT), Malaysia (Phone: +60 129617196; e-mail: amuinm@yahoo.com).

Sulong Ibrahim, Haramaini Ariffin and Mohd Faiz Mahamud are with the INOS, UMT, Malaysia (e-mail: sulong@umt.edu.my).

Jasnizat Saidin is with the Institute Marine Biotechnology, UMT, Malaysia (e-mail: ijaxzt@umt.edu.my).

II. Methodology

A. Sampling Strategy and DNA Extraction to Sequencing

In this study, sampling strategy was design based on three theories; (1) Abundance; (2) Environment differences, and; (3) Isolation by distance [7]. Ninety nipah leaf samples were obtained through out Malaysia with three sampling sites at each state in Peninsular Malaysia and six sampling site at Sabah and Sarawak, each. Ranging distance between one site to another was greater than 1 km. At each site, two sampling point were appointed with the range of more than 3 m, to avoid duplication from same individual sample. In this study, the word individual is referred to terminal shoots or shoots cluster that separated from each other by at least 3 m length regardless of the status of its mother or rhizome mother [8]. Leaf sample was taken for DNA extraction and morphology of the sampled individual was observed.

Nipah leaves were preserved in CTAB (Cetyl Trimethyl Ammonium Bromide) preservation [9]. DNA was extracted using DNeasy Plant Mini Kit (Qiagen, 69104) [4] followed as per manufacturer protocol with little modification and quantified using SmartSpec Plus Spectrophotometer (Bio-Rad) [10]. PCR (Polymerase Chain Reaction) was performed using GoTaq Flexi DNA Polymerase (Promega, M8295) with two manufactured primers; c (F) 5'-CGA AAT CGG TAG ACG CTA CG-3' and f (R) 5'-TTT GAA CTG GTG ACA CGA G-3' [4]. PCR cycled was programmed as follow: initial denatured step 1 cycled for 3 minutes at 97°C, 30 cycled of 97°C for 1 minute, 53°C for 1 minute and 72°C for 3 minutes. The final extension was set at 72°C for 7 minutes. Sample was soak at 4°C after final extension [4]. Purified PCR products were run on agarose gel to estimate the band length [10] prior purification using QIAquick PCR Purification Kit (Qiagen, 28104) followed as per manufacturer protocol with little modification. Purified nipah DNA samples were sent to FirstBase for sequencing.

B. Phylogenetic Tree Construction

Nipah amplified sequences were manually edited using CodonCode Aligner. *Typha angustifolia* isolate TA_Ch02 insertion number JF 319452.1 was used as an outgroup and *Nypa fruticans* chloroplast DNA for *trn*L–*trn*F intergenic

spacer insertion number AJ241282.1 was used as reference sequence. Both sequences obtained from NCBI (National Center for Biotechnology Information). Multiple alignment MUSCLE was used to align the amplified sequences via Mega 5.10. Nucleotide substitution analysis using Mega 5.10 showed that best model to use for nipah amplified genome was T92 (Tamura 3-Parameter) with uniform rate. Phylogenetic tree was constructed using two methods; (1) Maximum-likelihood (ML) method via Mega 5.10, and; (2) Bayesian method via MrBayes. ML method was initiated using 2,000 bootstrap replications, T92 model with uniform rates was assumed among sites, gaps/missing data was treated with partial deletion at SPR level 5 (Subtree-Pruning-Regrafting-Extension). MrBayes was run using mixed NST (number of substitution types), four by four nucleotide model with rates among sites was treated equally and number of generation was set at 2,000,000 with sample was recorded every 100 generations. MrBayes MCMC (Markov Chain Monte Carlo) reaction was monitored using TRACER. All parameters were ensured to achieve ESS value >100.

C. Nipah Ultimate Tree Construction and Phylogeographical Distribution

Originally obtained phylogenetic trees from two construction methods were analyzed and cross-referenced with each other to find similarity and phylogeographical distribution possibility pattern. Since there were no obvious pattern showed by neither of the approaches, a newly develop mode of strategy was engaged to suit the need for this study. The only obvious similarity between these two methods were both phylogenetic trees showed nipah samples were divided into 3 major clades, namely; Clade A, Clade A1 and Clade B. To achieve phylogeographical distribution possibility pattern, appearing frequencies evaluation was commenced. Appearing frequency for sample to appear in representing clade represented by 1 for each time were labeled for each sample. Appearing frequency percentage was count. Sample with more than 50% of appearing frequency in representing clade was labeled as absolute sample. For sample with 50% of appearing frequency, sample position was decided based on significant relevant characteristic of possible morphology (Fig. 1).

III. Results And Discussions

In nucleotide substitution model analysis via Mega 5.10, *trn*LF nipah amplified sequences exhibited transition/ transversion rate occurs at 1.51 ratios and nucleotide frequencies of f(A) = f(T) = 0.342 > f(C) = f(G) = 0.158 with rates of base substitution for each nucleotide pair of [r(CT) = r(GA) = 0.218] > [r(AG) = r(TC) = 0.100] > [r(AT) = r(TA) = r(CA) = r(GT) = 0.062] > [r(AC) = r(TG) = r(CG) = r(GC) = 0.029], while in MrBayes analysis, r(AG) = 0.3161 > r(CT) = 0.3142 > r(CG) = 0.1000 > r(GT) = 0.0920 > r(AC) = 0.0892 > r(AT) = 0.0885, where both analysis resolved transition occurs higher in nipah samples than transversion. Both Mega 5.10 and MrBayes agreed that nipah amplified samples were rich in AT bases, which obtained from f(A) = f(T) = 0.342 and pi(A) = 0.334 and pi(T) = 0.341 via Mega 5.10 and MrBayes, respectively. MrBayes exhibited that trnLF amplified taxa were stable in the analysis shown by average standard deviation of split frequencies at 0.004368 with average PSRF (Potential Scale Reduction Factor) = 1.000. Nipah trnLF amplified sampled evidenced the presence of heterogeneity among nipah samples with Arithmetic Total Mean = -2,569.17 and Harmonic Total Mean = -2,651.73. The average ESS (Effective Sample Size) value >100 was observed in TRACER for all MrBayes parameters.

Condensed consensus ML tree and consensus MrBayes tree indicated nipah can be presented in three major clade; Clade A, Clade A1 and Clade B with sixty-five samples belong to Clade A, seven samples belong to Clade A1 and eighteen samples belongs to Clade B, in both trees. ML tree indicated 99%, 63% and 98% of bootstrap value while MrBayes tree indicated 100%, 87% and 100% of credibility value for Clade A, Clade A1 and Clade B, respectively. Since both original tree exhibited similarity between samples that appeared in each clades, thus ultimate cladogram phylogenetic tree was constructed based on similarity exhibited by both original trees.

Nipah in Malaysia came from same ancestor which most possible distributed along Malaysia coastline abundantly during Lower Miocene [11]. Nipah ancestor seeds came to split point where it scattered around Peninsular Malaysia, Sabah and Sarawak. There were high possibilities that nipah ancestor seeds were firstly scattered around Peninsular Malaysia, and then expended its distribution towards Sabah and Sarawak. This condition was clearly described by nipah ultimate phylogenetic tree where nipah Peninsular Malaysia samples were only presented in Clade A and Clade B (Fig. 1). In Sabah, nipah ancestor seeds started to scatter around and evolved. This can be observed from the presented nipah Sabah samples in Clade A, Clade B and Clade A1 of nipah ultimate phylogenetic tree. The appearance of Clade A1 in nipah Sabah samples indicated a delayed evolution, where nipah from Sabah were still in evolution process to a more stable evolution stage. In Sarawak, nipah remain static at Clade A evolution stage, where nipah Sarawak samples only presented in Clade A (Fig. 2).

In this study, presented clade of nipah ultimate phylogenetic tree represented nipah evolution. Nipah evolution was evaluated from nucleotide changes based on the branch length of originally obtained phylogenetic trees. In nipah phylogenetic tree evaluation, nipah strongly indicates the existence of variation within the species which clearly observed from highly supported Clade A and Clade B, while

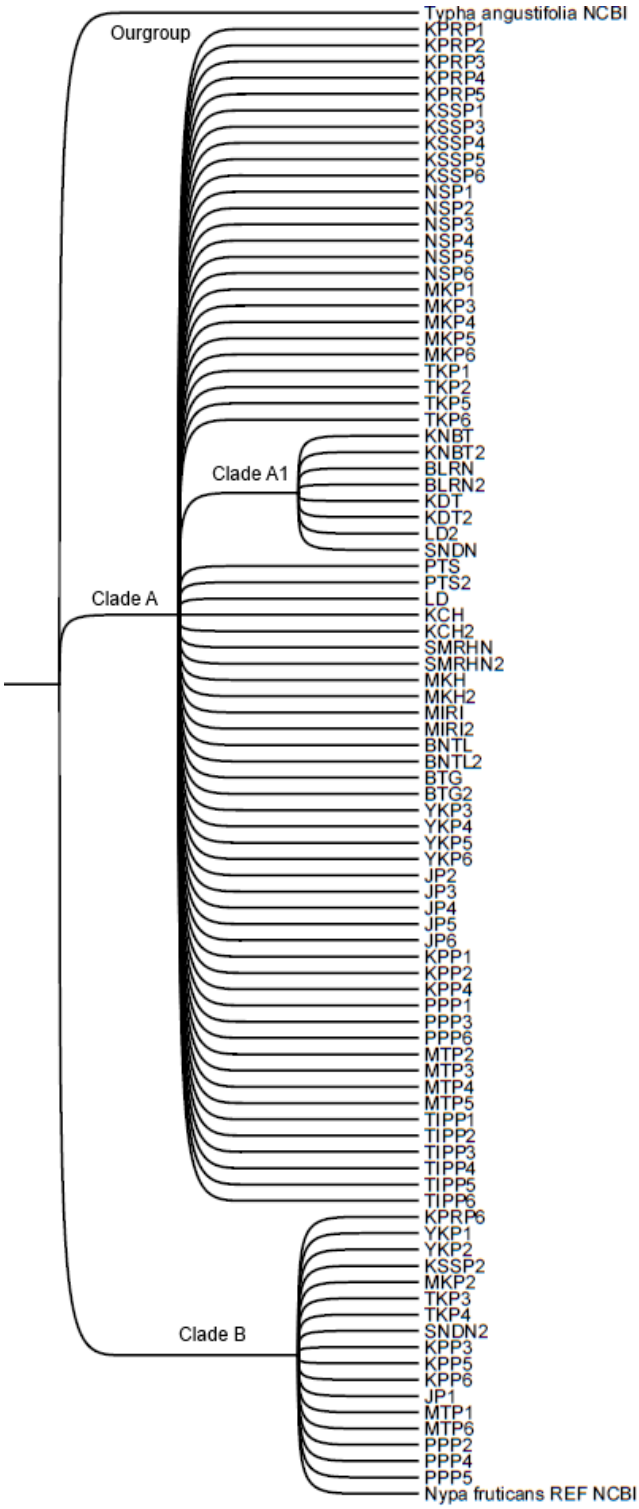

Fig. 1 Nipah ultimate phylogenetic tree (cladogram) constructed from two originally obtained phylogenetic tree attained from two methods derived from *trn*L-*trn*F sequences dataset. The cladogram showed nipah *trn*L-*trn*F dataset can be divided into three varieties; Clade A, Clade A1 and Clade B.

Clade A1 was moderately and acceptably high supported by both original trees. Since Clade A1 reside in Clade A, there are three evolution possibility happen in Clade A1; (1) Samples in Clade A1 is evolving to Clade B; (2) Samples in Clade A1 is evolving to Clade A; (3) Samples in Clade A1 is further evolving to create a new variety in Nipah species. In either ways, nipah had shown varieties presented in its species which indicates the evolution did occurred.

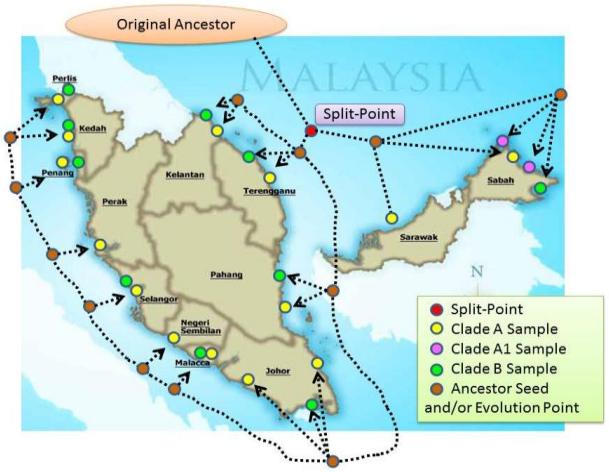

Fig. 2 Nipah phylodistribution in Malaysia interpreted from nipah *trn*L-*trn*F cladogram ultimate tree based on base nucleotide changes or branch length obtained from Maximum-Likelihood and Bayesian phylogenetic tree.

In previous nipah variety study, it was indicated that nipah had shown low genetic diversity within nipah sample from Thailand and almost no evolution occurred across nipah samples from China, Vietnam and Thailand [8]. We partly agreed and disagreed with this statement. We agreed because in this study, nipah also had shown very low genetic diversity within samples in presented clade, but in contrast, we disagreed due to the fact that in this study, nipah had shown genetic diversity across the samples by presented in three different clades that highly supported by bootstrap value, probability value and nucleotide changes in branch length. Given that, we can concluded that nipah in Malaysia had evolved to two certain varieties and one possible variety.

Nipah *trn*LF amplified genome patterned out that nipah evolved from Clade B to Clade A where nucleotide changes occurred higher in Clade A (0.660258 in ML tree and 0.687222 in MrBayes tree) than Clade B (0.068535 in ML tree and 0.109187 in MrBayes tree) with Clade A1 (0.661604 in ML tree and 0.689163 in MrBayes tree) most likely to be evolution adjusted stage. Nipah evolution may influenced by a few factors including; (1) Natural selection; (2) Genetic drift; (3) Isolation by distance, and; (4) Founder effect [12]–[15]. Nipah *trn*LF amplified genome pointed that nipah is evolved from small in size and length but produced more seed cluster (represented by clade B where nipah frond length can grow up to 3 m 10 cm and the stalk can grow up to 89 cm with the production of 2 to 3 seed cluster per nipah individual) to being big in size and length but produce less seed cluster (represented by Clade A where nipah frond length can grow up to 8 m 47 cm and the stalk can grow up to 1 m 80.2 cm with the production of 1 to 2 seed cluster per nipah individual) with Clade A1 represented adjusting evolution stage where Clade

A1 morphology was noticeable for being in between Clade A and Clade B possible morphology with some characteristic shared with either those two main clades.

IV. Conclusion

In conclusion, we had identified that nipah had shown genetic variability across nipah samples in Malaysia. This finding is strongly supported by both original phylogenetic trees with both original trees agreed that nipah in Malaysia can be divided into two strongly supported varieties (Clade A and Clade B) and one moderately and acceptably high supported variety (Clade A1) by using *trn*L–*trn*F amplified genome as targeted genome with Clade A1 most possible to be inter-evolutionary stage. Base substitution or nucleotide changes were parallel with this finding where distinctive nucleotide changes differences can be observed between obtained clades. Nipah *trn*L – *trn*F amplified genome indicated that nipah may evolved from being small in size and length but produced more seed cluster to being big in size and length but produced less seed cluster. This study also attained that nipah in Malaysia may firstly scattered and settled in Peninsular Malaysia before established themselves in Sabah and Sarawak. This phylogeographical pattern of nipah was reflected by the obtained clades in phylogenetic study where Clade A1 only appeared to be in Sabah samples. This study indicated *trn*L – *trn*F may be useful as genomic markers to study the Arecaceae line down to species level. Further study should be done to investigate mode and tempo of evolution occurred in nipah and verifying the evolution pattern either by using combination matrixes and/or by using other genomic marker and/or by cross-combined between genomic markers. In order to investigate nipah evolution, we also suggest that in the future, nipah evolution study should include some other taxa from closely related Arecaceae line.

Acknowledgment

We would like to thank Ministry of Science, Technology and Innovation (Malaysia) to fund this project, Institute Oceanography and Environment of Universiti Malaysia Terengganu and States Forestry Department (Malaysia) for endless support.

References

[1] L. S. Hamilton and D. H. Murphy, "Use and Management of Nipa Palm (Nypa fruticans, Arecaceae): A Review," Econ. Bot., vol. 42, no. 2, pp. 206–213, 1988. http://dx.doi.org/10.1007/BF02858921

[2] K. Tsuji, M. N. F. Ghazalli, Z. Ariffin, M. S. Nordin, M. I. Khaidizar, M. E. Dulloo, and L. S. Sebastian, "Biological and Ethnobotanical Characteristics of Nipa Palm (Nypa fructicans Wurmb.): A Review," Sains Malaysiana, vol. 40, no. 12, pp. 1407–1412, 2011.

[3] P. B. Tomlinson, The Botany of Mangroves. Cambridge: Cambridge University Press, 1986, p. 419.

[4] W. J. Baker, C. B. Asmussen, S. C. Barrow, J. Dransfield, and T. A. Hedderson, "A Phylogenetic Study of the Palm Family (Palmae) Based on Chloroplast DNA Sequences From the trnL-trnF Region," Plant Syst. Evol., vol. 219, pp. 111–126, 1999. http://dx.doi.org/10.1007/BF01090303

[5] W. J. Baker, T. A. Hedderson, and J. Dransfield, "Molecular Phylogenetics of Subfamily Calamoideae (Palmae) Based on nrDNA ITS and cpDNA rps16 Intron Sequence Data," Mol. Phylogenet. Evol., vol. 14, no. 2, pp. 195–217, Mar. 2000. http://dx.doi.org/10.1006/mpev.1999.0697

[6] J. P. Huelsenbeck and F. Ronquist, "MRBAYES: Bayesian Inference of Phylogenetic Trees," Bioinformatics, vol. 17, no. 8, pp. 754–755, Aug. 2001. http://dx.doi.org/10.1093/bioinformatics/17.8.754

[7] J. Silvertown and D. Charlesworth, Introduction to Plant Population Biology, Fourth Edi. Oxford: Blackwell Science Ltd, 2001, pp. 1–347.

[8] S. Jian, J. Ban, H. Ren, and H. Yan, "Low genetic variation detected within the widespread mangrove species Nypa fruticans (Palmae) from Southeast Asia," Aquat. Bot., vol. 92, no. 1, pp. 23–27, Jan. 2010. http://dx.doi.org/10.1016/j.aquabot.2009.09.003

[9] D. V Jobes, D. L. Hurley, and L. B. Thien, "Plant DNA Isolation: A Method to Efficiently Remove Polyphenolics, Polysaccharides, and RNA," Taxon, vol. 44, no. 3, pp. 379–386, 1995. http://dx.doi.org/10.2307/1223408

[10] K. Weising, H. Nybom, K. Wolff, and G. Kahl, DNA Fingerprinting in Plants: Principles, Methods, and Applications, Second Edi. Boca Raton: Taylor & Francis Group, 2005, pp. 1–444. http://dx.doi.org/10.1201/9781420040043.ch1

[11] R. C. Mehrotra, R. P. Tiwari, and B. I. Mazumder, "Nypa megafossils from the Tertiary sediments of Northeast India," Geobios, vol. 36, pp. 83–92, 2003. http://dx.doi.org/10.1016/S0016-6995(02)00107-9

[12] D. Briggs and S. M. Walters, Plant Variation and Evolution, Third Edit. Cambridge, United Kingdom: Cambridge University Press, 1997, pp. 1–512.

[13] A. Poiani, Ed., Pragmatic Evolution: Applications of Evolutionary Theory. New York, United States of America: Cambridge University Press, 2012, pp. 1–340.

[14] E. . Wiley and B. S. Lieberman, Phylogenetics: Theory and Practice of Phylogenetic Systematics, Second Edi. Hoboken, New Jersey: John Wiley & Sons, Inc., Publication, 2011, pp. 1–406.

[15] A. Young, D. Boshier, and T. Boyle, Eds., Forest Conservation Genetics: Principles and Practice. Collingwood, Australia: CSIRO Publishing, 2000, pp. 1–352.