ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2011) **9**(2); 193–196
doi:10.1017/S1479262111000396

# Molecular adaptation of the chloroplast *matK* gene in *Nymphaea tetragona*, a critically rare and endangered plant of India

Jeremy Dkhar, Suman Kumaria* and Pramod Tandon

*Plant Biotechnology Laboratory, Centre for Advanced Studies in Botany, North Eastern Hill University, Shillong 793022, India*

## Abstract

Sustainable utilization of plant genetic resources for food and agriculture has been increasingly discussed at both national and international forums. Besides exploitation, conservation of plant genetic resources has become an integral part of these discussions. Conservation aims at maintaining the diversity of living organisms, their habitat and the interrelationship between organisms and their environment. For achieving such goals, appropriate conservation strategies have to be adopted. Determining the genetic makeup of a particular plant species is of critical importance when planning a suitable conservation strategy. In this study, we sequenced the chloroplast *trnK* intron, *matK* and *rbcL* gene aimed at understanding the rarity of *Nymphaea tetragona*, a critically rare and endangered plant of India found at only one location. We extended our investigation to other *Nymphaea* species such as *N. nouchali*, *N. pubescens* and *N. rubra* that are commonly available throughout India. Interestingly, *matK* gene of *N. tetragona* revealed high number of non-synonymous substitutions. Molecular evolutionary analysis indicated that three of these sites may be under mild selective pressures. Such adaptive changes at the DNA and protein sequence level of *matK* gene may have been associated with the colonization of *N. tetragona*, suggesting that it could have migrated from China.

**Keywords:** *matK*; molecular adaptation; molecular evolutionary analysis; *Nymphaea tetragona*

## Introduction

*Nymphaea tetragona* represents one of the naturally occurring species found in India (Mitra, 1990). Its distribution is restricted to one particular location found at Nongkrem, Meghalaya, India (25°28′N–91°52′E). It is considered to be rare and endangered and has been included for recovery programmes (Ganeshaiah, 2005). Besides *N. tetragona*, there are an additional nine species (both wild and cultivated) of the genus *Nymphaea* occurring in India (Mitra, 1990). However, traditional classification of the genus in India has received unconvincing treatments with some names inaccurately used (Cook, 1996). To vindicate Cook's remarks, molecular taxonomic revision of four Indian representatives of the genus *Nymphaea* viz. *N. nouchali*, *N. pubescens*, *N. rubra* and *N. tetragona* based on internal transcribed spacers (ITS), *trnK* intron and *matK* gene was reported (Dkhar *et al*., 2010). Molecular evidence suggested probable misidentification of one specimen of *N. nouchali* and *N. tetragona*. Furthermore, a comparison of the genetic closeness among the Indian, Chinese and Russian material of *N. tetragona* based on the ITS region was also evaluated. The results indicated a close relationship between the Indian and the Chinese representatives, which is further supported by the morphological similarities shared between them (eFloras, 2008).

An interesting observation made from the previous study was the relatively higher number of nonsynonymous substitutions as compared to synonymous substitutions detected in the *matK* gene of *N. tetragona*. The higher rate of nonsynonymous substitutions is usually inferred as an indication of selective pressures acting on protein-coding sequences. The chloroplast *matK* gene,

* Corresponding author. E-mail: sumankhatrikumaria@hotmail.com

coding for the maturase enzyme, has been reported to evolve under positive selection in some lineages of land plants (Hao *et al.*, 2010). Hao *et al.* (2010) showed that several regions (N-terminal region, RT domain and domain X) of *matK* experience molecular adaptation, which fine-tunes maturase performance.

In the present study, to gain insight into whether an adaptive evolutionary process is linked to the colonization of new habitats by *N. tetragona*, we tested whether there was selective pressure in *matK* using phylogenetic and maximum likelihood analyses of codon and branch site substitution models. This estimation was extended to members of the order Nymphaeales, not included in the study of Hao *et al.* (2010), representing a lineage at the base of the angiospermic tree.

## Materials and methods

### *Taxon sampling and nucleotide sequence data*

Sequence data of *matK* gene and *trnK* intron of four wild Indian representatives of the genus *Nymphaea*, viz. *N. nouchali*, *N. pubescens*, *N. rubra* and *N. tetragona* was taken from our own previous study (Dkhar *et al.*, 2010). The *rbcL* gene was amplified and sequenced following the method adopted in another study (Dkhar *et al.*, in press). In addition, barring *Nymphaea*, nucleotide sequence data of members of the order Nymphaeales (*Barclaya longifolia*, *Euryale ferox*, *Nuphar advena*, *Ondinea purpurea* and *Victoria cruziana*), *Amborella trichopoda* and *Austrobaileya scandens,* identified as the most basal angiosperms, were retrieved from GenBank.

### *Molecular evolutionary analysis*

In order to conduct phylogenetic maximum likelihood analysis of selection in chloroplast genes, a phylogenetic tree was reconstructed based on the chloroplast *trnK* intron, *matK* and *rbcL* genes using maximum likelihood method implemented in Phylip (Felsentein, 1989).

Molecular adaptation tests on the chloroplast *matK* codon sites were performed using maximum likelihood models and programs included in PAML ver. 4.3 (Yang, 2007). We performed five site-specific codon substitution models: null models for testing positive selection (M1A, M7 and M8A) and models allowing for positive selection (M2A and M8). Furthermore, to provide evidence whether positive selection is linked to the colonization of *N. tetragona*, the branch site models (model A with $\omega$ fixed or estimated) implemented in PAML ver 4.3 were used. The likelihood ratio test (LRT) was used to compare these alternative models.

## Results and discussion

### *Positive selection in* matK *gene*

Maximum likelihood analysis of the combined datasets yielded a phylogenetic tree with log likelihood value of − 10 921.95 (Fig. 1). The reconstructed phylogenetic tree was identical in topology to those obtained in a study by Löhne *et al.* (2007).

For detecting positive selection using codon substitution models, three LRTs were performed comparing M1A (nearly neutral) with M2A (positive selection), M7 ($\beta$) with M8 ($\beta$ and $\omega_s \geq 1$), and M8A ($\beta$ and $\omega_s = 1$) with M8 ($\beta$ and $\omega_s \geq 1$). Because positive selection operates more strongly at some sites (Aguileta *et al.*, 2009), the distribution of sites influenced by selection for all three regions (N-terminal region, RT domain and domain X) of the *matK* gene was evaluated. For N-terminal region, model M2A indicated that 13.2% of the sites are under positive selection ($\omega = 1.778$). In model M8, 13.0% of the sites are under selective pressure ($\omega = 1.791$). For RT domain, model M8 estimated an $\omega$ value of 46.57 for 0.01% of codon sites. No positive sites were detected for domain X, a relatively more conserved and important functionally, as it determines the maturase activity of *matK* proteins (Hausner *et al.*, 2006). Two LRTs, M7-M8 and M8A-M8, were significant at the 0.01 and 0.05 significance level, respectively. Comparing M8 against the null model M8A is more robust than M1A *versus* M2A or M7 *versus* M8 (Swanson *et al.*, 2003).

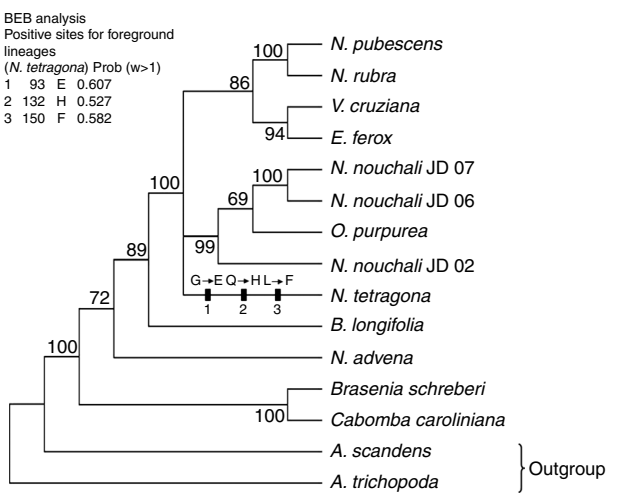


**Fig. 1.** Maximum likelihood tree of combined chloroplast DNA dataset (3859 bp) using Phylip. Three positive sites with probability for foreground (*N. tetragona*) estimated using branch-site models of PAML are indicated. Amino-acid replacements at these sites are also indicated at the branch (G: Glycine, E: Glutamic acid, Q: Glutamine, H: Histidine, L: Leucine, F: Phenylalanine). Numbers at nodes indicate bootstrap values. Prob, probability.

https://doi.org/10.1017/S1479262111000396 Published online by Cambridge University Press

**Table 1.** Estimation of log-likelihood (LnL) values and parameters for each region of the *matK* gene using branch-site models implemented in PAML selecting *N. tetragona* as foreground branch.

| | N-terminal region | | RT domain | | Domain X | |
|---|---|---|---|---|---|---|
| Model | LnL | Parameters | LnL | Parameters | LnL | Parameters |
| Foreground: *N. tetragona* | | | | | | |
| Model A ($\omega_2 = 1$ fixed) | −3148.00 | $p_0 = 0.73\,633$, $\omega_0 = 18\,596$ $p_1 = 0.25\,529$, $\omega_1 = 1.00\,000$ $p_2 + p_3\ 0.00\,839$, $\omega_2 = 1.00\,000$ | −229.30 | $p_0 = 1.00\,000$, $\omega_0 = 0.32\,206$ $p_1 = 0.00\,000$, $\omega_1 = 1.00\,000$ $p_2 + p_3 = 0.00\,000$, $\omega_2 = 1.00\,000$ | −1250.61 | $p_0 = 0.67\,267$, $\omega_0 = 0.11\,716$ $p_1 = 0.32\,733$, $\omega_1 = 1.00\,000$ $p_2 + p_3 = 0.00\,000$, $\omega_2 = 1.00\,000$ |
| Model A ($\omega_2$ = estimated) | −3147.80 | $p_0 = 0.73\,321$, $\omega_0 = 0.18\,697$ $p_1 = 0.24\,589$, $\omega_1 = 1.00\,000$ $p_2 + p_3 = 0.0209$, $\omega_2 = 23.09\,944$ | −229.30 | $p_0 = 1.00\,000$, $\omega_0 = 0.32\,206$ $p_1 = 0.00\,000$, $\omega_1 = 1.00\,000$ $p_2 + p_3 = 0.00\,000$, $\omega_2 = 1.17\,397$ | −1250.61 | $p_0 = 0.67\,267$, $\omega_0 = 0.11\,716$ $p_1 = 0.32\,733$, $\omega_1 = 1.00\,000$ $p_2 + p_3 = 0.00\,000$, $\omega_2 = 1.00\,000$ |

*p*, proportion of codon sites with ω; ω, d*N*/d*S*.

Similarly, branch-site model A ($\omega_2 = 1$ fixed) with branch-site model A ($\omega_2$ estimated) were compared. For N-terminal region, model A ($\omega_2$ estimated) for *N. tetragona* branch indicated that 2.09% of the sites are under selective pressure with $\omega_2 = 23.099$ (Table 1). Bayes Empirical Bayes (BEB) analysis showed three positive sites with probability value of 0.607, 0.527 and 0.582, respectively.

### *Selective pressure during colonization of* N. tetragona

*N. tetragona* is well distributed globally, found throughout China, Japan, Finland and Russia; but its location in India is restricted to one particular population. Our previous study indicated that the Indian plant is closely related to plants found in China (Dkhar *et al.*, 2010). Molecular evolutionary analysis suggested that the plant taxon migrated from China, and the migratory processes involved might have brought about some changes at both the DNA and protein sequence levels. This adaptive response would have rendered slight advantage for the plant to survive in new habitat. Such adaptive changes at the DNA and protein sequence levels of *rbcL* gene have been reported in *Schidea* (Kapralov and Filatov, 2006) and *Potamogeton* (Iida *et al.*, 2009).

### *Implications for conservation of* N. tetragona

Determining the genetic makeup of a particular plant species is of critical importance when planning a suitable conservation strategy. Here, we provided evidence for molecular adaptation of *matK* gene in *N. tetragona,* a critically rare and endangered plant of India. Such adaptive changes at the molecular level may have been associated with the adaptation of *N. tetragona* to varying ecological conditions. So, the threat to the existence of this plant taxon is primarily anthropogenic, brought about through large-scale cultivation at the vicinity of the location. Conservation of *N. tetragona* can be targeted at controlling such activities or the probable translocation to another site.

## Acknowledgements

The authors acknowledge financial support from Department of Biotechnology, Ministry of Science and Technology, India (file no. BT/PR-7055/BCE/08/437/2006).

## References

Aguileta G, Refrégier G, Yockteng R, Fournier E and Giraud T (2009) Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protests. *Infection, Genetics and Evolution* 9: 656–670.

Cook CDK (1996) *Aquatic and Wetland Plants of India*. Oxford: Oxford University Press.

Dkhar J, Kumaria S, Rama Rao S and Tandon P (2010) Molecular phylogenetics and taxonomic reassessment of four Indian representatives of the genus *Nymphaea*. *Aquatic Botany* 93: 135–139.

Dkhar J, Kumaria S and Tandon P Molecular evidence suggests that an Indian plant called *Nymphaea alba* var. *rubra* is a hybrid originating from *N. alba* and *N. odorata*. *Annales Botanici Fennici* (in press).

eFloras (2008) Published on the Internet http://www.efloras.org Missouri Botanical Garden, St. Louis, MO and Harvard University Herbaria, Cambridge, MA. Accessed 16 March 2010

Felsentein J (1989) Phylip-phylogeny inference package (version 3.2). *Cladistics* 5: 164–166.

Ganeshaiah KN (2005) Recovery of endangered and threatened species. Developing a national priority list of plants and insects. *Current Science* 89: 599–600.

Hao DC, Chen SL and Xiao PG (2010) Molecular evolution and positive Darwinian selection of the chloroplast maturase *matK*. *Journal of Plant Research* 123: 241–247.

Hausner G, Olson R, Simon D, Johnson I, Sanders ER, Karol KG, McCourt RM and Zimmerly S (2006) Origin and evolution of the chloroplast *trnK* (*matK*) intron: a model for evolution of group II intron RNA structures. *Molecular Biology and Evolution* 23: 380–391.

Iida S, Miyagi A, Aoki S, Ito M, Kadono Y and Kosuge K (2009) Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. *PLoS ONE* 4: e4633. doi:10.1371/journal.pone0004633

Kapralov MV and Filatov DA (2006) Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS ONE* 1: e8. doi:10.1371/journal.pone0000008

Löhne C, Borsch T and Wiersema JH (2007) Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers. *Botanical Journal of the Linnean Society* 154: 141–163.

Mitra RL (1990) Nymphaeaceae. In: Nayar MP, Thothathri K and Sanjappa M (eds) *Fascicles of flora of India, Fascicles 20*. Howrah: Botanical Survey of India, pp. 11–25.

Swanson WJ, Nielsen R and Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution* 20: 18–20.

Yang Z (2007) PAML4: a program package for phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.