A functional chromogen gene *C* from wild rice is involved in a different anthocyanin biosynthesis pathway in *indica* and *japonica*

Weihua Qiao^{1¶*}, Yanyan Wang^{1¶}, Rui Xu¹, Ziyi Yang¹, Yan Sun², Long Su², Lizhen Zhang², Junrui Wang¹, Jingfen Huang¹, Xiaoming Zheng¹, Shijia Liu³, Yunlu Tian³, Liangming Chen³, Xi Liu³, Jinhao Lan^{2*}, Qingwen Yang^{1*}

¹Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China.

²Qingdao Agricultural University, Qingdao, China.

³Nanjing Agricultural University, Nanjing, China.

*Corresponding authors

E-mail: qiaowehua@caas.cn(W Q), jinhao2005@163.com(J L), yangqingwen@caas.cn(Q Y)

¶ These authors contributed equally

Abstract

Accumulation of anthocyanin is a desirable trait to be selected in rice domestication, but the molecular mechanism of anthocyanin biosynthesis in rice remains largely unknown. In this study, a novel allele of chromogen gene *C*, *OrC1*, from *Oryza rufipongon* was cloned and identified as a determinant regulator of anthocyanin biosynthesis. Although *OrC1* functions in purple apiculus, leaf sheath and stigma in *indica* background, it only promotes purple apiculus in *japonica*. Transcriptome analysis revealed that *OrC1* regulates flavonoid biosynthesis pathway and activates a few bHLH and WD40 genes of ternary MYB-bHLH-WD40 complex in *indica*. Differentially expressed genes and metabolites were found in the *indica* and *japonica* backgrounds, indicating that *OrC1* activated the anthocyanin biosynthetic genes *OsCHI*, *OsF3H*, *OsANS*, *OsINS* and *OsANR* and produced six metabolites independently. Artificial selection and domestication of *C1* gene in rice occurred on the coding region in the two subspecies independently. Our results reveal the regulatory system and domestication of *C1*, provide new insights into MYB transcript factor involved in anthocyanin biosynthesis, and show the potential of engineering anthocyanin biosynthesis in rice.

Author summary

Accumulation of anthocyanin is a selection trait in rice domestication, whereas the mechanisms regulating the anthocyanin biosynthetic pathway in rice remain unresolved. Here, a novel allele of chromogen gene *C* from wild rice (*Oryza rufipongon*) was identified as a determinant regulator of anthocyanin biosynthesis. A key question is to what extent the involvement of the *C1* gene can explain coloration variability of cultivated rice, where anthocyanin accumulation has been eliminated by artificial selection. Our results reveal the functional chromogen gene *C* from wild rice causes different coloration phenotypes, regulates various anthocyanin biosynthetic genes and produces different metabolites in *indica* and *japonica*. Artificial selection and domestication of the *C1* gene in rice only occurs within the coding region of the two subspecies independently.

Introduction

Most seed plants have different colors in different parts of seeds mostly due to the anthocyanin pigmentation [1]. Anthocyanins and pro-anthocyanins are a class of flavonoid, as one of the largest groups of secondary metabolites, and are widely distributed in plants [2]. Anthocyanins contribute to not only multiple physiological roles in the responses of plant to biotic and abiotic stresses, but also commercial value of plant products. It was reported that various anthocyanin also have a profound impact on food quality beneficial to human health [3], which is of significant interest to both crop breeders and consumers.

Biosynthesis and accumulation of anthocyanin depend on the inherent genetic factors and external environmental factors. Inherent factors in anthocyanin biosynthesis include the structural and regulatory genes. Structural genes encode enzymes, include phenylalanine ammonia lyase (PAL), chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid-3-hydroxylase (F3H), flavonoid-3'-hydroxylase (F3'H), dihydroflavanol reductase (DFR), and anthocyanidin synthase (ANS). The expression of these structural genes is controlled by transcription factors (TFs) and has been well characterized in a range of plant species [4-6]. A ternary MBW complex, comprising R2R3-MYB TFs, basic-helix-loop-helix (bHLH) TFs and WDR (WD-repeat) proteins, was believed to tightly regulate the common pathway of anthocyanins and pro-anthocyanins biosyntheses [7-10]. The regulating network of anthocyanin biosynthesis is well documented in maize and Arabidopsis but much less in rice. Sun et al. (2018) reported that a *C-S-A* gene system regulated anthocyanin pigmentation in rice hull. In this system, *C1* encodes a R2R3-MYB transcription factor and acts as a color-producing gene, and *S1* encodes a bHLH protein that functions in a tissue-specific manner. *C1* interacts with *S1* and activates the expression of *A1*, which encodes a dihydroflavanol reductase [11].

The R2R3-MYB TFs were often identified as determinants of variation in anthocyanin pigmentation and have been identified in many higher plant species [12-17]. In a previous study, *OsC1*, *OsRb* and *OsDFR* were identified as the determinants of anthocyanin biosynthesis in rice leaves [18]. The R2R3-MYB gene *OsC1* was previously isolated from cultivated rice through comparative mapping between rice and maize or according to the nucleotide sequence homology with known maize orthologues [19, 20], and was also cloned from cultivated rice using various methods recently [21, 22]. Several studies determined that *OsC1* acted as a chromogen gene and mainly functioned in apiculus and leaf sheath [23-27]. The apiculus, as the remnant of awn, maintains its color in some modern varieties and seems not to have undergone artificial selection during domestication [20]. The purple sheath trait was reported to be a morphological marker in rice [24], which can be easily observed at the seedling stage, and have often been used in the screening of the authentic hybrids. However, molecular function of *OsC1* was not fully understood, the exact genetic determinants for purple apiculus and leaf sheath in rice remain to be unraveled.

Rice ancestor wild rice (*Oryza rufipogon*) has accumulated anthocyanin in various tissues, most of *O. rufipogon* appear to have purple or red pigment mainly on awn, apiculus, stigma, pericarp, pulvinus, leaf blade, leaf sheath, internode and palea, but most cultivated rice have lost these pigments due to artificial selection. It had been demonstrated that *OsC1* was also a domestication-related gene for the loss of pigments in cultivated rice [28]. Identification of the *C1* allele in wild rice is of great significance for understanding the origin and evolution of rice through investigation of the genes controlling color formation. In this study, a R2R3 MYB transcription factor was fine mapped in the wild rice using a set of chromosome segment substitution lines. The *C* gene from wild rice, *OrC1*, was cloned for functional analysis, transcriptome and metabolome profiling and was shown to be a functional allele for anthocyanin biosynthesis. Interestingly, we found that six anthocyanin metabolites were simultaneously regulated by *OrC1* in different genetic backgrounds. Our findings emphasize the importance and value of using wild relatives to uncover useful genes that have been lost during crop domestication in order to expand the genetic repertoire for rice domestication study and modern crop breeding.

Results

Fine mapping of *OrC1* using chromosome segment substitution lines

In our previous study, a set of Chinese common wild rice chromosome segment substitution lines (CSSL) was developed [29]. The donor parent, *O. rufipogon*, has a significant purple coloration in apiculus, leaf sheath, and stigma and black hull. The recipient parent, an *indica* variety 9311, has no purple pigment in the whole plant. In total 150 CSSLs was investigated under five environments (S1 Table) for purple coloration. One hundred twenty-one SSR and 62 InDel markers, which were polymorphic between the two parents and evenly distributed on all 12 chromosomes, were selected for genotyping (S2 Table). In the CSSL populations, the purple apiculus, leaf sheath and stigma were completely correlated, while the purple or black hull was independently segregated. Using SSR/InDel genotyping, one QTL related to purple apiculus, leaf sheath and stigma (hereafter named purple coloration trait) was identified under five environments and located near RM314 on chromosome 6. (Fig 1a)

One chromosome segment substitution line, CSSL52, which harbors this QTL and purple coloration trait, was selected for further fine mapping of purple genes. CSSL52 carried only one wild rice introgression segment and exhibited purple apiculus, leaf sheath and stigma (Fig 1a and Fig 2). We observed no significant difference in other agronomic traits such as plant height and heading date, between 9311 and CSSL52, indicating that the single introgression segment from wild rice was only involved in purple coloration. CSSL52 was back-crossed with 9311 to produce F₁ plants, which showed purple apiculus, leaf sheath and stigma (Fig 2). The color of apiculus, leaf sheath and stigma in F₂ population was completely correlated. A segregation ratio 3:1 of purple coloration to achromatic was observed in 1,964 F₂ individuals (S3 Table), indicating that the purple

coloration trait was controlled by a single dominant gene. Using F₂ segregating population, this gene was narrowed down to a 131-kb region between markers RM19552 and RM19561 with recessive class analysis (Fig 1c), in which there were 17 predicted open reading frames (ORF) (S4 Table). This region contains an expressed gene (*LOC_Os06g10350*) encoding a R2R3 MYB transcription factor that has been reported as the rice homolog of maize *C1* gene [20]. Then we compared the nucleotide sequences of CSSL52 and 9311 and *japonica* Nipponbare across the fine-mapped region. Sequence analysis of *Os06g10350* revealed that 9311 contained a 10-bp deletion at the start of the third exon, and Nipponbare contains a 3-bp deletion in the second exon (Fig 1d). Therefore, we focused on *Os06g10350* as a candidate for purple coloration and named it as *OrC1*. Furthermore, a near isogenic line (NIL) of *OrC1* was developed using the positive F₂ individual of CSSL52/9311.

Characterization of OrC1 from wild rice

OrC1 encodes a 272 amino-acid protein containing two highly conserved SANT domains, similar to the R2 and R3 motifs of R2R3-MYB transcriptional factors in plants. A phylogenetic tree was constructed using the amino acid sequences of *OrC1* and other anthocyanin biosynthesis related genes from several plant species. Phylogenetic analysis showed that *OrC1* exhibited highest homology to *OsC1* in cultivated rice and R2R3 MYB transcription factors in maize (Fig 3a). To investigate the expression patterns of *C1* in wild and cultivated rice, the levels of *C1* mRNA were quantified in different tissues including leaves, leaf sheaths, stems, and panicles of NIL-*OrC1* and 9311, at the booting stage and 3 days after flowering. No significant difference was detected between NIL and 9311. *C1* transcripts were detected at a high level in leaf sheaths, the lowest level of *OrC1* transcripts was detected in stems. Three days after flowering, the expression levels of *C1* in panicles was significantly increased in both NIL and 9311 (Fig 3b).

To examine the subcellular localization of OrC1 protein, an OrC1:GFP fusion gene was generated and transformed into rice protoplasts and tobacco leaves under the control of the cauliflower mosaic virus 35S promoter. As shown in Fig 3c, the control GFP protein was distributed throughout the entire cell, whereas the OrC1-GFP fluorescent signals were exclusively localized in nucleus, consistent with the function of OrC1 as a transcription factor.

Transcriptome analysis revealed differentially expressed genes between NIL and 9311

To further investigate the molecular mechanism of OrC1-mediated anthocyanin synthesis pathway, we performed a transcriptome analysis with the leaf sheath of 9311 and NIL at the seedling stage, because the purple color phenotype were most obvious at this stage. Three biological replicates were tested. The principal component analysis of the samples based on the number of fragments per kilobase of exon per million fragments mapped (FPKM) values showed that one replicate of NIL (OrC1-3) did not cluster with the other two (S1 Fig), so only two replicates were used for further analysis. In total, 2,388 differentially expressed genes (DEGs) were identified with the stringent criteria ($|\log_2(FC)| > 1$, and FDR < 0.05) (S5 Table). Of these DEGs, 1,225 were up-regulated and

1,163 were down-regulated in NIL compared with 9311. Furthermore, all DEGs were assigned to 98 KEGG. KEGG pathway enrichment analysis revealed that *OrCI* mostly affected the expression of genes involved in phenylpropanoid biosynthesis, diterpenoid biosynthesis, plant hormone signal transduction, and flavonoid biosynthesis (Fig 4a). To facilitate global analysis of gene expression, we investigated the possible mechanisms of significant DEGs by Gene Ontology (GO) analysis. The 2,388 DEGs were categorized into three main GO categories of biological process, cellular component, and molecular function. They were significantly assigned to biological process GO term, including cellular and metabolic processes. In terms of the cellular component GO term, they were mainly associated with cell, cell part, membrane and organelle. In respect to the biological process GO term, they were significantly involved in binding and catalytic activities (S2 Fig). The most enriched GO terms were metabolic process and catalytic activity.

Seven structural genes involved in the anthocyanin metabolic pathway were used for the validation of the sequencing results. The relative gene expression levels and FPKM (RNA-Seq) of the seven genes showed similar regulatory patterns in the NIL compared with 9311. A linear regression coefficient (R^2) of 0.9711 was obtained between the results of qRT-PCR and RNA-Seq (S3 Fig). These consistencies validated the results of RNA-Seq analysis. As shown in Fig 4c. All seven anthocyanin biosynthetic genes were consistently up-regulated in the NIL compared with those in 9311, among them, expression levels of *OsPAL*, *OsDFR*, *OsANS*, *OsF3'H*, and *OsF3H*, were significantly increased in the NIL (Fig 4b).

Anthocyanin biosynthesis are largely influenced by some key pathways and MBW complex. According to transcriptome results, 116, 84, and 19 DEGs were detected in phenylpropanoid biosynthesis, MAPK signaling and flavonoid biosynthesis pathways, respectively (Fig 5a). At the same time, five *R2R3-MYB* transcription factors, six *bHLH* proteins and four *OsWD40* proteins were also found in DEGs (Fig 5b), including the cloned gene *Rc* (*Os07g0211500*) which was involved in proanthocyanin synthesis.

Overexpression of *OrC1* results in purple apiculus in transgenic Nipponbare

To explore the biological function of *OrC1*, its ORF was transformed into *japonica* cultivar Nipponbare with the 35S promoter. More than 20 independently transgenic lines carrying 35S:*OrC1* (OE) were generated. The transgenic individuals were confirmed by PCR analysis with gene-specific primers. Ten positive transgenic lines were analyzed by qRT-PCR to determine the transgene expression level and two independent lines, OE1 and OE2, which differed in transcription levels were selected for additional analysis (S4 Fig). Under natural conditions, only the apiculus of T_1 plants showed a purple pigmentation phenotype, the sheath was as the same as wild type Nipponbare. Purple apiculus of transgenic plants could be observed when panicles were exposed to sunlight at the initial heading stage. It turned into brown color at the wax ripeness stage. T_2 transgenic plants showed a perfect segregation ratio 3:1 of purple or brown apiculus to normal. Both purple and brown apiculus could be easily distinguished at the fully ripen stage. There is no

difference between the two *OrCl* overexpression lines, OE1 and OE2, and no other phenotypes were different between OE and Nipponbare except apiculus color (Fig 6a). Seven above mentioned structural genes were also detected by qRT-PCR, only *OsANS*, *OsCHI* and *OsF3H* showed higher expression level in OE lines than Nipponbare (Fig 6b). Other TFs, especially MBW complex, might involve in different pigmentation of Nipponbare and 9311, we compared other bHLH and WDR genes in NIL, 9311 and Nipponbare. All the gene sequences in NIL is exactly the same as 9311, which indicated that only one coloration gene *OrCl* in NIL was from wild rice. Comparison between 9311 and Nipponbare showed that six bHLH genes and three WDR genes have non-synonymous SNP (S5 Fig). Therefore, we predicted that these genes might have different functions in anthocyanin biosynthesis in different tissues.

Identification of differential metabolites in the anthocyanin biosynthesis pathways

Liquid chromatography-electrospray ionization-tandem mass spectrometry analysis was used to quantify the anthocyanin content in leaf sheath of both NIL and 9311. A total of 28 metabolites of anthocyanins and pro-anthocyanidins were detected, and the relative contents for each metabolite were normalized before being subjected to downstream data analyses. Among them, 11 metabolites in NIL were significantly higher than that in 9311, two metabolites were higher in 9311 than that in NIL. For Nipponbare and transgenic plants, given the OE lines only have purple apiculus and no color in other organs, the leaf sheath and upper one third of the hull of two OE lines and Nipponbare were used for metabolite analysis. A total of 13 metabolites of anthocyanins were detected, and among them six metabolites in apiculus (hull) of OE were significantly higher than that in Nipponbare (Fig 7). In sheath, there is no significant difference between OE and Nipponbare. No difference was found in apiculus or sheath between the two OE lines.

The differentially accumulated metabolites (DAM) between pairs of samples (NIL-Sheath vs 9311-Sheath, OE-Hull vs Nip-Hull, OE-Hull vs OE-Sheath, OE-Sheath vs Nip-Sheath) were determined based on the variable importance in projection (VIP) ≥ 1 , and fold change ≥ 2 or ≤ 0.5 [30-31]. Comparative analysis of the four groups of DAMs revealed that six common anthocyanin metabolites were differentially accumulated in NIL-Sheath vs 9311-Sheath, OE-Hull vs Nip-Hull, and OE-Hull vs OE-Sheath, including rosinidin, delphinidin and four cyanidins (S6 Table). And all of them are up regulated in NIL compared with 9311, and in OE lines compared with Nipponbare.

Three genes were involved in delphinidin and cyanidin biosynthesis pathway in the flavonoid biosynthesis KEGG pathway, map 00941 (https://www.genome.jp/kegg-bin/show_pathway?map00941) (S6 Fig). According to our transcriptome profile results, two genes, *Os06g0626700* (*OsINS*) and *Os01g0372500* (*OsANS*), were upregulated and one gene *Os04g0630800* (*OsANR*) was down regulated in NIL compared with 9311 (S6 Fig, blue rectangle). *OsANS* was also upregulated in OE compared with Nipponbare (Fig 6). Therefore, we tested *OsINS* and *OsANR* in OE lines and Nipponbare. Consistent with expression profile in NIL and 9311, the *OsINS* was significantly upregulated and the *OsANR* was downregulated in OE lines compared with

Nipponbare, indicated that *OrCI* regulated some of the same genes in *indica* and *japonica* background.

Haplotype analysis

The promoter and coding regions of *CI* alleles from 180 rice accessions were isolated and analyzed. The haplotype analysis of promoter region showed that the variation in this region was not associate with coloration traits (S7 Fig). For coding region, 12 haplotypes (H_1 to H_12) were identified. Among them, H_2, 9, 10, 12 were functional, and H_1, 3, 4-8, 11 were non-functional with different deletions (Fig 8a). Only two *O. rufipogon* and seven *O. nivara* have non-functional allele. The combination of the phenotypes and genotypes of the association panel showed that 24 cultivated rice accessions with non-functional allele showed at least one coloration trait (purple hull or stigma), indicating the presence of other determinant factors besides *CI*. All *indica* and wild rice in H_2 displayed purple apiculus, leaf sheath and stigma, almost all *japonica* in H_2 only displayed purple apiculus (S7 Table). Haplotype network analysis showed that the most frequent non-functional alleles in *indica* and *japonica* are different, which were H_1 and H_7 in *indica* and *japonica*, respectively (Fig 8b), indicating multiple origins of *CI* in two subspecies.

Discussion

Although the spatiotemporal regulation of anthocyanin biosynthesis has been well documented and the chromogen *C* gene, a major coloration gene, has also been fine mapped in cultivated rice [20, 32, 33], the allele in wild rice or full functional allele of *C* gene in rice has not been cloned. Our study identified a functional *C* gene, which encodes a R2R3-MYB transcription factor in wild rice. Given the high heterogeneity in the genome of *O. rufipogon*, it is difficult to clone novel genes which have been lost or weakened in cultivated rice during domestication. CSSLs from inter specific hybridization represent a powerful and useful genetic resource for genome research [34]. Fine mapping was employed firstly in search of anthocyanin genes using a set of CSSLs [29]. The CSSLs in this study has been testified to be an excellent platform for large-scale gene discovery in wild rice [35-37].

Anthocyanin biosynthesis is affected by a variety of environmental factors, including light, temperature, and fertilization [38, 39]. In order to eliminate the environmental effects on coloration, the purple coloration of CSSL population was detected under five environments. A stably expressed QTL associated with purple coloration was fine mapped on same location of cultivated rice *OsCI*, which was located on chromosome 6 by GWAS [28], hence we named this gene *OrCI*. According to the phylogenetic analysis and sequence blast with homologs in *indica* 9311 and *japonica* Nipponbare, *OrCI* is a complete allele of *C* gene, encoding a R2R3 MYB transcriptor. Although the other organs of NIL such as stem and leaf have no purple coloration, the *CI* gene constitutively expressed in all tissues, and has consistent expression patterns in both NIL and 9311 (Fig 3b). We also sequenced the promoter region of *CI* gene in 180 rice accessions, no haplotype

was associated with purple traits (S7 Fig), indicating that the anthocyanin biosynthesis in rice was mostly controlled by C1 protein function but not at gene expression level. As domestication related gene, we deduced that the artificial selection occurs in its coding region but not the promoter region.

Anthocyanin biosynthesis in different rice tissues is primarily controlled by R2R3-MYB transcription factors and the ternary MBW complexes, which comprised of R2R3-MYBs, bHLHs and WD40 [40, 41]. In this study, the NIL-*OrC1* in *indica* background shown a co-segregated purple leaf sheath, stigmas and apiculus, the purple coloration is completely related and shown a 3:1 segregation in F₂ population of NIL/9311, which indicated that the coloration of three tissues was controlled by a single dominant gene *OrC1*. However, the OE lines of *OrC1* in *japonica* background only had a purple apiculus but non-colored leaf sheath and stigmas. The different purple traits in NIL and OE are due to their different MBW complex. At least six bHLH genes and three WDR genes have non-synonymous SNP between 9311 and Nipponbare (S5 Fig). The MBWs containing these bHLH or WD40 might have different functions in anthocyanin biosynthesis in different tissues. Moreover, both 9311 and Nipponbare have a non-functional allele of *Os04g0557500* (S5 Fig), which was reported as a tissue regulator and controls pigmentation in hull [11]. Hence, both NIL and OE accumulated pigments only in apiculus but not the hull. Previously studies demonstrated that *OsC1* is the determinant factor of anthocyanin biosynthesis in leaf sheath and apiculus [25, 26], but has not been reported to be involved in pigmentation in stigmas. It is likely that an MBW complex consisting of *OrC1* and a tissue-specific regulator regulated the anthocyanin biosynthesis in stigmas of NIL. Further work is needed to determine the exact bHLH TF combined with *OrC1*. Moreover, haplotype analysis showed that the functional *C1* allele led to anthocyanin accumulation in three tissues in *indica* and only in apiculus in *japonica* (Fig 8 and S7 Table), consistent with the results of *OrC1* function in 9311 and Nipponbare (Fig 2 and Fig 6). Besides *C1*, other redundant genes must be involved in the regulation of anthocyanin biosynthesis in rice stigma.

Transcriptome profiles revealed that almost all the structural genes in the flavonoid biosynthesis pathway were induced by *OrC1* in NIL, but only *ANS*, *CHI*, and *F3H* were induced in OE (Fig 4 and Fig 6). *OrC1* or MBW complex regulate different branches of flavonoid biosynthesis pathway by activating a subset of structural genes. In rice, structure genes are activated by MYB and MBW with redundancy [41], we believe that some bHLH TFs of 9311 is functional and some of Nipponbare is non-functional, the coloration of NIL was due to both *OrC1* and MBW complex, but the purple apiculus of OE lines was only caused by *OrC1* protein. Furthermore, *OrC1* is involved not only in flavonoid and phenylpropanoid (a basic skeleton of flavonoids) biosynthesis, but also in singling pathways like MAPK, and regulated bHLH and WD40 genes (Fig 5). This indicates that *C1* gene is involved in other physiological processes besides anthocyanin biosynthesis.

The exact metabolites of anthocyanin that *C1* gene produced is still not clear. Shin et al. (2006) transformed maize *C1* and *R-S* regulatory genes into rice using endosperm specific promoters, the produced flavonoids mostly are anthocyanins including dihydroquercetin (taxifolin),

dihydroisorhamnetin (3'-O-methyl taxifolin) and 3'-O-methyl quercetin [42]. We have detected six DAMs accumulated in both NIL and OE line (S6 Table). Because we only tested the anthocyanin metabolites, so other flavonoids metabolites are not excluded. Three genes involved in the common DAM biosynthesis pathway have the same expression pattern in NIL_vs_9311 and OE_vs_Nipponbare. Take together the seven structural gene expression profiles, we deduced that *OrCI* regulated *OsCHI*, *OsF3H*, *OsANS*, *OsINS*, and *OsANR* independently from the MBW complexes. These results set a foundation to understand the regulatory mechanisms of *OsCI* in the anthocyanin biosynthesis pathway.

The *CI* involved anthocyanin biosynthesis pathway pre-exists in wild rice but is absent in most cultivated rice, indicating a strong negative human selection of this trait. Haplotype network analysis showed that the functional mutations of *CI* had multiple origins and been selected independently in two subspecies (Fig 8). There are two main hypotheses for the cultivated rice domestication: single origin and multiple origin [43]. Some well-documented domestication genes of rice, such as *sh4* for seed shattering [44], and *prog1* for erect growth [45], supporting the single-origin hypothesis. Our haplotype analysis of *CI* supports multiple origin. More interestingly, the non-functional *CI* allele is much closer to *O. nivara* than *O. rufipogon*, which supports the theory that the origin of cultivated rice is *O. nivara*, an annual wild rice species that is generally considered an intermediate between *O. rufipogon* and cultivated rice; however, the direct ancestor of cultivated rice remains controversial.

Materials and Methods

Plant materials

A set of CSSLs produced from common wild rice (*O. rufipogon*) as the donor and an elite *indica* variety, 9311, as the recurrent parent was developed in our laboratory [29]. The CSSLs and 9311 were grown under five environmental conditions as shown in S1 Table. A panel of 180 rice accessions (S7 Table) including 89 *O. sativa indica*, 48 *O. sativa japonica*, 21 *O. rufipogon* and 22 *O. nivara*, was selected from a natural rice germplasm population that is maintained in our laboratory. The purple traits were recorded at two environments at Beijing and Sanya, respectively.

Sequence and phylogenetic analyses

The predicted amino acid sequences of *CI* in *O. rufipogon* and MYB homologous proteins in other species were downloaded from NCBI (<http://phytozome.jgi.doe.gov/pz/portal.html>). Multiple sequence alignments were performed with DNAMAN software. A phylogenetic tree was constructed using the Maximum Likelihood method and software MEGA version 5.1 [46]. The de novo genomic sequencing of the panel of 180 rice accessions was performed by our laboratory previously. *CI* gene

genomic sequence were isolated from the genomic sequencing data. The SSR primers used in this study were previously published [47], InDel primers were designed in our laboratory. The other primers used in this study were designed online at the NCBI website (<https://www.ncbi.nlm.nih.gov/>) based on the 9311 reference genome sequence. Alignments were performed on the Gramene (<http://www.gramene.org/>) website to ensure the accuracy of the location and the specificity of the primers. Sequences of all primers used in this study are shown in [S2 Table](#).

Subcellular localization

The full-length coding region of *OrC1* was inserted into the PAN580-GFP plasmid to generate 35S::OrC1-GFP constructs and transformed into rice protoplasts and tobacco leaves. The GFP fluorescence in rice protoplasts and leaf epidermal cells were observed using a laser scanning confocal microscope (LSM880, Leica)

RNA extraction and real time PCR

Total RNA was extracted from different rice tissues using the RNA RNeasy Plant Mini Kit (Qiagen, Beijing, China), followed by treatment with RNase-free DNase (TaKaRa, Dalian, China) to remove genomic DNA contamination. qRT-PCR was performed using an ABI 7500 real time PCR system (Applied Biosystems) following the manufacturer's instructions. The poplar endogenous control gene (Actin) was used as an internal control. Gene-specific primers are listed in [S2 Table](#). Relative expression was calculated using the $2^{-[\Delta][\Delta]Ct}$ method [48]. Each sample was amplified in triplicate.

Transcriptome analysis

The leaf sheath of 15-day seedlings of NIL-*OrC1* and 9311 were collected for RNA extraction. RNA samples were sent to Genedenovo Biotechnology Co., Ltd (Guangzhou, China) for RNA sequencing. The rRNA removed reads of each sample were then mapped to reference genome by TopHat2 [49]. The reconstruction of transcripts was carried out with Cufflinks (<http://cole-trapnell-lab.github.io/cufflinks/install/>) and TopHat2. Principal component analysis was performed with R package Rmodest (<http://www.r-project.org/>). Data analyses, including DEGs, GO and pathway enrichment, were performed by Genedenovo Biotechnology Co., Ltd.

Rice transformation

Agrobacterium-mediated transformation was performed by Biorun Biosciences Co., Ltd (Wuhan, China) following their standard procedures and protocols.

Anthocyanin metabolite determination

The leaf sheath of 15-day seedlings of NIL, 9311, OE lines and Nipponbare, the upper one third of hull at 10 days after flowering of OE and Nipponbare were collected. The sample extracts were analyzed using an LC-ESI-MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system, www.shimadzu.com.cn/; MS, Applied Biosystems 4500 Q TRAP, www.appliedbiosystems.com.cn/). The sample preparation, extract analysis, metabolite identification and quantification were performed at Wuhan MetWare Biotechnology Co., Ltd. (www.metware.cn) following their standard procedures and were previously described in details by [50-52]

Acknowledgements

We especially thank Dr. Peng Zhang (The University of Sydney, Australia) and Dr. Hao Chen (Huazhong Agricultural University, China) for critical reading of the manuscript. This work was supported by the National Key R&D Program of China (2016YFD0100101) and the National Natural Science Foundation of China (31471471).

Author contributions

WH Qiao, JH Lan and QW Yang conceived and designed the experiments. WH Qiao and YY Wang performed the experiments and wrote the paper. R Xu, ZY Yang, JR Wang and JF Huang analyzed the data. Y Sun, L Su and LZ Zhang performed fine mapping. SJ Liu, YL Tian, LM Chen and X Liu contributed to field investigation. All authors read and approved the final manuscript.

Supporting information

S1 Table. The rice crop locations used in this experiment

S2 Table. Informations of the primers used in this study

S3 Table. The segregation of purple coloration traits in the CSSL52/9311 F₂ population

S4 Table. Predicted open reading frames in the mapped location

S5 Table. DEGs in the transcriptome profile of 9311_vs_NIL

S6 Table. Differentially accumulated anthocyanins between 9311_vs_NIP and Nipponbare_vs_OE

S7 Table. Phenotypes of coloration of a panel of rice accessions

S1 Fig. Principal component analysis of the transcriptome samples. P93-11 and OrC1 represent those for 9311 and NIL, respectively, One replicate of NIL that is clearly different from other two was deleted.

S2 Fig. Gene ontology enrichment of DEGs in leaf sheath of NIL and 9311at the seedling stage.

S3 Fig. Regression analysis between the results of RNA-Seq and qRT-PCR experiments.

S4 Fig. Identification of overexpression lines of *OrC1*. 35S:OrC1 was transformed into Nipponbare by an *Agrobacterium*-mediated transformation. (a) Ten transgenic lines was confirmed by PCR analysis using genomic DNA with gene-specific primers. (b) Ten positive transgenic lines were analyzed by qRT-PCR to determine the OrC1 expression level. Nip:Nipponbare, OE: overexpression line.

S5 Fig. Comparison of bHLH and WDR genes between Nipponbare and 9311. The genomic sequence of each gene was from <http://oge-ensembl.gramene.org/index.html>, red rectangles indicate non-synonymous SNPs.

S6 Fig. KEGG pathway map of favonoid biosynthesis in 9311_vs_NIL transcriptome profile. The red and green rectangles indicate up and down regulation, respectively. Anthocyanin biosynthesis pathway was marked using a blue rectangle, ID of the three genes in each pathway were showed inside the blue rectangle, the number in each bracket behind ID represents $\log_2(FC)$.

S7 Fig. Genotype analysis of promoter region of *CI*. (a) Sequence polymorphism of different haplotypes of *CI* promoter from 180 accessions. No. of Acc.: number of accessions, I: *O. sativa indica*, J: *O. sativa japonica*, R: *O. rufipogon*, N: *O. nivara*. (b) Haplotype network of *CI* promoter.

Reference

1. Aizza L, Dornelas MC. A genomic approach to study anthocyanin synthesis and flower pigmentation in

- passionflowers. J Nucleic Acids. 2011; 37: 15-17. <https://doi.org/10.4061/2011/371517> PMID: 21772993
2. Xu WJ, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015; 20: 176-185. <https://doi.org/10.1016/j.tplants.2014.12.001>
3. Vinayagam R, Xu B. Antidiabetic properties of dietary flavonoids: a cellular mechanism review. Nutr. Metab. 2015; 12: 60. <https://doi.org/10.1186/s12986-015-0057-7>
4. Honda C, Kotoda N, Wada M, Kondo S, Kobayashi S, Soejima J, et al. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. Plant Physiol. Biochem. 2002; 40: 955-962. [https://doi.org/10.1016/S0981-9428\(02\)01454-7](https://doi.org/10.1016/S0981-9428(02)01454-7)
5. Grotewold E. The genetics and biochemistry of floral pigments. Ann. Rev. Plant Biol. 2006; 57: 761-780. <https://doi.org/10.1146/annurev.arplant.57.032905.105248>
6. Hugueney P, Provenzano S, Verriès C, Ferrandino A, Meudec E, Batelli G, et al. A novel cation-dependent O-methyltransferase involved in anthocyanin methylation in grapevine. Plant Physiol. 2009; 150: 2057-2070. <https://doi.org/10.1104/pp.109.140376>
7. de Vetten N, Quattrocchio F, Mol J, Koes. The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. Genes Dev. 1997; 11: 1422-1434. <https://doi.org/10.1101/gad.11.11.1422>
8. Baudry A, Heim MA, Dubreucq B, Caboche M, Weissshaar B, Lepiniec L. *TT2*, *TT8*, and *TTG1* synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant J. 2004; 39: 366-380. <https://doi.org/10.1111/j.1365-313X.2004.02138.x>
9. Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J. Exp. Bot. 2011; 62:2465-2483. <https://doi.org/10.1093/jxb/erq442>
10. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015; 20: 176-185. <https://doi.org/10.1016/j.tplants>.
11. Sun X, Zhang Z, Chen C, Wu W, Ren N, Jiang C, et al. The *C-S-A* gene system regulates hull pigmentation and reveals evolution of anthocyanin biosynthesis pathway in rice. J. Exp. Bot. 2018; 69: 1485-1498. <https://doi.org/10.1093/jxb/ery001>. PMID: 29361187.
12. Chagne D, Lin-Wang K, Espley RV, Volz RK, How NM, Rouse S, et al. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. Plant Physiol. 2013; 161: 225-239. <https://doi.org/10.1104/pp.112.206771>.
13. Jin W, Wang H, Li M, Wang J, Yang Y, Zhang X. The R2R3 MYB transcription factor *PavMYB10.1* involves in anthocyanin biosynthesis and determines fruit skin colour in sweet cherry (*Prunus avium* L.). Plant Biotechnol. J. 2016; 14: 2120-2133. <https://doi.org/10.1111/pbi.12568>. PMID: 27107393
14. Jian W, Cao H, Yuan S, Liu Y, Lu J, Lu W, et al. *SIMYB75*, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits. Hortic. Res. 2019; 6:22. <https://doi.org/10.1038/s41438-018-0098-y>
15. Tian J, Peng Z, Zhang J, Song T, Wan H, Zhang M, et al. *McMYB10* regulates coloration via activating McF3'H and later structural genes in ever-red leaf crabapple. Plant Biotechnol. J. 2015; 13: 948-961. <https://doi.org/10.1111/pbi.12331>.

16. Jun JH, Liu C, Xiao X, Dixon RA. The transcriptional repressor MYB2 regulates both spatial and temporal patterns of proanthocyanidin and anthocyanin pigmentation in *Medicago truncatula*. *Plant Cell*. 2015; 27: 2860-2879. <https://doi.org/10.1105/tpc.15.00476>.
17. Xu Z, Yang Q, Feng K, Yu X, Xiong A. *DcMYB113*, a root-specific R2R3-MYB, conditions anthocyanin biosynthesis and modification in carrot. *Plant Biotechnol. J.* 2020; 18:1585-1597. <https://doi.org/10.1111/pbi.13325>.
18. Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H, et al. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* 2007; 49: 91-102. <https://doi.org/10.1111/j.1365-313x.2006.02958.x>.
19. Mikami I, Takahashi A, Khin-Thidar, Sano Y. A candidate for C (chromogen for anthocyanin) gene. *Rice Genet Newsl.* 2000; 17: 54-56.
20. Saitoh K, Onishi K, Mikami I, Thidar K, Sano Y. Allelic diversification at the C (*OsCI*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics*. 2004; 168: 997-1007. <https://doi.org/10.1534/genetics>.
21. Liu X, Sun X, Wang W, Ding H, Liu W, Li G, et al. Fine mapping of *Pa-6* gene for purple apiculus in rice. *Plant Biotechnol. J.* 2012; 55: 218-225. <https://doi.org/10.1007/s12374-011-0276-z>
22. Zhao SS, Wang CH, Ma J, Wang S, Tian P, Wang JL, et al. Map-based cloning and functional analysis of the chromogen gene *C* in rice (*Oryza sativa* L.). *J. Plant Biol.* 2016; 59: 496-505. <https://doi.org/10.1007/s12374-016-0227-9>.
23. Saitoh K, Onishi K, Mikami I, Thidar K, Sano Y. Allelic diversification at the C (*OsCI*) locus of wild and cultivated rice: nucleotide changes associated with phenotypes. *Genetics*. 2004; 168: 997-1007. <https://doi.org/10.1534/genetics>.
24. Gao DY, He B, Zhou YH, Sun LH. Genetic and molecular analysis of a purple sheath somaclonal mutant in *japonica* rice. *Plant Cell Rep.* 2011; 30: 901-911. <https://doi.org/10.1007/s00299-011-1004-3>.
25. Fan FJ, Fan YY, Du JH. Fine mapping of C (chromogen for anthocyanin) gene in rice. *Rice Sci.* 2008 15: 1-6. [https://doi.org/10.1016/S1672-6308\(08\)60012-8](https://doi.org/10.1016/S1672-6308(08)60012-8).
26. Chin H, Wu Y, Hour A, Hong C, Lin Y. Genetic and evolutionary analysis of purple leaf sheath in rice. *Rice*. 2016; 9: 8. <https://doi.org/10.1186/s12284-016-0080-y>
27. Hu W, Zhou T, Han Z, Tan C, Xing, Y. Dominant complementary interaction between *OsCI* and two tightly linked genes, *Rb1* and *Rb2*, controls the purple leaf sheath in rice. *Theor Appl Genet.* 2020; <https://doi.org/10.1007/s00122-020-03617-w>.
28. Huang X, Kurata N, Wei X, Wang Z, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*. 2012; 490: 497-501. <https://doi.org/10.1038/nature11532>.
29. Qiao WH, Qi L, Cheng ZJ, Su L, Li J, Sun Y, et al. Development and characterization of chromosome segment substitution lines derived from *Oryza rufipogon* in the genetic background of *O. sativa* spp. *indica* cultivar 9311. *BMC Genomics*. 2016; 17: 580. <https://doi.org/10.1186/s12864-016-2987-5>.
30. Yuan H, Zeng X, Shi J, Xu Q, Wang Y, Jabu D, et al. Time-course comparative metabolite profiling under osmotic stress in tolerant and sensitive Tibetan hullless barley. *BioMed Research International*, 2018; 9415409:

1-12. <https://doi.org/10.1155/2018/9415409>

31. Zhang Q, Wang L, Liu Z, Zhao Z, Zhao J, Wang Z, et al. Transcriptome and metabolome profiling unveil the mechanisms of *Ziziphus jujuba* Mill. peel coloration. Food Chemistry. 2020; 312:125903. <https://doi.org/10.1016/j.foodchem.2019.125903>
32. Liu X, Sun X, Wang W, Ding H, Liu W, Li G, et al. Fine mapping of *Pa-6* gene for purple apiculus in rice. Plant Biotechnol. J. 2012; 55: 218-225. <https://doi.org/10.1007/s12374-011-0276-z>
33. Zhao SS, Wang CH, Ma J, Wang S, Tian P, Wang JL, et al. Map-based cloning and functional analysis of the chromogen gene *C* in rice (*Oryza sativa* L.). J. Plant Biol. 2016; 59: 496-505. <https://doi.org/10.1007/s12374-016-0227-9>.
34. Ali ML, Sanchez PL, Yu S, Lorieux M, Eizenga G. Chromosome segment substitution lines: A powerful tool for the introgression of valuable genes from *Oryza* wild species into cultivated rice (*O. sativa*). Rice. 2010; 3: 218-34. <https://doi.org/10.1007/s12284-010-9058-3>.
35. Qi L, Ding YB, Zheng XM, Xu R, Zhang LZ, Wang YY, et al. Fine mapping and identification of a novel locus *qGL12.2* controls grain length in wild rice (*Oryza rufipogon* Griff.). Theor Appl Genet. 2018; 131: 1497-1508. <https://doi.org/10.1007/s00122-018-3093-7>.
36. Li J, Xu R, Wang C, Qi L, Zheng XM, Wang W, et al. A heading date QTL, *qHD7.2*, from wild rice (*Oryza rufipogon*) delays flowering and shortens panicle length under long-day conditions. Sci. Rep. 2018; 8: 2928. <https://doi.org/10.1038/s41598-018-21330-z>.
37. Zhang LZ, Huang JF, Wang YY, Xu R, Yang, ZY, Zhao Z, et al. Identification and genetic analysis of *qCL1.2*, a novel allele of the “green revolution” gene *SD1* from wild rice (*Oryza rufipogon*) that enhances plant height. BMC Genet. 2020; 21: 62. <https://doi.org/10.1186/s12863-020-00868-w>
38. Jaakola, L. New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci. 2013; 18: 477-483. <https://doi.org/10.1016/j.tplants.2013.06.003>.
39. Albert N, Davies K, Lewis D, Zhang H, Montefiori M, Brendolise C, et al. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell. 2014; 26: 962-980. <https://doi.org/10.4161/psb.29526>.
40. Sakamoto W, Ohmori T, Kageyama K, Miyazaki C, Saito A, Murata M, et al. The purple leaf (*Pl*) locus of rice: the *Pl(W)* allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. Plant Cell Physiol. 2001; 42: 982-991. <https://doi.org/10.1093/pcp/pce128>. PMID: 11577193.
41. Zheng J, Wu H, Zhu HB, Huang C, Liu C, Chang YS, et al. Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. New Phytologist. 2019; 223: 705-721. <https://doi.org/10.1111/nph.15807> PMID: 30891753
42. Shin Y, Park H, Yim S, Baek N, Lee C, An G, et al. Transgenic rice lines expressing maize *C1* and *R-S* regulatory genes produce various flavonoids in the endosperm. Plant Biotechnol. J. 2006; 4: 303-315. <https://doi.org/10.1111/j.1467-7652.2006.00182.x>
43. Choi J, Platts A, Fuller D, Hsing Y, Wing R, Purugganan M. The rice paradox: multiple origins but single domestication in Asian rice. Mol. Biol. Evol. 2017; 34: 969-979. <https://doi.org/10.1093/molbev/msx049>

44. Li CB, Zhou A, Sang T. Rice domestication by reducing shattering. *Science*. 2006; 311: 1936-1939. <https://doi.org/10.1126/science.1123604>
45. Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, et al. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* 2008; 40: 1360-1364. <https://doi.org/10.1038/ng.197>
46. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 2011; 28: 2731-2739. <https://doi.org/10.1093/molbev/msr121>
47. Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor Appl Genet.* 2000; 100: 713-22. <https://doi.org/10.1007/s001220051343>
48. Livak K, Schmittgen T. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-[ΔΔCt]} method. *Methods*. 2001; 25: 402-408. <https://doi.org/10.1006/meth.2001.1262>
49. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley Ryan, Salzberg S. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 2013; 14: 36. <https://doi.org/10.1186/gb-2013-14-4-r36>
50. Chen W, Gong L, Guo Z, Wang WS, Zhang HY, Liu XQ, Yu SB. A Novel Integrated Method for Large-Scale Detection, Identification, and Quantification of Widely Targeted Metabolites: Application in the Study of Rice Metabolomics. *Mol. Plant.* 2013; 6: 1769-1780. <https://doi.org/10.1093/mp/sst080>
51. Zhu Z, Schultz AW, Wang J, Johnson CH, Yannone SM, Patti GJ, et al. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. *Nat Protoc.* 2013; 8: 451-456. <https://doi.org/10.1038/nprot.2013.004>
52. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28: 27-30. <https://doi.org/10.1093/nar/28.1.27>

Fig 1. Fine mapping of *OrCl*. (a) Genetic map of chromosome 6 of CSSL52. The red star represents the marker closely linked with the QTL for the purple coloration trait. (b,c) The location of *OrCl* was narrowed down to a 131-kb interval between markers RM19552 and RM19561. The number of recombinants obtained is indicated under the marker names, and the number of individuals (n) used in mapping are shown on the left. (d) Seventeen predicted open reading frames (ORF) were identified in the fine mapped region. ORF14 encodes a homolog of *Cl* gene. The deletions in 9311 and Nipponbare alleles were shown in (d).

Fig 2. Phenotype of CSSL52, 9311 and F₁ individual. (a-e) CSSL52 has purple apiculus (c), leaf sheath (d) and stigma (a-c) compared with 9311. (f) The F₁ of CSSL52/9311 showed the same purple coloration trait as CSSL52.

Fig 3. Characterization of *OrCl*. (a) Phylogenetic analysis of wild rice *OrCl* and other R2R3-MYB proteins. (b) Expression pattern of *OrCl* gene in different tissues at two time points. (c) Subcellular localization of *OrCl* in rice leaf protoplasts and tobacco leaf.

Fig 4. *OrCl* regulates gene expressions. (a) Top 20 significant enrichment pathway; (b) Relative expression levels of seven structure genes involved in anthocyanin synthesis determined by qRT-PCR. The values represent the mean $\pm$ s.d. *P < 0.05 and **P < 0.01 indicate significant differences in two-tailed Student's t-tests. Three biological replicates. The actin gene was used for normalization.

Fig 5. Heat map diagram of expression levels for anthocyanin biosynthesis related genes in (a) three KEGG pathways and (b) MBW complex. The heat map was drawn according to FPKM values. Columns and rows in the heat map represent samples and genes, respectively. The KEGG pathway or gene ID were displayed on the left. Color scale indicates fold changes in gene expression.

Fig 6. Overexpression of *OrCl* results in purple apiculus of Nipponbare. (a) Phenotypes of transgenic *OrCl* overexpression line (OE) and wild type Nipponbare (NIP) plants. The purple coloration was only found in the apiculus of transgenic plant. From left to right: whole plant, young spikelet, mature spikelet, mature and young seeds of NIP and OE. (b) The relative expression levels of seven structure genes determined by qRT-PCR in NIP and two OE. Three biological replicates. The values represent the mean $\pm$ s.d. *P < 0.05 and **P < 0.01 indicate significant differences in two-tailed Student's t-tests. The actin gene was used for normalization.

Fig 7. Identification of the differentially accumulated metabolites between NIL and 9311 (above), and between OE and Nipponbare (below). Red * indicates a significant increase at fold change ≥ 2 , and blue * indicates a significant decrease at fold change ≤ 0.5 .

Fig 8. Genotype analysis of coding region of *Cl*. (a) Sequence polymorphism of different haplotypes of *Cl* in 180 accessions. No. of Acc.: number of accessions, I: *indica*, J: *japonica*, R: *O. rufipogon*, N: *O. nivara*, F: functional, Non-F: non-functional. (b) Haplotype network of *Cl*.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

a

b

Figure 6

Figure 7

a

[illegible]

b

Figure 8