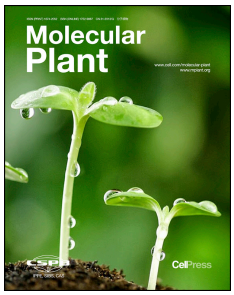
Accepted Manuscript



Regulation of seed vigor by manipulation of raffinose family oligosaccharides (RFOs) in maize and Arabidopsis

Tao Li, Yumin Zhang, Dong Wang, Ying Liu, Lynnette M.A. Dirk, Jack Goodman, A. Bruce Downie, Jianmin Wang, Guoying Wang, Tianyong Zhao

PII: S1674-2052(17)30333-7
DOI: 10.1016/j.molp.2017.10.014
Reference: MOLP 542

To appear in: *MOLECULAR PLANT*
Accepted Date: 31 October 2017

Please cite this article as: **Li T., Zhang Y., Wang D., Liu Y., Dirk L.M.A., Goodman J., Downie A.B., Wang J., Wang G., and Zhao T.** (2017). Regulation of seed vigor by manipulation of raffinose family oligosaccharides (RFOs) in maize and Arabidopsis. Mol. Plant. doi: 10.1016/j.molp.2017.10.014.

Title Page

Title:

Regulation of seed vigor by manipulation of raffinose family oligosaccharides (RFOs) in maize and Arabidopsis

Tao Li[a,b]· Yumin Zhang[a,b]· Dong Wang[a,b]· Ying Liu[a,b]· Lynnette M.A. Dirk[c]· Jack Goodman[d] A. Bruce Downie[c]· Jianmin Wang[e]· Guoying Wang[e*]· Tianyong Zhao[a,b*]

[a]State Key Laboratory of Crop Stress Biology for Arid Areas, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, 712100, China.

[b]The Key Laboratory of Biology and Genetics Improvement of Maize in Arid Area of Northwest Region, Ministry of Agriculture, Northwest A&F University, Yangling, Shaanxi, 712100, China.

[c]Department of Horticulture, Seed Biology, College of Agriculture, Food, and Environment, University of Kentucky, Lexington, KY 40546, USA

[d]Department of Plant and Soil Sciences, College of Agriculture, Food, and Environment, University of Kentucky, Lexington, KY 40546, USA

[e]Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Running title: RFOs regulation of seed vigor

Summary: Raffinose, the only member of the RFOs in maize, regulates maize seed vigor. Unlike maize, seeds of Arabidopsis also synthesize higher order RFOs (stachyose or verbascose) in addition to raffinose. The total amount of RFOs and the ratio of RFOs to sucrose, especially the ratio of higher order RFOs to sucrose control Arabidopsis seed vigor.

*Corresponding authors

Correspondence to Tianyong Zhao, or Guoying Wang

Tianyong Zhao, Ph.D, Professor, Email: tzhao2@nwafu.edu.cn

Guoying Wang, Ph.D, Professor, Email: gywang@caas.net.cn

ABSTRACT

Raffinose family oligosaccharides (RFOs) accumulate in seeds during maturation desiccation in many plant species. It is still unclear whether RFOs have a role in establishing seed vigor. GALACTINOL-, RAFFINOSE- and STACHYOSE-SYNTHASE (GOLS, RS and STS) are enzymes responsible for RFOs biosynthesis. Only raffinose is detected in maize seeds and a unique maize *RS* gene (*ZmRS*) was identified. Two independent *mutator* (*Mu*)-interrupted *zmrs* lines, containing no raffinose but hyper-accumulating galactinol, had significantly reduced seed vigor, compared to null segregant (NS) controls. Unlike maize, Arabidopsis seeds contain several RFOs (raffinose, stachyose and verbascose). Manipulation of Arabidopsis RFOs content by overexpressing *ZmGOLS2*, *ZmRS* or *AtSTS* demonstrated that co-overexpression of *ZmGOLS2* and *ZmRS*, or overexpression of *ZmGOLS2* alone, significantly increased total RFOs and enhanced seed vigor. Surprisingly, while over-expression of *ZmRS* increased raffinose it dramatically decreased seed vigor, galactinol, stachyose and verbascose amounts, while the mutant (*atrs5*) was similar to WT in all respects except for stachyose, which it accumulated. Total RFOs, RFOs:sucrose ratio, but not absolute individual RFOs amounts, positively correlated with Arabidopsis seed vigor, with stachyose and verbascose contributing more than raffinose. These findings provide new and contrasting information about the requirement for RFOs for seed vigor of a monocot and a dicot.

INTRODUCTION

Raffinose family oligosaccharides (RFOs) are thought to play a role in promoting seed vigor, seed longevity in storage, and plant abiotic stress-tolerance based on positive correlations between RFOs accumulation in the late stage of seed development or in leaves when plants encounter abiotic stress and subsequent survival of these stresses (Downie et al., 2003; Egert et al., 2015; Koster and Leopold, 1988). The first committed step of RFOs biosynthesis is the GALACTINOL SYNTHASE (GOLS; inositol 3-alpha-galactosyltransferase; EC 2.4.1.123) catalyzed production of galactinol using *myo*-inositol and UDP-galactose as substrates (Saravitz et al., 1987). The second step uses RAFFINOSE SYNTHASE (RS; galactinol-sucrose galactosyltransferase; EC 2.4.1.82) to generate raffinose using galactinol and sucrose as substrates (Peterbauer et al., 2002). RFOs with a higher degree of polymerization (DP) exceeding 3 monosaccharides are synthesized by the further addition of galactose moieties, donated by galactinol, to lower DP RFOs. For example, stachyose is synthesized by STACHYOSE SYNTHASE (STS; galactinol-raffinose galactosyltransferase; EC 2.4.1.67) using the substrates galactinol and raffinose (Gangl et al., 2015). Alternatively, the higher DP RFOs can also be produced by the action of a GALACTAN: GALACTAN GALACTOSYL TRANSFERASE (GGT) in those species in which this enzyme exists (Bachmann and Keller, 1995; Peterbauer and Richter, 1998; Tapernoux-Luthi et al., 2004); the terminal galactosyl residue is transferred from one RFOs member to another to form one higher (HDP) and one lower DP RFOs.

Despite years of study, it remains unclear whether galactinol, or any RFOs, are definitively required for, and directly involved in, seed vigor (Dierking and Bilyeu, 2009) or vegetative drought- or cold-tolerance (Amiard et al., 2003; Zuther et al., 2004). If required, the possible mechanisms of action (peripheral metabolic involvement (Kim et al., 2016); free radicle scavenging (Nishizawa et al., 2008); sucrose crystallization inhibition (Leinen and Labuza, 2006); water replacement (Martinez-Villaluenga et al., 2008) and the cellular component(s) affected (lipids (Hincha et al., 2003); proteins (Wendorf et al., 2004); general metabolic dampening (Sun and Leopold, 1997); or combinations of these) remains obscure.

GOLS has been extensively studied using overexpression of various *GOLS* genes from a host of

73 species in transgenic Arabidopsis or tobacco plants which has generally increased galactinol- and
74 raffinose-content while concurrently enhancing abiotic stress tolerance (Gu et al., 2016; Himuro et
75 al., 2014; Shimosaka and Ozawa, 2015; Taji et al., 2002; Zhuo et al., 2013). Presumably, redundancy
76 among the members of the *GOLS* gene family in plants has resulted in only a few studies reporting
77 results from *gols* mutants. In Arabidopsis, *atgols1* and *atgols1 atgols2* double mutant seeds were
78 negatively correlated with seed longevity (de Souza Vidigal et al., 2016). Those that have examined
79 *gols* mutant vegetative stress tolerance have only recorded a biochemical phenotype (reduction in
80 galactinol and raffinose quantities after heat shock (Panikulangara et al., 2004); or none at all
81 (Nishizawa et al., 2008); under multiple abiotic stresses. A single report has documented a reduction
82 in the induced systemic resistance against the fungal pathogen *Pseudomonas chlororaphis* O6, a
83 biotic stress, in *atgols1* (Cho et al., 2010).

84 Compared to the well-studied GOLS, little is known about RS, which shares high sequence
85 homology with ALKALINE-α-GALACTOSIDASE (AGA; EC 3.2.1.22) (Peters et al., 2010). To
86 date, while several putative RS genes have been reported, few of these have been confirmed to have
87 the ability to synthesize raffinose in vitro or in vivo (Egert et al., 2013; Gangl et al., 2015; Lahuta et
88 al., 2014; Li et al., 2007; Peterbauer et al., 2002; Sui et al., 2012). In maize there were ten *RS* genes
89 predicted but none have been characterized for RS activity (Zhou et al., 2012). In the model plant,
90 *Arabidopsis thaliana*, there are six putative *RS* genes (*AtRS1-6*) (Nishizawa et al., 2008), but, to date,
91 only AtRS4, identified as a STACHYOSE SYNTHASE (AtSTS) with some raffinose synthetic
92 capacity (Gangl et al., 2015) and AtRS5, a RAFFINOSE SYNTHASE without the capacity to
93 synthesize HDP RFOs (Egert et al., 2013), have been biochemically verified to produce raffinose.
94 The *atrs5* mutation abolished raffinose in the leaves but only reduced the raffinose content in seeds,
95 leading to speculation that a second RS exists in this species (Egert et al., 2013), a role filled by
96 AtRS4. The *atrs4* mutant was devoid of stachyose, but had greater raffinose in the seeds, while the
97 double knockout *atrs4 atrs5* mutant seeds lacked both raffinose and stachyose but had much more
98 galactinol (Gangl and Tenhaken, 2016). In keeping with a role for RFOs in seed vigor, the
99 Arabidopsis *atrs4* mutant seeds (no stachyose; increased raffinose) completed germination faster
100 than control seeds (Gangl et al., 2015); while the *atrs4 atrs5* double mutant seeds exhibited a 5

day-delayed completion of germination in darkness which was alleviated in the light or partially alleviated by the addition of galactose to the media in darkness, both novel findings. To our knowledge, the *atrs4* mutant in Arabidopsis is the sole example of the recovery and characterization of a mutation in a *STACHYOSE SYNTHASE* (Gangl et al., 2015) while no reports have explored mutation of enzymes involved in RFOs with a degree of polymerization greater than Stachyose.

Seed vigor is a multigenic trait, as emphasized by the existence of a variety of quantitative trait loci (QTL) contributing to this seed characteristic (Cheng et al., 2013; Cui et al., 2002; Dargahi et al., 2014; Liu et al., 2014; Xie et al., 2014). This complicates and restrains the genetic modification and the traditional breeding of crops for improvement of seed vigor. RFOs content is positively correlated with seed desiccation-tolerance of soybean (Koster and Leopold, 1988) but is dispensable for desiccation tolerance in Arabidopsis seeds (Gangl and Tenhaken, 2016). A positive correlation between raffinose content in embryos and seed vigor was declared in maize seeds (Bernal-Lugo and Leopold, 1992). Despite these studies, the knockout mutants in genes involved in RFOs biosynthesis in Arabidopsis have failed to show any phenotypic changes during normal growth other than perturbations in soluble sugar quantities (Egert et al., 2013; Gangl et al., 2015). In addition, the function of galactinol and each individual member of RFOs, such as raffinose, stachyose or verbascose, has not been specifically investigated or has not been distinguished from one another.

In this study, we confirm that raffinose is the only RFOs member stored in maize seed and demonstrate that *zmrs* seeds, which lack raffinose, although surviving desiccation, display significantly reduced seed vigor. As well, we corroborated that Arabidopsis seeds contain several members of RFOs, including raffinose, stachyose and verbascose. The total amount of RFOs and the ratio of RFOs to sucrose, but not the absolute amount of individual RFOs members, determined Arabidopsis seed vigor. Galactinol, a substrate for RFOs synthesis and correlated with seed longevity (de Souza Vidigal et al., 2016), does not directly contribute to seed vigor in either maize or Arabidopsis. We provide evidence that, despite being a multigenic trait, seed vigor can be altered positively through careful manipulation of RFOs composition relative to sucrose amounts.

RESULTS

Identification and Characterization of a Maize *RAFFINOSE SYNTHASE* gene

To identify maize RAFFINOSE SYNTHASE (ZmRS), four RAFFINOSE SYNTHASE (RS) protein sequences from other species, known to have raffinose synthetic activity, were used to BLAST the maize B73 genome (AtRS5, NP_198855; CsRS, E15707; PsRS, CAD20127; OsRS, XP_015621501) (http://www.maizegdb.org/). These RS protein sequences showed high homology to a single maize protein (GRMZM2G150906-P01) and this maize protein shared a greater evolutionary relationship with RSs than STSs from other species (Figure. 1A). Sequence alignment analysis showed that ZmRS shared 60.42%, 59.45%, 59.70% and 78.91% identity with AtRS5, CsRS, PsRS and OsRS, respectively (Supplemental Figure 1). Two known conserved motifs, DD×W and K×D were present in AGA, as well as in ZmRS and other RS (Supplemental Figure 1) (Carmi et al., 2003). Notably, two conserved motifs (FM×LGTEA××LG and SGDP×GT×WLQGOHMVHC) present in all RS sequences were absent in AGA amino acid sequences (Supplemental Figure 1). This indicated that GRMZM2G150906 was most likely encoding a RS rather than an AGA or a STS.

To characterize the biochemical function of the putative *ZmRS* gene product, the *ZmRS* CDS was cloned into the *E. coli* expression vector *pET-21d* in frame with a carboxy-terminal hexahistidyl tag (Supplemental Figure 2A). The ZmRS protein is predicted to contain 790 amino acids with a calculated molecular mass of 85 kDa. After induction by IPTG, a recombinant presumptive ZmRS:His_6 protein band of approximately 85 kDa was detected (Supplemental Figure 2B arrowhead). In the presence of sucrose and galactinol, the crude bacterial extract containing ZmRS:His_6 protein produced raffinose and *myo*-inositol whereas the control crude extract from *E. coli* cultures transformed with an empty *pET-21d* vector failed to produce either in the same time frame (Figure 1B). In addition to the raffinose synthetic activity, crude bacterial extract with the ZmRS:His_6 protein also hydrolyzed galactinol and released *myo*-inositol if sucrose was not present whereas the bacterial lysate from the empty vector was incapable of such hydrolysis (Figure 1C). Even though ZmRS shared high-homology with AGA, the recombinant protein was incapable of hydrolyzing raffinose (Figure 1D).

ZmRS gene expression and the protein accumulation were analyzed by real time RT-PCR and western blot analysis, respectively. Both mRNA and protein expression of *ZmRS* were determined to occur

predominantly in leaves and in embryos (Supplemental Figures 3A and 3B). Raffinose accumulated to much greater amounts in embryos compared to endosperms in mature seeds (Supplemental Figure 3C) and ZmRS protein abundance was much greater in embryos than in endosperms at the late stage of seed development (Supplemental Figure 3D). In embryos, the *ZmRS* mRNA and ZmRS protein started to accumulate at 30 days after pollination (DAP) and peaked between 35-45 DAP (Figure 2A) and 45 DAP (Figure 2B), respectively, and yet the protein was already undetectable in 55 DAP embryos. In overall trends, raffinose amounts were more closely correlated with *ZmRS* mRNA expression level (Figures 2A and 2C) than the corresponding protein accumulation (Figure 2B) during development. After 40 DAP, raffinose amounts were trended toward a decline, possibly due to the dramatically increased expression at 45 DAP of the *ZmAGA1* encoding an alkaline-galactosidase that hydrolyzes raffinose (Figure 2C; Supplemental Figure 4). During seed imbibition, both embryo *ZmRS* mRNA (Figure 2A) and raffinose (Figure 2C) trended steadily downward whereas the ZmRS protein was below the level of detection (Figure 2B).

The role of raffinose in maize was investigated using two *mutator* (*Mu*)-inserted *zmrs* mutant lines, which were obtained from different resources from different inbreds (specifically W22 and B73) and which were backcrossed twice in respective parental lines and subsequently selfed and genotyped to track homozygosity for both the *Mu*-insertional mutant and its corresponding null segregant (NS) (Figure 3A and Supplemental Figure 5A). PCR determined the region of *Mu* insertion in the *ZmRS* gene for both mutant alleles (Figure 3B and Supplemental Figure 5B). A 2.2 kb and a 0.55 kb DNA fragment was amplified from the *zmrs-1* mutant and NS plant, respectively (Figure 3B) using primers F2a and R2a (Figure 3A); indicating that the *Mu* had inserted between the binding sites of these primers (Figure 3A). PCR and sequencing of the resultant amplicons using a *Mu*-specific primer (TIR) with corresponding gene-specific primer (F2a, R2a) revealed a typical 9-bp duplication flanking the *Mu element* (Figures 3A and 3B). Transcriptional competency of *ZmRS* in the mutant plant was determined by RT-PCR using different primer sets anchored on different regions of the *ZmRS* CDS (Figure 3A and Supplemental Figure 5A). While transcription of the cDNA encoded by the first exon was inhibited in *zmrs-1*, a similar *ZmRS* transcript abundance was detected in both NS and mutant plants when the RT-PCR was performed (Figure 3C) using primers F3a and R3a (Figure

3A). As anticipated, RT-PCR using R2a and TIR primers did not produce an amplicon from NS cDNA, but, unexpectedly, did generate a clear amplicon from the homozygous *zmrs-1* mutant plant (Figure 3C). This indicated that the insertion of the *Mu* interrupted the transcription of *ZmRS* gene but there may be a cryptic promoter inside the *Mu* which can initiate the transcription of its downstream sequences present in the *zmrs-1* mutant. Similar characterization by PCR and RT-PCR of the *zmrs-2* mutant determined the *Mu* site was instead in the second exon and caused premature termination of *ZmRS* transcription and did not use another start site from within the *Mu* as had *zmrs-1* (Supplemental Figures 5A to 5C). By using western blot hybridization, detectable ZmRS protein accumulation was evident in the leaves of the NS but not in those of either mutant (Figure 3D and Supplemental Figure 5D). Consistent with the absence of a putative unique RS enzyme, no raffinose (the product of the enzyme-catalyzed reaction) but greater galactinol (a required substrate) were detected in the mutant leaves, embryos and endosperm (Figure 3E and Supplemental Figure 5E) as compared with respective NS tissues. Within the limits of the current methods' detection levels, these results strongly support the hypothesis that the *ZmRS* gene (GRMZM2G150906) product is uniquely responsible for raffinose synthesis in maize leaves and seeds.

Despite the loss of raffinose in both mutant lines, the 1000-seed weights between the mutants and their null segregants, within each inbred line, were statistically identical (W22 null: 186.48±9.72 g; W22 *zmrs-1*: 184.78±15.24 g; B73 null: 164.18±9.38 g; B73 *zmrs-2*: 179.23±14.18 g). Plant stature and leaf appearance were also indistinguishable when the null and *zmrs* mutant were compared within each inbred line (Supplemental Figure 6).

Raffinose enhances maize seed vigor which improves seedling growth

Seed vigor of NS and *zmrs* was tested using accelerated aging (AA) treatment, a method which is widely adopted for testing seed vigor and longevity (Bueso et al., 2014; de Souza Vidigal et al., 2016; Han et al., 2014; Rajjou et al., 2008; Wang et al., 2016). The only significant differences for the completion of germination of unaged seeds between the *zmrs-1* and its corresponding NS occurred at 36 and 48 h after imbibition (HAI) (Figures 4A and 4B). Although AA treatment for both 3 and 6 days (AA3 and AA6, respectively) decreased the germination percentage of both NS and *zmrs-1*

seeds, the performance of *zmrs-1* seeds trended as poorer than that of its NS with significant differences between the genotypes within an AA treatment at 3 times points, namely 72 h AA3, and both 108 and 120 h AA6 (Figures 4A and 4B). *zmrs-2* seeds also completed germination to a lower percentage than its NS when aged for 3 days at 60 h and beyond (Supplemental Figures 7). In addition to the seed vigor, both root- and shoot-lengths of NS seedlings produced from those seeds were significantly longer than that of the *zmrs* mutants (the one exception being the roots of AA6 for *zmrs-1* and its NS) (Figure 4C; Supplemental Figure 7C). The sugar profile of the NS and *zmrs-1* embryos before or after AA treatment revealed no significant differences in sucrose content regardless of aging treatment (Figure 4D). The *myo*-inositol and galactinol contents were consistently greater and lesser, respectively, in NS embryos relative to *zmrs-1* embryos (Figure 4D). As shown in Figure 2E, the mutant had no detectable raffinose in the embryo (Figure 4D) and HDP RFOs were not present in maize. Taken together, these seed vigor data and sugar profiles supports the hypothesis that a complete loss of raffinose detrimentally influences seed vigor in maize; that increases in galactinol do not compensate for the loss of raffinose, and suggest that RS is potentially maintaining *myo*-inositol amounts in seeds via an unproductive cycle of galactinol synthesis/hydrolysis.

Total amount of RFOs and the ratio of RFOs to sucrose, not the individual RFOs members, determines Arabidopsis seed vigor

Raffinose, stachyose and verbascose are 3 major forms of RFOs that can be detected in seeds of Arabidopsis (Righetti et al., 2015) and other crop plants, e.g. soybean (*Glycine max*) (Blackman et al., 1992) and pea (*Pisum sativum*) (Peterbauer et al., 2003). To determine the effects of RFOs on Arabidopsis seed vigor, Arabidopsis lines lacking a RS (*atrs5*), or constitutively expressing *ZmGOLS2*, *ZmRS*, *ZmGOLS2/ZmRS*, or *AtSTS* (*At4g01970*) were generated and the expression of relevant genes was confirmed by RT-PCR or western blot analysis (Supplemental Figures 8, 9A and 10A). Previously generated (Gu et al., 2016) Arabidopsis lines, constitutively expressing *GFP* or *ZmGOLS2*, were also included in the experiment. These Arabidopsis lines (within a Figure) were grown at the same time, in the same place and harvested on the same day producing seedlots to be characterized for seed viability, seed AA tolerance and the content of each RFOs member (Figure 5; Supplemental Figures 9 and 10). Unaged seeds from all lines completed germination to a similar

239 extent as WT or *OEGFP* lines (Figure 5C; Supplemental Figures 9C and 10C). In all lines for which
240 seeds completed germination after the 3-day AA treatment to a greater percentage than those seeds of
241 WT and OEGFP, the seeds prior to the AA treatment had less sucrose and that was observed as a
242 negative correlation with tolerance to AA upon Pearson correlation analysis (Table 1; Figure 5E and
243 Supplemental Figures 9E).

244 The *OEZmRS* lines contained statistically significantly less galactinol, stachyose and verbascose and
245 more raffinose, relative to the WT seeds (Figures 5G to 5J; Supplemental Figures 9G to 9J) and;
246 upon seed aging for 3 days, completed germination more poorly than the WT or *OEGFP* lines
247 (Figure 5D and Supplemental Figure 9D). The *atrs* seeds had more galactinol, less stachyose, and
248 similar amounts of raffinose and verbascose to WT and *OEGFP* lines while, following aging, its seed
249 vigor was similar to that of the WT or *OEGFP* (Supplemental Figure 9D, 9G to 9J). The
250 OE*ZmGOLS2* seeds contained more galactinol, stachyose and verbascose and similar raffinose,
251 relative to the WT and *OEGFP* and produced seeds that withstood aging significantly better than the
252 controls (Figures 5D, 5G to 5J; Supplemental Figures 9D, 9G to 9J). Seeds that had poorer or greater
253 vigor differed regarding stachyose content, possessing less or more, respectively. To pursue this
254 observation linking stachyose accumulation with Arabidopsis seed vigor further, Arabidopsis plants
255 overexpressing the endogenous *STACHYOSE SYNTHASE* gene were created, and grown, along with
256 *OEGFP* and WT, for seeds. One line (*OEAtSTS-4*) failed to accumulate AtSTS noticeably more than
257 WT or OEGFP, while two other lines (*OEAtSTS-1* and *-9*) hyper accumulated STS protein
258 (Supplemental Figure 10A). Seeds from these lines were analyzed for sugar quantity and identity and
259 *OEAtSTS-1*, *-9* were found to contain similar sucrose and verbascose to WT and *OEGFP* lines, less
260 galactinol, and more *myo*-inositol, raffinose, and stachyose than the seeds of the two controls
261 (Supplemental Figure 10F to 10J). Line OEAtSTS-4 varied from WT and *OEGFP* lines only with
262 respect to galactinol amounts, of which it had more. Regardless of sugar amount, the *OEAtSTS* lines
263 and the *STS* co-suppressed line, were all equal to WT and *OEGFP* seeds regarding final germination
264 percentage after aging except for 84 hours after imbibition when *OEAtSTS-1* and *-9* attained a
265 significantly greater percentage germination following aging (Supplemental Figure 10D).

266 Three double overexpressing lines of *ZmGOLS2*/*ZmRS*, independent with regards to the *ZmRS* event

267 by the transformation of *ZmGOLS2* over-expressing plants, were generated and, along with WT,
268 *OEGFP*, *OE ZmGOLS2*, and *OEZmRS-25*, grown at the same time, in the same place and harvested
269 on the same day producing seedlots to be characterized. There were more viable seeds in
270 *ZmGOLS2*/*ZmRS* double expressing lines than in any other line after AA treatment as determined by
271 tetrazolium assay (Figure 5A). After 3 days of AA treatment, seeds of *ZmGOLS2*/*ZmRS* double
272 expressing lines had the greatest germination percentage among all lines. This was not due simply to
273 the mitigation of ill effects from *OEZmRS* by concurrent *OEZmGOLS2* because seeds of the
274 *ZmGOLS2* line, although with significantly greater percentage germination than WT or OEGFP, were
275 surpassed by those of OE*ZmGOLS2*/*ZmRS* expressing lines. The seeds of the OE*ZmRS-25* line again
276 had significantly lower germination percentages than all other lines (Figures 5B and 5D). The seeds
277 of the double OE lines contained less sucrose, galactinol, stachyose, and verbascose than the WT or
278 *OEGFP* seeds but had greater amounts of raffinose (Figures 5E to 5J). This sugar profile of the
279 double over-expressors (relative to WT and OEGFP seeds) is the same as for the *OEZmRS* lines (both
280 have more raffinose than the controls) except that the double over-expressors have less sucrose than
281 the controls or the *OEZmRS* lines, and yet seed vigor of the *OEZmRS* lines was much worse, and the
282 double OE lines, much better, than WT and OEGFP seeds.
283 If the RFOs play a role in Arabidopsis seed vigor, then it may be reasonable to expect that alterations
284 leading to departures from WT relative abundances (and flux through the various RFOs precursors
285 and members) may be fraught with consequences. It is obvious that seed vigor cannot be attributed to
286 any one RFOs (Table 1). To determine if there may be a discernable pattern between the relative
287 abundance (percent composition) of sugars involved in making, and comprising the RFOs, each
288 member of RFOs, sucrose, *myo*-inositol and galactinol and the seed AA tolerance (determined by the
289 germination percentage at 84 h after imbibition following 3 days AA) were compared among all
290 Arabidopsis lines (Figure 6A). The original sugar content of each Arabidopsis line is listed
291 (Supplemental Dataset 2). Additionally, the presence of verbascose in Arabidopsis seeds was
292 validated using LC-MS/MS with high resolution accurate mass (Supplemental Figures 11-14).

293 The relative abundance of RFOs in OE*ZmGOLS2/ZmRS* expressing seeds was the greatest among all
294 lines and its AA-resistant germination percentage was also the greatest. The second greatest RFOs

295 percent composition and AA-resistant germination percentage was that of *OEZmGOLS2*. The
296 *OEAtSTS* lines have the next greatest RFOs relative abundance and AA-resistant germination
297 percentage. Finally, the *OEZmRS* lines had the least RFOs percent composition as compared to other
298 OE lines, but have greater RFOs percent composition than WT or OEGFP lines, yet produced the
299 seeds with the worst vigor. In addition, galactinol percent composition was not evidently linked with
300 Arabidopsis seed vigor (Figure 6A, Table 1) while sucrose percent composition followed a trend
301 opposite to total RFOs relative abundance. Statistical analysis using Pearson correlation was
302 performed to determine the relationship between each individual sugar, the total amount of RFOs, the
303 raffinose: sucrose ratio, the HDP RFOs: sucrose ratio and seed AA tolerance (Table 1). Total RFOs
304 was positively correlated with seed AA tolerance ($P<0.05$). Furthermore, sucrose content is
305 negatively correlated with seed AA tolerance ($P<0.01$). There is no statistically significant correlation
306 between any individual RFOs and seed AA tolerance. Likely due to the competition for the substrate
307 galactinol and the direct utilization of raffinose in the formation of stachyose, there is a negative
308 correlation between raffinose and stachyose or verbascose ($P<0.01$). Dualistic linear regression
309 analysis of effects of the ratios of raffinose/sucrose (X1) and HDP-RFOs/sucrose (X2) on seed vigor
310 (Y) defined the results as the equation provided in Figure 6B with a $R=0.737$ ($p=0.003$). According
311 to this formula, the contribution of HDP-RFOs to sucrose ratio (270) to seed AA tolerance is greater
312 than the raffinose to sucrose ratio (204) to seed AA tolerance. Thus, the carbohydrate distribution
313 within Arabidopsis seed as it relates to tolerance to AA requires further study to predict an optimum
314 target sucrose to RFOs ratio.

315 **_RS_ and _STS_ Have Evolved from the Same Ancestral Gene and Maize Has Lost the _STS_ Gene**
316 **During Evolution**

317 The protein sequences of RS and STS differ most regarding the α-amylase catalytic (AC) domain
318 from STS which is missing from RS (Supplemental Figure 15A). The coding region of the AC
319 domain of *AtSTS* gene was either deleted from *AtSTS* or inserted into the corresponding site of the
320 *ZmRS* gene (Supplemental Figure 15B). ZmRS, ΔZmRS (ZmRS with the AC domain of AtSTS),
321 AtSTS and ΔAtSTS (AtSTS with the AC domain deleted) were produced in bacteria (Supplemental
322 Figure 15C). ΔZmRS showed similar raffinose synthetic activity as ZmRS. AtSTS showed a very

weak raffinose synthetic activity, while ΔAtSTS did not show any raffinose synthetic activity (Supplemental Figure 15D). AtSTS showed a clear stachyose synthetic activity, while ΔAtSTS, ZmRS or ΔZmRS did not show any stachyose synthetic activity under the conditions tested (Supplemental Figure 15E).

Using AtSTS, AtRS (Gangl et al., 2015) or ZmRS as query sequences to search the genome of 7 dicot and 4 monocot species, RS was found in all species and STS were found in all dicot plant species. While *STS* gene was found in sorghum and foxtail millet (*Setaria italica*), it was not found in other monocot plant species, such as maize or rice (Supplemental Table 1). The RFOs profile of seeds of different plant species, at varying phylogenetic distances from maize, was examined (Supplemental Figure 16, Supplemental Table 1). Raffinose was found to be the only form of RFOs detectable in maize, teosinte (the undomesticated relative of maize), foxtail millet and rice (*Oryza sativa*), while raffinose and stachyose were detected in sorghum and all dicot plant species (Supplemental Figure 16, Supplemental Table 1). These data suggest that RS and STS have evolved from the same ancestral gene and that either the STS gene was lost in the maize ancestral genome during evolution, not during domestication or, that the teosinte lineage did not develop a STS before diverging from the ancestral group. The biosynthesis and function of RFOs in plants may be divided into two categories (Figure 7). Plants in the first category, such as maize, have raffinose as the sole RFOs member and some raffinose is necessary for seed vigor. Plants in the second category, such as Arabidopsis, contain both raffinose and higher DP RFOs members. Raffinose, in the second category of plants, may exert a protective role (as it does in the first category), act to draw down sucrose amounts through its synthesis, as well as act as a substrate for synthesis of higher DP RFOs. Regardless of the complexity between total RFOs amount and its constitution relative to sucrose, seed vigor can be positively manipulated in these seeds as well when the sucrose to RFOs ratios are altered (Figure 7).

DISCUSSION

Raffinose, the Only RFOs Member in Maize, is Directly Correlated with Maize Seed Vigor.

Ten genes were previously predicted as RS synthase using a Pfam search in maize (Zhou et al., 2012);

however, GRMZM2G100355 and GRMZM2G311684 were not protein coding sequences but transposable elements according to the MaizeGDB database. GRMZM2G340656, GRMZM2G127147 and GRMZM2G037265 have been identified as AGA (named as ZmAGA1, ZmAGA2 and ZmAGA3) by our group (Zhao et al., 2006). Phylogenetic analysis indicated that GRMZM2G050177, GRMZM2G077181 and GRMZM2G047292 were also placed into the AGA group and are missing the raffinose hallmark motifs described here (Figure 1A). One enzyme (GRMZM2G311756) groups into the alpha galactosidase clade and the last is GRMZM2G150906, a legitimate RS demonstrated to possess a raffinose biosynthetic capacity/galactinol hydrolase activity in this report.

Neither RS nor raffinose was detectable in the leaves or seeds of *zmrs* mutant plants, suggesting that *ZmRS* is a unique gene encoding a single RS in the maize genome (Figures 1A, 2 and Supplemental Figure 5). Raffinose is the only RFOs member detected in maize (Kuo et al., 1988; Obata et al., 2015), a result confirmed in this study (Supplemental Figure 16). Similar to Arabidopsis seeds devoid of raffinose (Gangl and Tenhaken, 2016), we can now definitively state that maize seeds, devoid of raffinose, do not require RFOs to assume desiccation tolerance. It has been previously reported that raffinose content is positively correlated with maize seed vigor and that a decrease in raffinose is associated with the decline of seed vigor during storage, while sucrose amounts remained unchanged (Bernal-Lugo and Leopold, 1992; Obendorf, 1997). We have demonstrated here, using independent mutant (*zmrs*) maize plants from two different backgrounds, that a complete lack of raffinose synthesis has negative consequences for maize seed vigor. Both *zmrs* mutants accumulated greater amounts of galactinol but no raffinose in all tissues tested (Figures 3E, 4D and Supplemental Figure 5E). The only reported uses of galactinol are to; 1) synthesize RFOs and; 2) potentially act as a protective molecule in some species (Nishizawa et al., 2008). Because *zmrs* mutant seeds, hyper-accumulating galactinol but without raffinose, demonstrated poor seed vigor, galactinol amount, and/or metabolic flux to it, are not directly correlated with maize seed vigor although galactinol may be positively associated with seed longevity, a separate seed attribute (de Souza Vidigal et al., 2016).

Total RFOs Amounts and the Ratio of Total RFOs to Sucrose, but not Galactinol, Determines Arabidopsis Seed Vigor.

While the influence of raffinose (or its synthesis) on maize seed vigor was clear, the production of RFOs and its relation to seed vigor in Arabidopsis was complex. Unlike maize, many dicot plants, including Arabidopsis, contain several RFOs members in their seeds (Supplemental Figure 16, Supplemental Table 1) and Arabidopsis contains 10 *GOLS* genes (Cho et al., 2010) and at least three different enzymes capable of synthesizing raffinose (Gangl and Tenhaken, 2016). Studies examining RFOs associations with seed vigor in these species have provided conflicting views, some failing to find associations (Bentsink et al., 2000; Buitink et al., 2000; Dierking and Bilyeu, 2009) while others do (Blochl et al., 2007; Vandecasteele et al., 2011). One reason for the discrepancies lies in the experimental approaches. Those without a correlation have compared lines/cells differing in RFOs contents (not complete RFOs absence) as test subjects or examined QTL associated with RFOs or sucrose accumulation individually, not as a ratio. Those elucidating a correlation do so using QTL and ratios of sucrose to total RFOs, or pharmacological inhibition of metabolite flux to all RFOs. These experimental observations then collide for several reasons. While a complete absence of raffinose gave a phenotype in maize, the residual raffinose titer required to mask this vigor phenotype may be small. But gene redundancy (Cho et al., 2010) and compensatory capacities (Gangl and Tenhaken, 2016) make the complete elimination of all RFOs in many species a daunting task. Additionally, perhaps the tacit expectation that RFOs amounts will additively influence seed vigor is inappropriate. Once some threshold RFOs is present, additional RFOs may not increase seed vigor additively (Obendorf, 1997). Finally, with multiple RFOs present in the seed, assigning any potential seed vigor benefit to a specific RFOs, potentially including galactinol, is difficult, and perhaps misplaced if it is the amalgam of all RFOs that is effective, potentially relative to sucrose abundance (Horbowicz and Obendorf, 1994). Indeed, the ratio of sucrose to RFOs has been identified to influence seed vigor (Obendorf, 1997) and it was just this ratio that was demonstrated to effectively predict seed longevity when *Medicago* seeds were examined using QTL (Vandecasteele et al., 2011).

There are some publications suggesting that galactinol positively regulates seed vigor in Arabidopsis

and some other plant species (de Souza Vidigal et al., 2016; Salvi et al., 2016). Galactinol was considered as a marker for seed storability in Arabidopsis, cabbage and tomato (de Souza Vidigal et al., 2016) and overexpression of the chickpea *GALACTINOL SYNTHASE* (*CaGOLS*) gene in Arabidopsis increased the galactinol and raffinose content and enhanced seed vigor (Salvi et al., 2016). Since both galactinol and raffinose were increased by over-expression of the *CaGOLS* gene, it is not clear whether galactinol, raffinose, or both directly enhanced Arabidopsis seed vigor. In addition, galactinol and raffinose are substrates for synthesis of higher DP RFOs, such as stachyose or verbascose. The change of stachyose or verbascose and their effects on Arabidopsis seed vigor was not investigated in *CaGOLS* expressing Arabidopsis (Salvi et al., 2016). This information is important because our evidence from overexpressing one or more genes whose products are involved in RFOs biosynthesis suggest that the total amount of RFOs and the ratio of RFOs/sucrose, but not galactinol, are responsible for Arabidopsis seed vigor (Table 1; Figure 6B). Because galactinol is the substrate for all RFOs synthesis, overexpression of the *GOLS* gene would generate more galactinol for the subsequent synthesis of RFOs. RS, STS or VES (VERBASCOSE SYNTHASE) would compete among each other for galactinol (Figure 7, Table 1) and this may explain why the content of many RFOs members changed when overexpressing any *RFOs SYNTHASE* gene in Arabidopsis (Figure 7). Furthermore, the apparent efficiency with which galactinol is utilized to make different RFOs differs among these enzymes. Note that the *ZmGOLS2OE* line has more galactinol, stachyose and verbascose but similar raffinose to WT. This could result if the efficiency of RS galactinol utilization is the poorest among the RFOs synthesizing enzymes.

Double knock out mutant seeds of *atrs4* (*stachyose synthase*) and *atrs5* (*raffinose synthase*), which were devoid of both raffinose and stachyose, but accumulated much more galactinol, showed a 5-day delayed completion of germination in darkness (Gangl and Tenhaken, 2016). This severe *atrs4*/*atrs5* Arabidopsis mutant seed germination phenotype was dramatically alleviated by exogenously supplied galactose, a phenomenon the authors suggested might be due to galactose, arising from the action of galactosidases on RFOs (but apparently not galactinol) and generating a potent metabolite (galactose) fueling the completion of germination in Arabidopsis (Gangl and Tenhaken, 2016). The authors present a model in which there is an interplay between galactose from RFOs and

PHYTOCHROME INTERACTING FACTOR6 (PIF6) in dark-germinating, positively-photoblastic, Arabidopsis seeds (Gangl and Tenhaken, 2016). The drastic delay (5 days) in the completion of germination in darkness for the *atrs4*/*atrs5* double mutant Arabidopsis seeds (Gangl and Tenhaken, 2016) was not observed for *zmrs* mutant seeds completing germination in darkness. Only one of the two *zmrs* mutant lines produced seeds for which there was any delay in the completion of germination in darkness, relative to the null segregant line (Fig. 4B; Supplemental Figure 7B). The delay in this line was not severe and, although it was statistically significant, it was evident for only 36- and 48-HAI, after which the germination percentages between the mutant and null segregant lines were identical (Fig. 4B). The *zmrs* lines had no detectable RFOs to act as a galactose-producing substrate but the time to complete germination was not severely influenced. Potentially then, maize seeds do not depend on galactose to the same extent as Arabidopsis to complete seed germination, probably due to the photoblastic differences between the two species seeds. A recent study using both *Medicago truncatula* and pea *abi5* mutants showed that the mutant seeds are sensitive to AA treatment and the RFOs amounts in embryonic axis are less than that of WT (Zinsmeister et al., 2016). There are reports that the mass ratio of sucrose to raffinose influence maize seed vigor (Brenac et al., 1997a; Brenac et al., 1997b). Our findings are in support of these publications' conclusions that the total amount of RFOs, and the ratio of RFOs/sucrose, not galactinol, is important for Arabidopsis seed vigor.

The Evolution of RFOs Synthases.

Raffinose is found in mature seeds of many plant species while higher DP RFOs, such as stachyose and verbascose accumulate only in certain plant species (Supplemental Table 1) (Janecek et al., 2011; Kuo et al., 1988). Multiple *RS* genes have been predicted and reported in Arabidopsis, even though some of them have not yet been confirmed to embody a raffinose biosynthetic capacity (Egert et al., 2013; Gangl and Tenhaken, 2016; Nishizawa et al., 2008). Unlike Arabidopsis, and despite reports to the contrary (Zhou et al., 2012) there appears to be a single, functional RS in maize. In our hands, using *AtSTS* (STACHYOSE SYNTHASE of Arabidopsis) (Gangl et al., 2015) as a query to search the maize B73 genome, we failed to find convincing evidence for a *STACHYOSE SYNTHASE* gene

460 in the maize genome. Neither have we (or others) detected RFOs of a DP exceeding raffinose in
461 maize (Supplemental Figure 16).

462 The STS gene was found in foxtail millet genome (*SiSTS*) (Supplemental Table 1), however, no
463 detectable amount of stachyose was produced (Supplemental Figure 16). Evidence from RNA-Seq
464 reads in the NCBI database demonstrates that the gene is transcribed and properly spliced but it is
465 still possible that the mRNA is not translated. Or, if the mRNA is translated, the protein has lost its
466 previous stachyose synthetic activity. A similar scenario exists for defective or non-functional
467 invertase genes (Ruan, 2014). Whether the *SiSTS* gene has developed other functions during
468 evolution needs to be further investigated.

469 A phylogenetic configuration of the GOLS, RS and STS proteins in the plant kingdom indicates that
470 GOLS and RS are relatively conserved in the monocotyledonous- compared to the
471 dicotyledonous-plants, while the higher DP RFOs synthases may not have co-evolved with GOLS
472 and RS. The RFOs biosynthetic pathway evolution might be split into two parts, the synthesis of
473 raffinose and the synthesis of higher DP RFOs (Sengupta et al., 2015). The insertion of the AC
474 domain of AtSTS in ZmRS did not enable ZmRS to gain STS activity (Supplemental Figure 15), also
475 suggesting that besides the AC domain, other specific sites of STS are essential for STS activity.
476 These data support the contention that STS may not have co-evolved with RS.

477 Maize has but one RS and lacks other alpha-galactosyl-containing oligosaccharides (e.g. planteose
478 (Gurusinghe and Bradford, 2001) or cyclitols (Horbowicz and Obendorf, 1994)) that may otherwise
479 have obfuscated the results. We propose here that while RFOs are not necessary for plant survival or
480 seed viability, a complete lack of RFOs is detrimental to seed vigor. Why plant species differ with
481 regard to the number of genes encoding RFOs biosynthetic enzymes and/or the variety of RFOs
482 produced is not clear. The explanation that wild plant species are often challenged by extreme
483 environmental stress and so reduction in the RFOs variety is an artifact of domestication does not
484 bear scrutiny when one considers that the seeds of the wild relative of maize, teosinte, also contained
485 only raffinose. In different cultivars or different breeding lines of the same crop species, the RFOs
486 content varies (Kuo et al., 1988) and *RFOs synthase* gene variants exist (Dierking and Bilyeu, 2008;

Peterbauer and Richter, 2001). Whether seed vigor varies in species/cultivars with different RFOs amounts, ratios, and RFOs synthase gene variants needs to be investigated further.

Regulation of Plant Seed Vigor Through Manipulation of RFOs.

We and other groups have confirmed that overexpression of *GOLS* genes enhanced the abiotic stress tolerance of transgenic plants (Gu et al., 2016; Shimosaka and Ozawa, 2015; Sun et al., 2013). It is worth mentioning that constitutive expression of *ZmGOLS2* genes in Arabidopsis did not cause any adverse effects to the plant under normal conditions (Gu et al., 2016). These findings imply that it may be feasible to manipulate RFOs metabolism in crop plants to enhance seed vigor, but while feasible, the mechanism imparting greater tolerance is not well understood (i.e. strictly speaking, protection may not result due to greater RFOs amounts if their ratio to sucrose amounts remains the same). We are currently working on overexpression of *AtSTS* (*AtRS4*), *ZmGOLS2/ZmRS, ZmGOLS2/AtSTS* in maize plants to see if over production of raffinose, or generation of some stachyose in maize seed would increase maize seed vigor.

RFOs can not be digested by monogastric animals and thus, are considered to be anti-nutritional. Hence, there are efforts focused on reducing the RFOs content in crop seeds (Dierking and Bilyeu, 2008; Yang et al., 2015). By contrast, RFOs are an important energy resource for beneficial microflora in the small intestine (Grmanova et al., 2010; Rada et al., 2008). Understanding the RFOs metabolic pathway and the physiological function of these oligosaccharides would assist researchers to modulate RFOs in crop plants for these multifarious goals.

METHODS

Identification of Maize RAFFINOSE SYNTHASE.

To identify the maize RAFFINOSE SYNTHASE, four protein sequences known to have exhibited raffinose synthetic activity *in vitro* (AtRS5; CsRS; PsRS; and OsRS) were used as queries to blast against the Maize GDB (http://www.maizegdb.org).

Phylogenetic Analysis

To determine the phylogeny of the putative ZmRS (GRMZM2G150906-P01) in the

galactosyl-transferase family and 38 additional protein sequences including 17 AGAs (PLANT ALKALINE α-GALACTOSIDASE), 12 AGALs (EUKARYOTIC-α-GALACTOSIDASE), 5 STSs (STACHYOSE SYNTHASE) and 4 RSs (RAFFINOSE SYNTHASE) were aligned using CLUSTAL W (Thompson et al., 1994) (Supplemental dataset 1) for input to MEGA5 (Figure 1A) (Tamura et al., 2011). The sequence-level similarities among these enzymes were computed according to the p-distance (Nei and Kumar, 2000) and their evolutionary relationship was inferred using the Neighbor-Joining method (Saitou and Nei, 1987), and the robustness of the resultant tree was indicated as bootstrap values from 1000 iterations on the tree branches. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The tree was drawn to scale, with branch lengths proportional to the associated evolutionary divergence. Overall, the tree in Figure 1A was generated based on 39 protein sequences, being 79-aa length consistently since all positions containing gaps and missing data of their original sequences were eliminated from the dataset.

Vector Construction.

For testing the enzyme activity of ZmRS, a prokaryotic expression vector was constructed by amplifying the *ZmRS* ORF from RT-PCR using a pair of primers (ZmRS-CRF-BamHI and ZmRS-CRR-HindIII; Supplemental Table 2) and cDNA synthesized from maize leaf RNA. Using the BamHI and HindIII designed in the primers, the amplicons were then digested, gel purified and directionally ligated into *pET-21d* vector (Takara, Japan).

For construction of AtSTS bacterial expression vector, primers AtRS4-cF and AtRS4-cR1 were used for RT-PCR amplification of the coding region of AtSTS from cDNA synthesized from Arabidopsis leaf RNA. Using the SacI-NotI designed in the primers, the amplicons were then digested, gel purified and directionally ligated into *pET-21d* vector (Takara, Japan).

For construction of the expression vector for ΔZmRS protein (ZmRS with the α-amylase catalytic domain of AtSTS, (AC) inserted), overlap extension PCR was applied. In addition to amplifying the AC domain from the *AtSTS* bacteria expression vector with ZmRS-ADD-F2 and ZmRS-ADD-R2 primers, the two flanking *ZmRS* amplicons were made with the complementary AC primers

(ZmRS-ADD-R1 and ZmRS-ADD-F2, respectively) and 5'-ZmRS-CRF-BamHI and 3'-ZmRS-CRR-Hind□, respectively, from the *ZmRS* bacterial expression vector. A final PCR reaction with all 3 amplicons and the 2 restriction-enzyme-site-containing primers, produced a 2510bp ΔZmRS CDS with the AC of AtSTS inserted between nucleotides 936 and 937 for ligation into *pET-21d*.

For construction of the expression vector for the ΔAtSTS protein (*AtSTS* without the α-amylase catalytic domain (AC)), primers AtRS4-cF and DeleteR1 were used for PCR amplification of the 5'coding region of *AtSTS*. Primers DeleteF2 and AtRS4-cR1 were used for PCR amplification of the 3'coding region of *AtSTS*. The two PCR amplicons were used as templates for overlap extension PCR to obtain *ΔAtSTS*. The PCR product of *ΔAtSTS* (2394 bp) was then cloned into SacI-NotI site of *pET-21d* vector.

For construction of the *ZmRS* expression vector for Arabidopsis transformation, the *ZmRS* CDS was amplified by PCR from the prokaryotic *pET-ZmRS* vector (above) using primers OXZmRS-CRF-BamHI and OXZmRS-CRR-XbaI prior to digestion and ligation into *pCSGFPBT* vector where expression is driven by the Cauliflower mosaic virus 35S promoter (Gu et al., 2016). Similarly, the *AtSTS* genomic DNA sequence was amplified by PCR from Arabidopsis genomic DNA (Col-0) using OEAtSTS-CRF-NcoI and OEAtSTS-CRR-SpeI primers for subsequent ligation into *Nco*I-*Spe*I sites of *pCAMBIA1303* vector where it was also expressed under the control of the Cauliflower mosaic virus 35S promoter.

For construction of the *ZmRS* expression vector used for generation of ZmGOLS2/ZmRS double expressing Arabidopsis line, the *ZmRS* ORF was amplified by PCR from *pET-ZmRS* vector using primer pair (OEZmRS-double-F and OEZmRS-double-R) and cloned into *BamH*I-*Spe*I sites of pBI111L (Shao et al., 2012).

Enzyme Activity Assay of ZmRS

The ZmRS, ΔZmRS, AtSTS or ΔAtSTS expression vector was transformed into *Escherichia coli* (Rosetta gami2, DE3; EMD Millipore) cells. Bacterial cultures were grown to $OD_{600\ nm}$=0.6 at 37°C, then supplemented with or without 0.1 mM isopropyl β-D-thiogalactoside (IPTG) and growth

continued overnight at 25°C. *E. coli* cells were then collected and lysed in PBS buffer (50 mM, containing 150 mM NaCl, pH 7.4) using an ultrasonic cell disruptor. The lysate were then centrifuged at 15,000×g at 4°C for 20 min.

Raffinose synthetic assays were 5.3 mM galactinol (Sigma, USA), and 2.8 mM sucrose; galactinol hydrolytic activity assays were 10.6 mM galactinol; stachyose synthetic activity assays were 5.3 mM galactinol, 16.8 mM raffinose; while raffinose hydrolytic activity assays were 16.8 mM raffinose. All reactions were performed in a 50 μL reaction system (25 μL 2x stock buffer and 25 μL crude extract) containing 25 mM HEPES-KOH (pH 7.0).

The reaction mixture for ZmRS activity or AtSTS activity assay was respectively incubated at 37 □ or 25 □ for two h, then 500 μL of 80% (v/v) ethanol was added to stop the reaction. The mixture was boiled for 10 min, centrifuged at 15, 000×g at room temperature for 10 min. The supernatant was then diluted 5 fold with water, frozen and lyophilized to dryness under vacuum. The dry powder was reconstituted in 100 μL of double distilled water and stored at -80 □ until HPLC analysis.

Identification of *Mu*-inserted *zmrs* Mutant Plants

Maize (*Zea mays* L.) inbred line B73 was maintained in the lab. The seeds of the *zmrs-1* mutant and its W22 background were obtained from the Maize Genetics Cooperation Stock Center using the UniformMu Transposon Resource (http://www.maizegdb.org/uniformmu; (Settles et al., 2007) for identification of UFMu-09411 using the *ZmRS* gene model ID (GRMZM2G150906) as the query. The *zmrs-2* mutant seeds in the B73 background were obtained from the Barkan lab (Williams-Carrier et al., 2010). All the maize lines used in this study were outcrossed twice at the Northwest A&F University, China to their respective WT and then selfed to recover homozygous mutants and counterpart Null Segregants (NS) which were used as controls. Genomic DNA was isolated from maize leaves using the CTAB method (Porebski et al., 1997). PCR was performed to characterize the genotype of the plants using *ZmRS* gene-specific primers (F1a, F2a, F3a, R1a, R2a, R3a, F1b, F2b, F3b, F4b, R1b, R2b; Supplemental Table 2) and *mutator*-anchor primer TIR. These primers were also used for PCR or RT-PCR analysis of *ZmRS* gene expression.

Plant Material and Growth Conditions

Arabidopsis thaliana ecotype Columbia plants were transformed by floral dip according to the vacuum infiltration method using *Agrobacterium tumefaciens* strain GV3101 (Clough and Bent, 1998). Arabidopsis were grown in a culture room which was set for a 16 h photoperiod at minimum of 200 $\mu mol \cdot m^{-2}\ s^{-1}$ and a day/night temperature of 20ºC/18 ºC. Seeds of sorghum (*Sorghum bicolor*) and foxtail millet (*Setaria italica*) were provided by Dr. Junfeng Zhao (Millet Research Institute, Shanxi Academy of Agricultural Sciences). Seeds of teosinte (*Zea mays* ssp. Mexicana) were provided by Dr. Dongwei Guo and seeds of rice (*Oryza sativa* spp. japonica) were provided by Dr. Kunming Chen (Northwest A&F University).

Sample Collection

For each genotype, seeds were gathered from separate cobs/siliques and combined before analysis. For tissue specific expression analysis of the *ZmRS* gene, different tissues were harvested from mature plants of B73 grown in the field. The developing seeds were detached from ears at the 14, 22, 30, 35, 40, 45, 50, 55 days after control pollination (DAP), then the seeds were separated into two parts (embryo and endosperm, without episperm). Corn seed imbibition was performed on wet filter paper in a plastic box at 28ºC in darkness. Seed embryos and endosperms were collected at 0, 12, 24, 36, 48, 60 hours after imbibition (HAI).

Accelerated aging Treatment of Seeds

High temperature and high moisture conditions were used to evaluate the vigor and storability of mature seeds of corn and Arabidopsis (Rajjou et al., 2008). The *zmrs* seeds and their NS siblings were sterilized in 10% (v/v) NaClO for 5 min and then washed with deionized water three times. The seeds were dried on filter paper for 6 h and then packaged in nylon fabric. Three replications of 25 maize seeds each were then tested for germination (no accelerated aging; NAA) or incubated in a hermetic dryer at 98% relative humidity (RH) (controlled by saturated K_2SO_4 solution) at 42ºC for 3 or 6 days. After this aging treatment, the seeds were desiccated at room temperature for 24 h. Seed germination was performed on wet filter paper in a plastic box at 28ºC in darkness. The number of seeds completing germination were counted every 12 h and the shoot/root length was measured after

620 imbibition for 120 h.

621 Three replications of 50 Arabidopsis seeds each were tested for seed germination (no accelerated
622 aging; NAA) or placed in a hermetic dryer at 83-85% RH (controlling by saturated KCl solution) at
623 42ºC for 3 days. After such aging treatment, seeds were desiccated at room temperature for 24 h. An
624 aliquot of these seeds were tested for viability (see **Tetrazolium Assay of the Arabidopsis Seed**
625 **Viability**). Other seeds were surface-sterilized with 70% (v/v) ethanol for 30 s, followed by washing
626 with 2.6% (v/v) NaClO containing 0.03% (v/v) Tween-20 for 5 min, and then 3 replications of 50
627 seeds each were sown on GM plates for germination at 22ºC. Germination was determined every 12
628 h.

629 **RNA Extraction, RT-PCR, and Real Time RT-PCR**

630 Total RNA was extracted from embryos, endosperms and other tissues of maize using a published
631 protocol (Jaakola et al., 2001). TRIzol Reagent (Takara, Japan) was used for extraction of RNA from
632 Arabidopsis leaves. The Total RNAs were digested with DNase□. The RNA concentrations were
633 determined by a Nanodrop 200 spectrophotometer. The cDNA was obtained using the Transcriptor
634 1st Strand cDNA Synthesis kit (Roche, Switzerland) and then diluted 50 fold for RT-PCR and
635 RT-qPCR template. Real-time RT-PCR was performed using Fast Start Essential DNA Green master
636 Mix (Roche, Switzerland) in a CFX96 Touch™ system (Bio-Rad, USA). The expression of tested
637 genes was normalized to that of constitutively expressed genes such as *ACTIN2* in Arabidopsis and
638 *GAPDH* in maize. The experiments were repeated at least three times with independent biological
639 samples. The RT-PCR and real time RT-PCR reaction was performed with the primers listed in
640 Supplemental Table 2.

641 **Western Blot Analysis of ZmRS and AtSTS Protein Expression**

642 Polyclonal antibody against RS or STS was generated in immunized rabbits. ZmRS/AtSTS-His6
643 fusion protein expressed in *E. coli* Rosetta gami 2 (DE3) cells was purified using an electrodialysis
644 method. Crude extract were separated by SDS-PAGE, then the recombinant protein strips was cut off
645 and soaked in Tris-Glycine buffer (25 mM Tris pH 12.5, 250 mM glycine and 0.5% w/v SDS) in
646 dialysis bags (8 kD-14 kD, Solarbio, China). The rabbits were immunized every 14 days on four

different occasions. For the first immunization, the 1:1 (v/v) mixture of Freund's complete adjuvant (Sigma, USA) and purified protein (150 μg) was used to inject the rabbit. For the following immunizations, an equal volume mixture of Freund's incomplete adjuvant (Sigma, USA) and purified protein were used to inject the rabbit. Serum was extracted two weeks following the final boost. The polyclonal antibody for ZmGAPDH was purchased from CWBIO (China).

Western blot analysis of ZmRS protein expression in maize leaves, seed embryos, and ZmRS/AtSTS protein expression in transgenic Arabidopsis leaves was performed following a published protocol (Gu et al., 2016). Protein detection used a Western Bright™ ECL Kit (Advansta, USA).

Soluble carbohydrate Extraction, Analysis of Total Soluble Sugar, HPLC-ELSD Analysis of Sugar Components, LC-MS/MS Validation of Verbascose Presence

Soluble sugar extraction followed a published protocol with minor revision (Zhao et al., 2004). Maize tissues (0.5 g for leaves or endosperms, 0.1 g for embryos) were grounded into powder in liquid nitrogen. Five aliquots of 1 mL of 80% (v/v) ethanol containing 200 $\mu g \cdot mL^{-1}$ lactose was added and homogenized to a slurry. Another 2 mL of 80% ethanol was used to wash the mortar. Arabidopsis seeds (0.2 g) were grounded in liquid nitrogen. Then 3 mL of 80% (v/v) ethanol containing 200 $\mu g \cdot mL^{-1}$ lactose was added and homogenized to a slurry, Another 4 mL of 80% ethanol was used to wash the mortar. The suspensions were heated at 80□ for 30 min, then centrifuged at 16,000×g to collect the supernatants. The tubes containing sugar extracts were opened and incubated in a water bath at 95□ until the ethanol was evaporated. The remaining sugar solution (about 500 μL) was diluted with 5-fold volumes of water and lyophilized to dryness under a vacuum. The dry powder was reconstituted in 1 mL double distilled water and stored at -80□. A Waters X-bridge amide column (Waters, USA) was washed by methanol : H_2O (90:10) as the mobile phase at speed of 0.5 $mL \cdot min^{-1}$ for separation of soluble sugar components. An evaporative light-scattering detector (ELSD, Waters 2424) was applied to monitor the sugar signal.

An Arabidopsis mature, dehydrated seed sugar extract and a standard solution containing verbascose were analyzed (in that order to avoid residual verbascose contamination i.e. "Ghost peaks") by hydrophilic interaction chromatography (HILIC) on a Waters Acquity UPLC coupled to a Waters

Synapt G2 (q-ToF) mass spectrometer. Chromatographic separation was obtained using a Waters BEH Amide UPLC column (1.7um, 2.1mm x 150mm; 30°C). The mobile phase employed a mixture of water (Fisher Optima) containing 0.1% formic acid (solvent A; Fisher) and acetonitrile (Fisher Optima) containing 0.1% formic acid (solvent B) in a linear gradient from 80% B to 50% B at a flow rate of 0.35mL•min^{-1}. The high resolution mass spectrometer was operated in negative ion electrospray mode with a resolving power of ~14,000 and scanned from 100 to 1000 Da in 0.3 s. Leucine enkephalin was used to provide a lock mass (m/z 554.2615).

Tetrazolium Assay of the Arabidopsis Seed Viability

Tetrazolium assay was performed following a published protocol with modification (Salvi et al., 2016). Seeds treated with or without artificial aging were initially sterilized with 10% hypochlorous acid (containing 0.03% Tween-20) for 5 mins and then washed by sterilized distilled water five times. The seeds were then incubated in 1% tetrazolium solution (50 mM phosphate buffer, pH-7.0) in darkness at 30□ for 48 h. Seeds were washed by sterilized distilled water three times after staining and then were immersed in clearing agent (mix lactic acid: phenol: glycerine: water in a ratio of 1:1:2:1) for 4 h to remove the seed coat pigments. Seed viability was determined by the staining intensity of red 2, 3, 5 triphenyl formazan which was generated from the reduction of the tetrazolium by dehydrogenases in the live cells.

Statistical Analysis

When experiments had equal numbers of replications an analysis of variance was used to determine statistically significant deviations among average values. Otherwise (e.g. coleoptile lengths from seedlings after seed AA when different numbers of seeds complete germination) a general linear model was used. In either case, if the test was significant (α=0.05) Tukey's multiple comparison test was used to identify significantly deviating means. The Pearson correlation analysis was conducted with the averaged absolute amount of each individual sugar ($\mu g \cdot mg^{-1}$ DW) and the seed AA tolerance (relative to WT) (n=18). The resultant Pearson correlation coefficient was analyzed using two tailed method (IBM SPSS Statistics 19). The dualistic linear regression analysis was performed with IBM SPSS Statistics 19. The data were graphed using SigmaPlot 10.

ACCESSION NUMBERS

Sequence data from this article can be found in GenBank under the following accession numbers: *ZmRS:* BT063253, *ZmGOLS2*: AF497509, *AtRS5*: NM_123403, *AtSTS*: NM_116428, *AtACTIN2*: NM_112764 and *ZmGAPDH*: XM_008679567.

Accession numbers for sequences used to construct the tree in Figure 1A are as follows: CsRS (*Cucumis sativus*; E15707); AtRS5 (*Arabidopsis thaliana*; NP_198855); PsRS (*Pisum sativum*; CAD20127); OsRS (*Oryza sativa*; XP_015621501); PsSTS (*Pisum sativum*; CAC38094); VaSTS (*Vigna angularis*; CAB64363); AmSTS (*Alonsoa meridionalis*; CAD31704); SaSTS (*Stachys affinis*; CAC86963); AtSTS (*Arabidopsis thaliana*; NP_192106); AtSIP1 (*Arabidopsis thaliana*; NP_175970); LeAGA1 (*Lycopersicon esculentum*; AAN32954); CmAGA1 (*Cucumis melo*; AAM75139); ZmAGA3 (*Zea mays*; AAQ07253); OsAGA1 (*Oryza sativa*; XP_483143); ZmAGA1 (*Zea mays*; AAQ07251); OsSIP1 (*Oryza sativa*; XP_477103); HvSIP1 (*Hordeum vulgare*; Q40077); CmAGA2 (*Cucumis melo*; AAM75140); PaSIP1 (*Persea americana*;CAB77245); AtSIP2 (*Arabidopsis thaliana*; NP_191311); BoSIP1 (*Brassica oleracea*; CAA55893); ZmSIP2 (*Zea mays*; AAQ07252); AtSIP3 (*Arabidopsis thaliana*;NP_001190347) ; SsSIP1 (*Sulfolobus solfataricus*; AAK43227); ZmRS7 (*Zea mays*; XP_008669826); ZmRS2 (*Zea mays*; ONM02661); ZmRS3 (*Zea mays*; XP_008665643); UvAGAL1 (*Umbelopsis vinacea*; BAA33931); CaAGAL1 (*Coffea arabica*; Q42656); GmAGAL1 (*Glycine max*; AAA73963); CtAGAL1 (*Cyamopsis tetragonoloba*; P14749); AtAGAL2 (*Arabidopsis thaliana; NP_001031855*); LeAGAL (*Lycopersicon esculentum*; AAF04591); AtAGAL1 (*Arabidopsis thaliana*; NP_191190); AnAGAL1 (*Aspergillus niger*; CAB46229); ScAGAL1 (*Saccharomyces cerevisiae*; P04824); HsAGAL1 (*Homo sapiens*; P06280); ZmRS5 (*Zea mays*; XP_008671237).

AUTHOR CONTRIBUTIONS

T.L., Y.Z., D.W., Y.L., L.D., B.D., J.G., and J.W. performed research, T.L., L.D., J.G., and B.D. analyzed the data, T.Z., B.D., and G.W. conceived the experiments and wrote the article.

ACKNOWLEDGEMENTS

We acknowledge Drs. Ruolin Yang, Hongchang Cui from Northwest A&F University for useful

discussions about the manuscript. This research was funded by the Special fund for transgenic research from Ministry of Agriculture in China (2014ZX0800920B) and the NSFC (31671776) to T.Z. We wish to thank the Maize Genetics COOP Stock Center for providing the maize mutants and the Arabidopsis Biological Resource Center for the Arabidopsis mutants. Mr. Leandro Reis, during a summer internship sponsored by the Brazilian Scientific Mobility Program, identified one of the maize Mu insertion flanks for which we are grateful. Prof Glen Aiken, USDA-FAPRU, University of Kentucky, kindly provided access to the Waters Synapt G2 (q-ToF) mass spectrometer. The authors declare no conflicts of interest.

COMPETING FINANCIAL INTERESTS

We are not aware of any competing financial interests.

REFERENCES

Amiard, V., Morvan-Bertrand, A., Billard, J.P., Huault, C., Keller, F., and Prud'homme, M.P. (2003). Fructans, but not the sucrosyl-galactosides, raffinose and loliose, are affected by drought stress in perennial ryegrass. Plant Physiol 132:2218-2229.

Antonio, C., Larson, T., Gilday, A., Graham, I., Bergström, E., and Thomas-Oates, J. (2008). Hydrophilic interaction chromatography/electrospray mass spectrometry analysis of carbohydrate-related metabolites from Arabidopsis thaliana leaf tissue. Rapid Communications in Mass Spectrometry 22:1399-1407.

Bachmann, M., and Keller, F. (1995). Metabolism of the Raffinose Family Oligosaccharides in Leaves of Ajuga-Reptans L - Intercellular and Intracellular Compartmentation. Plant Physiol 109:991-998.

Bentsink, L., Alonso-Blanco, C., Vreugdenhil, D., Tesnier, K., Groot, S.P.C., and Koornneef, M. (2000). Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of Arabidopsis. Plant Physiol 124:1595-1604.

Bernal-Lugo, I., and Leopold, A.C. (1992). Changes in Soluble Carbohydrates during Seed Storage. Plant Physiol 98:1207-1210.

Blackman, S.A., Obendorf, R.L., and Leopold, A.C. (1992). Maturation Proteins and Sugars in Desiccation Tolerance of Developing Soybean Seeds. Plant Physiol 100:225-230.

Blochl, A., Peterbauer, T., and Richter, A. (2007). Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. J Plant Physiol 164:1093-1096.

Brenac, P., Horbowicz, M., Downer, S.M., Dickerman, A.M., Smith, M.E., and Obendorf, R.L. (1997a). Raffinose accumulation related to desiccation tolerance during maize (Zea mays L) seed development and maturation. J Plant Physiol 150:481-488.

Brenac, P., Smith, M.E., and Obendorf, R.L. (1997b). Raffinose accumulation in maize embryos in the absence of a fully functional Vp1 gene product. Planta 203:222-228.

Bueso, E., Munoz-Bertomeu, J., Campos, F., Brunaud, V., Martinez, L., Sayas, E., Ballester, P., Yenush, L., and Serrano, R. (2014). ARABIDOPSIS THALIANA HOMEOBOX25 uncovers a role for Gibberellins in seed longevity. Plant physiology 164:999-1010.

Buitink, J., Leprince, O., Hemminga, M.A., and Hoekstra, F.A. (2000). Molecular mobility in the cytoplasm: An approach to describe and predict lifespan of dry germplasm. P Natl Acad Sci USA 97:2385-2390.

Carmi, N., Zhang, G., Petreikov, M., Gao, Z., Eyal, Y., Granot, D., and Schaffer, A.A. (2003). Cloning and functional expression of alkaline alpha-galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. The Plant journal : for cell and molecular biology 33:97-106.

Cheng, X., Cheng, J., Huang, X., Lai, Y., Wang, L., Du, W., Wang, Z., and Zhang, H. (2013). Dynamic quantitative trait loci analysis of seed reserve utilization during three germination stages in rice. PLoS One 8:e80002.

770 Cho, S.M., Kang, E.Y., Kim, M.S., Yoo, S.J., Im, Y.J., Kim, Y.C., Yang, K.Y., Kim, K.Y., Kim, K.S., Choi, Y.S., et al. (2010).
771 Jasmonate-dependent expression of a galactinol synthase gene is involved in priming of systemic fungal resistance
772 in Arabidopsis thaliana. Botany 88:452-461.
773 Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of
774 Arabidopsis thaliana. Plant J 16:735-743.
775 Cui, H., Peng, B., Xing, Z., Xu, G., Yu, B., and Zhang, Q. (2002). Molecular dissection of seedling-vigor and associated
776 physiological traits in rice. Theor Appl Genet 105:745-753.
777 Dargahi, H., Tanya, P., and Srinives, P. (2014). Mapping of the genomic regions controlling seed storability in soybean
778 (Glycine max L.). J Genet 93:365-370.
779 de Souza Vidigal, D., Willems, L., van Arkel, J., Dekkers, B.J., Hilhorst, H.W., and Bentsink, L. (2016). Galactinol as marker
780 for seed longevity. Plant Sci 246:112-118.
781 Dierking, E.C., and Bilyeu, K.D. (2008). Association of a Soybean Raffinose Synthase Gene with Low Raffinose and
782 Stachyose Seed Phenotype. Plant Genome-Us 1:135-145.
783 Dierking, E.C., and Bilyeu, K.D. (2009). Raffinose and stachyose metabolism are not required for efficient soybean seed
784 germination. J Plant Physiol 166:1329-1335.
785 Downie, B., Gurusinghe, S., Dahal, P., Thacker, R.R., Snyder, J.C., Nonogaki, H., Yim, K., Fukanaga, K., Alvarado, V., and
786 Bradford, K.J. (2003). Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before
787 maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiol
788 131:1347-1359.
789 Egert, A., Eicher, B., Keller, F., and Peters, S. (2015). Evidence for water deficit-induced mass increases of raffinose family
790 oligosaccharides (RFOs) in the leaves of three Craterostigma resurrection plant species. Frontiers in physiology
791 6:206.
792 Egert, A., Keller, F., and Peters, S. (2013). Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is
793 mediated by a single raffinose synthase (RS5, At5g40390). Bmc Plant Biol 13.
794 Gangl, R., Behmuller, R., and Tenhaken, R. (2015). Molecular cloning of AtRS4, a seed specific multifunctional RFO
795 synthase/galactosylhydrolase in Arabidopsis thaliana. Front Plant Sci 6:789.
796 Gangl, R., and Tenhaken, R. (2016). Raffinose Family Oligosaccharides Act As Galactose Stores in Seeds and Are Required
797 for Rapid Germination of Arabidopsis in the Dark. Front Plant Sci 7:1115.
798 Grmanova, M., Rada, V., Sirotek, K., and Vlkova, E. (2010). Naturally occurring prebiotic oligosaccharides in poultry feed
799 mixtures. Folia Microbiol (Praha) 55:326-328.
800 Gu, L., Zhang, Y., Zhang, M., Li, T., Dirk, L.M., Downie, B., and Zhao, T. (2016). ZmGOLS2, a target of transcription factor
801 ZmDREB2A, offers similar protection against abiotic stress as ZmDREB2A. Plant molecular biology 90:157-170.
802 Gurusinghe, S., and Bradford, K.J. (2001). Galactosyl-sucrose oligosaccharides and potential longevity of primed seeds.
803 Seed Sci Res 11:121-133.
804 Han, Z., Ku, L., Zhang, Z., Zhang, J., Guo, S., Liu, H., Zhao, R., Ren, Z., Zhang, L., Su, H., et al. (2014). QTLs for seed
805 vigor-related traits identified in maize seeds germinated under artificial aging conditions. PLoS One 9:e92535.
806 Himuro, Y., Ishiyama, K., Mori, F., Gondo, T., Takahashi, F., Shinozaki, K., Kobayashi, M., and Akashi, R. (2014). Arabidopsis
807 galactinol synthase AtGolS2 improves drought tolerance in the monocot model Brachypodium distachyon. J Plant
808 Physiol 171:1127-1131.
809 Hincha, D.K., Zuther, E., and Heyer, A.G. (2003). The preservation of liposomes by raffinose family oligosaccharides
810 during drying is mediated by effects on fusion and lipid phase transitions. Biochim Biophys Acta 1612:172-177.
811 Horbowicz, M., and Obendorf, R.L. (1994). Seed desiccation tolerance and storability: dependence on
812 flatulence-producing oligosaccharides and cyclitols--review and survey. Seed Science Research 4:385-405.
813 Jaakola, L., Pirttila, A.M., Halonen, M., and Hohtola, A. (2001). Isolation of high quality RNA from bilberry (Vaccinium
814 myrtillus L.) fruit. Mol Biotechnol 19:201-203.
815 Janecek, S., Lanta, V., Klimesova, J., and Dolezal, J. (2011). Effect of abandonment and plant classification on
816 carbohydrate reserves of meadow plants. Plant biology 13:243-251.
817 Kim, H.S., Cha, E., Kim, Y., Jeon, Y.H., Olson, B.H., Byun, Y., and Park, H.D. (2016). Raffinose, a plant galactoside, inhibits
818 Pseudomonas aeruginosa biofilm formation via binding to LecA and decreasing cellular cyclic diguanylate levels. Sci
819 Rep-Uk 6.
820 Koster, K.L., and Leopold, A.C. (1988). Sugars and desiccation tolerance in seeds. Plant Physiol 88:829-832.
821 Kuo, T.M., VanMiddlesworth, J.F., and Wolf, W.J. (1988). Content of Raffinose Oligosaccharides and Sucrose in Various
822 Plant Seeds. J. Agric. Food Chem 36:32-36.
823 Lahuta, L.B., Pluskota, W.E., Stelmaszewska, J., and Szablinska, J. (2014). Dehydration induces expression of GALACTINOL
824 SYNTHASE and RAFFINOSE SYNTHASE in seedlings of pea (Pisum sativum L.). J Plant Physiol 171:1306-1314.
825 Leinen, K.M., and Labuza, T.P. (2006). Crystallization inhibition of an amorphous sucrose system using raffinose. J
826 Zhejiang Univ Sci B 7:85-89.

827 Li, S.H., Li, T.P., Kim, W.D., Kitaoka, M., Yoshida, S., Nakajima, M., and Kobayashi, H. (2007). Characterization of raffinose
828 synthase from rice (Oryza sativa L. var. Nipponbare). Biotechnol Lett 29:635-640.
829 Liu, L., Lai, Y., Cheng, J., Wang, L., Du, W., Wang, Z., and Zhang, H. (2014). Dynamic quantitative trait locus analysis of
830 seed vigor at three maturity stages in rice. PloS one 9:e115732.
831 Martinez-Villaluenga, C., Frias, J., and Vidal-Valverde, C. (2008). Alpha-galactosides: Antinutritional factors or functional
832 ingredients? Crit Rev Food Sci 48:301-316.
833 Nei, M., and Kumar, S. (2000). Molecular evolution and phylogenetics: Oxford University Press.
834 Nishizawa, A., Yabuta, Y., and Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants
835 from oxidative damage[Plant Physiol 147:1251-1263.
836 Obata, T., Witt, S., Lisec, J., Palacios-Rojas, N., Florez-Sarasa, I., Yousfi, S., Araus, J.L., Cairns, J.E., and Fernie, A.R. (2015).
837 Metabolite Profiles of Maize Leaves in Drought, Heat, and Combined Stress Field Trials Reveal the Relationship
838 between Metabolism and Grain Yield. Plant Physiol 169:2665-2683.
839 Obendorf, R.L. (1997). Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. Seed Sci Res 7:63-74.
840 Panikulangara, T.J., Eggers-Schumacher, G., Wunderlich, M., Stransky, H., and Schoffl, F. (2004). Galactinol synthase1. A
841 novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in
842 arabidopsis. Plant Physiol 136:3148-3158.
843 Peterbauer, T., Karner, U., Mucha, J., Mach, L., Jones, D.A., Hedley, C.L., and Richter, A. (2003). Enzymatic control of the
844 accumulation of verbascose in pea seeds. Plant Cell Environ 26:1385-1391.
845 Peterbauer, T., Mach, L., Mucha, J., and Richter, A. (2002). Functional expression of a cDNA encoding pea (Pisum sativum
846 L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of
847 raffinose synthesis. Planta 215:839-846.
848 Peterbauer, T., and Richter, A. (1998). Galactosylononitol and stachyose synthesis in seeds of adzuki bean - Purification
849 and characterization of stachyose synthase. Plant Physiol 117:165-172.
850 Peterbauer, T., and Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl
851 cyclitols in seeds. Seed Sci Res 11:185-197.
852 Peters, S., Egert, A., Stieger, B., and Keller, F. (2010). Functional Identification of Arabidopsis ATSIP2 (At3g57520) as an
853 Alkaline alpha-Galactosidase with a Substrate Specificity for Raffinose and an Apparent Sink-Specific Expression
854 Pattern. Plant Cell Physiol 51:1815-1819.
855 Porebski, S., Bailey, L.G., and Baum, B.R. (1997). Modification of a CTAB DNA extraction protocol for plants containing
856 high polysaccharide and polyphenol components. Plant Mol Biol Rep 15:8-15.
857 Rada, V., Nevoral, J., Trojanova, I., Tomankova, E., Smehilova, M., and Killer, J. (2008). Growth of infant faecal
858 bifidobacteria and clostridia on prebiotic oligosaccharides in in vitro conditions. Anaerobe 14:205-208.
859 Rajjou, L., Lovigny, Y., Groot, S.P., Belghazi, M., Job, C., and Job, D. (2008). Proteome-wide characterization of seed aging
860 in Arabidopsis: a comparison between artificial and natural aging protocols. Plant physiology 148:620-641.
861 Righetti, K., Vu, J.L., Pelletier, S., Vu, B.L., Glaab, E., Lalanne, D., Pasha, A., Patel, R.V., Provart, N.J., Verdier, J., et al. (2015).
862 Inference of Longevity-Related Genes from a Robust Coexpression Network of Seed Maturation Identifies
863 Regulators Linking Seed Storability to Biotic Defense-Related Pathways. Plant Cell 27:2692-2708.
864 Ruan, Y.L. (2014). Sucrose metabolism: gateway to diverse carbon use and sugar signaling. Annual review of plant
865 biology 65:33-67.
866 Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol
867 Biol Evol 4:406-425.
868 Salvi, P., Saxena, S.C., Petla, B.P., Kamble, N.U., Kaur, H., Verma, P., Rao, V., Ghosh, S., and Majee, M. (2016). Differentially
869 expressed galactinol synthase(s) in chickpea are implicated in seed vigor and longevity by limiting the age induced
870 ROS accumulation. Scientific reports 6:35088.
871 Saravitz, D.M., Pharr, D.M., and Carter, T.E. (1987). Galactinol synthase activity and soluble sugars in developing seeds of
872 four soybean genotypes. Plant physiology 83:185-189.
873 Sengupta, S., Mukherjee, S., Basak, P., and Majumder, A.L. (2015). Significance of galactinol and raffinose family
874 oligosaccharide synthesis in plants. Front Plant Sci 6:656.
875 Settles, A.M., Holding, D.R., Tan, B.C., Latshaw, S.P., Liu, J., Suzuki, M., Li, L., O'Brien, B.A., Fajardo, D.S., Wroclawska, E.,
876 et al. (2007). Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. BMC
877 Genomics 8:116.
878 Shao, J., Liu, X., Wang, R., Zhang, G., and Yu, F. (2012). The over-expression of an Arabidopsis B3 transcription factor,
879 ABS2/NGAL1, leads to the loss of flower petals. PLoS One 7:e49861.
880 Shimosaka, E., and Ozawa, K. (2015). Overexpression of cold-inducible wheat galactinol synthase confers tolerance to
881 chilling stress in transgenic rice. Breeding science 65:363-371.
882 Sui, X.L., Meng, F.Z., Wang, H.Y., Wei, Y.X., Li, R.F., Wang, Z.Y., Hu, L.P., Wang, S.H., and Zhang, Z.X. (2012). Molecular
883 cloning, characteristics and low temperature response of raffinose synthase gene in Cucumis sativus L. J Plant
884 Physiol 169:1883-1891.

Sun, W.Q., and Leopold, A.C. (1997). Cytoplasmic vitrification acid survival of anhydrobiotic organisms. Comp Biochem Phys A 117:327-333.

Sun, Z.B., Qi, X.Y., Wang, Z.L., Li, P.H., Wu, C.X., Zhang, H., and Zhao, Y.X. (2013). Overexpression of TsGOLS2, a galactinol synthase, in Arabidopsis thaliana enhances tolerance to high salinity and osmotic stresses. Plant Physiol Bioch 69:82-89.

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2002). Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J 29:417-426.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739.

Tapernoux-Luthi, E.M., Bohm, A., and Keller, F. (2004). Cloning, functional expression, and characterization of the raffinose oligosaccharide chain elongation enzyme, galactan:galactan galactosyltransferase, from common bugle leaves. Plant Physiol 134:1377-1387.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680.

Vandecasteele, C., Teulat-Merah, B., Morere-Le Paven, M.C., Leprince, O., Vu, B.L., Viau, L., Ledroit, L., Pelletier, S., Payet, N., Satour, P., et al. (2011). Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in Medicago truncatula. Plant Cell Environ 34:1473-1487.

Wang, B., Zhang, Z., Fu, Z., Liu, Z., Hu, Y., and Tang, J. (2016). Comparative QTL analysis of maize seed artificial aging between an immortalized F_2 population and its corresponding RILs. The Crop Journal 4:30-39.

Wendorf, J.R., Radke, C.J., and Blanch, H.W. (2004). Reduced protein adsorption at solid interfaces by sugar excipients. Biotechnol Bioeng 87:565-573.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D.B., Monde, R.A., Coalter, R., and Barkan, A. (2010). Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. Plant J 63:167-177.

Xie, L., Tan, Z., Zhou, Y., Xu, R., Feng, L., Xing, Y., and Qi, X. (2014). Identification and fine mapping of quantitative trait loci for seed vigor in germination and seedling establishment in rice. Journal of integrative plant biology 56:749-759.

Yang, W., Zhang, Y., Zhou, X., Zhang, W., Xu, X., Chen, R., Meng, Q., Yuan, J., Yang, P., and Yao, B. (2015). Production of a Highly Protease-Resistant Fungal alpha-Galactosidase in Transgenic Maize Seeds for Simplified Feed Processing. Plos One 10:e0129294.

Zhao, T.Y., Corum, J.W., Mullen, J., Meeley, R.B., Helentjaris, T., Martin, D., and Downie, B. (2006). An alkaline alpha-galactosidase transcript is present in maize seeds and cultured embryo cells, and accumulates during stress. Seed Sci Res 16:107-121.

Zhao, T.Y., Thacker, R., Iii, J.W.C., Snyder, J.C., Meeley, R.B., Obendorf, R.L., and Downie, B. (2004). Expression of the maize GALACTINOL SYNTHASE gene family: (I) Expression of two different genes during seed development and germination. Physiologia Plantarum 121:634–646.

Zhou, M.L., Zhang, Q., Zhou, M., Sun, Z.M., Zhu, X.M., Shao, J.R., Tang, Y.X., and Wu, Y.M. (2012). Genome-wide identification of genes involved in raffinose metabolism in Maize. Glycobiology 22:1775-1785.

Zhuo, C.L., Wang, T., Lu, S.Y., Zhao, Y.Q., Li, X.G., and Guo, Z.F. (2013). A cold responsive galactinol synthase gene from Medicago falcata (MfGolS1) is induced by myo-inositol and confers multiple tolerances to abiotic stresses. Physiol Plantarum 149:67-78.

Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B.L., Dubois-Laurent, C., Geoffriau, E., Le Signor, C., and Dalmais, M. (2016). ABI5 is a regulator of seed maturation and longevity in legumes. Plant Cell 28:2735-2754.

Zuther, E., Buchel, K., Hundertmark, M., Stitt, M., Hincha, D.K., and Heyer, A.G. (2004). The role of raffinose in the cold acclimation response of Arabidopsis thaliana. Febs Lett 576:169-173.

FIGURE LEGENDS

Figure 1. Identification of RAFFINOSE SYNTHASE (RS) in maize. (A) Evolutionary relationships of RFOs synthetic- and select RFOs-hydrolytic enzymes from various taxa. An asterisk (*) indicates

938 the RFOs SYNTHASES predicted by a Pfam search published by others (Zhou et al., 2012), and
939 corresponding *GRMZM* gene IDs are given. The scale bar represents 0.1 amino acid substitutions per
940 site. (B-D) Characterization of the enzyme activity (raffinose [Raf] synthetic in B, galactinol [Gol]
941 hydrolytic in C, and Raf hydrolytic in D) of crude lysates from *E. coli* expressing either empty vector
942 (VC) or ZmRS:His_6 (recombinant maize RS) as determined by HPLC-ELSD in light scattering units
943 (LSU). Other sugars used or detected in the assay are abbreviated as follows: Suc, sucrose; Lac,
944 lactose; *myo*, *myo*-inositol; and Gal, galactose.
945
946 Figure 2. Raffinose and ZmRS expression were concurrently accumulated in embryos at late stage of
947 seed development. (A) *ZmRS* mRNA accumulation in maize seed embryos (B73 inbred line) during
948 development and imbibition as detected by real time RT-PCR. The expression of *ZmRS* was
949 normalized to GAPDH expression. Data are means ±SEM (n=3). Different letters over the bars
950 indicate significant differences among means (Tukey's test). (B) ZmRS protein abundance (western
951 blot in top panel) in B73-inbred-line embryos during development and imbibition. The western blot
952 analysis of GAPDH protein (bottom panel) with the same extracts is used to demonstrate equal
953 protein loading. (C) The raffinose content in seed embryos (B73 inbred line) during development and
954 imbibition as detected by HPLC-ELSD. Data are means ±SEM (n=3). Different letters over the bars
955 indicate significant differences among means (Tukey's test).
956
957 Figure 3. *Mutator* interrupted *zmrs-1* maize mutant (W22 background) abolished raffinose
958 production both in leaf and seed. (A) Structure of the ZmRS gene and *mutator* insertion of the *zmrs-1*
959 mutant. Exons are shown as black boxes and introns as lines. The 5'UTR and 3'UTR are white boxes.
960 The *mutator* insertion site and primer sites are indicated. (B) PCR genotyping of NS and *zmrs-1*
961 mutant plants. (C) RT-PCR analysis of *ZmRS* transcript abundance in NS and *zmrs-1* plants. (D)
962 Western blot analysis of RS protein accumulation (top panel) in NS and *zmrs-1* plants. The bottom
963 panel is western blot analysis of GAPDH protein from the same extracts, demonstrating equal protein
964 loading. (E) HPLC-ELSD analysis of sugar profiles in leaves, embryos and endosperms. Detected
965 sugars are sucrose (Suc), *myo*-inositol (*myo*), raffinose (Raf), and, galactinol (Gol) and are measured

966 in light scattering units (LSU).

967

968 Figure 4. *Mutator* interrupted *zmrs-1* maize mutant (W22 background) showed lower seed vigor than
969 its null segregant (NS) control. (A) Photographs of one replicate of NS and *zmrs-1* seeds/seedlings
970 after imbibition for 120 h without accelerated aging (AA) or following AA for 3 or 6 days (AA3 or
971 AA6, respectively). The white bar is equivalent to 5 cm. (B) Comparison of seed germination
972 percentages between NS and *zmrs-1* seeds that were either left untreated or AA-treated. Seeds were
973 imbibed at 28°C in the dark and the completion of germination was monitored every 12 h from 36 h
974 to 120 h. There are 3 replicates for each treatment and 50 seeds for each replicate. One and two
975 asterisks denote significance relative to NS within the same AA treatment (Student's T-test) at $p<0.05$
976 and <0.01, respectively. (C) A comparison of root length and shoot length of seedlings generated in B.
977 Values are means ±SEM (n values for each sample provided above the bars) and significance
978 determined and denoted as in B after removing the bias of slower completion of germination which
979 could artificially lower the *zmrs-1* mean. (D) Comparison of Suc, *myo*, Gol and Raf contents from
980 embryos between NS and *zmrs-1* that were treated with or without AA. Values are means ± SEM
981 (n=3). Different letters indicate significant difference among different groups (Tukey's test).

982

983 Figure 5. Co-overexpression of *ZmGOLS2* and *ZmRS*, or overexpression of *ZmGOLS2* alone,
984 significantly enhanced seed vigor while single-overexpression of *ZmRS* dramatically decreased seed
985 vigor. (A) Representative photographs of the unaged (no accelerated aging (AA) treatment) and aged
986 (3 d AA treatment) Arabidopsis seeds that had been stained by tetrazolium. Genotypes included were
987 as follows: WT, wild type (Col-0); OEGFP, *GFP* over-expressing Arabidopsis line; OEZmRS, *ZmRS*
988 over-expressing Arabidopsis line; OEZmGOLS2, *ZmGOLS2* over-expressing Arabidopsis line;
989 OEZmGOLS2/ZmRS, *ZmGOLS2* and *ZmRS* double over-expressing Arabidopsis lines (only
990 homozygotes OEZmGOLS2 plants transformed with *ZmRS* overexpression vector). (B)
991 Representative photographs of the seeds after 156 h of imbibition from different Arabidopsis lines
992 that were treated with or without aging. (C-D) Comparison of seed germination percentages between
993 WT and different Arabidopsis lines that were treated either without (C) or with (D) AA treatment.

994 There are 6 replicates for each treatment and there are 50 seeds for each replicate. *p<0.05, **p<0.01
995 relative to WT within the same aging treatment (Dunnett's test). (E-J) Comparison of each individual
996 sugar content among different unaged Arabidopsis seeds. (A) sucrose; (E) *myo*-inositol; (F)
997 galactinol; (G) raffinose; (I) stachyose; (J) verbascose. Different letters indicate significant
998 differences among different lines (Tukey's test). Values are means ±SEM; n=3.
999
1000 Figure 6. The effects of total RFOs, each individual sugar, raffinose: sucrose ratio or HDP RFOs:
1001 sucrose ratio on Arabidopsis seed vigor. (A) Comparison of total RFOs, each individual sugar and
1002 AA tolerance among different Arabidopsis lines. The data of figure 5 and supplemental figures 9-10
1003 were recalculated as a percentage of the total of the 6 detected sugars. AA tolerance was calculated as
1004 the germination percentage at 84 hours after imbibition and then normalized to WT. The data of
1005 *OEAtSTS-4* shown in supplemental figure 10 was excluded in this figure because the *AtSTS* was not
1006 overexpressed in that line. Different letters indicate significance between each group (Tukey's test).
1007 ** indicate significant difference of seed AA tolerance between different lines. (B) Dualistic linear
1008 regression analysis of effects of the ratios of raffinose/sucrose (X1) and HDP-RFOs/sucrose (X2) on
1009 seed vigor (Y). The results are summarized as Y=204 X1 + 270 X2 + 21 (R=0.737).
1010
1011 Figure 7. Simplified RFOs synthetic metabolism and its effects on seed vigor in maize and
1012 Arabidopsis. GOLS, GALACTINOL SYNTHASE; RS, RAFFINOSE SYNTHASE; STS,
1013 STACHYOSE SYNTHASE; VES, VERBASCOSE SYNTHASE.

Table 1 Pearson correlation matrix of Sugar content and seed AA tolerance of Arabidopsis

	Sugar content (μg · mg^{-1} Dw)											
	SUC	*MYO*	GOL	RAF	STA	VER	HDP-RFO	Total RFO	RAF/SUC	HDP-RFO/ SUC	Total RFO/SUC	AA tolerance
SUC	1											
MYO	.035	1										
GOL	.228	.321	1									
RAF	-.356	-.047	-.839**	1								
STA	-.090	.292	.704**	-.797**	1							
VER	.021	.612**	.855**	-.592**	.693**	1						
HDP-RFO	-.069	.387	.782**	-.794**	.985**	.808**	1					
Total RFO	-.642**	.327	-.530*	.775**	-.252	-.106	-.232	1				
RAF/SUC	-.553*	-.088	-.790**	.963**	-.712**	-.568*	-.719**	.794**	1			
HDP-RFO/SUC	-.286	.401	.693**	-.651**	.943**	.814**	.968**	-.037	-.553*	1		
Total RFO/SUC	-.869**	.185	-.445	.683**	-.171	-.092	-.162	.925**	.798**	.061	1	
AA tolerance	-.810**	-.013	-.106	.258	.072	.003	.060	.476*	.473*	.210	.718**	1

**indicate significance on $p<0.01$ level (Two tailed), *indicate significance on $p<0.05$ level (Two tailed). SUC, sucrose; *MYO*, *myo*-inositol; GOL, Galactinol; RAF, Raffinose; STA, Stachyose; VER, verbascose; HDP-RFO, stachyose+verbacose; Total RFO, raffinose+stachyose+verbascose.