

REVIEW ARTICLE

Water lilies as emerging models for Darwin's abominable mystery

Fei Chen¹, Xing Liu¹, Cuiwei Yu², Yuchu Chen², Haibao Tang¹ and Liangsheng Zhang¹

Water lilies are not only highly favored aquatic ornamental plants with cultural and economic importance but they also occupy a critical evolutionary space that is crucial for understanding the origin and early evolutionary trajectory of flowering plants. The birth and rapid radiation of flowering plants has interested many scientists and was considered 'an abominable mystery' by Charles Darwin. In searching for the angiosperm evolutionary origin and its underlying mechanisms, the genome of *Amborella* has shed some light on the molecular features of one of the basal angiosperm lineages; however, little is known regarding the genetics and genomics of another basal angiosperm lineage, namely, the water lily. In this study, we reviewed current molecular research and note that water lily research has entered the genomic era. We propose that the genome of the water lily is critical for studying the contentious relationship of basal angiosperms and Darwin's 'abominable mystery'. Four pantropical water lilies, especially the recently sequenced *Nymphaea colorata*, have characteristics such as small size, rapid growth rate and numerous seeds and can act as the best model for understanding the origin of angiosperms. The water lily genome is also valuable for revealing the genetics of ornamental traits and will largely accelerate the molecular breeding of water lilies.

Horticulture Research (2017) 4, 17051; doi:10.1038/hortres.2017.51; Published online 4 October 2017

INTRODUCTION

Ornamentals, cultural symbols and economic value

Water lilies are beautiful aquatic flowering plants that are distributed worldwide. These plants are found in the aquatic section of nearly every botanic garden because of their highly valued ornamental features. The almost full spectrum of petal colors range from black to white, making water lilies the most diversely colored flowering plants (Figure 1). The lovely cup-like flower shapes and floating leaves such as the famous *Victoria* and Amazon water lilies are also favored ornamental characteristics. In Bangladesh and Sri Lanka, water lilies were chosen as the national flower because they are regarded as a symbol of truth, purity, and discipline.

Beyond being beautiful ornamental plants, water lilies have been utilized as an ingredient in many products, including beneficial and cosmetic substances, soap, perfume, hand cream, flower tea bags, and traditional medicine.¹ In Asian countries, some water lilies such as *Brasenia schreberi*, *Euryale ferox*, *Nymphaea* spp. are traditional vegetables with edible parts including young leaves, stems, and seeds. Several *Nymphaea* species have also been used to purify heavy metal-contaminated water and soap-polluted wastewater.²

Critical evolutionary place

In taxonomy, plants categorized in the Order Nymphaeales share the common name water lily.³ Water lilies are divided into three families: Hydatellaceae, Cabombaceae, and Nymphaeaceae.⁴ The family Nymphaeaceae has the most species of the three families and consists of six genera: *Barclaya*, *Euryale*, *Nuphar*, *Nymphaea*,

Ondinea, and *Victoria*.^{4,5} Floral organs differ greatly among each family in the order Nymphaeales. In the genus *Nymphaea*, flowers are composed of 4 sepals, 50 to 70 petals, 30 to 40 carpels, and 120 to 250 stamens. These characteristics are often regarded as the most primitive angiosperm floral characteristics, as seen in various ancestral flowering plant fossils.⁶

In the tree of plant life, basal angiosperms consisting of three orders Nymphaeales, Amborellales, and Austrobaileyales, have long been regarded as the basal branches of angiosperms using both molecular phylogenetic and developmental classifications.^{7,8} Although multiple lines of evidence support *Amborella* as the basal-most angiosperm,^{7,9–11} the water lily-basal or *Amrorella*-water lily co-basal theories cannot yet be ruled out.^{12–15} The genomic sequences of the water lily may be critical in resolving the early evolution of angiosperms, because among all basal angiosperms only the genome of *Amborella* is currently known.

Limited genetic and genomic analysis of water lilies

Despite the importance of water lilies in phylogenetic research and as an aquatic ornamental plant, limited genetic and genomic information is available. Previous chromosome number and size studies have provided the karyotype background of approximately 65 water lily species^{16,17} (Table 1). Only two homologs of *INO* genes,¹⁸ two reference genes for expression studies,¹⁹ six floral organ identity genes,²⁰ and ABC model genes²¹ have been cloned (Table 1). Genetic markers, such as the *matK* genes⁴ and inter-simple sequence repeats²² have been applied in DNA barcoding of the water lily germplasm.

At the omics level, genome-wide expressed sequence tags (ESTs) were generated in 2006 from the yellow water lily *Nuphar*

¹Center for Genomics and Biotechnology; State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops; Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops; Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology; Fujian Agriculture and Forestry University, Fuzhou 350002, China and ²Zhejiang Humanities Landscape Co., LTD, Hangzhou 310030, China.

Correspondence: L Zhang (fafuzhang@163.com)

Received: 14 May 2017; Revised: 30 June 2017; Accepted: 26 July 2017



Figure 1. Water lilies are ornamental plants with beautiful flowers and leaves: (a) *Nymphaea* 'Hermine', (b) *N.* 'Marliacea Chromatella', (c) *N.* 'Wanvisa', (d) *N.* 'Gigantea Hybrid1', (e) *N. colorata*, (f) *N.* 'Muang Wiboonlak', (g) *N.* 'Piyalarp', (h) *N.* 'Agkee Sri Non', (i) leaf ornamental Victoria water lily.

advena for genome duplication analysis.²³ Later, the transcriptomes from seven tissues/organs were sequenced and analyzed from the same species²⁴ (Table 1). Recently, the transcriptome of six samples from two coloring stages of the beautiful blue water lily *Nymphaea* 'King of Siam' were sequenced, together with metabolic analysis, to reveal the blue flower's formation.²⁵ So far, no water lily genome has been reported.

The water lily holds the key to Darwin's abominable mystery. The origin and rapid massive expansion of flowering plants in a relatively short geological time, which resulted in most current-day flora, fascinated Charles Darwin, who called it an 'abominable mystery' and the 'most perplexing phenomenon', beyond which there was 'nothing... more extraordinary'.²⁶ Over the 137 years since this expansion was proposed by Charles Darwin in 1879, evolutionary biologists have long attempted to reconstruct the early history of angiosperms. One of the most critical questions to solving this mystery is to determine which lineage is the most basal angiosperm. So far, there have been several hypotheses.

From ANITA to ANA basal hypothesis to Amborella and water lily co-basal hypothesis. In 1999, relying on molecular phylogenetic methods, several groups proposed that the *Amborella*, *Nymphaeales*, and *Illiciales-Trimeniaceae-Austrobaileya* (ANITA) clade is the extant basal angiosperm^{8,27,28} (Figure 2). However, these phylogenetic trees were all based on a single gene or a few genes, mainly from chloroplasts.²⁸ In 2005, based on several plastid, mitochondrial, and nuclear genes, researchers proposed that the *Amborella*, *Nymphaeaceae*, and *Austrobaileya* (ANA) clade (Figure 2) were the basal sister clades to all other angiosperms.²⁹ This classification sets either *Amborella* or *Amborella* and *Nymphaeales* as the sister to all other angiosperms.²⁹ However, it was not clear which was the most basal angiosperm. Recent releases of new genome sequences has greatly improved phylogenomic or phylotranscriptomic analysis for species tree reconstruction.³⁰ A phylogenetic analysis of 61 plastid genes first reported *Nymphaeales* and the *Amborella*, the extant relatives, as the most basal lineage of flowering plants.³¹ This was later supported by two phylotranscriptomic analyses.^{9,10} In the last few years, phylogeneticists have attempted to resolve which angiosperm is the most basal.

Amborella as the most basal angiosperm. Unlike single gene-based phylogenetics, when using three mitochondrial genes, one chloroplast gene and one nuclear gene, an early phylogenetic

analysis placed *Amborella*, and not water lilies, as the most basal angiosperm branch⁹ (Figure 2). This species tree topology is well supported by two recent phylotranscriptomic analyses^{9,10} using

Table 1. Available molecular research on water lilies

	Taxon name	Chromosome (n)	Genome size (Mb)	Genetic study	Transcriptome
Nymphaeaceae	<i>Nymphaea alba</i> L.	42 [1]	1950 [6]	INO gene [8], ITS2+matK [9]	
	<i>Nymphaea amazonum</i> Mart. & Zucc.	9 [1], ~ 18 [2]	821.52 [2]	ITS2+matK [9]	
	<i>Nymphaea ampla</i> (Salisb.) DC.	14 [2]	772.62 [2]		
	<i>Nymphaea atrans</i> S. W. L. Jacobs	42 [1]	1408.32 [2]		
	<i>Nymphaea bissetii</i> hort.	42 [1]			
	<i>Nymphaea caerulea</i> Savigny	14 [1]		ITS2+matK [9]	
	<i>Nymphaea carpentariae</i> 'Julia Leu'	~ 42 [2]	1447.44 [2]		
	<i>Nymphaea candida</i> C. Presl	56 [1]	1936.44 [2]		
	<i>Nymphaea capensis</i> Thunb.	14 [1]		trnH-psbA, rpoC1 [10]	
	<i>Nymphaea colorata</i> Peter	14 [2]	489 [2]	INO gene [8]	
	<i>Nymphaea conardii</i> Wiersema	9 [1]			
	<i>Nymphaea daubeniana</i> Hort. ex O. Thomas.	28 [1]			
	<i>Nymphaea dentatamagnifica</i> Bisset.	42 [1]			
	<i>Nymphaea gardneriana</i> Planch.	14 [1]			
	<i>Nymphaea gigantea</i> Hook.	112 [1]	2709.06 [2]		
	<i>Nymphaea heudelotii</i> Planch.	14 [1]			
	<i>Nymphaea immutabilis</i> S. W. L. Jacobs	42 [1]	1408.32 [2]		
	<i>Nymphaea jamesoniana</i> Planch.	14 [1]		ITS2+matK [9]	
	<i>Nymphaea japono-koreana</i> Nakai	56 [1]			
	<i>Nymphaea lasiophylla</i> Mart. & Zucc.	9 [1]			
	<i>Nymphaea lingulata</i> Wiersema	9 [1]			
	<i>Nymphaea lotus</i> L.	28 [1]	1779.96/1682.16 [2]	ITS2+matK [9], trnH-psbA, rpoC1 [10]	
	<i>Nymphaea mexicana</i> Zucc.	28 [1]	586.80 [2]		
	<i>Nymphaea micrantha</i> Guill. & Perr.	14 [2]	889.98 [2]	ITS2+matK [9]	
	<i>Nymphaea minuta</i>	14 [2]	449.88 [2]		
	<i>Nymphaea nouchali</i> Burm.	38 [1], 42 [2]	1193.16 [2]	ITS2+matK [9], trnH-psbA, rpoC1 [10]	
	<i>Nymphaea nouchali</i> var. <i>caerulea</i> (Savigny) Verdc.	14 [1]	567.24 [2]		
	<i>Nymphaea novogranatensis</i> Wiersema	14 [1]		ITS2+matK [9]	
	<i>Nymphaea odorata</i> Aiton	28 [1], 56 [2]	1574.58 [2]	ITS2+matK [9]	
	<i>Nymphaea oxypetala</i> Planch.	42 [1]		ITS2+matK [9]	
	<i>Nymphaea pubescens</i> Willd.	28 [2]	1975.56 [2]		
	<i>Nymphaea prolifera</i> Wiersema	9 [1]			
	<i>Nymphaea pubescens</i> Willd.	12 [1]		ITS2+matK [9]	
	<i>Nymphaea rubra</i> Roxb. ex Andrews	42 [1]		ITS2+matK [9]	
	<i>Nymphaea rudgeana</i> G. Mey.	21 [1]	792.18 [2]		
	<i>Nymphaea stellata</i> var. <i>versicolor</i> (Sims) Hook. & Thomson	28 [1]			
	<i>Nymphaea sturtevantii</i> J. N. Gerard	28 [1]			
	<i>Nymphaea tenerinervia</i> Casp.	10 [1]			
	<i>Nymphaea tetragona</i> Georgi	42 [1]		ITS2+matK [9]	
	<i>Nymphaea tetragona</i> subsp. <i>Leibergii</i> / <i>Nymphaea leibergii</i>	56 [1]			
	<i>Nymphaea thermarum</i> E. Fisch.	14 [1]	498.78 [2]		
	<i>Nymphaea tuberosa</i> / <i>Nymphaea odorata</i> subsp. <i>Tuberosa</i>	42 [1]			
	<i>Nymphaea violacea</i>	56 [2]	1770.18 [2]		
	<i>Euryale ferox</i>	29 [1]	870.42 [2]	ITS2+matK [9]	
	<i>Nuphar advena</i> (Aiton) W. T. Aiton	17 [1]	2709.06/2718.84 [2]	ITS2+matK [9]	y [11]
<i>Nuphar lutea</i> (L.) Sm.	17 [1]	2875.32 [2]	ITS2+matK [9]		
<i>Nuphar microphylla</i> (Pers.) Fernald	17 [1]		ITS2+matK [9]		
<i>Nuphar polysepalum</i> Engelm.	17 [1]	3080.70/3070.92 [2]	ITS2+matK [9]		
<i>Nuphar pumila</i> (Timm) DC.	17 [1]		ITS2+matK [9]		
<i>Nuphar variegata</i> Durand	17 [1]		ITS2+matK [9]		
<i>Nuphar</i> × <i>spenneriana</i> Gaudin	17 [1]	2581.92 [2]			
<i>Nuphar intermedia</i> Ledeb.	17 [1]				
<i>Nuphar japonica</i> DC.	17 [1]	2699.28 [2]	ITS2+matK [9]		
<i>Nuphar subintegerrima</i> Makino	17 [1]		ITS2+matK [9]		
<i>Barclaya longifolia</i>	17 [1]		ITS2+matK [9]		
<i>Victoria cruziana</i>	12 [1]	4009.80 [2]	ITS2+matK [9]		
<i>Victoria yamasu</i>	12 [1]				
<i>Victoria amazonica</i>	10 [1]	4557.48 [2]	ITS2+matK [9]		
<i>Victoria</i> 'Longwood Hybrid'	11 [2]	4303.20 [2]	ITS2+matK [9]		
Cabombaceae	<i>Cabomba caroliniana</i>	52 [2]	3471.9 [2]	ITS2+matK [9]	
	<i>Brasenia schreberi</i>	36 [3], 40 [1], [2]	1193.16 [2]	ITS2+matK [9]	
Hydatellaceae	<i>Trithuria submersa</i>	28 [1]	~ 2680 [7]	ITS2+matK [9]	y [12]
	<i>Trithuria inconspicua</i>	12 [1]		ITS2+matK [9]	
	<i>Trithuria konkanensis</i>	20 [4]			
	<i>Trithuria australis</i>	7 [5]		ITS2+matK [9]	

[1] <http://ccdb.tau.ac.il/> [2] Pellicer *et al.*,¹⁷ [3] Diao *et al.*,¹⁶ [4] Gaikward *et al.*,⁵² [5] Iles *et al.*,⁵³ [6] Viallette-Guiraud *et al.*,⁴⁷ [7] Kynast *et al.*,⁵⁴ [8] Yamada *et al.*,¹⁸ [9] Biswal *et al.*,⁴ [10] Chaveerach *et al.*,²² [11] <http://sra.dnanexus.com> [12] Marques *et al.*,⁵⁵

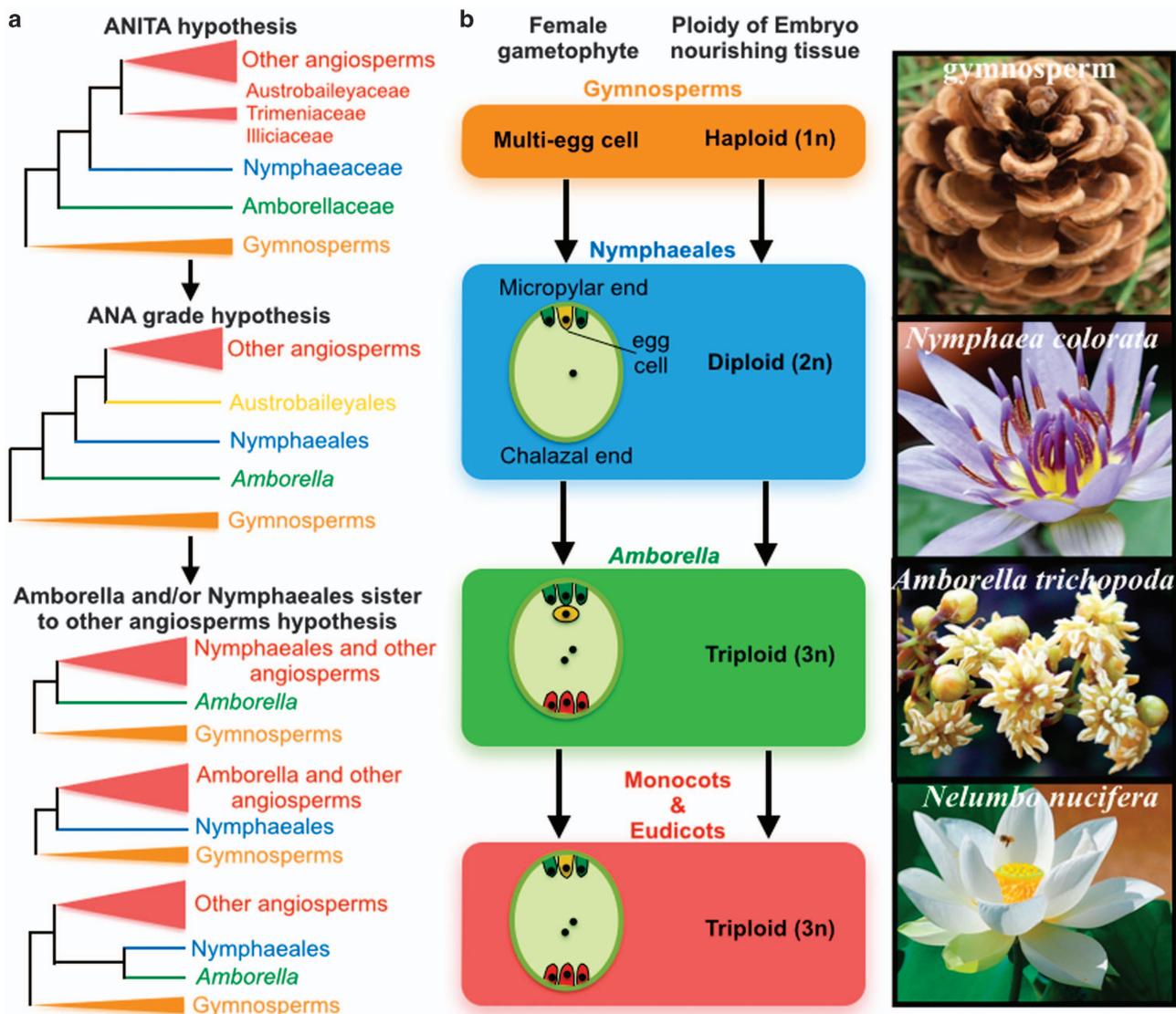


Figure 2. Phylogenetic uncertainty among *Amborella*, water lily, and other angiosperms. (a) Hypothesized phylogenetic relationships of basal angiosperms. (b) Developmental evidence suggests water lily as the most basal angiosperm.

nuclear genes and one phylogenomic analysis using plastid and mitochondria genes.

Water lilies and Amborella as the basal sister to all other angiosperms. In other studies, *Amborella* and water lilies have been thought to form sister groups that both represent the first lineage to all other angiosperms. Relying on both nuclear and plastid genes, Xi and colleagues in 2014 placed *Amborella* and water lilies as sister groups using the coalescent-based phylogenetic method, and these sister groups serve as the most basal angiosperm clade³² (Figure 2).

Water lilies as the most basal angiosperms. There is still evidence to support water lilies as the most basal angiosperm. Relying on concatenation-based phylogenetic analysis of the whole chloroplast coding genes and using the transversion of the third position of the codon, researchers found that the water lily was the earliest branch of all extant angiosperms.¹³ A comparison of the female gametophyte and the embryo-nourishing tissue ploidy also suggested that *Amborella* was an exception in the ANITA group, which contained triploid endosperm and nine cells in the

embryo sac and is thereby closest to monocots and eudicots^{13,33} (Figure 2). In addition, water lilies contain fewer stomatal modifications from the ancestral angiosperm stomata, whereas *Amborella* exhibited extensive modifications of stomata.³⁴ In addition, the first known fossil flower of a water lily is from the early cretaceous period, approximately 125–115 million years ago.³⁵ Another Jurassic fossil with flowers and other above-ground organs including the archaefructus is also placed within Nymphaeales.⁶

Phylogenetic signals hold the key for basal angiosperm phylogeny. A major concern in phylogenomics is the selection of the best phylogenetic signals, which are now generally regarded to be low/single-copy nuclear genes³⁶ that should fulfill two important criteria: high neutrality and low saturation.¹³ For the selected genes, position 1 and position 2 codons lack synonymous mutation rates and suffer extremely low neutrality.¹³ Researchers found that position 3 transversion rates are suitable for both shallow and deep phylogenetic tree constructions.¹³ Based on this position 3 transversion, most single-gene-based trees placed the water lily as the most basal lineage of angiosperms.¹³ For species tree construction for angiosperms, we suggest the utilization of

both protein sequences and nucleotide sequences as a more accurate method for land plant species tree construction.¹⁰ Most importantly, phylogenetic signals for species tree reconstruction should be genome-wide and contain a large number of signals but not rare genes or a limited number of signals.

Concatenation VS coalescent methods. In recent years, phylogenomics have relied on both the concatenation method and coalescent method.^{14,32} Although the coalescent method is theoretically sound to explain incomplete lineage sorting, both theories and applications show that concatenation could yield misleading results when highly conflicting gene trees exist, due to incomplete lineage sorting. These two methods have been under heated debate regarding the effects of tree estimation error in phylogenomics.³⁷⁻⁴¹ Strong phylogenetic signals are still needed for more accurate species tree inference.⁴²

Until recently, our understanding of plant phylogeny has largely depended on studies of plastid, mitochondrial, and ribosomal genes. However, recently, using large-scale comparisons of dozens of genes comprising thousands of DNA bases, phylogenomics have reshuffled most of our long-established trees of life, such as trees for eukaryotic life,⁴³ bird life,⁴⁴ fish life,⁴⁵ and major nodes of eudicots of plants.⁴⁶ The availability of the *Amborella* genome²¹ has shed light on basal angiosperm tree construction, but definite resolution of basal angiosperm phylogeny has not been resolved.

Pantropical water lilies could serve as the model for studying basal angiosperms

To understand basal angiosperm evolution and the radiation of angiosperms, a good model species is needed. Among all basal angiosperms, the enormous genome size for Austrobaileyales, ~7050 Mb,⁴⁷ is a major challenge for genome decoding and genetic experiments. A slow growth rate and the woodiness of Austrobaileyale plants and *Amborella* may also be challenging due to the difficulty of producing experimental materials. In the water lily order, *Cabomba* displays multiple features as a model for

basal angiosperms, such as small size and rapid vegetative growth, but its large genome size, 3290 Mb,⁴⁷ excludes it from gene functional studies, as large genomes usually harbor redundant gene copies, are highly heterogenous and thereby not appropriate for gene functional studies. Although *Trithuria* species grow into small herbs, their genomes are still too large for genetic studies. Luckily, four pantropical diploid ($2n=28$) water lilies (or subgenus *Brachyceras*) may be good choices, as they have the smallest genomes, *N. caerulea* = 567.24 Mb, *N. colorata* = 489 Mb, *N. minuta* = 449.88 Mb, *N. thermanum* = 498.78 Mb.¹⁷ The native habitats of all four water lilies are in Africa, and all are annual plants (Table 2). *N. caerulea* and *N. colorata* are famous ornamental water lilies and have been widely used to breed new cultivars. *N. minuta* and *N. thermanum* are minute water lilies with thumb-sized flowers. Unlike hardy water lilies (a in Figure 1), these four tropical water lilies are easy to cultivate and maintain hundreds of plants in a single green house; it is easy to trigger flowering via temperature control (below 18 °C). All four water lilies can produce hundreds of seeds in a single flower (Figure 3) and can be used to generate a large mutant library. These plants are also easy to self-pollinate in nature to generate pure lines, and can also easily be cross-pollinated. They have a relatively short life cycle of approximately three months from seed to seed in tropical regions. In addition, *N. thermanum* has recently been well studied for its potential as a model system for basal angiosperms.⁴⁸ These characteristics make these four water lilies the best candidates for genome sequencing and the best model for functional studies.

The water lily genome for basic evolutionary research and applied horticulture

Based on the advantages of pantropical water lilies, we launched a genome-sequencing project of *N. colorata* using a third-generation single-molecule real-time sequencing method. We produced half of the reads > 20 kb, and this has facilitated the assembly of complex repeating sequences and GC-rich regions

Table 2. Characteristics of four pantropical water lilies

Species	Genome size	Chromosome	Classification	Distribution	Flowers in diameter
<i>Nymphaea caerulea</i>	567.24 Mb	$2n=28$	<i>Nymphaea</i> , <i>Brachyceras</i>	East Africa	10–15 cm
<i>Nymphaea colorata</i>	489 Mb	$2n=28$	<i>Nymphaea</i> , <i>Brachyceras</i>	tropical East Africa	11–14 cm
<i>Nymphaea minuta</i>	449.88 Mb	$2n=28$	<i>Nymphaea</i> , <i>Brachyceras</i>	Madagascar	2 cm
<i>Nymphaea thermanum</i>	498.78 Mb	$2n=28$	<i>Nymphaea</i> , <i>Brachyceras</i>	Rwanda, Africa	10–15 cm

The four listed genome sizes and chromosome counts have been reported.¹⁷

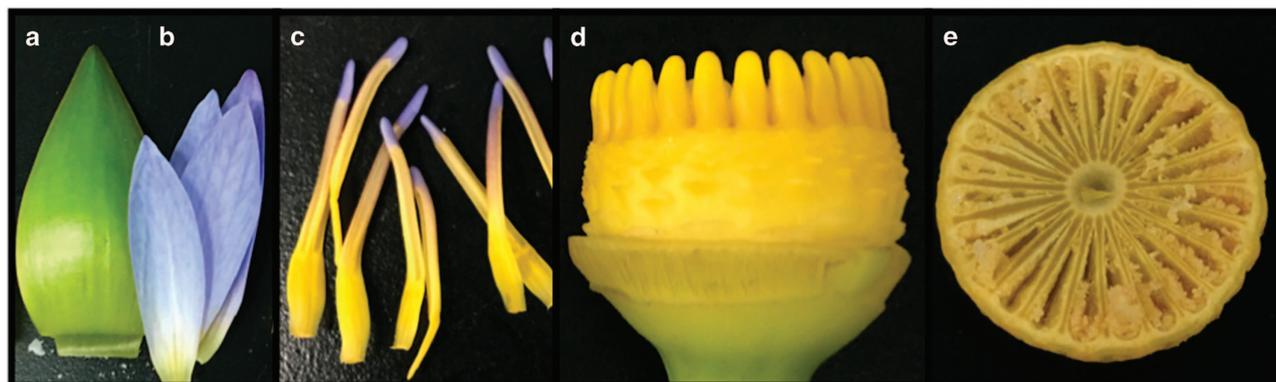


Figure 3. Floral organs of a typical tropical water lily. (a) petal, (b) sepal, (c) stamen, (d) carpels on the receptacle, (e) numerous young seeds.

Table 3. Sequenced aquatic plants

Scientific name	Common name	Classification	Ornamental plant	Habitat	Genome size
<i>Nymphaea</i> spp.	Water lilies	Basal angiosperms	Partially yes	Floating leaf type	~400 Mb–2.7 Gb
<i>Nelumbo nucifera</i>	Sacred lotus	Basal eudicot	Yes	Emerged plant	929 Mb
<i>Utricularia gibba</i>	Floating bladderwort	Eudicot, Lentibulariaceae	No	Floating plant	82 Mb
<i>Zostera marina</i>	Seagrass	Monocot, Zosteraceae	No	Submerged plant	202.3 Mb
<i>Spirodela polyrhiza</i>	Duckweed	Monocot, Araceae	No	Free-floating	158 Mb
<i>Zizania latifolia</i>	Manchurian wildrice	Monocot, Poaceae	No	Wetland plant	586 Mb
<i>Oryza sativa</i>	Rice	Monocot, Poaceae	No	Wetland plant	420 Mb

The genome of *Nymphaea colorata* was sequenced recently by our team. Other genomes have been sequenced and are publicly available.

that are usually highly fragmented or even unassembled in next-generation sequencing projects.⁴⁹ We have annotated the genes and other key DNA elements using multiple tools. The future reference water lily genome will provide genomic information for reconstructing the karyotype of an angiosperm ancestor, with the species trees of basal angiosperms and early massive radiation of angiosperms. This availability of the water lily reference genome will greatly help us to understand Charles Darwin's 'abominable mystery', the early evolution trajectory of angiosperms, the aquatic life style of angiosperms, the evolution of a 4-celled embryo sac and diploid endosperm, the comparative analyses of genes and other elements such as conserved non-coding elements and telomeres. The water lily genome is also needed to revisit the age of angiosperms and whether they evolved 0.1 billion-years ago⁵⁰ or 0.2 billion years ago.⁵¹ The reference genome will also provide genetic information for breeders and geneticists. Currently, only seven aquatic plants have their genomes decoded, and only two aquatic ornamental plants, the water lily and the sacred lotus, have sequenced genomes (Table 3). The similar appearance of the water lily and the lotus does not actually indicate a tight relationship; the former is a basal angiosperm and the latter is a eudicot. Thus, the genome of the water lily will serve as a template to accelerate genomic studies of other aquatic ornamentals.

CONCLUSIONS

Upgrading sequencing technologies and bioinformatics tools have provided high-resolution genomic details, showing great potential for understanding the large questions in biology (including Darwin's famous abominable mystery), and are valuable resources for molecular breeding. Although the genetics and genomics of water lilies are incipient, four pantropical water lilies, especially *N. colorata*, show great potential as a model system to study basal angiosperms; their genomes will greatly enhance our current knowledge, including Charles Darwin's abominable mystery, the early evolutionary trajectory of angiosperms, the aquatic life style of angiosperms, and molecular breeding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (81502437) and a start-up fund from the Fujian Agriculture and Forestry University to LZ. We acknowledge Dr Xianxian Yu from Xuchang University for providing the *N. colorata* picture in Figure 1.

REFERENCES

- Raja MKMM, Sethiya NK, Mishra SH. A comprehensive review on *Nymphaea stellata*: a traditionally used bitter. *J Adv Pharm Technol Res* 2010; **1**: 311–319.
- Tani M, Sawada A, Oyabu T, Seiryu K. Ability of water lilies to purify water polluted by soap and their application in domestic sewage disposal facilities. *Sens Mater* 2006; **18**: 91–101.
- Friedman WE. Hydatellaceae are water lilies with gymnospermous tendencies. *Nature* 2008; **453**: 94–97.
- Biswal DK, Debnath M, Kumar S, Tandon P. Phylogenetic reconstruction in the Order Nymphaeales: ITS2 secondary structure analysis and in silico testing of maturase k (*matK*) as a potential marker for DNA bar coding. *BMC Bioinformatics* 2012; **13**: S26.
- Les DH, Schneider EL, Padgett DJ, Soltis PS, Soltis DE, Zanis M. Phylogeny, classification and floral evolution of water lilies (Nymphaeaceae; Nymphaeales): a synthesis of non-molecular *rbcl*, *matK*, and 18S rDNA data. *Syst Bot* 1999; **24**: 28–46.
- Gandolfo MA, Nixon KC, Crepet WL. Cretaceous flowers of Nymphaeaceae and implications for complex insect entrapment pollination mechanisms in early angiosperms. *Proc Natl Acad Sci USA* 2004; **101**: 8056–8060.
- Soltis D, Soltis P, Nickrent D *et al.* Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann Miss Bot Gard* 2015; **71**: 607–630.
- Qiu Y, Lee J, Bernasconi-quadroni F, Zimmer EA, Chen Z, Savolainen V. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 1999; **402**: 404–407.
- Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat Commun* 2014; **5**: 4956.
- Wickett NJ, Mirarab S, Nguyen N *et al.* Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc Natl Acad Sci USA* 2014; **111**: E4859–E4868.
- Drew BT, Ruhfel BR, Smith SA *et al.* Another look at the root of the angiosperms reveals a familiar tale. *Syst Biol* 2014; **63**: 368–382.
- Leebens-Mack J, Raubeson LA, Cui L *et al.* Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Mol Bio Evol* 2005; **22**: 1948–1963.
- Yang X, Tuskan GA, Tschaplinski TJ, Cheng Z-M. Third-codon transversion rate-based *Nymphaea* basal angiosperm phylogeny -- concordance with developmental evidence. *Nat Preced* 2007, 1–20.
- Simmons MP, Gatesy J. Coalescence vs. concatenation: sophisticated analyses vs. first principles applied to rooting the angiosperms. *Mol Phylogenet Evol* 2015; **91**: 98–122.
- Goremykin VV, Nikiforova SV, Cavalieri D, Pindo DM, Lockhart P. The root of flowering plants and total evidence. *Syst Biol* 2015; **64**: 879–891.
- Diao Y, Chen L, Yang G *et al.* Nuclear DNA C-values in 12 species in Nymphaeales. *Caryologia* 2006; **59**: 25–30.
- Pellicer J, Kelly LJ, Magdalena C, Leitch IJ. Insights into the dynamics of genome size and chromosome evolution in the early diverging angiosperm lineage Nymphaeales (water lilies). *Genome* 2013; **56**: 437–449.
- Yamada T, Ito M, Kato M. Expression pattern of INNER NO OUTER homologue in *Nymphaea* (water lily family, Nymphaeaceae). *Dev Genes Evol* 2003; **213**: 510–513.
- Luo H, Chen S, Wan H, Chen F, Gu C, Liu Z. Candidate reference genes for gene expression studies in water lily. *Anal Biochem* 2010; **404**: 100–102.
- Luo H, Chen S, Jiang J *et al.* The expression of floral organ identity genes in contrasting water lily cultivars. *Plant Cell Rep* 2011; **30**: 1909–1918.
- Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. *Science* 2013; **342**: 1241089.
- Chaveerach A, Taneer T, Sudmoon R. Molecular identification and barcodes for the genus *Nymphaea*. *Acta Biol Hungarica* 2011; **62**: 328–340.

- 23 Cui L, Wall PK, Leebens-mack JH *et al.* Widespread genome duplications throughout the history of flowering plants. *Genome Res* 2006; **16**: 738–749.
- 24 Yoo MJ, Chanderbali AS, Altman NS, Soltis PS, Soltis DE. Evolutionary trends in the floral transcriptome: Insights from one of the basalmost angiosperms, the water lily *Nuphar advena* (Nymphaeaceae). *Plant J* 2010; **64**: 687–698.
- 25 Wu Q, Wu J, Li S-S *et al.* Transcriptome sequencing and metabolite analysis for revealing the blue flower formation in waterlily. *BMC Genomics* 2016; **17**: 897.
- 26 Friedman WE. The meaning of Darwin's 'abominable mystery'. *Am J Bot* 2009; **96**: 5–21.
- 27 Soltis PS, Soltis DE, Chase MW. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 1999; **402**: 402–404.
- 28 Kuzoff RK, Gasser CS. Recent progress in reconstructing angiosperm phylogeny. *Trends Plant Sci* 2000; **5**: 330–336.
- 29 Qiu Y, Dombrowska O, Lee J *et al.* Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *Int J Plant Sci* 2005; **166**: 815–842.
- 30 Delsuc F, Brinkmann H, Philippe H. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 2005; **6**: 361–375.
- 31 Goremykin VV, Nikiforova SV, Biggs PJ *et al.* The evolutionary root of flowering plants. *Syst Biol* 2013; **62**: 50–61.
- 32 Xi Z, Liu L, Rest JS, Davis CC. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Syst Biol* 2014; **63**: 919–932.
- 33 Friedman WE. Embryological evidence for developmental lability during early angiosperm evolution. *Nature* 2006; **441**: 337–340.
- 34 Carpenter KJ. Stomatal architecture and evolution in basal angiosperms. *Am J Bot* 2005; **92**: 1596–1615.
- 35 Friis E, Pedersen K, Crane PR. Fossil evidence of water lilies (Nymphaeales) in the early cretaceous. *Nature* 2001; **410**: 357–360.
- 36 Sang T. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Crit Rev Biochem Mol Biol* 2002; **37**: 121–147.
- 37 Betancur-R R, Naylor GJP, Ortí G. Conserved genes, sampling error, and phylogenomic inference. *Syst Biol* 2014; **0**: 1–6.
- 38 Mirarab S, Bayzid MS, Boussau B, Warnow T. Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* 2014; **346**: 1250463.
- 39 Liu L, Edwards SV. Comment on 'Statistical binning enables an accurate coalescent-based estimation of the avian tree'. *Science* 2014; **350**: 171.
- 40 Springer MS, Gatesy J. The gene tree delusion. *Mol Phylogenet Evol* 2016; **94**: 1–33.
- 41 Edwards SV, Xi Z, Janke A *et al.* Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Mol Phylogenet Evol* 2016; **94**: 447–462.
- 42 Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 2013; **497**: 327–331.
- 43 Burki F, Shalchian-Tabrizi K, Minge M *et al.* Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2007; **2**: e790.
- 44 Prum RO, Berv JS, Dornburg A *et al.* A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 2015; **526**: 569–573.
- 45 Ricardo B-R, Broughton RE, Wiley EO *et al.* The tree of life and a new classification of bony fishes. *PLOS Curr Tree Life* 2013, 1–52.
- 46 Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytol* 2017; **214**: 1338–1354.
- 47 Vialette-Guiraud ACM, Alaux M, Legeai F *et al.* *Cabomba* as a model for studies of early angiosperm evolution. *Ann Bot* 2011; **108**: 589–598.
- 48 Povilus RA, Losada JM, Friedman WE. Floral biology and ovule and seed ontogeny of *Nymphaea thermarum*, a water lily at the brink of extinction with potential as a model system for basal angiosperms. *Ann Bot* 2015; **115**: 211–226.
- 49 VanBuren R, Bryant D, Edger PP *et al.* Single-molecule sequencing of the desiccation-tolerant grass *Oropetium thomaeanum*. *Nature* 2015; **527**: 508–511.
- 50 Buggs RJA. The deepening of Darwin's abominable mystery. *Nat Eco Evol* 2017; **1**: 169.
- 51 Herendeen PS, Friis EM, Pedersen KR, Crane PR. Palaeobotanical redux: revisiting the age of the angiosperms. *Nat Plants* 2017; **3**: 17015.
- 52 Gaikward SP, Yadav SR. Further morphotaxonomical contribution to the understanding of the family Hydatellaceae. *J Swamy Bot* 2003; **20**: 1–10.
- 53 Iles WJD, Rudall PJ, Sokoloff DD *et al.* Molecular phylogenetics of Hydatellaceae (Nymphaeales): sexual-system homoplasy and a new sectional classification. *Am J Bot* 2012; **99**: 663–676.
- 54 Kynast RG, Joseph JA, Ramsay MM, Rudall PJ. Chromosome behavior at the base of the angiosperm radiation: karyology of *Trithuria submersa* (Hydatellaceae, Nymphaeales). *Am J Bot* 2014; **101**: 1447–1455.
- 55 Marques I, Montgomery SA, Barker MS *et al.* Transcriptome-derived evidence supports recent polyploidization and a major phylogeographic division in *Trithuria submersa* (Hydatellaceae, Nymphaeales). *New Phytol* 2016; **210**: 310–323.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017