SSRG evolution in the genus Oryza

Karine Elise Janner de Freitas
Universidade Federal de Pelotas
Railson Schreinert dos Santos
IFFarroupilha
Carlos Busanello
Universidade Federal de Pelotas
Filipe de Carvalho Victoria
Universidade Federal do Pampa - Campus Sao Gabriel
Jennifer Luz Lopes
Universidade Federal de Pelotas
Rod A. Wing
University of Arizona
Antonio Costa De Oliveira (acostol@terra.com.br)
Universidade Federal de Pelotas https://orcid.org/0000-0001-8835-8071

Short communication

Keywords: Leersia perrieri, phylogeny, starch synthesis, cooking quality.

DOI: https://doi.org/10.21203/rs.3.rs-45627/v1

Abstract

Background: Cooking quality is an important attribute in Common/Asian rice (*Oryza sativa* L.) varieties, being highly dependent on grain starch composition. This composition is known to be highly dependent on a cultivar's genetics, but the way in which their genes express different phenotypes is not well understood. Further analysis of variation of grain quality genes using new information obtained from the wild relatives of rice should provide important insights into the evolution and potential use of these genetic resources.

Findings: The analysis of the protein sequences of grain quality genes across the *Oryza* suggest that the deletion/mutation of amino acids in active sites result in variations that can negatively affect specific steps of starch biosynthesis in the endosperm. As observed in *O. sativa* subsp. *japonica*, the lower amylose content is probably related to the absence of a C-terminal domain in PUL, characterizing what we know as *japonica* genotypes. On the other hand, the complete deletion of some genes in the wild species do not affect the amylose content, as observed in the absence of GBSSII in starch biosynthesis of *O. meridionalis*, SSIV2 in *O. glaberrima* and *DPE1* in *O. brachyantha* and *O. nivara* in which such modifications seem not to affect the final endosperm starch composition.

Conclusion: Here we present new insights for obtaining new starch-specific rice phenotypes, considering structural protein features that include both the absence and duplication of copies, once again denoting that *Oryza* species are a rich source of variability for use in plant breeding.

Findings

Common rice (*Oryza sativa* L.) is a food of great importance worldwide, especially in Asian countries, where it is an important part of local culture. Being widely consumed and having different forms of preparation makes "quality" something different in each country around the world. Nevertheless, no matter what grain quality means, its demand is increasingly becoming a priority for international export markets worldwide.

Today cooking behavior has become one of the most important research components in several rice breeding programs where characteristics such as amylose content (AC) and gelatinization temperature (GT), which have major effects on cooking quality (CQ) and consumption, are controlled by physicochemical properties of starch in rice grain endosperm (Pandey et al., 2012).

The ratio of amylose to amylopectin as well as the structure of amylopectin itself can vary greatly between different rice genotypes (Yu et al., 2011). Generally, grains with higher amylose content present a harder non-sticky texture after cooking, being preferred in several countries. Such a feature is usually evaluated during grain development in different cultivars (Walter et al., 2008). However, the genetic events that lead to this type of grain are not well understood and genotypes that deliver such grains are not easily obtained. That is the reason why it is so important to understand the behavior of grain-quality-related genes, which enable more efficient and precise breeding applications.

The 27 known *Oryza* species span over 15 million years of evolution which we can take advantage of, since it constitutes a rich source of genetic variation. Though a better understanding of the genomic differences between these species is essential for such a purpose, the recent publication of the genomes of 13 rice species has opened the door to a series of new studies that make it possible to enrich the germplasm that can be used for breeding (Santos et al., 2017; Stein et al., 2018). The possibility of using these wild species to improve grain quality should also be considered, but what would be the first genes to start such an analysis?

Considering the importance of Starch Synthesis-Related Genes (SSRGs) in the control of CQ and the limited exploration of the information recently made available to the scientific community on *Oryza* genomes, an evolutionary analysis is needed to reveal the role of adaptive mechanisms before and after rice domestication. It will thus help to understand the complexity of the evolution of enzymes involved in the starch synthesis pathways, and further provide the basis for approaches that can generate new phenotypes trough new strategies to modify starch synthesis. We therefore selected a set of SSRGs according Zeng et al. (2017) to explore their evolution across the genus *Oryza*. The Additional file 1: Table S1 present all SSRGs identified in 11 *Oryza* species and *Leersia perrieri.*

Agpase Subunits

The phylogenetic tree of the AGPase genes, which include both large (AGPL1, AGPL3, AGPL4) and small (AGPS2a) subunits, identified 48 genes across the *Oryza* and the outgroup (*L. perrieri*), which revealed 3 different groups (Fig. 1 and Additional file 2: Figure S1).

The first group is formed by the *AGPS2a* genes with both large exon and intron structures. This group is believed to have the highest similarity to the ancestor of every *Oryza* AGPase gene. *O. meridionalis* (AA) also contains the same large exon structure in the second group formed by *AGPL4* and likewise, in the third clade which is a mixture formed by *AGPL3* and *AGPL1*, respectively.

The evolution of large and small ADP proteins subunits in *Oryza* was markedly different, probably due to different rates of selection pressure that were denoting diversification in AGPS2a (Figure S1B). Recombination analysis based in the alignment do not show any evidence of recombination in the AGP partition. On the other hand, positive selective pressure (dN/dS > 1) was detected in the sequence alignments, suggesting evidences of diversifying selection (Fig. S3). Small subunits were under higher purifying selection than the large subunits, being thus responsible for most of the diversity of AGPase gene and allies (Batra et al, 2017; Georgelis et al., 2007). These would be the regions that concentrate most of the positive selection, since they also demonstrate most of the variability. One explanation for this would be that in large subunits most duplications occur when compared to small subunits of AGPase genes (Georgellis et al., 2008). However, NHR cannot yet be ruled out, since our data are in accordance with previous reports that indicate that NHR are probably more frequent than MEI in *Oryza* species (Bai et al., 2016). On the other hand, contrasting evolutionary patterns are expected among paralogues, and in AGPase, some duplications have been accompanied by a change of cellular

compartmentalization (e.g. from plastid to cytosolic) or changes in expression with subsequent modification of regulation properties (Corbi et al., 2012).

Regarding to gene position, all *Oryza* species have AGPS2a positioned in Chr 8, of note, we observed that *OMERAGPS2A* and *ONIVAGPS2A* are located on two different chromosomes (i.e. Chr 9 and Chr 4, respectively (Fig. 1)), being necessary investigate what led to these genes change their position or what mechanism was involved in that. Possible differential Mobile Element Insertion (MEI) events related to these loci was investigated, a region of 50 kb up- and downstream of these genes were aligned, showing high similarity between *OMERAGPS2A* (AA) to AGPS2a of other species, which means that this change in position probably did not occur through TE insertion (Fig. S2).

On the other hand, Non-Homologous Recombination (NHR) is likely to have occurred in this region, placing this large block (upstream + gene + downstream) on Chr 9. The locus from *O. nivara* in Chr 4 has only a small ortholog block that corresponds to the end of the upstream region and the start of the downstream region. Small up- and downstream fragments similar to specific LTR-TEs were found using the Rice Transposable Elements database (RiTE-db), but it is unlikely that these are responsible for a translocation event. As previously reported, the most frequent events responsible for changing copy number variations and gene position to other chromosomes are mediated by either transposons, through MEI, or NHR, for both *Oryza* and *Arabidopsis* (Bai et al., 2016; Freeling et al., 2018).

In *Oryza* and other plants, the *AGPase* protein subunit is characterized by a core region that is important for catalytic activity, called the nucleotidyl transferase domain (NTP_transferase) that is important in providing the substrate for starch biosynthesis. The conserved motifs of the four analyzed *ADP-glucose pyrophosphorylase* subunits form a signature pattern, revealing that motifs 9 and 10 are not detectable in the NTP_transferase domain of *AGPS2a*; the same occurs for motifs 8 and 9 in *AGPL4*; 7, 8 and 9 in *AGPL1*; and 3, 6, 8 and 9 in *AGPL3* which are not found in some *Oryza* species. The absence of specific motifs can affect the endosperm starch synthesis limiting the reaction converting Glucose 1-Phosphate (Glc-1-P) and Adenosine triphosphate (ATP) to ADP-glucose and inorganic pyrophosphate (PPi) in amyloplasts, directly reflecting the control of carbon flux into the starch accumulation pathway, consequently causing a shrunken endosperm in rice (Smith et al. 1997; Pandey et al., 2012; Qu et al., 2018).

Starch Synthesis (ss) Genes

A total of 92 protein coding SS genes were found across the 12 genome data set, while its phylogenetic analysis allowed the identification of nine different clades based on sequence similarity. Clades I, II, III, IV, V, VI, VII, VIII, and IX typically represent SSIV1, SSI, Waxy, SSIII1, SSIII2, GBSSII/ALK, SSII2, SSII1 and SSIV2, respectively (Fig. 1 and Additional file 5: Figure S4).

The phylogenetic analysis showed that in most *Oryza* species, SS isoforms have undergone different degrees of gene duplication, something that is also observed in most plant species. *Oryza* clades I, IV, V, IX possess a different genetic origin from clades II, III, VI, VII and VIII and, since paralogous genes tend to

slowly accumulate variations over time, it is easy to notice a large variation when we compare SS genes between these two clades (Patron and Killing, 2005; Deschamps et al., 2008; Ball et al., 2011; Guo et al., 2019). The distinct spatial pattern of starch deposition within a caryopsis, which is also related to differences in the temporal expression pattern between early (SSIII1, SSII2, GBSSII) and late (ALK, SSIII2, Waxy) expressed genes (Hirose and Terao, 2004), is probably the result of variations accumulated over time. Overall, the phylogenetic tree analysis reveals a highly conserved structure for both gene and amino acid sequences, suggesting a strong evolutionary relationship between species in each SS.

Some genes that have long exons near the 5' or 3' UTRs, as observed in few SS proteins of *L. perrieri, O. longistaminata*, *O. brachyantha* and *O. meridionalis*, seem to be ancestors of other species SSs. *OMERSSIV1_2D* (Clade I) is a probable result of a sub-functionalization since it does not contain motifs 4, 7, and 8, that represent the catalytic domain of starch synthase (Glyco_transf_5) and (Glyco_transf_1).

Another recent duplication was identified in the *O. meridionalis SSIII1* gene, but in this case both the original and duplicated copies look functional, containing all the motifs that are part of its characteristic domain, however the large size of *OMERSSIII1_1D* (7,844 bp longer than the original copy) is something that deserves more investigation, especially when we take into account the highly conserved profile of these genes. It is also important to notice that the same large domain occurs in duplicated copies of *SSII2* and *Waxy* in the outgroup *L. perrieri.*

Taking into account that sequence variation in SSRGs have a great influence in rice amylose content, gelatinization temperature, and amylopectin chain length (Kasem et al., 2011), although important, it is hard to understand the roles of each SS isoform in each of the characters, due to the high sequence variation among these genes. Also, it is even more complicated when we consider its diversity of genes in starch biosynthesis. The structural features of the genes and duplicated copies denote that these species are a rich source of variability that can improve starch quantity and quality, mainly through modifications of amylopectin synthesis chains B2 and B4 (Pandey et al., 2012).

Expressed specifically in the developing rice endosperm and leaves, SSIII 1 and 2 include 3 other repeated domains in addition to the starch synthase domain. An N-terminal Carbohydrate-Binding Module (CBM) domain is a contiguous amino acid sequence within a carbohydrate-active enzyme with carbohydrate-binding activity. Although no lack of protein motifs were observed that could affect the catalytic domain in SSIII, in *O. sativa* this domain synthesizes long chains, and a deficiency in SSIII1 that is the second major enzyme (Fujita et al., 2006), can indirectly enhance both the SS-I and GBSS-I gene transcripts. On the other hand, a survey of amino acid motifs of SS isoforms reveals that certain motifs are absent in certain *Oryza* species, as it is possible to notice in OsINDSSIV1, OLONSSIV1, OBARSSII2 and OMERSSII2, which are part of the two C-terminal domains. This may affect the catalytic performance of the chain-elongation reaction of α-1-4-glucosidic linkage, which can further complicate the interplay between SS, SBE and DBE (Myers et al., 2000; Nakamura, 2002).

Waxy is believed to be the main enzyme that controls high amylose content in *Oryza* species and, with GBSSII, present tissue-specific expression in a complementary manner between endosperm and non-

endosperm tissues, causing different characteristics with respect to amylose content, and branch length distribution in amylopectin (Wang et al., 2019). Thus, the differential action of these two enzymes affect the final amylose content in the endosperm. Despite this, the absence of GBSSII (Table S1) does not influence the high content of amylose in the endosperm (about 35%) of *O. meridionalis* (Mondal & Henry, 2018). Despite the evolutionary advantage that the presence of the two enzymes (Waxy and GBSSII) confer for starch biosynthesis, Waxy enzymes without GBSSII seem to be enough for high amylose accumulation in *Oryza* endosperm, something that brings new perspectives for the improvement of this complex network (Vrinten and Nakamura, 2000; Tian et al., 2009; Wang et al., 2019).

On the other hand, the loss of SSIV2 in *O. glaberrima* during evolution does not eliminate the ability of chloroplasts in producing starch granules, since features in the N-terminal extension of SSIV enable the interaction with other proteins contributing to granule initiation. In Arabidopsis, when the SSIV glucosyl transferase domain is absent, a significant reduction of starch synthesis is observed (Szydlowski et al., 2009; Zeeman et al., 2010).

Some *Oryza* species and *L. perrieri* show changes in chromosome position of the SS genes relative to *O. sativa* (Fig. 1), such as OMERSSII1 from Chr 10 to 4 (Additional file 6: Figure S5), OMERSSII2 from Chr 2 to 6, ONIVSSII2 from Chr 2 to 6 (Additional file 7: Figure S6) and OLONSSIV2 from Chr 5 to 9 (Additional file 8: Figure S7). An alignment analysis shows that for OMERSSII1 and OMERSSII2, the change did not occur through a differential TE insertion, since an analysis of 50 kb upstream and downstream of each gene shows a lack of or just partial synteny (fragments from approximately 40 Kb) between the other *Oryza* loci. In case of partial sinteny, a significant presence of TEs in this region was not identified using the with RiTE-DB. Interestingly, OMERSSII2 contained an inverted region of 50 kb that denotes an unusual rearrangement by translocation and inversion of blocks up- and downstream of the gene (Additional file 6: Figure S5, Additional file 7: Figure S6).

Interestingly, recombination events were found both for ALK and Waxy gene copies (Additional file 9: Figure S8, Additional file 10: Figure S9), being the strongest evidence for the last gene family. However, the same was not observed for the other SS copies, where 117 were found to be under positive selection (Additional file 11: Figure S10) with no recombination events were detected. This agrees with previous reports, in which the diversification in these genes was suggested to be driven by a large number of duplication events instead of recombination events (Nougué et al, 2014).

Debranching Enzymes (dbe)

The DBE genes are classified as DPE1 (*Disproportionating enzyme*), PUL (*Pullulanase*) and ISA (*Isoamylase*). In total 35 genes DBE were identified in *Oryza* and *L. perrieri* 12 genome data set (Fig. 1 and Additional file 12: Figure S11). The phylogenetic analysis showed that the DBE proteins can be grouped in two clades, one that comprises DPE1 (Group I) and the other consisting of a mixed group composed of PUL and ISA genes (Group II).

DPE1, despite forming a conserved clade, presents some variations in its two subgroups. First, *O. meridionalis* (AA) shows the longest gene structure, with more than eight exons, being the longest in its 5' UTR, something that contrasts with the usual short structure of DPE1 genes. Despite that, *OMERDPE1*, *OBARDPE1* and *OLONDPE1* lack motif 9, which is part of glycoside hydrolase family 77 domain (Glico_transf_77), a domain responsible for cleaving the starch granule into smaller glucan molecules. Additionally, this protein is part of Group 1, which comprise enzymes that act in the initial phase of endosperm development (Tian et al., 2009), playing an important role in grain quality improvement programs (Zeng et al., 2017). On the other hand, a total absence of DPE1 was observed in *O. nivara* and *O. brachyantha*. Although there is not much clarity about the performance of DPE1 in *Oryza* chloroplasts, it is known that *Arabidopsis* plants lacking the plastidic DPE1 accumulate maltooligosaccharides (maltotriose-maltoheptaose), but not maltose, an important carbohydrate in starch formation (Critchley et al., 2001).

Completely different from DPE1, regarding its phylogenetic position and structure, but also showing an important influence in the final portion of the starch synthesis pathway in *Oryza*, the enzymes PUL and ISA catalyze different reactions, but both have a conserved gene structure. Although they play unique roles in regulating the crystallization and degradation of starch, the enzymes have a close relationship in *Oryza* and share, as expected, the N-terminal O-Glycosyl hydrolase (CBM_48) and central domain alpha-amylase (Aamy), in which both degrade amylopectin. However, in some species like *O. sativa* v.g. *japonica* and *O. longistaminata*, there is still an absence of the C-terminal domain DUF_3372 domain (Fig. S4), which characterizes the Pullulanase, and usually cleaves the α-1,6-linkages of polyglucans in pullulan. This absence may affect the final endosperm amylose content. The main gene that controls amylose is Waxy, but as starch synthesis is a fine regulatory network, together with other enzymes like PUL, AGPase, SSI, ALK, and SSIII2, they control the final content of amylose (AC). However, in the absence of pullulan degradation, the final starch content may be lower, and consequently the AC (Tian et al., 2009). Exactly what is perceived in the *O. sativa* v.g. *japonica* genotypes that have amylose content around 10–22% (low AC) while *O. sativa* v.g. *indica* show 18–32% (high AC) (Lang and Buu, 2004; Ayabe et al., 2009).

On the other hand, ISA, different from PUL, contains long and frequent introns in its gene structure, besides it also possesses every single motif that form the formerly discussed protein signature. We identified an event in *O. glaberrima* where *PUL* (Fig. 1) is duplicated and translocated from Chr 4 to Chr 6 (Additional file 13: Figure S12). Although NHR constitutes a relatively frequent event in *Oryza* genomes, one might think that MEI could also be the responsible for such duplication and translocation, since these events frequently generate syntenic failures between homologue chromosomes when comparing different species (Ammiraju et al., 2008), though here we show (Fig. S12) that it is not possible that MEI insertion could have occurred in these *PUL* genes. Neither any recombination inference was found. However, 43 sites were observed to be under positive selection (Additional file 14: Figure S13) in DBE, ISA and PUL gene copies phylogeny. The same event could also have occurred in the other genes that have different chromosome positions (Fig. 1). Both Nougué et al. (2014) and Qu et al (2018) reports DBE

homologue diversification could be explained by the strong positive selection under these genes, as well support its prominence along the complex evolutionary history of starch biosynthesis pathway.

Starch Branching Enzymes (sbe)

In total, 24 SBE genes were identified in *Oryza* and *L. perrieri* (Fig. 1 and Additional file 15: Figure S14). Positive selection was found in the gene alignment (Additional file 16: Figure S15), but no recombination events were detected. According to the position of *L. perrieri* in the phylogenetic tree, SBEs are defined as a mixed clade, that present a very conserved gene structure and protein signature that comprises SBE3 and SBE1. Although the conserved motif analysis showed that motif 9 is not present in OPUNSBE3 and OMERSBE3, they contain many more exons than the SBE3s of other *Oryza* species. *Oryza* species present multiple SBE isoforms, more than shown here, but these are the major genes involved in the synthesis of amylopectin (Zeng et al., 2017). SBE proteins are characterized by a modular architecture composed of an N-terminal domain with a carbohydrate-binding module family 48 (CBM48), a central α-amylase domain, as well as a α-amylase C-terminal domain. Both C and N termini play important roles in determining the substrate preference, catalytic capacity and chain length transfer (Kuriki et al., 1997). The importance of SBE1 in synthesis of B1, B2, B3 chains of amylopectin has been reported in rice mutants (Satoh et al., 2003a, b), while others show that SBE3 has a role in the synthesis of 1–6 branching linkage (Chen et al., 2004). In *Oryza*, these two enzymes are in the same clade. Some residues of binding sites for maltopentaose and glucose were not conserved between SBEI and SBEII isoforms, however these residues were mainly found in SBEIII, which seems to be the reason for such a close proximity between SBE1 and SBE3 in *Oryza* (Qu et al., 2018).

In summary, we identified and characterized SSRG homologs in the wild relatives of rice. Using phylogenetics and comparative genomics analyses we offer insights for the use of their gene variations in plant breeding. We confirmed the relative conservation of SSRGs between species within the AA-, BB- and FF-genomes, but structural analysis of these proteins suggest that deletions/mutations of amino acids in some active sites can result in structural variation that may negatively affect specific phases of starch biosynthesis. Direct modification of the endosperm, as usually observed in *O. sativa* v.g. *japonica*, which possesses lower AC, can likely be related to the absence of PUL C-terminal domain. The complete deletion of some genes appears not to affect the final composition of starch in the endosperm, as observed for GBSSII in *O. meridionalis*, SSIV2 in *O. glaberrima*, and *DPE1* in *O. brachyantha* and *O. nivara.*

The analysis of structural features points to both absence and duplicated copies of some motifs that can modify metabolic activity, denoting that the use of different *Oryza* species can be a rich source of variability for starch-targeted improvement in rice. These genes should now be further investigated by phenotyping different mutants and through the characterization of starch content of both wild *Oryza* genotypes and near isogenic lines (NILs) of *O. sativa* containing introgressions of these wild relatives. Such an analysis will help us to reveal the role of each variation of these genes thereby contributing greatly to the simplification of the improvement processes that involve this complex path.

List Of Abbreviations

AC: amylose content; SSRGs: Starch synthesis-related genes; AGPase: ADP-glucose pyrophosphorylase; Aamy: alpha-amylase; CBM: Carbohydrate-Binding Module; Chr: chromosome; CQ: cooking quality; SS: Starch Synthase; DBE: Debranching enzyme; dN/dS: ratio of synonymous/non-synonymous mutations; DPE1: Disproportionating enzyme; Glc-1-P: glucose 1-phosphate; GT: gelatinization temperature; ISA: Isoamylase; PPi: inorganic pyrophosphate; PUL: Pullulanase; SBE: Starch Branching enzyme; TE: Transposable Element; RiTE-db: Rice Transposable Elements database; LTR TEs: Long Terminal Repeats Transposable Elements; NHR: non-homologous recombination; NIL: near isogenic lines; MEI: mobile element insertion; Glico_transf: glycoside hydrolase family.

Declarations

CONSENT FOR PUBLICATION

Not applicable.

ADDITIONAL FILES

Methods and other additional files are available at Rice Online.

COMPETING INTERESTS

The authors declare no competing interests.

FUNDING

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) -Finance Code 001, by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and by the Fundação de Amparo a Pesquisa do Rio Grande do Sul (FAPERGS).

ACKNOWLEDGMENTS

We would like to thank the Arizona Genomics Institute and its School of Plant Sciences, Ecology & Evolutionary Biology for the providing data and for contributing to the review of details related to the use of these. No conflict of interest declared.

AUTHOR CONTRIBUTIONS

K.E.J.F conducted the major bioinformatics analyses, data interpretation and wrote the manuscript. R.S.S., J.L.L., R.W., and C. B. contributed to performing the research and revising the manuscript, F.C.V conduct the selection pressure and *in silico* recombination analysis and A.C.O. conceived the study and supervised the research.

References

Ammiraju JS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, ... & Talag J (2008) Dynamic evolution of Oryza genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. The Plant Cell 20: 3191-3209. doi: 10.1105/tpc.108.063727

Ayabe S, Kasai M, Ohishi K, Hatae K (2009) Textural properties and structures of starches from *indica* and *japonica* rice with similar amylose content. F Sci Tec Res. 15:299-306. doi: 10.3136/fstr.15.299

Ball S, Colleoni C, Cenci U, Raj JN, Tirtiaux C (2011) The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis. J Exp Bot. 62: 1775–1801. doi:10.1093/jxb/erq411

Bai Z, Chen J, Liao Y, Wang M, Liu R, Ge S, ... & Chen M, (2016) The impact and origin of copy number variations in the Oryza species. BMC genomics. 17: 261. doi: 10.1186/s12864-016-2589-2

Batra R, Saripalli G, Mohan A, Gupta S, Gill KS, Varadwaj PK, Balyan HS and Gupta PK (2017) Comparative Analysis of AGPase Genes and Encoded Proteins in Eight Monocots and Three Dicots with Emphasis on Wheat. *Front. Plant Sci*. 8:19. doi: 10.3389/fpls.2017.00019

Chen MH, Huang LF, Li HM, Chen YR, Yu SM (2004) Signal peptide-dependent targeting of a rice α-amylase and cargo proteins to plastids and extracellular compartments of plant cells. Plant Physiol. 135:1367–77. doi: 10.1104/pp.104.042184

Corbi J, Dutheil JY, Damerval C, Tenaillon MI, Manicacci D (2012) Accelerated evolution and coevolution drove the evolutionary history of AGPase sub-units during angiosperm radiation. *Ann Bot* 109: 693–708. doi:10.1093/aob/mcr303

Critchley JH, Zeeman SC, Takaha T, Smith AM, Smith SM (2001) A critical role for disproportionating enzyme in starch breakdown is revealed by a knock-out mutation in *Arabidopsis*. Plant J 26:89–100. Doi: 10.1046/j.1365-313x.2001.01012.x

Deschamps P, Colleoni C, Nakamura Y, ... & Suzuki E (2008) Metabolic Symbiosis and the Birth of the Plant Kingdom. *Mol Biol Evol* 25: 536–548, https://doi.org/10.1093/molbev/msm280

Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D, (2008) Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. Genome Res. 18: 1924–37. doi: 10.1101/gr.081026.108

Fujita N, Yoshida M, Asakura N, Ohdan T, Miyao A, Hirochika H, Nakamura Y (2006). Function and characterization of starch synthase lusing mutants in rice. Plant Physiol 140: 1070–1084. doi: 10.1104/pp.105.071845

Georgelis N, Braun EL, Shaw JR, Hannah CL (2007). The two AGPase subunits evolve at different rates in angiosperms, yet they are equally sensitive to activity-altering amino acid changes when expressed in bacteria. *Plant Cell.* 19, 1458–1147. doi: 10.1105/tpc.106.049676

Georgelis N, Braun EL, Hannah, LC (2008). Duplications and functional divergence of ADP-glucose pyrophosphorylase genes in plants. *BMC Evol. Biol.* 8:232. doi: 10.1186/1471-2148-8-232

Guo H, Jiao Y, Tan X, Wang X, Huang X, Jin H, Paterson AH (2019) Gene duplication and genetic innovation in cereal genomes. Gen Res 29: 261-269. doi: 10.1101/gr.237511.118

Hirose T & Terao T (2004) A comprehensive expression analysis of the starch synthase gene family in rice (Oryza sativa L.). Planta. 220: 9-16. doi: 10.1007/s00425-004-1314-6

Kasem S, Waters DLE, Rice NF, Shapter FM, & Henry RJ (2011) The endosperm morphology of rice and its wild relatives as observed by scanning electron microscopy. Rice. 4:12-20. doi: 10.1007/s12284-011-9060-4

Kuriki T, Stewart DC, Preiss J (1997) Construction of chimeric enzymes out of maize endosperm branching enzymes I and II: activity and properties. J Biol Chem. 272: 28999–9004. doi: 10.1074/jbc.272.46.28999

Lang NT, Buu BC (2004) Quantitative analysis on amylase content by DNA markers through backcross populations of rice (*Oryza sativa* L.). Omon rice Res Inst 12: 13–18.

Mondal TK, Henry RJ (2018) The Wild Oryza Genomes: Springer International Publishing.

Myers AM, Morell MK, James MG, Ball SG (2000) Recent progress toward understanding the biosynthesis of the amylopectin crystal. Plant Physiol 122: 989–98. Doi 10.1104/pp.122.4.989

Nakamura Y (2002) Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. Plant Cell Physiol 43: 718–25. doi: 10.1093/pcp/pcf091

Nougué O, Corbi J, Ball SG, Manicacci D, Tenaillon M (2014) Molecular evolution accompanying functional divergence of duplicated genes along the plant starch biosynthesis pathway.BMC Evol Biol 14: 103. doi.org/10.1186/1471-2148-14-103

Pandey MK, Rani NS, Madhav MS, Sundaram RM, Varaprasad GS, Sivaranjani AKP... & Kumar A (2012) Different isoforms of starch-synthesizing enzymes controlling amylose and amylopectin content in rice (Oryza sativa L.). Biotec Adv 30: 1697-1706. doi: 10.1016/j.biotechadv.2012.08.011

Patron NJ & Keeling PJ (2005) Common Evolutionary Origin of Starch Biosynthetic Enzymes in Green and Red Algae1. J Phyc 41: 1131–1141. doi:10.1111/j.1529-8817.2005.00135.x

Qu, J, Xu S, Zhang Z, Chen G, Zhong Y, Liu L, ... & Guo D (2018) Evolutionary, structural and expression analysis of core genes involved in starch synthesis. Sci Rep. 8: 12736. doi: 10.1038/s41598-018-30411-y

Satoh H, Nishi A, Fujita N, Kubo A, Nakamura Y, ... & Kawasaki T (2003a) Isolation and characterization of starch mutantsin rice. J Appl Glycosci. 50:225–30. doi: 10.5458/jag.50.225

Santos RS dos, Farias, da DR, Pegoraro C, Rombaldi CV, Fukao T, Wing RA, de Oliveira AC (2017) Evolutionary analysis of the SUB1 locus across the *Oryza* genomes. Rice 10. doi: 10.1186/s12284-016-0140-3

Satoh H, Nishi A, Yamashita K, Takemoto Y, Tanaka Y, ... & Hosaka Y (2003b) Starch-branching enzyme I-deficient mutation specifically affects the structure and properties of starch in rice endosperm. Plant Physiol. 133:1111–21. doi: 10.1104/pp.103.021527

Smith AM, Denyer K, Martin C (1997) The synthesis of the starch granule. Plant Mol. Biol. 48: 67– 87. doi: 10.1146/annurev.arplant.48.1.67

Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, ... & Gao D (2018) Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. Nat Genet. 50: 285-296. doi: 10.1038/s41588-018-0040-0.

Szydlowski N, Ragel P, Raynaud S, Lucas MM, Roldan I, Montero M, … & Mérida A (2009) Starch granule initiation in Arabidopsis requires the presence of either class IV or class III starch synthases. P Cell. 21: 2443e2457. Doi: 10.1105/tpc.109.06652

Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, ... & Wang Y. (2009) Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. Proc Nat Acad Sci 51: 21760-21765. doi: 10.1073/pnas.0912396106

Vrinten PL, Nakamura T (2000) Wheat granule-boundstarch synthase I and II are encoded by separate genesthat are expressed in different tissues. Plant Physiol 122: 255–264. Doi: 10.1104/pp.122.1.255

Yu G, Olsen KM, Schaal BA (2011) Molecular evolution of the endosperm starchsynthesis pathway genes in rice (*Oryza sativa* L.) and its wild ancestor, *O. rufipogon* L. Mol Biol Evol 28:659–671. doi: 10.1093/molbev/msq243.

Wang W, Wei X, ... & Jiao G (2019) *GBSS-BINDING PROTEIN*, encoding a CBM48 domain-containing protein, affects rice quality and yield. J Int Plant Biol. doi: 10.1111/jipb.12866

Walter M, Marchezan E, Avila LA (2008) Arroz: composição e características nutricionais. Ciência Rural. 38:1184-1192. doi: 10.1590/S0103-84782008000400049

Zeeman SC, Kossmann J, Smith AM (2010) Starch: its metabolism, evolution, and biotechnological modification in plants. Ann Rev Plant Biol. 61: 209-234. Doi: 10.1146/annurev-arplant-042809-112301

Zeng D, Tian Z, ... & Rao Y (2017) Rational design of high-yield and superior-quality rice. Nat Plants 3: 17031. doi: 10.1038/nplants.2017.31

Figures

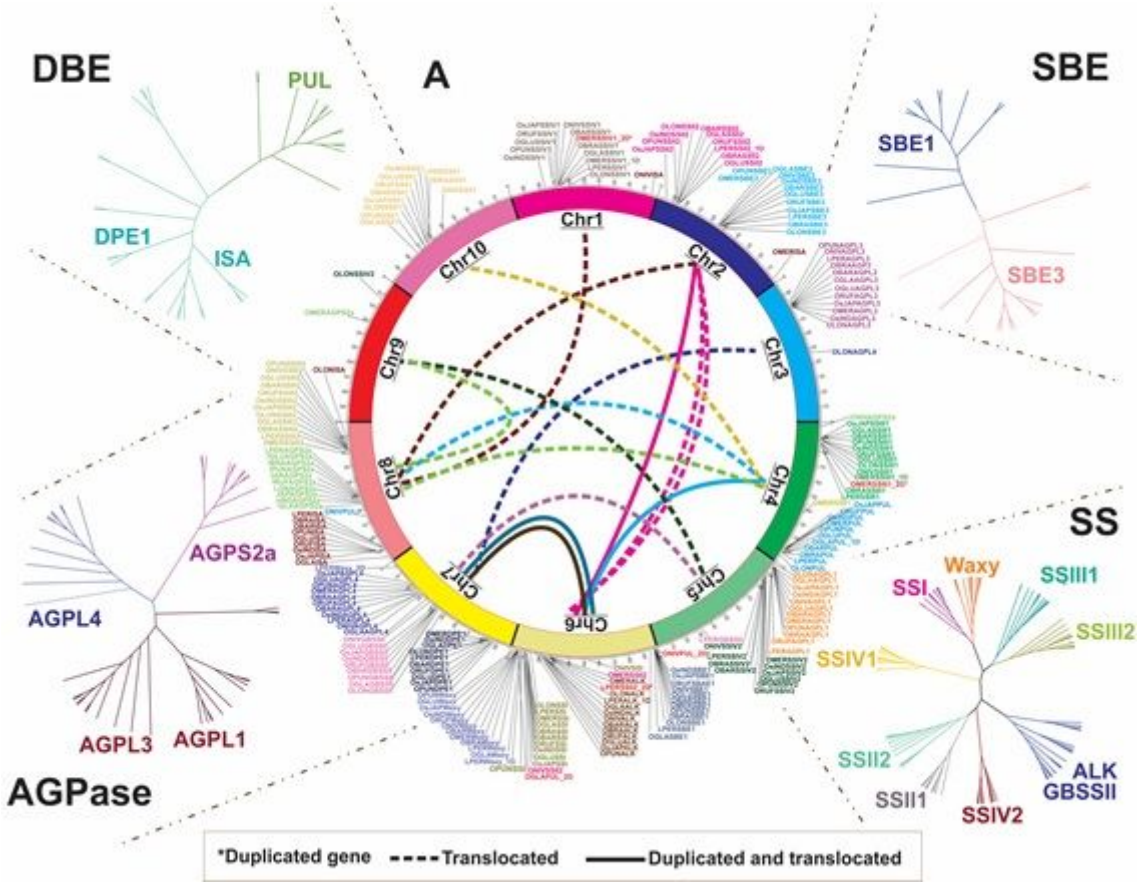

Figure 1

Gene localization (center) and phylogeny of SSR proteins. Center. Circular map of SSRGs in a single representation from chromosomes 1 to 10 in Oryza species. Duplicated genes are marked with an asterisk. Around, from the upper left, clockwise. Phylogenetic analysis of DBE, SBE, SS and AGPase proteins. Clade groups are indicated by different colors followed by the name.

Supplementary Files

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

- Additionalfile1table.docx
- Additionalfile2FigureS1.pdf
- Additionalfile3FigureS2.pdf
- Additionalfile4FigureS3.pdf

- Additionalfile5FigureS4.pdf
- Additionalfile6FigureS5.pdf
- Additionalfile7FigureS6.pdf
- Additionalfile8FigureS7.pdf
- Additionalfile9FigureS8.pdf
- Additionalfile10FigureS9.pdf
- Additionalfile11FigureS10.pdf
- Additionalfile12FigureS11.pdf
- Additionalfile13FigureS12.pdf
- Additionalfile14FigureS13.pdf
- Additionallyfile15FigureS14.pdf
- Additionallyfile16FigureS15.pdf