Tandemly Duplicated Rubisco Activase Genes of Cereals Show Differential Evolution and Response to Heat Stress

Ragupathi Nagarajan
Washington State University
Kaviraj Singh Kahlon
Washington State University
Amita Mohan
Washington State University
Kulvinder S. Gill
ksgill@wsu.edu

Washington State University https://orcid.org/0000-0001-7370-6184

Research Article

Keywords: Rubisco Activase, Cereal Crops, Gene Duplication, Heat Stress, Photosynthesis, Molecular Evolution, and Regulation of Gene Expression

Posted Date: August 2nd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4676428/v1

Version of Record: A version of this preprint was published at Plant Molecular Biology on December 21st, 2024. See the published version at https://doi.org/10.1007/s11103-024-01515-z.

Abstract

Heat stress affects various components of photosynthetic machinery of which Rubisco activation inhibition due to heat sensitive Rubisco activase (RCA) is the most prominent. Detailed comparison of RCA coding genes identified a tandem duplication event in the grass family lineage where the duplicated genes showed very different evolutionary pattern. One of the two genes showed high level of sequence conservation whereas the second copy, although present only 1.5kb away, was highly variable among various plant species because of loss of introns, alternative splicing and loss of the last exon coding redox regulated C-terminal extension domain. Gene specific expression analysis, both at the transcription as well as the protein level, showed very different expression pattern of the two RCA copies. Expression of the highly conserved copy was higher under normal plant growing conditions that decreased many folds under heat stress with substantial genotypic variation, but the variable copy showed much higher expression under heat stress conditions across all grass species. The cultivated rice has only one functional gene as the second copy became nonfunctional due to multiple deletions but *Oryza brachyantha* and *Oryza australiensis*still have two functional *Rca* genes. Detailed analysis of the promoter region of the two copies among various plant species showed insertion of several transposable elements harboring heat responsive elements in the heat inducible copy of the gene. The conserved RCA copy of wheat didn't have any transposable insertions whereas in that of maize has one heat shock element and sorghum had two. It would be interesting to study if the higher level of heat stress tolerance observed in sorghum and maize is associated with the differences observed for RCA.

Introduction

The optimal temperature for photosynthesis varies among crops, from 25-30°C in wheat to 30-35°C in rice (Nagai and Makino, 2009). Every degree-Celsius increase over the optimal temperature may result in up to 6% reduction in grain yield of wheat and 3.2% of rice (Zhao et al., 2017). Although many plant metabolic pathways are affected by heat stress, the effect on photosynthesis is critical as any constraint on the process will directly impact development, growth, biomass production, and crop yield. Plant photosynthetic machinery involves complex components and processes, many of which including Rubisco activation, enzymes and proteins of photosystems, membrane permeability and ATP generation are vulnerable to high temperature (Allakhverdiev et al. 2008). The critical temperature and its duration affecting each photosynthetic process often differ among species (Feller et al., 1998). Plants exposed to temperatures above 40°C for extended periods show irreversible damage to the permeability of thylakoid membranes and electron transport chain, resulting in photosynthesis inhibition. The photosystems and electron transport are usually not affected until 40°C, whereas for many species including some important crop plants including wheat the Rubisco activation level and photosynthetic rate (Pn) start to decline even at moderately elevated temperatures (Feller et al., 1998).

The functional state of Rubisco is maintained by Rubisco activase (RCA) that removes inhibitory sugar phosphates from the active sites of Rubisco with energy derived from ATP hydrolysis (Salvucci et al., 1985). RCA is a nuclear encoded, cytosol synthesized chloroplast protein, generally consisted of two

immunologically related polypeptides called α isoform (45 to 46 kDa) and β isoform (41 to 43 kDa). The size difference between the two isoforms is due to the presence of a redox regulated C-terminal extension (CTE) domain that is present in the α isoform but is lacking in the β form. In most plants, both α and β isoforms of RCA are capable of activating Rubisco (Werneke et al., 1989; Rundle and Zielinski, 1991; To et al., 1999; Salvucci et al., 2003). Shown in Arabidopsis, the two cysteine (Cys) residues located in the CTE domain of RCA α isoform make a disulfide bond under oxidizing conditions. This significantly increases its sensitivity to physiological level of ADP/ATP ratio and affects both ATP hydrolysis and Rubisco activation (Zhang and Portis, 1999; Zhang et al., 2002). The major function of larger isoform is to regulate the activity of smaller RCA isoforms under low light/dark conditions in order to conserve energy. The equimolar concentrations of the larger and the smaller isoforms, completely inhibited the activity of smaller isoform under higher ADP/ATP ratio in light conditions (Zhang et al., 2001). Interestingly, tobacco, which expresses only the β isoform, is also sensitive to physiological levels of ADP/ATP ratio and responds similar to species with both isoforms (Carmo-Silva and Salvucci, 2013). Selection for genotypes with higher *Rca* expression levels at normal conditions in soybean, maize, and some other crops showed a positive correlation with Rubisco activity, photosynthetic rate, and seed yield (Morales et al., 1999; Yin et al., 2010, 2014).

In-vitro analysis has shown that while the RUBISCO enzyme is stable even at 50^{o}C, RCA is very sensitive to heat stress and its activity start to decline at temperatures over 35^{o}C (Crafts-Brandner and Salvucci, 2000). As a result, at higher temperatures RCA is unable to keep pace with faster Rubisco deactivation thus constrains photosynthesis (Crafts-Brandner and Salvucci, 2000). In crops including cotton and wheat, even a moderate level of heat stress decreases the Rubisco activation levels, which correlated with reduced net photosynthesis, increased RuBP level, and decreased levels of 3-phosphoglycerate in leaves (Law and Crafts-Brandner, 1999). Independent studies on various crops have shown that photosynthetic heat acclimation is directly associated with the acclimation of heat sensitive RCA to high temperature (Law and Crafts-Brandner, 1999; Yamori et al., 2006). This conclusion was further supported by the studies showing increase in the thermal tolerance of Arabidopsis photosynthesis by improving RCA thermostability (Kurek et al., 2007; Kumar et al., 2009).

Previously, we showed that the *Rca* gene structure is highly conserved in higher plants, and the common grass ancestor *Rca*, similar to Arabidopsis *AtRca1*, contained seven exons and six introns, enabling expression of the RCA α isoform with redox regulated CTE domain coded from the seventh exon (Nagarajan and Gill, 2018). The alternatively spliced mRNA of the same gene contains a premature stop codon and produces a shorter redox insensitive RCA β isoform (Werneke et al., 1989). We also reported that the *Rca* gene is tandemly duplicated (*Rca1* and *Rca2*) among grass species and the duplication event probably occurred in the common grass ancestor, but before the divergence of subfamilies (Nagarajan and Gill, 2018). Heat stress induced changes in the RCA expression pattern at protein level were reported earlier in rice (Scafaro et al., 2010; Wang et al., 2010), wheat (Feller et al., 1998; Law and Crafts-Brandner, 2001; Ristic et al., 2009), and maize (Crafts-Brandner and Salvucci, 2002; Ristic et al., 2009). High temperature stress altered the ratio of α and β isoform levels in rice (Wang et al., 2010), but

induced new RCA isoform expression in wheat and maize (Law and Crafts-Brandner, 2001; Crafts-Brandner and Salvucci, 2002; Ristic et al., 2009). A recent study in wheat showed that heat stress altered the expression pattern of *Rca* genes (Scafaro et al., 2019). Later, (Degen *et al.*, 2020) showed a single amino acid substitution (M159I) in TaRCA2-β isoform could increase the thermal stability of the enzymatic activity by up to 5°C. However, the heat stress induced expression pattern of RCA isoforms in important cereal crops and key evolutionary changes causing the differential gene expression are still not well studied. Earlier, we proposed an evolutionary model for tandemly duplicated *Rca* genes of the grass species (Nagarajan and Gill, 2018). We also showed that after the duplication event, one of the gene copies, *Rca1*, have lost the third, the fifth and the sixth introns, which also abolished the alternative splicing mechanism of the *Rca* transcripts. Other changes observed in Rca1 among various grass species included nonfunctional *Rca1* in some Oryzoideae species; 2) loss of *Rca1* sequences coding the redox regulated CTE domain in Pooideae including that in Brachypodium, barley, and wheat and the loss of various introns (Ragarajan and Gill, 2018). 3) loss of *Rca2* sequences coding for the redox regulated CTE domain in Panicoideae species (maize, sorghum and foxtail millet), and 4) loss of various introns in both *Rca1* and *Rca2* of these crops. Whereas, the second gene copy, *Rca2*, did not amass any major changes in its gene structure although loss of some introns and sequences coding for the redox regulated CTE domain in Panicoideae species including maize, sorghum and foxtail millet was observed.

In the present study, first, we studied the origin of the tandem duplication of *Rca* genes. Specifically, we tracked the evolution of the rice *OsRca1* by comparing the sequences of various Oryzoideae species. Second, we studied the effect of heat stress on the expression pattern of tandemly duplicated cereal *Rca* gene copies both at the mRNA as well as at the protein level and investigated possible causes (i.e. promoter elements) of the variation. Finally, we performed a phylogenetic study to analyze the rate of amino acid substitution and the evolutionary relationship of different monocot RCA isoforms. The information was then used to understand the role of tandemly duplicated *Rca* genes in Rubisco activation, and photosynthetic acclimation to heat stress in important cereal crops.

Results

Tandem duplication of Rca gene occurred after the divergence of Poales families

With the objective to accurately determine origin of the tandem duplication in the RCA gene, we evaluated the *Rca* gene structure in three distantly related monocot species: banana (*Musa acuminata*), date palm (*Phoenix dactylifera*) and pineapple (*Ananus comosus;* family: Bromeliaceae; order Poales). Unlike the grass family, no tandem duplication of the *Rca* gene was observed in any of these species. Instead, these species contained three *Rca* gene copies on separate chromosomes (table S1). The *Rca* gene copies in banana were located on chromosomes 2, 10 and 11, whereas in date palm, the copies were present on two different unplaced scaffolds NW_008246537.1 (1.6 Mbp) and NW_008246711.1 (0.506 Mbp) and were not tandemly duplicated as was seen in grasses. The gene copies in pineapple were located in the linkage groups (LG) LG14, LG18, and LG23. Otherwise, the structure of the gene copies was similar to those of other higher plants. In all three species, the exon sizes and relative order

were highly conserved among the three copies as was seen in other plant species. Major variation was however observed for intron length (Fig. 2b). Based on these observations, we concluded that the gene duplication event occurred after the divergence of the Poales families and is probably specific to the Poaceae.

Rca gene structure among cereals

After evaluating Rca gene structure among various grass family members, we identified and sequence confirmed *Rca* gene copies from seven (*O. brachyantha*, *O. australiensis*, *O. sativa*, *T. aestivum*, *Z. mays*, *S. bicolor*, and *Setaria italica*) species. Since rice showed major variation for the *Rca* gene structure, we also obtained sequence of seven cultivated rice accessions and two wild rice species *(Oryza australiensis* and *Oryza brachyantha*). For each accession, both genomic and cDNA copies were cloned and sequenced. While the *Rca2* sequence from one of the accessions was reported earlier (To et al., 1999; Scafaro et al., 2016), we cloned and sequenced the *Rca1* gene from *O. australiensis* (GenBank Acc. MH115971), and *O. brachyantha* (GenBank Acc. MH220418), and *Rca2* gene from *O. brachyantha* (GenBank Acc. MH660917).

The structural comparison of *Rca1* gene copy across various cultivated and wild rice species revealed presence of large deletions in the exons 2, 3, 4, 5, 6 and 7 which appeared to have caused frameshift mutation in the *Rca1* of cultivated rice species (Fig. 3). In other species, between 11 to 13 independent deletions were detected in all four exons, resulting in 157 to a total loss of 218 nucleotides. These *Rca1* deletions appear to have occurred multiple times, but in different time periods during the evolution. For example, *O. barthi* and *O. glumipatula* have a 15 bp deletion on the third exon, whereas *O. nivara*, *O. sativa*, and *O. rufipogon* showed a 39 bp deletion (Fig. 3; Box1). Similarly, the fourth exon of *Rca1* of *O. nivara*, *O. sativa*, and *O. rufipogon* contains two additional deletions of 13 and 24 nucleotides (Fig. 3; Box 2), which were absent in other wild species. Additionally, species specific indels in the *Rca1* gene were also identified. The six bp addition in the second exon of *O. punctata Rca1*, a 25 bp deletion in the fourth exon of *O. nivara*, and a nine bp instead of the 14 bp deletion in the second exon of *O. glumipatula Rca1* occurred probably after the divergence of the species. However, it is not clear how the 24 bp deletion is absent in *O. sativa* ssp. *indica* but present in most other species including *O. sativa* ssp. *japonica*. Interestingly, the intron lengths of *Rca1* were determined before the major *Oryza* species divergence and maintained thereafter (Fig. 3). Alignment of RCA2 amino acid sequences of these species showed 100% similarity, except for *O. brachyantha* and *O. punctata*, which had only 95.5% and 97.4% similarities with the other *Oryza* species respectively.

Rca1 gene structure in Oryzoideae subfamily

From the comparative analysis of *Rca* genes in Poaceae, it was evident that both Pooideae and Panicoideae subfamilies have two functional *Rca* gene copies present in tandem (Fig. 1). The *OsRca1* copy contained multiple small deletions that seem to have occurred after the divergence of these sub-families. Comparison of the two wild and seven cultivated rice species revealed the presence of a putative *Rca1* gene in all these species. Structurally, the *Rca1* gene in *O. australiensis*, *O. brachyantha*,

and *O. punctata* contained four exons with slight size variation for the three introns (Fig. 3). The remaining five species including the two accessions of cultivated rice showed length of the three introns were remarkably conserved. The exons of these five species showed 11 to 13 deletions (Fig. 3). Nine of these deletions were common to all five species and the remaining were specific to one or more species. At the predicted protein level, the largest protein among these three species was predicted to be 466aa in *O. punctata*. The predicted protein in *O. australiensis* is 465aa as compared to 458 in *O. brachyantha*. Unlike the RCA1 of Pooideae members, these predicted peptides also contain the light sensitive CTE domain as identified in the RCA1 of the Panicoideae members.

Thermostability conferring amino acids in RCA1 isoform of cereals

Previous studies showed that *Arabidopsis* transgenic plants expressing RCA variants with certain amino acid substitutions were heat tolerant, and the modified Arabidopsis RCA showed up to 10°C increase in stability and Rubisco activation activity (Kurek et al., 2007). One such substitution T274R (Threonine 274 Arginine) in the variant 183H12, which alone could provide the maximal heat stability and activity, was identified in all three copies of TaRCA1 and ZmRCA2 (Fig. 6). Other natural variants observed in the same position were glutamine (Q) in SbRCA1, SbRCA2, and ZmRCA1, and lysine (K) in ObRCA1 and ObRCA2 of *O. brachyantha*, and in all of the RCA2 isoforms of rice and wheat. Based on these results, we hypothesize that the isoform expressed during heat stress may have better thermal stability than the isoforms expressed at normal conditions because of the putative thermal tolerant amino acid substitutions. The role of these natural RCA variants in thermal tolerance is not known, and further experiments are needed to identify the significance of these sequence changes in providing thermal tolerance.

Phylogenetic analysis of monocot RCA

Phylogenetic analysis performed with the alignment of RCA protein from distantly related monocots, using the MEGA7 maximum likelihood method, revealed different clusters of RCAs, based on the sequence similarity and evolutionary pattern (Fig. 7). In deep nodes, the phylogenetic relationship is unclear, and the bootstrap (BS) values are low (< 70%). The comparative analysis showed how the duplicated *Rca* genes evolved in cereals. Thus, we did not collapse the deep nodes with lower statistical support. Based on the phylogenetic analysis, RCA isoforms of banana, date palm and pineapple, and the cereal isoforms, formed distinct clades, with cereals further comprising of RCA1 and RCA2 from the tandemly duplicated genes. Although the RCA isoforms of banana (MaRCA1, 2 and 3), date palm (PdRCA1 and 2), and pineapple (AcRCA1, 2 and 3) were clustered together, the subgroup pattern within the cluster suggests that the genes coding these isoforms were duplicated or triplicated after the divergence of monocot orders. Similarly, the RCA1 and RCA2 of cereals formed different clades, but within each clade, the isoforms of closely related species, or species from a shared family, showed less divergence with high BS values. The branch and individual leaf lengths show the rate of substitution, the RCA1 isoforms of wild rice and Panicoideae species showed a higher rate of changes than did the RCA1

of Pooideae. Together, these results suggest that after the divergence for a common monocot ancestor, the *Rca* gene copy number increased either through polyploidization or tandem duplication and evolved at different rates in the studied species.

Expression pattern of RCA isoforms in cereals

Detailed analysis of 302 ESTs from banana (*Musa acuminata*) and 331 from pineapple (*Ananus comosus*) showed that in both banana and pineapple, only one of the three copies codes for RCA-α, whereas the other two copies code for only the β isoform (Supplementary Table S1). EST sequences showed no evidence of alternative splicing in the RCA-α coding copy, suggesting a possible loss of the splice junction during evolution. In comparison, alternate splicing was observed in date palm (*Phoenix dactylifera)* based on the analysis of 305 ESTs. One of the two *Rca* copies, *PdRca1*, codes for both RCA-α and RCA-β isoforms via alternative splicing, whereas the other gene copy (*PdRca2*) codes only for the RCA-α isoform.

Evaluating the EST data, *MaRca1-α, MaRca2-β*, and *MaRca3-β* of Banana has 101, 101 and 100 ESTs respectively; *PdRca1-α, PdRca1-β and PdRca2-α* of date palm has 103, 102 and 100 ESTs respectively and *AcRca1-α, AcRca2-β and AcRca3-β* of pineapple has 110, 115 and 106 ESTs respectively. Based on this data, the three splice variants appear to be equally expressed in these three plants. With the objective to compare expression of the gene(s) at the RNA level with that at the protein level, and to identify the relative contribution of the two tandemly duplicated *Rca* gene copies, we performed gene specific expression analyses using quantitative Real Time PCR (qRT-PCR) for both *Rca1* and *Rca2* genes of *O. brachyantha*, *O. australiensis*, *O. sativa*, *T. aestivum*, *Z. mays*, and *S. bicolor* at normal as well as at high temperature conditions (Fig. 5). In cultivated rice (*O. sativa*), either under heat stress or normal plant growth conditions, no expression was observed for the *Rca1* copy in either of the two cultivars that were used for the analysis (Fig. 5a). *O. australiensis* showed some expression of *Rca1* under normal growth conditions and the expression reduced both under 35^{o}C and 45^{o}C treatment. *O. brachyantha* on the other hand showed a different expression pattern for *Rca1* where its expression was very low under normal plant growth conditions but increased under the heat stress conditions. *Rca1* expression was however significantly less than that for *Rca2*. The *Rca2* copy in cultivated rice showed very high expression level under normal plant growth conditions and its expression reduced dramatically under heat stress conditions. There was significant difference between the two cultivated rice lines for the expression of *Rca2* copy. The two rice wild relatives showed dramatically higher expression for *Rca2* under heat stress. Even between the two wild relatives, *O. brachyantha* showed significantly higher expression at 45^{o}C treatment as compared to *O. australiensis*.

Heat stress also affected the alternative splicing mechanism, as the reduction in the levels of transcripts coding RCA2α varied during heat stress (Fig. 4). In rice, *OsRca2* gene expresses *OsRca2α* and *OsRca2β* transcripts via alternate splicing mechanism. At normal conditions, the level of *OsRca2β* transcripts was many hundred-folds higher than that of the *OsRca2α* transcripts, but at 42°C, the *OsRca2α* transcript level was either equal or higher than that of the *OsRca2β* transcripts. Interestingly, *ObRca2α* transcripts

were stable at all studied temperatures, however, *ObRca2β* showed more than two-fold reduction at higher temperatures. These results suggest that the temperature increase, at least among various rice species, favors transcripts coding RCA2α isoform. Even though the transcripts encoding OsRCA2α and OsRCA2β showed a significant reduction at 45°C in rice, the quantity of both the isoforms from the heat stressed leaves is comparable or higher to their control leaves (Fig. 4).

The expression pattern of the *Rca* gene copies in wheat was very different from that observed in rice. *Rca1* showed essentially no expression at normal plant growth conditions, but its expression increased dramatically under heat stress conditions (Fig. 4b). The expression pattern was very similar among the three cultivars (Chinese Spring, Giza168 and PBW343) although significant differences were observed among cultivars. In all three cultivars, transcript level for *Rca2* copy was significant at all three temperature conditions although its level at 37^{o}C was significantly higher than either of the other two conditions. Genotypic differences were significant among the three genotypes at each temperature condition. Between the two isoforms of *Rca2*, transcript level of *Rca2α* isoform was much lower than the *Rca2β* form and stayed about the same over the three temperature treatments (Fig. 4b). The *Rca2β* form showed significant increase at 37^{o}C treatment but reduction at 42^{o}C.

Both in maize and sorghum, *Rca1* copy made only the alpha isoform of the gene whereas the *Rca2* copy made the beta form (Fig. 4c and d). In both of these species, the beta form was made at all three temperatures although significant differences were observed both among temperatures as well as among genotypes. The alpha form showed essentially no accumulation at 25^{o}C, slight increase at 35^{o}C treatment and dramatic increase at 45^{o}C. In general, the genotypic differences for relative expression of the two isoforms were much higher than that observed among the wheat genotypes.

Using the same plant samples as that used for real-time expression analysis, RCA protein level was also studied using immunoblot analysis with an RCA specific antibody (Fig. 5). Except for wheat, all of the studied cereals expressed a 42 kDa RCAβ isoform at normal as well as under heat stress conditions whereas the 46 kDa RCAα isoform was expressed only during the heat stress conditions (Fig. 5). In comparison to Rubisco, the expression level of wheat RCAβ was significantly higher at the normal temperature as compared to that under heat stress. Relative to RCAβ, expression of the RCAα is much lower at normal conditions but is undetectable after the heat treatment. Although it is difficult to accurately quantify, expression levels of both Rubisco, as well as RCA of rice, maize, and sorghum, increased after 45°C treatment as compared to that at the normal temperature. Wheat, on the other hand did not show the 46 kDa RCAα isoform under heat stress conditions along with a significant reduction in the RCAβ level. With the current analysis, it was not possible to determine if the 42 kDa RCAβ isoform of wheat is expressed from *TaRca1* or the *TaRca2* copy.

Regulatory elements in the Rca genes of cereals

As the two *Rca* gene copies of cereals showed very different expression pattern, we looked for differences in expression control elements between promoters of the two copies. We evaluated

regulatory elements in ~ 1.5 kb region upstream of the transcription start site of the two *Rca* gene copies of the studies species along with the intronic sequences, with a focus on heat regulated elements (material and methods). Along with the core promoter elements, this analysis showed various putative cis-elements including that for circadian rhythm, light, hormone and biotic and abiotic stress responsives (Table S2). The *Rca1* copy showed insertion of heat shock elements (HSE) in its promoter with wild rice (*O. brachyantha*) having three HSEs. These HSEs were missing in the *Rca1* promoter of cultivated rice (*O. sativa*). Interestingly, *Rca2* of *O. sativa* have HSEs in its promoter and in the 5th intron. Transposable element (TE) of size 366bp, 347bp and 320 bp were present in the promoter of *Rca1*, promoter of *Rca2* and in the 5th intron of *Rca2*, respectively (Fig. 1). In wheat, *Rca1* promoter of the *A* and the *B* genome showed one HSEs and the *D* genome has two. There were no HSEs present in the promoter of *Rca2* of wheat but promoter of D genome showed insertion of 162 bp TE. In maize (*Z. mays*) and sorghum (*S. bicolor*), both *Rca1* and *Rca2* had insertion of HSEs with sorghum having maximum of 5 HSEs in the promoter of *Rca1*. Promoter of *Rca2* had two HSEs in sorghum and one in maize. TE were also present in the promoter of *Rca1* in both maize and sorghum (Fig. 1)

Discussion

Adaptation of organisms to the changing environmental conditions via modifications in the basic biological processes is a common feature of evolution. For better adaptation, an organism often relies on new gene mutations which may develop neo-functions, however, that could also cause detrimental effects if the mutated gene is essential for the organism's survival. Gene and genome duplications provide the base genetic material for the origin of novel genes or trait development during evolution.

Tandem duplication of *Rca* genes was first reported in barley (Rundle and Zielinski, 1991) and later in wheat (Carmo-Silva *et al.*, 2015). In rice, DNA gel blot analysis with cDNA coding RCA as a probe showed the presence of more than one *Rca* gene in the genome (Zhang and Komatsu, 2000). We (Nagarajan and Gill, 2018) showed that the tandem duplication event occurred in the common grass ancestor and the structural changes then accumulated after the divergence of the Poaceae families. In this study we have shown that this tandem duplication of *Rca* gene is Poaceae-specific and probably occurred after the divergence of Poales order as one (Bromeliaceae) of the 16 families did not have the duplication. It would be possible to pinpoint the exact timeline of the duplication event once the genome sequence information for the other 14 families is available.

Phylogenetic and fossil data analyses have suggested that the grass sub-families probably diverged from a common ancestor between 50-70 million years ago (MYA) (Kellogg, 2001; Salse *et al.*, 2008). The tandem duplication event of *Rca* genes might have occurred before the divergence of the grass sub-families as the Oryzoideae, Pooideae, and Panicoideae members have the duplicated genes (Fig. 2). The conserved heat induced expression pattern of *Rca1* in all three sub-families suggests that the promoter elements required for heat responsive expression might have integrated into the *Rca1* promoter of the grass ancestor. Later, species specific TE insertions added additional HSEs in the promoter region and further altered the gene expression. Effect of TE insertions on the neighboring gene expression is a well-

documented phenomenon in plants and other organisms (see reviews, Feschotte, 2008; Lisch, 2013; De Souza et al., 2013). Similarly, integration of biotic and abiotic responsive cis-regulatory elements in the promoter regions of genes by TE was also reported in plants (Beguiristain *et al.*, 2001; Naito *et al.*, 2009), including HSEs (Ito *et al.*, 2013). The insertions of TEs with HSE elements in duplicated *Rca* gene regions of cereals could be the source for evolutionary adaptation through altered gene expression. Heat stress induced changes in the RCA expression pattern were reported earlier in rice (Scafaro *et al.*, 2010; Wang *et al.*, 2010), wheat (Feller *et al.*, 1998; Law and Crafts-Brandner, 2001; Ristic *et al.*, 2009), and maize (Crafts-Brandner and Salvucci, 2002; Ristic *et al.*, 2009). Our copy specific *Rca* transcript quantification along with immunoblot analysis have specifically revealed expression pattern of each of the *Rca* copies and of its splice varients. We have also shown differences and similarlities among *Rca* of various species for a differential response to heat stress. In comparison to maize and sorghum, the wheat *Rca* showed almost an opposite accumulation pattern of different *Rca* isoforms and splice varients (Figures 4&5). Role of these differences in explaining differences in heat stress tolerance among wheat, maize and sorghum needs to be validated.

One of the important findings of this study is the identification of a functional *Rca1* gene copy among wild rice species. The genus *Oryza* is composed of 23 recognized species, including two cultivated (*O. sativa* and *O. glaberrima*) and 21 wild rice species with 10 different genomes (AA, BB, CC, BBCC, CCDD, EE, FF, GG, HHJJ, and HHKK) (Ge *et al.*, 1999; Vaughan *et al.*, 2003). Our analysis on the non-functionalization of *Rca1* showed that the species *O. sativa*, *O. glaberrima*, *O. barthii*, *O. glumipatula*, *O. nivara*, and *O. rufipogon* with AA genome have multiple deletions in the exonic sequences thus have non-functional *Rca1*. Interestingly, *O. punctata* with BB genome, *O. australiensis* with EE, and *O. brachyantha* with FF genome contain a complete and functional *Rca1* gene. These results suggest that the non-functionalization of *Rca1* in the AA genome species occurred after the divergence of *Oryza* species and also specifically in the AA genome lineage. Recent research on thermal tolerance of wild rice species *O. australiensis* showed unaffected net photosynthetic and leaf elongation rates, even at 45°C, whereas *O. meridionalis* and cultivated rice *O. sativa* showed intermediate tolerance to > 50% decline in net photosynthesis and leaf growth rates (Scafaro *et al.*, 2016). Rubisco activation levels of these species at high temperature correlated with the net photosynthesis and leaf elongation rate. Interestingly, the thermal stability of recombinant OaRCA2 was similar to the OsRCA2 of cultivated rice in the *in vitro* assays, but the leaf extracted RCA of *O. australiensis* showed no Rubisco activation inhibition at temperatures as high as 42°C, whereas rice RCA inhibition occurred at temperatures above 34°C. The authors characterized the *OaRca2* of *O. australiensis* and compared its thermal stability with that of rice *OsRca2*. They speculated that some unknown factor(s) in the leaf extract could contribute to its higher thermal activity and stability (Scafaro *et al.*, 2016). Based on our *Rca1* comparative analysis of *Oryza* sp. and heat stress induced expression of *OaRca1* of *O. australiensis*, we propose that the presence of OaRCA1 could be one of the reasons for increased thermal stability of Rubisco activation and unaffected net photosynthetic observed in the above-mentioned study. However, further experiments are needed to characterize the *OaRca1* and utilize this gene in rice breeding through alien introgression to improve photosynthetic thermal tolerance of rice cultivars.

Various heat shock elements (HSEs) present in the sequences upstream of the transcription start sites (TSS) were identified in *ObRca1* of *O. brachyantha*, *OsRca2* of rice, all three copies of wheat *TaRca1*, both *ZmRca1* and *ZmRca2* of maize, and both *SbRca1* and *SbRca2* of sorghum (Fig. 1). The number of HSEs also varied among species with the promoter of *SbRca1* containing up to five, *ObRca1* containing three, *TaRca1-D* and *SbRca2* containing two, and the remaining ones showing only a single HSE in the upstream region. It is tempting to speculate a correlation of higher level of heat stress tolerance in sorghum and wild rice with higher number of HSE's but additional experimentation is needed to confirm.

Previous studies have shown that certain transposable elements are preferentially inserted into the regulatory sequences of various genes and play an important role in the expression control of the neighboring genes (Makarevitch *et al.*, 2015). In some cases, heat stress induced gene expression was acquired from TE insertion events (Ito *et al.*, 2011). The duplicated *Rca* genes of closely related yet evolutionarily diverged cereals, and their heat stress induced expression pattern may serve as an ideal model for understanding different TE insertions and their role in the stress regulated expression. To identify TEs in the genic region, we used the duplicated gene sequences, along with 2Kb sequence flanking on both sides, as the query to search the crop specific TE database. We identified independent TE insertion events in the promoter sequence of all six *Rca1* copies of the studied species (Table 2). Similarly, there were TE insertions in the promoter and the fifth intron of rice *OsRca2*, and in the promoter of *TaRca2-D* copy (Fig. 1). Interestingly, many of these TEs contain HSEs; these TEs inserted in different time periods of evolution served as a potential source for the HSEs and may have caused heat stress induced expression pattern in the *Rca1* genes. A detailed list of the identified TEs, their sizes, and locations relative to the genes is given in Table 2.

In conclusion, the comparative and expression analyses of RCA coding genes of important cereal crops revealed a tandem gene duplication, which occurred in the common grass ancestor subsequently led to neo-functionalization and evolved to respond differently during heat stress. One of the two copies of the tandem duplication maintained a very high level of conservation whereas the second copy showed tremendous divergence to evolve species specific function of the gene. We identified the insertion of various transposable elements harboring heat responsive elements in the promoter region. These results suggest that one of the duplicated *Rca* genes showing divergence, which could be responsible for the heat induced expression pattern and thermal stability of these heat tolerant cereal crops. Understanding these evolutionary changes are important as this knowledge can be used to engineer future plants with better and wider adaptability.

Materials and Methods

Gene model prediction and validation

The genomic regions comprising *Rca* genes of banana, date palm, and pineapple were searched using the well characterized 474 aa sequence of AtRCA1 (AT2G39730.1) (https://www.arabidopsis.org/) as the query in Phytozome 11 (https://phytozome.jgi.doe.gov) and NCBI database. We selected banana, date

palm, and pineapple because of its position in the monocot evolution with sequenced genome at the time of analysis. Gene models were developed using the predicted transcripts for each of these genes from NCBI database (Table S1). The *Rca* gene models for rice (cv. Nipponbare), wheat (cv. Chinese spring), maize (inbred B73), and sorghum (inbred BTx623) were adopted from our previous study (Nagarajan and Gill, 2018). To identify the *Rca* genes of wild rice species, AtRCA1 was used as a query sequence to search 11 *Oryza* species (Jacquemin et al., 2013) of Ensembl plants DNA database with default settings (http://plants.ensembl.org/index.html). For the *Oryza Rca1* structural analysis, we did not include *O. meridionalis*, *O. glaberrima*, and *O. longistaminata*, because the sequences for the region was not complete in these three species. The criteria used for searching true orthologs in this study were described elsewhere (Dhaliwal et al., 2014). For identification of cis-regulatory elements in the promoter region, we used the 1.5 Kb upstream sequence from ATG start codon for each gene to search in PlantCARE database (Lescot et al., 2002). The transposable elements for *Oryza* species were identified from RiTE-DB (http://www.genome.arizona.edu/cgi-bin/rite/index.cgi); wheat from cereal repeat database (http://wheat.pw.usda.gov/ggpages/Repeats/), maize from MaizeGDB (http://www.maizegdb.org), and sorghum from TREP database (http://botserv2.uzh.ch/kelldata/trep-db/index.html). For validation of the gene models, cDNA sequences coding the mature isoform of Rca were PCR amplified using *attB* sites flanking transcript specific primers (Supplementary table S3) with PrimeSTAR GXL polymerase (Clontech-Takara, USA) and cloned in pDONR201 vector using BP Clonase-II enzyme (Invitrogen, USA) as per the manufacturer's recommendation. Multiple clones per construct were sequenced and the data was validated with the previously predicted gene models. *Rca* gene from banana, date palm, and pineapple was included in the analysis due to their position in the monocot evolution. Gene models were developed using the predicted transcripts for each of these genes from NCBI database (Table S1).

To identify the *Rca* genes of wild rice species, AtRCA1 was used as a query sequence to search 11 *Oryza* species (Jacquemin et al., 2013) of Ensembl plants DNA database with default settings (http://plants.ensembl.org/index.html). For the *Oryza Rca1* structural analysis, we did not include *O. meridionalis*, *O. glaberrima*, and *O. longistaminata*, because the sequences for the region was not complete in these three species.

Plant growth condition and heat treatment

Rice cultivars Nipponbare (GSOR 100) and N22 (PI 389102), maize inbred lines B73 (PI 550473) and Mo22 (PI 558534), and sorghum genotypes BTx623 (PI 564163), SC1019 (PI 656071), SC1047 (PI 656072) and B35 (PI 147224) were received from Germplasm Resources Information Network (https://www.ars-grin.gov). The wild rice species *Oryza brachyantha* (IRGC 101232) and *Oryza australiensis* (IRGC 101397) used in this study were received from Gerald E. Edwards, Washington State University, Pullman, WA, USA. The wheat cultivars Chinese Spring, Giza168, and PBW343 were used in the experiment. All plant materials were grown in Conviron PGR15 growth chamber equipped with high-intensity discharge lamps capable of Photosynthetically Active Radiation (PAR) of > 600 µmol/m^2/sec before treating at high temperatures. Except for the rice and wild rice species, all plants were planted in 4

x 9-inch-deep pots using Sunshine LC1 potting mixture (SunGro Horticulture, Bellevue, WA, USA) supplemented with 14g Osmocote 14-14-14 slow releasing fertilizer. Cultivated and wild rice species were grown as mentioned in Giuliani et al. (2013). The wheat plants were grown under controlled temperature of 24/18°C with 16/8 hours of the day/night cycle. For rice maize, and sorghum genotypes the temperature was 25/22°C with 12/12 hours day/night cycle. Heat shock treatments were given to plants with fully expanded leaves (four-five weeks old). The chamber temperature was gradually increased (5°C/hour) from morning to reach the target temperatures and maintained for four hours before collecting leaf samples. To match the circadian rhythm, samples from control sets were also collected at the same time period. Leaf samples were frozen immediately in liquid nitrogen and stored in -80°C freezers for subsequent RNA and protein extractions.

Immunoblot analysis

Protein extraction and immunoblot analyses were performed as described in Yamori and von Caemmerer (2011) with some modifications. Protein concentrations were measured using Bradford reagent (Bio-Rad). For Fig. 4, 2ul of the total protein was loaded per well on 1D-SDS-PAGE. Immunoblots were probed first with cotton anti-Rubisco activase (Agrisera, Catalog # AS10-700) antibody and later with horseradish peroxidase-conjugated secondary antibody as per the manufacturers' recommendation. Immunological bindings were detected using Pierce ECL Western Blotting Substrate (Thermo Scientific) with Kodak X-ray film developing system.

RNA extraction, cDNA preparation, and Real Time (RT) PCR analysis

Total RNA from the normal and heat stress treated leaf samples were extracted using the hot phenol RNA extraction method. In brief, 100–200 mg frozen tissue samples were ground in liquid nitrogen and 0.75 ml of preheated (80°C) extraction buffer containing a mix of 100 mM Tris-HCl (pH 8.0), 100 mM Lithium Chloride (LiCl), 20 mM Ethylenediaminetetraacetic acid (EDTA), 1% Sodium dodecyl sulfate (SDS) prepared in Diethyl pyrocarbonate (DEPC) treated water and equal volume of phenol (pH 4.8) was added. Before centrifugation, the homogenates were vortexed, chloroform (0.35 ml) was added and vortexed till the mix turn whitish green. The resulted supernatant was separated, precipitated with 8M LiCl in ice for two hours and spun at 4°C for 30 min. The RNA pellets were washed with 2M LiCl and 70% ethanol and air dried. The pellets were dissolved in DEPC treated water and DNAse digestion was performed using Turbo™ DNAse kit (Thermo Fisher Scientific; Catalog # AM2238) as per the manufacturer's recommendation. Purified RNA was loaded on a 1% agarose gel to check the quality and quantity of the sample. First strand cDNA was synthesized from the purified RNA samples using M-MLV Reverse transcriptase (Promega; Catalog # M1701) with anchored-oligo $(dT)_{18}$ primer as per the manufacturer's recommendation. The cDNA samples were diluted 50 times and used in the RT-PCR reactions. All the RT-PCR analysis were performed using the KAPA SYBR® FAST qPCR master mix (Kapa Biosystems Inc; Catalog # KK4609) on LightCycler® 480 (Roche Diagnostics), as per the manufacturer's instructions. The sequences of all the PCR primers used for the RT-PCR are given in Table S3. The

relative expression for each sample was calculated using $2^{-\Delta CT}$, where ΔC_T = (C_T gene of interest – C_T internal control) (Schmittgen and Livak 2008) .

Phylogenetic analysis

Phylogenetic studies were performed using the putative RCA amino acid sequences retrieved from the annotated gene models. The complete list of the sequences used is listed in the SI. The analysis was carried out using the Maximum likelihood method in the Molecular Evolutionary Genetics Analysis version 7 (MEGA7) (Kumar et al., 2016). Alignment for tree construction was done using MUSCLE with the Gonnet scoring matrix. Reliability of the obtained trees was tested using 1000 bootstrap replicates.

Declarations

Conflict of interest statement

The authors declare no conflict of interest.

Author contributions

RN and KSG designed the research, RN and KSK performed the experiments and data analysis and wrote the initial draft of manuscript. KSK, AM and KSG critically revised and prepared the final draft of the manuscript.

Funding

This work was supported by the USDA National Institute of Food and Agriculture, Hatch (project number WNP00449) and United States Agency for International Development Feed the Future Innovation Lab-Climate Resilient Wheat (Grant number AID-OAA-A-13-00008).

Author contributions:

RN and KSG designed the research, RN and KSK performed the experiments and data analysis and wrote the initial draft of manuscript. KSK, AM and KSG critically revised and prepared the final draft of the manuscript.

Acknowledgements

We sincerely thank Dr. G.N. Mohan Kumar, Department of Horticulture, Washington State University for providing advice on protein extraction and immunoblot analyses, Sheila Fitzgerald for proofreading the

manuscript, and Dr. Gerald Edwards and Dr. Rita Giuliani, School of Biological Sciences, Washington State University for providing wild rice species and protocol for growing rice plants under control conditions.

References

1. **Allakhverdiev SI, Kreslavski VD, Klimov V V., Los DA, Carpentier R, Mohanty P**. 2008. Heat stress: An overview of molecular responses in photosynthesis. Photosynthesis Research **98**, 541–550.
2. **Ashraf M, Harris PJC**. 2013. Photosynthesis under stressful environments: An overview. Photosynthetica **51**, 163–190.
3. **Beguiristain T, Grandbastien M-A, Puigdomènech P, Casacuberta JM**. 2001. Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiology **127**, 212–221.
4. **Carmo-silva AE, Salvucci ME**. 2013. The regulatory properties of Rubisco activase differ among species and affect photosynthetic induction during light transitions. Plant physiology **161**, 1645–55.
5. **Carmo-Silva E, Scales JC, Madgwick PJ, Parry M a J**. 2015. Optimizing Rubisco and its regulation for greater resource use efficiency. Plant, cell & environment **38**, 1817–32.
6. **Chase MW, Christenhusz MJM, Fay MF, *et al.*** 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society **181**, 1–20.
7. **Crafts-Brandner SJ, Salvucci ME**. 2000. Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO2. Proceedings of the National Academy of Sciences of the United States of America **97**, 13430–5.
8. **Crafts-Brandner SJ, Salvucci ME**. 2002. Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. Plant physiology **129**, 1773–1780.
9. **Dhaliwal AK, Mohan A, Gill KS**. 2014. Comparative analysis of ABCB1 reveals novel structural and functional conservation between monocots and dicots. Frontiers in Plant Science **5**, 657.
10. **Feller U, Crafts-Brandner SJ, Salvucci ME**. 1998. Moderately High Temperatures Inhibit Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (Rubisco) Activase-Mediated Activation of Rubisco. Plant physiology **116**, 539–546.
11. **Feschotte C**. 2008. Transposable elements and the evolution of regulatory networks. Nature Reviews Genetics **9**, 397–405.
12. **Ge S, Sang T, Lu BR, Hong DY**. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proceedings of the National Academy of Sciences of the United States of America **96**, 14400–14405.
13. **Giuliani R, Koteyeva N, Voznesenskaya E, Evans MA, Cousins AB, Edwards GE**. 2013. Coordination of Leaf Photosynthesis, Transpiration, and Structural Traits in Rice and Wild Relatives (Genus Oryza). Plant physiology **162**, 1632–51.

14. **Guo Y-L, Ge S**. 2005. Molecular phylogeny of *Oryzeae* (*Poaceae*) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. American journal of botany **92**, 1548–58.
15. **Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, Paszkowski J**. 2011. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature **472**, 115–119.
16. **Ito H, Yoshida T, Tsukahara S, Kawabe A**. 2013. Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae. Gene **518**, 256–261.
17. **Jacquemin J, Bhatia D, Singh K, Wing RA**. 2013. The International Oryza Map Alignment Project: Development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. Current Opinion in Plant Biology **16**, 147–156.
18. **Kellogg EA**. 2001. Evolutionary history of the grasses. Plant physiology **125**, 1198–1205.
19. **Kobza J, Edwards GE**. 1987. Influences of leaf temperature on photosynthetic carbon metabolism in wheat. Plant physiology **83**, 69–74.
20. **Kumar A, Li C, Portis A**. 2009. Arabidopsis thaliana expressing a thermostable chimeric Rubisco activase exhibits enhanced growth and higher rates of photosynthesis at moderately high temperatures. Photosynthesis research **100**, 143–53.
21. **Kumar S, Stecher G, Tamura K**. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution **33**, 1870–1874.
22. **Kurek I, Chang TK, Bertain SM, Madrigal A, Liu L, Lassner MW, Zhu G**. 2007. Enhanced Thermostability of Arabidopsis Rubisco activase improves photosynthesis and growth rates under moderate heat stress. The Plant cell **19**, 3230–41.
23. **Law RD, Crafts-Brandner SJ**. 1999. Inhibition and acclimation of photosynthesis to heat stress is closely correlated with activation of ribulose-1,5-bisphosphate Carboxylase/Oxygenase. Plant physiology **120**, 173–82.
24. **Law RD, Crafts-Brandner SJ**. 2001. High temperature stress increases the expression of wheat leaf ribulose-1,5-bisphosphate carboxylase/oxygenase activase protein. Archives of biochemistry and biophysics **386**, 261–7.
25. **Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S**. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic acids research **30**, 325–327.
26. **Lisch D**. 2013. How important are transposons for plant evolution? Nature Reviews Genetics **14**, 49–61.
27. **Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM**. 2015. Transposable elements contribute to activation of maize genes in response to abiotic stress. PLoS Genetics **11**.
28. **Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, Kilian B**. 2014. Sequencing of chloroplast genomes from wheat, barley, rye and their relatives provides a detailed insight into the evolution of the triticeae tribe. PLoS ONE **9**.

29. **Morales A, Ortega-Delgado ML, Molina-Galan J, de Jimenez ES**. 1999. Importance of Rubisco activase in maize productivity based on mass selection procedure. Journal of Experimental Botany **50**, 823–829.
30. **Nagai T, Makino A**. 2009. Differences between rice and wheat in temperature responses of photosynthesis and plant growth. Plant & cell physiology **50**, 744–55.
31. **Nagarajan R, Gill KS**. 2018. Evolution of Rubisco activase gene in plants. Plant Molecular Biology **96**, 69–87.
32. **Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR**. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature **461**, 1130–1134.
33. **Portis AR, Salvucci ME, Ogren WL**. 1986. Activation of Ribulose bisphosphate Carboxylase/Oxygenase at physiological CO2 and Ribulosebisphosphate concentrations by Rubisco activase. Plant physiology **82**, 967–971.
34. **Ristic Z, Momcilovic I, Bukovnik U, Prasad PVV, Fu J, Deridder BP, Elthon TE, Mladenov N**. 2009. Rubisco activase and wheat productivity under heat-stress conditions. Journal of experimental botany **60**, 4003–14.
35. **Rundle S, Zielinski R**. 1991. Organization and expression of two tandemly oriented genes encoding ribulosebisphosphate carboxylase/oxygenase activase in barley. The Journal of biological chemistry **266**, 4677–4685.
36. **Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C**. 2008. Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. The plant cell **20**, 11–24.
37. **Salvucci ME, van de Loo FJ, Stecher D**. 2003. Two isoforms of Rubisco activase in cotton, the products of separate genes not alternative splicing. Planta **216**, 736–44.
38. **Salvucci ME, Portis AR, Ogren WL**. 1985. A soluble chloroplast protein catalyzes ribulosebisphosphate carboxylase/oxygenase activation in vivo. Photosynthesis research **7**, 193–201.
39. **Scafaro AP, Bautsoens N, den Boer B, Van Rie J, Gallé A**. 2019. A Conserved Sequence from Heat-Adapted Species Improves Rubisco Activase Thermostability in Wheat. Plant physiology **181**, 43–54.
40. **Scafaro AP, Gallé A, Van Rie J, Carmo-Silva E, Salvucci ME, Atwell BJ**. 2016. Heat tolerance in a wild Oryza species is attributed to maintenance of Rubisco activation by a thermally stable Rubisco activase ortholog. New Phytologist **211**, 899–911.
41. **Scafaro AP, Haynes PA, Atwell BJ**. 2010. Physiological and molecular changes in Oryza meridionalis Ng., a heat-tolerant species of wild rice. Journal of Experimental Botany **61**, 191–202.
42. **De Souza FSJ, Franchini LF, Rubinstein M**. 2013. Exaptation of transposable elements into novel Cis-regulatory elements: Is the evidence always strong? Molecular Biology and Evolution **30**, 1239–1251.

43. **Swigoňová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J**. 2004. Close split of sorghum and maize genome progenitors. Genome Research **14**, 1916–1923.
44. **To KY, Suen DF, Chen SC**. 1999. Molecular characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase activase in rice leaves. Planta **209**, 66–76.
45. **Vaughan DA, Morishima H, Kadowaki K**. 2003. Diversity in the Oryza genus. Current Opinion in Plant Biology **6**, 139–146.
46. **Vogel JP, Garvin DF, Mockler TC, *et al.*** 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature **463**, 763–768.
47. **Wang D, Li X-F, Zhou Z-J, Feng X-P, Yang W-J, Jiang D-A**. 2010. Two Rubisco activase isoforms may play different roles in photosynthetic heat acclimation in the rice plant. Physiologia plantarum **139**, 55–67.
48. **Werneke JM, Chatfield JM, Ogren WL**. 1989. Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and Arabidopsis. The Plant cell **1**, 815–825.
49. **Yamori W, von Caemmerer S**. 2011. Quantification of Rubisco activase content in leaf extracts. Methods in molecular biology (Clifton, N.J.) **684**, 383–391.
50. **Yamori W, Suzuki K, Noguchi K, Nakai M, Terashima I**. 2006. Effects of Rubisco kinetics and Rubisco activation state on the temperature dependence of the photosynthetic rate in spinach leaves from contrasting growth temperatures. Plant, Cell and Environment **29**, 1659–1670.
51. **Yin Z, Meng F, Song H, Wang X, Xu X, Yu D**. 2010. Expression quantitative trait loci analysis of two genes encoding rubisco activase in soybean. Plant physiology **152**, 1625–37.
52. **Yin Z, Zhang Z, Deng D, Chao M, Gao Q, Wang Y, Yang Z, Bian Y, Hao D, Xu C**. 2014. Characterization of Rubisco Activase Genes in Maize: An α-Isoform Gene Functions alongside a β-Isoform Gene. Plant physiology **164**, 2096–106.
53. **Zhang N, Kallis RP, Ewy RG, Portis AR**. 2002. Light modulation of Rubisco in Arabidopsis requires a capacity for redox regulation of the larger Rubisco activase isoform. Proceedings of the National Academy of Sciences of the United States of America **99**, 3330–3334.
54. **Zhang Z, Komatsu S**. 2000. Molecular cloning and characterization of cDNAs encoding two isoforms of ribulose-1,5-bisphosphate carboxylase/oxygenase activase in rice (Oryza sativa L.). Journal of Biochemistry **128**, 383–389.
55. **Zhang N, Portis AR**. 1999. Mechanism of light regulation of Rubisco: a specific role for the larger Rubisco activase isoform involving reductive activation by thioredoxin-f. Proceedings of the National Academy of Sciences of the United States of America **96**, 9438–43.
56. **Zhang N, Schürmann P, Portis AR**. 2001. Characterization of the regulatory function of the 46-kDa isoform of Rubisco activase from Arabidopsis. Photosynthesis Research **68**, 29–37.
57. **Zhao C, Liu B, Piao S, *et al.*** 2017. Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of Sciences **114**, 9326–9331.

Tables

Table 1: Structure and expression of RCA coding genes of distantly related monocot species.

Gene name	Exon (E) and intron (I) sizes (bp)[a]													Protein[b]	Isoform[c]
	E1	I1	E2	I2	E3	I3	E4	I4	E5	I5	E6	I6	E7	(aa)	(kDa)
Musaceae															
M. acuminata															
MaRca1	45	108	320	98	106	73	86	876	473	96	282	94	116	475	46.03
MaRca2	45	90	320	82	106	83	86	71	473	80	287			438	42.17
MaRca3	45	78	320	82	106	79	86	89	473	81	281			438	42.22
Arecoideae															
P. dactylifera															
PdRca1-α	45	91	320	104	106	90	86	1213	473	142	282	163	116	475	46.37
PdRca1-β	45	91	320	104	106	90	86	1213	473	142	287			438	42.27
PdRca2-α	45	566	320	118	106	117	86	804	473	625	282	148	119	476	46.64
Bromelioideae															
A. comosus															
AcRca1	45	1246	323	242	106	91	86	225	473	432	282	281	116	478	46.27
AcRca2	48	823	320	111	109	94	86	626	473	174	296			443	42.67
AcRca3	45	255	320	88	106	115	86	493	473	261	287			438	42.53

Note:

a. E and I followed by a number in the column head represent corresponding exon and intron numbers of the gene

b. Total number of amino acids including cTP.

c. Molecular weight of the mature protein (without cTP) were predicted by ChloroP 1.1(Emanuelsson et al. 1999).

Table 2. List of transposable elements identified in the grass *Rca* genic region

Gene	TE type	Size (bp)	Region	E value	No. of HSEs (sequence)
ObRca1	Mutator-like (MULE)	336	Promoter	0	2 (aaaaaattta; caaatttttt)
OsRca2	Mutator-like (MULE)	347	Promoter	0	1 (aaaaaatttg)
OsRca2	MuDR-like (MITE)	320	Intron (5)	0	1 (gtaatttttt)
TaRca1-A	Gypsy (LTR)	319	Promoter	3e-43	0
TaRca1-B	Mariner (MITE)	239	Promoter	1e-18	0
TaRca1-D	Copia (LTR)	967	Promoter	0	0
TaRca2-D	Mariner (MITE)	162	Promoter	5e-21	0
ZmRca1	Class II/III TE	368	Promoter	0	1 (agataattcg)
SbRca1	CACTA-consensus	584	Promoter	2e-97	1 (acaaaatttc)

Figures

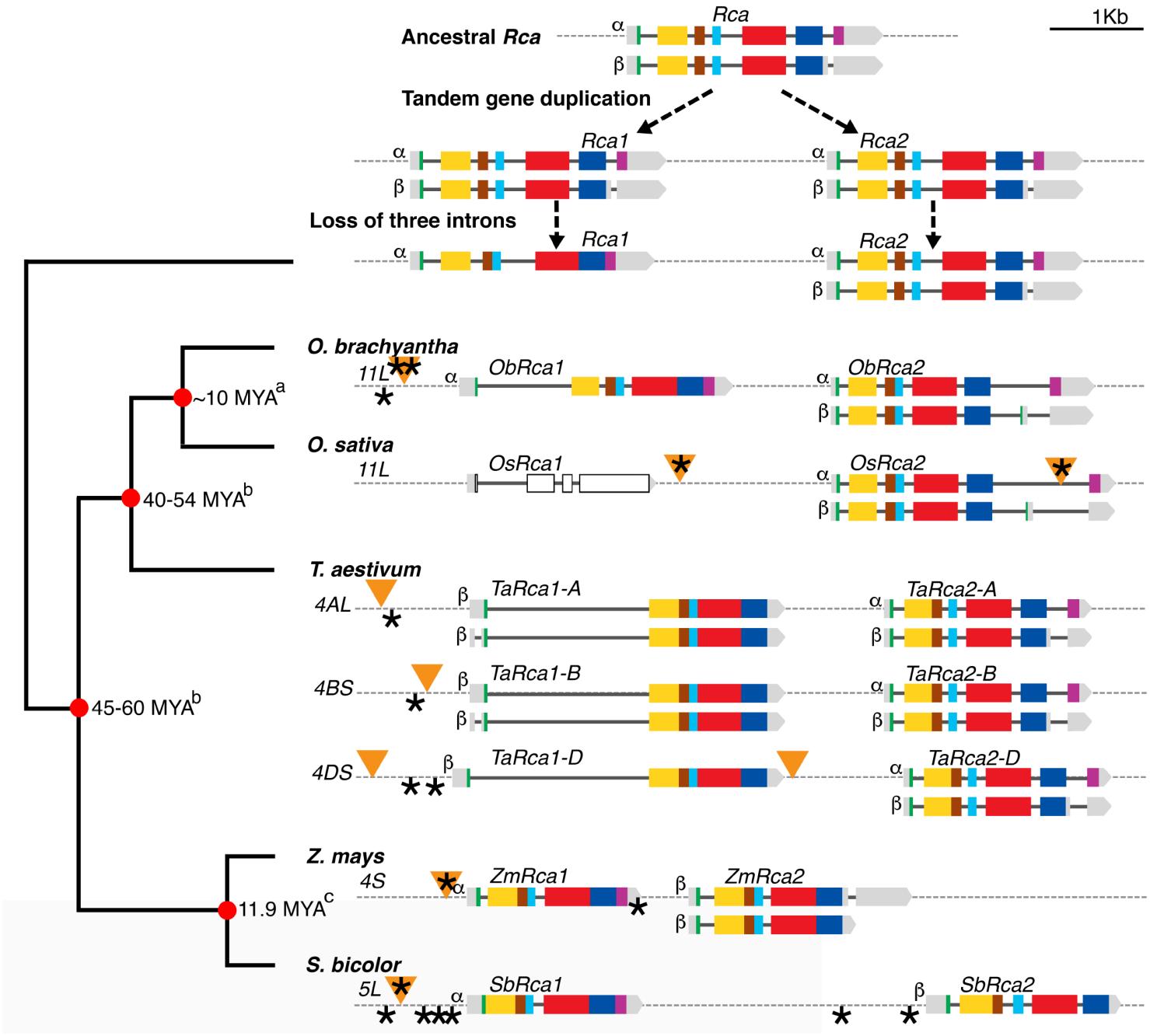

Figure 1

Evolution of Rubisco activase gene in cereals. Comparison of *Rca* gene models from cereals showing structural differences, accumulation of heat shock elements and insertion various transposable elements with an evolutionarily predicted timeline. The *Rca* gene evolution model for grasses is adapted from Nagarajan and Gill (2017). The color bars on and below the gray dotted lines represent gene models and its corresponding alternate mRNA splicing models. Species with multiple gray lines represent RCA coding genes on different chromosomes. The a and bsymbols in the gene models denote the type of isoforms coded from the corresponding gene copy. In gene models, the thin black bar represents introns and thick light gray bars represent UTR regions. The green, yellow, brown, sky blue,

red, navy blue and purple color bars represent exons from 1 to 7 of the ancestral *Rca*gene, respectively. Bars with multiple colors represent fused exons that lost introns in comparison to the ancestral *Rca* gene model. The exons of non-functional rice *OsRca1* is shown as white bars. Inverted triangles in orange depict transposable element insertions and asterisk symbol (*) below the gray dotted lines represent heat shock elements in the promoter region. The asterisk symbol on the inverted orange triangles show the presence of heat shock elements in the transposable element sequences. The gene model at the top indicates the putative common grass ancestor *Rca* gene and predicted evolutionary changes based on *Rca*genes of different cereal crops. The evolutionary timescale on species divergence presented on the left side of gene models were based on previous estimates - “a” (Guo and Ge, 2005), “b” (Vogel *et al.*, 2010), (Middleton *et al.*, 2014), “c” (Swigoňová *et al.*, 2004). The gene models were drawn to scale.

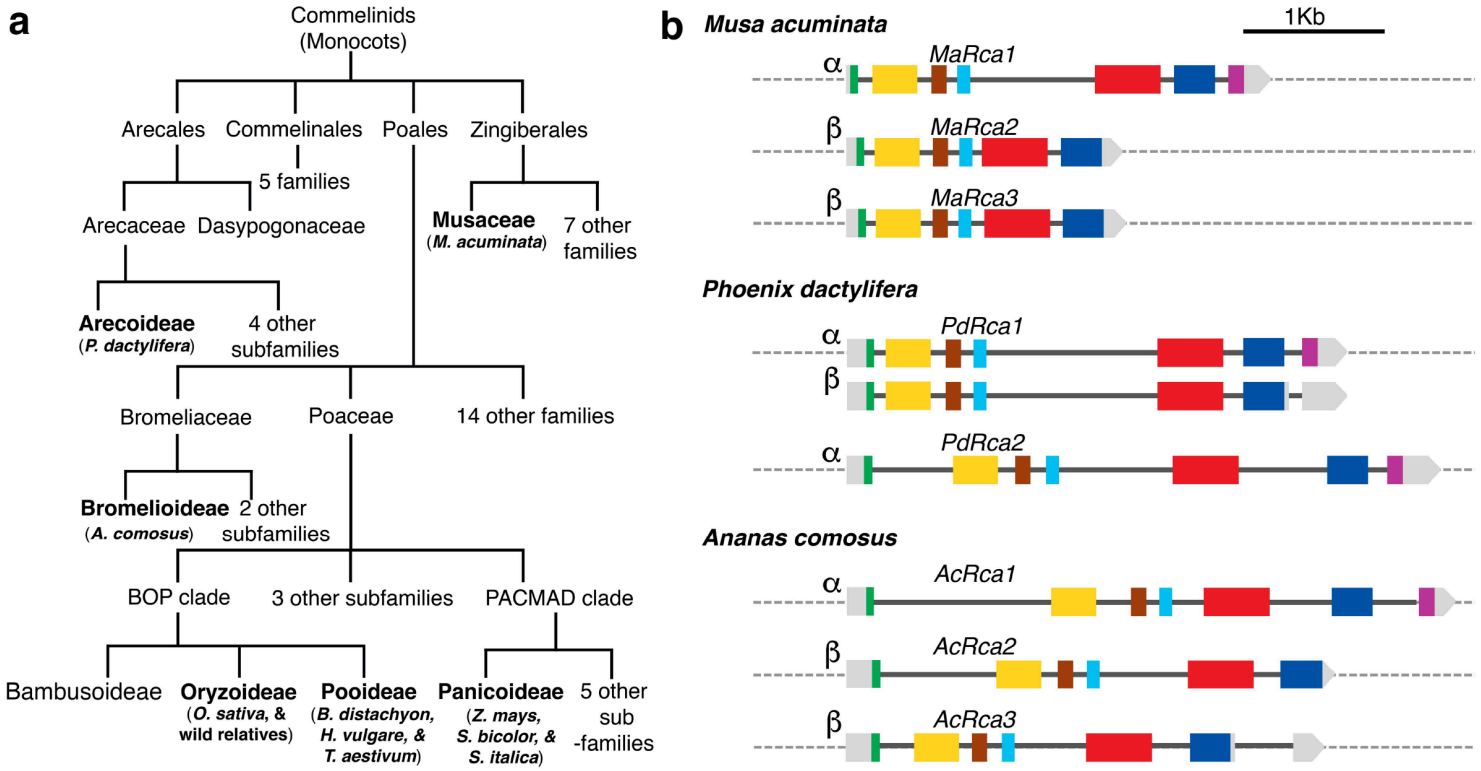

Figure 2

Structure of Rubisco activase genes in different monocot species. a. Taxonomical classification showing different monocot species (bold letters) of Commelinids clade used in this study. Angiosperm Phylogeny Group IV (APG-IV) system (Chase *et al.*, 2016) was followed to show the classification. **b.** RCA coding gene models of banana (*Musa acuminata*), date palm (*Phoenix dactylifera*), and pineapple (*Ananas comosus*) were drawn based on the comparative analysis of genomic and cDNA sequences (see materials and methods). The color bars on and below the gray dotted lines represent gene models and its corresponding alternate mRNA splicing models. The aand bin front of each gene model denote the type of isoform coded from that gene model. Species with two or more gray dotted lines represent RCA coding genes on different genomic locations or chromosomes. In gene models, the thin black bars

represent introns and light gray bars represent UTR regions. The green, yellow, brown, sky blue, red, navy blue and purple color bars represent exons 1 to 7 respectively. The gene models were drawn to scale.

Fig. 3

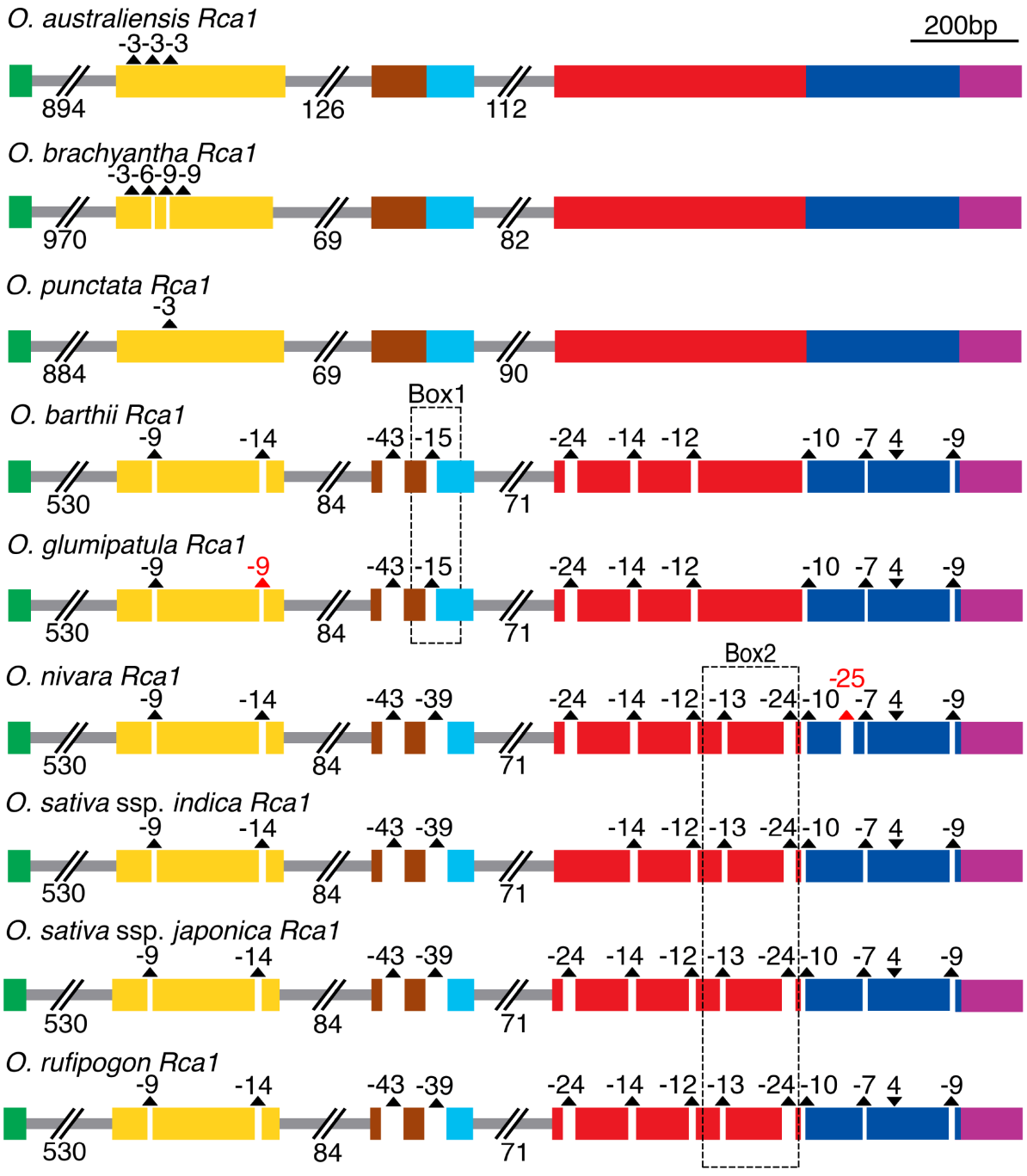

Figure 3

Comparison of sequence deletions between the *Rca1* gene of Oryza species. Comparative analysis showing nucleotide deletions and insertions in the *Rca1* gene of different Oryza species. In gene models, green, yellow, brown, sky blue, red, navy blue and purple color bars represent corresponding exonic

regions from the ancestral *Rca* gene exon shown in Fig.1. Grey bars represent introns and their sizes (in basepairs) are mentioned below the double slash (//) of each gray bar. Upward black triangles above the gaps show nucleotide deletions and downward black triangles show insertions, corresponding to the *Rca1* sequence of *O. brachyantha*. Red color triangles represent unique deletions or insertions identified in that particular species. The numbers above the triangles represent the nucleotide count on that particular addition or deletion (-) site. Box1 and Box2 represent unique deletions found in the box overlapping species.

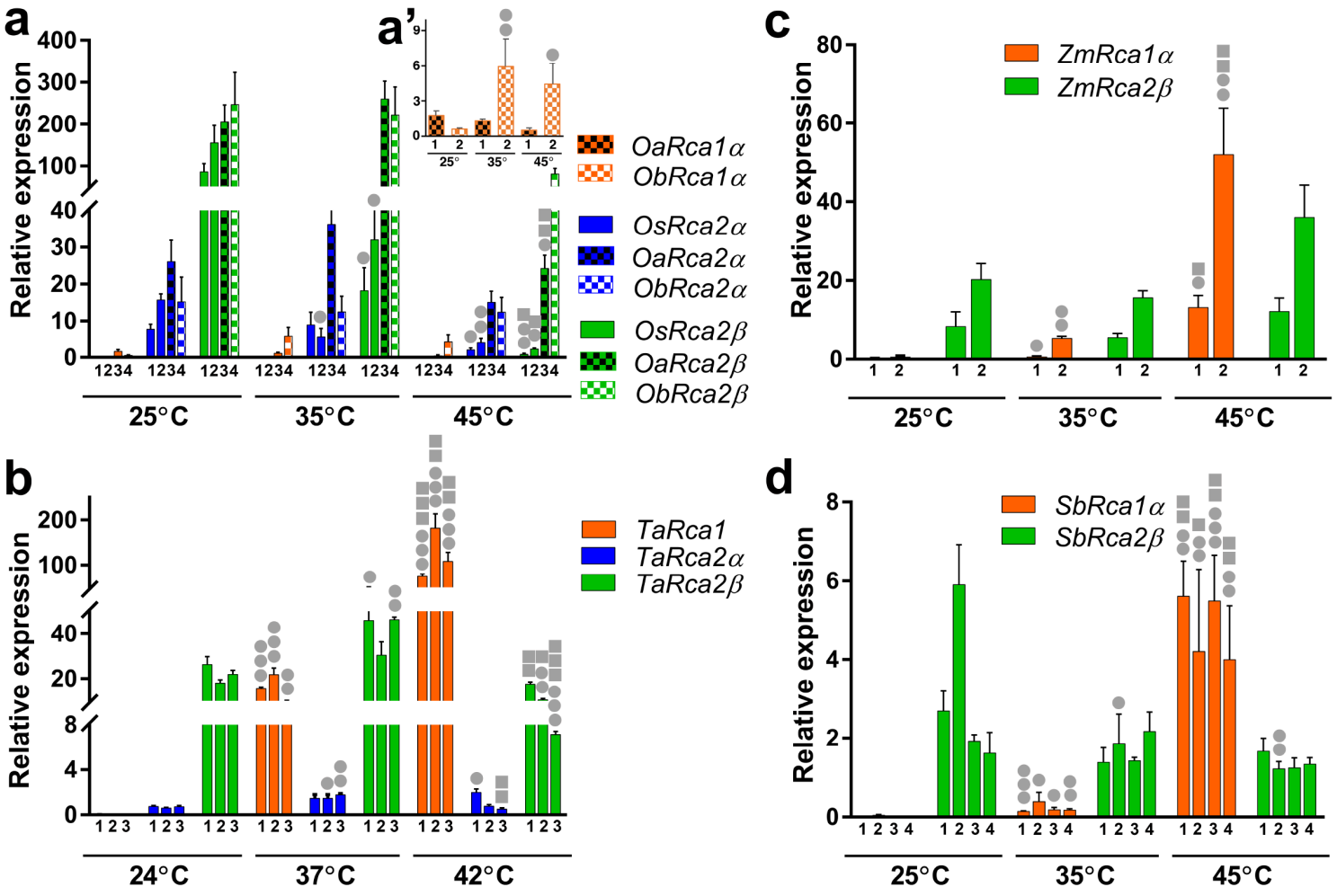

Figure 4

Expression pattern of various *Rca* transcripts in cereals. Quantitative Real-Time PCR (qRT-PCR) analysis showing the transcript levels of tandemly duplicated *Rca* genes of different genotypes of rice, wheat, maize and sorghum at normal (24/25°C), moderate (35/37°C) and severe heat stress (42/45°C) conditions. The *Rca* transcripts of rice, wild rice, maize, and sorghum were quantified using gene and alternate splicing specific primers. In case of wheat, the primers were specific to *Rca1*, *Rca2* and its alternatively spliced transcripts but not to the A, B and D homoeologous copies. The columns on the chart indicate the mean relative expression of *Rca* transcripts and the number on the x- axis represent different genotypes. Data represent the means and standard errors (SE) of three biological replicates (n=3). A complete list of primers is shown in Supplemental table S3. **a**. Effect of heat stress on the *Rca*

genes of two rice cultivars Nipponbare and N22 (column #1 and 2) and two wild species *O. australiensis*, and *O. brachyantha*, (column #3 and 4). **b**. Response of wheat *TaRca1* and *TaRca2* expression to high temperature stress in cultivar Chinese Spring (column #1), Giza168 (column #2) and PBW343 (column #3). **c.** Expression pattern of *ZmRca1* and *ZmRca1* at normal and elevated temperatures in maize inbred lines B73 (column #1) and Mo22 (column #2). **d.** Effect of temperature stress on the expression of *SbRca1* and *SbRca2* of B35 (column #1), BTx623 (column #2), SC1019 (column #3), and SC1047 (column #4).

Fig. 5

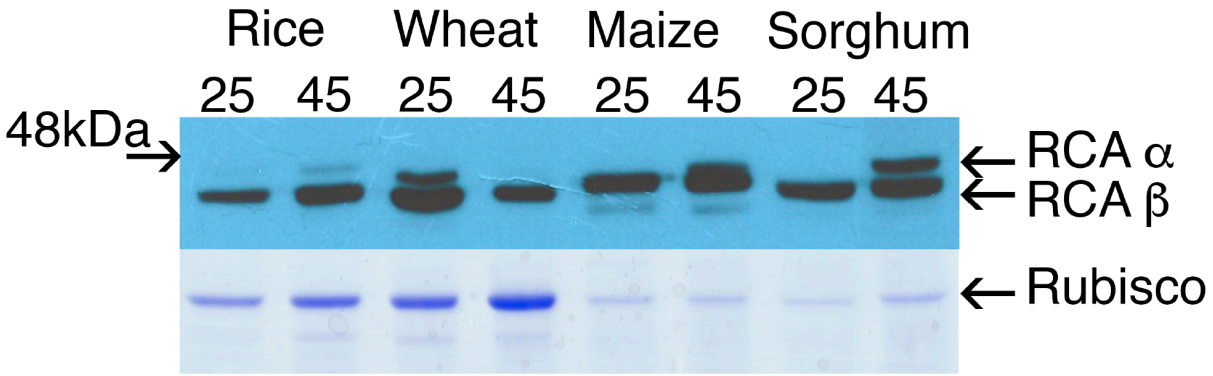

Figure 5

Immunoblot analysis showing expression pattern of cereal RCA isoforms at elevated temperatures. Proteins were extracted from control (25°C) and heat stressed (45°C) leaf samples of four weeks old rice (cv. Nipponbare), wheat (cv. Chinese spring), maize (B73 inbred) and sorghum (BTx623 inbred) lines. Plants grown under control conditions (25°C day/21°C night) were gradually exposed (5°C / 30 min) to heat stress and maintained at 45°C (45°C) for four hours. Leaf samples from control and heat-treated plants were collected at noon (six hours after the chamber lights were on) to avoid any effect of circadian rhythm on the gene expression. Samples were frozen immediately in liquid nitrogen and the total protein was extracted (see material and methods for details). An equal amount of total protein (2 mg) was loaded in 1D-SDS-PAGE for immunoblot analysis. Anti-RCA specific polyclonal antibody raised in rabbits (Agrisera, Catalog # AS10-700) was used as primary to detect RCA isoforms. The Coomassie Brilliant Blue R250 stained gel below the immunoblot shows the Rubisco large subunit.

Fig. 6

```
ZmRCA2          221 PRVPIIVTGNDFSTLYAPLIRDGRMDKFYWAPTREDRVGVCKGIFRSDGVPDEDVVRLVD 280
SbRCA2          223 PRVPIIVTGNDFSTLYAPLIRDGRMDKFYWAPTRDDRIGVCKGIFRTDGVPDEHVVQLVD 282
AtRCA1          218 ARVPIICTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDKIKDEDIVTLVD 277
AtRCA1-183H12   218 ARVPIICTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDKIKDEDIVRLVD 277
ZmRCA1          220 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDGVDEEHVVQLVD 279
SbRCA1          221 ARVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDGVDEEHVVQLVD 280
ObRCA1          223 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVAKGIFRTDKVPDDAIVKLVD 282
TaRCA1-A        222 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDNVPDEAVVRLVD 281
TaRCA1-D        222 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDNVPDEAVVRLVD 281
TaRCA1-B        222 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTREDRIGVCKGIFRTDNVPDEAVVRLVD 281
TaRCA2-B        219 PRVPIVVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFQTDNVSDESVVKIVD 278
TaRCA2-A        219 PRVPIVVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFQTDNVSDESVVKIVD 278
TaRCA2-D        219 PRVPIVVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRIGVCKGIFQTDNVSDESVVKIVD 278
OsRCA2          220 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRVGVCKGIFRTDNVPDEDIVKIVD 279
ObRCA2          220 PRVPIIVTGNDFSTLYAPLIRDGRMEKFYWAPTRDDRVGVCKGIFRTDNVPEDDIVKVVD 279
                     ****: ******************:********:**:**.****::* : :: :*  :**
```

Figure 6

Comparison of grass RCA sequences with wild and thermostable variants of Arabidopsis RCA. Multiple sequence alignment showing a part of mature cereal RCA isoform (excluding chloroplast transit peptide) sequences with Arabidopsis wild type (AtRCA1) and thermostable RCA variant (AtRCA1-183H12). The amino acid substitution T274R, which provided thermostability of RCA in AtRCA1-183H12 variant, and its corresponding amino acid in cereal RCA isoforms are shown in the box. The exact Arginine (R) amino acid is marked in bold and underlined.

Fig. 7

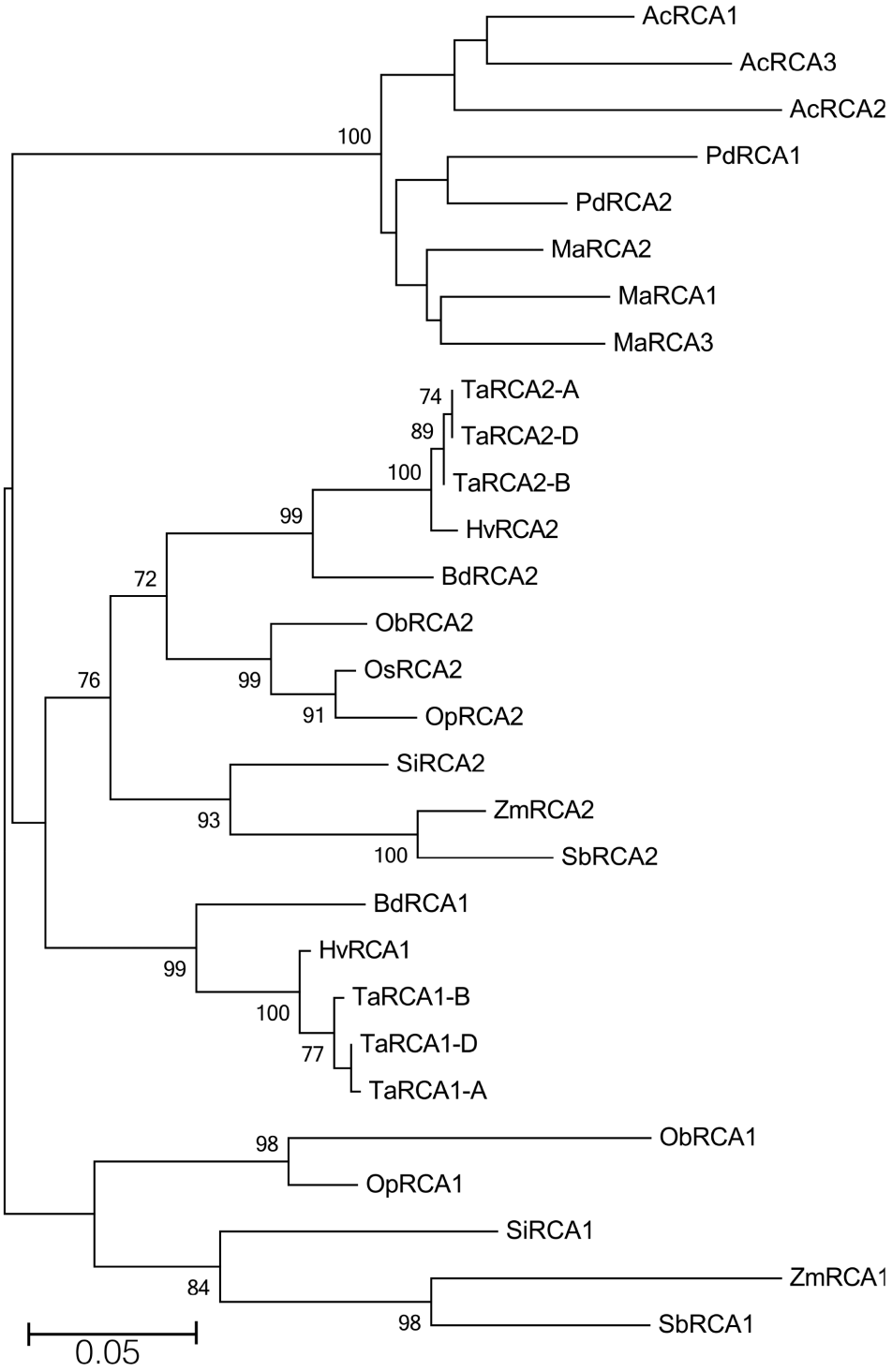

Figure 7

Maximum Likelihood phylogenetic tree showing relationship between different monocot RCA proteins. Monocot species used for the analysis are *Ananas comosus* (Ac), *Phoenix dactylifera* (Pd), *Musa acuminata* (Ma), *Triticum aestivum* (Ta), *Hordeum vulgare* (Hv), *Brachypodium distachyon* (Bd), *Oryza sativa* (Os), *Oryza brachyantha* (Ob), *Oryza puctata* (Op), *Setaria italica*(Si), *Zea mays* (Zm), and *Sorghum*

bicolor (Sb). Scale bar represents the number of amino acid changes per site and the numbers at the nodes represent bootstrap values larger than 70%.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx
- TableS2.docx
- TableS3.docx