Molecular Plant
Research Article

CellPress
PARTNER JOURNAL

The *Coix* Genome Provides Insights into Panicoideae Evolution and Papery Hull Domestication

Chao Guo[1,7], Yanan Wang[2,7], Aiguo Yang[3,7], Jun He[1,7], Chaowen Xiao[4], Shanhua Lv[5], Fengming Han[6], Yibing Yuan[1], Yuan Yuan[2], Xiaolong Dong[1], Juan Guo[2], Yawen Yang[1], Hailan Liu[1], Ningzhi Zuo[1], Yaxi Hu[1], Kangxu Zhao[1], Zhengbo Jiang[1], Xing Wang[1], Tingting Jiang[1], Yaou Shen[1], Moju Cao[1], Yuan Wang[1], Zhaobo Long[1], Tingzhao Rong[1], Luqi Huang[2,*] and Shufeng Zhou[1,*]

[1]State Key Laboratory of Exploration and Utilization of Crop Gene Resources in Southwest China, Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture, Maize Research Institute of Sichuan Agricultural University, Chengdu 611130, China

[2]National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, State Key Laboratory Breeding Base of Dao-di Herbs, Beijing 100700, China

[3]Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China

[4]Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610064, China

[5]College of Agronomy, Liaocheng University, Liaocheng 252000, China

[6]Biomarker Technologies Corporation, Beijing 101300, China

[7]These authors contributed equally to this article.

***Correspondence: Luqi Huang (huangluqi01@126.com), Shufeng Zhou (shufengzhou@163.com)**

https://doi.org/10.1016/j.molp.2019.11.008

ABSTRACT

***Coix* is a grass crop domesticated as early as the Neolithic era. It is still widely cultivated for both highly nutritional food and medicinal use. However, the genetic study and breeding of this crop are hindered by the lack of a sequenced genome. Here, we report *de novo* sequencing and assembly of the 1619-Mb genome of *Coix*, and annotation of 75.39% repeats and 39 629 protein-coding genes. Comparative genomics analysis showed that *Coix* is more closely related to sorghum than maize, but intriguingly only *Coix* and maize had a recent genome duplication event, which was not detected in sorghum. We further constructed a genetic map and mapped several important traits, especially the strength of hull. Selection of papery hull (thin: easy dehulling) from the stony hull (thick: difficult dehulling) in wild progenitors was a key step in *Coix* domestication. The papery hull makes seed easier to process and germinate. Anatomic and global transcriptome analysis revealed that the papery hull is a result of inhibition of cell division and wall biogenesis. We also successfully demonstrated that seed hull pressure resistance is controlled by two major quantitative trait loci (QTLs), which are associated with hull thickness and color, respectively. The two QTLs were further fine mapped within intervals of 250 kb and 146 kb, respectively. These resources provide a platform for evolutionary studies and will facilitate molecular breeding of this important crop.**

Key words: *Coix* genome, Comparative genomics, Panicoideae evolution, hull domestication, QTL mapping

Guo C., Wang Y., Yang A., He J., Xiao C., Lv S., Han F., Yuan Y., Yuan Y., Dong X., Guo J., Yang Y., Liu H., Zuo N., Hu Y., Zhao K., Jiang Z., Wang X., Jiang T., Shen Y., Cao M., Wang Y., Long Z., Rong T., Huang L., and Zhou S. (2020). The *Coix* Genome Provides Insights into Panicoideae Evolution and Papery Hull Domestication. Mol. Plant. **13**, 309–320.

INTRODUCTION

Coix lacryma-jobi L., commonly known as adlay or Job's tears, is a widely cultivated crop in East and Southeast Asia (Arora, 1977). In China, people have used *Coix* seeds for food and brewing beer since ~8000 years ago (Yang and Jiang, 2010; Wang et al., 2016). *Coix* seeds have higher protein content than most of the common cereals, and are a nutritionally balanced food (Ottoboni et al.,

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, CAS.

Category	Number	N50 (kb)	Size (Mb)	Percentage of assembly
Contigs	4519	751	1619	NA
Correction	3005	2241	1615	99.8
Anchored	2385	NA	1595	98.8
Oriented	1000	NA	1500	92.9
Predicted coding genes	39 629	NA	56.6	3.5
Repeat sequences	527 996	NA	1218	75.4

Table 1. Summary of *C. aquatica* Genome Assembly and Annotation.
NA, not applicable.

1990; Liu et al., 2015; Zhu, 2017). In addition, *Coix* is a key ingredient in traditional Chinese herbal medicine. The Kanglaite drug extracted from *Coix* seeds has been widely used to cure cancer and reduce the side effects of other treatments (Normile, 2003; Zhang et al., 2019). With the re-recognition of *Coix*'s nutrition and medical function, *Coix* demand is increasing rapidly, and it has now been introduced to almost all tropical and subtropical zones of the world (Lim, 2012).

In traditional taxonomy, *Coix*, *Sorghum*, *Zea*, and *Setaria* are all Panicoideae (Flora of China Editorial Committee, 1978). Based on characteristic inflorescence, *Sorghum* and *Setaria* are classified into Andropogoneae and Paniceae, respectively. *Coix* and *Zea* belong to Maydeae. *Coix* is the only genus in Asia ever considered as a possible progenitor of maize (Hallauer et al., 2010). However, analysis of repetitive sequences indicated that *Coix* was closer to *Sorghum* than to *Zea* (Cai et al., 2014). In addition, *Zea mays* has a recent whole-genome duplication (WGD) event, which is associated with ancient tetraploids and ongoing gene loss (Schnable et al., 2011; Zhang et al., 2012). Whether it is also this case in *Coix* needs to be investigated. Even now, the evolutionary relationship of these genera in Panicoideae is still controversial, owing to lack of a sequenced *Coix* genome.

Wild *Coix* grain is encapsulated in a stony hull and the cultivated *Coix* in a papery one (Flora of China Editorial Committee, 1978). The stony hull is thick and difficult to dehull, whereas the papery one is thin and easily removed. Papery hull is considered to be a key domestication trait in *Coix*. A stony hull in the wild species contributes to their fitness in adverse environments. However, stony hulls impede seed germination, production and processing in agriculture (Singh et al., 2013; Sun et al., 2015). During domestication of seed crops, the stony hull of the wild progenitors would be one of the most notable obstacles faced by ancient humans (Mangelsdorf and Reeves, 1938). Hence, selection of papery hulls and even exposed kernels is a key step in domestication of many important seed crops including barley (Staudt, 1961), maize (Dorweiler et al., 1993; Dorweiler and Doebley, 1997), and wheat (Salamini et al., 2002). This trait is usually controlled by a small number of key genes as revealed in maize (Wang et al., 2005), legume (Sun et al., 2015; Chai et al., 2016), barley (Taketa et al., 2008), and oil palm (Singh et al., 2013). However, whether this is also applied to the hull domestication of *Coix* still needs to be investigated.

Despite the importance of *Coix*, the related genomic and genetic research is very limited. Here, we report a chromosomal-scale genome assembly of *Coix aquatica*, a wild relative of the cultivated *Coix chinensis*, as well as a genetic map and seed hull transcriptome. Comparative genomic analysis showed that *Coix* is more closely related to sorghum than maize. Using this -omics information, we successfully revealed the physiology, transcriptome, and QTLs behind the key step in *Coix* domestication: from stony hull to papery hull.

RESULTS

Genome Sequencing, Assembly, and Annotation

The genome of *C. aquatica* Daheishan (DHS, $2n = 2x = 20$), which was selected from wild *Coix* and used as a perennial grass for herbivores (Supplemental Figure 1), was *de novo* sequenced and assembled using integrated technologies of Illumina and SMRT. We obtained 82 Gb of data, which contained 8 783 914 clean reads with an average length of 9 kb, N50 of 14 kb, and N90 of 6 kb (Supplemental Figure 2A). K-mer analysis showed that the *C. aquatica* genome size was about 1.66 Gb (Supplemental Figure 2B), which is almost identical to the estimated 1.68 Gb using flow cytometry (Cai et al., 2014). The assembled genome is 1.62 Gb, containing 4519 contigs with an N50 of 751 kb (Table 1). The assembled sequences cover 99.84% of 807 129 741 Illumina short reads, 97.82% of 458 core eukaryotic genes, and 95.76% of 1440 core land-plant genes, respectively (Supplemental Figure 3A). Using Hi-C technology, the contig N50 was further improved to 2.24 Mb, and 1595-Mb contig sequences were anchored onto ten pseudo-chromosomes with a maximum and minimum length of 184 Mb and 136 Mb, respectively (Supplemental Table 1). Among them, the order and direction of 1500-Mb contig sequences were determined (Table 1). The Hi-C assembly accuracy was also confirmed by normalized contact matrix (Supplemental Figure 3B).

In total, 39 629 protein-coding genes were predicted by integrated approaches (Table 1), including *ab initio*, homologous, and prediction based on RNA sequencing (RNA-seq) (Supplemental Figure 4A). Among them, 37 336 genes were annotated at least by one of gene ontology (GO), KEGG, KOG, TrEMBL, and Nr databases (Supplemental Figure 4B). In *C. aquatica* genome, total exon length, mean total exon length per gene, and mean exon length were 56.6 Mb, 1428 bp, and 310 bp, respectively (Figure 1A), which were closer to that of *Sorghum bicolor* (Paterson et al., 2009) and *Z. mays* (Schnable et al., 2009) rather than *Setaria italica* (Zhang et al., 2012) and *Oryza sativa* (International Rice Genome Sequencing Project, 2005) (Figure 1A).

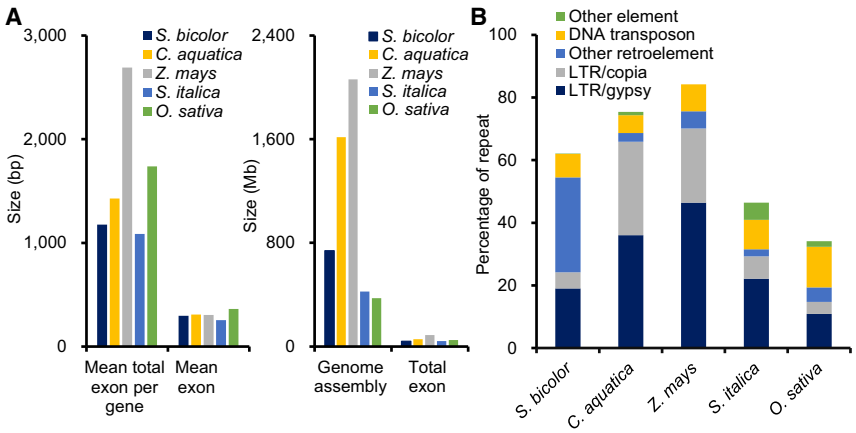

Figure 1. Comparative Analysis of the Genome Structure of *C. aquatica*, *S. bicolor*, *Z. mays*, *S. italica*, *O. sativa*, *H. vulgare*, and *A. thaliana*.

(A) Comparison of the protein-coding gene structure.

(B) Comparison of the repetitive element structure.

The *C. aquatica* genome has a total repetitive element content of 75.39%, comprising 68.64% retroelements, 5.72% DNA transposons, and 1.03% other elements (Table 1 and Figure 1B). Most of the retroelements are long terminal repeat (LTR), including 36.07% *gypsy* and 29.83% *copia* (Figure 1B). The repetitive element content in *C. aquatica* (75.39%) was between that of *S. bicolor* (62%) and *Z. mays* (84%), but far from *S. italica* (46%) and *O. sativa* (34%) (Figure 1B). However, the ratio of *gypsy*/*copia* in *C. aquatica* was the lowest in the five compared species, suggesting that *copia* amplification had higher contribution to genome evolution in *C. aquatica* than in the other species.

Evolutionary Analysis of *Coix*

Predicted genes in *C. aquatica*, *S. bicolor* (Paterson et al., 2009), *Z. mays* (Schnable et al., 2009), *S. italica* (Zhang et al., 2012), *O. sativa* (International Rice Genome Sequencing Project, 2005), *Hordeum vulgare* (Mascher et al., 2017), and *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative, 2000) were clustered into 19 589 gene families, which covered 34 169 genes of *C. aquatica*. These genes were classified into 5881 single-copy, 7558 multiple-copy, 2755 unique, and 17 975 other genes (Figure 2A). Among them, 2755 unique genes were clustered into 655 gene families mainly involved in photosynthesis, ribosome biogenesis, oxidative phosphorylation, and protein processing in endoplasmic reticulum (Supplemental Table 2). These genes may explain the high biomass and protein content of *C. aquatica* (Supplemental Figure 1A and 1D; Ottoboni et al., 1990; Liu et al., 2015). A phylogenetic tree was constructed based on 2882 shared single-copy genes (Figure 2A). The results showed that Panicoideae was split ~49.6 million years ago (Mya), *S. italica* was split ~23.9 Mya, and *Z. mays* was split ~13.1 Mya (Figure 2A), which are close to the reported ~48, ~27, and ~13 Mya, respectively (Zhang et al., 2012). *C. aquatica* and *S. bicolor* were located in the same branch and diverged ~10.0 Mya, which was very close to the divergent time of *Z. mays* (Figure 2A). In addition, *C. aquatica* and *S. bicolor* shared the largest number of gene families, followed by *C. aquatica* and *Z. mays*, *C. aquatica* and *S. italica*, and *C. aquatica* and *O. sativa* (Figure 2B). Furthermore, the synteny between *C. aquatica* and *S. bicolor* was higher than that between *C. aquatica* and *Z. mays* (Figure 2C). Together, these results suggest that *C. aquatica* is more closely related to *S. bicolor* than to *Z. mays* on the whole-genome level. We detected a recent WGD event in *Z. mays*, which was not found in *S. bicolor* (Figure 2D), in agreement with previous reports (Schnable et al., 2011; Zhang et al., 2012). Interestingly, *C. aquatica* has a recent segmental genome duplication event that is independent from the one in *Z. mays* (Figure 2D). However, we cannot determine the exact order of genome duplication and divergent times among *Z. mays*, *C. aquatica*, and *S. bicolor* because they are very close to each other (Figure 2A and 2D), and the molecular clock may run at different rates in different lineages or over different time spans.

We found 1647 gene families expanded and 603 gene families contracted in *C. aquatica* (Figure 2A). The expanded gene families were mainly involved in photosynthesis, ABC transporters, oxidative phosphorylation, ribosome, isoquinoline alkaloid biosynthesis, plant–pathogen interaction, and other pathways (Supplemental Table 3), which may contribute to high biomass and biotic resistance of *C. aquatica*. Among them, there were 41 NB-ARC domain gene families, and the total number of genes in *C. aquatica* is more than other compared species excluding *H. vulgare* (Supplemental Figure 5A). Moreover, another family, GF_58, involved in benzoxazinoid (BX) biosynthesis, was also expanded (Supplemental Figure 5B). It has been documented that BXs play important roles in plant resistance to pests and diseases (Ahmad et al., 2011; Bruijn et al., 2018).

Characterization and Transcriptome of *Coix* Seed Hull

The selection of papery hull was a milestone in *Coix* domestication. Thus, we characterized the differences between two contrasting hull types in wild DHS (stony hull) and cultivated Xiaobaike (XBK, papery hull) (Figure 3A and Supplemental Figure 6B). We found that the hull of XBK (~171 μm) was significantly thinner than that of DHS (~460 μm) (Figure 3B). We also measured seed hull pressure resistance (SHPR), the minimum strength required to break the hull of a mature seed in dehulling. The SHPR of XBK was only one-fiftieth that of DHS (Figure 3C). This makes seeds of cultivated *Coix* much easier to dehull by human beings, especially in ancient times when there were few effective processing tools available. Furthermore, the germination rate of hulled seeds was 91.3% in XBK but only 25.7% in DHS, while the corresponding values for dehulled seeds were about the same (90%) (Supplemental Figure 8A and 8B). Hindering of dehulling and seedling establishment by the stony hull may ascribe to its thickness.

To further investigate the structural differences between papery hull and stony hull, we compared cross-sections of seed hulls (Figure 3D). XBK hull has fewer vascular bundles than

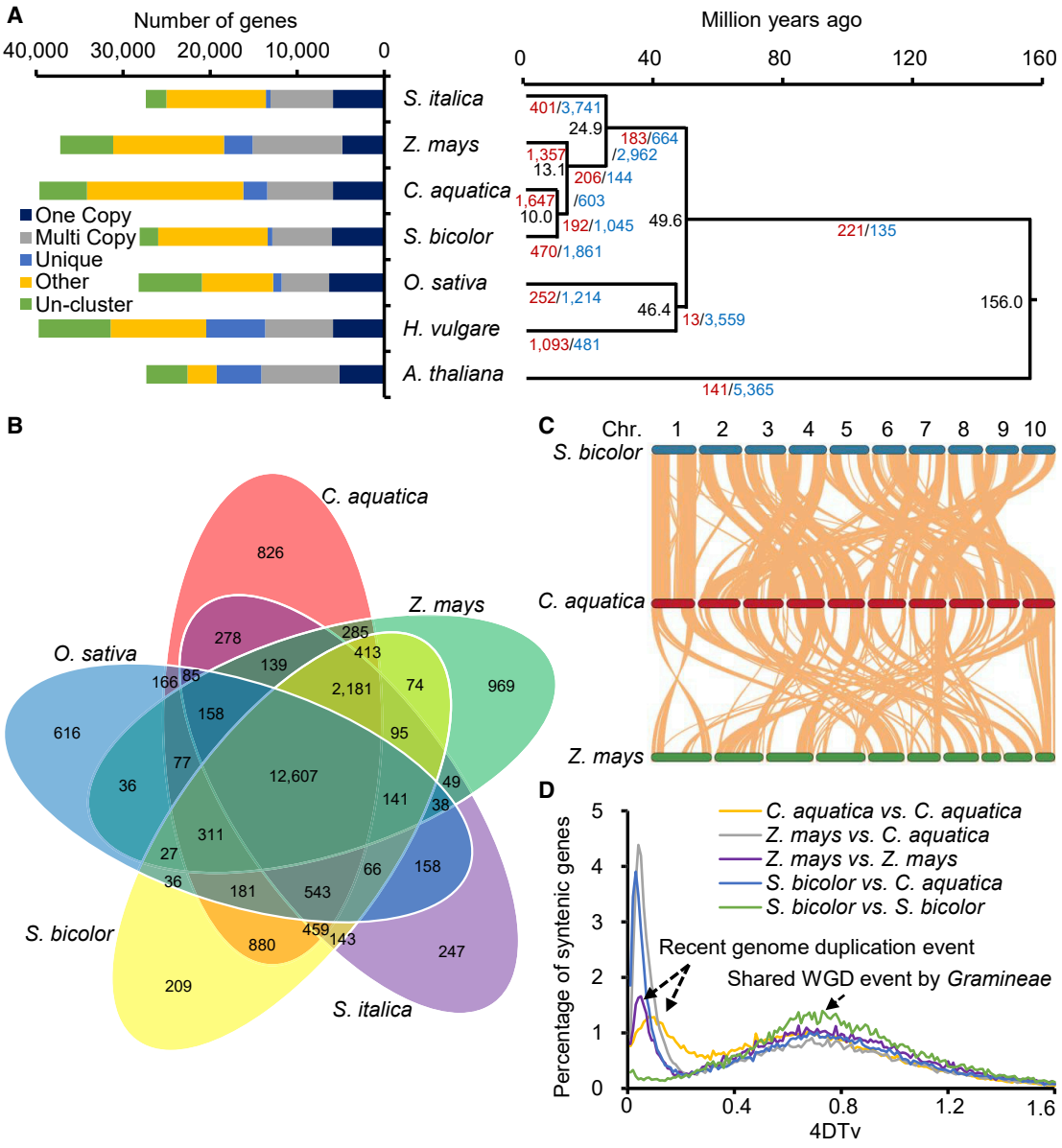

Figure 2. Evolutionary Analysis of the *C. aquatica* genome.

(A) Left: family clustering of predicted genes. Right: phylogenetic tree based on shared single-copy gene families. Divergent times are indicated above the tree, and the number of expanded and contracted gene families are shown in red and blue numbers on the branches, respectively.

(B) Number of shared and unique gene families.

(C) Syntenic blocks between *C. aquatica*, *S. bicolor*, and *Z. mays*.

(D) Distribution of four-fold degenerate sites of the third codons (4DTv) for percentage of syntenic genes from *C. aquatica*, *S. bicolor*, and *Z. mays*. Two whole-genome duplication (WGD) events were indicated by the peaks (4DTv = 0.08 and 0.68).

DHS hull, and sclerenchyma cells around the vascular bundles in XBK hull were significantly smaller than in DHS. In addition, these differences were also found between other wild and cultivated *Coix* accessions (Supplemental Figure 7 and Supplemental Table 4). To track the developmental process of the hull, we divided *Coix* hull development into five stages according to the shape and growing speed of the hull (Figure 3E). At stage 2 (Supplemental Figure 8C), hull thickness of DHS and XBK was still comparable but XBK hull started to have fewer vascular bundles than DHS hull. The sclerenchyma cells around the vascular bundles in XBK hull were slightly smaller than in DHS. At stage 3 (Supplemental Figure 8C), the XBK hull became significantly thinner than the DHS hull. The sclerenchyma cells in XBK hull were less developed while they were well enlarged in DHS. These differences were maintained to the maturing stages (stage 4 and stage 5) (Supplemental Figure 8C). These results showed that the number of vascular bundles and size of sclerenchyma cells determined the difference in hull thickness between wild and cultivated *Coix*.

At stage 4, *Coix* hull stops enlarging and has the shape and thickness of a mature hull (Supplemental Figure 8C). Thus,

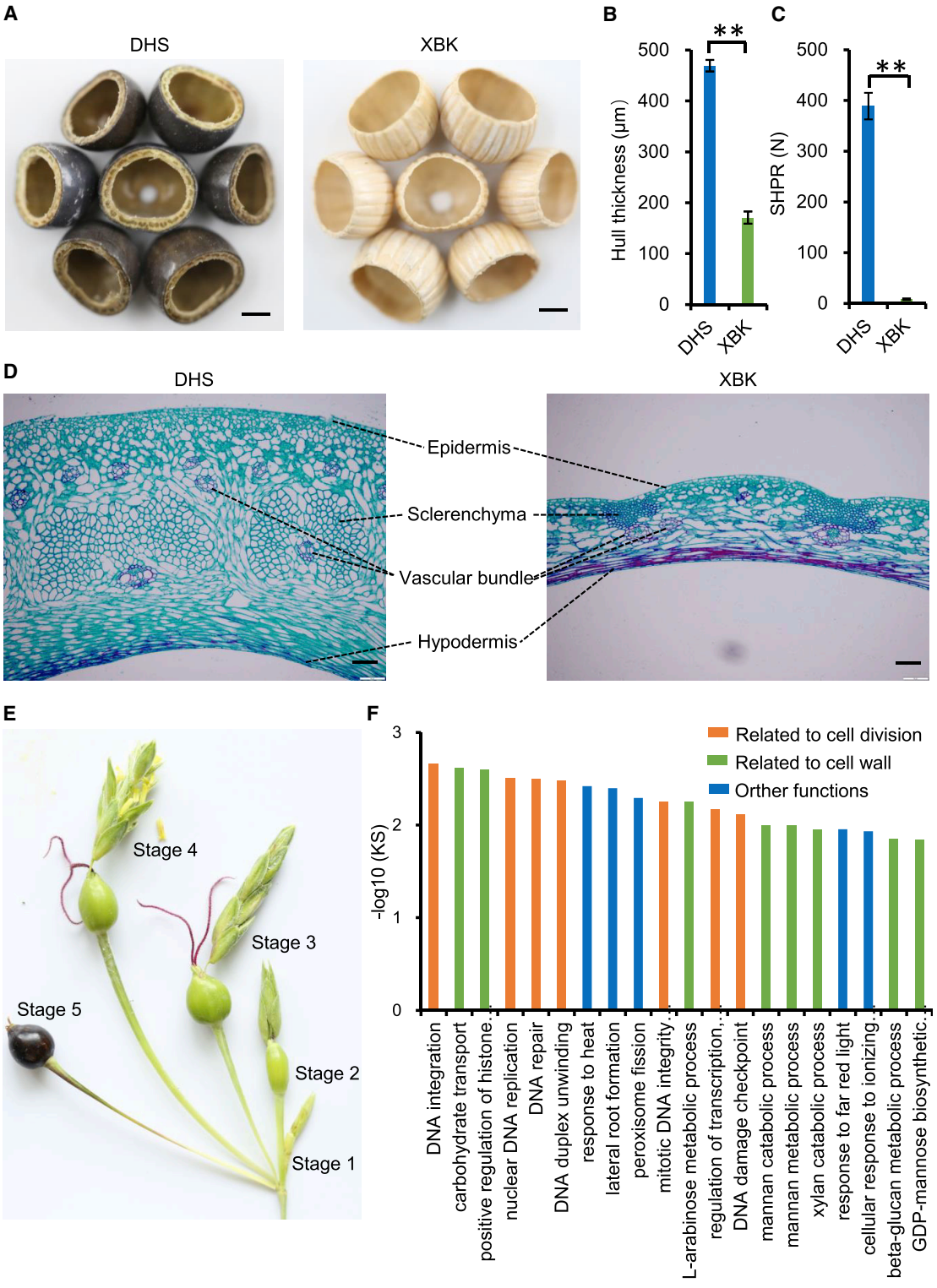

Figure 3. Comparison of Seed Hull of Daheishan (Stony Hull) and Xiaobaike (Papery Hull).

(A) Hull cross section of Daheishan (DHS) and Xiaobaike (XBK). Scale bar represents 2 mm.

(B) Hull thickness of DHS and XBK (mean ± SD; two-sided *t*-test ($n = 30$), $^{**}p \leq 0.01$).

(C) SHPR of DHS and XBK (mean ± SD; two-sided *t*-test ($n = 51$), $^{**}p \leq 0.01$).

(D) Hull structure of DHS and XBK. Scale bar represents 50 μm.

(E) Hulls at different developmental stages co-exist in a typical *Coix* inflorescence. Stage 1, emerging; stage 2, fast-enlarging; stage 3, late-enlarging; stage 4, early-maturing; stage 5, late-maturing.

(F) The 20 most enriched biological process GO terms among DEGs in stage-2 hull.

transcriptomes of hulls at stages 1–3 (Supplemental Figure 9A) were analyzed to find differentially expressed genes (DEGs) associated with hull formation between DHS and XBK. The number of DEGs at stage 2 was greater than that at stage 1 and stage 3 in cell division, carbohydrate transport and metabolism, and cell wall biogenesis (Supplemental Figure 9B). GO enrichment showed that these DEGs at stage 2 were mainly involved in DNA integration, carbohydrate transport, positive regulation of histone acetylation, nuclear DNA replication, mitotic DNA integrity checkpoint, L-arabinose metabolic, and xylan catabolic (Figure 3F and Supplemental Figure 10). By clustering these DEGs into 20 subclusters (Supplemental Figure 11A), we found that subcluster 18 was the most significantly downregulated in XBK. These genes were enriched in glutathione metabolic, L-arabinose metabolic, regulation of transcription from RNA polymerase II promoter, and DNA strand elongation involved in mitotic DNA replication (Supplemental Figure 11B). These results and the morphology differences suggested that inhibition of cell division and wall biogenesis contributed to the thinner hull in XBK.

Mapping of QTLs Contributing to SHPR

A genetic map was constructed with an "immortalized F_2" population derived from a cross between DHS and XBK. The map contains 230 Indel markers and covers 1570.12 cM with an average interval of 6.83 cM (Supplemental Figure 12 and Supplemental Table 5). Ninety percent of the markers are consistent with the physical map (Supplemental Table 5). Using the genetic map and F_2 population, QTL mapping of SHPR, heading date, plant height, and number of tillers was conducted (Supplemental Figure 12 and Supplemental Table 6). In the F_2 population, besides the parental phenotypes (thick black hull and thin white hull, hereafter termed TKB hull and TNW hull), two recombinant phenotypes (thick white hull and thin black hull, hereafter TKW hull and TNB hull) were also observed (Figure 4A and 4B; Supplemental Table 7). The average SHPR of the four hull types was significantly different, with the highest to the lowest being TKB, TKW, TNB, and TNW hull, respectively (Figure 4B), revealing that thick/thin is more associated with SHPR than black/white. To resolve the puzzle of the association between hull color and SHPR, we measured the contents of mineral elements in different hull types. We found that the contents of multiple mineral elements, including Ca, Cu, K, Mg, Si, P, Zn, and S, were higher in the white hulls than in the black ones (Supplemental Figure 13).

We consistently detected two QTLs for SHPR. One was *Ccph1* (*Coix chinensis papery hull 1*) flanked by markers yyIndel0226 and yyIndel0732 in Chr. 8, and the other was *Ccph2* (*Coix chinensis papery hull 2*) neighboring marker yyIndel0715 in Chr. 6 (Supplemental Figure 12). *Ccph1* explained 61%, 69%, and 74% of phenotypic variance in three independent experiments, and the corresponding values for *Ccph2* were 8%, 6%, and 5%, respectively (Supplemental Table 6). Furthermore, we constructed a $BC_{4(DHS)}$ population by backcrossing an F_1 plant to the double-recessive parent XBK (*ph1ph1ph2ph2*) four times. The substituted chromosome segments containing QTLs of SHPR in the population were confirmed by the 230 Indel markers of the genetic map (Supplemental Figure 14 and Supplemental Table 8). We found that *Ccph1* and *Ccph2* segments were both heterozygous in all the tested TKB hull plants (*PH1ph1PH2ph2*). *Ccph1* segments were heterozygous and *Ccph2* segments homozygous in TKW hull plants (*PH1ph1ph2ph2*), and this was reversed in TNB plants (*ph1ph1PH2ph2*). Together, these results convincingly demonstrated that *Ccph1* and *Ccph2* coordinately regulate SHPR, and *Ccph1* specifically governs hull thickness while *Ccph2* solely regulates hull color.

Fine Mapping of *Ccph1* and *Ccph2*

To narrow down the region of *Ccph1*, we genotyped an additional 1276 F_2 individuals with markers yyIndel0226 and yyIndel0732. On the basis of the *Coix* genome, three new Indel markers (yyIndel0893, yyIndel0894, and yyIndel0903) were developed (Supplemental Table 9), and *Ccph1* was mapped within yyIndel0903 and yyIndel0894. With two additional single-nucleotide polymorphism (SNP) markers (ZN1 and PK1), we fine mapped *Ccph1* to an interval of 250 kb between ZN1 and PK1, using 14 286 F_2 plants (Figure 4C). The region contains *EVM0025867* encoding an uncharacterized protein, *EVM0018786*, related to DNA recombination and a gap (Supplemental Table 10). To find clues as to what genes may be in the gap, we checked the synteny sequence in *S. bicolor* genome corresponding to *Ccph1* region (Supplemental Figure 15). The *S. bicolor* segment contains only two putative protein-coding genes. One is *LOC110435847*, encoding an uncharacterized protein, and the other is *LOC8070685*, encoding a putative Harbinger Transposase Derived 1 (HARBI1). Similarly, *Ccph2* was fine mapped to a 146-kb region between two SNP markers, Co14 and Co13 (Figure 4D). The genome sequence of this region carries 10 predicted protein-coding genes (Supplemental Table 10), among which *EVM0013000* encodes a putative β subunit of adaptor protein (AP) complex 3.

DISCUSSION

The comprehensive information of the wild *C. aquatica* genome would provide an excellent resource for genetic improvements for this crop and its close relatives including sorghum and maize. Many valuable traits of wild *Coix* have been identified, such as higher medicinal components and protein level, long duration of staying green, and waterlogging resistance (Huang et al., 1995; Xi et al., 2016). Discovery and utilization of the functional genes for these traits would help to breed new elite *Coix* cultivars. Furthermore, as a helophyte, *Coix* has excellent disease resistance, especially against pathogens that tend to burst under a humid environment. This is probably due to the expanded gene families with NB-ARC domain and gene families involved in biosynthesis of defensive metabolites, such as GF_58 (Supplemental Figure 5A and 5B). These advantages of *Coix* may benefit resistance breeding of maize and sorghum, which suffer greatly from humid-induced diseases.

The mutation of *tga1* (*teosinte glume architecture 1*) in teosinte, the progenitor of maize, causes glume thinning, which leads to the exposed and readily utilized kernels in maize (Dorweiler et al., 1993; Dorweiler and Doebley, 1997; Wang et al., 2005). In *Coix*, we showed that a similar trait was controlled by a major QTL, *Ccph1*. Anatomic and transcriptome analysis suggested that recessive *Ccph1* in cultivated *Coix* leads to inhibition of cell division and wall biogenesis (Figure 3F; Supplemental

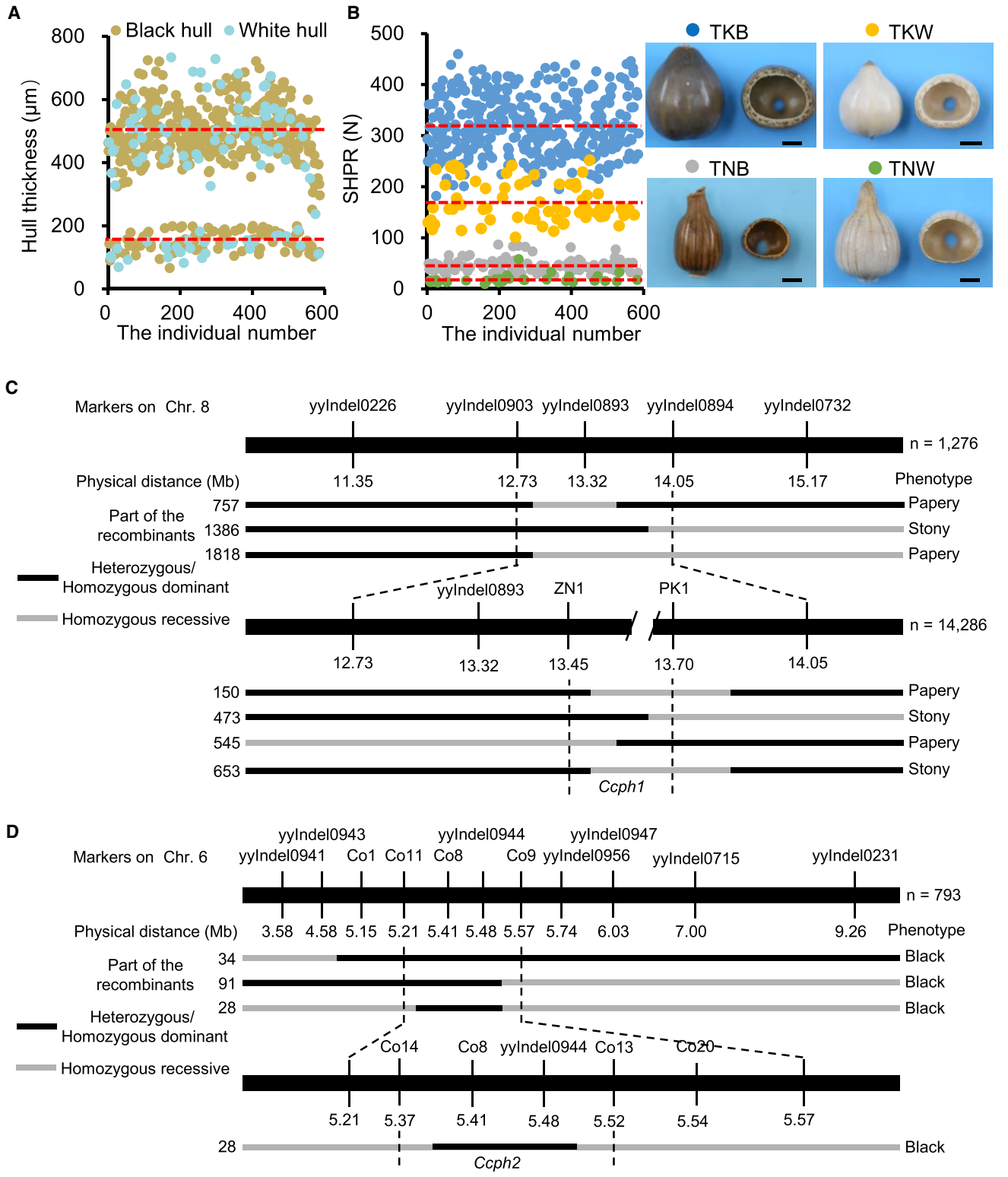

Figure 4. Fine Mapping of *Ccph1* and *Ccph2*.

(A) Distribution of hull thickness. Hull thickness was clearly clustered into two groups (mean. $n = 5$).

(B) Distribution of SHPR. Four types of hulls are clearly distinguished. TKB, thick black hull; TKW, thick white hull; TNB, thin black hull; TNW, thin white hull (mean, $n = 5$). Scale bar represents 2 mm. For **(A)** and **(B)**, 551 F_2 individuals in the "immortalized F_2" population were used, and the dotted red lines represent the average thickness and SHPR in different hull types.

(C) Fine mapping of *Ccph1*.

(D) Fine mapping of *Ccph2*.

Figures 9B and 11). We fine mapped *Ccph1* to a region of 250 kb (Figure 4C), which contains a gene with unknown function, a gene involved in DNA recombination, and possibly a gene of unknown function and a *HARBI1* in the gap revealed by the synteny between *Coix* and *Sorghum* genome (Supplemental Figure 15). However, which gene is the real *CcPH1* and how it regulates hull thickness still need further investigation. Interestingly, *Ccph2*, another major QTL of SHPR, is also associated with hull color. In the fine-mapped region of *Ccph2* (Supplemental Table 10), a gene *EVM0013000* encoding AP3-β was annotated. AP complexes display a conserved function of selection and packaging of cargo proteins in the eukaryotic cell (Feraru et al., 2010). Furthermore, mutation of AP3-β caused defective pigment biogenesis in fruit fly, mouse, and human (Kretzschmar et al., 2000). Based on the plant-specific role of AP3-β in vacuolar biogenesis and function (Feraru et al., 2010; Zwiewka et al., 2011), it is likely that AP3-β also regulates pigment accumulation in *Coix* hull. Besides, this gene could negatively affect the transportation and accumulation of mineral elements in the hull (Supplemental Figure 13). The higher accumulation of mineral elements in the white hulls might make them more fragile, thus leading to a lower SHPR.

The *C. aquatica* genome together with the mapping population, genetic map, and transcriptome data set have offered a powerful platform for evolutionary study and functional genomics, and will facilitate the molecular breeding of this important crop. Using these toolboxes, we demonstrated that papery hull in cultivated *Coix* might be caused by both abnormal cell division and development. We further found this trait to be controlled by two major QTLs, which are separately associated with hull thickness and color. These findings greatly enhanced our understanding of the domestication of this crop. As the first sequenced species in *Coix* genus of Gramineae, *C. aquatica* genome bridges the sequenced sorghum and maize, and will help in studies of the evolution of grass crops.

METHODS

Plant Materials

Daheishan (DHS, $2n = 2x = 20$), possessing a stony hull (thick and black), was selfed for seven generations from a wild *C. aquatica* plant collected from Jinghong, China. DHS is bred as a perennial grass (Supplemental Figures 1 and 6). In this study, it was used for whole-genome sequencing and as one parental line of the "immortalized F_2" population. Xiaobaike (XBK, $2n = 2x = 20$) was the most widely cultivated *Coix* in China and has a papery hull (thin and white). Besides hull feature, the two varieties have significant differences in many important agronomic traits, including plant height, number of tillers, and heading date (Supplemental Figure 6). In addition, a mini-core collection of 10 wild and 10 cultivated *Coix* accessions were used for anatomic analysis (Supplemental Figure 7 and Supplemental Table 4).

Preparation of Samples

Genomic DNA of DHS, XBK, 551 F_2 individuals, and 84 $BC_{4(DHS)}$ individuals were isolated from 7-day-old leaves using a conventional CTAB (cetyl trimethylammonium bromide) method. Total RNA of DHS and XBK were isolated from 2-month-old mixed tissues (root, stem, and leaf) for RNA-seq to predict encoding genes and develop markers. Total RNA of DHS and XBK were isolated from hulls at stage 1–3 for RNA-seq to analyze DEGs with three replicates. All samples were collected and flash-frozen in liquid nitrogen. Total RNAs were isolated using the E.Z.N.A. Plant RNA Kit (Omega Bio-Tek) according to the manufacturer's standard protocol. The quality of RNAs was then checked using an Agilent 2100 Bioanalyzer.

Genome Sequencing and Assembly

Genome sequencing of DHS was performed using SMRT sequencing on a PacBio RS II sequencer (Pacific Biosciences) following the manufacturer's standard protocol. We acquired about 50× clean reads using Pacific Biosciences SMART analysis software v1.2.

With the aid of module of Canu v1.5 (Koren et al., 2017), clean reads longer than 500 bp were selected with the settings 'genomeSize = 3500000000' and 'corOutCoverage = 80'. Reads overlapping were detected through a highly sensitive overlapper MHAP v2.12 ('corMhapSensitivity = low/normal/high'), and errors were corrected using falcon_sense method ('correctedErrorRate = 0.025'). The corrected reads were assembled into contigs using Canu v1.5. The iterative polishing of assembly was achieved by aligning Illumina paired-end reads (~72× PE150 reads) into contigs using Pilon v1.22 (Walker et al., 2014). Furthermore, the assembly was evaluated by mapping Illumina reads, recoveries of core eukaryotic genes from CEGMA v2.5 (Parra et al., 2007), and core land-plant genes from BUSCO v2 (Simão et al., 2015). According to the previous Hi-C assembly procedure (Lieberman-Aiden et al., 2009), the contigs were broken into fragments with a length of 500 kb and then were clustered by LACHESIS software (Burton et al., 2013) using Hi-C data. The number of the Hi-C read pairs between any two bins (500 kb each bin) was computed as the signal of contact intensity. The frequent contacts between adjacent DNA regions within contigs were confirmed by an invariable, strong and broad diagonal in the contact matrix after some manual adjustments (Marie-Nelly et al., 2014). Finally, the contigs were clustered, ordered, and oriented into pseudo-chromosomes by Hi-C data. The gaps of pseudo-chromosomes were filled, and Illumina reads were used to correct base-calling again. Finally, the interaction of adjacent Hi-C sequences was checked for assessing the assembly.

Genome Annotation

Using LTR FINDER v1.05 (Xu and Wang, 2007), MITE-Hunter (Han and Wessler, 2010), RepeatScout v1.0.5 (Price et al., 2005), and PILER-DF v2.4 (Edgar and Myers, 2005), transposable element (TE) was predicted by homolog-based and *de novo* strategies. These repeats were then merged together using Repbase (Jurka et al., 2005) to form a *C. aquatica* repetitive sequence database. Finally, the database was used to identify and annotate TEs by RepeatMasker v4.0.6 (Chen, 2004). *De novo*, homolog-based, and transcriptome-based strategies were combined to predict the gene model. Genscan (Burge and Karlin, 1997), Augustus v2.4 (Stanke and Waack, 2003), GlimmerHMM v3.0.4 (Majoros et al., 2004), GeneID v1.4 (Blanco et al., 2007), and SNAP v2006-07-28 (Korf, 2004) were used in the *de novo* prediction. GeMoMa v1.3.1 (Keilwagen et al., 2016) was used for homolog-based prediction. In the transcriptome-based prediction, unigenes were first parametrically and non-parametrically assembled, respectively, then were predicted by TransDecoder v2.0 (Haas and Papanicolaou, 2016), GeneMarkS-T v5.1 (Tang et al., 2015), and PASA v2.0.2 (Campbell et al., 2006), respectively. All the predicted gene structures were integrated into a consensus set with EVM v1.1.1 (Haas et al., 2008). The predicted genes were annotated according to alignments against databases including GO (Conesa et al., 2005), KEGG (Kanehisa and Goto, 2000), KOG (Tatusov et al., 2001), TrEMBL (Boeckmann et al., 2003) and Nr (Marchler-Bauer et al., 2010) using BLAST v2.2.31 (Altschul et al., 1990) (E value $\leq 1\times10^{-5}$). In addition, noncoding RNA, including microRNA, rRNA, and tRNA, were predicted by Infenal v1.1 (Nawrocki and Eddy, 2013) and tRNAscan-SE v1.3.1 (Lowe and Eddy, 1997) according to the Rfam (Griffiths-Jones et al., 2005) and miRbase (Griffiths-Jones et al., 2006) databases. Pseudogenes were predicted by GenBlastA v1.0.4 (She et al., 2009) alignment and GeneWise v2.4.1 (Birney et al., 2004).

Comparative Genomic Analysis

OrthoMCL (Li et al., 2003) was used to cluster the gene families. The sequences of shared single-copy gene families in *C. aquatica*, *S. bicolor* (Paterson et al., 2009), *Z. mays* (Schnable et al., 2009), *S. italica* (Zhang et al., 2012), *O. sativa* (International Rice Genome Sequencing Project, 2005), *H. vulgare* (Mascher et al., 2017), and *A. thaliana* (*Arabidopsis* Genome Initiative, 2000) were used to construct phylogenetic trees using PhyML (Guindon et al., 2010), and divergent time, including fossil time points of *O. sativa* versus *H. vulgare*, *A. thaliana* versus *S. bicolor*, *O. sativa* versus *S. italica*, and *S. italica* versus *Z. mays*, were estimated with Macmctree (Yang, 2007). Based on results of phylogenetic trees and gene family clustering, analysis of expanded and contracted gene family was performed with CAFÉ (De Bie et al., 2006) and annotated using Pfam (Finn et al., 2013). Syntenic block analysis was performed by MCScanX (Wang et al., 2012). The 4DTV method was adapted to identify genome duplication events in the *S. bicolor*, *Z. maize*, and *C. aquatica* genomes.

RNA Sequencing, Marker Development, and DEG Analysis

RNA-seq libraries of mixed tissues (root, stem, and leaf) and hull tissues of stage 1–3 were prepared and sequenced on a HiSeq 2500 system with a PE150 strategy following the manufacturer's instructions (Illumina, USA). All low-quality reads were filtered using in-house perl script as follows: (1) reads with $\geq$10% unidentified nucleotides (N); (2) reads with >10 nucleotides aligned to the adapter, allowing $\leq$10% mismatches; (3) reads with >50% bases having Phred quality <5. About 6 Gb of clean reads were aligned to genome using HISAT2 v2.0.4 (Sirén et al., 2014) with the default parameters, then were assembled by StringTie v1.3.0 (Pertea et al., 2015) with parameters "—max-intronlen 20 000 –min-intronlen 20 -p 10."

Based on the RNA-seq data of mixed tissues (root, stem, and leaf), the SNP and Indel were identified by GATK4 (McKenna et al., 2010) and SAMtools (Li et al., 2009) with default parameters. Intersection of the two methods was used as the candidate marker set for genetic map construction and fine mapping.

In DEG analysis, FPKM (Florea et al., 2013) was used as measurement of the expression level to calculate gene expression by StringTie v1.3.0 with the default parameters. Pearson's correlation coefficient was used to measure the correlation between biological replicates. DEseq (Wang et al., 2009) was then used to analyze DEGs at fold change $\geq$2 and false discovery rate (FDR) <0.01. Finally, these DEGs were performed functional enrichment using Blast2GO. The *P* value was calculated using two-sided Fisher's exact test and was further corrected with the Benjamini-Hochberg procedure.

Construction of Population

Using an F_1 plant derived from the cross between DHS and XBK, an F_2 population and a $BC_{4(DHS)}$ population were developed. The F_2 population contains 551 individuals. As *Coix* is perennial under controlled conditions and can be propagated through cutting at the stage of flowering time, the initial F_1 plant and all the F_2 individuals are permanently maintained. Thus, the F_2 population is an "immortalized F_2" population. The $BC_{4(DHS)}$ population was created by backcrossing the F_1 plant to the double-recessive parent XBK four times.

Phenotyping

The "immortalized F_2" population was grown in Jinghong, China in 2017–2018 and Chengdu, China in 2018. Heading date, plant height, and number of tillers at heading stage were recorded. Five mature seeds were randomly selected from each F_2 individual to measure SHPR and hull thickness, and color recorded. SHPR was recorded as the minimum pressure to break the hull of a mature seed by putting the seed laterally on a pressure detector (YYD-1 type, Zhejiang Top Instrument). To measure hull thickness, we sliced the hull crosswise at the middle position, then measured it under an anatomical microscope. The hull was classified as thick if thicker than 230 μm, or thin if thinner than 200 μm. Hull color was recorded as white or black.

Coix has an indefinite inflorescence, and hulls at different developmental stages co-exist in a typical *Coix* inflorescence (Figure 3E). Based on shape and growing speed, *Coix* hull development can be divided into five stages, namely emerging, fast-enlarging, late-enlarging, early-maturing, and late-maturing. Stage 2 is the vital and fastest growing stage. For anatomical observation, hulls at stages 2–5 were sliced crosswise as used to measure thickness and used for making paraffin sections. After staining with safranin O-fast green, the sections were observed under a microscope.

For the germination assay, hulled and dehulled seeds of DHS and XBK were cultivated in an incubator (28°C, relative humidity 80%, light/dark 16 h/8 h), and the germination rate was recorded every 2 days until the 20th day.

For measurement of mineral element contents of hull, 2-g hulls from each of 10 random individuals for each hull type (TKB, TKW, TNB, and TNW) in the "immortalized F_2" population were harvested and then mixed together. In total, four mixing pools, together with DHS and XBK, were tested using reported methods (Ma et al., 2003).

A two-sided *t*-test and least significant difference test were performed for corresponding phenotypic measures in IBM SPSS.

Construction of Genetic Map and QTL Mapping

Previously identified Indels from DHS and XBK were used to design primers with Primer 3, and 891 Indel markers were selected from the marker set to verify polymorphism between the parents. Finally, 230 reliable markers were used to genotype the 551 individuals in the "immortalized F_2" population. Linkage analysis and QTL mapping were carried out using QTL IciMapping (Meng et al., 2015). Markers were grouped at LOD = 3.0, and QTLs were mapped by the ICIM-ADD method with LOD $\geq$ 2.5. Phenotypic data from multiple environments together with genotypic data were used for QTL analysis. The QTLs that were found in at least two environments were considered candidates. To confirm the chromosome segments containing QTLs of SHPR, we subjected $BC_{4(DHS)}$ individuals to genotyping analysis using 230 markers covering the whole genome of DHS as illustrated in Supplemental Figure 14.

To fine map *Ccph1* and *Ccph2*, we developed and verified more Indel and SNP markers within flanking markers from a previous marker set of RNA-seq. At the same time, we used an additional F_2 population of 15 562 individuals to identify more recombinants in a linkage block. Primers of these markers are listed in the Supplemental Table 9.

ACCESSION NUMBERS

The *C. aquatica* genome assembly has been deposited in the National Center for Biotechnology Information under BioProject accession code PRJNA509097. Other sequence release data will be deposited under the same BioProject and will be available upon acceptance of the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This research was supported by the National Natural Science Foundation of China (grant no. 31971753, 81703657), Key Project at Central Government Level: the Ability Establishment of Sustainable Use for Valuable Chinese Medicine Resources (2060302), the Fundamental Research Funds for Central Non-profit Scientific Institution of China (Y2017XM06, ZZ13-YQ-085), the Key Science and Technology Program of Biotechnology

and Medicine of Sichuan Province (2018NZDZX0001), and the National Basic Research Program of China (2014CB138705).

AUTHOR CONTRIBUTIONS

C.G., Y.W., A.Y., and J.H. are joint first authors. S.Z., L.H., and T.R. designed and advised on the experiments and wrote the manuscript. C.G., J.H., and F.H. revised the manuscript. C.G. and Y.W. performed the bioinformatics analyses. A.Y., C.X., S.L., J.H., Yibing Yuan, Yuan Yuan, and X.D. constructed the mapping population. J.G., Y. Yang, H.L., and N.Z. carried out the phenotyping assay. Y.H., K.Z., Z.J., X.W., T.J., Y.S., M.C., Y.W., and Z.L. constructed the genetic map.

ACKNOWLEDGMENTS

We thank Dr. Chunming Liu (Institute of Crop Science, CAAS), Dr. Yinxin Li (Institute of Botany, CAS), and Dr. Chunxiang Fu (Qingdao Institute of Bioenergy and Bioprocess Technology, CAS) for discussion about the structural difference between stony and papery hull; Prof. Michael Freeling (University of California-Berkeley, USA.) for help about the WGD event analysis; and Dr Xu Zheng (Henan Agricultural University, China) for advice about the manuscript. No conflict of interest declared.

Received: June 15, 2019
Revised: September 24, 2019
Accepted: November 13, 2019
Published: November 24, 2019

REFERENCES

Ahmad, S., Veyrat, N., Gordon-Weeks, R., Zhang, Y., Martin, J., Smart, L., Glauser, G., Erb, M., Flors, V., and Frey, M. (2011). Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiol. **157**:317–327.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**:796–815.

Arora, R.K. (1977). Job's-tears (*Coix lacryma-jobi*)—a minor food and fodder crop of northeastern India. Econ. Bot. **31**:358–366.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. Genome Res. **14**:988–995.

Blanco, E., Parra, G., and Guigó, R. (2007). Using geneid to identify genes. Curr. Protoc. Bioinformatics **18**:4.3.1–4.3.28.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'donovan, C., and Phan, I. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. **31**:365–370.

Bruijn, W.J., Gruppen, H., and Vincken, J.-P. (2018). Structure and biosynthesis of benzoxazinoids: plant defence metabolites with potential as antimicrobial scaffolds. Phytochemistry **155**:233–243.

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. **268**:78–94.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. Nat. Biotechnol. **31**:1119–1125.

Cai, Z., Liu, H., He, Q., Pu, M., Chen, J., Lai, J., Li, X., and Jin, W. (2014). Differential genome evolution and speciation of *Coix lacryma-jobi* L. and *Coix aquatica* Roxb. hybrid guangxi revealed by repetitive sequence analysis and fine karyotyping. BMC Genomics **15**:1025.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. BMC Genomics **7**:327.

Chai, M., Zhou, C., Molina, I., Fu, C., Nakashima, J., Li, G., Zhang, W., Park, J., Tang, Y., and Jiang, Q. (2016). A class II KNOX gene, KNOX4, controls seed physical dormancy. Proc. Natl. Acad. Sci. U S A **113**:6997–7002.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **5**:4–10.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**:3674–3676.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics **22**:1269–1271.

Dorweiler, J., Stec, A., Kermicle, J., and Doebley, J. (1993). *Teosinte glume architecture* 1: a genetic locus controlling a key step in maize evolution. Science **262**:233–235.

Dorweiler, J.E., and Doebley, J. (1997). Developmental analysis of *teosinte glume architecture* 1: a key locus in the evolution of maize (Poaceae). Am. J. Bot. **84**:1313–1322.

Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. Bioinformatics **21**:i152–i158.

Feraru, E., Paciorek, T., Feraru, M.I., Zwiewka, M., De Groodt, R., De Rycke, R., Kleine-Vehn, J., and Friml, J. (2010). The AP-3 β adaptin mediates the biogenesis and function of lytic vacuoles in *Arabidopsis*. Plant Cell **22**:2812–2824.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., and Mistry, J. (2013). Pfam: the protein families database. Nucleic Acids Res. **42**:D222–D230.

Flora of China Editorial Committee. (1978). Flora Reipublicae Popularis Sinicae (Beijing: Science Press).

Florea, L., Song, L., and Salzberg, S.L. (2013). Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. F1000Res. **2**:188.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. **33**:D121–D124.

Griffiths-Jones, S., Grocock, R.J., Van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. **34**:D140–D144.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59**:307–321.

Haas, B., and Papanicolaou, A. (2016). TransDecoder (find coding regions within transcripts). https://github.com/TransDecoder/TransDecoder/wiki.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. **9**:R7.

Hallauer, A.R., Carena, M.J., and Miranda Filho, J.B. (2010). Quantitative Genetics in Maize Breeding (New York: Springer Science and Business Media Press).

Han, Y., and Wessler, S.R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. **38**:e199.

Huang, H., Lu, P., Zhu, Y., and Li, Y. (1995). Ecotypes, diversities and utilization of the *Coix* varieties from China. Crop Germplasm **4**:4–8.

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. Nature **436**:793–800.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. **110**:462–467.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28**:27–30.

Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. **44**:e89.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736.

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics **5**:59.

Kretzschmar, D., Poeck, B., Roth, H., Ernst, R., Keller, A., Porsch, M., Strauss, R., and Pflugfelder, G.O. (2000). Defective pigment granule biogenesis and aberrant behavior caused by mutations in the *Drosophila* AP-3β adaptin gene ruby. Genetics **155**:213–223.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13**:2178–2189.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., and Dorschner, M.O. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science **326**:289–293.

Lim, T.K. (2012). Edible Medicinal and Non-medicinal Plants (Dordrecht: Springer Press).

Liu, X., Zhang, X., Rong, Y.-Z., Wu, J.-H., Yang, Y.-J., and Wang, Z.-W. (2015). Rapid determination of fat, protein and amino acid content in *Coix* seed using near-infrared spectroscopy technique. Food Anal. Methods **8**:334–342.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Ma, J.F., Higashitani, A., Sato, K., and Takeda, K. (2003). Genotypic variation in silicon concentration of barley grain. Plant Soil **249**:383–387.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**:2878–2879.

Mangelsdorf, P.C., and Reeves, R.G. (1938). The origin of maize. Proc. Natl. Acad. Sci. U S A **24**:303–312.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., and Gonzales, N.R. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. **39**:D225–D229.

Marie-Nelly, H., Marbouty, M., Cournac, A., Flot, J., Liti, G., Parodi, D.P., Syan, S., Guillén, N., Margeot, A., and Zimmer, C. (2014). High-quality genome (re)assembly using chromosomal contact data. Nat. Commun. **5**:5695.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., and Russell, J. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature **544**:427–433.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., and Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop J. **3**:269–283.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics **29**:2933–2935.

Normile, D. (2003). The new face of traditional Chinese medicine. Science **299**:188–190.

Ottoboni, L.M., Leite, A., Targon, M.L.N., Crozier, A., and Arruda, P. (1990). Characterization of the storage protein in seed of *Coix lacryma-jobi* var. Adlay. J. Agric. Food Chem. **38**:631–635.

Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics **23**:1061–1067.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., and Poliakov, A. (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457**:551–556.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. Bioinformatics **21**:i351–i358.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. Nat. Rev. Genet. **3**:429–441.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., and Graves, T.A. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326**:1112–1115.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. U S A **108**:4069–4074.

She, R., Chu, J.S.-C., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. **19**:143–149.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212.

Singh, R., Low, E.-T.L., Ooi, L.C.-L., Ong-Abdullah, M., Ting, N.-C., Nagappan, J., Nookiah, R., Amiruddin, M.D., Rosli, R., and Manaf, M.A.A. (2013). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature **500**:340–344.

Sirén, J., Välimäki, N., and Mäkinen, V. (2014). HISAT2-fast and sensitive alignment against general human population. IEEE/ACM Trans. Comput. Biol. Bioinform. **11**:375–388.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics **19**:ii215–ii225.

Staudt, G. (1961). The origin of cultivated barleys: a discussion. Econ. Bot. **15**:205–212.

Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., Zhang, X., Swarm, S.A., Zhou, L., and Zhang, Z.J. (2015). *GmHs1-1*, encoding a calcineurin-like protein, controls hard-seededness in soybean. Nat. Genet. **47**:939–943.

Taketa, S., Amano, S., Tsujino, Y., Sato, T., Saisho, D., Kakeda, K., Nomura, M., Suzuki, T., Matsumoto, T., and Sato, K. (2008). Barley grain with adhering hulls is controlled by an ERF family transcription factor gene regulating a lipid biosynthesis pathway. Proc. Natl. Acad. Sci. U S A **105**:4062–4067.

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. **43**:e78.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. **29**:22–28.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One **9**:e112963.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436**:714–719.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2009). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics **26**:136–138.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., and Guo, H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. **40**:e49.

Wang, J., Liu, L., Ball, T., Yu, L., Li, Y., and Xing, F. (2016). Revealing a 5,000-y-old beer recipe in China. Proc. Natl. Acad. Sci. U S A **113**:6444–6448.

Xi, X.-J., Zhu, Y.-G., Tong, Y.-P., Yang, X.-L., Tang, N.-N., Ma, S.-M., Li, S., and Cheng, Z. (2016). Assessment of the genetic diversity of different Job's tears (*Coix lacryma-jobi* L.) accessions and the active composition and anticancer effect of its seed oil. PLoS One **11**:e0153269.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268.

Yang, X., and Jiang, L. (2010). Starch grain analysis reveals ancient diet at Kuahuqiao site, Zhejiang Province. Chin. Sci. Bull. **55**:1150–1156.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**:1586–1591.

Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., Xie, M., Zeng, P., Yue, Z., and Wang, W. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. Nat. Biotechnol. **30**:549–554.

Zhang, P., Meng, X., Tang, X., Ren, L., and Liang, J. (2019). The effect of a coix seed oil injection on cancer pain relief. Support. Care Cancer **27**:461–465.

Zhu, F. (2017). *Coix*: chemical composition and health effects. Trends Food Sci. Technol. **61**:160–175.

Zwiewka, M., Feraru, E., Möller, B., Hwang, I., Feraru, M.I., Kleine-Vehn, J., Weijers, D., and Friml, J. (2011). The AP-3 adaptor complex is required for vacuolar function in *Arabidopsis*. Cell Res. **21**:1711–1722.