S1-3

Molecular phylogeny and evolution of tropical forest trees

Koichi Kamiya [1,4], Ko Harada [1], Satoshi Nanami [2], Akira Itoh [2], Bibian M. Diway [3] and Lucy Chong [3]

[1] Faculty of Agriculture, Ehime University, Tarumi Matsuyama Ehime 790-8566, Japan
[2] Graduate School of Science, Osaka City University, Sugimoto 3-3-138 Sumiyoshi-ku, Osaka 558-8585, Japan
[3] Botanical Research Centre, Forestry, Sarawak Forestry Corporation, Sarawak, Malaysia
[4] Author for correspondence (e-mail: kkamiya@agr.ehime-u.ac.jp)

Abstract Genetic diversity is important for all levels of biodiversity because the extents and patterns of genetic diversity influence population viability and adaptation in changing environments, the generation of new species, and the function of ecosystems as well. DNA molecular markers are also widely used for rapid identification of species and for investigating spatial and temporal patterns of gene flow, which are important for understanding the population stability. Here we synthesize knowledge from evolutionary and population genetic studies of tropical forest trees, especially for dipterocarp species in Southeast Asia, and give a brief introduction of the related theoretical population genetic background. First we focus on DNA divergence between species, and how the comprehensive molecular phylogeny of the dipterocarp species is useful. Second we highlight the research studying the extents and the geographic distributions of genetic variations within species, and then discuss how to elucidate the biogeographic history of Sundaland based on the pattern of DNA variations. Future opportunities using genetic data in tropical forestry and ecology are discussed.

Keywords Dipterocarpaceae, Genetic variation, Phylogeny, Phylogeography

Analysis of genetic variation

Genetic diversity defined by the extents of variations in genetic materials from individuals, populations or species is fundamentally important for maintaining the species' existence in changing environments. The variability of genes is generated by spontaneous mutations. Novel mutations will be eliminated from a population immediately if they are deleterious, or will become more widespread if they increase individual survival and fertility (natural selection). However, if mutations are neither deleterious nor advantageous, the frequencies of the mutations are changed by chance alone and are fixed or finally disappear in a finite population (the neutral theory of molecular evolution). Such fates of genetic mutants are mathematically and empirically studied in population genetics.

DNA molecular markers are now widely used in systematics and ecological studies. Many of the markers are considered to be selectively neutral. Each locus often has several different alleles, and the combination of alleles is called “the genotype”. Because the number of genotypes increases with the number of loci studied, a multilocus genotype is usually individual-unique, and therefore individuals can be distinguished from each other by their genotypes. Such analysis is used for investigating spatial and temporal patterns of gene flow, which are important for understanding the population stability.

Molecular phylogeny of dipterocarps

DNA sequences are now often used for reconstructing evolutionary relationships among different species. The data are obtained from many samples of species to estimate the nucleotide difference between two DNA sequences, *k* (see appendix), and the results of DNA sequence analysis are expressed in a phylogenetic tree. Molecular phylogeny is used for testing morphology-based classifications of living organisms, and has more recently been applied for studying evolutionary association of ecological and ecophysiological traits. Here we introduce molecular phylogenetic studies of Dipterocarpaceae in Southeast Asia, and then discuss opportunities and challenges in linking such genetic studies to ecology.

Dipterocarpaceae is the most dominant tree family in Southeast Asian tropics, and consists of 13 genera with about 470 species in Asia (Ashton 1982). Molecular phylogenetic studies of dipterocarp species have been conducted based on DNA sequences of chloroplast genome regions (Kamiya et al. 1998; Kajita et al. 1998; Tsumura et al. 2011) and the nuclear *PgiC* locus (Kamiya et al. 2005). The earlier studies mainly focused on the generic relationship in Asian Dipterocarpaceae, and showed that Asian genera of Dipterocarpaceae were divided into two groups. The first group consists of *Dryobalanops*, *Hopea*, *Neobalanocarpus*, *Parashorea*, *Shorea*, and *Dipterocarpus*, and the second consists of *Anisoptera*, *Vatica*, *Cotylelobium*, and *Upuna*. Generic delimitations in Dipterocarpaceae proposed based on morphological characters (summarized in Ashton 1982) are mostly supported by these molecular studies, except for *Shorea*, a genus that appeared to be paraphyletic. There are four groups in genus *Shorea*, classified by timber characters (i.e., Balau, White Meranti, Yellow Meranti, and Red Meranti). *Parashorea* is a sister to the *Shorea* clade that includes the species belonging to Balau, Yellow Meranti, and Red Meranti, while *Hopea* is allied with White Meranti. Phylogeny with additional DNA sequence data from *Shorea* botanical sections (Kamiya unpublished) suggests that *Pentacme* is sister to a clade comprising three timber groups of *Shorea* and *Parashorea*, and *Doona* is sister to the cluster including *Neobalanocarpus*, *Hopea*, *Shorea* White Meranti, and *Shorea roxburghii* (Fig. 1).

DNA barcode and its applications

Molecular data based phylogenetic analysis is a strong tool to test evolutionary relationships among living organisms, and the occurrence of DNA sequence(s) unique to a particular species could be used for species identification using routine molecular laboratory work. This may be particularly useful for tropical plant species of which herbarium materials are often limited. A DNA barcode uses one or several short nucleotide sequences to identify species through reference to a DNA sequence database. For plant barcoding, chloroplast DNA regions, such as *rbcL*, *matK*, and *trnH-psbA* intergenic spacer, are often used (Kress et al. 2005). Among them, the noncoding *trnH-psbA* is more variable, and so more helpful to find differences between closely related taxa.

As mentioned above, previous phylogenetic studies could help to resolve a group of species, such as genus and timber groups of *Shorea*. Are gene sequences obtained from one species always different from those of other species?

Figure 2 shows a phylogeny of 72 *Shorea* specimens of 35 species collected in Lambir Hills National Park. Nucleotide substitutions found in the *trnH-psbA* region could identify 11 distinct clusters, but many of them were shared among different species. For example, cluster B involved at least 16 different species of the Red Meranti. Only clusters C, D, E, I, and K correspond to a single species. It is not surprising that closely related species share the same DNA sequences if they have diverged recently and have not accumulated nucleotide substitutions which had been fixed within the respective species. Chloroplast capture through inter-species hybridization also leads to the presence of DNA sequences shared between species (Kamiya et al. 2011).

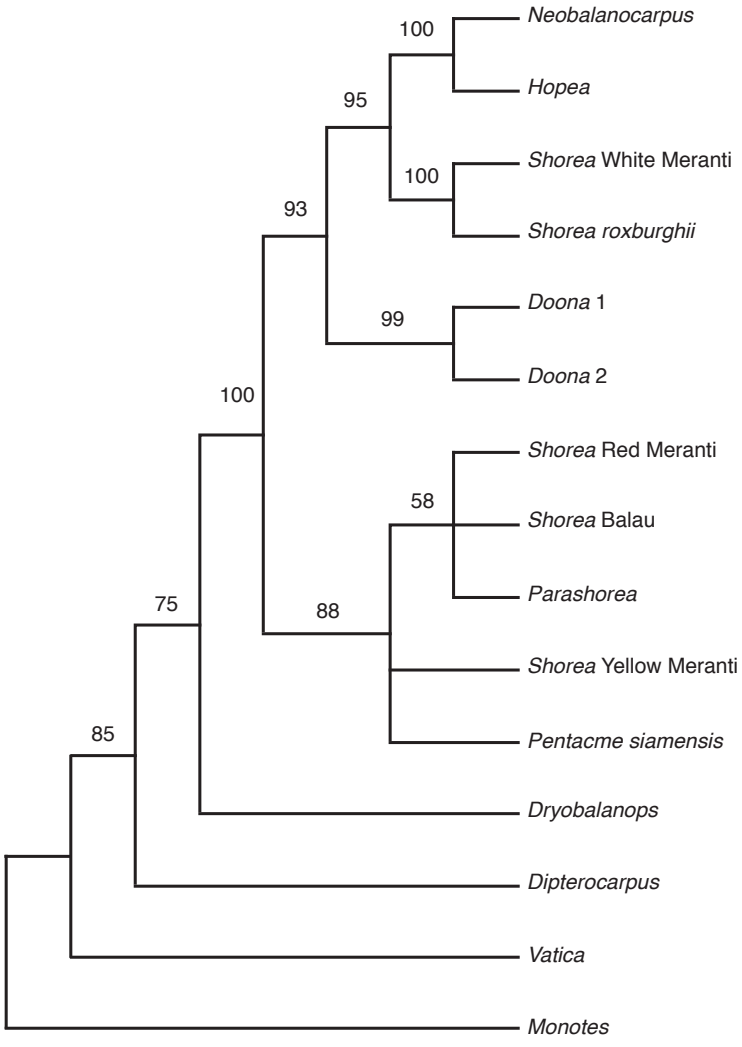

Fig. 1 Consensus of maximum parsimonious trees constructed from DNA sequences of chloroplast *trnK* region of 14 representative dipterocarp species. Numbers above branches show bootstrap values based on 1000 replications.

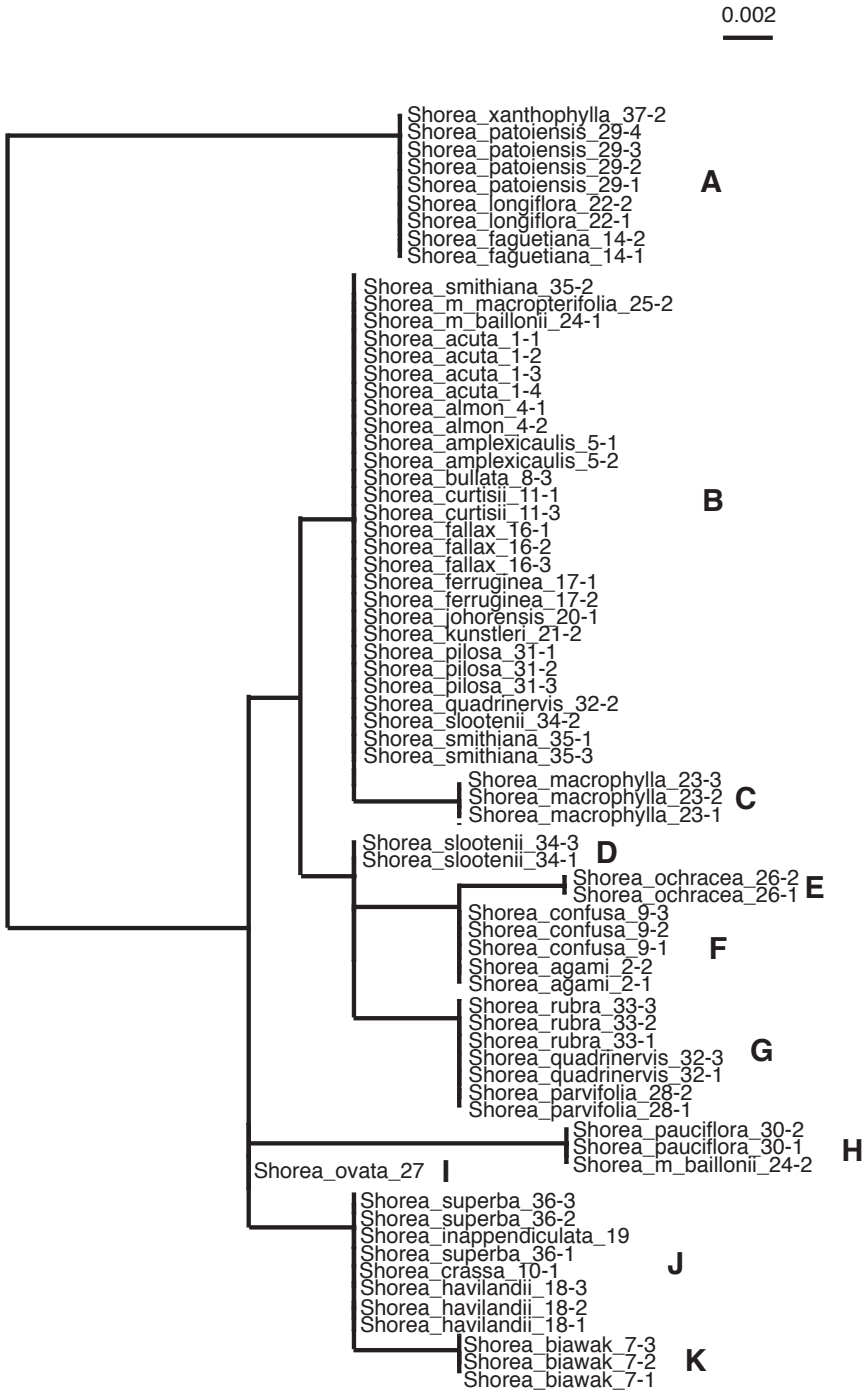

Fig. 2 Neighbor joining phylogeny based on chloroplast *trnH-psbA* sequences from 72 specimens comprising 35 *Shorea* species (data from Tsumura et al. 2011). Branch length is proportional to nucleotide divergence per site. Numbers after species name are ID codes.

In addition to elucidation of taxonomical uncertainty and species identification, plant DNA barcode data are also effectively used for ecological forensics. Kishimoto-Yamada et al. (2013) examined host ranges of herbivorous beetles based on identification of plant materials

collected from insect internal organs, and found that chrysomelid beetles fed on a wide variety of plant species, including ferns.

Tsumura et al. (2011) developed a chloroplast DNA sequence database of 84 *Shorea* species and methods for extracting DNA materials from veneers and sawn timbers. The study successfully found candidate species from veneer samples in the market using the chloroplast DNA analysis. Although this methodology is not always perfect due to physical degradation of the DNA remaining in the plywood samples, and will require well-trained personnel, this method can be utilized for tropical forest conservation; e.g., the method may enable us to identify products made from illegally logged timber (Tsumura et al. 2011).

The remarkable degree of species diversity of tropical rainforests is associated with their complicated ecological processes, such as interspecies interactions and niche partitioning. We have used modern analytical methods to explain how species richness in tropical rainforests reflects the evolution of physiological characteristics and functional traits. The method using phylogenetically independent contrast correlations between quantitative morphological traits has now become popular in ecology for finding evolutionarily associated traits (Webb et al. 2008). Blomberg et al. (2003) developed a method to detect phylogenetic signals in traits statistically. This analysis uses a descriptive statistic, *K*. *K* increases when the trait is more similar between evolutionarily related species. At present we are analyzing morphology, growth, mortality, and habitat preference data collected in the past 20 years at the Lambir 52-ha plot, together with our comprehensive DNA data set of dipterocarps, to examine the tendency for evolutionarily related species to closely resemble each other. A preliminary result suggests that growth traits, such as maximum dbh, rates of growth, mortality and recruitment, are more similar between evolutionarily closely related species in Dipterocarpaceae, but there is not such a tendency for habitat preference.

Intraspecific genetic variations and phylogeographic structures of the variations

Some plant species are distributed widely, and the areas of distribution are often partitioned by geological barriers, such as deep seas and high mountain ranges. Different populations of a species are genetically diverged independently with a limited gene dispersal, and as a result the populations are genetically distinct from each other. Molecular phylogeography is the study of DNA variations to find genetic disjunctions between populations and to explain how the patterns are related to the geological and climatic histories. The islands Java, Sumatra, Borneo, and the southern tip of the Eurasian continent (the Malay Peninsula) are part of a continental shelf, namely Sundaland. The part of Sundaland that is now underwater would have been exposed during the Pleistocene, allowing plant dispersal between the islands separated by the sea today. In Dipterocarpaceae many species are locally endemic and are only found in limited regions, but some species occur throughout Borneo, Sumatra, and the Malay Peninsula. What are the extent and pattern of genetic variation within each Southeast Asian species, and how do they reflect the geological and climate histories of Sundaland?

So far we have studied the geological distributions of genetic variations of *Shorea curtisii* (Kamiya et al. 2012), *S. leprosula* (Ohtani et al. 2013) and *Dryobalanops aromatica* (Dwiyanti et al. 2014). Analysis of DNA variations of multiple individuals and populations in those studies suggest that distributions of genetic variations are correlated with the geographic regions. Strong genetic disjunctions are found between the Bornean and the Malay Peninsula-Sumatran populations of these species, but populations in the Peninsula and Sumatra are genetically similar.

In *Shorea curtisii*, populations in the Malay Peninsula showed a signature of a bottleneck followed by demographic expansion, suggesting that the populations were newly derived, probably after the Pleistocene, when a cooler and drier climate were relaxed. In contrast, two Bornean populations of the species had no signature of a population expansion, suggesting these to be relict populations. DNA data of *S. leprosula* and *D. aromatica* also indicated gene dispersal between Borneo and the Malay Peninsula-Sumatra, most probably in the Pleistocene, when the greater part of Sundaland was exposed above sea level. The migration events occurred eastward in *S. leprosula*, but both eastward and westward in *D. aromatica*. There might have been several different migration routes depending on the paleo-distribution of the species reflected by their habitat preferences. It should be noted that attempting to explain the complicated biogeographic distributions of this region based only on the DNA variation of a few tree species would be imprudent.

Phylogeographic patterns of DNA variation are often explained in relation to past geological and climatic events. This is because the distribution of genetic variations becomes structured when migrations between populations are limited by geological and climatic factors for a long time. However, phylogeographic patterns of DNA variation could also be explained by the patterns of environmental gradients. In this case, populations in different environments would evolve under different natural selection pressures. Regardless of the reasons why genetic distinctiveness between populations is created, a crossing between genetically distinctive individuals often generates infertile or maladaptive descendants. Therefore, transfer of forest reproductive materials from one population to another in silviculture and reforestation programs must take particular care to avoid genetic contamination. The Forest Research Institute Malaysia (FRIM) used DNA data of *Neobalanocarpus heimii* to build Forest Reproductive Materials (FRM) transfer guidelines (N. Tani, personal communication). Although this is the first step for detecting FRM transfer zones, provenance tests, transplant experiments, common garden experiments, and genomic analysis of adaptive genetic variations for important timber species will improve the guidelines in the future.

Appendix

The neutral theory of molecular evolution predicts that the number of neutral mutations fixed in populations or species (λ) corresponds to the neutral mutation rate (μ) (Kimura 1968). In a population with the size $2N$, the total number of mutations becomes $2N\mu$ per locus per generation, and each mutation will be fixed in a population with the probability $1/2N$, and hence $\lambda = 2N\mu \times 1/2N = \mu$. The nucleotide difference between two DNA sequences (k) increases with t, the time since the DNA sequences have diverged from the common ancestor, and therefore k is expected to be $2\mu t$.

Acknowledgements

The authors thank Dr. K. Ogino, Dr. H. S. Lee, and Dr. J. J. Kendawang for their guidance and encouragement of our study. This research was supported by the Global Environment Research Fund of the Ministry of the Environment, Japan (D-0901 and RF-1010), Grants-in-Aid for Scientific Research from MEXT and JSPS, Japan (20405011 and 26304027), and a grant from the Sumitomo Foundation, Japan (073343).

References

Ashton PS (1982) Dipterocarpaceae. Flora Malesiana 9:237–552

Blomberg SP, Garland Jr. T, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745

Dwiyanti FG, Kamiya K, Harada K (2014) Phylogeographic structure of the commercially important tropical tree species *Dryobalanops aromatica* Gaertn. F. (Dipterocarpaceae) revealed by microsatellite markers. Reinwardia 14:43–51

Kajita T, Kamiya K, Nakamura K, Tachida H, Wickneswari R, Tsumura Y, Yorshimaru H, Yamazaki T (1998) Molecular phylogeny of Dipterocarpaceae in Southeast Asia based on nucleotide sequences of *matK*, *trnL* intron, and *trnL-trnF* intergenic spacer region in chloroplast DNA. Mol Phylogenet Evol 10:202–209. doi:10.1006/mpev.1998.0516

Kamiya K, Harada K, Lee HS, Ashton PS, Kajita T, Yamazaki T, Ogino K (1998) Molecular phylogeny of dipterocarp species using nucleotide sequences of two non-coding regions in chloroplast DNA. Tropics 7:195–207

Kamiya K, Harada K, Tachida H, Ashton PS (2005) Phylogeny of *PgiC* gene in *Shorea* and its closely related genera (Dipterocarpaceae), the dominant trees in Southeast Asian tropical rain forests. Am J Bot 92:775–788. doi:10.3732/ajb.92.5.775

Kamiya K, Gan YY, Lum SKY, Khoo MS, Chua SC, Faizu NNH (2011) Morphological and molecular evidence of natural hybridization in *Shorea* (Dipterocarpaceae). Tree Genet Genomes 7:297–306. doi:10.1007/s11295-010-0332-8

Kamiya K, Nanami S, Kenzo T, Yoneda R, Diway B, Chong L, Azani MA, Majid MN, Lum SKY, Wong KM, Harada K (2012) Demographic history of *Shorea curtisii* (Dipterocarpaceae) inferred from chloroplast DNA sequence variations. Biotropica 44:577–585. doi:10.1111/j.1744-7429.2011.00834.x

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626. doi: 10.1038/217624a0

Kishimoto-Yamada K, Kamiya K, Meleng P, Diway B, Kaliang H, Chong L, Itioka T, Sakai S, Ito M (2013) Wide host ranges of herbivorous beetles? Insights from DNA bar coding. PLoS ONE 8:e74426. doi:10.1371/journal.pone.0074426

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102:8369–8374. doi:10.1073/pnas.0503123102

Ohtani M, Kondo T, Tani N, Ueno S, Lee LS, Ng KKS, Muhammad N, Finkeldey R, Na'iem M, Indrioko S, Kamiya K, Harada K, Diway B, Khoo E, Kawamura K, Tsumura Y (2013) Nuclear and chloroplast DNA phylogeography reveals Pleistocene divergence and subsequent secondary contact of two genetic lineages of the tropical rainforest tree species *Shorea leprosula* (Dipterocarpaceae) in South-East Asia. Mol Ecol 22:2264–2279. doi:10.1111/mec.12243

Tsumura Y, Kado T, Yoshida K, Abe H, Ohtani M, Taguchi Y, Fukue Y, Tani N, Ueno S, Yoshimura K, Kamiya K, Harada K, Takeuchi Y, Diway B, Finkeldey R, Na’iem M, Indrioko S, Ng KKS, Muhammad N, Lee SL (2011) Molecular database for classifying *Shorea* species (Dipterocarpaceae) and techniques for checking the legitimacy of timber and wood products. J Plant Res 124:35–48. doi:10.1007/s10265-010-0348-z

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24:2098–2100. doi:10.1093/bioinformatics/btn358