MDPI

Review

Studies on Lotus Genomics and the Contribution to Its Breeding

Huanhuan Qi [ID]**, Feng Yu** [ID]**, Jiao Deng** [†] **and Pingfang Yang *** [ID]

State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Science, Hubei University,
Wuhan 430062, China; qihuanhuan0911@163.com (H.Q.); yufeng@hubu.edu.cn (F.Y.); ddj613@163.com (J.D.)
* Correspondence: yangpf@hubu.edu.cn; Tel.: +86-27-88661237
† Current Address: Research Center of Buckwheat Industry Technology, School of Life Sciences,
  Guizhou Normal University, Guiyang 550001, China.

Abstract: Lotus (*Nelumbo nucifera*), under the Nelumbonaceae family, is one of the relict plants possessing important scientific research and economic values. Because of this, much attention has been paid to this species on both its biology and breeding among the scientific community. In the last decade, the genome of lotus has been sequenced, and several high-quality genome assemblies are available, which have significantly facilitated functional genomics studies in lotus. Meanwhile, re-sequencing of the natural and genetic populations along with different levels of omics studies have not only helped to classify the germplasm resources but also to identify the domestication of selected regions and genes controlling different horticultural traits. This review summarizes the latest progress of all these studies on lotus and discusses their potential application in lotus breeding.

Keywords: lotus; genome; variant; germplasm; breeding; omics

1. Introduction

Food shortage has become a rising challenge with the increase of the world's population and decrease of natural resources. It is incredibly significant to breed crops with high yields, of good quality, and high-stress resistance to ascertain food security since crops provide a staple food supply for the world. To achieve this, it is necessary to obtain a deeper understanding of the crops' genetic background, especially their genome information. Since the first flower plant, *Arabidopsis thaliana* genome was sequenced in 2000 [1], more and more plant whole genomes have been sequenced and deposited in databases, which are available to the public [2], including *Nelumbo* genome database (http://nelumbo.biocloud.net/nelumbo/home) [3]. Third-generation sequencing, which can produce long sequence read, has shown its advantages over next-generation sequencing (NGS) in generating high continuity reference genome assemblies [4].

Genomics is the cornerstone of breeding, and studies based on whole-genome sequencing and genome-wide association study have greatly driven forward genomics-assisted breeding in many crops [5,6]. Cloning and functional analysis of genes associated with important agronomic traits in rice (*Oryza Sativa*), soybean (*Glycine max*), and tomato (*Solanum lycopersicum*) have also demonstrated that high-quality genomes are prerequisite to clarify variations in each species [7–11]. However, population genetic analysis relied on a single reference genome that lost variant information, especially in the highly polymorphic region. Pan-genome contains the totality of genome sequence information of the target species and covers more comprehensive variant information. Pan-genomes have been constructed in many plants, such as rice, maize, brassica, and soybean, and applied to identify causal genes [12–15]. Pan-genome or graph pan-genome is obtaining new references along with the upgrading of sequencing. The information on genome maps, domestication, improvement-related genes, and regulation pathway promotes the understanding of plant evolution and accelerate breeding [15].

Lotus is one of the relict plants retaining the original morphology of its ancestors, as well as *Ginkgo biloba*, *Liriodendron*, and *Metasequoia glyptostroboides*. It belongs to the

Nelumbo genus of the Nelumbonaceae family, which includes two species, namely Asian lotus (*Nelumbo nucifera* Gaertn.) and American lotus (*Nelumbo lutea* Pear.). The two species are named for their different geographical distributions. Asian lotus is mainly distributed in Asia and the north of Oceania, while American lotus is distributed in North America and South America. The plant morphology differs between them. Asian lotus is a tall plant, with oval leaves and seeds, and red or white flower colors, whereas American lotus is a short plant, nearly round and with dark green leaves, spherical seeds, and yellow flowers [16]. There is no strict reproductive isolation between them, and the life cycles are similar at about five months. Asian lotus is commonly called lotus and has more than 3000 years of cultivation history as a horticultural crop [17]. Lotus seeds and rhizomes have rich nutritional value and unique health-care function. Lotus seeds contain starch, proteins, amino acids, polysaccharides, polyphenols, alkaloids, and mineral elements. Lotus rhizome has a high vitamin C content. During the long period of domestication and artificial selection, about 4500 lotus cultivars have been obtained up till the present [18]. These cultivars have been planted to produce edible vegetables, snacks, beverages, restorative materials, and ornamental flowers, which impact human life and economic development. The lotus industry is also important for rural revitalization in the Yangtze River, Pearl River, and Huang Huai river basins. The cultivated lotus is generally divided into rhizome lotus, seed lotus, and flower lotus based on their different usage. The notable feature of the rhizome lotus is the enlarged rhizome but with few flowers. It can be divided into power and crisp type according to the taste of the rhizome. Different varieties were bred to meet the taste of the different regions of people or for further usage. The main breeding goal of rhizome lotus is to improve the yield and quality of the rhizome. Seed lotus is mainly for lotus seed production, with high yield, good quality, and disease resistance being the breeding goals. Flower lotus is preferred for ornamental use, and it has distinct flower colors and shapes. During long cultivation, ornamental lotus with different flower morphologies were obtained, including few-petaled, double-petaled, petaloidy, and thousand-petalled flowers. Red, pink, yellow, and white are the main flower colors. Currently, the breeding objective is mainly aimed at flower shape and color, yield or quality of lotus seed and rhizome, and wide adaptability.

As a basal eudicot species, lotus plays an essential role in studying plant evolution and phylogeny. It is adapted to the aquatic environment, while its relatives are shrubs or trees living on land. Water lily lies at the phylogenetic position of the base angiosperm and has similar living conditions and flowers. However, its genomes are vastly different [19]. Lotus has unique features such as water-repellent self-cleaning function, multi-seed production, and flower thermogenesis, which may relate to flower protogyny or provide a warm environment for pollination [16]. Because of its importance in plant phylogeny and wide application, lotus has gained increasing attention from the scientific community. Since the release of the first version of two lotus reference genomes [20,21], genome-based investigations have been conducted continuously. Subsequently, the high-resolution genetic map and BioNano optical map were applied to improve the accuracy and assembly of the lotus genome [22]. A hybrid assembly was completed using PacBio sequencing data and previously published short reads [23]. High-quality genome assembly of "Taikonglian NO. 3" and American lotus genome were also recently generated [24,25]. High-throughput re-sequencing of different lotus cultivars has been utilized to identify numerous molecular markers, promoting marker-assisted selection. Moreover, "omics" approaches such as transcriptomics, proteomics, and metabolomics were applied in elucidating molecular regulatory networks of yield, quality, and response to stress in lotus. Here, we briefly review the latest progress of studies on the lotus genome, and how genome information could be used in lotus breeding. Meanwhile, the existing challenge and potential prospects are also discussed.

2. Sequencing, Assembly and Annotation of Lotus Genome

Lotus occupies a crucial phylogenetic position in flowering plants. The high-quality reference genome of lotus plays a vital role in studying the origin of eudicot and lotus molecular breeding. In the last decade, some lotus varieties were sequenced by different platforms, which resulted in a different version of the genome assembly and annotation (Table 1). Based on NGS, a wild lotus, "China antique (CA)", was successfully sequenced and assembled [20]. The total sequenced genome length of "CA" is 804 Mb, of which 543.4 Mb (67.6%) were anchored to nine megascaffolds. The contig N50 was 38.8 Kb and the scaffold N50 was 3.4 Mb. The heterozygosity of "CA" genome is 0.03%, and the repetitive sequence is about 57%. A total of 26,685 protein-coding genes were predicted, with the average length of a gene being 6561 bp. Simultaneously, another wild strain of lotus, "Chinese Taizi" was assembled through NGS technology. The final assembled genome size is 792 Mb with the contig N50 39.3 Kb and scaffold N50 986.5 Kb [21]. The length of transposable elements is 392 Mb (49.48%), and 36,385 protein-coding genes were annotated. One WGD event -λ in lotus instead of the paleo-hexaploid arrangement (γ WGD) event that occurred in core eudicots was predicted [20,21]. These two genomes were further anchored to eight pseudo-chromosomes by constructing a higher resolution genetic map and physical maps [22].

Table 1. Comparison of assembled lotus genomes.

Items	Year	Sequencing Technology	Final Assembly (Mb)	Contig N50	Number of Genes	Repeat Sequences	Ref.
China Antique v1.0	2013	Illumina, 454	804	38.8 Kb	26685	57%	[20]
Taizi	2013	Illumina Hiseq2000	792	39.3 Kb	40348	49.48% (TEs)	[21]
China Antique v2.0	2020	Pacbio Sequel, Illumina	821.2	484.3 Kb	32124	58.50%	[23]
Taikonglian NO.3	2022	Nanopore	807	5.1 Mb	28274	63.11%	[24]
American lotus	2022	Pacbio RSII, Hi-C	843	1.34 Mb	31382	81.00%	[25]

With the advent of a new sequencing platform, the genome of "CA" was re-assembled using 11.9 Gb long-read data from PacBio Sequel, and 94.2 Gb previously sequenced short-read data [23]. The new assembly of "CA" is 807.6 Mb with the contig N50 being 484.3 Kb, which has significantly increased the quality of the genome. The ratio of repetitive sequence (58.5%) was similar to the first version. Moreover, a cultivated lotus, "Taikonglian NO. 3", was also assembled using the Oxford Nanopore sequencing platform (57.9 Gb raw data) with the contig N50 being 5.1 Mb, and eight chromosomes were anchored based on high-throughput chromatin conformation capture (Hi-C) data [24]. Another lotus species, American lotus, was recently assembled using PacBio RSII (74.6 Gb raw data) and Hi-C (50.32 Gb raw data), and the total length is 843 Mb while contig N50 is 1.34 Mb [25]. These data demonstrate that long-read sequencing technology has greatly improved the quality of the lotus genome. The successful assembly of the genome in Asian lotus, including wild and cultivar varieties, and American lotus will assist the investigation of functional genomics as well as molecular breeding in lotus.

3. Study on the Potential Adaptive Evolution and Domestication of Lotus

The availability of lotus reference genome information has facilitated the resequencing of different lotus germplasms. Several studies were conducted on how the lotus genome was subjected to adaptive evolution and artificial selection. Although it is known that there are only two species of lotus, namely Asian lotus and American lotus, except for the difference in flower color, their plant architecture and morphology are very similar. Based on molecular phylogeny analysis, significant genetic differentiation between American and Asian lotus was verified [25–28]. De-novo deep sequencing of the American lotus showed that its genome size is 843 Mb, and an approximate 81% repeat sequence was identified (Table 1), which is larger than the genome of Asian lotus. It is interesting to investigate the dramatic difference in repeat sequence between them because most protein-

coding genes show a one-to-one synteny pattern. A total of 29,533 structure variations (SVs) were detected between two lotus species, with the SV-associated genes overexpressed in 'regulation of mitotic cell cycle', and 'protein transporter activity' [25]. Meanwhile, this study also showed that the selection on an *MYB* gene might contribute to the color difference between Asian and American lotus [25]. It is still an open question about when the two species diverged during the evolution and how they could keep high similarity in the independent geographical evolution. The wild lotus is distributed widely worldwide and maintains higher genomic diversity than cultivated lotus. Tropical and temperate lotus are the two ecotypes of Asian lotus. The comparison of the genome of these two ecotypes showed that a total of 453 genes were subjected to selection, including *cyp714a* genes that may relate to rhizome morphogenesis and a 10-Mb region in chromosome 1 that might play key roles in environmental adaption; including a homolog gene of *at5g2394* in *Arabidopsis* encoding an acyltransferase protein [24]. By comparing their expressional patterns, the genes encoding granule-bound starch synthases, storage organ development, *COSTAN-like* gene family, vernalization, as well as cold response genes may relate to ecotypic differentiation [26].

It is very important to know the genetic backgrounds of parental lines in breeding. The origin, classification, and evolution of cultivated lotus were investigated through population re-sequence analysis. A total of 18 lotus accessions, including categories of American, seed, rhizome, flower, wild, and Thai lotus, were re-sequenced, based on which phylogenetic tree was constructed. The results indicated that the rhizome lotus had a closer relationship with wild lotus. In contrast, seed and flower lotus were admixed [26], which could be supported by re-sequencing of an enlarged population containing 296 accessions of different germplasm (58 wild, 163 rhizome, 39 flower, 32 seed lotus varieties) [28]. Further re-sequencing of 69 lotus accessions showed that flower lotus might mix with rhizome or seed lotus [27]. All studies showed a low genetic variation in rhizome lotus, while higher genetic diversity in seed lotus. The origin of different subgroups is controversial, which is possible because the same accession of lotus has other names which were then divided into different subgroups by various people. Based on this genomic diversity, the potential domestication signals of cultivar subgroups could be speculated because the selected genomic regions had lower nucleotide diversity. When subgroups of seed, rhizome, and flower lotus were compared with wild lotus subgroup, a total of 1214, 95, and 37 artificially selected regions containing 2176, 77, and 24 genes were identified in seed, rhizome, and flower lotus, respectively [27]. Several of these selected genes were involved in key developmental processes associated with different organs. For example, a *SUPER-MAN like* gene affecting seed weight and size and a *legumin A-like* gene involved in storage protein synthesis were identified in the subgroup of seed lotus, while an *expansin-A 13-like* gene was identified in the subgroup of rhizome lotus [27]. These specifically selected genes controlling agronomic traits in different subgroups are also possible targets for lotus breeding. Meanwhile, different types of molecular markers have been developed, which may further facilitate the clarification of the relationship between different subgroups and maker-assisted breeding of new lotus varieties [29–38].

4. Identification of Genes with Potential Application in Lotus Breeding

As the largest aquatic vegetable in China, lotus is mainly bred through traditional cross-breeding and physical and chemical mutation as supplementation, based on which thousands of varieties have been obtained [39]. However, the selection of high-quality varieties was mainly based on the breeders' experience, because the mechanisms underlying each economic trait remained unclear. With the development of genomics and molecular genetics of the lotus, genome-based breeding is gradually becoming an effective method for lotus. Causal genes regulating essential traits, such as flower color and shape, rhizome yield, and seed quality, have been widely studied.

Flower color, shape, and flowering time are important traits that determine the ornamental value of lotus. There are three different colors in lotus, red and white in Asian

lotus and yellow in American lotus. The red color in Asian lotus is determined by the contents of anthocyanin [40,41], which is controlled by key enzyme encoding genes, and their regulating transcription factors (TFs) such as *MYB*, *basic-Helix-Loop-Helix* (*bHLH*), *WD40* in its biosynthetic pathway. Among all the enzyme encoding genes in this pathway, *NnANS* and *NnUFGT* seem to be the decisive two genes [42,43]. Several TFs including 5 *MYB*, 2 *bHLH*, and one *WD-repeat* genes, may be involved in the regulation of anthocyanin biosynthesis in lotus based on a transcriptome analysis [43]. Among them, a *bHLH* gene *NnTT8* was verified to regulate anthocyanin biosynthesis [44], whereas the yellow color of American lotus is determined by carotenoid, and no anthocyanin was detected [25,45]. Further analysis indicated that the difference in the coding region between *NnMYB5* (Genbank accession, KU198697) and *NlMYB5* (Genbank accession, KU198698) is the main reason for the different colors in the two species. Flower morphology is another factor that determines the ornamental value of lotus. Flower development is controlled by intricate gene-regulatory networks, and many vital genes that control flowering time have been identified in flowering plant species. However, the molecular regulation mechanism has not been well characterized in lotus. Comparative transcriptomic analysis of different bud development stages in temperate and tropical lotus identified 147 lotus flowering-time associated genes that participate in photoperiod, gibberellic acid and vernalization pathways [46]. The *MADS-box* TFs are widely involved in plant growth and development. A total of 44 *MADS-box* genes were identified in lotus, and based on the selected candidates, *NnMADS14* (*SEPALLATA3* (*SEP3*) homolog gene) was identified to be related to floral organogenesis in lotus [47]. Lotus possesses distinct types of flower morphology, and the floral organ petaloid phenomenon is universal. Comparative transcriptomic analysis identified many hormonal signal transduction pathway genes and *MADS-box* genes; *AGAMOUS*(*AG*) was predicted as the candidate which was gene related to carpel petaloidy [48,49]. Genome-wide DNA methylation analysis showed that different flower organs exhibited different methylation levels, while *plant U-box* (*PUB33*) homolog gene might play crucial roles in the stamen petaloid [50]. Furthermore, *NnFTIP1* was proven to interact with *NnFT1* and regulate the flowering time in lotus [51].

The rhizome is the main edible part of lotus. It is important to explore the mechanisms underlying rhizome formation and expansion in rhizome lotus breeding. Comparative transcriptomic and proteomic analyses focusing on rhizome development have been conducted to dig out the key genes and pathways critical for the crucial physiological process [52–54]. Furthermore, re-sequencing of the natural and genetic F_2 populations has also identified several genetic regions and candidate genes that might be involved in lotus rhizome enlargement [55]. A systematic analysis was conducted on one candidate gene *CONSTANS-LIKE 5* (*COL5*). Functional analysis in the potato system indicated that *NnCOL5* might be positively associated with rhizome enlargement by regulating the expression of *CO-FT* genes and the GA signaling pathway [56]. In addition, one SNP was identified in another candidate gene *NnADAP* of *AP2* subfamily, which is closely associated with rhizome enlargement phenotype and the soluble sugar content [57]. There is a big difference between temperate and tropical lotus, especially the rhizome's morphology. Many genes were highly differentiated between them, such as *APL* homologs and granule-bound starch synthases genes [26]. Temperate lotus is distributed at 20° north latitude and shows a significant annual growth cycle, whereas tropical lotus is distributed south of 17° north latitude and exhibits perennial growth. Asian wild lotus can be further divided into temperate, subtropical, and tropical types and is distributed in northeast China, the Yangtze River and Pearl River Basin, Thailand and India [27]. Different lotus groups are subject to different selection pressure, such as light, temperature, UV, and soil types. The genes underlying selection were discussed by integrating population genetics and omics data. Several genes related to photosynthesis and DNA repair were selected, such as NAD + ADP ribosyltranferase, 8-oxoguanine-DNA glycosylase 1 and DNA polymerase epsilon subunit B2. The *vacuolar iron transporter* (*VIT*) family gene, *nodulin-like 21* gene encoding vacuolar iron transporter, may be related to metal ion metabolism [24]. The homolog gene of *Arabidopsis VIN3* in lotus was predicted to

be related to flowering time and dormancy, with higher expression in temperate lotus than in tropical lotus [26].

Lotus seeds are rich in nutrients and functional compounds such as alkaloids, flavonoids, and polyphenols [58,59]. They are consumed "as both food and medicine" [60]. It is essential to increase the yield and nutrition of lotus seed. The main factors determining lotus seed yield are the seed size and the number of lotus seeds per seedpod. Transcriptome analysis on the cotyledon of "CA" and "Jianxuan-17 (JX-17)" seeds at different developmental stages identified 8437 differentially expressed genes (DEGs). Many DEGs are involved in the brassinosteroid biosynthesis pathway, and further analysis predicted two *AGPase* genes as candidate genes affecting lotus seed yield [61]. It seems that phytohormones are involved in lotus seed development. A combination of metabolomic and proteomic methods revealed that 15 DAP (Day After Pollination) was a switch time point from the physiological active to the nutrition accumulation stage [62]. Starch is the primary nutritional component in mature lotus seed [63]. Its contents and the proportion of amylose and amylopectin could largely determine the nutritional value and taste of lotus cotyledon, respectively. ADP-glucose pyrophosphorylase (AGPase) plays an important role in regulating starch biosynthesis. The evolution of *AGPase* genes experienced a purification selection, and *NnAGPL2a* and *NnAGPS1a* were the candidate genes related to starch content [64]. Starch branching enzyme (SBE) genes are key regulatory genes during starch synthesis, and *NnSBEI* and *NnSBEIII* were identified as related to the chain length of amylopectin in lotus [65]. In addition, comparative metabolomics between wild germplasm "CA" and domesticated cultivar "JX-17" indicated that the seed yield and the content of metabolites showed trade-offs [66]. For nutritional and medicinal values of lotus seed, the metabolomics-assisted strategy might be applied in lotus breeding in the future [67]. Seed dormancy is one of the domestication traits. The classical stay-green *G* gene controlling seed dormancy was cloned in domestication and as improvement genes in soybean, rice, and tomato. G gene interacts with *NCEDS* and *SPY* and in turn, regulates abscisic acid (ABA) synthesis [68]. *NnDREB1* and *NnPER1* were identified from lotus and may be involved in the ABA signal transduction pathway and then modulate longevity and dormancy [69].

Except for the above breeding objectives, there are other diversified breeding objectives, such as resistance to submerging and high antioxidant content. Lotus has evolved novel features to adapt to aquatic lifestyle. Many putative copper-dependent proteins, especially *COG2132* gene family, expand in lotus and form a separate phylogenetic clade having functions distinct from *Arabidopsis* [20]. Research has shown that although lotus grows in water, it is actually "afraid" of water. A time-course submergence experiment and RNA-seq analysis showed lotus has a low tolerance to complete submergence stress, and took two major strategies to cope with submergence stress in different stages. In the early stage (3~6 h) it initiates a low oxygen "escape" strategy (LOES), with the rapid accumulation of ethylene, rapid elongation of petioles, and significantly increases the density of aerenchyma and *ERF-VII* genes while lotus innate immunity genes become elevated; In the later stage (24~120 h), it starts a "breath holding" mode to limit its anaerobic respiration to the lowest level [70]. Flooding is serious abiotic stress affecting plant growth and can be classified into waterlogging and submergence. During the rainy season, the lotus is vulnerable to submergence. It is necessary to cultivate lotus varieties that are resistant to flooding to promote economic value.

WRKY TFs play key roles in modulating plant biotic and abiotic stress response and secondary metabolic regulations. A total of 65 *WRKY* genes were identified in lotus, and they were regulated by salicylic acid (SA) and jasmonic acid (JA), of which *NnWRKY40a* and *NnWRKY40b* were significantly induced by JA and promoted benzylisoquinoline alkaloid (BIA) biosynthesis [71]. Lotus predominantly accumulates BIA, and the leaf and embryo have different alkaloid components that may be caused by two cluster *CYP80* genes synthetic bis-BIAs, and aporphine-type BIAs, respectively. Five TFs (3 *MYBs*, one ethylene-response factor, and one *bHLH*) were identified as the regulator involved in the BIA biosynthetic pathway in lotus [72].

5. Conclusions and Perspective

The new varieties of lotus with high yield, wide adaptability, and stress resistance play a vital role in improving the economic value of this important horticulture crop. The variations identification, functional gene cloning, and metabolites alterations among diverse germplasm resources were investigated in the past decades, driven by the progressively improved genome information which could facilitate breeding practices in lotus (Figure 1). However, a high-quality reference genome is the limiting factor that will affect the molecular breeding process. Improvement of the lotus reference genome will be a requisite in the future, directly affecting the accuracy of molecular markers and the efficiency of cloning functional genes. Gapless reference genomes and pan-genomes have become the new reference, based on which plentiful information of genomes such as open chromatin and more variant information can be explored. With the explosive growth of large-omics data, deep learning can be used to mine biological information and decipher gene regulation networks. Moreover, a sound genetic transformation system has not yet been well established in lotus, which still restricts the validation of gene function and genome-based gene editing, further hindering breeding strategies. Few studies on epigenomics, such as histone marks, accessible chromatin regions, and genomic interactions, have been conducted and are needed in future. Based on these investigations, collection of wild lotus germplasm and classification of both wild and cultivated germplasm, analysis of domestication, and identification of molecular markers and genes closely linked to important agronomic traits, should be systematically conducted in the coming years. Combining these and developing multiple breeding targets will speed up the breeding efficiency in lotus.
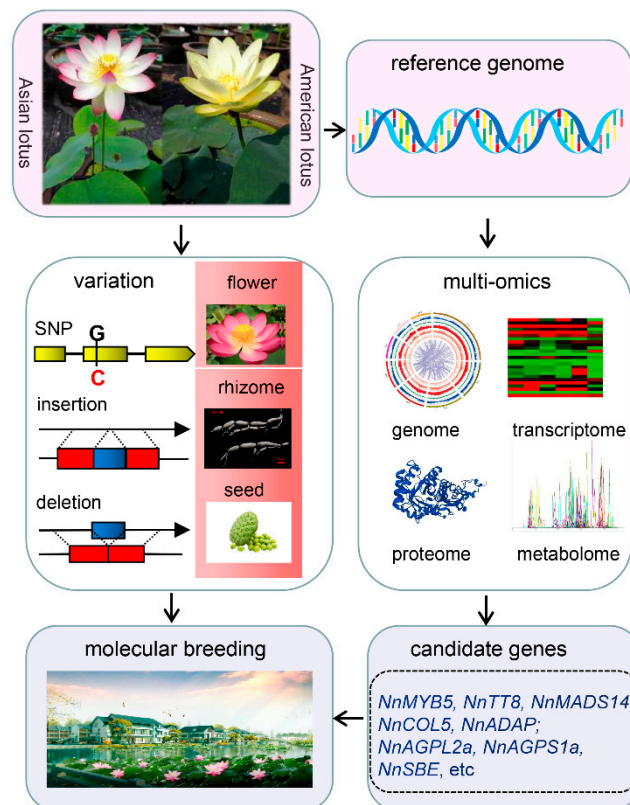

Figure 1. Flowchart of the molecular breeding process of lotus.

References

1. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815. [CrossRef] [PubMed]
2. Chen, F.; Dong, W.; Zhang, J.; Guo, X.; Chen, J.; Wang, Z.; Lin, Z.; Tang, H.; Zhang, L. The sequenced angiosperm genomes and genome databases. *Front. Plant Sci.* **2018**, *9*, 418. [CrossRef] [PubMed]
3. Li, H.; Yang, X.; Zhang, Y.; Gao, Z.; Liang, Y.; Chen, J.; Shi, T. Nelumbo genome database, an integrative resource for gene expression and variants of *Nelumbo nucifera*. *Sci. Data* **2021**, *8*, 38. [CrossRef] [PubMed]
4. van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666–681. [CrossRef]
5. Liang, Y.; Liu, H.J.; Yan, J.; Tian, F. Natural Variation in Crops, Realized Understanding, Continuing Promise. *Annu. Rev. Plant Biol.* **2021**, *72*, 357–385. [CrossRef]
6. Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef]
7. Huang, X.; Kurata, N.; Wei, X.; Wang, Z.X.; Wang, A.; Zhao, Q.; Zhao, Y.; Liu, K.; Lu, H.; Li, W.; et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **2012**, *490*, 497–501. [CrossRef]
8. Liu, S.; Zhang, M.; Feng, F.; Tian, Z. Toward a "Green Revolution" for Soybean. *Mol. Plant* **2020**, *13*, 688–697. [CrossRef]
9. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **2012**, *485*, 635–641. [CrossRef]
10. Bolger, A.; Scossa, F.; Bolger, M.E.; Lanz, C.; Maumus, F.; Tohge, T.; Quesneville, H.; Alseekh, S.; Sørensen, I.; Lichtenstein, G.; et al. The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nat. Genet.* **2014**, *46*, 1034–1038. [CrossRef]
11. Lin, T.; Zhu, G.; Zhang, J.; Xu, X.; Yu, Q.; Zheng, Z.; Zhang, Z.; Lun, Y.; Li, S.; Wang, X.; et al. Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* **2014**, *46*, 1220–1226. [CrossRef]
12. Tay Fernandez, C.G.; Nestor, B.J.; Danilevicz, M.F.; Gill, M.; Petereit, J.; Bayer, P.E.; Finnegan, P.M.; Batley, J.; Edwards, D. Pangenomes as a resource to accelerate breeding of under-utilised crop species. *Int. J. Mol. Sci.* **2022**, *23*, 2671. [CrossRef] [PubMed]
13. Gao, L.; Gonda, I.; Sun, H.; Ma, Q.; Bao, K.; Tieman, D.M.; Burzynski-Chang, E.A.; Fish, T.L.; Stromberg, K.A.; Sacks, G.L.; et al. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **2019**, *51*, 1044–1051. [CrossRef] [PubMed]
14. Liu, Y.; Du, H.; Li, P.; Shen, Y.; Peng, H.; Liu, S.; Zhou, G.A.; Zhang, H.; Liu, Z.; Shi, M. Pan-Genome of Wild and Cultivated Soybeans. *Cell* **2020**, *182*, 162–176. [CrossRef] [PubMed]
15. Huang, X.; Huang, S.; Han, B.; Li, J. The integrated genomics of crop domestication and breeding. *Cell*, 2022; *Advance online publication*. [CrossRef]
16. Lin, Z.; Zhang, C.; Cao, D.; Damaris, R.N.; Yang, P. The latest studies on lotus (*Nelumbo nucifera*)-an emerging horticultural model plant. *Int. J. Mol. Sci.* **2019**, *20*, 3680. [CrossRef]
17. Shen-Miller, J. Sacred lotus, the long-living fruits of China Antique. *Seed Sci. Res.* **2002**, *12*, 131–143. [CrossRef]
18. Liu, L.; Li, Y.; Min, J.; Xiang, Y.; Tian, D. Analysis of the cultivar names and characteristics of global lotus (*Nelumbo*). *Hans J. Agric. Sci.* **2019**, *9*, 163–181.
19. Zhang, L.; Chen, F.; Zhang, X.; Li, Z.; Zhao, Y.; Lohaus, R.; Chang, X.; Dong, W.; Ho, S.; Liu, X.; et al. The water lily genome and the early evolution of flowering plants. *Nature* **2020**, *577*, 79–84. [CrossRef]
20. Ming, R.; VanBuren, R.; Liu, Y.; Yang, M.; Han, Y.; Li, L.T.; Zhang, Q.; Kim, M.J.; Schatz, M.C.; Campbell, M.; et al. Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* **2013**, *14*, R41. [CrossRef]
21. Wang, Y.; Fan, G.; Liu, Y.; Sun, F.; Shi, C.; Liu, X.; Peng, J.; Chen, W.; Huang, X.; Cheng, S.; et al. The sacred lotus genome provides insights into the evolution of flowering plants. *Plant J.* **2013**, *76*, 557–567. [CrossRef]

22. Gui, S.; Peng, J.; Wang, X.; Wu, Z.; Cao, R.; Salse, J.; Zhang, H.; Zhu, Z.; Xia, Q.; Quan, Z.; et al. Improving *Nelumbo nucifera* genome assemblies using high-resolution genetic maps and BioNano genome mapping reveals ancient chromosome rearrangements. *Plant J.* **2018**, *94*, 721–734. [CrossRef] [PubMed]

23. Shi, T.; Rahmani, R.S.; Gugger, P.F.; Wang, M.; Li, H.; Zhang, Y.; Li, Z.; Wang, Q.; Van de Peer, Y.; Marchal, K.; et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* **2020**, *37*, 2394–2413. [CrossRef] [PubMed]

24. Zheng, X.; Wang, T.; Cheng, T.; Zhao, L.; Zheng, X.; Zhu, F.; Dong, C.; Xu, J.; Xie, K.; Hu, Z.; et al. Genomic variation reveals demographic history and biological adaptation of the ancient relictual, lotus (*Nelumbo* Adans). *Hortic. Res.* **2022**, *9*, uhac029. [CrossRef] [PubMed]

25. Zheng, P.; Sun, H.; Liu, J.; Lin, J.; Zhang, X.; Qin, Y.; Zhang, W.; Xu, X.; Deng, X.; Yang, D.; et al. Comparative analyses of American and Asian lotus genomes reveal insights into petal color, carpel thermogenesis and domestication. *Plant J.* **2022**, *110*, 1498–1515. [CrossRef] [PubMed]

26. Huang, L.; Yang, M.; Li, L.; Li, H.; Yang, D.; Shi, T.; Yang, P. Whole genome re-sequencing reveals evolutionary patterns of sacred lotus (*Nelumbo nucifera*). *J. Integr. Plant Biol.* **2018**, *60*, 2–15. [CrossRef]

27. Li, Y.; Zhu, F.L.; Zheng, X.W.; Hu, M.L.; Dong, C.; Diao, Y.; Wang, Y.W.; Xie, K.Q.; Hu, Z.L. Comparative population genomics reveals genetic divergence and selection in lotus, *Nelumbo nucifera. BMC Genom.* **2020**, *21*, 146. [CrossRef]

28. Liu, Z.; Zhu, H.; Zhou, J.; Jiang, S.; Wang, Y.; Kuang, J.; Ji, Q.; Peng, J.; Wang, J.; Gao, L.; et al. Resequencing of 296 cultivated and wild lotus accessions unravels its evolution and breeding history. *Plant J.* **2020**, *104*, 1673–1684. [CrossRef]

29. Liu, Z.; Zhu, H.; Liu, Y.; Kuang, J.; Zhou, K.; Liang, F.; Liu, Z.; Wang, D.; Ke, W. Construction of a high-density, high-quality genetic map of cultivated lotus (*Nelumbo nucifera*) using next-generation sequencing. *BMC Genom.* **2016**, *17*, 466. [CrossRef]

30. Chen, Y.; Zhou, R.; Lin, X.; Wu, K.; Qian, X.; Huang, S. ISSR analysis of genetic diversity in sacred lotus cultivars. *Aquat. Bot.* **2008**, *89*, 311–316. [CrossRef]

31. Guo, H.B.; Li, S.M.; Peng, J.; Ke, W.D. Genetic diversity of *nelumbo* accessions revealed by RAPD. *Genet. Resour. Crop Evol.* **2007**, *54*, 741–748. [CrossRef]

32. Han, Y.C.; Teng, C.Z.; Sheng, Z.; Zhou, M.Q.; Hu, Z.L.; Song, Y.C. Genetic variation and clonal diversity in populations of *Nelumbo nucifera* (nelumbonaceae) in central china detected by ISSR markers. *Aquat. Bot.* **2007**, *86*, 69–75. [CrossRef]

33. Hu, J.; Pan, L.; Liu, H.; Wang, S.; Wu, Z.; Ke, W.; Ding, Y. Comparative analysis of genetic diversity in sacred lotus (*Nelumbo nucifera* Gaertn.) using AFLP and SSR markers. *Mol. Biol. Rep.* **2012**, *39*, 3637–3647. [CrossRef] [PubMed]

34. Li, Z.; Liu, X.; Gituru, R.W.; Juntawong, N.; Zhou, M.; Chen, L. Genetic diversity and classification of *nelumbo* germplasm of different origins by RAPD and ISSR analysis. *Sci. Hortic.* **2010**, *125*, 724–732. [CrossRef]

35. Pan, L.; Quan, Z.; Li, S.; Liu, H.; Huang, X.; Ke, W.; Ding, Y. Isolation and characterization of microsatellite markers in the sacred lotus (*Nelumbo nucifera* gaertn.). *Mol. Ecol. Notes* **2007**, *7*, 1054–1056. [CrossRef]

36. Yang, M.; Xu, L.; Liu, Y.; Yang, P. RNA-Seq Uncovers SNPs and alternative aplicing events in Asian lotus (*Nelumbo nucifera*). *PLoS ONE* **2015**, *10*, e0125702.

37. Hu, J.; Gui, S.; Zhu, Z.; Wang, X.; Ke, W.; Ding, Y. Genome-wide identification of SSR and SNP markers based on whole-genome re-sequencing of a Thailand wild sacred lotus (*Nelumbo nucifera*). *PLoS ONE* **2015**, *10*, e0143765. [CrossRef]

38. Zhang, Q.; Zhang, X.; Liu, J.; Mao, C.; Chen, S.; Zhang, Y.; Leng, L. Identification of copy number variation and population analysis of the sacred lotus (*Nelumbo nucifera*). *Biosci. Biotechnol. Biochem.* **2020**, *84*, 2037–2044. [CrossRef]

39. Zhang, X.; Wang, Q. A preliminary investigation on the morphological and biological characteristics of lotus varieties. *Acta Hortic. Sin.* **1966**, *5*, 89–100.

40. Yang, R.Z.; Wei, X.L.; Gao, F.F.; Wang, L.S.; Zhang, H.J.; Xu, Y.J.; Li, C.H.; Ge, Y.X.; Zhang, J.J.; Zhang, J. Simultaneous analysis of anthocyanins and flavonols in petals of lotus (*Nelumbo*) cultivars by high-performance liquid chromatography-photodiode array detection/electrospray ionization mass spectrometry. *J. Chromatogr. A* **2009**, *1216*, 106–112. [CrossRef]

41. Deng, J.; Chen, S.; Yin, X.; Wang, K.; Liu, Y.; Li, S.; Yang, P. Systematic qualitative and quantitative assessment of anthocyanins, flavones and flavonols in the petals of 108 lotus (*Nelumbo nucifera*) cultivars. *Food Chem.* **2013**, *139*, 307–312. [CrossRef]

42. Deng, J.; Fu, Z.; Chen, S.; Damaris, R.N.; Wang, K.; Li, T.; Yang, P. Proteomic and epigenetic analyses of lotus (*Nelumbo nucifera*) petals between red and white cultivars. *Plant Cell Physiol.* **2015**, *56*, 1546–1555. [CrossRef] [PubMed]

43. Deng, J.; Li, J.; Su, M.; Chen, L.; Yang, P. Characterization of key genes involved in anthocyanins biosynthesis in *Nelumbo nucifera* through RNA-Seq. *Aquat. Bot.* **2021**, *21*, 103428. [CrossRef]

44. Deng, J.; Li, J.; Su, M.; Lin, Z.; Chen, L.; Yang, P. A *bHLH* gene *NnTT8* of *Nelumbo nucifera* regulates anthocyanin biosynthesis. *Plant Physiol. Biochem.* **2021**, *158*, 518–523. [CrossRef] [PubMed]

45. Sun, S.; Gugger, P.; Wang, Q.; Chen, J. Identification of a R2R3-MYB gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (*Nelumbo* Adans.). *PeerJ* **2016**, *4*, e2369. [CrossRef]

46. Yang, M.; Zhu, L.; Xu, L.; Liu, Y. Comparative transcriptomic analysis of the regulation of flowering in temperate and tropical lotus (*Nelumbo nucifera*) by RNA-Seq. *Ann. Appl. Biol.* **2014**, *165*, 73–95. [CrossRef]

47. Lin, Z.; Cao, D.; Damaris, R.N.; Yang, P. Genome-wide identification of *MADS-box* gene family in sacred lotus (*Nelumbo nucifera*) identifies a *SEPALLATA* homolog gene involved in floral development. *BMC Plant Biol.* **2020**, *20*, 497. [CrossRef]

48. Lin, Z.; Damaris, R.N.; Shi, T.; Li, J.; Yang, P. Transcriptomic analysis identifies the key genes involved in stamen petaloid in lotus (*Nelumbo nucifera*). *BMC Genom.* **2018**, *19*, 554. [CrossRef]

49. Lin, Z.; Cao, D.; Damaris, R.N.; Yang, P. Comparative transcriptomic analysis provides insight into carpel petaloidy in lotus (*Nelumbo nucifera*). *PeerJ* **2021**, *9*, e12322. [CrossRef]

50. Lin, Z.; Liu, M.; Damaris, R.N.; Nyong'a, T.M.; Cao, D.; Ou, K.; Yang, P. Genome-wide DNA methylation profiling in the lotus (*Nelumbo nucifera*) flower showing its contribution to the stamen petaloid. *Plants* **2019**, *8*, 135. [CrossRef]

51. Zhang, L.; Zhang, F.; Liu, F.; Shen, J.; Wang, J.; Jiang, M.; Zhang, D.; Yang, P.; Chen, Y.; Song, S. The lotus *NnFTIP1* and *NnFT1* regulate flowering time in *Arabidopsis*. *Plant Sci.* **2021**, *302*, 110677. [CrossRef]

52. Cheng, L.; Li, S.; Yin, J.; Li, L.; Chen, X. Genome-wide analysis of differentially expressed genes relevant to rhizome formation in lotus root (*Nelumbo nucifera* Gaertn). *PLoS ONE* **2013**, *8*, e67116. [CrossRef]

53. Yang, M.; Zhu, L.; Pan, C.; Xu, L.; Liu, Y.; Ke, W.; Yang, P. Transcriptomic analysis of the regulation of rhizome formation in temperate and tropical lotus (*Nelumbo nucifera*). *Sci. Rep.* **2015**, *5*, 13059. [CrossRef]

54. Cao, D.; Damaris, R.N.; Zhang, Y.; Liu, M.; Li, M.; Yang, P. Proteomic analysis showing the signaling pathways involved in the rhizome enlargement process in *Nelumbo nucifera*. *BMC Genom.* **2019**, *20*, 766. [CrossRef] [PubMed]

55. Huang, L.; Li, M.; Cao, D.; Yang, P. Genetic dissection of rhizome yield-related traits in *Nelumbo nucifera* through genetic linkage map construction and QTL mapping. *Plant Physiol. Biochem.* **2021**, *160*, 155–165. [CrossRef] [PubMed]

56. Cao, D.; Lin, Z.; Huang, L.; Damaris, R.N.; Li, M.; Yang, P. A *CONSTANS-LIKE* gene of *Nelumbo nucifera* could promote potato tuberization. *Planta* **2021**, *253*, 65. [CrossRef] [PubMed]

57. Cao, D.; Lin, Z.; Huang, L.; Damaris, R.N.; Yang, P. Genome-wide analysis of *AP2/ERF* superfamily in lotus (*Nelumbo nucifera*) and the association between *NnADAP* and rhizome morphology. *BMC Genom.* **2021**, *22*, 171. [CrossRef] [PubMed]

58. Limwachiranon, J.; Huang, H.; Shi, Z.; Li, L.; Luo, Z. Lotus flavonoids and phenolic acids, health promotion and safe consumption dosages. *Compr. Rev. Food Sci. Food Saf.* **2018**, *17*, 458–471. [CrossRef]

59. Yu, Y.; Wei, X.; Liu, Y.; Dong, G.; Hao, C.; Zhang, J.; Jiang, J.; Cheng, J.; Liu, A.; Chen, S. Identification and quantification of oligomeric proanthocyanidins, alkaloids, and flavonoids in lotus seeds, A potentially rich source of bioactive compounds. *Food Chem.* **2022**, *379*, 132124. [CrossRef]

60. Zhang, Y.; Lu, X.; Zeng, S.; Huang, X.; Guo, Z.; Zheng, Y.; Tian, Y.; Zheng, B. Nutritional composition, physiological functions and processing of lotus (*Nelumbo nucifera* gaertn.) seeds, a review. *Phytochem. Rev.* **2015**, *14*, 321–334. [CrossRef]

61. Li, J.; Tao, S.; Huang, L.; He, D.; Maraga, N.T.; Yang, P. Systematic transcriptomic analysis provides insights into lotus (*Nelumbo nucifera*) seed development. *Plant Growth Regul.* **2018**, *86*, 339–350. [CrossRef]

62. Wang, L.; Fu, J.; Li, M.; Fragner, L.; Weckwerth, W.; Yang, P. Metabolomic and proteomic profiles reveal the dynamics of primary metabolism during seed development of lotus (*Nelumbo nucifera*). *Front. Plant Sci.* **2016**, *7*, 750. [CrossRef]

63. Shen-Miller, J.; Mudgett, M.B.; Schopf, J.W.; Clarke, S.; Berger, R. Exceptional seed longevity and robust growth, ancient sacred lotus from china. *Am. J. Bot.* **1995**, *82*, 1367–1380. [CrossRef]

64. Sun, H.; Li, J.; Song, H.; Yang, D.; Deng, X.; Liu, J.; Wang, Y.; Ma, J.; Xiong, Y.; Liu, Y.; et al. Comprehensive analysis of *AGPase* genes uncovers their potential roles in starch biosynthesis in lotus seed. *BMC Plant Biol.* **2020**, *20*, 457. [CrossRef] [PubMed]

65. Zhu, F.; Sun, H.; Diao, Y.; Zheng, X.; Xie, K.; Hu, Z. Genetic diversity, functional properties and expression analysis of *NnSBE* genes involved in starch synthesis of lotus (*Nelumbo nucifera* Gaertn.). *PeerJ* **2019**, *7*, e7750. [CrossRef] [PubMed]

66. Qi, H.; Yu, F.; Damaris, R.N.; Yang, P. Metabolomics analyses of cotyledon and plumule showing the potential domestic selection in lotus breeding. *Molecules* **2021**, *26*, 913. [CrossRef]

67. Zhu, M.; Liu, T.; Guo, M. Current advances in the metabolomics study on lotus Seeds. *Front. Plant Sci.* **2016**, *7*, 891. [CrossRef]

68. Wang, M.; Li, W.; Fang, C.; Xu, F.; Liu, Y.; Wang, Z.; Yang, R.; Zhang, M.; Liu, S.; Lu, S.; et al. Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* **2018**, *50*, 1435–1441. [CrossRef]

69. Liu, M.; Lin, Z.; Yang, P.; He, D. Research progress on the longevity mechanism of *Nelumbo nucifera* seeds. *Plant Sci. J.* **2019**, *37*, 396–403.

70. Deng, X.; Yang, D.; Sun, H.; Liu, J.; Song, H.; Xiong, Y.; Wang, Y.; Ma, J.; Zhang, M.; Li, J. Time-course analysis and transcriptomic identification of key response strategies to complete submergence in *Nelumbo nucifera*. *Hortic. Res.* **2022**, *9*, uhac001. [CrossRef]

71. Li, J.; Xiong, Y.; Li, Y.; Ye, S.; Yin, Q.; Gao, S.; Yang, D.; Yang, M.; Palva, E.T.; Deng, X. Comprehensive Analysis and Functional Studies of WRKY Transcription Factors in *Nelumbo nucifera*. *Int. J. Mol. Sci.* **2019**, *20*, 5006. [CrossRef]

72. Deng, X.; Zhao, L.; Fang, T.; Xiong, Y.; Ogutu, C.; Yang, D.; Vimolmangkang, S.; Liu, Y.; Han, Y. Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. *Hortic. Res.* **2018**, *5*, 29. [CrossRef] [PubMed]