

## Genome and Evolution of the Sacred Lotus\*

Partha P Majumder\*\*

(Received 14 June 2016)

### Abstract

The ornamental lotus is of iconic significance in two major cultures of the world; Hinduism and Buddhism. It is also of considerable medicinal and economic importance. The sacred lotus split from the ancestor of the core eudicots about 140 million years ago. It possesses 16 (2n) chromosomes. Its estimated genome size is about 900 Mb. It was originally a land plant that had adapted itself to surviving in water. A large number of multi-copper oxidase family proteins with root-specific expression in the lotus reflects its adaptation to limited nutrient availability in an aquatic environment. The seeds of the lotus can remain dormant for over 1000 years. One annexin gene regulates seed thermotolerance and germination. The sacred lotus, therefore, possesses some unique genomic characteristics.

**Key words:** Aquatic adaptation, Dormancy, Eudicot, Evolutionary split, Whole genome duplication

### 1. INTRODUCTION

Sacred lotus (*Nelumbo nucifera*) is an ornamental plant of agricultural, medicinal, cultural and religious importance. It is an angiosperm. It belongs to Nelumbonaceae; a family of basal eudicot plants that contains only one genus, *Nelumbo*. There are only two species within the Nelumbonaceae family, sacred lotus and American lotus. Its closest relatives are shrubs or trees belonging to the families Proteaceae and Platanaceae. Lotus was a land plant that has adapted to aquatic environments.

It is primarily found in East Asia and Northern Australia. It was domesticated in Asia about 7,000 years ago. In China and other Asian countries, the lotus is an economically important crop. It is used as food and is cultivated for its edible rhizomes, seeds and leaves. Its buds, flowers, anthers, stamens, fruits, leaves, stalks, rhizomes and roots are used as herbal medicines in China and elsewhere for treatment of cancer,

depression, diarrhoea, heart problems, hypertension and insomnia.

This basal eudicot species is especially important from an evolutionary perspective as it occupies a critical phylogenetic position in flowering plants. The two hallmark genomic studies (Ming et al., 2013; Wang et al., 2013) on the sacred lotus studied a wild strain from Central China (Wang et al., 2013) and the “China Antique” (Ming et al., 2013). Both studies came to similar conclusions. There were some differences in the details of their inferences presumably resulting from differences of genomic technologies and methodologies of data analyses adopted by them.

### 2. GENOME STRUCTURE

The diploid chromosome set (2n) of the sacred lotus is 16. The estimated genome size is between 879-929 Mb, based on assembled draft sequences of 792-803 Mb. Approximately 50% of sequences are related to transposable elements;

\* This summary is based on work published by other scientists which was presented by the author during the Symposium.

\*\* National Institute of Biomedical Genomics, Kalyani, Email: ppm1@nibmg.ac.in

represented LTR superfamilies are *Gypsy* and *Copia*, with more *Copia* elements present in the lotus genome. The genome notably lacks the paleo-triplication observed in other eudicots, but reveals a lineage-specific duplication. A whole genome duplication event in the sacred lotus likely took place about 10 million years ago. The genome has evidence of slow evolution, with a 30% slower nucleotide mutation rate than observed in the grape plant. The average gene density is 1 gene per 30 kb. In the human the average gene density is 1 per 70-100 kb. Thus, the sacred lotus genome is more gene-rich than the human.

### 3. RELATIONSHIP TO OTHER GENOMES

Gene families identified by clustering encoded proteins revealed that the genes of the sacred lotus can be classified into 13,834 families. Of these, 2075 gene families were specific to sacred lotus and 9481 families were shared among sacred lotus, *Arabidopsis*, grape, and soybean. Comparison of the sacred lotus genome with those of maize, rice, potato, grape, soybean and *Arabidopsis* showed that the sacred lotus genome split from the ancestor of the core eudicots about 140 million years ago.

### 4. PROTEIN-CODING GENES AND EXPRESSION

The estimated number of protein-coding genes varied between the two studies; these were 26685 and 36385. A large proportion (84%) of the genes was also represented in RNA sequencing data. The number of genes that were found expressed across all tissues surveyed was 14477. Expression of 3094 protein coding genes was tissue-specific; 1910 genes showed expression only in rhizomes and 841 only in roots.

### 5. ADAPTATION TO AQUATIC ENVIRONMENT

The lotus is particularly noted for its seed longevity [an <<1300-yr-old lotus fruit, recovered from a dry lakebed in northeastern China, is the

oldest germinated and directly <sup>14</sup>C-dated fruit known; Shen-Miller et al., 2002] and exceptional water repellency, known as the lotus effect. Submersed plant growth presents unique physiological challenges. An astonishing number of putative copper-dependent proteins, of which (a) 63 proteins contain at least one COX2 domain, (b) 55 contain a 'copper-binding-like' domain, and (c) 4 contain polyphenol oxidases. This feature is attributable to expansions in COG2132, a family of multi-copper oxidases. Most plant genomes encode one or two members of COG2132, whereas lotus has at least 16 members due to whole genome duplication and repeated tandem duplications. These sixteen COG2132 multi-copper oxidase family proteins have root-specific expression; these are involved in root meristem phosphate starvation, reflecting adaptation to limited nutrient availability in an aquatic environment.

### 6. DORMANCY AND THERMOTOLERANCE OF THE SEED

A sacred lotus seed can remain dormant for hundreds of years and then germinate when placed into optimal conditions. Sacred lotus seeds can also withstand extremely high temperatures. Annexin proteins play important roles in this process. The peroxidase activity of annexins protects membranes against peroxidation. One annexin gene, NnANN1, had been identified in sacred lotus and shown to regulate seed thermotolerance and germination (Chu et al., 2012).

### 7. STARCH ASSOCIATED GENES

The sacred lotus has high levels of starch. Seventy genes that are potentially involved in starch synthesis were examined in 21 plant species. Of these 70 genes, 35 localized to a duplicated genomic region, indicating that the expansion of starch genes was likely caused by the recent whole genome duplication. This expansion may have resulted in high levels of

amylose in sacred lotus. No such gene expansion was observed in the potato. Therefore, it has been speculated that processes other than gene expansion (e.g., changes in gene expression levels) may play important roles in starch synthesis in potato.

### 8. EPILOGUE

Sequencing of the sacred lotus genome has provided insights into the genomic basis of the extraordinary properties of this plant, that include long seed dormancy, amazing hydrophobicity and high starch content.

### REFERENCES

- Chu, P., Chen, H., Zhou, Y., Li, Y., Ding, Y., Jiang, L., Tsang, E.W., Wu, K., Huang, S. Proteomic and functional analyses of *Nelumbo nucifera* annexins involved in seed thermotolerance and germination vigor. *Planta*, 235 (2012):1271–1288.
- Ming, R., Vanburen, R., Liu, Y., Yang, M., Han, Y., Li, L.T., Zhang, Q., Kim, M.J., Schatz, M.C., Campbell, M., Li, J., Bowers, J.E., Tang, H., Lyons, E., Ferguson, A.A., Narzisi, G., Nelson, D.R., Blaby-Haas, C.E., Gschwend, A.R., Jiao, Y., Der, J.P., Zeng, F., Han, J., Min, X.J., Hudson, K.A., Singh, R., Grennan, A.K., Karpowicz, S.J., Watling, J.R., Ito, K., et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.) *Genome Biology* 14(2013): R41.
- Shen-Miller, J., William Schopf, J., Harbottle, G., Cao, R.-J., Ouyang, S., Zhou, K.-s., Southon, J.R., Liu, G.-h. Long-living lotus: germination and soil  $\alpha$ -irradiation of centuries-old fruits, and cultivation, growth, and phenotypic abnormalities of offspring. *American Journal of Botany* 89(2002): 236-247.
- Wang, Y., Fan, G., Liu, Y., Sun, F., Shi, C., Liu, X., Peng, J., Chen, W., Huang X., Cheng, S., Liu, Y., Liang, X., Zhu, H., Bian, C., Zhong, L., Lv, T., Dong, H., Liu, W., Zhong, X., Chen, J., Quan, Z., Wang, Z., Tan, B., Lin, C., Mu, F., Xu, X., Ding, Y., Guo, A.Y., Wang, J., Ke, W. The sacred lotus genome provides insights into the evolution of flowering plants. *The Plant Journal* 76(2013):557–567.