Article

Genetic Conservation of CBS Domain Containing Protein Family in Oryza Species and Their Association with Abiotic Stress Responses

Surabhi Tomar 1, Ashish Subba 1, Meenu Bala 2, Anil Kumar Singh 2,3, Ashwani Pareek 4,5* and Sneh Lata Singla-Pareek 1,4

Abstract: Crop Wild Relatives (CWRs) form a comprehensive gene pool that can answer the queries related to plant domestication, speciation, and ecological adaptation. The genus ‘Oryza’ comprises about 27 species, of which two are cultivated, while the remaining are wild. Here, we have attempted to understand the conservation and diversification of the genes encoding Cystathionine β-synthase (CBS) domain-containing proteins (CDCPs) in domesticated and CWRs of rice. Few members of CDCPs were previously identified to be stress-responsive and associated with multiple stress tolerance in rice. Through genome-wide analysis of eleven rice genomes, we identified a total of 36 genes encoding CDCPs in O. longistaminata, 38 in O. glaberrima, 39 each in O. rufipogon, O. glumaepatula, O. brachyantha, O. punctata, and O. sativa subsp. japonica, 40 each in O. barthii and O. meridionalis, 41 in O. nivara, and 42 in O. sativa subsp. indica. Gene duplication analysis as well as non-synonymous and synonymous substitutions in the duplicated gene pairs indicated that this family is shaped majorly by the negative or purifying selection pressure through the long-term evolution process. We identified the presence of two additional hetero-domains, namely TerCH and CoatomerE (specifically in O. sativa subsp. indica), which were not reported previously in plant CDCPs. The in silico expression analysis revealed some of the members to be highly inducive specifically in salt-tolerant genotype in response to salinity. The cis-regulatory element analysis predicted the presence of numerous stress as well as a few phytohormone-responsive elements in their promoter region. The data presented in this study would be helpful in the characterization of these CDCPs from rice, particularly in relation to abiotic stress tolerance.

Keywords: Cystathionine β-synthase; CDCPs; Oryza species; promoter analysis; stress-responsive genes; wild rice

1. Introduction

With a constant rise in the annual global rice consumption (about 504.3 million metric tons in the year 2020–2021), there is a need to increase its production to feed the rising world population, which is estimated to reach more than 9 billion by 2050 [1,2]. However, the climate change, soil degradation, and sensitivity of rice crops to various biotic and abiotic stresses pose a major challenge to meet this demand. In this scenario, to understand the mechanism of abiotic stress tolerance in rice, our lab previously carried out a...
comparative study between the transcriptome of salt-tolerant Pokkali and salt-sensitive IR64 rice genotypes, and identified many genes which showed differential regulation under salt stress treatment. Among these are the genes encoding members of a protein family containing the conserved Cystathionine β-Synthase (CBS) domain [3]. Subsequently, our lab carried out the genome-wide identification of the CBS domain containing proteins (CDCPs) in rice (Oryza sativa subsp. japonica) and Arabidopsis [4], and identified two CDCP members to impart tolerance to multiple abiotic stresses when overexpressed constitutively in tobacco [5,6].

The CBS domain, first identified by Bateman [7], consists of about 60 amino acid residues, and is conserved across all kingdoms of life. This domain generally exists as tandem repeats, mostly in pairs or quads, in the polypeptide. While some CDCPs are composed of only CBS domains, others possess additional hetero-domain(s). CBS domains are known to possess an affinity for various ligands, mainly the adenosine nucleotides, and have regulatory functions based on ligand-induced conformational changes [8,9]. Mutations in the CBS domain of different CDCPs in humans have been identified to be associated with several hereditary disorders [10,11], which implies an indispensable role of CBS domains or broadly CDCPs. However, despite such clinical significance, the studies on CDCPs have remained scarce in the plant kingdom.

Few groups working on plant CDCPs have shown the involvement of CDCP members in diverse functions, e.g., Degenerated Panicle and Partial Sterility (DPS1)/OsCBSDUF1, a CDCP member from rice, has been recently reported to have a role in ROS-dependent cuticle development in leaf and anther, and in regulation of the leaf senescence [12,13]. The overexpression of soybean CDCP genes, namely GmCBS21 and GmCBSDF3, have been reported to impart tolerance to low nitrogen and multiple abiotic stresses, respectively [14,15]. In Arabidopsis, AtCBSX1 and AtCBSX3 have been identified to interact with the thioredoxins (Trxs) in the chloroplast and mitochondria, respectively, and thereby regulate the ROS levels in the respective organelles [16,17]. The chloride channel sub-family of CDCPs in plants has been inferred to function in nitrate transport and Cl− sequestration [18].

In the field of crop breeding, it has been realized that improvement of most of the major crops, including rice, has reached their plateau due to the narrow genetic base of the cultivated species. Hence, the researchers are gradually turning their attention towards the exploration and utilization of Crop Wild Relatives (CWRs), which encompasses a wide reservoir of genetic diversity valuable for the improvement of the crop, particularly for tolerance to different abiotic as well as biotic stresses with enhanced yield potential [19–21]. The genus ‘Oryza’ comprises about 27 species having distinct ecological adaptations. These include two independently domesticated cultivated rice species, viz., O. sativa (Asian rice) and O. glaberrima (African rice), whereas the remaining species are wild. These 27 Oryza species are estimated to have evolved over 15 million years ago and have diverged into 11 genome types comprising six diploid (AA, BB, CC, EE, FF, and GG) and five polyploid (BBCC, CCDD, HHJJ, HHKK, and KKLL) genomes [1,19]. The phylogenetic relationship among the Oryza species with different genome types is depicted in Figure 1. Both the cultivated rice species are diploid (AA genome), which have evolved through a series of events, such as introgression, natural selection, and breeding [22]. Therefore, considering the importance of CDCPs, particularly in conferring abiotic stress tolerance in plants, the present study was aimed to identify the CDCP families in the available genomes of ten Oryza species, comprising both wild and cultivated rice, to understand their evolution, conservation/diversification as well as their association with abiotic stress tolerance.
Figure 1. Phylogenetic tree depicting the evolution of different Oryza species. The representative species possessing each genome type, namely AA, BB, CC, EE, FF, GG, BBCC, CCDD, HHJJ, HHKK, and KKLL, including the ones with the completely sequenced genome (analyzed in this study) have been incorporated into the tree. The figure was generated using the TimeTree database (http://www.timetree.org/, accessed on 21 October 2021) and visualized using iTol (https://itol.embl.de/upload.cgi, accessed on 21 October 2021).

2. Results and Discussion

2.1. Number of Genes Encoding CDCPs Varies in Oryza Species

The whole-genome analysis for the genes encoding the CBS domain (Pfam id: PF00571) containing proteins (CDCPs) in 10 different Oryza species using the previously annotated sequences of CDCPs (encoded by 37 genes) from O. sativa subsp. japonica as queries [4] identified a total of 36 CDCP genes in O. longistaminata, 38 in O. glaberrima, 39 each in O. rufipogon, O. glumaepatula, O. brachyantha, and O. punctata, 40 each in O. barthii and O. meridionalis, 41 CDCPs in O. nivara, and 42 in O. sativa subsp. indica. Moreover, we also re-analyzed the genome sequence of O. sativa subsp. japonica and identified two new genes encoding CDCPs, unannotated in the earlier study by Kushwaha et al. [4], thus making a total of 39 CDCP genes in this species (Table 1). We classified these newly identified CDCPs from all 10 species into different subfamilies based on the presence of CBS domains in pairs or quads, the presence or absence of other associated hetero-domain(s) in the proteins, and their sequence identity with the ones from O. sativa subsp. japonica, which has been annotated by Kushwaha et al. [4]. Notably, we have incorporated some changes for updating in the previously classified order of CDCPs by Kushwaha et al. [4], specifically in the protein members containing only one pair or two pairs of CBS domains (CBSX and CBSCBS, respectively, discussed in the following paragraphs).
In addition to different hetero-domains reported earlier by Kushwaha et al. [4] in the CDCPs from *O. sativa* subsp. *japonica*, which includes the CNNM/DUF21 (PF01595), CorC_HlyC (PF03471), Chloride Channel (CLC; PF00654), Inosine-5′-Monophosphate Dehydrogenase (IMPDH; PF00478), Sugar Isomerase (SIS; PF01380), Pentatricopeptide Repeat (PPR; PF01535), and Phox and Bem1 (PB1; PF00564) domains, three new domains, namely Coatomer epsilon subunit (CoatomerE; PF04733), terC (PF03741), and Carbohydrate binding domain (CBD; PF16561), were also found to be present in some CDCP members from *Oryza* species (Figure 2). The CoatomerE domain existed in association with a pair of CBS domains in *O. sativa* subsp. *indica* (OsICBSCoatomer E), while in *O. longistaminata*, this domain was found to be present in chloride channel members of CDCPs (OICBSCLC5).

In yeast and mammals, the CoatomerE domain is found among the seven subunits of Coat Protein Complex1 (COP1) [23], which is involved in the early retrograde transport of proteins from Golgi to Endoplasmic Reticulum as well as in intra Golgi transport [24]. The TerCH domain, involved in natural resistance to xenobiotic compounds in bacteria [25], was found only in *O. sativa* subsp. *indica*, in association with the CBS_CorC_HylC domain at the C-terminus. In *Arabidopsis*, a protein containing the lone TerCH domain has been reported to participate in the thylakoid membrane biogenesis and the *de novo* synthesis of the Photosystem II core proteins, while its knockout resulted in chlorophyll-deficient lethal seedlings [26,27].

**Figure 2.** Schematic representation (unscaled) of the domain organization of different CDCPs in rice. CBS domains have been found to occur alone as well as in association with other domains in the polypeptide. ‘CBS’ denotes a CBS domain. Other domains include CNNM (or DUF21), CorC_HlyC, voltage chloride channel (Voltage CLC), IMPDH, Sugar isomerase (SIS), Pentatricopeptide repeat (PPR), Phox and Bem1(PB1), Carbohydrate binding domain (CBD) Coatomer epsilon subunit (CoatomerE), and TerCH. Note that in IMPDH, a pair of CBS domains occurs within the IMPDH domain.
Table 1. New classification and the distribution of CDCPs in 11 genomes from 10 *Oryza* species. Note that the 'old classification' represents the previous nomenclature and classification of CDCPs from *O. sativa* subsp. *japonica* genome by Kushwaha et al. [4], which has been updated with a few changes based on updated protein domain prediction and the identification of new CDCP members. The blank space denotes the absence of the corresponding ortholog in the respective species and '-' in the old classification column represents that the CDCP member was not previously identified.

| Old Classification | New Classification | O. sativa, indica | O. brachyantha | O. glaberimana | O. nivara | O. meridionalis | O. longistaminata | O. sativa, japonica |
|-------------------|-------------------|------------------|--------------|--------------|----------|---------------|------------------|------------------|
| CBSX1             | CBSX1             | LOC_Os08g22149   | ORBART08G30130 | ORB08G19240  | ORGLA08G308800 | ORUFI08G11440   | CNIVA08G10850   | OMERI08G88600   |
| CBSX2             | CBSX2             | LOC_Os08g22170   | ORBART08G300150 | ORB08G10450  | ORGLA08G200600 | ORUFI08G20780   | OPUNC08G0550    | CNIVA08G05700   |
| CBSX3             | CBSX3             | LOC_Os02g572800  | ORBART02G27390 | ORB02G44900  | ORGLA02G202600 | ORUFI02G288900 | OPUNC02G34550   | CNIVA02G40140   |
| CBSX4             | CBSX4             | LOC_Os03g526900  | ORBART03G35390 | ORB03G408700 | ORGLA03G203100 | ORUFI03G247800  | OPUNC03G206600 | CNIVA03G07900   |
| CBSX5             | CBSX5             | LOC_Os04g950100  | ORBART04G015100 | ORB04G114100 | ORGLA04G201270 | ORUFI04G201970  | OPUNC04G015600 | CNIVA04G201220   |
| CBSX6             | CBSX6             | LOC_Os01g448600  | ORBART01G228500 | ORB01G303220 | ORGLA01G307400 | ORUFI01G268600  | OPUNC01G228900 | CNIVA01G286810   |
| CBSX7             | CBSX7             | LOC_Os01g699000  | ORBART01G242100 | ORB01G508000 | ORGLA01G235900 | ORUFI01G453700  | OPUNC01G409000 | CNIVA01G47160   |
| CBSX8             | CBSX8             | LOC_Os02g141700  | ORBART02G214100 | ORB02G281500 | ORGLA02G2031800 | ORUFI02G288000  | OPUNC02G195500  | CNIVA02G243900   |
| CBSX9             | CBSX9             | LOC_Os03g653600  | ORBART03G446600 | ORB03G128140 | ORGLA03G123150 | ORUFI03G156500  | OPUNC03G123150 | CNIVA03G166500   |
| CBSX10            | CBSX10            | LOC_Os04g228400  | ORBART04G254400 | ORB04G328900 | ORGLA04G2011600 | ORUFI04G246400  | OPUNC04G235700 | CNIVA04G203510   |
| CBSX11            | CBSX11            | LOC_Os04g313400  | ORBART04G101600 | ORB04G174400 | ORGLA04G2072200 | ORUFI04G117440  | OPUNC04G208200  | CNIVA04G207900   |
| CBSX12            | CBSX12            | LOC_Os05g083100  | ORBART05G209200 | ORB05G367900 | ORGLA05G2061200 | ORUFI05G201480  | OPUNC05G273500  | CNIVA05G283000   |
| CBSX13            | CBSX13            | LOC_Os01g305600  | ORBART01G237300 | ORB01G331600 | ORGLA01G1925900 | ORUFI01G286900  | OPUNC01G238500  | CNIVA01G267100   |
| CBSX14            | CBSX14            | LOC_Os01g104800  | ORBART01G209600 | ORB01G290500 | ORGLA01G1925900 | ORUFI01G236700  | OPUNC01G231010  | CNIVA01G225600   |
| CBSX15            | CBSX15            | LOC_Os08g104200  | ORBART08G209600 | ORB08G175800 | ORGLA08G2361000 | ORUFI08G143600  | OPUNC08G123150  | CNIVA08G114400   |
| CBSX7/CBSX8       | CBSX7/CBSX8       | LOC_Os01g402000  | ORBART01G209600 | ORB01G209600 | ORGLA01G1925900 | ORUFI01G236700  | OPUNC01G231010  | CNIVA01G225600   |
| CBSX9/CBSX10      | CBSX9/CBSX10      | LOC_Os01g924000  | ORBART01G242100 | ORB01G509800 | ORGLA01G2341100 | ORUFI01G454600  | OPUNC01G410000  | CNIVA01G472500   |
| CBSX11/CBSX12     | CBSX11/CBSX12     | LOC_Os01g312400  | ORBART01G101600 | ORB01G471400 | ORGLA01G0237200 | ORUFI01G411990  | OPUNC01G208200  | CNIVA01G207900   |
| CBSX13/CBSX14     | CBSX13/CBSX14     | LOC_Os01g942500  | ORBART01G237300 | ORB01G331600 | ORGLA01G1925900 | ORUFI01G286900  | OPUNC01G238500  | CNIVA01G267100   |
| CBSX15/CBSX16     | CBSX15/CBSX16     | LOC_Os08g120700  | ORBART08G209600 | ORB08G175800 | ORGLA08G2361000 | ORUFI08G143600  | OPUNC08G123150  | CNIVA08G114400   |
| CBSX17/CBSX18     | CBSX17/CBSX18     | LOC_Os01g655000  | ORBART01G390500 | ORB01G478400 | ORGLA01G9239900 | ORUFI01G242100  | OPUNC01G237800  | CNIVA01G439000   |
| CBSX19/CBSX20     | CBSX19/CBSX20     | LOC_Os01g926000  | ORBART01G279900 | ORB01G371400 | ORGLA01G2031200 | ORUFI01G310500  | OPUNC01G279800  | CNIVA01G239500   |
| CBSX21/CBSX22     | CBSX21/CBSX22     | LOC_Os02g351900  | ORBART02G206700 | ORB02G282400 | ORGLA02G2017200 | ORUFI02G217000  | OPUNC02G218500  | CNIVA02G226500   |
| CBSX23/CBSX24     | CBSX23/CBSX24     | LOC_Os03g489400  | ORBART03G305700 | ORB03G376500 | ORGLA03G2081000 | ORUFI03G318100  | OPUNC03G278700  | CNIVA03G319200   |

Note: '-' in the old classification column represents that the CDCP member was not previously identified.
| Old Classification | New | O. sativa, japonica | O. barthii | O. brachyantha | O. glaberrima | O. rufipogon | O. punctata | O. nivara | O. meridionalis | O. longistaminata | O. sativa, indica | O. glumaepatula |
|--------------------|-----|--------------------|------------|---------------|---------------|---------------|--------------|---------------|---------------|----------------|-----------------|-----------------|
| CBJSCLC5           | CBJSCLC5 | LOC_Os04g52510 | OBART04G22750 | OB04G34170 | ORGLA04G20235000 | ORUFI04G28900 | OPUNC04G24730 | ONIVA04G25550 | OMERI04G22690 | KN538912.1_FGP009 | IG0SQA017236 | OGLUM04G27200 |
| CBJSCLC6           | CBJSCLC6 | LOC_Os08g20570 | OBART08G09810 | OB08G38730 | ORGLA08G20384800 | ORUFI08G11100 | OPUNC08G299140 | ONIVA08G10550 | OMERI08G11970 | KN538923.1_FGP002 | IG0SQA028422 | OGLUM08G10690 |
| CBJSCLC7           | CBJSCLC7 | LOC_Os12g25200 | OBART12G10660 | OB12G19240 | ORGLA12G20995000 | ORUFI12G11740 | OPUNC12G29550 | ONIVA12G11240 | OMERI12G207260 | KN540194.1_FGP001 | IG0SQA036265 | OGLUM12G11730 |
| CBJSCLC8           | CBJSCLC8 | LOC_Os08g38960 | OBART08G19330 | OB08G26380 | ORGLA08G17070000 | ORUFI08G21610 | OPUNC08G17380 | ONIVA08G21390 | OMERI08G15970 | KN539998.1_FGP007 | IG0SQA028930 | OGLUM08G20420 |
| CBJSCLC9           | CBJSCLC9 | LOC_Os02g48880 | OBART02G30500 | OB02G37640 | ORGLA02G26400000 | ORUFI02G232200 | OPUNC02G28140 | ONIVA02G33300 | OMERI02G29590 | KN539828.1_FGP002 | IG0SQA038723 | OGLUM02G31200 |
| CBJSCLC10          | CBJSCLC10 | LOC_Os04g36560 | OBART04G13660 | OB04G20940 | ORGLA04G10740000 | ORUFI04G14940 | OPUNC04G11570 | ONIVA04G11880 | OMERI04G12600 | KN538758.1_FGP005 | IG0SQA015026 | OGLUM04g13430 |
|                   |         |                   | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 | OB04G209800 |
Previously, Kushwaha et al. [4] used Pfam (version 21.0) for their analysis as well as the systematic classification of the CDCPs. However, with the rapid advancements and up-gradation of different tools, the classification needed to be modified and updated as per the latest predictions. We, therefore, made some changes in the pre-existing classification and the same has been used for all the subsequent analyses. The re-classification for different CDCPs from *Oryza* species has been systematically arranged and presented in Table 1. The previously reported CBSX7 and CBSCBS1 are the products of a single gene (LOC_Os01g40420) in *O. sativa* subsp. *japonica*, of which the longest isoform (CBSCBS1) harbors two pairs of CBS domains. As a result, we have reclassified it and its homologs in other *Oryza* species as CBSCBS1. Additionally, the previously reported CBSCBS5 has now been predicted to possess a single pair of CBS domains, and therefore, we have reclassified it as the new CBSX7. Additionally, we observed the carbohydrate binding domain in two members of CDCPs, which were earlier classified as CBSX8 and CBSCBS4 by Kushwaha et al. [4]. These two proteins have now been renamed as CBSCBSCBD1 and CBSCBSCBD2, respectively, to distinguish them from CBSCBS members. CBSCBSCBD2 was observed to be absent in *O. brachyantha*. The orthologs of CBSCBSCBD in *Arabidopsis* have been reported to function as hybrid βγ-subunits of SnRK complexes [28] which regulate various cellular processes, including plant growth and stress responses [29]. Consequently, the newly identified CDCP genes in *O. sativa* subsp. *japonica* with the locus id LOC_Os08g41740 and its respective homologs in other *Oryza* species have now been classified as CBSX8. The previously named CBSX10 protein was found to possess two pairs of CBS domains; hence, it has been renamed as the new CBSCBS4. Another newly identified CDCP (gene id: LOC_Os10g35630) and its homologs in other *Oryza* species possess a pair of CBS domains and have now been classified as CBSX10.

We noticed the presence of CBS domains either in pairs or quads in all the CDCPs from different *Oryza* species, except in the genome of *O. meridionalis* and *O. nivara*, where we found two genes (gene id: OMERI02G33320 and ONIVA05G14030, respectively) that encodes for a protein containing only one CBS domain, and a gene in *O. meridionalis* (OMERI05G12070) to encode a protein containing three CBS domains. We anticipated the possible loss of one CBS domain from these proteins during the course of evolution and named the former two as OmCBSX13 and OnCBSX15, respectively, while the latter containing three CBS domains is classified as OmCBSCBS5.

The differences in the total number of genes encoding CDCPs across 11 genomes from 10 *Oryza* species were observed to be mainly due to the gain or loss of CBSX and/or CBSCBS members during evolution (Table 1). However, CBSCLCs showed high conservation in all these species, except in *O. meridionalis*, where two CBSCLCs were found to be absent, while *O. barthii* possessed an additional CBSLC. The absence of a member of either CBSDFUF or CBSCBSFB sub-family was also observed in five *Oryza* species, namely *O. sativa* subsp. *japonica*, *O. barthii*, *O. glaberrima*, *O. rufipogon*, and *O. meridionalis*. Interestingly, CBSIMPDH, CBSSIS, CBSSPR, and CBSDFUFCH were found to exist as a lone gene, but conserved in the genome of all the *Oryza* species studied, except *O. glaberrima*, which contains two members of CBSDFUFCH, and *O. longistaminata*, in which CBSSIS1 was annotated to encode a truncated protein without CBS domain, so it was not included in the present study.

### 2.2. Phylogenetic Analysis of CDCPs in Cultivated and Wild Rice

To determine the evolutionary relationship among CDCP members in the cultivated and wild species of rice, a phylogenetic analysis was performed based on protein sequence alignment. The phylogenetic tree distributed all the CDCPs from 11 *Oryza* genomes into 14 major clades (referred herein as C-1 to C-14). The orthologous CDCPs from different species clustered together in the same clade, except a few CDCPs, namely ObaCBSIMPDH1, OglCBSIMPDH4, OICBSIMPDH, OICBSX2, OmCBSX13, OnCBSX15, and OgbCBSBS7, which were found to cluster distantly from the rest of their respective members (Figure 3).
Figure 3. Phylogenetic tree depicting the relationship among various CDCPs identified in the 11 genomes from 10 *Oryza* sp. The tree comprises of fourteen major clades (C-1 to C-14), and each clade is depicted in a distinct color. Green circles represent the bootstrap values (1000 replicates).

The proteins containing only a single pair of CBS domains (CBSX1 to CBSX12) clustered into four distinct clades, implying functional diversification among these members. The orthologs of CBSX1 and CBSX2 in different rice species formed sister groups (in C-5). Similarly, the orthologs of CBSX3 and CBSX5, and CBSX4 and CBSX6 clustered in C-8 and C-7, respectively, indicating that the proteins in each cluster have descended from the common ancestor. The orthologs of the newly classified CBSX7 to CBSX12 clustered together in the clade C-13, suggesting that the updated classification provided in this analysis is consistent with the evolution of these CDCPs. Notably, the protein sequence length of the members of CBSX7 to CBSX12 is exceptionally longer than that of the remaining CBSX members. The CBSX14, identified only in *O. meridionalis*, was found to have 100% sequence identity with OmCBSX3 and clustered with the CBSX3 proteins. Conversely, the CBSX13 and CBSX15, present only in *O. meridionalis* and *O. nivara*, respectively, were observed to be clustered distantly from all other CBSX members. The orthologs of CBSCBS1 to CBSCBS4 containing only two pairs of CBS domains, including CBSCBS5 and CBSCBS6, that were
identified only in *O. meridionalis* and *O. glaberrima*, respectively, clustered together in C-11, signifying the common ancestor of these members. Moreover, the sequence alignment showed 100% identity between OgbCBSCBS6 and OgbCBSCBS3. However, the CBSCBS7 present only in *O. glaberrima* clustered with CBSCBSPB3 members, which implies that OgbCBSCBS7 might be a CBSCBSPB member that has lost its PB1 domain during the course of evolution. The orthologs of CBSCBSCBD1 and CBSCBSCBD2 possessing an additional CBD, which was previously classified as CBSX8 and CBSCBS4, respectively, clustered together in C-6.

Our previous phylogenetic study on CBSCLCs from various plant species clustered these CDCP members into two distinct groups: one consisting of a majority of these members, while others consisting of few members with higher identity to prokaryotic CLCs [18]. Likewise, in the present study, we observed the orthologs of CBSCLC1, CBSCLC3, CBSCLC4, CBSCLC5, CBSCLC6, CBSCLC7, and CBSCLC10 to jointly form a major CBSCLC clade (C-14) in *Oryza* species. On the other hand, the orthologs of CBSCLC2, CBSCLC8, and CBSCLC9 together formed a minor CBSCLC clade (C-10), and these proteins appeared to be more identical to prokaryotic CLCs (data not shown). This suggests that the two groups of CLCs have arisen from different ancestors. The C-14 cluster of CBSCLCs also included CBSIMPDH1 from *O. barthii* and *O. glumaepatula*, implying their distant relationship. However, the CBSIMPDHs from the rest of the species clustered independently into C-1. The orthologs of CBSCBSPB and CBSDUF family members clustered independently into C-12 and C-9, respectively. Likewise, the orthologs of CBSSIS1, CBSPPR, and CBSDUFCH1 formed their distinct clades, C-2, C-3, and C-4, respectively.

Among the newly identified CDCPs in this study that possess additional hetero-domain, CBSCoatomerE present specifically in *O. sativa* subsp. *indica* is clustered with the orthologs of CBSCLC5 (C-14), which also include CBSCLC5 ortholog from *O. longistaminata* possessing additional CoatomerE domain. The CBSTerCH, present only in *O. sativa* subsp. *indica*, clustered distantly from the rest of the CDCPs. The sequence alignment showed 100% identity between the newly identified ObaCBSCLC11 and the previously known OgbDUFCH2 and OgbDUFCH1. Accordingly, these proteins clustered together in their respective sub-clades.

When we analyzed the branching pattern of CDCPs from 11 different genomes of *Oryza* species in each sub-clade and clade, we observed no consistent evolutionary relationship pattern among these *Oryza* species. Such inconsistency in the phylogenetic relationship among *Oryza* species with AA genomes has also been noted previously [1]. Nevertheless, we observed *O. brachyantha* (FF) CDCPs as early divergent or distant members in the majority of the clades, followed by *O. punctata* (BB) (Figure 3). Among the species with the AA genome, *O. longistaminata*, which has been regarded as the most ancestral species [30–32], appeared to be evolutionarily distant from the rest in most of the sub-clades or clades. Similarly, *O. brachyantha* has also been perceived to be distant from the rest of the species [33]. *O. longistaminata* is known to possess unique morphological features, such as self-incompatibility, rhizomatous, and the presence of distinct ligule, making it different from the rest of the AA genome species of *Oryza*. Moreover, *O. longistaminata* shows higher heterozygosity and a greater percentage of the presence of transposable elements (TEs) [34], indicating the accumulation of greater genomic variations from the rest of the species. TEs are known to comprise a major portion of plant genetic material and these potential endogenous mutation-causing agents have a significant role in the evolution of their respective host species [35]. TEs can bring about changes ranging from loss of function of any gene to complete reprogramming of the regulatory circuitry. Additionally, it has also been reported that TEs tend to be selectively removed from gene regions in the case of cultivated rice, and if present, they are more likely to occur in the intronic regions suggesting that cultivated rice species possess similar TE arrangements in their respective genomes [36].

In the case of the cultivated species, *O. sativa* and *O. glaberrima*, the orthologous genes associated with their domestication were reported to have undergone convergent, yet
independent selection, and they share a high syntenic relationship [37]. In the present study, we also observed a closer relation between *O. sativa* subsp. *japonica* and *O. glaberrima*. The diversification in the *Oryza* species has occurred within a narrow time scale of about 15 million years. As such, analyzing the phylogenetic relationship for any multigene family proteins in domesticated and wild species would facilitate a better understanding of the evolution of the *Oryza* species.

2.3. Gene Structural Organization and Protein Motif Analysis of Different CDCPs

Following the phylogenetic analysis of the CDCPs in different *Oryza* species, we analyzed their gene structure as well the conservation of different protein motifs, to gain insight into their molecular diversity (Figures 4–8). Generally, it has been found that a loss or gain of intron leads to structural complexity which functions as an important evolutionary force in the case of large protein families [38]. In the case of plants, genes with higher expression levels tend to have longer introns and untranslated regions in comparison to the low expressing ones [39]. In our analysis, we observed the orthologs of each CDCP from different *Oryza* species to exhibit high conservation in both gene structure and protein motifs. Corresponding to the phylogenetic study, the variations in gene structure and protein motif conservation were majorly observed in the CDCP members from *O. longistaminata*, followed by the ones from *O. brachyantha*, suggesting these two species as the distant ancestors, as reported previously [30,33].

In the case of CDCPs with only CBS domains, CBSX2 and CBSX4 from *O. longistaminata* were found to have longer sequences with additional motifs, while a large fraction of protein was found to be missing in CBSX5 from *O. longistaminata* and *O. barthii* (Figure 4). The CBSCBS3 from *O. longistaminata* also exhibited longer protein size with the presence of additional motifs, while a large region was missing in its other CDCPs, namely, OICBSCBS2 and OICBSX11. The CBSX8 orthologs from *O. brachyantha*, *O. punctata*, and *O. barthii* exhibited differences in both the gene structure and motifs arrangement. The CBSX10 and CBSX12 orthologs from *O. brachyantha* were observed to have lost a few motifs, while CBSX9 from both *O. brachyantha* and *O. barthii* showed loss of two motifs. Although OmCBSCBS5 was clustered with the members of CBSCBS1, it showed differences in both gene structure and motifs from CBSCBS1 orthologs (Figure 5). In CBSCBSCBD1 members, additional motifs were observed in orthologs from *O. longistaminata* and *O. barthii* (Figure 5).

In the case of CBSCLC sub-family, as CBSCLC2, CBSCLC8, and CBSCLC9 were identified to form distinct clades from the rest of the members in the phylogenetic tree, these proteins also exhibited different motif patterns from the remaining CLCs. A large region was observed to be missing in CBSCLC2 from *O. sativa* subsp. *japonica* and in CBSCLC5 from *O. brachyantha*, while CBSCLC5 from *O. glumaepatula* showed relatively longer gene and protein sequences (Figure 6).

Among CBSCBSBP family members, CBSCBSBP5 orthologs from *O. longistaminata* and *O. brachyantha* were found to be missing in CBSX5 from *O. longistaminata* and *O. barthii* (Figure 4). The CBSCCBS3 orthologs from *O. longistaminata* also exhibited longer protein size with the presence of additional motifs, while a large region was missing in its other CDCPs, namely, OICBSCBS2 and OICBSX11. The CBSX8 orthologs from *O. brachyantha*, *O. punctata*, and *O. barthii* exhibited differences in both the gene structure and motifs arrangement. The CBSX10 and CBSX12 orthologs from *O. brachyantha* were observed to have lost a few motifs, while CBSX9 from both *O. brachyantha* and *O. barthii* showed loss of two motifs. Although OmCBSCBS5 was clustered with the members of CBSCBS1, it showed differences in both gene structure and motifs from CBSCBS1 orthologs (Figure 5). In CBSCBSCBD1 members, additional motifs were observed in orthologs from *O. longistaminata* and *O. barthii* (Figure 5).

In the case of CBSCLC sub-family, as CBSCLC2, CBSCLC8, and CBSCLC9 were identified to form distinct clades from the rest of the members in the phylogenetic tree, these proteins also exhibited different motif patterns from the remaining CLCs. A large region was observed to be missing in CBSCLC2 from *O. sativa* subsp. *japonica* and in CBSCLC5 from *O. brachyantha*, while CBSCLC5 from *O. glumaepatula* showed relatively longer gene and protein sequences (Figure 6).

Among CBSCBSBP family members, CBSCBSBP5 orthologs from *O. longistaminata* and *O. brachyantha* were found to be missing in CBSX5 from *O. longistaminata* and *O. barthii* (Figure 4). The CBSCCBS3 orthologs from *O. longistaminata* also exhibited longer protein size with the presence of additional motifs, while a large region was missing in its other CDCPs, namely, OICBSCBS2 and OICBSX11. The CBSX8 orthologs from *O. brachyantha*, *O. punctata*, and *O. barthii* exhibited differences in both the gene structure and motifs arrangement. The CBSX10 and CBSX12 orthologs from *O. brachyantha* were observed to have lost a few motifs, while CBSX9 from both *O. brachyantha* and *O. barthii* showed loss of two motifs. Although OmCBSCBS5 was clustered with the members of CBSCBS1, it showed differences in both gene structure and motifs from CBSCBS1 orthologs (Figure 5). In CBSCBSCBD1 members, additional motifs were observed in orthologs from *O. longistaminata* and *O. barthii* (Figure 5).

In the case of CBSCLC sub-family, as CBSCLC2, CBSCLC8, and CBSCLC9 were identified to form distinct clades from the rest of the members in the phylogenetic tree, these proteins also exhibited different motif patterns from the remaining CLCs. A large region was observed to be missing in CBSCLC2 from *O. sativa* subsp. *japonica* and in CBSCLC5 from *O. brachyantha*, while CBSCLC5 from *O. glumaepatula* showed relatively longer gene and protein sequences (Figure 6).

Among CBSCBSBP family members, CBSCBSBP5 orthologs from *O. longistaminata* and *O. brachyantha* were found to be missing in CBSX5 from *O. longistaminata* and *O. barthii* (Figure 4). The CBSCCBS3 orthologs from *O. longistaminata* also exhibited longer protein size with the presence of additional motifs, while a large region was missing in its other CDCPs, namely, OICBSCBS2 and OICBSX11. The CBSX8 orthologs from *O. brachyantha*, *O. punctata*, and *O. barthii* exhibited differences in both the gene structure and motifs arrangement. The CBSX10 and CBSX12 orthologs from *O. brachyantha* were observed to have lost a few motifs, while CBSX9 from both *O. brachyantha* and *O. barthii* showed loss of two motifs. Although OmCBSCBS5 was clustered with the members of CBSCBS1, it showed differences in both gene structure and motifs from CBSCBS1 orthologs (Figure 5). In CBSCBSCBD1 members, additional motifs were observed in orthologs from *O. longistaminata* and *O. barthii* (Figure 5).

In the case of CBSCLC sub-family, as CBSCLC2, CBSCLC8, and CBSCLC9 were identified to form distinct clades from the rest of the members in the phylogenetic tree, these proteins also exhibited different motif patterns from the remaining CLCs. A large region was observed to be missing in CBSCLC2 from *O. sativa* subsp. *japonica* and in CBSCLC5 from *O. brachyantha*, while CBSCLC5 from *O. glumaepatula* showed relatively longer gene and protein sequences (Figure 6).
**Figure 4.** Schematic representation of the conservation of gene structure (left panel) and protein motifs (right panel) in the members of CDCPs containing only a single pair of CBS domains (CBSX1–CBSX6; excluding CBSX7–CBSX12, which are presented in Figure 5a) from different *Oryza* species. The length of UTR, exon, and intron has been depicted in proportion to the actual sizes, which is also indicated using a scale at the bottom. The order of different CDCPs is kept as per their phylogenetic relationship. The conserved motifs are predicted using the MEME suite.
Figure 5. Schematic representation of the conservation of gene structure (left panel) and protein motifs (right panel) in CDCPs containing (a) only a single CBS domain containing proteins (CBSX7–CBSX12) and (b) only two pairs of CBS domains, from different *Oryza* species. The length of UTR, exon, and intron has been depicted in proportion to the actual sizes which is also indicated using a scale at the bottom. The order of different CDCPs is kept as per their phylogenetic relationship. The conserved motifs are predicted through the MEME suite.
proteins also exhibited different motif patterns from the remaining CLCs. A large region was observed to be missing in CBSCLC2 from *O. sativa* subsp. *japonica* and in CBSCLC5 from *O. brachyantha*, while CBSCLC5 from *O. glumaepatula* showed relatively longer gene and protein sequences (Figure 6).

Figure 6. Schematic representation of the conservation of gene structure (left panel) and protein motifs (right panel) in CBSCLC members of the CDCP family from different *Oryza* species. Much similitude can be observed in terms of the genetic organization of different orthologs harboring the chloride channel domain. The length of UTR, exon, and intron has been depicted in proportion to the sequence lengths as indicated by the scale at the bottom. The order of the different CDCPs is kept as per their phylogenetic relationship. The conserved motifs are predicted through the MEME suite.
Figure 7. Schematic representation of the conservation of gene structure and protein motifs in CBSCBSPB and CBSDUF members of the CDCP family from different Oryza species. The length of UTR, exon, and intron are represented in correspondence to their respective sequence sizes. The different CDCPs are clustered according to their phylogenetic relationship. The conserved motifs are predicted using the MEME suite.
Figure 8. Schematic representation of the conservation of gene structure and protein motifs in (a) CBSIMPDH, (b) CBSDUFCH, (c) CBSSIS (d) CBSPPR, and (e) CBSCBSCBD members of CDCP family from different Oryza species. The respective length of UTR, exon, and intron for each ortholog are well in proportion to actual sequence lengths. The order of the different CDCPs is kept by their phylogenetic relationship. The conserved motifs are predicted through the MEME suite.

Although the orthologs often possess high sequence similarity at the protein level, there may have differences acquired in the length of 5′ and 3′ UTR at the gene level, which are known to subsequently provide functional specificity [40]. Additionally, in the case of plants, the developmental transitions, as well as response towards environmental adversities, are modulated through transcriptional reprogramming, which is substantially coordinated by the UTRs [40]. During evolution, with an increase in genome sizes, the UTRs have been found to lengthen, particularly the 3′ UTR [41]. Moreover, some 5′ UTRs and 3′ UTRs harbor certain regulatory sequences that may act on the stability and localization of mRNA, as well as its translational efficiency [42–44]. The efficient translation facilitating 5′ UTRs have been observed to be short, with less GC content and secondary structures. On the
contrary, longer and highly structured 5′ UTRs are more often than not associated with the genes regulating highly specific developmental processes, particularly in a tissue-specific manner [45]. In congruency with the above statements, we observed marked differences in the length of UTRs between the orthologs of CDCPs belonging to different subfamilies. Unprecedently long 5′ UTRs were observed in the case of all the orthologs of CBSX3 and CBSX5. Longer 5′ UTRs were also observed in OnCBSX1, OmCBSX4, OrCBSCLC7, OnCBSCLC7, OsCBSCLC2, OsCBSCLC3, OmCBSCLC5, ObaCBSCLC8, and OmCBSDF3 among their respective orthologs from other Oryza species, supporting a length-dependent functional precision as reported by Srivastava et al. [40]. Interestingly, only in O. sativa subsp. japonica, we noticed that some CDCP members were devoid of either 5′UTR or 3′ UTR, or both, such as in OsCBSCLC10, ObaCBSX7, and OsCBSX12, suggesting a complex and alternative post-translational regulatory mechanism for such genes [44].

2.4. Gene Duplication and Synteny Analysis in Various Oryza Species

Gene duplication events and their subsequent retention contribute to the evolution of novel functions and stress adaptation in plants [46–49]. Using the MCScanX algorithm, which considers both homology and genomic distribution to evaluate the collinearity and synteny, we detected duplications in the CDCP genes in all 10 rice genomes, which ranged from 1–6 in number (Figure 9; Table 2). Since the annotation for O. longistaminata is available only up to scaffold level, it was excluded from the gene duplication analysis. Our analysis suggested that the whole genome or segmental duplication events have led to the expansion of the CDCP gene family in all the Oryza species.

To understand the selection pressure on the duplicated CDCP genes, the non-synonymous (Ka) and synonymous substitutions (Ks), and also the Ka/Ks ratios were calculated for the duplicated gene pairs. The value of Ka/Ks = 1 suggests that the genes have undergone a neutral selection, while <1 and >1 values suggest negative and positive selection, respectively [50]. We observed Ka/Ks values < 1 for all the genes encoding CDCPs, indicating that these genes in all the Oryza species have experienced negative or purifying selection pressure during evolution (Table 2). Prevalence of negative selection indicates optimization of genetic structures through long-term evolutionary processes such that any mutational change in genes leads to a reduction in biological fitness. Negative selection, thus, maintains the fixation of genetic characters in a population by removing deleterious mutations [51]. Additionally, in the paralogous gene pairs encoding the CDCPs, the purifying selection pressure appears to act more strongly on the non-synonymous mutations than the silent mutations. Thus, preserving the functional properties of CDCPs in all the 10 Oryza genomes evaluated even after duplication, as a further adaptive advantage by a mutation on a non-synonymous site might be an unlikely event. Several other gene families in Oryza sp., such as WRKY [52], ALOG domain [53], DUF-221 domain containing gene family [54], F-box, and NB-ARC gene families [55], have also been found to be shaped majorly by negative selection. The gene duplication has contributed to the expansion of the gene family through segmental or whole-genome duplication in rice species [56]. In a previous study on rice, duplicated blocks resulting from a whole-genome duplication event were found to cover about 60% of the genome [57]. The large-scale duplication of the rice genome was also reported by Wang et al. [58] in Oryza sativa subsp. indica.
Figure 9. Gene duplication analysis of CDCP genes in different Oryza species. The duplication analysis was carried out in 10 genomes of 9 Oryza species, namely (a) O. meridionalis (b) O. sativa subsp. japonica (c) O. glaberrima (d) O. nivara (e) O. glumaepatula (f) O. punctata (g) O. barthii (h) O. rufipogon (i) O. brachyantha, and (j) O. sativa subsp. indica. The chromosome number is shown in the middle of each chromosome and the genomic location for each gene has been marked on the respective chromosomes. The grey lines in the background represent the collinear blocks within each genome. The red-colored lines highlight the duplication events present. The gene duplication analysis was carried out using the MCScanX toolkit.
Table 2. The ratio of the number of non-synonymous substitutions per non-synonymous site (Ka) and the number of synonymous substitutions per synonymous site (Ks) in the same time period (Ka/Ks ratio) in duplicated gene pairs encoding CBS domain containing proteins in *Oryza* sp.

| Genome          | Paralogous Gene Pairs               | Ka       | Ks       | Ka/Ks    | Type of Selection | Type of Duplication |
|-----------------|-------------------------------------|----------|----------|----------|-------------------|----------------------|
| *O. barthii*    |                                     |          |          |          |                   |                      |
| OBART01G42110/OBART04G10160 | ObCBS CBS2/ObCBS CBS3 | 0.318665 | 1.884767 | 0.169074 | Negative          | Segmental            |
| OBART01G42010/OBART08G21410 | ObCBSX7/ObCBSX8 | 0.418751 | 1.481156 | 0.282719 | Negative          | Segmental            |
| OBART11G04430/OBART12G04230 | CBS CBSPB2/CBS CBSPB4 | 0.13844  | 0.685464 | 0.201966 | Negative          | Segmental            |
| OBART02G20670/OBART04G13680 | CBS CLC3/CBS CLC10 | 0.073549 | 0.819377 | 0.089762 | Negative          | Segmental            |
| OBART08G09810/OBART08G09800 | Oba CBS CLC6/Oba CBS CLC11 | 0        | 0        |          |                   | Tandem               |
| *O. brachyantha*|                                     |          |          |          |                   |                      |
| OB01G50980/OB04G17440 | ObrCBS CBS2/ObrCBS CBS3 | 0.280887 | 2.127978 | 0.131997 | Negative          | Segmental            |
| OB01G50800/OB08G28150 | ObrCBSX7/ObrCBSX8 | 0.412242 | 1.49645  | 0.27548  | Negative          | Segmental            |
| OB10G21840/OB02G32800 | ObrCBSX10/ObrCBSX11 | 0.420248 | 0.90595  | 0.463875 | Negative          | Segmental            |
| OB11G13640/OB12G14180 | ObrCBS CBSPB2/ObrCBS CBSPB4 | 0.12394  | 0.770187 | 0.160921 | Negative          | Segmental            |
| OB02G28240/OB04G20940 | ObrCBS CLC3/ObrCBS CLC10 | 0.076766 | 0.80562  | 0.095288 | Negative          | Segmental            |
| OB02G37640/OB08G26380 | ObrCBS CLC9/ObrCBS CLC9 | 0.168039 | 1.389784 | 0.12091  | Negative          | Segmental            |
| *O. glaberrima* |                                     |          |          |          |                   |                      |
| ORGLA01G0361100/ ORGLA04G072300 | OgCBS CBS2/OgCBS CBS3 | 0.312057 | 2.522365 | 0.123716 | Negative          | Segmental            |
| ORGLA11G0041900/ ORGLA12G0039900 | OgCBS CBSPB2/OgCBS CBSPB4 | 0.148377 | 0.637685 | 0.232681 | Negative          | Segmental            |
| ORGLA02G0177300/ ORGLA04G0107400 | OgCBS CLC3/OgCBS CLC10 | 0.073549 | 0.81441  | 0.09031  | Negative          | Segmental            |
| *O. glumaeapatula* |                                     |          |          |          |                   |                      |
| OGLUM01G46320/OGLUM04G09770 | OgI CBS CBS2/OgI CBS CBS3 | 0.301647 | 2.187936 | 0.137868 | Negative          | Segmental            |
| OGLUM01G46190/OGLUM08G22650 | OgI CBSX7/OgI CBSX8 | 0.512293 | 1.151735 | 0.444801 | Negative          | Segmental            |
| OGLUM11G04170/OGLUM12G05000 | Og I CBS CBSPB2/OgI CBS CBSPB4 | 0.152228 | 0.644932 | 0.236037 | Negative          | Segmental            |
| OGLUM02G20940/OGLUM04G13430 | OgI CBS CLC3/OgI CBS CLC10 | 0.085222 | 0.857666 | 0.099366 | Negative          | Segmental            |
| *O. sativa subsp. indica* |                                     |          |          |          |                   |                      |
| BGIOSGA000232/BGIOSGA015211 | OsIb CBS CBS2/OsI CBS CBS3 | 0.314794 | 2.122648 | 0.148302 | Negative          | Segmental            |
| BGIOSGA000236/BGIOSGA026597 | OsI CBSX7/OsI CBSX8 | 0.465974 | 1.220924 | 0.381657 | Negative          | Segmental            |
| Genome Paralogous Gene Pairs | Ka   | Ks   | Ka/Ks | Type of Selection | Type of Duplication |
|-------------------------------|------|------|-------|-------------------|---------------------|
| BGIOSGA0033216/BGIOSGA008689 OsICBSCBS7/OsICBSX11 | 0.371022 | 0.695562 | 0.533414 | Negative | Segmental |
| BGIOSGA0034423/BGIOSGA036554 OsICBSCBSPB2/OsICBSCBSPB4 | 0.149293 | 0.639354 | 0.233506 | Negative | Segmental |
| BGIOSGA006252/BGIOSGA015026 OsICBSCCLC3/OsICBSCLC10 | 0.06991 | 0.819761 | 0.085281 | Negative | Segmental |
| BGIOSGA008689/BGIOSGA014082 OsICBSX11/OsICBSX12 | 0.487428 | 0.914739 | 0.53286 | Negative | Segmental |
| BGIOSGA017236/BGIOSGA017237 OsICBSCBSC/OSICBSCcoatamerE | 0.177 | 0.251 | 0.705 | Negative | Tandem |

*The sequences of the pair are 100% identical.*
To study synteny relationships of the CDCP genes from the cultivated species, *O. sativa* subsp. *japonica*, with the genes from other rice genomes, orthologous genes were identified between genomes of *O. sativa japonica* and each of the other rice species using MCScan Toolkit (Supplementary Figure S1). Most of the orthologous gene pairs were collinear, manifesting conservation of synteny blocks. A total of six, seven, six, seven, five, nine, six, and seven non-collinear orthologous pairs were found between *O. sativa* subsp. *japonica* CDCP genes and that of *O. barthii*, *O. brachyantha*, *O. glaberrima*, *O. glumaepatula*, *O. sativa* subsp. *indica*, *O. meridionalis*, *O. nivara*, *O. punctata*, and *O. rufipogon*, respectively. Few of the CDCP genes have two orthologous genes—one collinear and another non-collinear—which indicates that the genes in collinear and non-collinear positions might have evolved from common ancestors. For instance, each of *OsJCBSCBSPB2* and *OsJCBSCBSPB4* genes showed orthology with both *CBSBSPB2* and *CBSBSPB4* of *O. barthii*, *O. brachyantha*, *O. glaberrima*, *O. glumaepatula*, *O. sativa* subsp. *indica*, *O. nivara*, *O. punctata*, and *O. rufipogon*.

2.5. Analysis of Cis-Elements in the Promoter Sequence of Genes Encoding CDCPs

To delve into the evolution and functional divergence of the CDCPs, the 2 kb up-stream promoter regions of all the genes encoding CDCPs (except *OsJCBSCLC10*, whose chromosomal location is not annotated in the genome) from *O. sativa* subsp. *japonica* were analyzed using the PlantCARE tool. Different cis-acting regulatory elements pertaining to both plant development and stress responses were predicted to be present in their promoter sequences (Figure 10; Supplementary Table S1). The motifs related to plant growth and development included Box 4, G-Box, SP1 and GT1 (light-responsive), zein metabolism and CAT motif (meristem specific expression), RY element (seed-specific), and GCN4 motif (endosperm specific). Additionally, the auxin-responsive AuxRR-core and TGA-element, the gibberellin-responsive GARE motif and P-box motifs, and the salicylic acid-responsive TCA-elements were also identified. The light-responsive Box 4 and G-box cis-regulatory elements were found abundantly in the promoter sequences of most of the genes encoding CDCPs. Specifically, the G-box is a hexameric DNA motif associated with the transcription induction of genes in response to light as well as senescence in the leaf [59,60]. This regulatory element was found in the promoters of all the genes encoding CDCPs, except for *OsCBSX1*, *OsCBSX5*, *OsCBSX6*, *OsCBSCLC5*, and *OsCBSDF2* genes. The Sp1 light-responsive element was predominantly found in the promoter sequences of the *OsCBSX* subfamily, indicating their photoperiod-dependent mode of functioning. Additionally, the presence of other light-responsive elements in the promoters of CDCP genes suggests circadian control as well as a putative role in photomorphogenesis.

It may be noted that the promoter of *CBSX1*, *CBSX9*, *OsCBSBSCBD1*, and *CBSDF1* contained an E2Fb transcription factor binding motif. In *Arabidopsis*, the E2Fb in complex with the Dimerization Partner (DP) has been reported to be involved in the DNA repair process during cell division [61]. The E2Fb has been reported to stimulate cell division [62] and is important during the post-mitotic state to resolve organ size [63]. Additionally, the E2Fa and E2Fb transcription factors have also been reported to be induced by a protein kinase Target of Rapamycin (TOR), which in turn is regulated through a small GTPase ROP2 protein in light-dependent as well as in auxin-dependent manner [64]. This suggests the putative role of *OsCBSCBSCBD1* in cell division and DNA repair.
Figure 10. Analysis of the cis-regulatory elements present in the 2 kb upstream promoter region of different CDCP genes from *O. sativa* subsp. *japonica.* (a) Motifs related to plant growth and development and (b) motifs related to plant defense and stress response. Each colored box and its length in the bar represent a specific motif and its frequency of occurrence in the promoter, respectively.

The promoter sequences of CDCP genes were also found to contain a higher frequency of stress-responsive cis-regulatory elements, such as ARE (anaerobic responsive element), MBS (MYB transcription factor binding site), MYB (stress-responsive), ABRE (ABA-Abscisic acid-responsive element), LTR (low temperature-responsive), TC-motif (stress responsive-
ness), CGTCA and TGACG motifs (Me-JA responsive), GC motif (Anoxia), STRE (stress-responsive), WRE3 (heat-responsive), WUN (wounding), and ERE (ethylene-responsive). The ARE motifs have been established as key motifs in the promoter regions of anaerobically induced proteins (ANPs) [65]. Except for OsCBSX2, OsCBSX4, OsCBSX8, OsCBSCBS4, OsCBSCLC4, and OsCBSDUF1, the promoters of all other CDCP genes were found to contain at least one ARE element. The ability to cope with varying degrees of oxygen limitations occurs through the induction of genes encoding for enzymes participating in carbohydrate metabolism, fermentation as well as survival pathways [66]. Although CDCPs do not have a defined role in the case of oxygen stress to date, the presence of AREs suggests that these proteins might have important functions under anaerobic growth conditions.

ABRE is another major cis-acting regulatory element well-known to have an indispensable role in acclimation to abiotic stresses [67]. It is also known to regulate seed maturation and dormancy [68]. We observed this element to be present in different genes encoding CDCPs, except OsJCBSX5, OsJCBSX11, OsJCBSCLC1, OsCBSCLC5, OsCBSCLC6, OsCBSCLC2, OsCBSCLC3, and OsCBSCLC4. Moreover, the SA-responsive TCA element was found in the promoter sequences of many CDCP genes, including OsJCBSX2–OsJCBSX4, all CBSCLCs (except OsJCBSCLC5 and OsJCBSCLC6), OsJCBSPPR1, and two members of the CBSCBSPB subfamily (OsCBSCBSPB1 and OsCBSCBSPB4).

A maize Viviparous 1 gene (VP1), a homolog of Arabidopsis ABI3, is known to carry out two important functions, including embryo maturation and germination repression [70]. The VP1 is also known to alter the ABA-responsive gene expression including the maturation associated genes [71–73] by directly binding to the RY cis-regulatory element [74]. We observed the presence of the RY element in three members of the CBSX family, namely OsCBSX5, OsCBSX9, and OsCBSX11, as well as in OsCBSDUF1. The presence of the RY element along with ABRE in the case of OsCBSX9 and OsCBSDUF1 suggests that these genes might be involved in seed maturation and/or seed dormancy as well in imparting stress tolerance.

Altogether, the promoter sequence analysis of genes encoding CDCPs from Oryza sativa subsp. japonica indicates their involvement in diverse growth and developmental processes as well as in stress responses.

2.6. Developmental and Stress-Responsive Regulation of Genes Encoding CDCPs

To understand the anatomical, developmental, and stress-responsive expression of different CDCP genes in the cultivated O. sativa subsp. japonica, the gene expression profiles were retrieved from the GENEVESTIGATOR database. For a comprehensive analysis, datasets from both Affymetrix and RNA-seq experiments were taken into consideration. However, it should be noted that the expression data for OsCBSX11, OsCBSCLC1, OsCBSCLC10, and OsCBSDUFCH1 were unavailable in the database; hence, they could not be analyzed in the present study. We found marked variations between transcripts of different genes encoding CDCPs in various tissues (Figure 11a). The OsCBSX4, OsCBSBS1, OsCBSBSB1, OsCBSDUF1, and OsCBSIMPDH1 exhibited consistently higher expression levels in all the tissues, suggesting their prominent role in diverse growth and development processes. Though not many studies have been conducted on these CDCPs, Zafar et al. [12] recently reported the role of OsCBSDUF1 (they termed it as Degenerated Panicle and Partial Sterility (DPS1)) in seed setting via regulation of ROS homeostasis. This finding strengthens our observation that these CDCPs have a crucial role in plant growth and developmental processes. In another instance, we found OsCBSX9 to have higher expression levels in the endosperm, embryo, and seedling root, indicating its role in seed
storage or germination. Likewise, OsCBSCBS4 also showed elevated expression only in the endosperm and embryo.

![Expression heatmap](image)

**Figure 11.** The in silico expression analysis of different CDCP genes in *O. sativa* subsp. *japonica* in (a) different tissues during the course of development and (b) under different stress conditions. Expression data were retrieved from the publicly available Affymetrix as well as RNA-seq datasets from the Genevestigator. The data in (a) are presented as linear expression value as used in the Genevestigator, and the data in (b) are presented as Log2 of fold change. The expression data in (b) are filtered with p-value < 0.05, while the data with p-value > 0.05 are left empty as grey boxes. The data have been depicted as a heatmap created using the MeV tool. The color scale above each heatmap shows the level of expression.

Duplication events are known to increase the expression diversity, thereby enabling specialized or distinct evolutionary patterns for the development of an organism [75]. The duplicated pair of OsCBSCBS2 and OsCBSCBS3 showed such a variation in terms of their expression levels. OsCBSCBS2 maintained relatively lower but constant expression...
levels in all the tissues, whereas OsCBSCBS3 showed higher expression in a tissue-specific manner. This is in correlation with the fact that plant genes involved in signal transduction, transcriptional regulation, and stress response favor to have paralogs [48]. More often, paralogs have a propensity to bifurcate the ancestral functions in a way that both the gene copies become vital for the organism and are, therefore, retained by the system through the dynamic process of evolution [76,77]. Alternatively, one duplicate copy of the gene might give rise to advanced expression patterns or new functions (neo-functionalization) [78,79]. The duplicated pair of OsCBSBSPB2 and OsCBSCBSB4 showed such a variation with the latter showing higher fold expression in the case of seedling root, leaf, and flag leaf tissues in comparison to its duplicate copy, suggesting neo-functionalization. Hoffman and Palmgren [80] previously attempted to establish a link between the retention of paralogous genes and their expression. They observed a correlation between the Ka/Ks ratio and the expression of the paralogous genes in Arabidopsis. We also found similar observations in two sets of paralogs from O. sativa subsp. japonica, OsCBSCBS2 and OsCBSCBS3, and OsCBSCBSB2 and OsCBSCBSB4. Although both these paralogs showed differences in tissue specificity, which was found to be more in the case of OsCBSCBS2 and OsCBSCBS3, as indicated by its lower Ka/Ks ratio in comparison to OsCBSBSPB2 and OsCBSCBSB4 gene pair, suggesting a difference in the intensity of the purifying selection acting during evolution.

Furthermore, to gain insight into the function and relevance of CDCPs in plant survival under abiotic stresses, their gene expression data under various abiotic stresses, namely, anoxia, cold, drought, salinity, and heat stresses, were also retrieved and analyzed (Figure 11b). We observed higher expression levels of OsCBSX1, OsCBSX3, OsCBSCLC3, OsCBSCLC9, OsCBSIS1, and OsCBSX10 in anoxic conditions at variable time periods ranging from 12–30 hours. Similarly, different CDCP genes such as OsCBSX2, OsCBSX9, OsCBSCLC4, OsCBSCLC5, and OsCBSDUF1 exhibited higher expression levels under cold conditions. Additionally, significantly higher expression levels were observed in the case of OsCBSX9 and OsCBSCBS4 under drought as well as salinity stress conditions. Apart from these genes, many other CDCPs such as OsCBSX4, OsCBSX5, OsCBSDUF2, and many OsCBSCLCs exhibited relatively higher expression under heat stress conditions. This observation correlated well with the role of OsCBSX4 in tolerance against multiple stress responses [5]. Additionally, overexpression of Soybean genes, namely GmCBS21 and GmCBS11, have been reported to enhance tolerance to low nitrogen and multiple abiotic stresses, respectively [14,15]. It should also be noted that substantial difference in expression levels between the gene duplicate pairs—OsCBSX2 and OsCBSX3; OsCBSBSPB2 and OsCBSCBSB4—could also be seen suggesting divergence in their functions.

The difference in expression levels in the case of different tissues could be linked with the presence of different cis-regulatory elements. In the case of cotton, it has been reported that about 40% of homologs formed from a whole-genome duplication event have transcriptional divergence primarily because of the difference in cis-regulatory elements [81]. On comparing the same, we observed a marked difference in the cis-regulatory elements present in the promoters of the duplicated pairs—OsCBSBSP2/OsCBSBSP3, OsCBSCLC3/OsCBSCLC9, and OsCBSCBSB2/OsCBSCBSB4 (Figure 10). Additionally, it has been reported that genes having similar patterns of expression possess alike motifs in their respective promoter regions [82]. For example, the drought stress responsiveness observed in the case of OsCBSX9, OsCBSX12, and OsCBSX4 could be attributed to the presence of a MYB binding site for drought inducibility (MBS) in their promoter regions. Likewise, higher fold expression levels in response to hypoxia stress were observed in the case of OsCBSX1, OsCBSX3, OsCBSCLC3, OsCBSCLC9, OsCBSIS1, and OsCBSX10 converge at the point of presence of ARE in their respective promoter regions. Taken together, it can be concluded that the cis-elements play a vital role in shaping the expression divergence of genes during the course of evolution.

We further analyzed the expression of selected CDCP genes by qRT-PCR in two contrasting rice genotypes (Oryza sativa subsp. Indica), namely, IR64 (salt-sensitive) and
Pokkali (salt-tolerant), in response to salinity treatment (200 mM NaCl) in the shoot tissues at the seedling stage. Importantly, the expression of many members, namely, OsCBSX3, OsCBSX4, OsCBSX7, OsCBSBS2, OsCBSBS3, OsCBSBSBD1, OsCBSCLCI, OsCBSCLC4-7, OsCBSPPR1, and OsCBSBSBP4, were highly induced in salt-tolerant Pokkali, while salt-sensitive IR64 rice did not manifest the induction of most of these genes (Figure 12). Such differential expression of the members of the CDCP genes implies their possible association with the salinity tolerance traits in the tolerant genotype.

Figure 12. The qRT-PCR based expression analysis of different CDCP genes in the shoot tissues of the seedlings of IR64 and Pokkali in response to salinity treatment (200 mM NaCl). The expression data are expressed as relative fold change against the untreated control samples. The experiment was repeated twice with three replicates in each case. Error bar represents the standard deviation ($n = 6$).

3. Materials and Methods
3.1. Data Retrieval and Sequence Analysis

The HMM (Hidden Markov Model) profile of the CBS domain (PF00571) was retrieved from the Pfam database, (http://pfam.xfam.org/; EMBL-EBI, UK; accessed on 24 June 2021), which was used to identify the full-length protein sequences of CDCPs in the rice genome (*O. sativa* subsp. *japonica*) from MSU Rice Genome Annotation Project Database (http://rice.uga.edu; MSU, USA; accessed on 24 June 2021). The protein blast (blastp) search was performed using each rice CDCP sequence as a query in the Gramene database (https://www.gramene.org, accessed on 26 June 2021) for the identification of homologous sequences in *O. sativa* subsp. *indica* and nine other *Oryza* species, namely *O. barthii*, *O. brachyantha*, *O. glumaepatula*, *O. glaberrima*, *O. longistaminata*, *O. meridionalis*, *O. nivara*, *O. punctata*, and *O. rufipogon*. All the retrieved sequences were examined with the Pfam 34.0 tool (http://pfam.xfam.org/, accessed on 7 July 2021) for the presence of CBS and other hetero-domains, and the proteins were named according to Kushwaha et al. [4] with some changes in the previous classification system.

3.2. Phylogenetic Analysis

Multiple sequence alignment and the construction of a phylogenetic tree was performed using MEGA7.0 [83]. The sequences were aligned following MUSCLE method (default parameters) and the evolutionary tree was constructed following a Neighbor-
joining method (1000 bootstrap replicates). The tree was visualized and annotated using iTol (https://itol.embl.de/upload.cgi, accessed on 24 July 2021).

3.3. Gene Structure and Motif Analysis

Structural visualization of the genes was performed using Gene Structure Display Server 2.0 (GSDS 2.0) (http://gsds.cbi.pku.edu.cn/; Peking University; accessed on 15 October 2021). The putative CDCP sequences were analyzed for the presence of conserved motifs using the Multiple Expectation Maximization for Motif Elicitation (MEME) program (http://meme-suite.org/, accessed on 20 July 2021) with the default parameters and the maximum number of motifs was set as 6.

3.4. Chromosomal Distribution, Duplication Analysis, and Synteny

Information of chromosome coordinates and annotations (.gff files) of the rice species were obtained from the plant Ensembl database (https://plants.ensembl.org/index.html, accessed on 7 July 2021). The full-length CDS and the protein sequences were also downloaded from the Ensembl database. To investigate gene duplication events of CDCPs within a species, the MCScanX toolkit [84] with its default parameters was employed using the output of blastp homology search within 10 rice species. To evaluate selection pressure between the paralogous genes pairs, the non-synonymous substitution rate (Ka), synonymous substitution rate (Ks), and the pairwise Ka/Ks ratios were estimated using TBtools [85]. TBtools was also used to generate Circos plots to visualize the paralogous pairs. The annotation for O. longistaminata was available only up to scaffold level; thus, it was excluded from the gene duplication analysis.

3.5. In Silico Promoter Analysis of CDCPs

The 2 kb upstream promoter sequences from the transcription start site of different CDCP genes were retrieved from Oryza sativa subsp. japonica using the RAP-DB database (https://rapdb.dna.affrc.go.jp/, accessed on 28 July 2021) and analyzed using the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, accessed on 28 July 2021) for the presence of different developmental-associated as well as stress-responsive cis-regulatory elements in these promoter sequences.

3.6. Developmental and Stress Mediated Expression Profiling of Different CDCPs in O. sativa

The expression profiles of different CDCP genes of Oryza sativa subsp. japonica under various abiotic stress conditions, such as heat, cold, salinity, anoxia, and drought, as well as in various tissue at different developmental stages, were retrieved from the publicly available Genevestigator database (https://genevestigator.com; NEBION, Switzerland; accessed on 22 July 2021). The data were presented in the form of heatmaps using MeV 4.9.0 tool [86].

For qRT-PCR based expression analysis, 10-day old seedlings of IR64 and Pokkali were treated with 200 mM of NaCl for 6 hr. Total RNA was isolated from the shoot tissues using TRIZOL reagent (ThermoFisher Scientific, Waltham, MA, USA), followed by cDNA preparation using RevertAid First Strand cDNA Synthesis Kit (ThermoFisher Scientific, Waltham, MA, USA) and qRT-PCR (Applied Biosystems 7500, Foster City, CA, USA). The eukaryotic Elongation Factor-1a (eEF-1a) was used as an internal control for normalization in the qRT-PCR [87]. The gene expression data were analyzed based on the comparative CT method [88].

4. Conclusions

The CBS domain containing proteins constitute a large superfamily with only a few members characterized functionally. In the present study, we explored the CDCP family in the genomes of six wild species of Oryza with diploid AA genome, namely, O. longistaminata, O. rufipogon, O. glumaepatula, O. meridionalis, O. nivara, and O. barthii, along with two domesticated species, namely O. sativa (subsp. Japonica and indica) and O. glaberrima. These
eight species form a pioneer gene pool and can be crossed easily for breeding purposes. Other than these, two more species, *O. brachyantha* (FF genome) and *O. punctata* (BB genome), were also included in this study to identify the evolutionary relationships among these stress-responsive CDCP genes in different *Oryza* species. Through this comparative analysis, we identified three previously unreported hetero-domains associated with CBS domains in CDCPs, namely, TerCH, CoatomerE, and CBD. Additionally, we also identified other new members in previously known CDCP subfamilies. Their phylogenetic analysis suggested CDCPs possess high sequence conservation, as also indicated by their gene structure organization. The gene expression analysis revealed their differential expression under a single as well as multiple stresses, suggesting their involvement in various stress regulatory pathways. The expression of some members was also observed to be differential in the salt-tolerant and salt-sensitive rice genotypes in response to salinity. Altogether, this study provides novel insights into the classification, evolutionary conservations, and functional divergence of the members of the CDCP family across different *Oryza* species, which in the future can help researchers in pursuing functional characterization of these proteins. The stress-responsiveness of some members of CDCP genes noted in this study encourages their further study for improving stress tolerance in domesticated *Oryza* species.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23031687/s1.

**Author Contributions:** S.L.S.-P. conceived the idea and designed the study. S.T., A.S. and M.B. performed the data analysis and wrote the manuscript. S.L.S.-P., A.K.S. and A.P. critically read and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funded by the internal grants of ICGEB.

**Institutional Review Board Statement:** Nor applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets supporting the conclusions of this article are included within the article and its additional files. The sequence data for all the *Oryza* species were obtained from Gramene data resource (https://gramene.org (accessed on 15 September 2021)). For *O. sativa* subsp. *japonica*, the sequences were also retrieved from the RGAP (http://rice.plantbiology.msu.edu/ (accessed on 15 September 2021)).

**Acknowledgments:** S.L.S.-P. acknowledges ICGEB for core grant support. A.S. acknowledges the Department of Biotechnology, and S.T. and M.B. acknowledge the University Grants Commission, Government of India, for providing research fellowship during their research work.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Stein, J.C.; Yu, Y.; Copetti, D.; Zwickl, D.J.; Zhang, L.; Zhang, C.; Chougule, K.; Gao, D.; Iwata, A.; Goicoechea, J.L.; et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover, and innovation across the genus *Oryza*. *Nat. Genet.* 2018, 50, 285–296. [CrossRef]

2. Agarwal, P.; Parida, S.K.; Raghuvanshi, S.; Kapoor, S.; Khurana, P.; Tyagi, A.K. Rice improvement through genome-based functional analysis and molecular breeding in India. *Rice* 2016, 9, 1. [CrossRef] [PubMed]

3. Kumari, S.; nee Sabharwal, V.P.; Kushwaha, H.R.; Sopory, S.K.; Pareek, A. Transcriptome map for seedling stage specific salinity stress response indicates a specific set of genes as candidate for saline tolerance in *Oryza sativa* L. *Funct. Integr. Genom. 2009*, 9, 109–123. [CrossRef] [PubMed]

4. Kushwaha, H.R.; Singh, A.K.; Sopory, S.K.; Pareek, A. Genome wide expression analysis of CBS domain containing proteins in *Arabidopsis thaliana* (L.) Heynh and *Oryza sativa* L. reveals their developmental and stress regulation. *BMC Genom.* 2009, 10, 200. [CrossRef] [PubMed]

5. Singh, A.K.; Kumar, R.; Pareek, A.; Sopory, S.K.; Singla-Pareek, S.L. Overexpression of rice CBS domain containing protein improves salinity, oxidative, and heavy metal tolerance in transgenic tobacco. *Mol. Biotechnol.* 2012, 52, 205–216. [CrossRef]

6. Kumar, R.; Subba, A.; Kaur, C.; Ariyadasa, T.U.; Sharan, A.; Pareek, A.; Singla-Pareek, S.L. OsCBS-CBSPB4 is a two cystathionine-β-synthase domain-containing protein from rice that functions in abiotic stress tolerance. *Curr. Genom.* 2018, 19, 50–59. [CrossRef]
7. Bateman, A. The structure of a domain common to archaeabacteria and the homocystinuria disease protein. *Trends Biochem. Sci.* 1997, 22, 12–13. [CrossRef]

8. Baykov, A.A.; Tuominen, H.K.; Latbi, R. The CBS domain: A protein module with an emerging prominent role in regulation. *ACS Chem. Biol.* 2011, 6, 1156–1163. [CrossRef]

9. Ereño-Orbea, J.; Oyenarte, I.; Martínez-Cruz, L.A. CBS domains: Ligand binding sites and conformational variability. *Arch. Biochem. Biophys.* 2013, 540, 70–81. [CrossRef]

10. Ignoul, S.; Eggermont, J. CBS domains: Structure, function, and pathology in human proteins. *Am. J. Physiol. Cell Physiol.* 2005, 289, C1369–C1378. [CrossRef]

11. Giménez-Mascarel, P.; González-Recio, I.; Fernández-Rodríguez, C.; Oyenarte, I.; Müller, D.; Martínez-Chantar, M.L.; Martínez-Cruz, L.A. Current structural knowledge on the CNNM family of magnesium transport mediators. *Int. J. Mol. Sci.* 2019, 20, 1135. [CrossRef]

12. Zafar, S.A.; Patil, S.B.; Uzair, M.; Fang, J.; Zhao, J.; Guo, T.; Yuan, S.; Uzair, M.; Luo, Q.; Shi, J.; et al. DEGENERATED PANICLE AND PARTIAL STERILITY 1 (DPS1) encodes a cystathionine β-synthase domain containing protein required for anther cuticle and panicle development in rice. *New Phytol.* 2020, 225, 356–375. [CrossRef]

13. Zafar, S.A.; Uzair, M.; Khan, M.R.; Patil, S.B.; Fang, J.; Zhao, J.; Singla-Pareek, S.L.; Singla-Pareek, A.; Li, X. DPS1 regulates cuticle development and leaf senescence in rice. *Food Energy Secur.* 2021, 10, e273. [CrossRef]

14. Hao, Q.; Wang, W.; Zhang, C.; Chen, H.; Chen, L.; Yuan, S.; Chen, S.; Zhang, X.; Zhou, H. Identification and comparative analysis of CBS domain-containing proteins in Soybean (*Glycine max*) and the primary function of GmCBS21 in enhanced tolerance to low nitrogen stress. *Int. J. Mol. Sci.* 2016, 17, 620. [CrossRef]

15. Hao, Q.; Yang, Y.; Shan, Z.; Chen, H.; Zhang, F.; Chen, L.; Yuan, S.; Chen, S.; Yang, Z.; et al. Genome-wide investigation and expression profiling under abiotic stresses of a Soybean unknown function (DUF21) and Cystathionine-β-Synthase (CBS) domain-containing protein family. *Biochem. Genet.* 2020, 59, 83–113. [CrossRef] [PubMed]

16. Yoo, K.S.; Ok, S.H.; Jeong, B.C.; Jung, K.W.; Cui, M.H.; Hyoung, S.; Lee, M.-R.; Song, H.K.; Shin, J.S. Single cystathionine β-synthase domain-containing proteins modulate development by regulating the thioredoxin system in Arabidopsis. *Plant Cell* 2011, 23, 3577–3594. [CrossRef]

17. Shin, J.S.; So, W.M.; Kim, S.Y.; Noh, M.; Hyoung, S.; Yoo, K.S.; Shin, J.S. CBSX3-Trxo-2 regulates ROS generation of mitochondrial complex II (succinate dehydrogenase) in Arabidopsis. *Plant Sci.* 2020, 294, 110458. [CrossRef]

18. Subba, A.; Tomar, S.; Pareek, A.; Singla-Pareek, S.L. The chloride channels: Silently serving the plants. *Physiol. Plant.* 2020, 171, 688–702. [CrossRef]

19. Jena, K.K. The species of the genus *Oryza* and transfer of useful genes from wild species into cultivated rice, *O. sativa*. *Breed. Sci.* 2010, 60, 518–523. [CrossRef]

20. Hajjar, R.; Hodgkin, T. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Europhya* 2007, 156, 1–13. [CrossRef]

21. Dempewolf, H.; Baute, G.; Anderson, J.; Kilian, B.; Smith, C.; Guarino, L. Past and future use of wild relatives in crop breeding. *Physiol. Plant.* 2014, 215, 48–58. [CrossRef]

22. Atwell, B.J.; Wang, H.; Scalfaro, A.P. Could abiotic stress tolerance in wild relatives of rice be used to improve *Oryza sativa*? *Plant Sci.* 2014, 215, 1070–1082. [CrossRef]

23. Ahn, H.K.; Kang, Y.W.; Lim, H.M.; Hwang, I.; Pai, H.S. Physiological functions of the COPI complex in higher plants. *Mol. Cells* 2015, 38, 866–875. [CrossRef]

24. Paul, M.J.; Frigerio, L. Coated vesicles in plant cells. *Semin. Cell Dev. Biol.* 2007, 18, 471–478. [CrossRef]

25. Anantharaman, V.; Iyer, L.M.; Aravind, L. Ter-dependent stress response systems: Novel pathways related to metal sensing, and production of a nucleoside-like metabolite, and DNA-processing. *Mol. Biol. Evol.* 2012, 8, 3142–3165. [CrossRef]

26. Kwon, K.C.; Cho, M.H. Deletion of the chloroplast-localized AtTerC gene product in Arabidopsis thaliana leads to loss of the thylakoid membrane and to seedling lethality. *Plant J.* 2008, 55, 428–442. [CrossRef]

27. Schneider, A.; Steinberger, I.; Strissel, H.; Kunz, H.H.; Manavski, N.; Meurer, J.; Burkhard, G.; Jarzombiski, S.; Schünemann, D.; Geimer, S.; et al. The *Arabidopsis* tellurite resistance C protein together with ALB3 is involved in photosystem II protein synthesis. *Plant J.* 2014, 78, 344–356. [CrossRef]

28. Ramon, M.; Ruelens, P.; Li, Y.; Sheen, J.; Geuten, K.; Rolland, F. The hybrid four-CBS-domain KINβy subunit functions as the canonical γ subunit of the plant energy sensor SnRK1. *Plant J.* 2013, 75, 11–25. [CrossRef] [PubMed]

29. Polge, C.; Thomas, M. SNI1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? *Trends Plant Sci.* 2007, 12, 20–28. [CrossRef]

30. Cheng, C.; Tsuchimoto, S.; Ohsubo, H.; Ohsubo, E. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes Genet. Syst.* 2002, 77, 323–334. [CrossRef] [PubMed]

31. Iwamoto, M.; Nagashima, H.; Nagamine, T.; Higo, H.; Higo, K. p-SINE1-like intron of the CatA catalase homologs and phylogenetic relationships among AA-genome *Oryza* and related species. *Theor. Appl. Genet.* 1999, 98, 853–861. [CrossRef]

32. Zhang, Q.; Kochert, G. Independent amplification of two classes of Tourists in some *Oryza* species. *Genetica* 1997, 101, 145–152. [CrossRef] [PubMed]

33. Li, Y.; Yuan, F.; Wen, Z.; Li, Y.; Wang, F.; Zhu, T.; Zhuo, W.; Jin, X.; Wang, Y.; Zhao, H.; et al. Genome-wide survey and expression analysis of the OSCA gene family in rice. *BMC Plant Biol.* 2015, 15, 261. [CrossRef] [PubMed]
64. Li, X.; Cai, W.; Liu, Y.; Li, H.; Fu, L.; Liu, Z.; Xu, L.; Liu, H.; Xu, T.; Xiong, Y. Differential TOR activation and cell proliferation in Arabidopsis root and shoot apexes. *Proc. Natl. Acad. Sci. USA* 2017, 114, 2765–2770. [CrossRef]

65. Mohanty, B.; Krishnan, S.P.; Swarup, S.; Bajic, V.B. Detection and preliminary analysis of motifs in promoters of anaerobically induced genes of different plant species. *Ann. Bot.* 2005, 96, 669–681. [CrossRef]

66. Loreti, E.; Valeri, M.C.; Novi, G.; Perata, P. Gene regulation and survival under hypoxia requires starch availability and metabolism. *Plant Physiol.* 2018, 176, 1286–1298. [CrossRef]

67. Fujita, Y.; Fujita, M.; Satoh, R.; Maruyama, K.; Parvez, M.M.; Seki, M.; Hiratsu, K.; Ohme-Takagi, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. AREB1 is a transcriptional activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in Arabidopsis. *Plant Cell* 2005, 17, 3470–3488. [CrossRef] [PubMed]

68. Nakabayashi, K.; Okamoto, M.; Koshiba, T.; Kamiya, Y.; Nambara, E. Genome-wide profiling of stored mRNA in Arabidopsis thaliana seed germination: Epigenetic and genetic regulation of transcription in seed. *Plant J.* 2005, 41, 697–709. [CrossRef]

69. Dong, X. SA, JA, ethylene, and disease resistance in plants. *Curr. Opin. Plant Biol.* 1998, 1, 316–323. [CrossRef]

70. Fan, J.; Niu, X.; Wang, Y.; Ren, G.; Zhuo, T.; Yang, Y.; Lu, B.R.; Liu, Y. Short, direct repeats (SDRs)-mediated post-transcriptional processing of a transcription factor gene OsVPI in rice (*Oryza sativa*). *J. Exp. Bot.* 2007, 58, 3811–3817. [CrossRef]

71. McCarty, D.R.; Hattori, T.; Carson, C.B.; Vasil, V.; Lazar, M.; Vasil, I.K. The *Viviparous-1* developmental gene of maize encodes a novel transcriptional activator. *Cell* 1991, 66, 895–905. [CrossRef] [PubMed]

72. Hattori, T.; Vasil, V.; Rosenkranz, L.; Hannnah, L.C.; McCarty, D.R.; Vasil, I.K. The *Viviparous-1* gene and abscisic acid activate the C1 regulatory gene for anthocyanin biosynthesis during seed maturation in maize. *Genes Dev.* 1992, 6, 609–618. [CrossRef] [PubMed]

73. McCarty, D.R. Genetic control and integration of maturation and germination pathways in seed development. *Annu. Rev. Plant Biol.* 1995, 46, 71–93. [CrossRef]

74. Suzuki, M.; Wang, H.H.; McCarty, D.R. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol.* 2007, 143, 902–911. [CrossRef]

75. Li, W.H.; Yang, J.; Gu, X. Expression divergence between duplicate genes. *Trends Genet.* 2005, 21, 602–607. [CrossRef] [PubMed]

76. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. B Biol. Sci.* 1994, 256, 119–124. [PubMed]

77. Force, A.; Lynch, M.; Pickett, F.B.; Amores, A.; Yan, Y.L.; Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999, 151, 1531–1545. [CrossRef]

78. Hughes, T.E.; Langdale, J.A.; Kelly, S. The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome Res.* 2014, 24, 1348–1355. [CrossRef] [PubMed]

79. Wang, B.; Du, X.; Wang, H.; Jin, C.; Gao, C.; Liu, J.; Zhang, Q. Comparative studies on duplicated *tdrd7* paralogs in teleosts: Molecular evolution caused neo-functionalization. *Comp. Biochem. Physiol. Part D Genom. Proteom.* 2019, 30, 347–357. [CrossRef]

80. Hoffmann, R.D.; Palmgren, M. Purifying selection acts on coding and non-coding sequences of paralogous genes in Arabidopsis thaliana. *BMC Genom.* 2016, 17, 456. [CrossRef]

81. Chaudhary, B.; Flagel, L.; Stupar, R.M.; Udall, J.A.; Verma, N.; Springer, N.M.; Wendel, J.F. Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). *Genetics* 2009, 182, 503–517. [CrossRef]

82. Vilo, J.; Brazma, A.; Jonassen, I.; Robinson, A.; Ukkonen, E. Mining for putative regulatory elements in the yeast genome using gene expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2000, 8, 384–394.

83. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 2018, 35, 1547–1549. [CrossRef] [PubMed]

84. Wang, Y.; Tang, H.; DeBarry, J.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012, 40, e49. [CrossRef]

85. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. Ttools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Biol. 2020, 13, 1194–1202. [CrossRef]

86. Saeed, A.I.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; et al. TM4: A free, open-source system for microarray data management and analysis. *BioTechniques* 2003, 34, 374–378. [CrossRef]

87. Jain, M.; Nijhawan, A.; Tyagi, A.K.; Khurana, J.P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* 2006, 345, 646–651. [CrossRef] [PubMed]

88. Schmittgen, T.D.; Livak, K.J. Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 2008, 3, 1101–1108. [CrossRef] [PubMed]