Identification of candidate genes for the seed coat colour change in a *Brachypodium distachyon* mutant induced by gamma radiation using whole-genome re-sequencing

Journal:	<i>Genome</i>
Manuscript ID	gen-2016-0145.R2
Manuscript Type:	Article
Date Submitted by the Author:	22-Jan-2017
Complete List of Authors:	Lee, Man Bo; KOREA UNIVERSITY, Dept. of Biotechnology Kim, Dae Yeon; Korea University Seo, Yong Weon; KOREA UNIVERSITY,
Keyword:	Brachypodium distachyon, gamma radiation, SNP, InDel, seed coat colour

Identification of candidate genes for the seed coat colour change in a *Brachypodium distachyon* mutant induced by gamma radiation using whole-genome re-sequencing

Man Bo Lee, Dae Yeon Kim, and Yong Weon Seo

M.B. Lee. Department of Biotechnology, College of Life Sciences & Biotechnology, Korea University, Seoul, 02841, Korea

D.Y. Kim and Y.W. Seo. Division of Biotechnology, College of Life Sciences & Biotechnology, Korea University, Seoul, 02841, Korea

Corresponding author: Yong Weon Seo (email: seoag@korea.ac.kr)

email

Man Bo Lee: manbo615@korea.ac.kr

Dae Yeon Kim: dykim@korea.ac.kr

Yong Weon Seo: seoag@korea.ac.kr

Abstract: *Brachypodium distachyon* has been proposed as a model plant for agriculturally important cereal crops such as wheat and barley. Seed coat colour change from brown-red to yellow was observed in a mutant line (142-3) of *B. distachyon*, which was induced by chronic gamma radiation. In addition, dwarf phenotypes were observed in each of the lines 142-3, 421-2, and 1376-1. In order to identify causal mutations for the seed coat colour change, the three mutant lines and the wild type were subjected to whole-genome re-sequencing. After removing natural variations, 906, 1057, and 978 DNA polymorphisms were detected in 142-3, 421-2, and 1376-1, respectively. A total of 13 high-risk DNA polymorphisms were identified in mutant 142-3. Based on a comparison with DNA polymorphisms in 421-2 and 1376-1, candidate causal mutations for the seed coat colour change in 142-3 were selected. In the two independent *A. thaliana* lines carrying T-DNA insertions in the *AtCHI*, seed colour change was observed. We propose a frameshift mutation in *BdCH11* as a causal mutation responsible for seed colour change in 142-3. The DNA polymorphism information for these mutant lines can be utilized for functional genomics in *B. distachyon* and cereal crops.

Key words: *Brachypodium distachyon*, gamma radiation, SNP, InDel, seed coat colour

Introduction

Grasses provide staple foods for humans, animal feed, and sustainable energy sources. The grass family (Poaceae) dominates ecological and agricultural systems. The subfamily Pooideae includes numerous economically important cereals such as wheat, barley, oats, and rye (Hands and Drea 2012). *Brachypodium distachyon* was the first member of the subfamily Pooideae to be sequenced. The high-quality genome sequence of *B. distachyon* provides a template for analysis of the large genomes of economically important crops such as wheat (Vogel et al. 2010). The *B. distachyon* genome can be utilized for structural genomic studies in grasses because *B. distachyon* is a relative of an exceptionally large number of important crops, and has close phylogenetic relationships with other crops (Kellogg 2001; Opanowicz et al. 2008).

Seed colour is an important agronomic trait that is associated with pre-harvest sprouting and dormancy (Himi et al. 2002). In wheat, *Tamyb10*, which is considered a red grain colour regulator, controls flavonoid (including anthocyanin) biosynthesis genes such as *chalcone isomerase (CHI)* and *dihydroflavonol 4-reductase (DFR)* (Himi et al. 2011; Himi et al. 2012). In rice, *Rc*, which encodes a basic helix-loop-helix protein, controls proanthocyanidin synthesis in red grain pericarp (Sweeney et al. 2006). *B. distachyon* is an attractive functional model for the study of small-grain temperate cereals and related grasses. The presence of a shallow crease in the *B. distachyon* grain is similar to that in wheat. The grain composition appears to be close to that of oats and the high β -glucan content in endosperm cell walls is similar to that in barley (Guillon et al. 2011; Hands and Drea 2012).

As a consequence of the development of next-generation sequencing (NGS) technologies, it is now possible to re-sequence entire plant genomes economically and productively. NGS technologies produce enormous amounts of data on DNA polymorphisms on a genomic scale, including information on single-nucleotide polymorphisms (SNPs) and inserts and deletions (InDels) (Fu et al. 2016). NGS technologies have been used for a variety of applications, such as developing SNP-based markers and identifying causal mutations (Uchida et al. 2011; Varshney et al. 2009). For example, the ethyl methane sulfonate-induced mutations responsible for phenotypes of interest were identified through bulked segregant analysis in *Arabidopsis thaliana* (Schneeberger et al. 2009).

Gamma radiation is a more widely used mutagen compared to other ionizing radiations because of its

ready availability and power of penetration (Marcu et al. 2013). Gamma radiation produces reactive oxygen species, which cause nucleotide substitutions, as well as inducing random inserts, deletions, inversions, and translocations in plant genomes (Morita et al. 2009). The present study aimed to discover the genomic variations and DNA polymorphisms in a *B. distachyon* mutant line with altered seed coat colour, induced by chronic gamma radiation. Furthermore, mutant-specific DNA polymorphisms were identified and probable causal mutations for seed coat colour change are proposed.

Materials and methods

Plant materials

Mutants were induced by chronic gamma radiation (^{60}Co), which was described in the previous study (Lee et al. 2013). The M_3 seeds derived from each M_2 plant were harvested and planted to obtain subsequent M_3 , M_4 , and M_5 generations. The 142-3 mutant line showed altered dehulled seed coat colour (Fig. 1A) and reduced lignin content (Lee et al. 2016). In addition, 142-3, 421-2, and 1376-1 mutant lines showed reductions in plant height as compared to that of the wild type (WT) (PI 254867, Bd21) (Fig. 1B).

Whole-genome re-sequencing

The mutant lines 142-3 ($M_{4;5}$), 421-2 ($M_{3;4}$), and 1376-1 ($M_{3;4}$) and the WT were used for whole-genome re-sequencing. Genomic DNA was extracted from fresh leaves using the CTAB method, as described by Kim et al. (2013). To eliminate natural variation, genomic DNA was extracted from three WT plants and the same amounts of genomic DNA were mixed. The genomic DNA of each mutant line was extracted from a single plant. After purifying genomic DNA, the purified DNA was randomly sheared to produce 400–500 bp DNA fragments. The average DNA fragment size was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). The DNA fragments were ligated with index adapters of the Illumina TrueSeq End Repair-Kit and the AMPure XP Beads Purification Kit (Beckman Coulter Genomics, Danvers, MA, USA). After size fractionation of the ligation products on a 2% agarose gel, selected fragments were amplified via PCR using adapter-specific primer sets. The DNA was isolated using AMPure XP Beads, and the resulting

libraries were assessed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and sequenced by 100 bp paired-end sequencing using the Illumina HiSeq 2500 platform. Base calling was performed with the Illumina pipeline using default settings. To produce high-quality reads, sequence reads were filtered using sickle (v1.33) by retaining the bases with a minimum Phred quality score of 30 for each sample. Using BWA ver. 0.7.12, the remaining reads were mapped to the Purple false brome genome assembly *Brachypodium distachyon* v3.1 (<http://genome.jgi.doe.gov/pages/dynamicOrganismDownload.jsf?organism=Bdistachyon>) (Vogel et al. 2010). Local realignment was performed using GATK ver. 2.3-9, and then duplicates were marked using Picard ver. 1.98 (<http://picard.sourceforge.net>) after mapping. Variant calling was performed using GATK ver. 2.3-9. Variants with a GATK QUAL score of less than 30 and average read depth of less than 9 were excluded. Heterologous variants were excluded. Subsequent annotation of variants was performed using SnpEff (v.4.1).

Identification of seed coat colour using *Arabidopsis thaliana* T-DNA knock-out mutants

Two *At3g55120* [best hit (1.3E-07) with *BdCH11*] knock-out lines (SALK_034145 and GK-176H03) and two *At1g27680* [best hit (6.9E-41) with *BdApL1* (large subunit of *ADP-glucose pyrophosphorylase*)] knock-out lines (SALK_029682C and SALK_008527) were obtained from the *Arabidopsis* Information Resource (Alonso et al. 2003; Rosso et al. 2003). The Col-0 ecotype was used as the wild type. Gene-specific primers (S1-6, G1, and G2) and T-DNA-specific primers (LB and RB) were used for the identification of putative mutant lines. In order to detect *Atchi* allele, S1 (5'-GAGAGCATTCATGGTGGGGT) and S2 (5'-CTCGGACACCTGCGTAAGTT) were used for SALK_034145 and G1 (5'-TCCCCTACCGGCTCTCTTAC) and G2 (5'-ACATGTCAGCCCAGTCCAAC) were used for GK-176H03. In order to detect *Atapl2* allele, S3 (5'-AATCCATCAACTGCCACCGA) and S4 (5'-GGTTCCCCAGAACGCACTAA) were used for SALK_029682C and S5 (5'-GCAGTGTGCAGCACTCAATC) and S6 (5'-CAGGACTGTTGGCCTTGAT) were used for SALK_008527. T-DNA-specific primer LB is LBa1_pROK2 (http://signal.salk.edu/tdna_protocols.html) and RB is 3144 (https://www.gabi-kat.de/db/genotyping_details.php?lineid=176H03&genecode=At3g55120). Seed coat colour were investigated at the fully mature stage.

Anthocyanin measurement

Seeds of 142-3 ($M_{4.5}$) and the WT, and 2-week-old seedlings of 142-3 ($M_{4.5}$), 421-2 ($M_{4.5}$), 1376-1 ($M_{4.5}$), and the WT were used for quantification of the anthocyanin content. Relative anthocyanin content per gram of each sample was determined as described by Mancinelli et al. (1988). The samples (50 mg for each sample) were ground in liquid nitrogen and extracted with 250 μ l of methanol (1% HCl) at 4°C overnight. Subsequently, 250 μ l of distilled water and 250 μ l of chloroform were added. After centrifugation (3000 rpm for 3 min at 25°C), the upper aqueous phase was used for determining relative anthocyanin level using a spectrophotometer at 535 nm.

Results

Whole-genome re-sequencing

Whole-genome re-sequencing of 142-3, 421-2, 1376-1, and the WT produced 103247238, 132810180, 110330530, and 105088784 reads, respectively (Table 1) (the raw data will be publicly available in Dec. 2017 at <http://nabic.rda.go.kr/>). More than 92 million trimmed reads for each sample were mapped to the *B. distachyon* reference genome. Mapping rates of the non-trimmed reads (total reads) ranged from 77.25% to 88.89%. Mapping rates of high-quality trimmed reads of each sample were greater than 96%. The mapped reads of 142-3, 421-2, 1376-1, and the WT covered 98.12%, 97.77%, 97.89%, and 98.79% of the reference genome, respectively. Lines 142-3, 421-2, and 1376-1 and the WT showed a mean depth of coverage of 32.95 $\times$, 36.89 $\times$, and 35.33 $\times$, and 34.05 $\times$, respectively.

The DNA polymorphisms were filtered to minimize possible false-positive detection using a minimum GATK QUAL score of 30 and minimum average read depth of 9. Heterozygous SNPs and InDels were also excluded since heterozygous SNPs and InDels contain a large number of false-positives. Furthermore, the casual mutations responsible for the seed coat colour could be homozygous DNA polymorphisms. The yellow seed phenotype was constantly observed in all M_4 and M_5 populations that were directly derived from the 142-3 line. Although the mutants and WT are derived from the standard line Bd21, natural variations were present in the mutants and WT compared to the *B. distachyon* reference sequence (Bd21). In order to eliminate natural

variations, the DNA polymorphisms identified at the same position in both mutant and WT genomes were excluded from the mutant data.

The number of genome-wide DNA polymorphisms in 142-3, 421-2, and 1376-1 were 906 (838 SNPs and 68 InDels), 1057 (972 SNPs and 85 InDels), and 978 (903 SNPs and 75 InDels), respectively (Table 2). The frequency of DNA polymorphisms ranged from 3.34 per Mb in the 142-3 genome to 3.90 per Mb in the 421-2 genome. Among the three mutant lines, the variant frequency was highest in 421-2. Deletions were detected more frequently than inserts in all three mutants. The distribution of DNA polymorphisms in the different regions of genes, such as upstream (5 kb), 5' UTR, exon, splice site, intron, 3' UTR, downstream (5 kb), and intergenic regions, is shown in Table 3. In the 5' UTR, exon, splice site, and 3' UTR regions, the proportions of DNA polymorphisms ranged from only 0.12% to 1.41%. In all three mutant genomes, approximately 50% (from 49.39% to 51.42%) of the DNA polymorphisms were detected in intergenic regions.

Candidate genes associated with seed coat colour change

For detecting DNA polymorphisms related to the yellow seed coat colour of 142-3, we selected high-risk DNA polymorphisms that were considered to potentially affect gene function, such as missense mutations, nonsense mutations, splicing donor site changes, splicing acceptor site changes, and frameshift mutations. A total of 13 high-risk DNA polymorphisms (1 deletion and 12 SNPs) in 12 different genes were detected in 142-3 (Table 4). Twelve of the 13 high-risk DNA polymorphisms in 142-3 were used for validation. In the present study, stringent criteria for removing false-positive detection were applied. In particular, DNA polymorphisms with an average read depth ≥ 9 were selected. Target-specific primers were designed using Primer3 (Table S1¹). PCR products were sequenced using the Sanger technique. Of the 12 high-risk DNA polymorphisms, 11 (91.67%) were validated.

DNA polymorphisms that were identified in both 142-3 and 421-2, and/or 1376-1 are marked in Table 4 (indicated by “a” and “b”, respectively). We assumed that the marked DNA polymorphisms in Table 4 are not involved in the yellow seed coat colour of 142-3. Since the yellow seed coat colour was observed only in 142-3,

¹ Supplementary data is available on the journal Web site.

we assumed that 142-3-specific DNA polymorphisms are responsible for this phenotype. The marked DNA polymorphisms might be involved in the dwarf phenotype since all three mutants (142-3, 421-2, and 1376-1) had reduced plant heights (Fig. 1B), or could be natural variations that were not excluded. Three of the 13 high-risk DNA polymorphisms (1 deletion and 2 SNPs) were only detected in 142-3. A frameshift mutation caused by a 14 base deletion was identified in *Bradi1g03840* (*BdCH11*); one missense mutation was identified in *Bradi1g53500* (*BdApL1*); and the other missense mutation was detected in an unidentified protein.

A. thaliana lines carrying T-DNA insertions in the *AtCHI* or *AtApL2* genes were used for identification of seed coat colour (Fig. 2). PCR analysis was performed with gene-specific primers (S1-6, G1, and G2) and T-DNA-specific primers (LB and RB). *Atchi* mutant alleles were detected from SALK_034145 and GK-176H03 and *Atapl2* mutant alleles were detected from SALK_029682C and SALK_008527. On the basis of PCR analysis, homozygous lines were detected, with the exception of the SALK_034145 line. In the SALK_034145 line, the PCR products were amplified with gene-specific primers but no PCR products were amplified with T-DNA-specific primers. It is seemed that the left border might be deleted from the genome in the SALK_034145 line. Yellow seeds were observed in the *Atchi* mutant lines but brown seeds were observed in the *Atapl2* mutant lines and the WT.

Anthocyanin contents

The seeds of 142-3 and the WT were used for anthocyanin quantification (Fig. 3A). Relative anthocyanin level per gram of seeds was decreased by 38.18% in the yellow seeds of 142-3 compared to the WT. The reduction in anthocyanin accumulation resulted in the yellow seed coat colour in 142-3. Two-week-old seedlings of M_{4;5} lines for each of 142-3, 421-2, 1376-1, and the WT were used for anthocyanin quantification (Fig. 3B). Relative anthocyanin level per gram of seedlings was reduced in 142-3 compared to the WT.

Discussion

Whole-genome re-sequencing

In order to analyze genome-wide DNA polymorphisms, massive and high-quality sequence data were

produced through NGS technology. More than 100 million reads were produced by whole-genome re-sequencing of each mutant and the WT using the Illumina Hiseq 2500 platform. High-quality reads were used for mapping. The mean depths of coverage were greater than 30× in each mutant and the WT (Table 1). Numerous researchers have reported genome-wide DNA polymorphisms using data with a mean depth of coverage of approximately 30×. Fu et al. (2016) reported that genome-wide DNA polymorphisms were discovered by whole-genome re-sequencing in rice varieties using data with mean depths of coverage of 27.83× and 28.72× produced by the Illumina Hiseq 2000 platform. Similarly, Jain et al. (2014) reported that DNA polymorphisms were identified in genes associated with drought and salinity stress in rice cultivars using data with mean depths of coverage from 25× to 27× produced by the Illumina Hiseq 2000 platform.

In order to minimize possible false-positive detection, in the present study, DNA polymorphisms were filtered using stringent criteria. We selected only DNA polymorphisms with a GATK QUAL score >30 and an average read depth ≥9. In addition, heterozygous DNA polymorphisms were removed. The validation rate of the high-risk DNA polymorphisms in 142-3 was 91.67%. Prior to selecting DNA polymorphisms with an average read depth ≥9, DNA polymorphisms with an average read depth ≥8 had been selected. Initially, 18 high-risk DNA polymorphisms in 142-3 had been used for validation and the validation rate had been 72.22% (13 out of 18) (data not shown). By raising the cut-off value of the average read depth, we obtained a reasonably high validation rate, and the total numbers of DNA polymorphisms in 142-3, 421-2, and 1376-1 were reduced from 1217, 1371, and 1292 to 906, 1057, and 978, respectively (data not shown). Numerous measures for removing false-positives have been applied by other researchers. Fu et al. (2016) reported that DNA polymorphisms with coverage ≥10 and ≤100 were selected and heterozygous DNA polymorphisms were removed for detecting DNA polymorphisms in the hybrid weakness genes of rice variety RGD-7S. Similarly, Lestari et al. (2014) reported that DNA polymorphisms with mapped reads of depth 5–100 and DNA polymorphisms with a quality score greater than 100 using the SAMTools program were selected for detecting DNA polymorphisms.

DNA polymorphism frequency and distribution in mutants induced by gamma radiation

B. distachyon mutant-specific DNA polymorphisms caused by gamma radiation were identified. Natural variations were removed using DNA polymorphisms in the WT by comparing with the reference

sequence. The number of DNA polymorphisms in mutant lines ranged from 906 to 1057 (Table 2). The frequency of DNA polymorphisms in mutant genomes ranged from 3.34 to 3.90 per Mb. Hwang et al. (2014) reported genome-wide DNA polymorphisms in an early-maturing rice mutant (EMT1) induced by gamma radiation and the WT, where a total of 136821 SNPs and 22599 InDels were identified as EMT1-specific DNA polymorphisms. The frequency of DNA polymorphisms observed in the present study was relatively lower than that in EMT1. This lower frequency of DNA polymorphisms might be attributable to numerous factors, including different plant species, different dose of gamma radiation treatment, use of different programs for variant calling, and the application of highly stringent selection criteria. In addition, DNA polymorphisms generated by gamma radiation have been identified using Targeting Induced Local Lesions IN Genome analysis (TILLING). Using TILLING in the rice cultivar Dongan, Chun et al. (2012) reported a DNA polymorphism frequency of 4.90 per Mb in *OASAI* from 1350 mutants derived from callus exposed to gamma radiation, which is similar to our result for DNA polymorphisms induced by gamma radiation.

Among the *B. distachyon* mutants induced by gamma radiation, approximately 1% (from 1.11% to 1.41%) of the DNA polymorphisms were detected in exon regions (Table 3). The DNA polymorphisms were mainly detected in the intergenic, upstream, and downstream regions in all three mutants. Hwang et al. (2014) reported that SNPs mainly occurred in the upstream (54.66%) and intron (18.74%) regions in EMT1. This identification of a high proportion of DNA polymorphisms in non-genic regions, as affected by gamma irradiation, is similar to our results.

Seed coat colour change

Whole-genome sequencing has been applied to the discovery of causal genes responsible for a number of phenotypes of interest (Uchida et al. 2011). In rice, the two causal SNPs of a pale green leaf mutant were identified using the MutMap method. This method utilizes whole-genome re-sequencing data of pooled DNA from a segregating population of plants that exhibit a certain trait (Abe et al. 2012). In maize, a total of 271 candidate SNPs for drought tolerance were identified by comparative whole-genome re-sequencing data of 16 inbred lines, including three extremely drought-sensitive lines and three extremely drought-tolerant lines (Xu et al. 2014). In the present study, candidate genes for seed coat colour change in mutant 142-3 were identified

through whole-genome re-sequencing. The candidate DNA polymorphisms were detected by comparison with the DNA polymorphisms identified in WT and another two mutants that showed a dwarf phenotype (Table 4).

Numerous previous studies have reported that flavonoid (including anthocyanin) biosynthesis genes are associated with seed colouration. For example, transformation of the *DFR* gene into brown-coloured rice resulted in red-coloured rice (Furukawa et al. 2007). *Tamyb10*, which is associated with red-coloured wheat, induces the expression of flavonoid-related genes (Himi et al. 2011). *Rc*, which is associated with red-coloured rice, is known as a negative regulator of anthocyanin production (Sweeney et al. 2006). In the flavonoid biosynthetic pathway, naringenin chalcone produced by chalcone synthase is rapidly isomerized by CHI to form the flavanone naringenin (Schijlen et al. 2004). A mutation in *CHI* can cause an alteration in pigmentation and flavonoid components in seed. Seed coat colour change from brown-red to yellow could be ascribed to an alteration in anthocyanin content. Suppression of *NtCH1* by RNA interference was shown to induce a reduction in anthocyanin content and change in flavonoid components in the flower petals of transgenic tobacco (Nishihara et al. 2005). Disruption of *CHI* and *DFR* by transposon insertions caused a change in flower colour to yellow in carnations (Itoh et al. 2002). In an *A. thaliana* *tt5* mutant, ionizing radiation was shown to generate inversion within *AtCHI* (Shirley et al. 1992). The *tt5* mutant was deficient in anthocyanin content in the leaves and showed yellow seed coat colour. A deficiency in oxidized flavonoid compounds results in the yellow seed coat colour of the *tt5* mutant compared to the dark brown seed coat colour of the WT (Jiang et al. 2015). Our results for anthocyanin content in 142-3 possessing a mutation in *BdCH1* are thus consistent with the findings of previous studies. The anthocyanin contents of seeds and seedlings were reduced in 142-3 compared to the WT (Fig. 3). Moreover, yellow seeds were observed in the two independent *A. thaliana* lines carrying T-DNA insertions in the *AtCHI* (Fig. 2). It is highly seemed that a frameshift mutation in *BdCH1* is a causal mutation responsible for seed colour change in 142-3.

Although *BdApL1* was selected as a candidate gene responsible for the seed colour change in 142-3 using whole-genome re-sequencing, there is no information in the literature regarding the relationship between *ApL* and seed colouration. ADP-glucose pyrophosphorylase catalyzes the first and limiting step in starch biosynthesis in plants. ADP-glucose pyrophosphorylase as a heterotetramer consists of two small and two large subunits (Ventriglia et al. 2008). In *A. thaliana* seeds, *ATApL1* is expressed in all the embryo cells, whereas other *ATApL* genes, including *ATApL2*, are expressed in the pro-vascular cells (Crevillén et al. 2005). In addition,

yellow seeds were not observed in the two independent *A. thaliana* lines carrying T-DNA insertions in the *AtApL2* (Fig. 2).

Whole-genome re-sequencing was performed to identify the causal DNA polymorphisms responsible for the yellow seed coat colour change in 142-3. Natural variations and DNA polymorphisms that are unrelated to seed colour were removed based on the whole-genome re-sequencing data of the WT and other mutants. Seed colour change was identified in *A. thaliana* lines carrying T-DNA insertions in the *AtCHI*. A frameshift mutation in *BdCHI1* appears to be responsible for the yellow seed colouration. This work provides a valuable genomics resource for studying genomic mutation via gamma irradiation, and it can be applied to other grasses and cereals.

Acknowledgements

We thank *Arabidopsis* Biological Resource Center for distributing seeds of *Arabidopsis thaliana* mutant lines. This study was supported by the Cooperative Research Program for Agriculture Science & Technology Development (Project Nos. PJ01102102, PJ01103501), Rural Development Administration, Republic of Korea.

References

- Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S., and Terauchi, R. 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**(2): 174–178. doi:10.1038/nbt.2095.
- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. 2003. Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. *Science*, **301**(5633): 653–657. doi:10.1126/science.1086391.
- Chun, J.B., Ha, B.-K., Jang, D.-S., Song, M., Lee, K.J., Kim, J.-B., Kim, S.H., Kang, S.-Y., Lee, G.-J., Seo, Y.W., and Kim, D.S. 2012. Identification of mutations in *OAS1* gene from a gamma-irradiated rice mutant population. *Plant Breed.* **131**(2): 276–281. doi:10.1111/j.1439-0523.2011.01933.x.
- Crevillén, P., Ventriglia, T., Pinto, F., Orea, A., Mérida, Á., and Romero, J.M. 2005. Differential Pattern of Expression and Sugar Regulation of *Arabidopsis thaliana* ADP-glucose Pyrophosphorylase-encoding Genes. *J. Biol. Chem.* **280**(9): 8143–8149. doi:10.1074/jbc.M411713200.
- Dixon, R.A., and Paiva, N.L. 1995. Stress-Induced Phenylpropanoid Metabolism. *The Plant Cell*, **7**(7): 1085–1097. doi:10.1105/tpc.7.7.1085.
- Fu, C.-Y., Liu, W.-G., Liu, D.-L., Li, J.-H., Zhu, M.-S., Liao, Y.-L., Liu, Z.-R., Zeng, X.-Q., and Wang, F. 2016. Genome-wide DNA polymorphism in the indica rice varieties RGD-7S and Taifeng B as revealed by whole genome re-sequencing. *Genome*, **59**(3): 197–207. doi:10.1139/gen-2015-0101.
- Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Takamure, I., and Kadowaki, K.-I. 2007. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **49**(1): 91–102. doi:10.1111/j.1365-313X.2006.02958.x.
- Guillon, F., Bouchet, B., Jamme, F., Robert, P., Quémener, B., Barron, C., Larré, C., Dumas, P., and Saulnier, L. 2011. *Brachypodium distachyon* grain: characterization of endosperm cell walls. *J. Exp. Bot.* **62**(3): 2011.

- 1001–1015. doi:10.1093/jxb/erq332.
- Hands, P., and Drea, S. 2012. A comparative view of grain development in *Brachypodium distachyon*. J. Cereal Sci. **56**(1): 2–8. doi:dx.doi.org/10.1016/j.jcs.2011.12.010.
- Himi, E., Mares, D.J., Yanagisawa, A., and Noda, K. 2002. Effect of grain colour gene (R) on grain dormancy and sensitivity of the embryo to abscisic acid (ABA) in wheat. J. Exp. Bot. **53**(374): 1569–1574. doi:10.1093/jxb/erf005.
- Himi, E., Maekawa, M., Miura, H., and Noda, K. 2011. Development of PCR markers for *Tamyb10* related to *R-1*, red grain color gene in wheat. Theor. Appl. Genet. **122**(8): 1561–1576. doi:10.1007/s00122-011-1555-2.
- Himi, E., Yamashita, Y., Haruyama, N., Yanagisawa, T., Maekawa, M., and Taketa, S. 2012. *Ant28* gene for proanthocyanidin synthesis encoding the R2R3 MYB domain protein (Hvmyb10) highly affects grain dormancy in barley. Euphytica, **188**(1): 141–151. doi:10.1007/s10681-011-0552-5.
- Hwang, S.-G., Hwang, J.G., Kim, D.S., and Jang, C.S. 2014. Genome-wide DNA polymorphism and transcriptome analysis of an early-maturing rice mutant. Genetica, **142**(1): 73–85. doi:10.1007/s10709-013-9755-0.
- Itoh, Y., Higeta, D., Suzuki, A., Yoshida, H., and Ozeki, Y. 2002. Excision of transposable elements from the *chalcone isomerase* and *dihydroflavonol 4-reductase* genes may contribute to the variegation of the yellow-flowered carnation (*Dianthus caryophyllus*). Plant Cell Physiol. **43**(5): 578–585. doi:10.1093/pcp/pcf065.
- Jain, M., Moharana, K.C., Shankar, R., Kumari, R., and Garg, R. 2014. Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. Plant Biotechnol. J. **12**(2): 253–264. doi:10.1111/pbi.12133.
- Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R.A., and Pang, Y. 2015. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. J. Exp. Bot. **66**(22): 7165–7179. doi:10.1093/jxb/erv413.
- Kellogg, E.A. 2001. Evolutionary history of the grasses. Plant Physiol. **125**(3): 1198–1205. doi:10.1104/pp.125.3.1198.
- Kim, J.Y. 2013. Identification and functional analysis of *s-adenosyl methionine synthetase* (*HvSAMS*) genes in

- early maturing barley (*Hordeum vulgare* subsp. *vulgare*). Plant Breed. Biotechnol. **1**(2): 178–195. doi:10.9787/PBB.2013.1.2.178.
- Lee, M.B., Kim, D.Y., Jeon, W.B., Hong, M.J., Lee, Y.J., Bold, O., Jang, J.H., Kim, Y.J., Kang, S.Y., Kim, D.S., Kim, J.B., Chun, J.B., and Seo, Y.W. 2013. Effect of gamma radiation on growth and lignin content in *Brachypodium distachyon*. J. Crop. Sci. Biotechnol. **16**(2): 105–110. doi:10.1007/s12892-013-0022-9.
- Lee, M.B., Kim, J.Y., and Seo, Y.W. 2016. Identification of lignin-deficient *Brachypodium distachyon* (L.) Beauv. mutants induced by gamma radiation. J. Sci. Food Agric. doi:10.1002/jsfa.8024.
- Mancinelli, A.L., Hoff, A.M., and Cottrell, M. 1988. Anthocyanin production in Chl-rich and Chl-poor seedlings. Plant Physiol. **86**(3): 652–654. doi:dx.doi.org/10.1104/pp.86.3.652.
- Marcu, D., Cristea, V., and Daraban, L. 2013. Dose-dependent effects of gamma radiation on lettuce (*Lactuca sativa* var. *capitata*) seedlings. Int. J. Radiat. Biol. **89**(3): 219–223. doi:10.3109/095553002.2013.734946.
- Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., and Nishimura, M. 2009. Molecular characterization of mutations induced by gamma irradiation in rice. Genes Genet. Syst. **84**(5): 361–370. doi:10.1266/ggs.84.361.
- Ngaki, M.N., Louie, G.V., Philippe, R.N., Manning, G., Pojer, F., Bowman, M.E., Li, L., Larsen, E., Wurtele, E.S., and Noel, J.P. 2012. Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. Nature, **485**(7399): 530–533. doi:10.1038/nature11009.
- Nishihara, M., Nakatsuka, T., and Yamamura, S. 2005. Flavonoid components and flower color change in transgenic tobacco plants by suppression of *chalcone isomerase* gene. FEBS Lett. **579**(27): 6074–6078. doi:dx.doi.org/10.1016/j.febslet.2005.09.073.
- Opanowicz, M., Vain, P., Draper, J., Parker, D., and Doonan, J.H. 2008. *Brachypodium distachyon*: making hay with a wild grass. Trends Plant Sci. **13**(4): 172–177. doi:dx.doi.org/10.1016/j.tplants.2008.01.007.
- Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. 2003. An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53**(1): 247–259. doi:10.1023/B:PLAN.0000009297.37235.4a.
- Schijlen, E.G.W.M., Ric de Vos, C.H., van Tunen, A.J., and Bovy, A.G. 2004. Modification of flavonoid biosynthesis in crop plants. Phytochemistry, **65**(19): 2631–2648. doi:dx.doi.org/10.1016/j.phytochem.2004.07.028.

- Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jorgensen, J.-E., Weigel, D., and Andersen, S.U. 2009. SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Meth.* **6**(8): 550–551. doi:10.1038/nmeth0809-550.
- Shirley, B.W., Hanley, S., and Goodman, H.M. 1992. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *The Plant Cell*, **4**(3): 333–347. doi:10.1105/tpc.4.3.333.
- Sweeney, M.T., Thomson, M.J., Pfeil, B.E., and McCouch, S. 2006. Caught Red-Handed: *Rc* Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. *The Plant Cell*, **18**(2): 283–294. doi:10.1105/tpc.105.038430.
- Uchida, N., Sakamoto, T., Kurata, T., and Tasaka, M. 2011. Identification of EMS-induced causal mutations in a non-reference *Arabidopsis thaliana* accession by whole genome sequencing. *Plant Cell Physiol.* **52**(4): 716–722. doi:10.1093/pcp/pcr029.
- Varshney, R.K., Nayak, S.N., May, G.D., and Jackson, S.A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* **27**(9): 522–530. doi:dx.doi.org/10.1016/j.tibtech.2009.05.006.
- Ventriglia, T., Kuhn, M.L., Ruiz, M.T., Ribeiro-Pedro, M., Valverde, F., Ballicora, M.A., Preiss, J., and Romero, J.M. 2008. Two *Arabidopsis* ADP-Glucose Pyrophosphorylase Large Subunits (*APL1* and *APL2*) are catalytic. *Plant Physiol.* **148**(1): 65–76. doi:10.1104/pp.108.122846.
- Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., and Bevan, M.W. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**(7282): 763–768. doi:10.1038/nature08747.
- Xu, J., Yuan, Y., Xu, Y., Zhang, G., Guo, X., Wu, F., Wang, Q., Rong, T., Pan, G., Cao, M., Tang, Q., Gao, S., Liu, Y., Wang, J., Lan, H., and Lu, Y. 2014. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biol.* **14**(1): 1–15. doi:10.1186/1471-2229-14-83.

Table 1. Mapping of wild type (WT) and mutant lines by comparison with the *Brachypodium distachyon* reference genome sequence.

	Total reads	Mapping rate	Average depth
WT	105088784	88.89%	34.05
142-3	103247238	86.73%	32.95
421-2	132810180	77.25%	36.89
1376-1	110330530	87.93%	35.33

Draft

Table 2. DNA polymorphisms in individual mutant lines.

	Total variants	SNP	Inserts	Deletions
142-3	906	838	20	48
421-2	1057	972	18	67
1376-1	978	903	17	58

Draft

Table 3. The number and ratio of variants within different gene regions in mutant lines.

	142-3		421-2		1376-1	
	Count	Percent (%)	Count	Percent (%)	Count	Percent (%)
Upstream	397	24.48	416	25.46	413	25.20
5' UTR	2	0.12	3	0.18	3	0.18
Exon	18	1.11	23	1.41	23	1.40
Splice site	5	0.31	9	0.55	8	0.49
Intron	50	3.08	59	3.61	45	2.75
3' UTR	3	0.18	6	0.37	6	0.37
Downstream	313	19.30	311	19.03	305	18.61
Intergenic	834	51.42	807	49.39	836	51.01
Total	1622		1634		1639	

Table 4. The DNA polymorphisms with a high-risk to gene function in mutant 142-3.

Chromosome	Position	Common ^a	Reference	Alteration	Mutation	Gene	Description
Bd1	915341	ab	C	T	Missense	Bradi1 g01368	Protein kinase domain (Pkinase), Legume lectin domain (Lectin_legB)
Bd1	915345	ab	T	C	Splice acceptor	Bradi1 g01368	Protein kinase domain (Pkinase), Legume lectin domain (Lectin_legB)
Bd1	2566754		AAGGAC TCGTCG GTG	A	Frameshift	Bradi1 g03840	Chalcone isomerase 1
Bd1	22573967		G	C	Missense	Bradi1 g27436	Unknown
Bd1	52008546		C	T	Missense	Bradi1 g53500	ADP-glucose pyrophosphorylase, large subunit
Bd2	1669337	b	G	A	Missense	Bradi2 g02503	Wall-associated receptor kinase galacturonan-binding (GUB_WAK_bind), Wall-associated receptor kinase C-terminal (WAK_assoc)
Bd3	527521	ab	G	A	Missense	Bradi3 g00926	Leucine-rich repeat-containing protein
Bd3	4590339	a	T	C	Missense	Bradi3 g06320	Arachidonic acid epoxigenase activity
Bd3	34896726	ab	T	C	Missense	Bradi3 g32715	Unknown
Bd4	8589233	ab	T	C	Missense	Bradi4 g09250	Zinc-binding in reverse transcriptase (zf- RVT)
Bd4	8973568	ab	T	G	Missense	Bradi4 g09567	Transcription factor MYB48-related
Bd4	9678804	a	C	T	Missense	Bradi4 g10040	Leucine-rich repeat-containing protein
Bd4	9895822	ab	A	G	Missense	Bradi4 g10207	Leucine-rich repeat-containing protein

^a “a” and “b” indicate that the DNA polymorphism was also detected in 421-2 and 1376-1, respectively.

Fig. 1. Mutants induced by chronic gamma radiation. Dehulled seed coat colour of the 142-3 $M_{4;5}$ mutant line changed from brown-red to yellow (A). Plant heights were reduced in mutant lines ($M_{2;3}$) compared to the wild type (B).

Draft

Fig. 2. T-DNA knock-out *Arabidopsis thaliana* lines associated with the candidate genes for seed colour change in 142-3. (A) Schematic representation of the *AtCHI* gene [*At3g55120*, best hit (1.3E-07) with *BdCHI1* in *A. thaliana*]. Arrows indicate the locations of T-DNA inserts and PCR primers. Boxes indicate exons. (B) Schematic representation of the *ATApL2* gene [*At1g27680*, best hit (6.9E-41) with *BdApL1* in *A. thaliana*]. (C) PCR analysis of T-DNA knock-out mutant lines (M) and wild type (WT). Gene-specific primers (S1-6, G1, and G2) and T-DNA-specific primers (LB and RB) were used for homologous line detection. S1 and S2 for SALK_034145; G1 and G2 for GK-176H03; S3 and S4 for SALK_029682C; S5 and S6 for SALK_008527. In the SALK_034145 line, the left border might be deleted. (D) Seed coat colour of the T-DNA knock-out lines. A yellow seed phenotype was observed in the *AtCHI* knock-out lines.

Fig. 3. Anthocyanin accumulation in the 142-3 mutant line. Relative anthocyanin level per gram of seeds (A) and 2-week old seedlings (B) were determined from mutant lines and the wild type (WT). The experiment was repeated three times for seeds and two times for seedlings. Values are means and SE.

Draft

Table S1. Primers to validate whole genome re-sequencing results using high-risk DNA polymorphisms in 142-3.

Gene	Forward primer	Reverse primer
Bradi1g01368	ACAGCAATCCCCAGAAGAGC	CGGTTTCGTGCTTTCCGTTT
Bradi1g01368	ACAGCAATCCCCAGAAGAGC	CGGTTTCGTGCTTTCCGTTT
Bradi1g03840	ACAGGCGAGTTCGAGAAGTT	CTGCCAGTACTTGACGCAGT
Bradi1g27436	TTTGATCTGCTGAGGCGCA	GTCACCTCGATTGCCAAGGA
Bradi1g53500	GATGCAAGGCTCAAACGCAT	AAAATGCCACCTGCAGGACT
Bradi2g02503	AGTGGTGGTTCAACGTCTCC	ATGAAACTCACCACCGCACT
Bradi3g00926	AAATACCGTCAGCCAAGGCA	TCTGCCATCCACCCTTGAAC
Bradi3g06320	CGGTCTGAAGAACCTGCACT	GACACACTGCAAGGCAACAG
Bradi3g32715	GTGTTGCCTAGCGGAAGAGT	TGCCATCTTGCTTCAGGAGG
Bradi4g09250	TTCTGGAAAGCGCCCACTAA	AGGCCATAAAAGCCTGGACT
Bradi4g09567	GTGATCGCCGTTGGGATTTC	TTCGTCCAAGTCTTGTTCTGGA
Bradi4g10040	CGCTCTGGCAACTACCAAAC	ACCTTGTGTCTGTTGGCATCT
Bradi4g10207	TGGCTTGGTACCACCCATTT	TCCACAAATAAGCGCCTCAGA