The molecular genetic study about awnedness of rice

(イネの芒に関する分子遺伝学的研究)

Laboratory of Molecular Biosystem,

Division of Molecular Cell Function,

Department of Bioengineering Sciences,

Graduate School of Bioagricultural Science,

Nagoya University, Nagoya, Japan

Kanako UEHARA

March 2017

Contents

Chapter 1

General introduction

Through the long domestication history, cultivated plants contribute to human health and prosperity. It is because human selected the species that have beneficial traits for agriculture from wild species over a long time period. In other words, human took an effort to improve wild species to be more manageable, to a higher yield and better taste. For example, the fruit of tomato (*Solanum lycopersicum*) has been selected larger and larger than its ancestor (Lin et al. 2014), and *Brassica oleracea* has been selected to represent the extraordinary diversity such as cabbage, kale, broccoli and so on (Maggioni et al. 2010). Among the agricultural products, cereals are the most important foods for human. Not only fruits or vegetables but also cereals have been domesticated. The wild progenitors of the major cereals, wheat (*Triticum aestivum*), maize (*Zea mays*) and rice (*Oryza sativa*), show weed like structure and physiological traits (Doebley et al. 2006). Among those, wheat and maize have single origin, Middle East and Mexico respectively, while rice has two origins in Asia and Africa (Salamini et al. 2002; Heerwaarden et al. 2010; Flint-garcia 2013). Rice belongs to the genus *Oryza*, which includes 22 wild species and 2 cultivated rice species (Brar and Khush 1997; Park et al. 2003; Vaughan et al. 2003). These 24 species can be classified into 9 sub-groups based

on the genomic structure; those are AA, BB, CC, BBCC, CCDD, EE, FF, GG and HHJJ (Khush 1997). AA group consists of 8 species including both of two cultivated rice species, *O. sativa* and *O. glaberrima*. *O. sativa* is considered to have been domesticated from *O. rufipogon* approximately 8,000 years ago in Asia (Fuller et al. 2010; Khush 1997; Park et al. 2003), and *O. glaberrima* is derived from *O. barthii* in Africa 2,000–3,000 years ago (Khush 1997; Cai and Morishima 2002). The comparison between two cultivated species does not only tell us of their independent domestication pathway, but also provides information whether the same gene or different gene has been targeted for domestication at separated region. Recent high-throughput genome-wide genetic study on 446 *O. rufipgon* accessions and 1,083 *O. sativa* varieties revealed that *O. sativa* ssp. *japonica* was firstly domesticated from a limited population of *O. rufipogon* around the middle area of the Pearl River in southern China, and also suggested that there is the strong genetic bottleneck by which genetic diversities were largely reduced during the domestication (Huang et al. 2010). This result supported the idea that modern cultivars have lost a large number of genes for promoting to be beneficial for current agricultural condition.

Domestication is the genetic modification process of a wild species to produce a new form of a plant to meet human needs. As for the rice, human selected the traits e.g., less shattering, suppressed spreading panicle, repressed prostrate growth habit, and awnless seeds (Khush 2001; Jin et al. 2008; Huang et al. 2009, 2012). These human-selected traits are common in Asian and African rice known as the "domestication syndrome". This may suggest that, regardless of cultivar or location, such traits significantly contribute to increase the harvest and to make ease of field management. The domestication-related genes in rice have been identified by quantitative trait loci (QTL), analyzing the chromosome segment substitution lines (CSSL), and genome wide association study (GWAS) (Eshed and Zamir 1995; Nadeau et al. 2000; Sweeney and McCouch 2007; Huang et al. 2010; Sang and Ge 2013). Most of them have been found in Asian rice, and interestingly, some genes are reported which were selected for regulating the same traits in African rice. It is suggested that the same genes are targeted for the selection in spite of two geographically isolated domestication processes. For example, shattering was regulated by *Shattering 4* (*Sh4*) gene encoding a Myb3 transcription factor (Li et al. 2006), and brown pericarps color in wild rice was

regulated by *Rc* gene which is a regulatory protein in the proanthocyanidin synthesis pathway (Gross et al. 2011). By identifying these genes, we could see the evidence that *O. sativa* and *O. glaberrima* were artificially selected, though they have different mutations, in the same genes for same domestication traits. However, no example has been reported that the common traits obtained from geographically independent domestication are caused by selection of different genes.

In this doctoral thesis, I showed the first example that the common trait “awnless” seeds in Asian and African rice, was attained by the selection of different genes respectively. In chapter 2, I represented the evidence that the different loci of wild species contribute to producing awn in *O. sativa* and *O. glaberrima* by comparison of phenotypes and genotypes of multiple CSSL which have the same genetic background of *O. sativa* ssp. *japonica* cv. Koshihikari. This comparison also showed that the Asian and African cultivated rice in two areas lost the awn phenotype by selection of different genes. In chapter 3, I identified the responsible gene for awn elongation named *RAE2* and performed its functional analysis. Comparison of the sequences of RAE2 among Asian and African rice, including wild and cultivated species, revealed the artificial

selection of *RAE2* occurred only in Asia. Together, I discussed the domestication process of awnless phenotype in Asian and African rice.

References

Brar DS, and Khush GS. (1997) Alien introgression in rice. *Plant Mol. Biol.* **35**: 35–47.

Cai HW, and Morishima H. (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor. App. Genet.* **104**:1217-1228.

Doebley JF, Gaut BS, Smith BD. (2006) The molecular genetics of crop domestication. *Cell* **127**:1309-1321.

Eshed Y, and Zamir D. (1995) An introgression line population of Lycopersicon pennellii in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **141**:1147–1162.

Flint-garcia SA. (2013) Genetics and Consequences of Crop Domestication. *J. Agr. Food Chem.* **61**: 8267−8276.

Fuller D, Sato Y, and Castillo C. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeol. Anthropol. Sci.* **2**:115–131.

Gross BL, Steffen FT, and Olsen KM. (2011) The molecular basis of white pericarps in African domesticated rice: Novel mutations at the Rc gene. *J. Evo Biol.* **23**: 2747-2753.

Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM. *et al.* (2010) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. USA* **108**:1088–1092.

Huang X, Qian Q, Liu Z, Sun H, He S, *et al*. (2009) Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* **41**: 494-497.

Huang X, Wei X, Sang T, Zhao Q, Feng Q, *et al*. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* **42**:961-967.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, *et al*. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**: 497-501.

Jin J, Huang W, Gao JP, Yang J, Shi M, *et al*. (2008) Genetic control of rice plant architecture under domestication. *Nat. Genet.* **40**: 1365-1369.

Khush GS. (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**: 25–34.

Khush GS. (2001) Green revolution : the way forward. *Genetics* **2**: 815-822.

Li C, Zhou A, and Sang T. (2006) Rice domestication by reducing shattering. *Science* **311**: 1936-1939.

Lin T, Zhu G, Zhang J, Xu X, Yu Q, *et al*. (2014) Genomic analyses provide insights into the history of tomato breeding. *Nat Genet.* **46**: 1220-1226.

Maggioni L, Bothmer VR, Poulsen G, and Branca F. (2010) Origin and Domestication of Cole Crops (Brassica oleracea L.): Linguistic and Literary Considerations. *Economic Botany* **64**: 109–123.

Nadeau JH, Singer JB, Matin A, and Lander ES. (2000) Analysing complex genetic traits with chromosome substitution strains. *Nat. Genet.* **24**: 221–225.

Park KC, Kim KH, Cho YS, Kang KH, Lee JK, and Kim NS. (2003) Genetic variations of AA genome Oryza species measured by MITE-AFLP. *Theor. Appl. Genet.* **107**: 203–209.

Salamini F, Özkan H, Brandolini A, Pregl R, and Martin W. (2002) Genetics and Geography of Wild Cereal Domestication in the Near East. *Genetics* **3**: 429-441.

Sang T and Ge S. (2013) Understanding rice domestication and implications for cultivar improvement. *Curr Opin Plant Biol.* **16**: 139-146.

Sweeney M and McCouch S. (2007) The complex history of the domestication of rice. *Ann Bot.* **100**: 951-957.

Vaughan D, Morishima H, and Kadowaki K. (2003) Diversity in the Oryza genus. *Curr. Opin. Plant Biol.* **6**: 139-146.

Chapter 2

Evaluation of awn phenotype in chromosome segment substitution lines (CSSL).

Introduction

An awn is the sharp, spine-like structure at the lemma tip of some species in the family poaceae (Fig. 1). The spinous architecture enhances to seed dispersal by attaching to the fur of mammals and it also protects against animal predation (Grundbacher 1963). In addition, some types of awns can move depending on changes of humidity during daily cycle. The awn movement in wild tetraploid wheat, which has a pair of awns on each spikelet, propels the seeds into the ground (Elbaum et al. 2007). Thus, the long-awned trait is considered to be important for habitat expansion and survival of wild poaceae. Moreover, it is reported that awns in barley and wheat cultivars can support grain filling by contributing photosynthesis (Grundbacher 1963; Kjack and Witters 1974; Takahashi et al, 1986). On the contrary, awns of rice have been removed through the domestication. It might be because that the awn in rice has no chlorenchyma and stomata that are needed for the photosynthesis (Tatsumi and Kawano 1972). In fact, the ablation of awn has only a limited effect on the grain maturation in rice (Tsudamori 1933). In addition, long awn having barbs hinders manual harvesting under agricultural condition. So most cultivated rice lost their awns as an obstacle thing

during the domestication (Tatsumi and Kawano 1972).

Two cultivated species have been domesticated in the genus *Oryza*. The domestication of those cultivated rice species occurred in Asia and Africa independently (Khush 1997). The domestication of Asian rice, *O. sativa* has been domesticated from wild species, *O. rufipogon* approximately 8,000 years ago (Fuller et al. 2010). Recent genome-wide genetic studies reveled that *O. sativa* has been derived polyphyletically from *O. rufipogon* originated from the middle area of the Pearl River in southern China (Cheng et al. 2007; Fuller et al. 2010; Huang et al. 2012). On the other hand, *O. glaberrima* has been domesticated from *O. barthii* in West Africa approximately 3,000 years ago (Linares 2002). Despite the independent histories of the domestications, most cultivars in both *O. sativa* and *O. glaberrima* have awnless grains, whereas their ancestral species, *O. rufipogon* and *O. barthii* have long awn (Chang 1977, Fig. 2). Several genetic analyses have been performed for detecting awn-related genes in the past decades (Sato et al. 1996; Xiong et al. 1999; Lorieux et al. 2000; Thomson et al. 2003; Yoshimura et al. 2010; Sang and Ge 2013). Based on these analyses, some genes regulating awn formation has been reported so far, such as *An-1*, *LABA1*, *OsETT2* and

DL (Luo et al. 2013, Hua et al. 2015, Toriba and Hirano 2014). However, most of the reports focus on allelic variation solely within Asian rice, and it remains unclear whether mutations in these genes might be responsible for the awnless phenotype in African cultivated rice *O. glaberrima*. I tried to reveal whether the awnless phenotypes in *O. sativa* and *O. glaberrima*, which have been independently achieved in Asia and Africa, caused by mutations in the same gene(s) or not.

Here I show the detection of three awn related locus, *Regulator of Awn Elongation 1* (*RAE1*), *RAE2* and *RAE3* on chromosome 4, 8 and 6 respectively by comparing many sets of chromosomal segment substitution lines (CSSL). These genes are involved in the losing awn during the domestication process of *O. sativa* and *O. glaberrima*. My data suggested that *O. sativa* has lost the function of both *RAE1* and *RAE2*, whereas *O. glaberrima* has achieved awnless phenotype by the loss of *RAE3* function in spite of having functional *RAE1* and *RAE2*. This is the first example that different genes were selected for the same phenotype in Asia and Africa.

Results

Long awn-inducing loci in Asian rice

CSSL is the series of backcrossed lines which have only small chromosome segment of donor parent in the background of recurrent parent. A set of CSSL covers the whole genome by different chromosome segment of donor parent (Fig. 3). CSSL is developed by crossing with different two lines and continue backcrossing of recurrent parent several times. Lines of CSSL are selected by marker assisted selection (MAS) (Fig. 4). I collected the 6 CSSL which have the same genetic background of *O. sativa* ssp. *japonica* cv. Koshihikari as recurrent parent. Donor parents are W0054 (*O. nivara*), W0106 (*O. rufipogon*), Kasalath (*O. sativa* ssp. *indica*) as Asian rice; IRGC104038 (*O. glaberrima*) and W0009 (*O. barthii*) as African rice; and IRGC105666 (*O. glumaepatula*) as from Latin America respectively. We named each CSSL like WBSL, RSL, KKSL, GLSL, WWKSL and RRESL (Furuta et al. 2016; Furuta et al. 2015; Ebitani et al. 2005; Shim et al. 2010; Bessho-Uehara et al. *in preparation*) (Table 1, Fig. 5-10). Because these CSSL have same background, I can compare the effect from the chromosomal segment of donor parent equally. So the comparison of these CSSL makes

me identify the responsible genes for awn formation effectively.

Firstly I compared the genotype and awn phenotype of Asian rice CSSL because many studies related to awn development have been done on Asian rice. WBSL were developed by crossing *O. nivara*, which is the closely related species of *O. rufipogon*, as a donor parent with *O. sativa* ssp. *japonica* cv. Koshihikari as a recurrent parent (Fig. 5) (Furuta et al. 2016). In WBSL chromosome 6 segment should be indispensable for surviving because of reproductive isolation, so most of all the lines have the part of chromosome 6 segment of *O. nivara*. There are 2 lines; both of WBSL10 and WBSL18 have formed long awn, whereas *O. sativa* does not have awn (Fig. 11A-D). WBSL10 has introgressed a chromosome segment derived from chromosome 4 of *O. nivara*. The substituted segment has spanned approximately 25 Mb from the short arm end of chromosome 4 (Fig. 11E). WBSL18 has introgressed mostly whole chromosome 8 of *O. nivara*. (Fig. 11F). Thus, the substituted two chromosome segments derived from *O. nivara* induced long awn formation in the *O. sativa*. These results implied that the loss of function of genes located on chromosome 4 and 8 of *O. nivara* lead awnless phenotype in cultivated rice during the domestication process. I named each gene on

chromosome 4 and 8 as *Regulator of Awn Elongation 1* (*RAE1*) and *RAE2* tentatively.

In the evaluation of RSL which is the crossing line with *O. rufipogon* and *O. sativa* ssp. *japonica* cv. Koshihikari, I found that RSL11 formed long awn (Fig. 6, 12A-C). RSL11 uniquely has possessed a chromosome segment derived from *O. rufipogon* spanning approximately 19 Mb from the short arm end of chromosome 4 (Fig. 12E). The overlap of the substituted region in chromosome 4 between RSL11 and WBSL10 suggested that *RAE1* seems to be functional in both of *O. rufipogon* and *O. nivara* although the RSLs having substituted segments in chromosome 8 did not exhibit awn formation at all (Fig. 12D, F). This result indicated that the gene located on chromosome 8 of *O. rufipogon* that would be *RAE2* has lost its function for awn production. These results of phenotypic analysis in WBSL and RSL including the chromosome segments of Asian wild species suggested that the awnless phenotype in *O. sativa* has achieved by the mutations in both of *RAE1* and *RAE2*.

Kasalath is the cultivated species of *O. sativa* ssp. *indica*, nontheless it has long awn. When I compared the KKSL which is the crossed with Kasalath and Koshihikari, it turned out that two lines, KKSL206 and 210 represent awn phenotype (Fig. 7, 13A-D).

KKSL210 possessed a segment of chromosome 4 derived from Kasalath overlapping the *RAE1* located region (Fig. 13G). On the other hand, KKSL206 have been substituted the long arm of chromosome 2 to Kasalath (Fig. 13F). It suggested that there might be new gene of awn production locating on this region. KKSL224 which has chromosome 8 segment of Kasalath did not show awn phenotype (Fig. 13G), it showed that Kasalath has also lost the function of *RAE2* as like W0106 (*O. rufipogon*).

Independently achieved awnless phenotype in Asian and African rice

To verify the functionality of *RAE1* and *RAE2* in African rice, next I observed awn phenotype in GLSL which is *O. glaberrima* CSSL (Shim et al. 2010) (Fig. 8). In the 34 lines of GLSLs, GLSL13 and GLSL25 showed long awn formation, despite both of parental species have no awns (Fig. 14A-D). GLSL13 possessed a chromosome segment derived from *O. glaberrima* which has spanned approximately 26 Mb from the short arm end of chromosome 4 in the genetic background of *O. sativa* cv. Koshihikari (Fig. 14E). This substituted chromosomal region is overlapped with the regions in WBSL10, RSL11 and KKSL210. On the other hand, GLSL25 has a chromosome

substituted segment in the region spanning 14 Mb from the long arm end of chromosome 8 which is overlapped with the chromosome substituted segment in WBSL18 (Fig. 14F). These results suggested that *RAE1* and *RAE2* in *O. glaberrima* are functional for inducing long awn, even though *O. glaberrima* did not show awn. My data also indicated that *O. glaberrima* has achieved awnless phenotype by the mutation(s) in other gene(s) during African rice domestication. The putative awn inducible gene that is involved with awnless phenotype in *O. glaberrima* was tentatively named as *RAE3*.

When I observed the awn phenotype in WWKSL which is *O. barthii* CSSL (Bessho-Uehara et al. *in prep*, Fig. 9, 15A-D), WWKSL14 and 29 showed awns. WWKSL14 possessed a short arm end of chromosome 4 segment derived from *O. barthii* which has spanned approximately 6 Mb in the genetic background of *O. sativa* cv. Koshihikari (Fig. 15E). This substituted chromosomal region is overlapped with the substituted regions in WBSL10 (spanning approximately 25 Mb from the short arm end), RSL11 (spanning approximately 19 Mb from the short arm end) and KKSL210. On the other hand, WWKSL29 has a substituted chromosomal segment in the region

spanning 18 Mb from the long arm end of chromosome 8 which is overlapped with the chromosome substituted segment in WBSL18 (substituted most of the chromosome 8) (Fig. 15F). These results suggested that *RAE1* and *RAE2* in *O. barthii* are functional for inducing long awn. It means that these 2 genes work for promoting awn elongation independently and they are common in Asian and African wild rice.

O. glumaepatula has functional *RAE1* and *RAE2* genes

I also conducted a survey of awn phenotype on CSSL named RRESL (Bessho-Uehara et al. *in prep*) whose donor parent is *O. glumaepatula* inhabiting in Latin America unrelated to domestication (Khush 1997). This wild species, *O. glumaepatula* has a long awn as long as *O. rufipogon* and *O. barthii* (Fig. 16B). In the 35 lines of RRESLs, RRESL14 and RRESL25 showed long awn phenotype (Fig. 16A-D). RRESL14 has possessed a chromosome 4 segment and RRESL25 has possessed a chromosome 8 segment respectively (Fig. 16E, F). This result showed that *O. glumaepatula* has functional *RAE1* and *RAE2*, and is suggested that *RAE1* and *RAE2* are broadly conserved to produce long awn in the wild rice species all over the world.

Identifying of the locus for *RAE3*

The existence of *RAE3* causing the absence of awn in *O. glaberrima* was indicated from the awn induction in two lines of GLSL. To identify *RAE3* locus, I firstly tried to find the suitable mapping population. That is *RAE1* and/or *RAE2* are fixed as *O. glaberrima* allele and an unknown locus for *RAE3* was segregating along with the appearance of awn. Prof. Atsushi Yoshimura in Kyusyu University kindly provided us backcrossed lines derived from crossing with awnless *O. sativa cv.* Taichung 65 (T65) and *O. glaberrima* Acc IRGC104038. This population was backcrossed 4 times by *O. glaberrima* as shown in Fig. 4. In such the BC_4F_1 population, theoretically 96.8% of genome statistically should be fixed as *O. glaberrima* genotype, whereas only remaining 3.2% heterologous region can segregate in next generation. The phenotypic analysis of BC_4F_2 from 54 BC_4F_1 lines was conducted to find the population which awn phenotype was segregated. We found only 1 population consists of 90 plants segregating the awn phenotype in 3:1 ratio, 64 plants showing long awn and 26 showing awnless (Fig. 17A). The population has possessed *O. sativa* chromosome segments in

chromosome 2, 5 and 6 in spite of *RAE1* and/or *RAE2* region have been fixed as *O. glaberrima* genotypes. Genotypes of chromosome 2, 5 and 6 were identified by using SSR markers RM341, RM6346 and RM20699 respectively. By the linkage analysis between the genotype and awn phenotype, I knew that long awn phenotype was highly linked only with *O. sativa* genotype in chromosome 6. Thus, awn induction by *RAE1* and *RAE2* of *O. glaberrima* allele were suppressed by introduction of chromosome 6 segment in T65which putatively harbors the awn inducing gene, *RAE3*. It means that *O. sativa* allele is dominant for awn formation, so I can see the effect of either *RAE1* or *RAE2* in the genetic background of T65 (*O. sativa*).

Fine mapping was carried out by using 6,000 BC_4F_3 plants derived from BC_4F_2 plants possessing heterologous genotype in the candidate locus. The candidate locus were narrowed down into about 240 kb region flanked by indel-based markers, KG28941 (28.94Mb) and KG29180 (29.12Mb), which includes 68 unique genes (Fig. 17B). These results revealed that *RAE3* located in the long arm of chromosome 6 controls awn formation in rice and a mutation in *RAE3* has eliminated awn in *O. glaberrima*.

Discussion

Independently achieved awnless phenotype of the seed in Asia and Africa

By comparing various CSSL in the same background, it makes possible to compare the effects of the locus derived from the donor parents. Comparison of 6 CSSL revealed that there are at least three genes *RAE1*, *RAE2* and *RAE3* that regulate awn formation. In addition, it was shown that loss of function allele *rae1* and *rae2* were selected in cultivated rice in Asia, whereas that *rae3* was selected in cultivated rice in Africa for producing awnless phenotype. It means that Asian rice cultivar, *O. sativa* and African rice cultivar, *O. glaberrima* have obtained awnless phenotype by mutations in different genes (Fig. 18). This is the first example showing that the common “awnless” phenotype in Asia and African rice, which was considered as "Domestication syndrome", is caused by the loss of function of different genes. Moreover, according to that *O. glumaepatula* represents long awn phenotype and RRESL also conserves *RAE1* and *RAE2*, I speculated *O. glumaepatula* has the set of *RAE1*, *RAE2* and *RAE3* same as other wild rice species. Considered that there are no lines showed awn formation in WWKSL (donor parent is *O. barthii*) and RRESL (donor parent is *O. glumaepatula*)

except for the lines having *RAE1*or *RAE2* even possessing the *RAE3* located region. It is suggested the possibility *RAE3* could not work single but could work with each *RAE1* or *RAE2* cooperatively.

RSL23 having no awn suggested that *RAE2* has been lost its function, and also QTL on chromosome 8 was not detected when used F_2 population crossed with 'Aijiao Nante' (*O. sativa* ssp. *indica*) cultivar and a common wild rice (*O. rufipogon*) accession named 'P16' (Xiong et al. 1999). However, I think it can be seen specifically in some accessions but not in all *O. rufipogon*. Because many studies reported that there is QTL peak on chromosome 8 related to awn length by using inbred lines of *O. rufipogon* and *O. sativa* (Sato et al. 1996; Cai and Morishima 2002; Thomson et al. 2003). Fawcett *et al*. (2013) reported that genome wide association study (GWAS) detected the peak on chromosome 8 related to awn length by using more than 30 resequenced genomes of *O. rufipogon* and *O. sativa*. These studies are clearly shown there is functional *RAE2* allele on many *O. rufipogon*.

Epistasis among the awn regulating genes

Awn formation in GLSL13 and GLSL25, which have functional *RAE1* and *RAE2* from *O. glaberrima* genome respectively, indicated that *RAE1* or *RAE2* can induce long awn independently. Furthermore, it was also showed that mutation in *RAE3* leads elimination of long awn in spite of functional *RAE1* and *RAE2* in *O. glaberrima. RAE1*, *RAE2* and *RAE3* are locating on different chromosome respectively. It means all three genes have mutual interaction via “trans” relationship directly or indirectly. These results suggested that *RAE3* genetically functions as a common upstream or downstream factor of both *RAE1* and *RAE2* that are thought to act redundantly in awn formation (Fig. 19). *RAE1* was identified a bHLH transcription factor (Furuta et al. 2015) by Luo *et al.* (2013) as *An-1*. On the other hand, *RAE2* candidate locus on chromosome 8 includes no bHLH transcription factor. Therefore, it is thought that *RAE1* and *RAE2* are genetically redundant but have different molecular functions in parallel pathway. Detailed analysis on these three awn regulating genes will help to understand molecular mechanisms regulating morphological development of plant organs and domestication process of rice in Asia and Africa.

Different awn inducing loci

KKSL206 (donor parent is Kasalath (*O. sativa* ssp. *indica*)) showed the medium awn length regardless it does not possess neither *RAE1* nor *RAE2*. This result suggested that Kasalath may get the function of the gene located on chromosome 2 after separating from *O. sativa* ssp. *japonica*. The more other region of *RAE1*, *RAE2* and *RAE3* were detected in other wild rice species; such as chromosome 5 on *O. meridionalis* (Matsushita et al. 2003), and chromosome 9 on *O. minuta* (Linh et al. 2008). It suggested that awn phenotype was regulated by many genes in different species and the rice speciation was occurred independently at the old era not recently. Further studies using different wild rice like *O. meridionalis* or *O. minuta* could reveal the whole picture of complicated mechanisms of awn regulation.

Materials and methods

Plant materials and growth condition

O. sativa ssp. *japonica* cv. Koshihikari and the set of CSSL which have the same chromosome background of Koshihikari; WBSL, RSL, KKSL, GLSL, WWKSL and RRESL were used in this study (shown in Table 1). The plant materials were grown in the greenhouse of the Laboratory of Plant Molecular Biosystem until 5^{th} leaf stage and then transplant to the research field of Nagoya University, Togo, Aichi, Japan. All materials were grown under natural day length and temperature along with the conservational method until harvesting.

Genotyping of CSSL

Each CSSL was genotyped using 149 single nucleotide polymorphisms (SNPs) via the AcycloPrime-FP Detection System and Fluorescence Polarization Analyzer (Perkin Elmer Life Science, Boston, MA, USA) (Table 2). The SNP markers, which were developed based on the Build 2 Pseudomolecules of cv. Nipponbare, were evenly distributed among the 12 rice chromosomes at an average marker interval of 2.63 Mb.

Phenotypic evaluation

Out of 10 individuals of each line, 5 individuals' phenotype was evaluated. The length of the awn at the tip of the primary branch seed in one panicle was measured. The length of 3 mm or more length was judged if it has awn.

Linkage analysis and fine mapping of *RAE3*

For *RAE3*, totally 6,000 BC_4F_3 plants were produced from BC_4F_2 lines with a heterologous genotype in the candidate *RAE3* locus in chromosome 6. Each mapping population was genotyped using SSRs and newly developed insertion/deletion (indel)-based markers targeting the putative location of the gene. Genomic DNA from the mapping population was extracted using the TPS method. SSRs, and indel markers were amplified using standard PCR protocols and run in 3% agarose gel with ethidium bromide. Primers used for fine mapping are listed in Table 3.

References

Cai HW, and Morishima H. (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor. App. Genet.* **104**:1217-1228.

Chang TT, Marciano AP, and Loresto GC. (1977) Morpho-agronomic variousness and economic potentials of *Oryza glaberrima* and wild species in the genus *Oryza*. *African Rice Species.* IRRI, Manila, Philippines.

Cheng SH, Zhuang JY, Fan YY, Du JH, and Cao LY. (2007) Progress in research and development on hybrid rice: a super-domesticate in China. *Ann. Bot.* **100**: 959–966.

Ebitani T, Takeuchi Y, Nonoue Y, Yamamoto T, Takeuchi K, and Yano M. (2005) Construction and Evaluation of Chromosome Segment Substitution Lines Carrying Overlapping Chromosome Segments of *indica* Rice Cultivar 'Kasalath' in a Genetic Background of *japonica* Elite Cultivar 'Koshihikari'. *Breed. Sci.* **55**: 65-73.

Elbaum R, Zaltzman L, Burgert I, and Fratzl P. (2007) The role of wheat awns in the seed dispersal unit. *Science* **316**: 884–886.

Fawcett JA, Kado T, Sasaki E, Takuno S, Yoshida K, et al. (2013) QTL Map Meets

Population Genomics : An Application to Rice. *Plos One* **8**: e83720.

Fuller D, Sato, Y, and Castillo C. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol. *Anthropol. Sci.* **2**: 115–131.

Furuta T, Komeda N, Asano K, Uehara K, Gamuyao R, et al. (2015) Convergent Loss of Awn in Two Cultivated Rice Species *Oryza sativa* and *Oryza glaberrima* Is Caused by Mutations in Different Loci. *G3:Genes, Genomes, Genetics* **5**: 2267-2274.

Furuta T, Uehara K, Shim RA, Shim J, Nagai K, et al. (2016) Development and evaluation of *Oryza nivara* chromosome segment substitution lines (CSSLs) in the background of *O. sativa* L. cv. Koshihikari. *Breeding Sci.* 66: 845-850.

Grundbacher FJ. (1963) The physiological function of the cereal awn. *Bot. Rev.* **29**:366-381.

Hua L, Wang DR, Tan L, Fu Y, Liu F, et al. (2015) *LABA1*, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice. *Plant Cell* **27**: 1875-1888.

Huang X, Kurata N, Wei X, Wang ZX, Wang A, *et al*. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**: 497-501.

Khush GS. (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* **35**: 25–34.

Kjack JL, and Witters RS. (1974) Physiological activity of awns in isolines of Atlas barley. *Crop Science* 14: 243–248.

Linares O. (2002) African rice (Oryza glaberrima): history and future potential. *Proc Natl Acad Sci U S A* **99**: 16360-16365.

Linh LH, Hang NT, Jin FX, Kang KH, Lee YT, et al. (2008) Introgression of a quantitative trait locus for spikelets per panicle from *Oryza minuta* to the *O. sativa* cultivar Hwaseongbyeo. *Plant Breeding* **127**: 262–267.

Lorieux M, Ndjiondjop M, and Ghesquière A. (2000) A first interspecific *Oryza sativa*× *Oryza glaberrima* microsatellite-based genetic linkage map. *Theor. Appl. Genet.* **100**: 593-601

Luo J, Liu H, Zhou T, and Gu B. (2013) An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. *Plant Cell* **25**: 3360-3376

Matsushita S, Sobrizal KT, Doi K, and Yoshimura A. (2003) Mapping of genes for awn

in rice using *Oryza meridionalis* introgression lines. *Rice Genetics Newsletter* **20**: 17-18

Sang T and Ge S. (2013) Understanding rice domestication and implications for cultivar improvement. *Curr Opin Plant Biol.* **16**: 139-146.

Sato S, Ishikawa S, Shimono M, and Sinjyo C. (1996) Genetic studies on an awnness gene An-4 on chromosome 8 in rice. *Breeding Science* **46**: 321-327.

Shim RA, Angeles ER, Ashikari M, and Takashi T. (2010) Development and evaluation of *Oryza glaberrima* Steud. chromosome segment substitution lines (CSSLs) in the background of *O. sativa* L. cv. Koshihikari. *Breed. Sci.* **60**: 613–619.

Takahashi N, Alwan H, Alterfa H, Sato T. (1986) Significant role of awn in rice plants(1): A survey of agricultural value of rice awn. *Reports of the Institute for Agricultural Research Tohoku University* **35**: 21–31.

Tastumi J, and Kawano K. (1972) 水稲の芒について,日本作物学会東海支部研究発表梗概 56: 11-15.

Toriba T, and Hirano HY. (2014) The DROOPING LEAF and OsETTIN2 genes

promote awn development in rice. *Plant J.* **77**:616-626.

Thomson M, Tai T, and McClung A. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar. *Theor. Appl. Genet* **107**: 479-493

Tsudamori M. (1933) 生理器官としての稲芒の価値, 日作紀, 5: 380-390

Xiong L, Liu K, Dai X, Xu C, and Zhang Q. (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor. Appl. Genet.* **98**: 243-251.

Yoshimura A, Nagayama H, and Kurakazu T. (2010) Introgression lines of rice (*Oryza sativa* L.) carrying a donor genome from the wild species, *O. glumaepatula* Steud. and *O. meridionalis* Ng. *Breed. Sci.* **60**: 597-603

Tables and Figs

Table 1. The list of 6 CSSLs which have same background.

CSSL name	Donor parent	Habitat	Recurrent parent	Line No.	References
WBSL	*O. nivara*	Asia	*O. sativa* ssp. *japonica* Koshihikari	27	Furuta et al. 2016
RSL	*O. rufipogon*			33	Furuta et al. 2014
KKSL	*O. sativa* ssp. *indica*			39	Ebitani et al. 2005
GLSL	*O. glaberrima*	Africa		35	Shim et al. 2010
WWKSL	*O. barthii*			40	Bessho-Uehara, *in prep*
RRESL	*O. glumaepatula*	Latin America		35	Bessho-Uehara, *in prep*

Table 2. A list of primers used for the SNP genotyping of CSSL in this study.

Chr.	Position (bp)	SNP_marker_name[a]	Forward-primer_seq	Reverse-primer_seq	SNP_primer_seq[b]
1	590339	SP-1177Ct	GTAGGTGCAAAGTGTGCC	CAGCATGGTGTACCAGGA	CTCTTTAGGGTGTTACCATG
1	1946270	SP-155Tg	TCTCCTTCCAGTCGATGC	TCCCAAATCGCCCTCTTTGC	AGCGACATGGTTGCTCTA
1	6522242	SP-2077Tc	CCTAATGGGCCGAATTTATAACG	GCGAAAGGCGAGGTTGATG	CAAAGTAATTTGGGATCTTTAC
1	8896815	SP-185Tc	GATGTACGTCGGACTATTCATC	CGACGTGCATGCTCATTAGC	TTAAACATATGTCCAAAGTCAA
1	11026011	SP-192Tg	TCCCTGCTGCTAGGATCTTG	ATCCATCAAGTCAGCAGGTG	GCTGTGAATGGATGCCAA
1	14312329	SP-5191Tc	GCTATCATTGGCTAGAACACAC	ACGGAGTTCCAGCGAATGAG	GATCGATCTTCATTAACTTCAATCT
1	17803741	SP-1208Tg	GAGGCCATTCTCGCAAC	ATCGACCGAGCATTCTAGC	TCACTCTCTCACCGATC
1	20366961	SP-2016Tc	GCTAGAACTAATTCGCAGTAGAG	CAATCTCGGCCACAATTACATAG	CACCGCCACTCAATTTAC
1	24412981	SP-1217Ct	CTCCAGTGATGTTTATTTCGTCC	CTAAACAGGTTGCATTTTGGTGC	TAACTGTTTGACTTATTTTATCA
1	25979477	SP-2478Ag	GTTCCAGATACACCATACGG	GCCTTAATATGGAGGAAAGCG	TAACCTCGGTTCAGGAATC
1	27840561	SP-242Ga	AGCACGGCTATGTAAGACTAAC	AGAACGTCATCGACAGCATG	TGGACCAAAACTTCACTTAGTA
1	31371175	SP-262Ga	GCAGCAGGACAAAGGCTAAC	ACCCTTCTTCAAGCTCCATC	TCACAACCGGACCAGATGAC
1	35598472	SP-1244Tc	CCTCCATTGATGTCGGTCA	AGCAGCCCAATTATCTTTGGAC	GGATATATTCACTCTTTTCACT
1	39677014	SP-2056Gt	TGCTGGTGAAATGTGGGTAG	CTCTGCAACCAACATTGCAC	CAACAAGCACACAAGATGAAA
1	42000585	SP-2079Ta	AACTCACCAGCTGGTAGGAG	GGCCAAGCCCAACTTAATCC	CACACCAATGGAGGATTGT
2	151976	SP-3541Gt	GAGATGGCGTGAGTACTAAG	GTGTAAAGGTGGCTTATAGC	TAAATCGGGCACTCTTTGTT
2	1758809	SP-3547Ag	TAGTTGCCTAACCAGTAGGC	CAATCCAGATTCATAGCCAGAG	GATTGACTACTTCCTCAATTTTC
2	3922571	SP-1387At	CCAAACCAACCGCTTCCAC	CCGCTTTATACACCCGCATC	ATGCGCGCTTTTCTTTCTTT
2	7030897	SP-1398Gt	CAACAGCTGTGCAGCAGTTG	ACTTTCTGAAGGCGCCTGC	GGTTTGGTTGATGAAACTAGCT
2	9256584	SP-29Gc	GGACAGGAACAAGAGAATGTG	AAGATGGTGAGACAGGGAGA	GTAACAGTTGTCTGGCTA
2	12146325	SP-33Gt	GCATGTGTCACATGGTTCTAAC	GGATGTCATGAACCCATCAC	TTGATTTTGAGGTAGCTTTG
2	14655638	SP-1413Tc	CCTCATTGGACGGTAAACGC	GATCTGCTGTTGGTGAAGCTC	ATCCACCACCTGGATG
2	18881343	SP-3783Gt	GTCCAGATCAGTGTGAGTGC	ACCTTTGAACCACTGCAACC	TTCTGCATTTCGTTGTGATG
2	22086348	SP-954Tg	GCAGCAAAAGGTTAACGACC	GCCTGCTGCATTCTCTAG	CACATATATAGGTAATTTTATGGA
2	25263253	SP-2941Ct	GTACCATGGTTGCTAAGTTGG	CACTACACCAAACCTTCTGAC	GGCTACCAAACCTCATTTTC
2	27919903	SP-1444Ag	ACCACACAAACAAGACTCAGG	GTACGGCGTCATGAATGATC	AATTTGTCTGCTCTGCTCT
2	29345357	SP-1450Ga	GGAACAACAGTCCAGGCTGA	TGAAGCATGGACATGTACGG	ATTGCTGGAGGCTCACAT
2	31272546	SP-1461Ca	CTGACTGCATTTGCTTTGAAGC	GGATGGTTCATCATCTGGTC	GAAGGAAAAGTTAAAGGATGAT
2	33189264	SP-1471Ta	TGGTGAGGGATGATTCATCC	GTATGGCAGTATGCTGGTTC	GCAAAGCAGTTACATATACT
2	35267219	SP-3571Ta	GAAGGCAGCCAGATGAGC	CACTCGCTCTGTATCAATTGG	CTCATCTCCTGGCTAATATC
3	577726	SP-111Ag	GCGTCGTGGATGATGACG	GCCTCCTCTTCGGCTTCG	GTAGACCTCGTCGTAGGA
3	3979349	SP-1289Cg	CATGCAAGAATTGACCTGGC	GCTTAACTGACGATCGGATC	GGTCTCGGAATTCCACT
3	6660541	SP-1302Ct	GTAGAGCCGGTAGATAGTC	TCAAACCGAGCGGTGAAC	GATGCCTCACTTTTTACAGA
3	8780254	SP-1311Ga	TAGGTAATGCTCCGCAAGTTC	CTTGACCTGCCTTAATGTGC	AGAGCATGGAACAGGACA
3	11073623	SP-1323Ta	CTATGGAAGAAGCTGCATCG	CATTGATGCATTGATTGGGAGAC	GATTGTTGAGCGATTTTCGAG
3	13984287	SP-1331Ga	CAATTGTTTCATCGGCCTCAG	TGCTGCAAATGGTGAAGTGC	CAGTATGTGATCTAGAGAAAAAC
3	15737302	SP-294Ct	GGCACGGATCTGATACAAG	TGCAGTTCTTGTGCCATCG	GAGCTACCCAGAAATTTAGC
3	19071470	SP-3143Ag	GATCTGCTTAACTCCAGTGC	GACATTGGCTCGGTTTAGAG	GTGGTTTCTTTGTCTTTATTTCTT
3	22287129	SP-306Tg	CATATTCTACAGCGTTCTCGTC	AACACCAAGGGCGATCGAG	ATACCGAGCCCAGCAAT

Chr.	Position (bp)	SNP_marker_name[a]	Forward-primer_seq	Reverse-primer_seq	SNP_primer_seq[b]
3	24488442	SP-3148Ac	GTCTGATCGATCGAGCAATC	TGCACAGTGGATTACGCTAC	TGTCCAATGCAACATCTCC
3	26352992	SP-3135Ag	TAGCAACTAGGACAGATGGC	CTCAGTTCCATCATCAGGTTAC	GATGTAAATCCGCAAACTGCATA
3	30021442	SP-1355Cg	CCTAGCAGCAAGATATGAACTG	GCCTGAATTCTGCAATACATTGC	CTATGGATGGATCCAGACA
3	31580078	SP-334Ct	GCGTGACATATAGCAATTCCG	GGTGTTCATCGCCATTAATGG	GCAATTACGACTCACGAATTCT
3	33640115	SP-348Ga	CCAGAGAACCCTTTTATGTACAG	CCGAGACATGATATAGTAACAGC	ATTGGATTTCGCATTTTGCG
3	35945011	SP-362Tc	CTGTGGTTAGCTCCTAAGGC	AGTCAGGACTCAACTCAAGC	TCTACAGCGCAGATTCGG
4	218809	SP-2590At	CTACACATTAGCTCGCTGGA	CACTGCACAACAATCAAGATCAG	CCATTACTTCTATACGTGATA
4	2471151	SP-2595Ga	GCCAAAACACTAGTGAGTGAG	CACCACAATGAGGTATCCATC	CATGTTTTTGTGCATCCTAGTT
4	6456488	SP-372Ct	CTGACAATTGATGCAGACGC	CTGCAGACTCCTTCTAAGC	CGTGACCTCATTCTCTGAAA
4	8285219	SP-3233Ag	CAAATGATTGGAGACCAGGAC	GATGGCATATCTCGTAGGATG	GTTATTTAGCTCATAATTTGCC
4	11949079	SP-375Tc	GCTTATTCTGACAGGTTGCC	CATATGACTGATTGGATGGAGC	TCCAGGTCCTAGCTTATTGC
4	19650008	SP-386Tc	AGTGCCTGTCAGGATCAG	GGATTCAACTGACGAGTATTGC	GTAATCTATCTTGATGGAGTATTA
4	23043128	SP-3269Tc	CCTTAAATTAGCAGGACCTGTG	GAAAACCGCTAATTGACCTCAG	AATAGTCAGTAGTCACCTGC
4	25877141	SP-402Ga	CTTGGTTCATTTCCTCCAGAAG	CATAACTGAGCACCTTGGTG	CTCTTAAACAACACTTTTGCCA
4	28349285	SP-409Ac	CAGAGCAAATTGGCTGCC	GTATGGCACCATGCTTATCC	CCATTTCCACAACTTTAGTTTG
4	32342809	SP-423Ga	AAGGAGAGCTTTGAGCACG	GGATCAACTTGCTTCAGCTC	CTGGAGGGTGTGCCTTC
4	34498729	SP-436Ct	CTCAGTAAGAAGGCACTCG	CTTGTAGCTAGCGCTTGTC	CTCTGAATGTATCAACCAACATA
5	40031	SP-487Ag	CCTTGATCGAGTTTGAGCTG	CCCATCAACCATGATTCCAC	ACATACCGTCCTCCTAC
5	1047930	SP-495Tc	CCAGTCAAAGCCTCAGACC	CTCTCAATGCAAGGTCTGAC	AGGATGACAAGGGCGATTA
5	3173133	SP-505Ct	ACAAGATCCTCAAGCATGGC	CATTCATGAACACCGCGTTC	CAATGTACTTGAAGTCGCTCAT
5	5073404	SP-516Tg	CTTGGTATAAAGCCGGTGTTC	CCAACGTCGAGGTCTGAAG	AATTGCAGTAGCCCTGCT
5	7305956	SP-1526Ga	GGTGATGCATGACCTAGG	CCACCTTGCTGGTGATCA	GCTCAGCTGAGACTTGTT
5	12908872	SP-526Gc	CCGACCGTCATTGTGTAGTG	GTTCTGGCCAATATGGCTTG	AGTCATCTTGGCCAACAGTT
5	15459785	SP-1537Ac	CAATGATGCCAGGTCTGACAT	TATGCAGGACACTCTCCATG	TGTACAGTCAATAATGCACTTG
5	18878114	SP-1546Tc	CTGCTTGAATGATTCAGTGGTTC	GTACCGTAGTTGAGCATTGATG	TGACCAGCAAATCTGGGA
5	21331373	SP-548Tc	CTGTTAGGTGGTAGTATTAGCC	AACCGAAGACATGGATTTCC	CGGAAAGCGGCGGAGA
5	24581478	SP-563Ct	ACTTGAGAGCCATGGACTTG	AGCTCACTGAGAGTAAGTGC	CAAGAAGTTAGCTTGCTGGA
5	26499154	SP-1573Ag	CATGTTCACGTAAGAGCACATG	TCAGTCACTCCGGTATCG	CCACAACACAAAATAGCATTAGTTTC
5	28699770	SP-3325At	CTCTGGCTGCAAGTCATCAC	CTTCTGAGGATGATTCCAGC	ACACATACCTTCTTGGGAGT
6	372258	SP-590Ag	CAAGTTCATGTCAAGCTCGG	GATGGAGCTTACAGCCAG	TGCCGCCAGGTACATG
6	2584447	SP-2505Ag	TCGTGGAGCTATGTCTGGAG	GTGGATGGCACACATTAATACAC	TACCACTAAACCATGCCTAG
6	4185220	SP-600Ct	GGACTGTCACTATCATAGGC	CGTCTGAAACTGTGCAATGTAC	CACTATCATAGGCACGTTTC
6	6290821	SP-2510Ct	GCCATGTAACACAAACAGTGG	TGGTGCTGGTACAGAATACC	ACAATTTTCCGACACTATCT
6	7554759	SP-611Tc	CTGAAGTATCCTGCTGACC	CTCCCTGAAGATATGCAACTTG	GTGTTGCCATTGAAACTGTG
6	10671175	SP-1603Tc	CCTAGTCCCTAAAGATCTCATG	GATAGACATGACGGAGAAGTG	GGGTGGTGTTATCTCTAGT
6	13689109	SP-3005Ct	CTTCTGAGTGAATCGACGAC	ATTTGCTCTGGTGGAGATGG	GGTATTAGCTAAGCAATCACA
6	16525923	SP-2270Ta	CCTAAACAGAAGTGCCAC	GGTGCACGATTGAACTCTATG	CCAGAGGAGGCTTCAGA
6	20235922	SP-3022Cg	GGGTGGATAAGTCAACTACTG	GGCCTAATATGTATGTTCCTTGC	TTTACGGCAAAACAGATCTT
6	22907151	SP-1613Ga	GAAATCACATCCGATGACTGG	CACTCATATGTGGATGTGTCAG	CAATGGCCATAAGGAAATGC
6	25711515	SP-3026Tc	GGACACTTATCTCAAGGTTTAGG	GCACCAAGTTCACTCATGGC	CAAAAGGGCAACAATCTCT

Chr.	Position (bp)	SNP_marker_name[a]	Forward-primer_seq	Reverse-primer_seq	SNP_primer_seq[b]
6	28329768	SP-1629Ac	TCCTTCTGTGTACTGGCATC	ACAACCGTGTGGATGCACTG	CCTAAGGGAGCTGAAACT
6	29975546	SP-673Ca	AATCGTGGGTGATGGGTCG	CATGCGGCGTAGAGTAGG	AAGGCCTGGAGGAAGG
7	260864	SP-676Ta	GTTATAACCCTGCACACATGAC	GTCGTTTCGTCGTCATTTCTG	CTGGATTTGTTTTACAAAAACCAA
7	1743980	SP-4168Ag	ACATGGTTGTCATCGATGTCC	ATCTTCGCTATGCTTCACACG	GTCTTCCTGCTGAAGATGAC
7	3898132	SP-683Ga	GGGATCTTCAGTAGTAGCAAC	GCCACTATAGTTGGGAATCATG	CAAAGGTATGGAACATAATACA
7	6843696	SP-1648Ta	CACGGGAACATCATACTGC	GTTAGGTGTCTTCGGCAAG	TGTAATGATACCTCACTCCA
7	10130032	SP-3066Ag	GCAGGGCCATCTTGTGATG	CAGCTTCACTCTCACATTCAG	TATGGAACAAAGGCGCCAT
7	13302019	SP-1657Ag	TCAGCCAGAGCGATCTCTG	CTATGCGGAATCATACGTAGC	CTCTTTCACCAGCAATAGCT
7	17327528	SP-2345Ga	ATGCTCGACATGGCATTGC	GCGATATCTACAACACCAAATCAG	TGCTTATGTGTGCCAGCA
7	18571959	SP-1662Ct	CGCAGCAGTAAGTAGGTTAC	CGCCAAATGGATGTAGAAATGC	ACTGACAGTCTGACACCAAA
7	21942076	SP-3095Gt	CATTTGGATGGATCGTGAGTG	GGAAACATGAGGCCGTACTG	TACTACGAGTCAACGGATTA
7	25490053	SP-741Tc	CAGAGCTGGCAACTTATGTAC	AGCTTCCTTAACTGGACCAC	TGAAGAAGCTGAAGGACTTG
7	27473590	SP-749At	GTCACAAGACACTACAACATGC	TGAAACCGCGGATTTAACGG	GAAACAACACAAGGAAATCC
7	29361700	SP-2356Ag	GTTCTGGAACTAGAATGTGACTG	GAACAAGTTTGGGCAGTGAG	TTGGCACTGCCTGTAG
8	195223	SP-761Tc	TGAAGATGTACACTGCGTCC	AGTTCTGCAGTTTCCACACG	CTTGCTAGTTGTACCCCAA
8	1669653	SP-976Ct	CAGAGATCAACACACCAATGC	CACCAAGACATGTCGGTTC	GCAATTACTCAACGGTAGGTAAA
8	3553188	SP-2551Tg	GGATAGCCTTGCACTTGGAG	CAGCAACAATGGCTAATTGCG	CTTTCAGTCAATCATTGGACA
8	5278855	SP-3598Ag	CCATACCTGAAGAAACTGTCC	GGATGTTAAACCTGAACACTGG	TCAGAGTTGAAGTCGATCCT
8	8094582	SP-795Ct	CATTGCATGTGGAGGCTTG	GCGGTAACAAACGAGCATG	CTTTGGCTGGCTGAAGAAA
8	10572009	SP-1707Tc	ACGTGCAAGGTGTATGCC	CTTGTCTACTGCTCACCGA	CAGCCTGTGTTACTCTGTT
8	14520452	SP-3608Cg	CGTGAATGAATGGCGATCTG	TGGGATTAGTGTGGCTCTTG	CCTCTTCAAGATAATAAATGCTA
8	18114324	SP-809Tc	GCTTACATCATGATGCACGC	CAGACTCCATTCTCAAGACC	GCTTTGAAGAAAAAGGCATC
8	19903905	SP-3199Tc	GAGGCCCTATAAGTTTGATTGG	CATGCGCACTTTCTTAAACGC	ACAAACGCAGACCCTCATAA
8	23167567	SP-2408Gc	CTCTTGTGGGAAAGTAGGAG	CAGTGGAGTAGCTAGCACAG	CTGAGTCCCCTTTGCT
8	25921712	SP-1734Ac	CTACTGTGCCGAGATTGC	AAGGAGAATACTTGGAGGCG	GTCACCAACAAGCTTGC
8	27758654	SP-844Tc	GTTTGCTTCCTAGCTACAAGC	ACAGTAACACGGCGAGTG	CAAATGTCTGAAACAGTTTGTAA
9	1	SP-3968Tg	TCGAGGTGACACCAAATGGC	GCATTGCAACTTGGAAGTTGTG	AACTTAGTTCACTATACTTGGAA
9	452497	SP-2183Ag	CAATGTACGATGTAAACGACAGC	GCTCCTCTGGAAGGTAAG	CCGTAAAATTTCTCAATTCCAAA
9	3066192	SP-866Ta	CAACATCAAACACAGAGAACGAG	GTTTGAAGAGTTCCTTTCTAGCC	CGGAGCCATACATTATCATC
9	5167966	SP-2129Tc	AGTGCTTCACGCGATAGTCT	GTTTCACCAACCTAGTTACGG	GTATTGAACGTTGGGTCCAT
9	6999573	SP-2144Ag	GACATCTGCCATTGTGTCC	GCGGTTGCAAATTCTGGTATTC	CAACTCCATGTAATGTGGATAC
9	10483155	SP-875Tc	GCGATTCGAAAGAAGCTAAC	GTTGCGAAGAGTACAAAGGC	CTAACAAACTATGAGGTTCAATATA
9	13253953	SP-2180Tc	CTAGGGAGTCAAATTACACCG	TACATTCTGGGCACCGTTCA	CATACCCCTTAACAAGAAAAC
9	15063825	SP-1747Ga	CAACAGCAAGATTGGGAAGC	GCAACTGCAATCAGTATACGG	CGGCTTTGTATTGCCAATTT
9	17880593	SP-909Tc	TGCAGTGCAGTACTACATCG	CAGCTACAGATGTACACCATAC	ATACCAATTTCTCTGACATG
9	19336407	SP-1761Ct	ACAGCCATGCTCGACTTCAG	GCAGCAGAGGAGATGTTCTC	AAATGAAATGAGCACCTCCT
9	20652100	SP-2125Tc	GGATGAACATGTAGGGTTCG	GCTATTCCATTTTGGTAAGCGC	CCAACTCCATTTGTCAGCA
10	94506	SP-924Gc	GTAAGTAAGTGTGTGTAGTATAGG	GCTCGACAGATGGACCTCTC	CCATAAACGGACATATATACG
10	2541199	SP-932Ag	AACGACAAAATTCTGCAGCAAC	CTACTAGTTTGTTGGAGAACC	GACAGCTTCTGCTTGTTTGT
10	7069227	SP-948Tg	AGCAATCTCATCAGCCAGATC	GATTTATGCTGCAGCTTGGC	GTTGACACTTCTAAAGCTG

Chr.	Position (bp)	SNP_marker_name[a]	Forward-primer_seq	Reverse-primer_seq	SNP_primer_seq[b]
10	9791595	SP-956Tc	GGCGAAAGAATTGGTGTTCAG	GGTGCTTAAAGATGAAAGGCGA	CCATATCATGAAATTCTGATGCAT
10	11824766	SP-3358Cg	GCCTTTCTCATGCTAACAGCA	GGCATACATCAGGACCTAG	AAGCGTTGGTGTTGACTAT
10	14502855	SP-1774Ag	CGATAAGCCAGGCATCCT	AAGGCATAAGCATGCTCGTCA	GTCAGAAAAATGAGGGAATAC
10	17555159	SP-981Tc	ATGGGCTGGTGCACTACTG	CCTTGGTCTGTTACATAGCG	CTTTGATATCCGGGAAGG
10	19468326	SP-992Gc	CCCTTGTAACTTGTTCTTCGTG	GCTTATCCTCCAAACGACAAC	CACATGAGCTAGATGATGAT
10	21171786	SP-1794Tg	CTCTTGTTTACTCGATTCGCCA	GAACGTGTAAGCTATATGGCAG	CCAAGTGATTCAAGAATTTTACA
10	22686557	SP-1013Ag	AATGGTCTTCCCTCCGTTTGC	GATTACCAGCTTGAGTACTCG	GTTTTTGTCTCCTTTTCTCAAA
11	827222	SP-2650Ga	GCTAATACCTTCCTATGAAAGCTC	CGCTCTGCAAAAGGCAAG	GTGTGTAATTGGAGACAAAGCA
11	2489967	SP-2816Tc	GCCATGGCGTATTATTAGCAC	GCATGGAATGGATCTGAACTG	CGCATTTCTTCAGTTTTCATCT
11	4005384	SP-2656Ta	GATGAAGCTCATCGAGCCT	GAGCTTGATATGGAGGAAACC	GTATCAAATATACCTCACGGAC
11	5687240	SP-1034Ga	GAATGTAGATTGACAGATTGCAC	CCTATATATGCTGGCTCCAC	CACATTACATTAGAGAGAGCA
11	7481468	SP-3417Ag	CGAGAAGAAGTATTCGACACC	CTCTGCATAAGTGAAAGTGAGC	CGTCTTTGTTTCCGTCGTAT
11	11922302	SP-3385Tc	CCACCCAATCCTTGCATGG	GAATTGGAGCTAGTAGTGAGC	TACATGCAAAACTGAACAGC
11	17424916	SP-2672Ag	GTGTTGGTCAATTGTGCCATC	GAACTGTCGGTTAACTTTGAGG	CAACAGCCTGGCTTTTGT
11	19050305	SP-1062At	TGCGCTTCTACATGACCTG	AAGTGACCTGCTGGAGC	CTGTGTAACTGAGAAGGTACAAA
11	21896773	SP-1073Tc	ACATACGGCGAGTGCAC	GAGAAGGTTGGAACCGTG	ATCTTGCGATGCAACCATA
11	24002145	SP-3403Ct	TGAGAGGTGTAGAGAAGAGC	AGCCTGTAGACCCTCACAG	CTGAAGAAGCATGCAACTCTT
11	25561274	SP-1085Tc	GGTTGGATATCATCTGAACTGC	CTCGATCACACTGAAACTGC	GGGCGGATATATATCCAAT
11	27656337	SP-5095Ga	GCAGTACCTTTGTCCTTAACG	TTTGGTGCTCCTTCCTCTGC	GTCATCTTGCCTTCCTGGG
12	322928	SP-2831Gc	TGTGGTTGGTTGATGTGGAG	CACAAACTCCACCAAACCAAG	CACTATTCCTTCCCGGATT
12	2225082	SP-1106Ac	CAATGTTTCAAGTGTGGGAGCT	CTCAGCTACCTAGCTTATGAG	TTGCTTCAGTAAGAGAGGAC
12	3849039	SP-1115Ac	AAGCATTGGCGCCTCATCG	TAGTGCCTAGGAGCTCTCTG	TTTCTGCCTGGAGAAACACA
12	5510849	SP-3499Ta	TCACAGAGGCCACTGTTCTG	CAGATGCCATGTTAGTCCTTC	ATCGATGAAATGCTGGGAGA
12	10041580	SP-1828Tc	AACGTTAGGCAGGTGCAG	GACGATGAACAAACTGGGAAC	CGTCATCGTCATATATAGCCT
12	12334210	SP-3455Ag	CTCACTTTGAGCATACTCCTC	CGTTGGTCCACAATCTGAC	ACTTGACTCCTCACTCAC
12	14689883	SP-2687At	ACTACAGTGATGCTTCTGGC	GTATTTGGCAGTGCGTTCC	GACGTAGTAGCCATATTGGA
12	17499935	SP-3747Cg	CTATTGACCTTGTTGCTGGTAC	CTCAAGCTTAAGTGTGTTGC	CAGTGTTTAAGCTTAACAAAGCTT
12	21257971	SP-2697Ct	ATGATTTGACTCCCGTAGTCC	GGATGAGAGCCTAAGCTTAC	GTAATCCTTGCTTAGTTAGTATTA
12	23828918	SP-1160Tc	ACGTCATGCACCCTGATCG	GGGCTTCCATGGTTCATGC	AACTCGTCGTTTCTAGGATC
12	27011209	SP-2705Ag	CTTCAGCGGGAACTCTGTC	GTTAAGGTATGATCCGTTGCC	TCTCTAGGCATATCCCTTTC

[a] Each SNP position is located at 1bp-downstream of the corresponding SNP_primer_seq.

[b] A capital letter and a small letter following each SNP_primer_name indicates a Koshihikari-allele and a rufipogon-allele of the corresponding SNP.

Table 3. Primers for fine-mapping of *RAE3*.

Marker name	Sequence	Chromosome
KG26425F	AAGACTTTCTTTTGCCGGAC	6
KG26425R	TGGGCTTTGATGAATGAAGT	6
KG27612F	TAGGTAGGAGTAGGCCGGAT	6
KG27612R	GCATGCACATATGTCACTGTGTAA	6
KG28331F	CGATCTCCTTTGCATCTTTC	6
KG28331R	GGTGGTTAGCACCTTTGTGT	6
KG28644F	CTCATGCACAAATCATTCCA	6
KG28644R	AGCCTGTGCACAAGTTCAG	6
KG28724F	GACTCGCCTAGAGATGCAAG	6
KG28724R	AAGCGGAGCCTAAGAAAGTC	6
KG28805F	GTTTCCAGATGCTGTCCATT	6
KG28805R	GCACTTACCAGGTTCCATTG	6
KG28941F	CTCCTCTGATCACCTCGCT	6
KG28941R	GGAGAGGAGCAGCTTCTTG	6
KG29231F	ACTCGTCGTCCACCTCAC	6
KG29231R	TGGTTGGGACGCTAGTTATC	6
KG29384F	GTTTTGCTCAGGCAAATGAT	6
KG29384R	GCACCCAAGATTTTATTGGA	6
KG29467F	TTGGGATTACCGAGTTCTTG	6
KG29467R	GTCTCTCCACCTTTCCACCT	6
KG29588F	AGTGGTGGACCAAAATTCAC	6
KG29588R	AAACCAAGCATGATAATCGG	6
KG29722F	ATGTTAGCCTTTTTCCTCCA	6
KG29722R	GGTCTTCGTTAGTCTTATGCATCT	6
KG29792F	GAACGAGTGCAGGTACGC	6
KG29792R	ACCTTGTACCACCTCATCCA	6
6KG30196F	CTTGTTTCCATTTTGTTGGG	6
6KG30196R	GGAGGAAGAAGAGGACGAAG	6
6KG30327F	CCCGAGCCAGACAACATTCC	6
6KG30327R	GAGGTGTGAGGTGAGGAAGATGC	6
6KG30986F	AGAATACTTTTACCACCCTTG	6
6KG30986R	GTTTAATTAATCCAGGTAGGTCCT	6

Figure 1. Awn is the spinous like structure at the tip of the cereal seeds. The picture of cereal seeds with long awn; barley (Morex), rice (GLSL25; a line of CSSL derived from crossing with *O. sativa* and *O. glaberrima*), wheat (Norin 61). Red arrows point the awn.

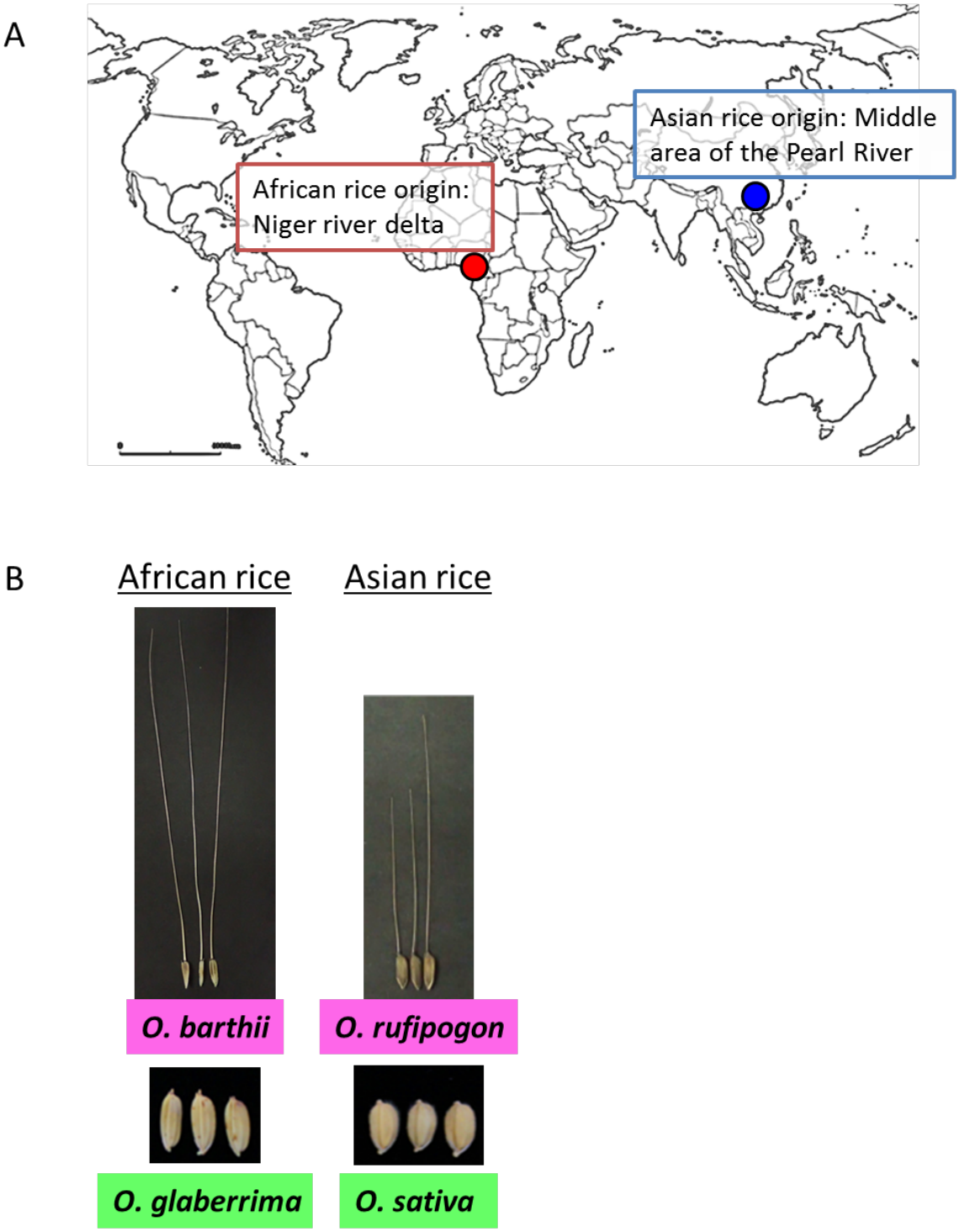


Figure 2. The origin of cultivated rice and seed phenotype of each species. (A) Red and blue circle represent the location of African rice and Asian rice domestication occurred, respectively. (B) *O. barthii*: wild rice, ancestor of african cultivar, *O. glaberrima*: cultivated rice in Africa, *O. rufipogon*: wild rice, ancestor of asian cultivar, *O. sativa*: cultivated rice in Asia

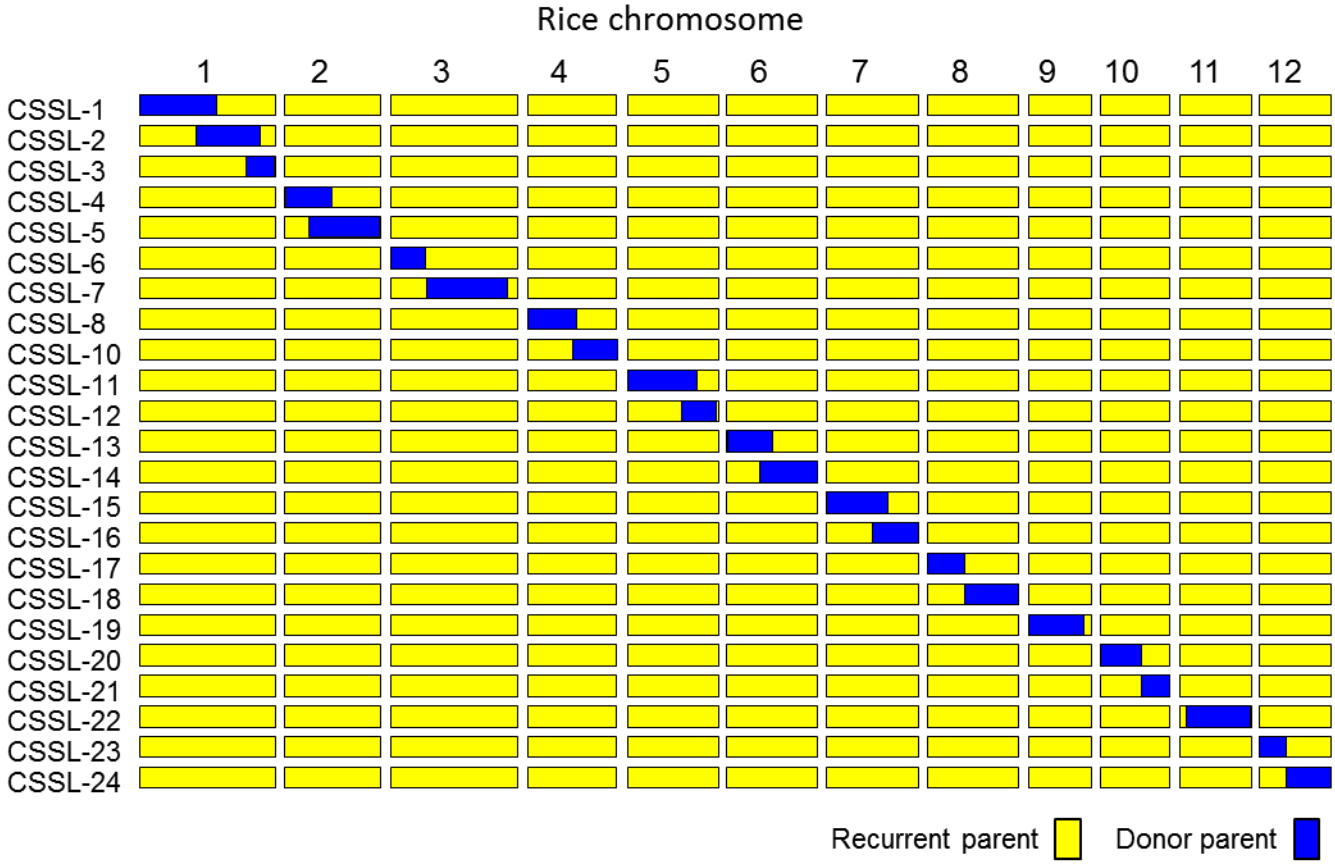


Figure 3. Example of graphical genotype of chromosome segment substitution lines (CSSL). Yellow square represent the chromosome segment of recurrent parent and blue represent the chromosome segment of donor parent each other. This example suggests that 24 lines cover the whole chromosome region from the edge of chromosome 1 to chromosome 12 by donor parent.

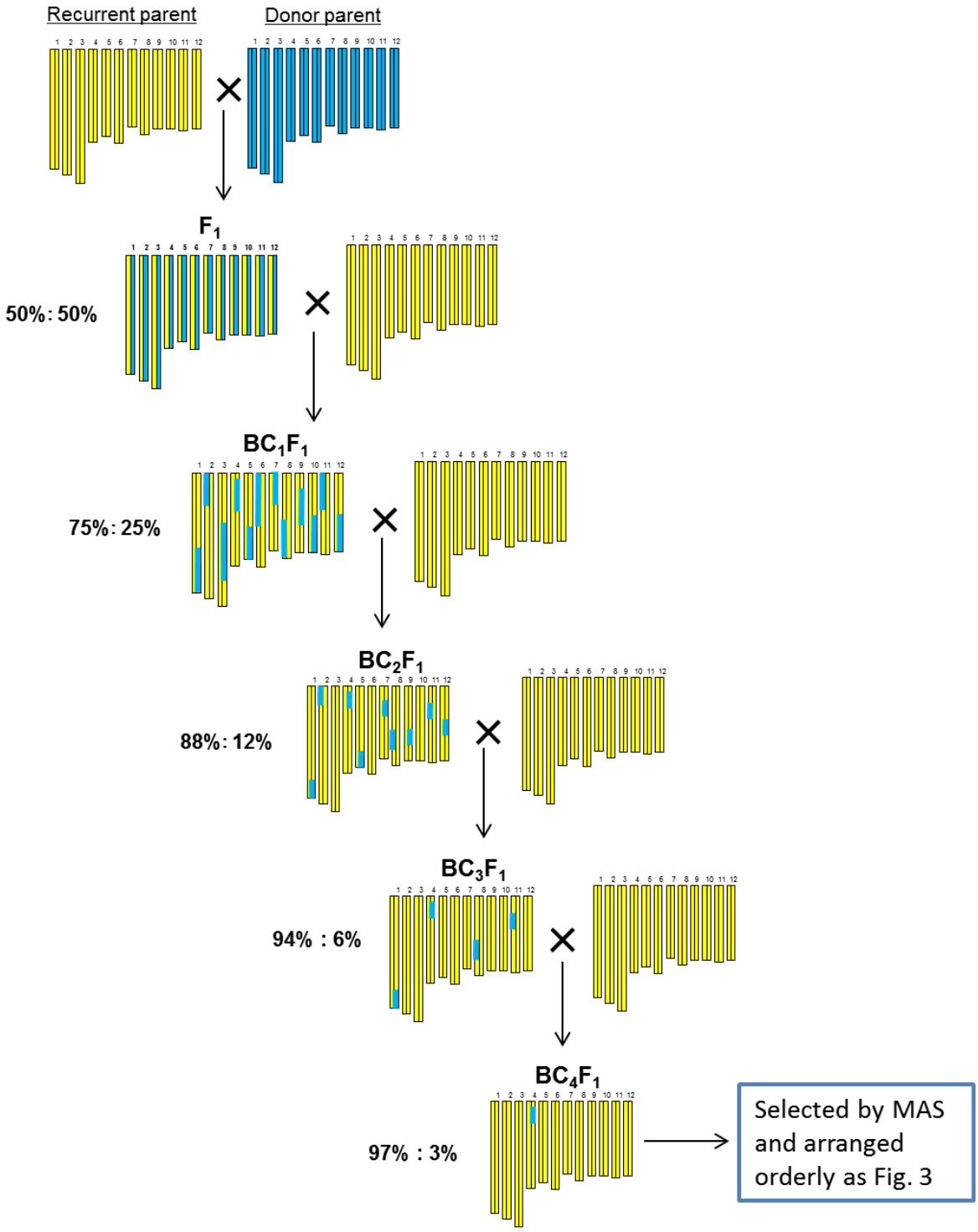


Figure 4. The schematic image of CSSL breeding. Yellow and blue chromosome segment represent recurrent and donor parent respectively. The percentage on the left side of the figure represents the ratio of the genomes of recurrent parent and donor parent. MAS means marker assisted selection.

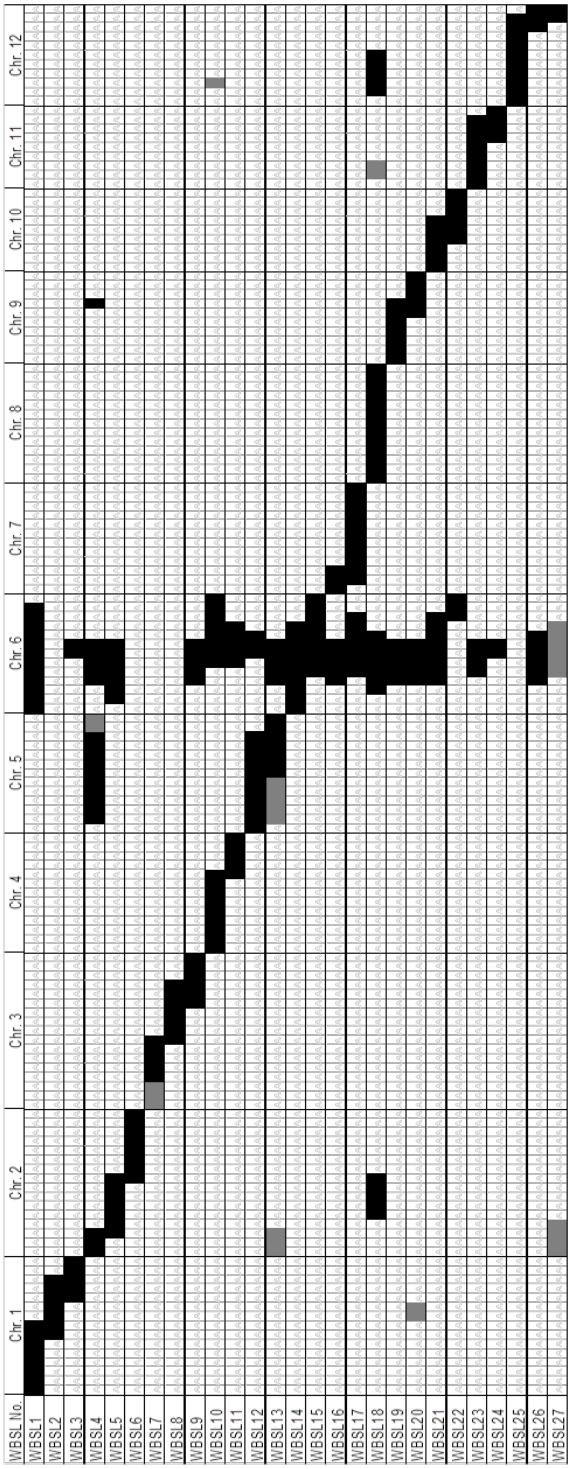


Figure 5. Graphical genotypes of WBSL. The set of WBSL consists of 27 lines harboring *O. nivara* chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

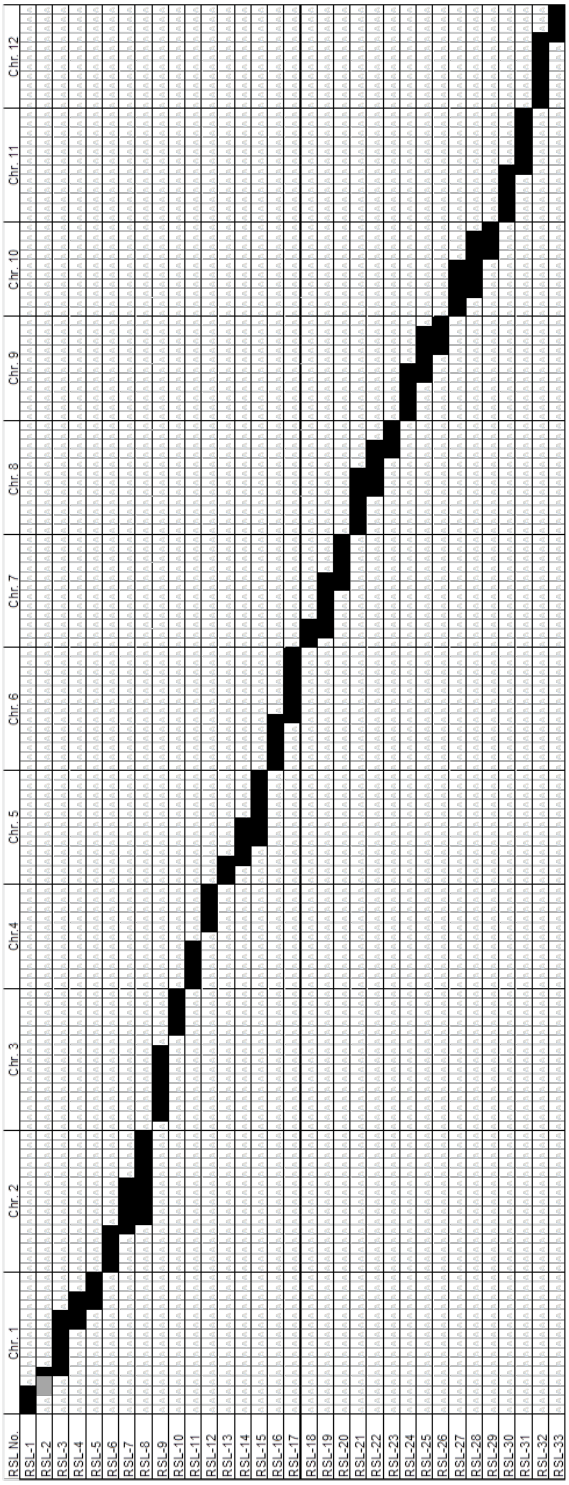


Figure 6. Graphical genotypes of RSL. The set of RSL consists of 33 lines harboring *O. rufipogon* chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

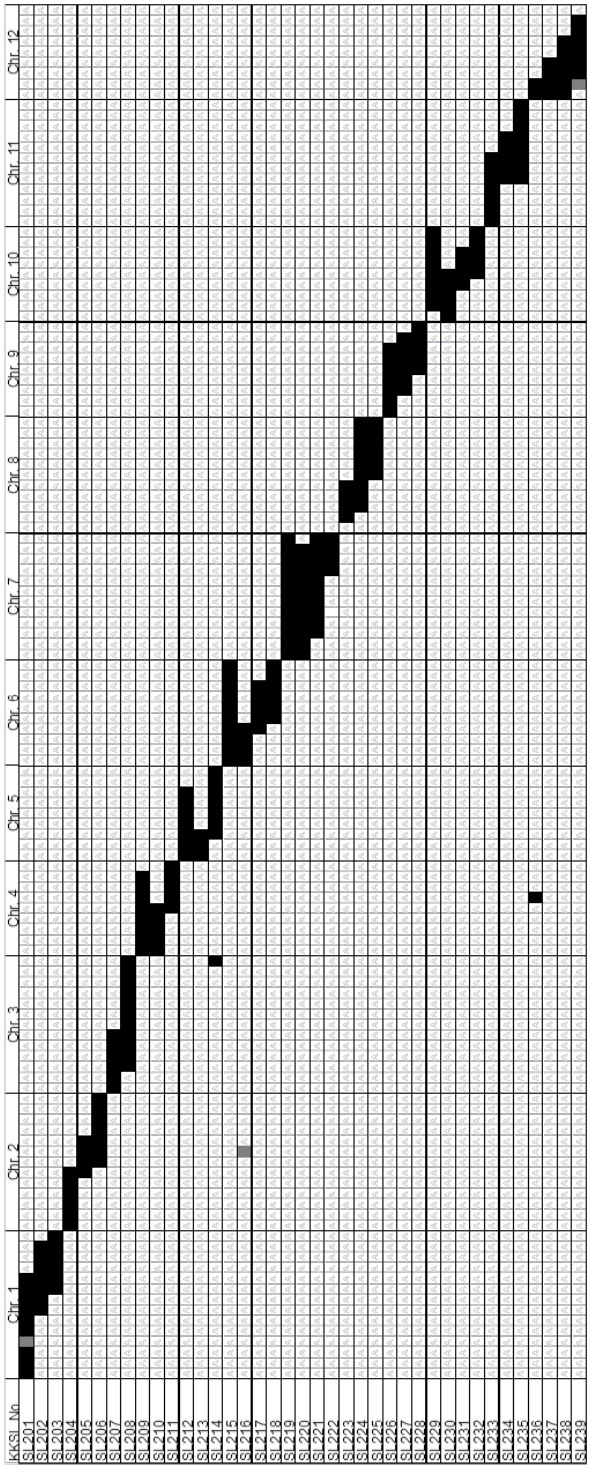


Figure 7. Graphical genotypes of KKSL. The set of KKSL consists of 39 lines harboring *O. sativa* ssp. *indica* var. Kasalath chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

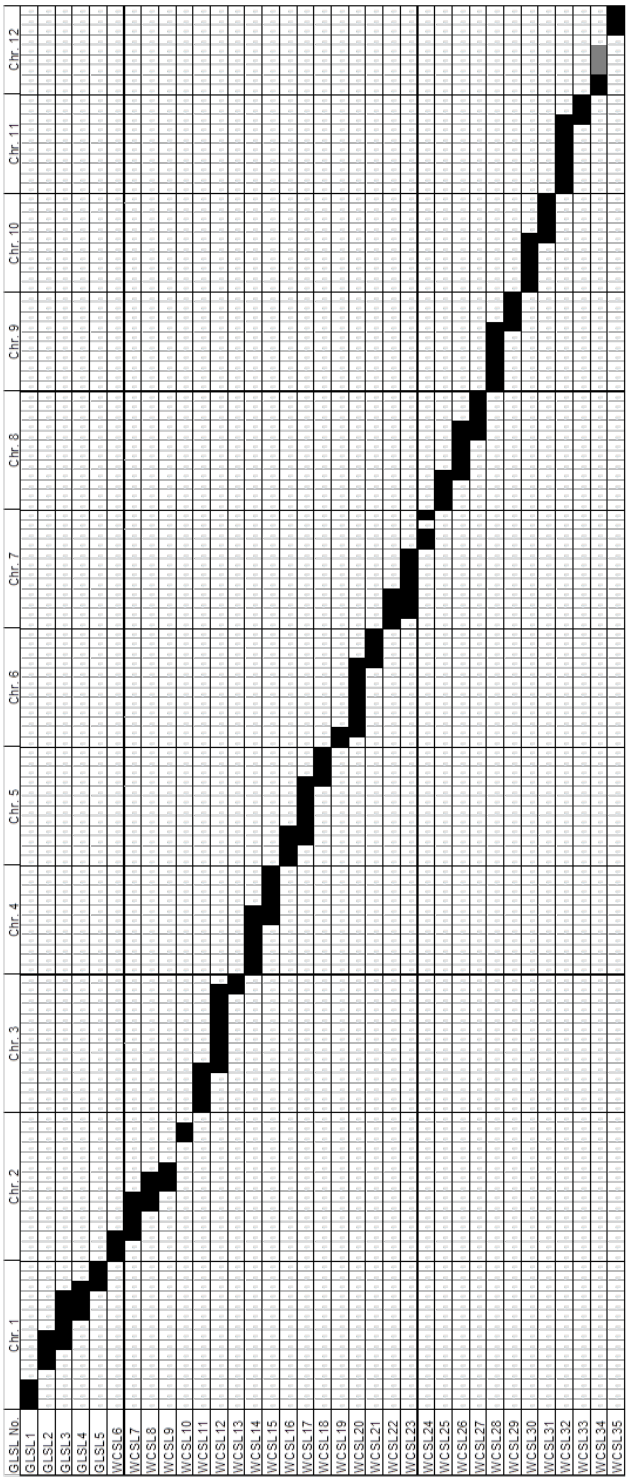


Figure 8. Graphical genotypes of GLSL. The set of GLSL consists of 35 lines harboring *O. glaberrima* chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

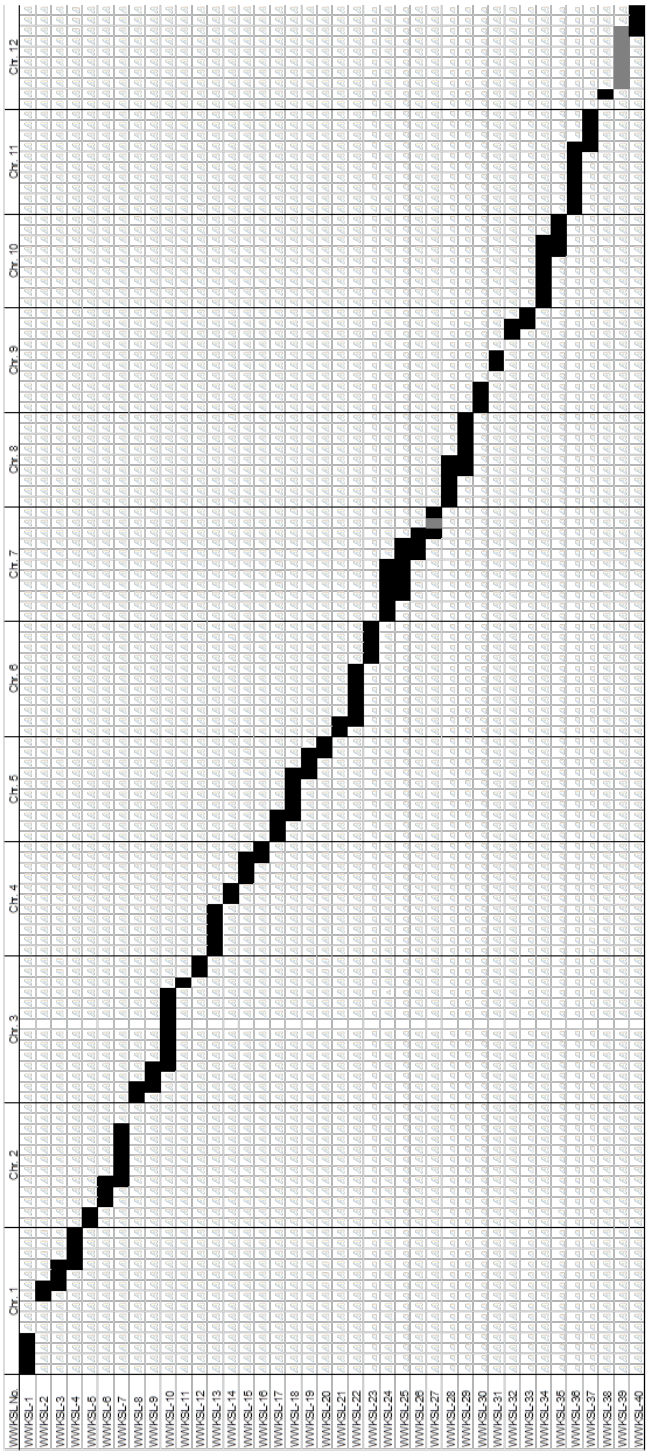


Figure 9. Graphical genotypes of WWKSL. The set of WWKSL consists of 40 lines harboring *O. barthii* chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

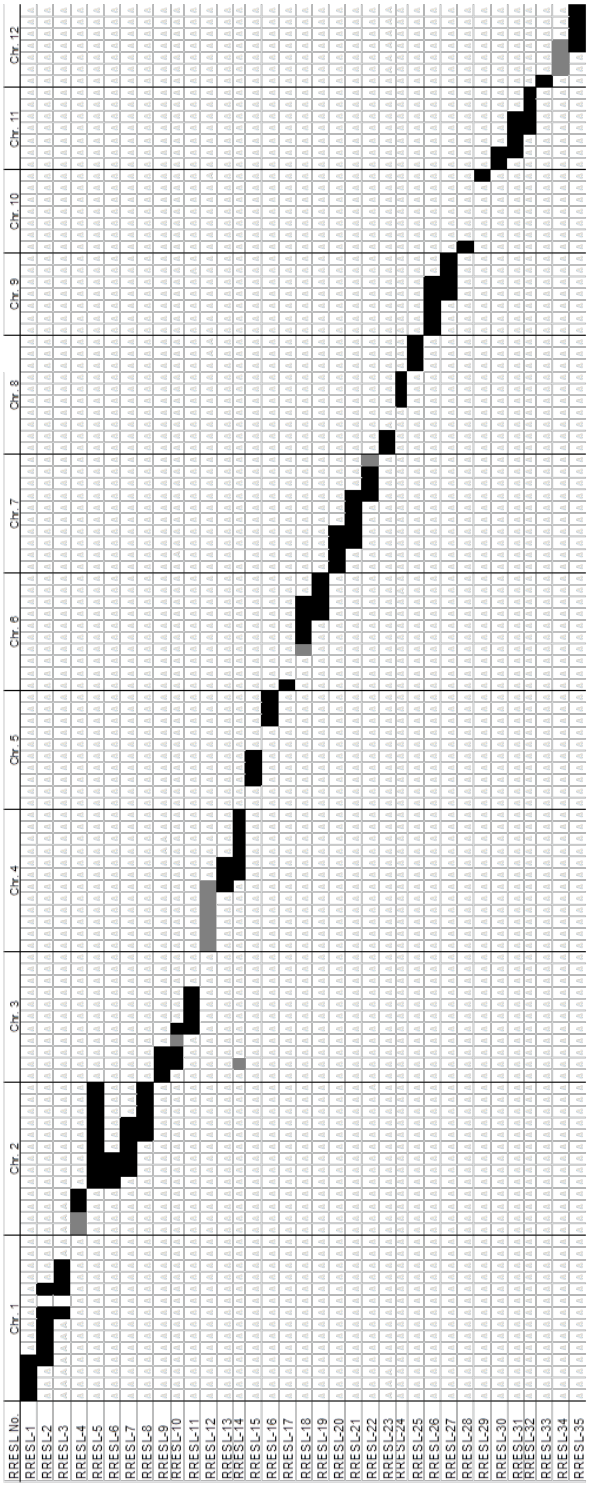


Figure 10. Graphical genotypes of RRESL. The set of RRESL consists of 35 lines harboring *O. glumaepatula* chromosome segments in the genetic background of *O. sativa* cv. Koshihikari. Black and white segments represent homozygous region each of the donor or recurrent parent, and gray segments represent the heterozygous region.

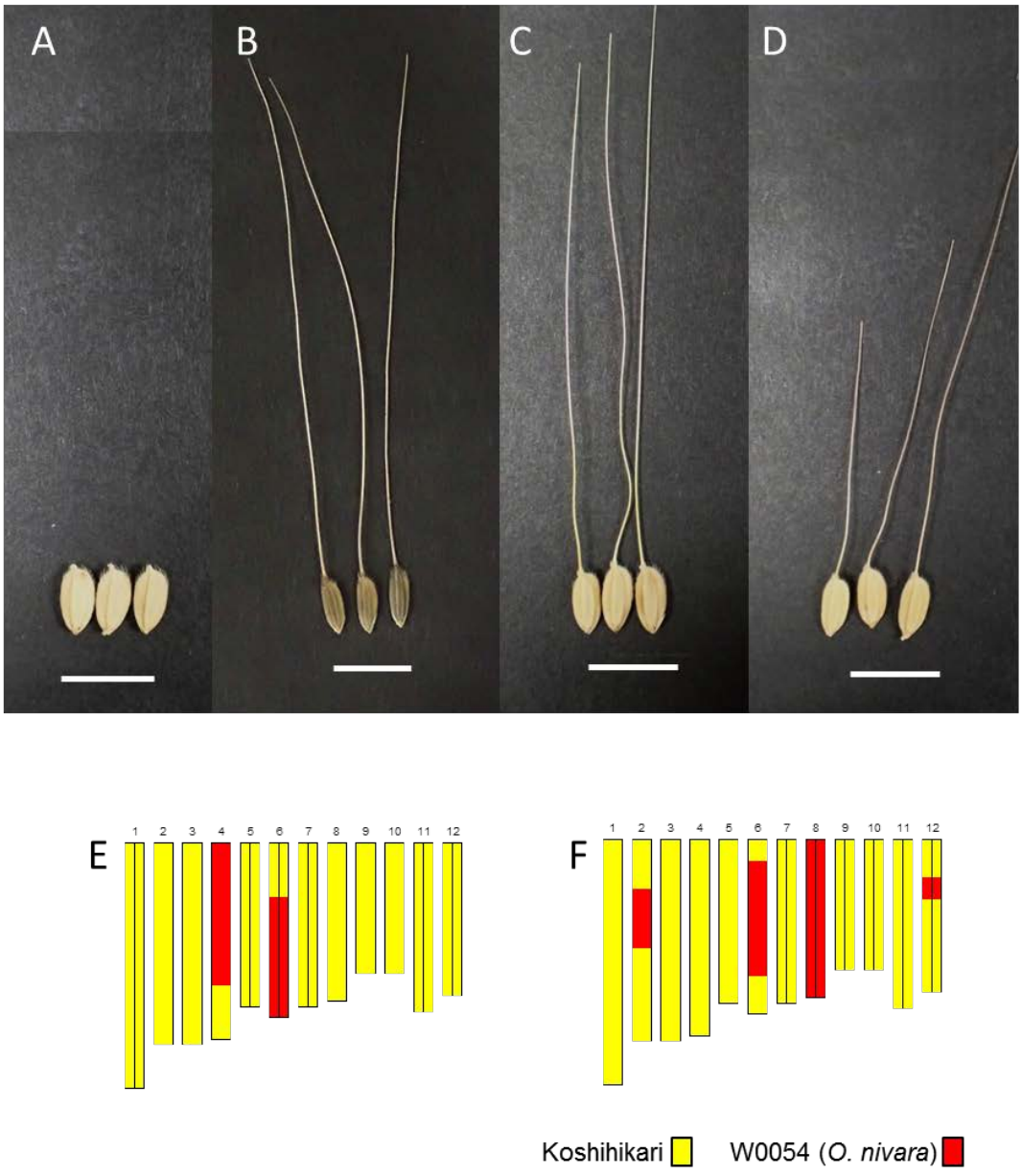


Figure 11. Two lines of WBSL have awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), W0054 (*O. nivara*) (B), WBSL10 (C) and WBSL18 (D). The below figures show graphical genotype of WBSL10 (E) and WBSL18 (F). Scale bars represent 1 cm.

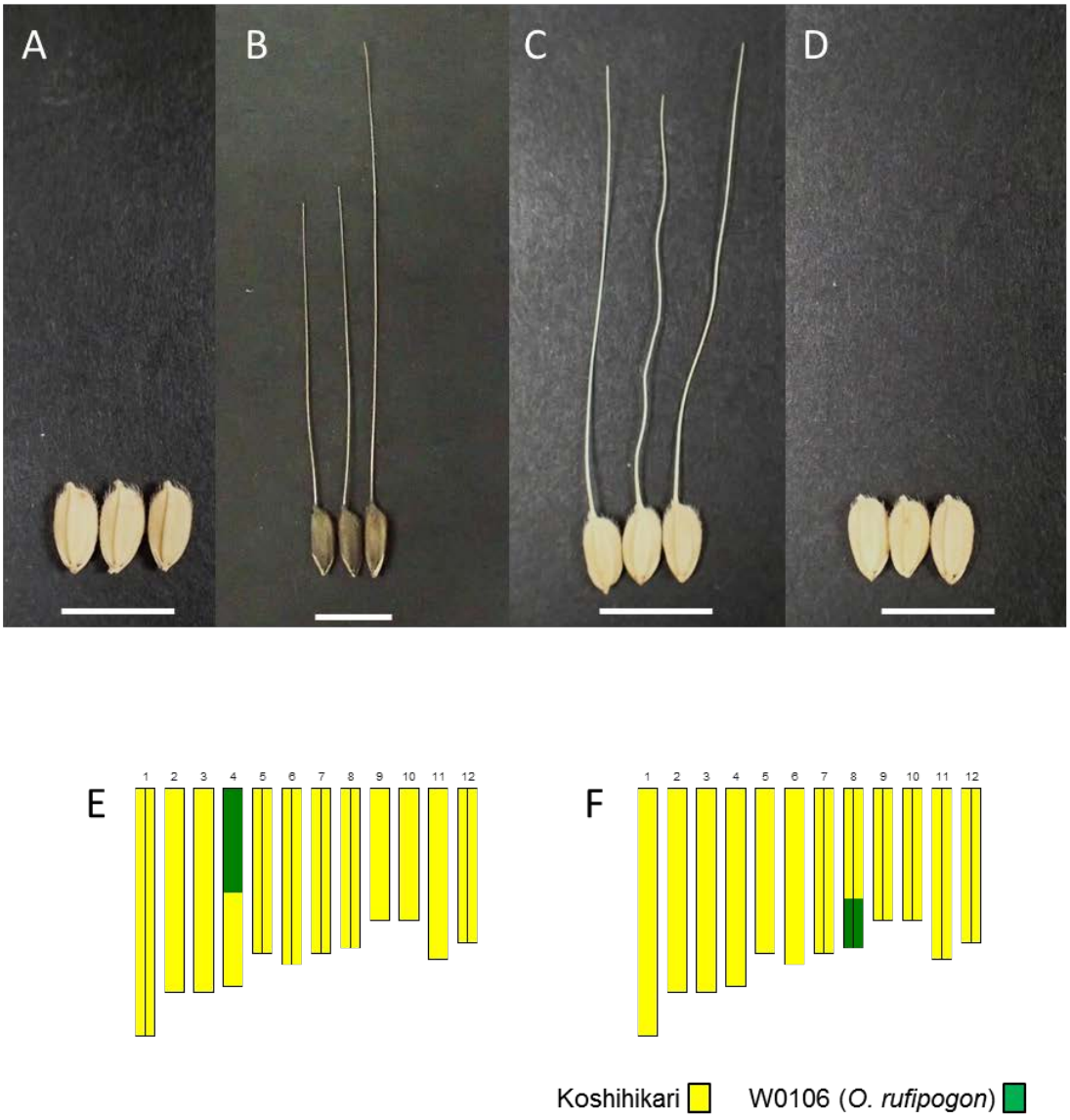


Figure 12. One line of RSL has awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), W0106 (*O. rufipogon*) (B), RSL11 (C) and RSL23 (D). The below figures show graphical genotype of RSL11 (E) and RSL23 (F). Scale bars represent 1 cm.

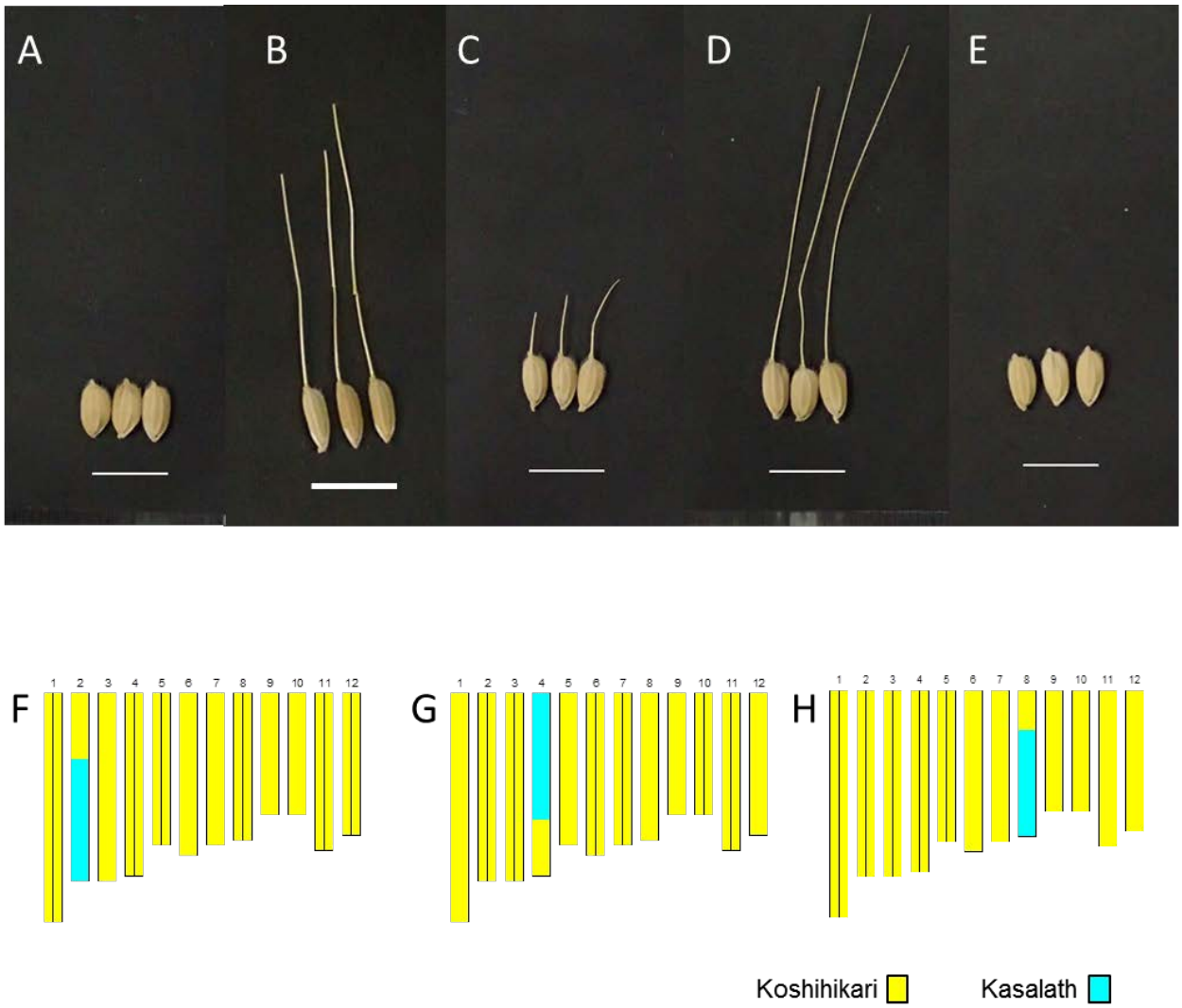


Figure 13. Two lines of KKSL have awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), *O. sativa* ssp. *indica* cv. Kasalath (B), KKSL206 (C), KKSL210 (D) and KKSL224 (E). The below figures show graphical genotype of KKSL206 (F), KKSL210 (G) and KKSL224 (H). Scale bars represent 1 cm.

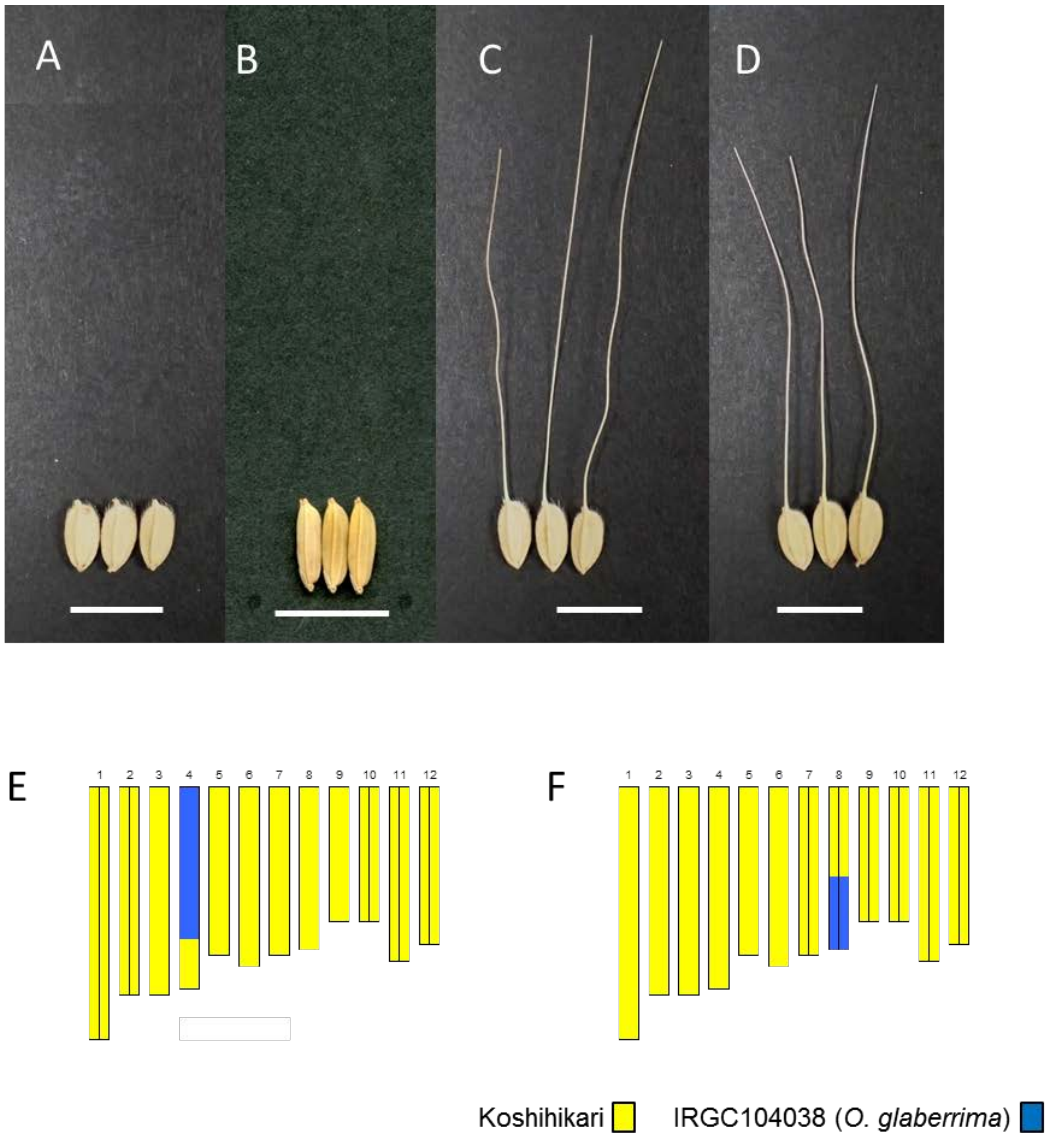


Figure 14. Two lines of GLSL have awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), IRGC104038 (*O. glaberrima*) (B), GLSL13 (C) and GLSL25 (D). The below figures show graphical genotype GLSL13 (E) and GLSL25 (F). Scale bars represent 1 cm.

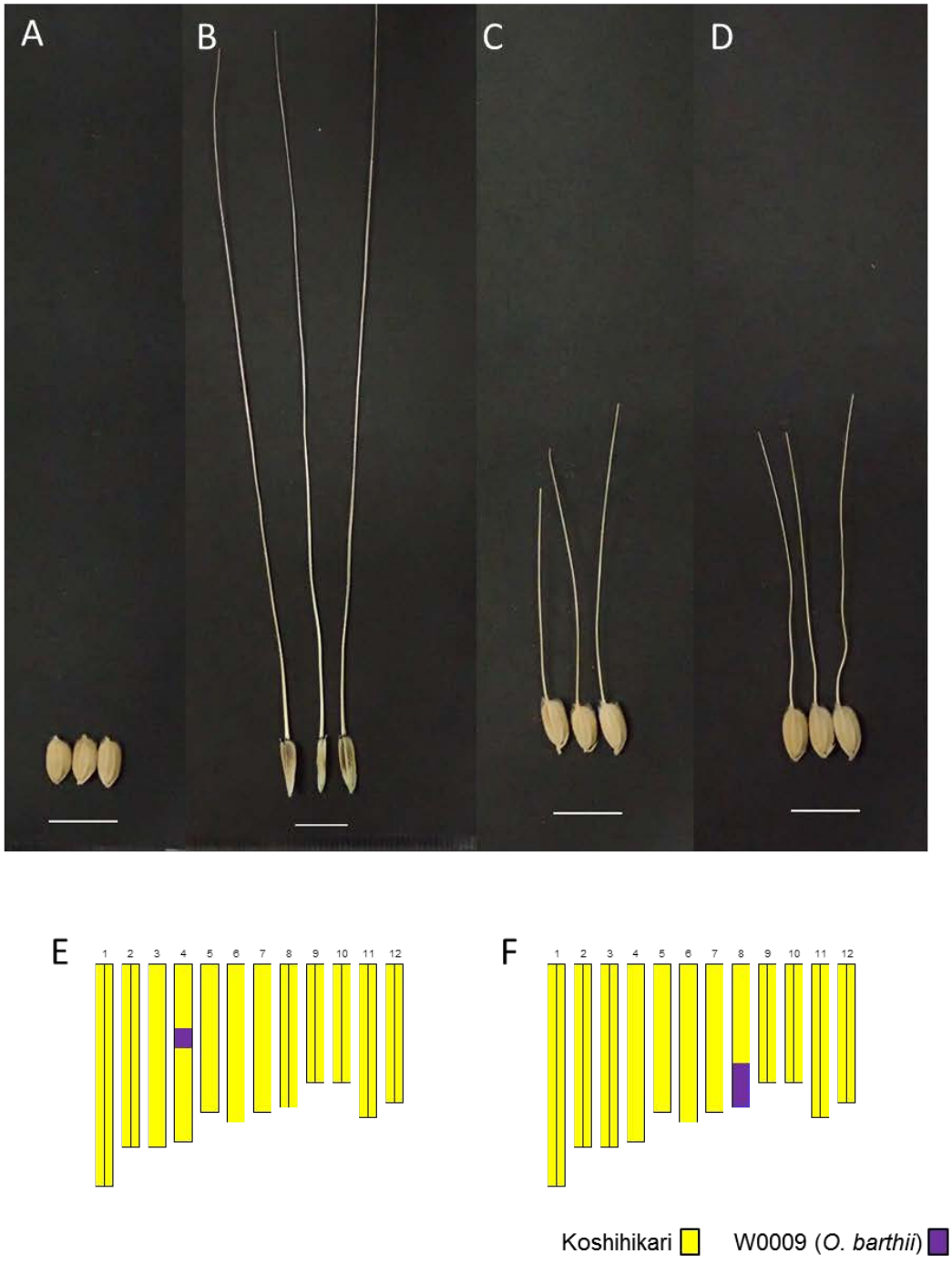


Figure 15. Two lines of WWKSL have awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), W0009 (*O. barthii*) (B), WWKSL14 (C) and WWKSL29 (D). The below figures show graphical genotype WWKSL14 (E) and WWKSL29 (F). Scale bars represent 1 cm.

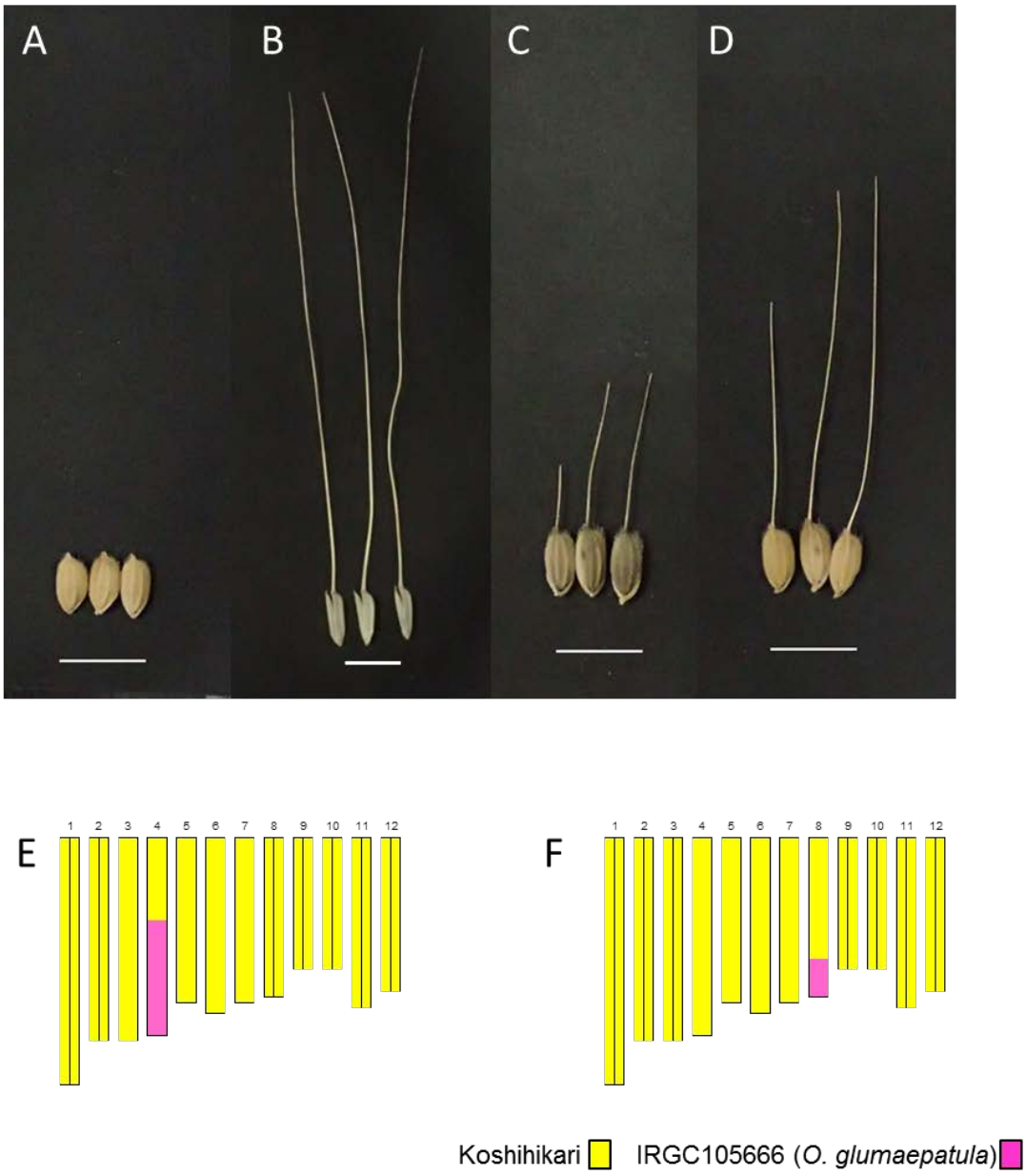


Figure 16. Two lines of RRESL have awn. Morphology of the seeds in *O. sativa* cv. Koshihikari (A), IRGC105666 (*O. glumaepatula*) (B), RRESL14 (C) and RRESL25 (D). The below figures show graphical genotype RRESL14 (E) and RRESL25 (F). Scale bars represent 1 cm.

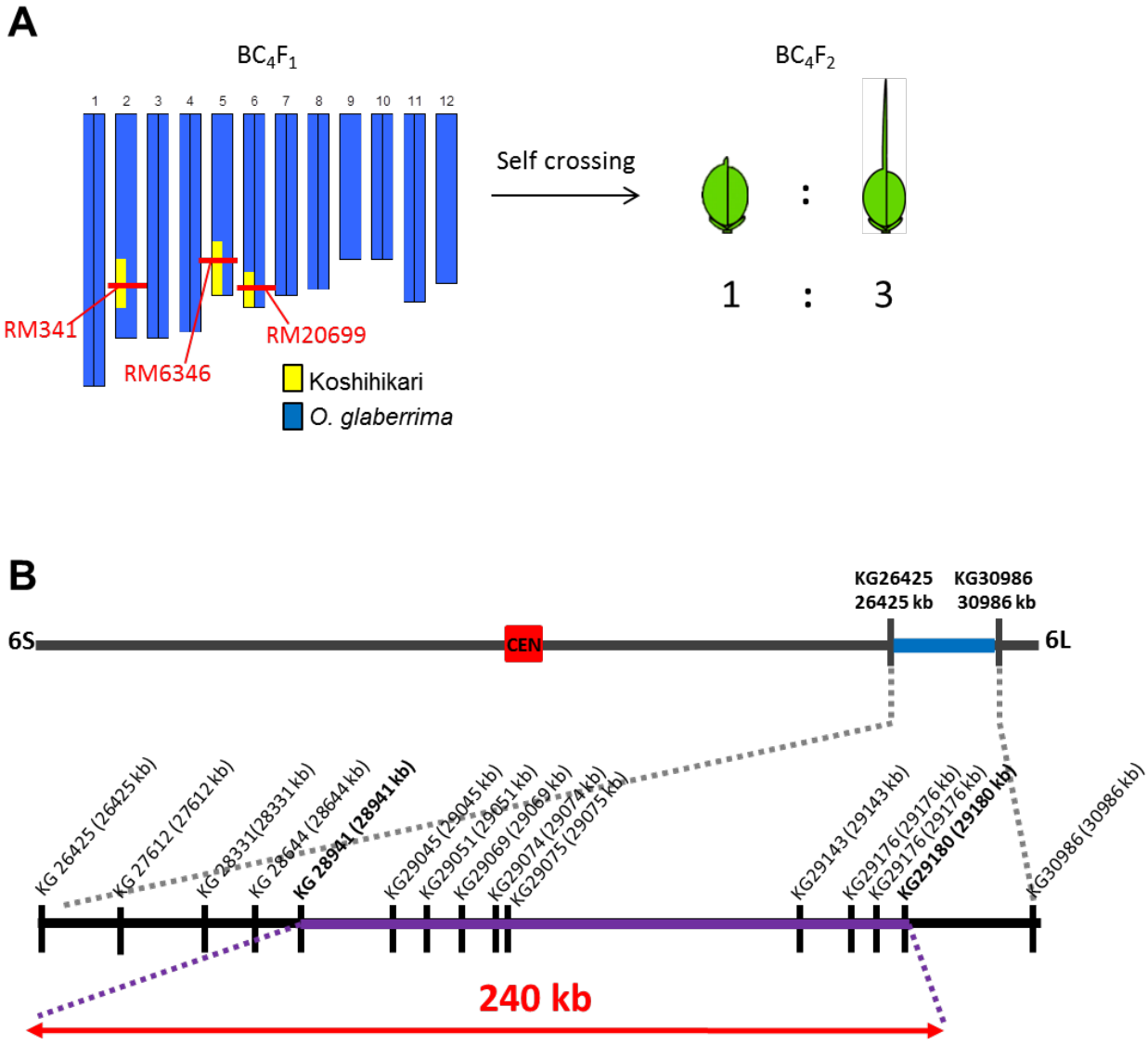


Figure 17. Fine mapping of *RAE3* on chromosome 6. (A) Graphical genotype of BC_4F_1 which have about 3.2% heterologous region in *O. glaberrima* genomic background. After self crossing of BC_4F_1, the awn phenotype segregate 1:3 in BC_4F_2. (B) *RAE3* mapping position in the approximately 240 Kb region flanked by KG28941 and KG29180 within the long arm of chromosome 6.

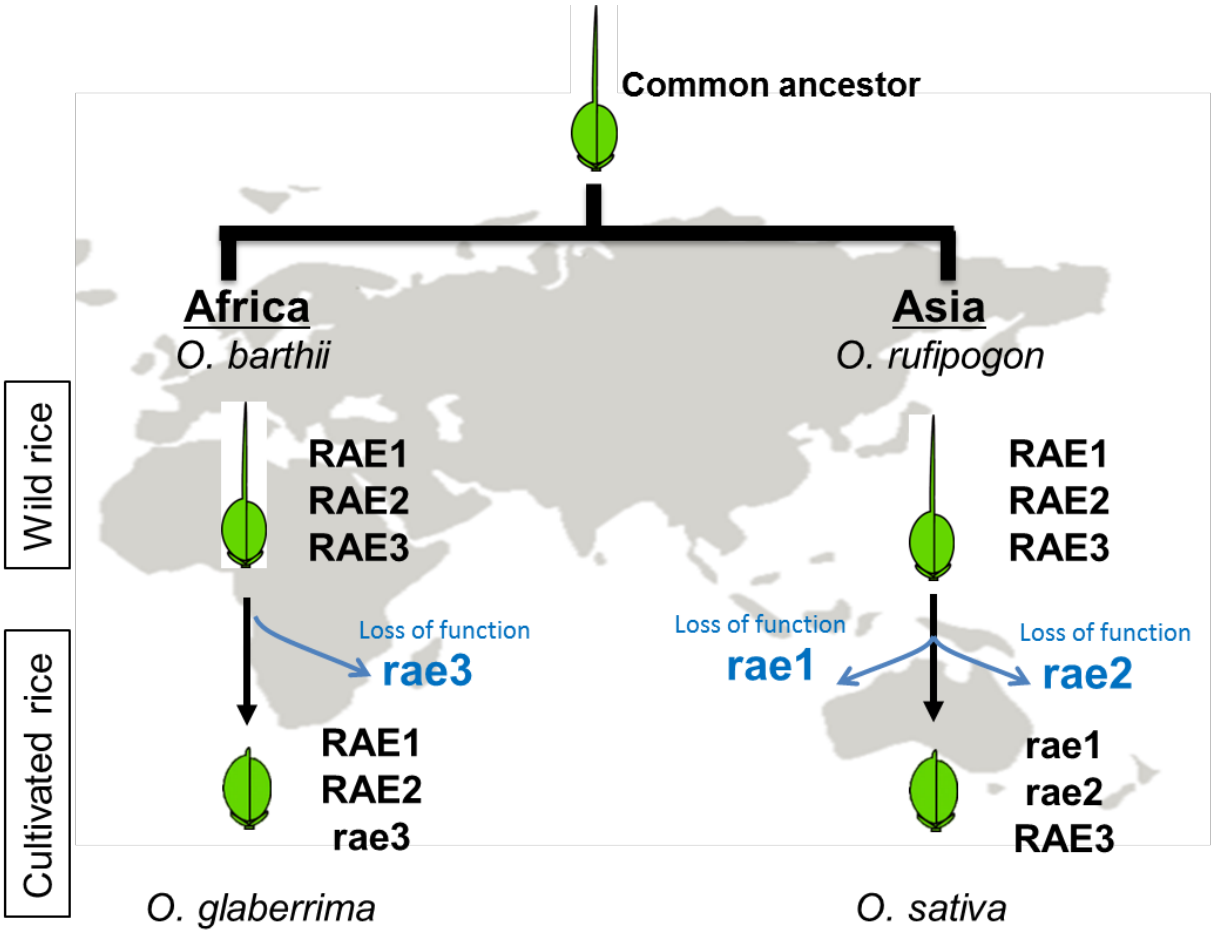


Figure 18. The selection pathway of awn responsible genes in Asia and Africa through the domestication.

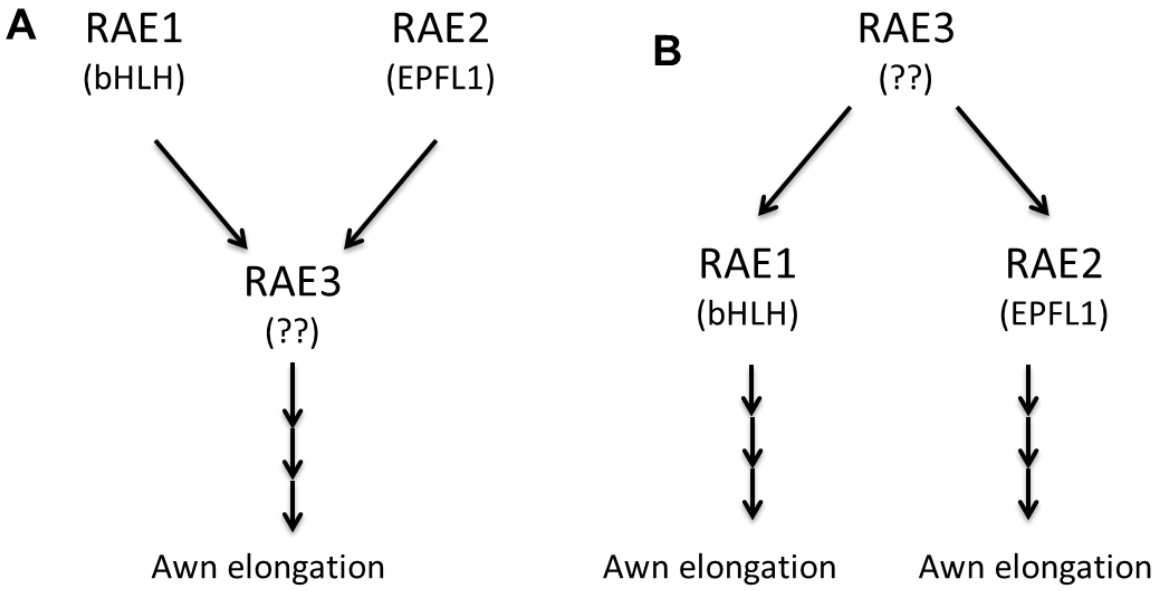


Figure 19. The model of mutual interaction among RAE1-RAE2-RAE3 for awn elongation. (A) This model shows RAE1 and RAE2 locate upstream of RAE3. (B) This model shows RAE1 and RAE2 locate downstream of RAE3.

Chapter 3

Identification of *Regulator of Awn Elongation 2* which is responsible for awn elongation.

Introduction

Awns of rice locate the tip of lemma of spikelet (Fig. 1A, B). There is a thick vascular bundle in the middle of awn and surrounded by parenchyma (Fig. 1C). Awn is thought as a modified leaf blade (Dahlgren et al. 1985). Many genes related to leaf development have been reported to regulate spikelet morphology such as *TONGARI-BOUSHI1* (*TOB1*) encoding *YABBY* gene (Tanaka et al. 2012), *OsETTIN2* encoding *ARF* gene (Toriba and Hirano 2014), and *KNOX* gene whose mutant showed hooded phenotype of spikelet in barley (Müller et al. 1995). However these gene's mutants affect not only on awn but whole spikelet morphology. Two genes named *An-1*and *LABA1* do not affect the whole spikelet morphology but regulate only awn length (Luo et al. 2013; Hua et al. 2015). In addition, I also found that two gene loci named *RAE1* and *RAE2* on chromosomes 4 and 8 respectively regulate only awn elongation but not other phenotype (Chapter 2). Furuta *et al.* (2015) identified *RAE1* gene by positional cloning using GLSL13 which possessed chromosome 4 segment of *O. glaberrima* in the genetic background of *O. sativa*. *RAE1* is the same gene previously reported *An-1* (Luo et al. 2013). This gene encodes bHLH transcription

factor and the expression level keeps low in *O. sativa* even the function conserved still (Furuta et al. 2015). On the other hand, *RAE2* gene has not been identified although QTL associated with awn elongation around the *RAE2* region on chromosome 8 have been reported (Sato et al. 1996; Cai and Morishima 2002b; Fawcett et al. 2013).

Here I show that *RAE2* gene encodes *EPIDERMAL PATTERNING FACTOR-LIKE 1* (*EPFL1*) and *O. glaberrima* allele can induces awn development. I also identified the Subtilisin-Like Protease 1 (SLP1) for processing the RAE2 peptide. This study also provides evidence that during the domestication of African rice and Asian rice, different genes were selected for the awnless phenotype.

Results

O. glaberrima has functional *RAE2*, but *O. sativa* has dysfunctional one

In six CSSL I examined in Chapter 2, several lines represent long awn phenotype that include functional RAE2 on chromosome 8. For an examination of domestication gene like RAE2, wild rice and cultivated rice are compared in usual strategy. However, RSL23 did not show awn phenotype and WBSL18 have some chromosomal region other than chromosome 8 (Chapter 2). Meanwhile, GLSL25 have only 11.5 Mb genomic region on chromosome 8 derived from *O. glaberrima* in the background of Koshihikari, and it showed long awn phenotype (Fig. 2A). It means that *RAE2* allele in *O. glaberrima* is functional. That is, to identify *RAE2* on chromosome 8 from *O. glaberrima* that induces awn elongation, I did the positional cloning by in a mapping population derived from introgression line, GLSL25 and Koshihikari.

Genetic linkage analysis using about ~8,000 F_2 individuals which derived from the cross between GLSL25 and Koshihikari delimited the candidate region into 80 kb which encompassed twelve predicted gene models in this region (Fig. 2B-D). I screened

an *O. glaberrima* BAC library and identified a clone Ogla0006B21 that covered the entire 80 kb candidate region from the BAC clone library using the SSR markers used for fine mapping of RAE2 (Fig. 3A). Five sub-clones, derived by partially digesting of Ogla0006B21, were systematically introduced into Taichung65 (T65), an awnless *O. sativa* ssp. *japonica* cultivar) (Fig. 3B). Two transgenic lines carrying the #33 sub-clone containing 29 kb and #89 containing 13kb fragment exhibited the awned phenotype (Fig. 3C-E). There is one ORF encoded *Os08g0485500* including in both sub-clones. To confirm this gene induces awn elongation or not, a construct carrying only the single gene *Os08g0485500* (*O. glaberrima* allele) was introduced into Nipponbare (awnless *O. sativa* ssp. *japonica*) (Fig. 4A). The resultant transgenic plant produced awns of comparable frequency and length to the awns of GLSL25 (Fig. 4B-D). We additionally transformed an RNA interference (RNAi) construct harboring a 3'UTR region of *Os08g0485500* CDS into GIL116 (Doi 1999), an awned introgression line carrying the chromosome 8 segment derived from IRGC104038 (*O. glaberrima*) in the T65 genetic background (Fig. 5A). Awn lengths of RNAi transgenic lines were significantly shorter than those of the controls, while awn frequencies were not different (Fig. 5B-D).

Together, these results indicate that *Os08g0485500* is *RAE2*, and that this gene acts to regulate awn elongation.

Coordination of *RAE2* expression with awn development

For observing the awn development, I compared the developmental stage of the spikelet in Koshihikari and GLSL25 by scanning electron microscopy (SEM). SEM observations showed that the lemma and palea morphology did not differ until the Sp7 stage between both (Fig. 6A, E). The awn primordium protruded at the distal end of the lemma of only GLSL25 from Sp8, and elongate until post Sp8 stage (Fig. 6B-D, 6F-H). Next I analyzed the expression pattern of *RAE2* at several organs in rice plant by qRT-PCR to determine its correlation with awn development. *RAE2* expression was about 10-fold higher in young spikelet of GLSL25 than in the other organs evaluated (leaf sheath, leaf blade, internode, and root) in contrast, *rae2* expression level in Koshihikari was low in all the organs (Fig. 7A). Moreover I performed RT-PCR using the materials classified the spikelet stage in detail. The result of RT-PCR showed that

the expression level of RAE2 was higher in younger stage of spikelet and gradually decreased depends on spikelet maturation (Fig. 7B). On the contrary to the result of Fig. 7A, the result did not represent the big difference of RAE2 expression in Fig. 7B in younger spikelet stage. Two possibilities was hypothesized; one is the difference of accuracy of the technique. qRT-PCR can detect the small difference quantitatively while RT-PCR was judged the data from the density of PCR band qualitatively. Second reason is that the material of qRT-PCR included much spikelet of Sp7 or Sp8 stage than RT-PCR one.

In situ hybridization showed similar expression patterns of both *RAE2* and *rae2* from Sp4 to Sp7 (Fig. 8A-C, 8F-H). *RAE2* transcripts, however, exhibited prolonged expression compared with *rae2* in the subsequent stages (Fig. 8D, I). When I observed the cross section of spikelet (Fig. 8K), GLSL25 showed especially high expression in the vascular bundles of the awn primordium but not in Koshihikari in post Sp8 stage (Fig. 8L-Q). Despite expression of *RAE2* in anthers, no obvious differences were observed for anther or pollen development in both Koshihikari and GLSL25 (data not shown). Together, these observations provide evidence of the importance of the

spatio-temporal regulation of *RAE2* expression in awn formation.

RAE2 encodes *EPIDERMAL PATTERNING FACTOR LIKE PROTEIN 1* in rice

Comparative sequence analysis of *RAE2* and *rae2* from *O. glaberrima* and *O. sativa* ssp. *japonica* respectively, revealed several SNPs and insertions in the promoter and coding region (Fig. 9A). It suggested the difference of expression pattern between *RAE2* and *rae2* derived from these SNPs in promoter region. A 2-bp insertion in the second exon of *rae2* results in a frame-shift mutation (Fig. 9B-D). Based on amino acid sequence analysis, *rae2* is predicted to encode an *EPIDERMAL PATTERNING FACTOR-LIKE1* (*EPFL1*) protein (Takata et al. 2013). This protein is a member of the EPF/EPFL family, a group of plant-specific secreted peptides that regulates a range of developmental processes (Hara et al. 2007; Hunt et al. 2009; Kondo et al. 2010; Sugano et al. 2010; Uchida et al. 2012). In *Arabidopsis thaliana* EPF/EPFLs are thus far the most well extensively studied such as Stomagen (also known as AtEPFL9) and its competitive factors; EPF1 and EPF2, while there are no reports about this family genes

in rice. Members of this peptide family share a conserved cysteine-rich region that mediates formation of disulfide bonds essential for functional conformation as a ligand (Marshall et al. 2011). Using cysteine-rich region sequences recognized mature peptide region, I evaluated the phylogenetic relationship of RAE2 with other members of the EPF/EPFL family from *Arabidopsis thaliana* and several grass species (*Brachypodium distachyon*, *Zea mayz*, *Sorghum bicolor*, *Triticum aestivum* and *Hordeum vulgare*) (Fig. 10A). Phylogenetic analysis revealed that *RAE2* is classified into the AtEPFL1-3 clade, a group of unknown function even in *A. thaliana*. Comparison of RAE2 and rae2 amino acid sequences with other EPFL relatives showed that all sequences except for rae2 contain six cysteine residues that are typical of EPFL peptides (Fig. 10B). Three-dimensional structure modeling clearly displayed structural similarity between RAE2 and Stomagen (Fig. 11). The loss of cysteine residues has been reported to cause dysfunctional activity of Stomagen due to the non-formation of the scaffold mediated by disulfide bonds (Ohki et al. 2011). Deletion of the two cysteine residues in the C-terminus of rae2 is therefore hypothesized to be the causal mutation for dysfunctional conformation.

Four types of RAE2 and their geographical distribution

Characterizing the diversity and frequency of nucleotide polymorphisms in domestication genes across divergent populations gives insight into the evolutionary history of rice. To understand *RAE2* variation across diverse accessions, I collaborated with Dr. Susan McCouch and Dr. Diane Wang in Cornell University, New York. We sequenced *RAE2* across a panel of 130 accessions made up of Asian (cultivated: n=42, wild: n=65) and African rice (cultivated: n=12, wild: n=11) (listed in Table 2). The result of sequence comparison among 130 accessions, we noticed that there are four types of RAE2 variants by classification based on cysteine number in mature peptide region (Fig. 12). As mentioned, the 2 bp insertion in the second exon of *rae2* was predicted to be the functional mutation in Koshihikari (Fig. 9B, D). Interestingly the insertion occurred in a highly variable GC-rich repeat region of second exon in our diversity panel (Fig. 13A). In addition to this, we found seven different-length polymorphisms in this region (Table 3). Among seven variants, four translated into functional RAE2 proteins (i.e. 6 cysteine residues), the other three variants are

putatively dysfunctional RAE2 proteins; either a truncated protein as in cv. Koshihikari (4 cysteine) or an extra-long protein (7 cysteine). We named these translated products according to the number of cysteine residues they harbor: 4C, 6C, and 7C, and grouped them into three protein-length classes (Fig. 12, Fig. 13B). That is, since the amino acid sequence of the mature peptide largely changed by frame shift after the insertion, it inferred that 4C or 7C RAE2 variants lose the function of RAE2. In addition, three singleton variants of RAE2 were identified that independently gave rise to translated products with 5 cysteine residues (5C), resulting from polymorphisms that occurred outside the GC-rich repeat region (Fig. 13A). All of these were predicted to give rise to medium length proteins but are compositionally divergent from the 6C medium peptide class (Fig. 13B). Based on these results, nucleotide variations in the GC-rich repeat region are responsible for functional cysteine-number variation in the RAE2 protein. The geographic distribution of the 130 diverse rice accessions reflects the fact that Asian rice varieties are widely planted around the world, while African rice is confined to Africa (Fig. 13C).

RAE2 loss of function has been selected in Asia but not in Africa

To understand the functionality of the RAE2 protein types (4C, 5C, 6C, 7C), we evaluated overexpression lines of each variant. For making 5C construction, I used NSFTV223 (listed in Table 2) genomic DNA in this experiments. Since our study suggested that conserved cysteine residues number is important for RAE2 conformation, we expect the other two alleles of 5C type RAE2 (derived from NSFTV673 and NSFTV762 in Table 2) lose those function. The result of transgenic plants, only the RAE2-6C type was able to induce awn elongation, while the others did not form awns, regardless of their *RAE2* expression levels (Fig. 14A, B). There are no notable traits other than awn elongation. Compared to the genomic RAE2 complementation test using own promoter (Fig. 4C, D), however, the percentage of awned seeds per panicle became lower and awn length shorter (Fig. 14C, D). This result might represent the secondary effects of overexpression analysis.

Moreover, two CSSL in cv. Koshihikari background from RAE2-6C and rae2-7C donor showed awned and awnless phenotype respectively (Fig. 15A, B). We evaluated the RAE2 variants found in two Asian wild rice; *O. nivara* (Acc. W0054) and *O.*

rufipogon (Acc. W0106) which were used as donor parents to develop two CSSL populations (WBSL and RSL, respectively) in the background of cv. Koshihikari (in Chapter 2). Phenotyping the two lines harboring the *RAE2* locus on chromosome 8 derived from each donor parent demonstrated that the WBSL18 derived from W0054 formed awns, but the RSL23 derived from W0106 did not (Fig. 15B). Consistent with our results from the comparative analysis of RAE2 alleles, W0054 had 6C RAE2 allele, while W0106 possessed dysfunctional 7C allele (Fig. 15C). These data supported the hypothesis that the number of cysteine residues in RAE2 directly affects awn development.

Interestingly, all African rice was found to carry the RAE2-6C type, despite the fact that cultivated African rice, *O. glaberrima* does not possess awns (Fig. 16A(i, ii), 16B(i, ii)). These results support previous our report that a different gene *RAE3*, was responsible for the awnless phenotype of *O. glaberrima* in Africa (Furuta et al. 2015). In Asia, dysfunctional RAE2 protein types were present in 32% of *O. rufipogon* (including *O. nivara*) wild accessions (Fig. 16A(iii)), while almost all individuals possessed awns (Fig. 16B(iii)). This distinguished *RAE2* from previously reported awn

domestication genes, *An-1* and/or *LABA1*, which were documented to persist as functional alleles in the vast majority of wild Asian rice populations. Although *O. sativa*, on the other hand, was nearly fixed (93%) for dysfunctional alleles at *RAE2* and most of them lose awn (Fig. 16A(iv), B (iv)), with significantly different frequencies of *RAE2* protein classes observed among the five subpopulations (Table 4).

RAE2 has been artificially selected in Asia through the domestication

To test for evidence of a selective sweep in the region of *RAE2*, nucleotide diversity (π) of *O. sativa* (n=67) relative to nucleotide diversity (π) of *O. rufipogon* (n=65) was estimated using data in 100-SNP sliding windows across chromosome 8 (listed in Table 5). A drastic decrease was observed in the ratio π *O. sativa*/ π *O. rufipogon* across a 1.5 Mb region flanking *RAE2* (Fig. 17A), consistent with a selective sweep in *O. sativa*. A second decrease in the ratio was observed 0.5 Mb downstream of *RAE2*, suggesting the possibility of a second target of selection nearby. Further analysis in this region including the drop around RAE2 and the next drop of it could reveal the

domestication block existence.

To investigate the relationships between wild and cultivated Asian rice we analyzed the genetic distance (*d*) between *O. sativa* and *O. rufipogon/ O. nivara* using the same 100-SNP windows. Genetic distance is a measure of the genetic divergence between species or populations. If the species with many similar alleles they showed small genetic distances, thus, they are closely related and have a recent common ancestor. The result of *d* across chromosome 8 revealed that the dysfunctional RAE2 types (4C, 7C) observed in wild accessions were likely the result of recent, back-introgression from *O. sativa* to *O. rufipogon* or *O. nivara.* This was supported by a decrease in *d* across the 1.5 Mb region surrounding *RAE2* in the dysfunctional class (4C, 7C) relative to *d* in all RAE2 types (Fig. 17B). This result is consistent with recent gene flow back from cultivated to wild Asian rice.

RAE2 is cleaved specifically in the spikelet

EPFL family peptides typically require post-translational cleavage from

pro-peptide to become a mature peptide (Fig. 18) (Wheeler and Irving 2011; Katsir et al. 2011). To test whether or not RAE2 is cleaved, we generated transgenic plants carrying *pACT*::*RAE2-3xFLAG*. Immunoblot analysis demonstrated that RAE2 was cleaved into a ~11 kD peptide only in the spikelet but not in the other organs (Fig. 19A, B). This size is consistent to the predicted mature RAE2 peptide (Fig. 19C), so I conclude that RAE2 become pro-peptide in all organs but genuine cleavage to mature peptide occurred only in spikelet specifically. This spikelet-specific cleavage of RAE2 was confirmed by mixing the recombinant RAE2 pro-peptide (RAE2-pro) with each of the protein extracts derived from various organs (Callus, Stem, Leaf and Spikelet) (Fig. 19D). In addition it is proved that the ~11 kD band contained the C-terminal region of RAE2 by using anti-RAE2 antibody made at the end of RAE2 C-terminus (Fig. 19E). Furthermore the cleavage of RAE2-pro was inhibited by protease inhibitor cocktail, Complete (Fig. 19F), these data suggest that RAE2-targeted protease(s) specifically functions in rice spikelets.

Identification of RAE2-targeted protease

To find the candidate protease(s), I compared the expression patterns of 63 members of the subtilisin-like protease family (listed in Table 6) which has been reported to cleave EPF/EPFL peptides in *A. thaliana* (*SDD1* (Groll et al. 2002), *CRSP* (Engineer et al. 2014)). Among the 63 proteases examined, *Os01g0702300*, named *SUBTILISIN-LIKE PROTEASE 1* (*SLP1*) that is specifically expressed in young inflorescence, lemma and palea according to Rice-Xpro (http://ricexpro.dna.affrc.go.jp/) was identified (Fig. 20A). Semi-quantitative PCR analysis of spikelet extract confirmed *SLP1* expression pattern experimentally (Fig. 20B). Moreover SLP1 was detected in spikelet extracts from Koshihikari by MALDI-MS/MS with iTRAQ (Fig. 20C). These results showed that this protease is the strong candidate for cleaving RAE2 in spikelet specifically. Comparison of *SLP1* sequences of *O. sativa* and *O. glaberrima* showed some SNPs in coding region, however, no deleterious change or frame shift could not be observed (Fig. 21A, B). Analysis of phylogenetic relationships classified this gene in the same clade as CRSP (Tripathi and Sowdhamini 2006, Fig. 22), supporting the hypothesis that *SLP1* is the most promising candidate gene targeting RAE2 peptide.

To determine whether the RAE2 could be cleaved by SLP1, I performed *in vitro*

processing assay using a synthetic RAE2 short peptide (synRAE2) spanning the predicted cleavage site (52-AGEEEKVRLGSSPPSCYSK-70) according to Stomagen (Sugano et al. 2010) and EPF2 (Engineer et al. 2014). With this peptide I made *in vitro*-synthesized SLP1 protein by wheat germ cell for *in vitro* processing assay. As a result, three cleaved peptide were detected (Fig. 23A). Cleavage of synRAE2 between amino acid positions G53 and E54, P65 and S66 or both sites by SLP1. This suggested that SLP1 cleaves 2 positions of this peptide, however according to the Stomagen and EPF2 reports the site between P65 and S66 is more appropriate as cleavage site. A series of mutated RAE2-pro (muRAE2 #1-#4) by *E. coli* were used for *in vitro* processing assay with *in vitro*-synthesized SLP1 to verify this cleavage site. muRAE2 #1 and #4, possessing 6 or 3 alanine-substituted amino acids adjacent to the predicted cleavage site, was not cleaved by SLP1 (Fig. 23B). In contrast, muRAE2 #2 and #3, each containing 3 alanine-substituted amino acids located away from predicted cleavage site, were cleaved (Fig. 23B). This experiment revealed that the alanine substitution closer to the cleavage point repressed RAE2 cleavage.

Detecting the cleavage point of RAE2

I also tested whether SLP1 digested the synthetic small peptides with a single amino acid substitution near the predicted cleavage site or not. SLP1 was unable to cleave synthetic peptides with mutations around the amino acid positions P65 and S66 (Fig. 24A). These results supported the alanine substitution experiments with muRAE2. That is, SLP1 cleaves RAE2 between the amino acid positions P65 and S66 in the spikelet specifically (Fig. 24B). Furthermore I made the RAE2 alanine substitution construct and transduced into Nipponbare (Fig. 25A). Five of 7 transgenic lines did not elongate awn at all while 2 lines showed short awn and low frequency of awned seeds in a panicle (Fig. 25B). The reason why this variation of the result occurred was hypothesized that the small existence of RAE2 mature peptide promoted slightly in 2 transgenic lines. To prove this, I have to check the transgene sequence and whether the peptide is cleaved or not in these 2 lines. The transgenic lines which containing RAE2 overexpression construct without pro-peptide region represented awn elongation (Fig. 25C, D). This result suggested that the 60 amino acids in mature peptide are sufficient for RAE2 function.

Discussion

Downstream pathway of RAE2 signal peptide

Our study identifies *RAE2* as *Os08g0485500,* a novel EPFL gene that is preferentially expressed during early panicle development and is cleaved by spikelet-localized protease SLP1 to mediate awn elongation. It is known that EPFL peptide signaling is transmitted via receptors such as ERECTA family receptor kinase (Lee et al. 2012, 2015). RAE2 might also work with this type of receptor in rice. Analyses of *RAE2* transgenic lines demonstrate that *O. sativa* ssp. *japonica* should carry a functional RAE2-targeted protease SLP1 and receptor(s) while *japonica* has the dysfunctional rae2 (4C type). The fact that the pathway is conserved, despite the dysfunctionality of the rae2 signal peptide in *O. sativa* ssp. *japonica*, suggests that other EPF/EPFL peptides work through the same components likely to contribute to other morphological characteristics. Identification of the receptor(s) of RAE2 might reveal the relationship among the EPF/EPFL peptides in rice and contribute to understand downstream factors affecting awn elongation (Fig. 26).

Artificial selection in *RAE2* was occurred specifically in Asia

The sequence comparison of Asian and African rice suggested that common rice ancestor had a set of genes including *RAE2*, however during the speciation, Asian rice and African rice took the different strategy for awnlessness. It was inferred that there is selective pressure to conserve RAE2 functionality in wild Asian rice, as its reading frame is preserved in most individuals despite nucleotide-level variation. Because other awn regulating genes (e.g. *RAE1/An-1* or *LABA1*) can mask *RAE2* loss-of-function, this raises the possibility that *RAE2* is conserved in the wild due to pleiotropic effects on other fitness phenotypes. With its hyper variable GC-rich region giving rise to frame shift and concomitantly occurred cysteine number diversity contributing to variation in awn formation. On the other hand, African rice kept functional *RAE2*, the cultivated African rice might acquire the mutation(s) in *RAE3*, a locus on chromosome 6 (Chapter 2; Furuta et al. 2015). Considering together, I made the model of *RAE2* selection and rice speciation in Africa and Asia (Fig. 27). In Africa, *RAE2* has not been selected for awn elongation. It inferred that *RAE2* has pleiotropic effects to survive in this region

due to on other fitness phenotypes as mentioned above. In Asia, as *O. sativa* ssp. *japonica* and *indica* have different variants of RAE2 respectively, we can estimate the mutation in GC-rich region have been occurred before or at a time of separation into two subspecies. Especially *indica* keeps single RAE2 protein length (196 AA) having 7C despite *japonica* has all types of RAE2 (4C, 5C, 6C and 7C), it is suggested that *indica* speciation happened only one time however *japonica* has been experienced many times. This difference in Africa and Asia can be considered because of the time length for the domestication of respective site. Further, there would be the speculation that Asian rice might have been introduced in the American continent by crossing with local species because there are only 4C or 7C types of RAE2. The story of *RAE2* is part of a larger narrative about human selection. As a trait that has been targeted for selection multiple times via multiple genes, the awn serves as a unique lens through which to study the domestication history of rice. Further archaeological study and the molecular network among *RAE1*, *RAE2* and *RAE3* might reveal the detail of awn missing pathway in Asia and Africa.

Other factors related to awn elongation

Awn elongation was triggered by low temperature or pest damage (Takahashi et al. 1986). Since awns have the role to disperse the seeds broadly, awn elongation might contribute as an escape strategy under such fluctuating environmental conditions to ensure their future propagation. The family member of EPF/EPFL peptide is known working as signaling molecule depends on environmental change. For example, EPF2 is cleaved more under high CO_2 condition by increasing CRSP expression level (Engineer et al. 2014). According to this, I speculated that RAE2 worked as the signaling molecule under lower temperature or when plant has damage from the insects.

Both molecular and genetic evidence demonstrate that RAE2 positively regulates awn elongation. Natural variation in *RAE2* plays crucial roles in the domestication and evolution of rice morphology. Identifying the receptor for RAE2 and investigating the relationship among *An-1*/*RAE1*, *LABA1*, *RAE2*, and *RAE3* will be important to further understand the molecular basis of the regulation of awn development in rice.

Materials and Methods

Fine mapping of *RAE2*

8,000 F2 plants derived from the cross between *O. sativa* ssp. *japonica* cv. Koshihikari and GLSL25, a chromosome segment substitution line (CSSL) carrying approximately 11.5 Mb of *O. glaberrima* Acc. The segregants were genotyped with respect to the DNA markers 8KG23935(distal) and 8KG24032(proximal) (showed in Table 1). Progeny in which a recombination had occurred between 8KG23941 and 8KG24021 were genotyped with de novo developed DNA markers to define the recombination site.

RAE2 construct using *O. glaberrima* BAC clone

The BAC clone Ogla0006B21 harboring the entire 80 kb candidate region was screened from the CG14 BAC library by the flanking markers used in the fine mapping of *RAE2*. Sequencing of the BAC clones was performed by shotgun sequencing using illumina HiSeq 2000. The BAC sequence was assembled using the GENETYX software package (GENETYX Co., Tokyo, Japan) and Clustalw ver. 2.0 with the default settings (Larkin et al., 2007). The annotated genes were compared with gene annotations in RAP-DB

(http://rapdb.dna.affrc.go.jp/viewer/gbrowse/build4). The BAC clone was partially digested with *Sau*3AI, yielding 10-30 kb fragments which were then sub-cloned into the binary vector TAC7. Five sub-clones that cover the entire 80 kb candidate region were selected and used for the complementation tests. Sub-clones #33 and #89 harbor the entire RAE2 gene.

RAE2 construct for complementation test, overexpression and RNAi silencing

The full gDNA sequence of RAE2 including 3 kb upstream of the start codon and 1 kb downstream of the stop codon of the gene was cloned into pENTR/D-TOPO (Invitrogen) and transferred into pGWB501 (Nakagawa et al. 2007) through Gateway cloning technology (Invitrogen) to develop the transformation construct *pRAE2:RAE2*.

For RAE2 overexpression and RNAi silencing, DNA fragments were PCR-amplified from cDNA and cloned into pCAMBIA1380 and pANDA vector (Miki et al. 2004, 2005) respectively. For the RNAi test, the awned CSSL line GIL116 carrying a chromosome 8 fragment of IRGC104038 in T65 background was used to suppress RAE2 expression. The constructs were used for the standard protocol of Agrobacterium

(strain EHA105)-mediated rice transformation. Transgenic lines were selected on Murashige and Skoog medium plates containing 50 mg hygromycin (Sigma).

Growth conditions

The non-transgenic plant materials were grown in the research field of Nagoya University, Japan under natural day length and temperature along with the conservational Japanese agriculture calendar.

For RAE2 complementation test, the awnless *O. sativa* ssp. *japonica* cv. Nipponbare and Taichung65 (T65) were used for transformation. For the RNAi test, the awned CSSL line GIL116 carrying a chromosome 8 fragment of IRGC104038 in T65 background was used to suppress RAE2 expression. Since Koshihikari have low regeneration ability in the mature seed culture system (Nishimura et al. 2005), Nipponbare and T65 were used instead. The transgenic plants were grown in isolated greenhouses under long day condition until the ten-leaf stage, and transferred to short day condition until flowering.

Phenotypic evaluation

Panicles of the parental plants (Koshihikari and each CSSL) and transgenic plants (BAC sub-clones and RAE2 genomic fragment complementation lines, RNAi lines and overexpression lines) were harvested after seed maturation. Panicles were sampled from 10 plants to measure awn length and frequency of awned seeds per panicle.

Scanning electron microscopy

The young panicles of Koshihikari and GLSL25 were fixed in starch-based glue for microscopic observation. The samples were viewed using the SEM (S-3000N, Hitachi, Tokyo, Japan) scanning electron microscope which was set at -5℃ inside temperature and at 3.2 kV.

Quantitative RT-PCR (qRT-PCR) analysis

Total RNA was extracted by RNeasy Plant Mini Kit (QIAGEN), whereas first-strand cDNA synthesis was performed using the Omniscript RT Kit (QIAGEN). StepOneTM Real-Time PCR system (Applied Biosystems) was used to analyze the relative

expression levels of the target genes (e.g. RAE2, rae2). Relative expression levels of the target genes were normalized to the levels of endogenous ubiquitin transcripts (Primers used are provided in Table 1). Each set of experiments was repeated three times, and the Comparative CT method (△△CT Method) was used to calculate the relative expression levels of the target genes. For qRT-PCR analysis of RAE2 and rae2, various plant parts (leaf blade, leaf sheath, stem, root and panicle) of Koshihikari and GLSL25 were used. Leaf blade, leaf sheath and roots were obtained from plants that were < 15 cm in height. Stems and panicles (<1 cm) were obtained from the plants 24 days after transplanting under short day condition.

Semi quantitative RT-PCR

RT-PCR was performed in a 50 μL solution containing a 2.5 μL aliquot of cDNA as the DNA template, 0.2 μM gene-specific primers (see Table 1), 10 mM deoxynucleotide triphosphates, 1 unit of ExTaq DNA polymerase (Takara), and reaction buffer. Amplifications of OsUBQ5 cDNAs was used as internal control. The reaction included an initial 5-min denaturation at 94°C, followed by 25 cycles of PCR (94°C for 30 s,

56°C for 30 s, and 72°C for 30 s), and a final 5-min extension at 72°C. The number of cycles used for amplification with each primer pair was adjusted to be in the linear range. All RT-PCR data are representative of at least three independent experiments.

in situ hybridization

Plant materials were fixed in 4% (wt/vol) paraformaldehyde and 0.25% (vol/vol) glutaraldehyde in 0.1 M sodium phosphate buffer (pH 7.2) overnight at 4°C, dehydrated through a graded ethanol series followed by a t-butanol series, and finally embedded in Paraplast Plus (Sherwood Medical). Microtome sections (8-10 μm) were mounted on adhesive glass slides (Matsunami Glass Ind., Ltd). Digoxygenin-labeled RNA probes were transcribed with T7 RNA polymerase. The probes were amplified using the respective primer set for RAE2 and rae2 and cloned into the pBluescript II SK+ and pBluescript II KS+ vectors. Hybridization and immunological detection of the hybridized probes were performed according to a described method (Kouchi et al. 1993) with some modifications.

Phylogenetic tree

RAE2 sequence was identified through reciprocal best-BLAST match searches of the Phytozome and National Center for Biotechnology Information (NCBI) databases. Accession numbers or locus IDs of EPF/EPFLs were derived from the NCBI database. Amino acid sequences for the C-terminal mature peptide region were aligned using the ClustalW program. The number of amino acid substitutions between each pair of EPF/EPFL proteins was estimated using the Jones-Taylor-Thornton (JTT) model with complete-deletion option. The phylogenetic tree was reconstructed by the neighbor-joining method. Bootstrap values were estimated (with 1000 replicates) to assess the relative support for each branch, and bootstrap values were labeled with cutoff at 50. To construct the phylogenetic tree, the neighbor-joining method in MEGA version 6.1 was used.

RAE2 3-D conformation modeling

Three-dimensional structures of RAE2 peptide were predicted by homology modeling system of the Mäestro (Schrödinger, NY, USA) software. The structure of Stomagen

(protein database ID: 2LIY) determined by NMR was used as the template. Pairwise alignment was improved manually by minor editing based on the secondary structure predictions and disulfide bonds were allowed to form during the modeling. The hypothetic structure of rae2 (OsEPFL1) could not be modelled because it lost two cysteine residues (C5 and C6) and did not fit the Stomagen template. Yellow: cysteine residues, red: disulfide bonds and atoms, blue: antiparallel beta sheets.

Diversity analysis

PCR products amplifying the RAE2 gene were assayed for polymorphisms using BigDye Terminator v3.1 Cycle Sequencing Kit and analyzed with CodonCode Aligner 6.0.2. All sequences were aligned to the rice reference genome (cv. Nipponbare) and predicted cDNAs were extracted and translated. SNPs and indels detected were used to construct RAE2 gene haplotypes (n=123 with full sequence). Polymorphisms with a minor allele count (MAC) <1 were filtered out for gene haplotype construction unless they represented a frameshift mutation. The geographical map displaying origin of diverse rice accessions was created using R package ‘maps.’

Selective Sweep Analysis

Sixty-seven *O. sativa* and 65 *O. rufipogon/O. nivara* accessions were analyzed for evidence of selective sweep. SNP information on the *O. sativa* set were extracted from re-sequencing data (unpublished, McCouch) and imputed for RAE2 protein variant (4C, 6C, 7C) using tag SNPs from a rice SNP array identified on an overlap set of *O. sativa* that were Sanger Sequenced in the diversity analysis (McCouch et al. 2016). SNP data on the *O. rufipogon/O. nivara* set were derived from Genotyping-By-Sequencing (GBS) information on chromosome 8 (unpublished, McCouch). An overlap SNP set between the re-sequencing data (1,137,573 markers on chromosome 8) and the GBS data (34,267 markers on chromosome 8) were used for estimation of nucleotide diversity and distance. π (nucleotide diversity) and d (distance between groups) statistics were calculated using sliding windows of 100 SNPs, with step size 2 variants, across chromosome 8 (Weir and Cockerham 1984). We enumerated the sequence differences between a given pair of DNA segments and calculated sequence differentiation using the Jukes-Cantor model (Li 1997). Genetic distances between population pairs and

nucleotide diversity within populations were estimated based on Nei (Nei 1973). To enable comparisons between different analyses, we estimated per-kb values of π and d by dividing the total value for a window by the reference map distance (in kb) between the first and last SNP. Since only sub-sets of sites on the chromosome was covered by sequence, this procedure results in a drastic underestimate of π and d. However, the degree of underestimation is the same across groups so values are comparable within our data set.

Protein extraction and immunoblot analysis

Crude protein extracts from several organs (e.g. callus, stem, leaf and spikelet) of pAct::RAE2-3xFLAG transgenic plants in the background of Nipponbare were prepared by grinding with liquid nitrogen. Total protein was extracted with 3.0 mL protein extraction buffer (20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 150 mM 2% Protease inhibitor cocktail (Complete, Roche), 0.1% TritonX). After centrifugation, supernatant mixed with an equal volume of 2× sample buffer (135 mM Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 0.2 w/v % bromphenol blue, 200 mM DTT) and boiling for 5 min.

Protein samples were separated by 15% SDS-PAGE and transferred to PVDF membrane (0.2 μm pore size, Millpore) by semi-dry blotting. The blots were treated with 5% skim milk in TBST (0.1w/v% Tween20, 2 mM Trisma Base, 13.7 mM NaCl , pH 7.4) for 1 h and subsequently incubated with anti-FLAG antibody (1:3,000) (v/v) (A8592, Sigma) for 2 h. Blots were washed three times with TBST for 10 min each. Goat anti-mouse IgG horseradish peroxidase–conjugated secondary antibody was incubated for 1 h, and blots were washed following the same procedure described above. All reactions were conducted at room temperature. Detection of peroxidase activity was performed according to the instruction manual from Pierce (Thermo Fisher, Massachusetts, USA).

Purification of recombinant RAE2

The recombinant RAE2 pro-peptide fused with 3xFLAG and a series of amino acid substitution-mutated peptides (muRAE2) were expressed in E. coli strain Rosetta (DE3) pLysS (NOVAGEN). The expressed recombinant proteins were purified by TALON beads (Clontech) according to the manufacturer instructions. The beads were washed 5

times with a wash buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1 % TritonX, 1 mM imidazole) and the recombinant proteins were collected using the elution buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1 % TritonX, 10 mM imidazole). The production of recombinant peptides was confirmed by 15% SDS-PAGE.

in vitro processing assay

To prepare the plant extracts, 1.0 g of each rice tissue (i.e., callus, leaf, stem, spikelet (<1 cm)) was collected and ground in liquid nitrogen following the procedure described above. The ground extract was centrifuged (15,000 rpm for 30 min at 4°C) and the resulting supernatant was used for the *in vitro* processing assay. For the assay, 0.5 μg of pro-RAE2-3xFLAG protein or other mutant proteins were mixed with 10 μg of each plant extract and incubated for 2 h at room temperature with or without 0.1% protease inhibitor cocktail (Complete, Roche). The concentration of recombinant peptides was determined using the Bio-Rad Protein Assay (Bio-Rad Laboratories). After incubation the peptides were separated by 15% (Fig. 20D, E) or 20% (Fig. 20A, B and F) SDS-PAGE. Immunoblotting was performed following the procedure described above.

We used anti-FLAG antibody (1:3,000) (v/v) for Fig. 20A, D, E and 24B, or anti-RAE2 antibody (1:1,000) (v/v) for Fig. 20F as primary antibody.

Specific antibody

A peptide antigen with 7-amino-acid (NH- 119 RDRLFDP 125 -COOH) in C-terminal region of RAE2 was synthesized, purified, and conjugated with keyhole limpet hemocyanin. The conjugate was injected into a rabbit to induce the production of anti-RAE2 polyclonal antibodies. These antibodies were purified from the rabbit serum using a HiTrap NHS-activated HP column (GE Healthcare) conjugated with the 7-amino-acid antigen peptide in accordance with the manufacturer's protocol.

Mass spectrometry

Proteins of panicle were separated by SDS-PAGE followed by CBB staining. The respective sized 70~75 KD bands were excised separately and then subjected to the in-gel digestions with trypsin and lysyl endopeptidase mixture (Promega) in 0.5M triethylammonium bicarbonate (pH8.5), 0.1% RapiGest (Waters) at 50°C for 1 hr. SLP1

product of in vitro transcription/translation system was immnoprecipitated using anti-DDDDK-tag mAb-Magnetic beads (MBL, Japan) and then eluate was performed same as above procedure. Two lots (#1 and #2) of panicle proteins and SLP1 product of *in vitro* translation were labeled by the iTRAQ reagent 114, 115, 116, respectively according to the instruction manual. The SLP1 synthetic five peptides mixture was labeled by iTRAQ reagent 117. iTRAQ reagents labeled samples were mixed and then were separated by reverse phase nano liquid chromatography (DiNa Nano LC system, KYA TECH Corporation, Tokyo) and were directly fractionated onto MALDI target plate with CHCA by a spotter (DiNa Map system, KYA TECH Corporation, Tokyo). iTRAQ mass spectrum analysis was same as above.

The SLP1 synthetic peptides used were as follows:

SLP1 52-71: QLPGVLAVIPDVLHKVHTTR

SLP1 72-80: SWDFLELER

SLP1 452-466: LGVKPAPVMAAFSSR

SLP1 543-560: SAIMTTAITGDNDSGKIR

SLP1 679-695: VTVYPPELSFESYGEER

in vitro transcription/translation

Protein synthesis was performed using the IN VITRO Transcription/Translation Reagents kit following the manufacturer's instructions (BioSieg, Tokushima, Japan). For in vitro transcription, the coding DNA sequence of FLAG tag (DYKDDDDK) was attached to the cDNA templates of SLP1by Tks Gflex DNA polymerase (TAKARA, Kyoto, Japan). Approximately 30 μg of RNA was prepared by T7 RNA polymerase-based transcription from the PCR product. The RNA samples were dissolved in 35 μl of RNase-free water and mixed with 10 μl of a wheat germ extract and 10 μl of amino acid mixture (BioSieg) at 16°C for 10 h. The synthesized proteins were confirmed by immunoblotting with an antibody against FLAG (Wako).

Synthetic peptide

All synthetic peptides were manufactured and purified to >95% purity by Biologica company. The RAE2 peptide (WT-RAE2) or mu-RAE2 peptides included the predicted cleavage site. The peptide sequences used were as follows:

WT-RAE2: AGEEEKVRLGSSPPSCYSK

muRAE2 (P64-G): AGEEEKVRLGSS**G**PSCYSK

muRAE2 (P65-G): AGEEEKVRLGSSP**G**SCYSK

muRAE2 (S66-A): AGEEEKVRLGSSPP**A**CYSK

muRAE2 (S66-T): AGEEEKVRLGSSPP**T**CYSK

muRAE2 (S66-D): AGEEEKVRLGSSPP**D**CYSK

muRAE2 (S66-N): AGEEEKVRLGSSPP**N**CYSK

Mass spectrometry: Detection of the cleavage site of RAE2.

SLP1 was prepared with in vitro transcription/ translation system. Cleavage assay was performed with SLP1 product and various synthetic RAE2 peptides in 50 mM Hepes-NaOH (pH8.0), 1mM DTT for 30min at 25°C. Assay mixture was filtrated by centrifugal filter devices (Amicon Ultra 10K device). Filtrate was applied on MonoSpin C18 (GL Sciences, Japan) and eluted with 70% acetonitrile-0.1% TFA. The eluate was spotted to MALDI plate with α-cyano-4-hydroxycinnamic acid (CHCA). MALDI-MS and MS/MS were performed on a SCIEX TOF/TOF™ 5800 System with version 4.1

software (Sciex). MS spectra were acquired in positive ion reflector mode and MS/MS spectra were acquired in positive ion mode with CID on. MS/MS data were analyzed by ProteinPilot Software 5.0 (Sciex).

References

Cai HW, and Morishima H. (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor. App. Genet.* **104**:1217-1228.

Dahlgren R, Clifford HT, and Yeo PF. (1985) The families of the monocotyledons: Structure, evolution and taxonomy *Springer*, New York.

Doi K. (1999) Construction and Utilization of *Oryza glaberrima* Introgression Lines in the Background of *O. sativa* L. *Ph.D thesis* Faculty of Agriculture, Kyushu Univ.

Engineer CB, Ghassemian M, Anderson JC, Peck SC, Hu H, and Schroeder JI. (2014) Carbonic anhydrases, EPF2 and a novel protease mediate CO2 control of stomatal development. *Nature* **513**: 246-250.

Fawcett JA, Kado T, Sasaki E, Takuno S, Yoshida K, et al. (2013) QTL Map Meets Population Genomics : An Application to Rice. *Plos One* **8**: e83720.

Furuta T, Komeda N, Asano K, Uehara K, Gamuyao R, et al. (2015) Convergent Loss of Awn in Two Cultivated Rice Species *Oryza sativa* and *Oryza glaberrima* Is Caused by Mutations in Different Loci. *G3:Genes, Genomes, Genetics* **5**: 2267-2274.

Groll UV, Berger D, and Altmann T. (2002) The Subtilisin-Like Serine Protease SDD1 Mediates Cell-to-Cell Signaling during *Arabidopsis* Stomatal Development. The *Plant cell* **14**: 1527-1539.

Hara K, Kajita R, Torii KU, Bergmann DC, and Kakimoto T. (2007) The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule. *Gene. Dev.* **21**: 1720-1725.

Hua L, Wang DR, Tan L, Fu Y, Liu F, et al. (2015) *LABA1*, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice. *Plant Cell* **27**: 1875-1888.

Hunt L, and Gray JE. (2009) Report The Signaling Peptide EPF2 Controls Asymmetric Cell Divisions during Stomatal Development. *Curr. Biol.* **19**: 864-869.

Jin J, Hua L, Zhu Z, Tan L, Zhao X, et al. (2016) GAD1 Encodes a Secreted Peptide That Regulates Grain Number, Grain Length, and Awn Development in Rice Domestication. *Plant Cell* **28**: 2453-2463.

Katsir L, Davies KA, Bergmann DC, and Laux T. (2011) Peptide Signaling in Plant Development. *Curr. Biol.* **21**: R356-R364.

Kondo T, Kajita R, Miyazaki A, Hokoyama M, Nakamura-Miura T, et al. (2010)

Stomatal Density is Controlled by a Mesophyll-Derived Signaling Molecule Stomata are composed of a pair of guard cells and a pore. *Plant Cell Physiol.* **51**: 1-8.

Lee JS, Kuroha T, Hnilova M, Khatayevich D, Kanaoka MM, et al. (2012) Direct interaction of ligand – receptor pairs specifying stomatal patterning. *Gene. Dev.* 26: 126-136.

Lee JS, Hnilova M, Maes M, Lin YC, Putarjunan A, et al. (2015) Competitive binding of antagonistic peptides fine-tunes stomatal patterning. *Nature* **522**: 439-443.

Luo J, Liu H, Zhou T, Gu B, Huang X, et al. (2013) *An-1* Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. *Plant Cell* **25**: 3360-76.

Marshall E, Costa LM, and Gutierrez-Marcos J. (2011) Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell–cell communication in plant reproduction and development. *J. Exp. Botany* **62**: 1677-1686.

Müller KJ, Romano N, Gerstner O, Garcia-Maroto F, Pozzi C, et al. (1995) The barley Hooded mutation caused by a duplication in a homeobox gene intron. *Nature* **374**: 727-730.

Ohki S, Takeuchi M, and Mori M. (2011) The NMR structure of stomagen reveals the basis of stomatal density regulation by plant peptide. *Nat. Comm.* **2**: 512-517.

Sato S, Ishikawa S, Shimono M, and Sinjyo C. (1996) Genetic studies on an awnness gene An-4 on chromosome 8 in rice. *Breeding Science* **46**: 321-327.

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, et al. (2010) Stomagen positively regulates stomatal density in Arabidopsis. *Nature* **463**: 241-246.

Takata N, Yokota K, Ohki S, Mori M, Taniguchi T, and Kurita M. (2013) Evolutionary Relationship and Structural Characterization of the EPF / EPFL Gene Family. *Plos One* **8**: 4-9.

Takahashi N, Alterfa HAH, and Sato T. (1986) Significant role of awn in rice [Oryza sativa] plants, 2: Effects of daylength and temperature on awn growth. *Tohoku Univ., Sendai (Japan). Inst. for Agricultural Research)*

Tanaka W, Toriba T, Ohmori Y, Yoshida A, Kawai A, et al. (2012) The YABBY gene TONGARI-BOUSHI1 is involved in lateral organ development and maintenance of meristem organization in the rice spikelet. *Plant Cell* **24**:80-95.

Toriba T, and Hirano H. (2014) The *DROOPING LEAF* and *OsETTIN2* genes promote

awn development in rice. *Plant Journal* **77**: 616-626.

Tripathi PL, and Sowdhamini R. (2006) Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics* **31**: 1-31.

Uchida N, Lee JS, Horst RJ, Lai HH, Kajita R, et al. (2012) Regulation of inflorescence architecture by intertissue layer ligand–receptor communication between endodermis and phloem. *Proc. Natl. Acad. Sci. USA* **109**: 6337-6342.

Uchida N, and Tasaka M. (2013) Regulation of plant vascular stem cells by endodermis derived EPFLfamily peptide hormones and phloem expressed ERECTA family receptor kinases. *J. Exp. Bot.* **64**: 5335-5343.

Wheeler JI, and Irving HR. (2011) Plant Signaling Peptides, *Springer*, New York

References in Materials and Methods

Kouchi H, and Hata S. (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* **238**: 106-119.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Li WH. (1997) Molecular Evolution. *Sinauer Associates* Sunderland, Massachusetts

McCouch SR, Wright MH, Tung CW, Maron LG, McNally KL, et al. (2016) Open access resources for genome-wide association mapping in rice. *Nat. Comm.* **7**: 10532.

Miki D, and Shimamoto K. (2004) Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant Cell* **45**: 490-495.

Miki D, Itoh R, Shimamoto K. (2005) RNA silencing of single and multiple members in a gene family of rice. *Plant physiol.* **138**: 1903-1913.

Nishimura A, Ashikari M, Lin S, Takashi T, Angeles ER, et al. (2005) Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. *Proc. Nat. Acad. Sci. U S A* **102**: 11940–11944.

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, et al. (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. and Bioengi.* **104**: 34-41.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci. U S A* **70**: 3321-3323.

Weir BS, and Cockerham CC. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Tables and Figs

Table 1. Primers used in this study.

Purpose	name	Sequence (5' -> 3')
Linkage mapping of RAE2	8KG23941	CACGCTTGTAAGGCTGAGTT
		ATTCCGTATCCGAAAACCTC
	8KG23994	TGGAACAACGTGAGATTGTC
		GTTCCTGATCAGATTGTTGC
	8KG23999	CATCCATCAACATGTCGTCG
		CGCCATGTATAGTGTGATTCCG
	8KG24021	TATCCTTCTTGGGTTCTTGC
		TGAATGTGGTGCATTTCATC
BAC screening	pk31	GCACCTCAGCCTGGTTTCAAG
	pk32	GTAGTAGTTTGGTTGTTCTCTTGC
RAE2 promoter sequencing	KU42	CCAAGATGACAGCATGCTACTG
	KU43	CCAATTCTTTGTAACAAAGGGTAG
RAE2 coding region cloning	KU32	CACCATGAGGACGGCGGCCACGCCGCCT
	KU35	TCAGGGGTCGAACAGGCG
RAE2 RNAi construct	RNAi-F	CACCGATAGATTCCGTGTAATAT
	RNAi-R	ATATTACACGGAATCTATC
qRT-PCR of RAE2	KU37	ATTTTGACCAGACCACCTCG
	KU38	CGCCCAGCTACTTATACCCA
qRT-PCR of ACT1	ACT1 RT-f	GGATCCATCTTGGCATCTCTCA
	ACT1 RT-r	GGGCCAGACTCGTCGTACTC
qRT-PCR of UBQ5	UBQ5 RT-f	AAACCCTAACGGGGAAGACCATAA
	UBQ5 RT-r	CCACAGTAATGGCGATCAAAATGA
in situ probe of RAE2	in situ-f	AGCTTCTTGGTAGGCGAGGTGT
	in situ-r	GAAGAAGACGGCGAGGAGGA
pACT::RAE2-3xFLAG	KU73	CGGGATCCATGAGGACGGCGGCCAC
	KU75	CCCAAGCTTGGGGTCGAACAGGCGGT
recombinant RAE2 pro-peptide	PRO-f	ATCGGATCCACGGCTCCCTCCTCCTCGCCGT
	PRO-r	CCCAAGCTTTCAGGGGTCGAAC
recombinant RAE2 mature peptide	MA-f	GATCGGATCCCGGCTGGGGTCGAGCCCGCC
	MA-r	CCCAAGCTTTCAGGGGTCGAAC
Alanine substituted construct #1	KU132	GAGGAGAAGGTGCGGGCGGCGGCGGCGGCGAGCTGCTACAGCAAGTGC
	KU133	CCCGTAGCACTTGCTGTAGCAGCTCGCCGCCGCCGCCGCCCGCACCTTCTC
Alanine substituted construct #2	KU111	GGGGAGGAGGAGGCGGCGCGGCTGGGGTCG
	KU112	CAGCCGCGCCGCCTCCTCCTCCCCAGCCAC
Alanine substituted construct #3	KU113	GAGGAGGAGAAGGCGGCGCTGGGGTCGAGC
	KU114	CCCCAGCGCCGCCTTCTCCTCCTCCCCAGC
Alanine substituted construct #4	KU121	GCTGTAGCAGCTCGCCGCGCTCGACCCCAG
	KU122	CTCGACCCCAGCCGCACCGCCGCCGCCTCCCCAGC
SLP1 cloning	KU127	CACCATGCAGACTTATGTGATCGTCTTTG
	KU139	GGAATTCCCTACCCGCAGGTCGTCTTG
semi qRT-PCR of SLP1	KU144	CGTGTCCCCTACAACATAATGTCC
	KU146	GATCTTGCCGCTGTCGTTGTC
in vitro translation of SLP1	FLAG-f	CCAGCAGGGAGGTACTATGCAGACTTATGTGATCGT
	FLAG-r	CCTTATGGCCGGATCCAAGAGCTCTTTTTTTTTTTTTACCCGCAGGTCGTCTTG

Table 2. Germplasm information used for RAE2 diversity study

Accession ID	Name	Species: subpopulation	Germplasm Repository	Origin	Awn Class (SES)	GC length	RAE2 protein length class
NSFTV7	Arias	*O. sativa: tropical japonica*	GSOR 301007	Indonesia	1	22	long/7C
NSFTV30	Chiem Chanh	*O. sativa: indica*	GSOR 301028	Vietnam	1	22	long/7C
NSFTV46	Dourado Agulha	*O. sativa: tropical japonica*	GSOR 301043	Brazil	5	20	short/4C
NSFTV50	DZ78	*O. sativa: aus*	GSOR 301046	Bangladesh	0	20	short/4C
NSFTV53	Firooz	*O. sativa: aromatic*	GSOR 301049	Iran	0	20	short/4C
NSFTV59	Gogo Lempuk	*O. sativa: tropical japonica*	GSOR 301055	Indonesia	9	24	med/6C
NSFTV75	Jambu	*O. sativa: tropical japonica*	GSOR 301068	Indonesia	5	20	short/4C
NSFTV93	Kitrana 508	*O. sativa: aromatic*	GSOR 301085	Madagascar	9	20	short/4C
NSFTV105	Mehr	*O. sativa: aus*	GSOR 301097	Iran	0	20	short/4C
NSFTV112	N12	*O. sativa: aromatic*	GSOR 301104	India	5	20	short/4C
NSFTV142	Shai-Kuh	*O. sativa: indica*	GSOR 301133	China	7	22	long/7C
NSFTV153	T26	*O. sativa: aus*	GSOR 301144	India	7	20	short/4C
NSFTV154	Ta Hung Ku	*O. sativa: temperate japonica*	GSOR 301145	China	9	24	med/6C
NSFTV160	NSF-TV 160	*O. sativa: aromatic*	GSOR 301151	Iran	1	20	short/4C
NSFTV173	Nipponbare	*O. sativa: temperate japonica*	GSOR 301164	Japan	1	20	short/4C
NSFTV200	P 737	*O. sativa: aus*	GSOR 301191	Pakistan	3	20	short/4C
NSFTV221	Sadri Belyi	*O. sativa: aromatic*	GSOR 301212	Azerbaijan	0	20	short/4C
NSFTV222	Paraiba Chines Nova	*O. sativa: indica*	GSOR 301213	Brazil	0	22	long/7C
NSFTV223	Priano Guaira	*O. sativa: tropical japonica*	GSOR 301214	Brazil	1	24	med/5C
NSFTV243	Tropical Rice	*O. sativa: temperate japonica*	GSOR 301233	Ecuador	0	22	long/7C
NSFTV250	Bulgare	*O. sativa: temperate japonica*	GSOR 301240	France	9	24	med/6C
NSFTV261	Shim Balte	*O. sativa: aus*	GSOR 301251	Iraq	9	20	short/4C
NSFTV265	Vialone	*O. sativa: temperate japonica*	GSOR 301255	Italy	3	22	long/7C
NSFTV269	Sundensis	*O. sativa: indica*	GSOR 301259	Kazakhstan	1	22	long/7C
NSFTV284	IR-44595	*O. sativa: indica*	GSOR 301274	Nepal	1	22	long/7C
NSFTV298	LD 24	*O. sativa: indica*	GSOR 301288	Sri Lanka	0	22	long/7C
NSFTV309	Manzano	*O. sativa: tropical japonica*	GSOR 301299	Zaire	0	20	short/4C
NSFTV310	R 101	*O. sativa: tropical japonica*	GSOR 301300	Zaire	0	22	long/7C
NSFTV337	Sabharaj	*O. sativa: indica*	GSOR 301327	Bangladesh	0	22	long/7C
NSFTV339	Yodanya	*O. sativa: indica*	GSOR 301329	Myanmar	0	22	long/7C
NSFTV349	Chang Ch'Sang Hsu Tao	*O. sativa: indica*	GSOR 301339	China	0	22	long/7C
NSFTV356	JC 117	*O. sativa: indica*	GSOR 301344	India	0	22	long/7C
NSFTV369	Sathi	*O. sativa: aus*	GSOR 301356	Pakistan	1	20	short/4C
NSFTV373	Lambayeque 1	*O. sativa: aromatic*	GSOR 301360	Peru	0	20	short/4C
NSFTV377	PR 304	*O. sativa: tropical japonica*	GSOR 301362	Puerto Rico	0	22	long/7C
NSFTV379	Wanica	*O. sativa: tropical japonica*	GSOR 301364	Suriname	0	22	long/7C
NSFTV380	Tainan-Iku No. 512	*O. sativa: temperate japonica*	GSOR 301365	Taiwan	0	20	short/4C
NSFTV381	325	*O. sativa: tropical japonica*	GSOR 301366	Taiwan	9	22	long/7C
NSFTV395	OS 6 (WC 10296)	*O. sativa: tropical japonica*	GSOR 301378	Zaire	1	22	long/7C
NSFTV398	93-11	*O. sativa: indica*	GSOR 301399	China	3	22	long/7C
NSFTV399	Spring	*O. sativa: tropical japonica*	GSOR 301381	United States	1	20	short/4C
NSFTV400	Yang Dao 6	*O. sativa: indica*	GSOR 301400	China	7	22	long/7C
NSFTV402		*O. spontanea*	IRGC80539	India	9	20	short/4C
NSFTV410		*O. nivara*	IRGC80759	Myanmar	9	21	med/6C
NSFTV413		*O. nivara*	IRGC81850	India	9	15	med/6C
NSFTV415		*O. spontanea*	IRGC81909	India	9	15	med/6C
NSFTV416		*O. spontanea*	IRGC81970	Thailand	0	22	long/7C
NSFTV422		*O. rufipogon*		Vietnam		15	med/6C
NSFTV427		*O. rufipogon*		China	9	21	med/6C
NSFTV431		*O. rufipogon*	IRGC82992	China	9	15	med/6C
NSFTV432		*O. rufipogon*		Thailand		21	long/7C
NSFTV433		*O. rufipogon*	IRGC83795	India	9	23	short/4C
NSFTV435		*O. rufipogon*	IRGC86448	Thailand		24	med/6C
NSFTV438 (438_B2_1_S2)		*O. rufipogon*		India	9	22	long/7C
NSFTV443		*O. nivara*	IRGC93183	Nepal	9	15	med/6C
NSFTV444		*O. nivara*	IRGC93188	Nepal	9	15	med/6C
NSFTV446		*O. spontanea*	IRGC93224	Nepal	9	15	med/6C
NSFTV450		*O. nivara*	IRGC100916	China	9	15	med/6C
NSFTV453		*O. rufipogon*	IRGC103404	Bangladesh	9	22	long/7C
NSFTV457 (457_B3_1_S2)		*O. nivara*		Bangladesh	9	27	med/6C
NSFTV461 (461_A1_1_S2)		*O. rufipogon*		China	9	20	short/4C
NSFTV467		*O. RUFIPOGON*	IRGC104624	China	5	22	med/6C
NSFTV472		*O. SPONTANEA*	IRGC104636	China		22	long/7C
NSFTV477		*O. SPONTANEA*	IRGC104967	China	9	22	long/7C
NSFTV481		*O. NIVARA*	IRGC105343	India	9	15	med/6C
NSFTV482		*O. RUFIPOGON*	IRGC105349	India	9	15	med/6C
NSFTV483 (483_C2_1_S2)		*O. RUFIPOGON*		Thailand	9	15	med/6C
NSFTV487 (487_C2_S2)		*O. NIVARA*		Sri Lanka	9	15	med/6C
NSFTV490		*O. RUFIPOGON*		Japan		24	med/6C
NSFTV492		*O. RUFIPOGON*		Japan	9	15	med/6C
NSFTV493		*O. NIVARA*	IRGC105706	Nepal	9	15	med/6C

Individuals used for RAE2 diversity study

Accession ID	Name	Species: subpopulation	Germplasm Repository	Origin	Awn Class (SES)	GC length	RAE2 protein length class
NSFTV494		*O. RUFIPOGON*	IRGC105711	India		15	med/6C
NSFTV495		*O. NIVARA*	IRGC105717	Cambodia	9	15	med/6C
NSFTV496		*O. RUFIPOGON*	IRGC105720	Cambodia		15	med/6C
NSFTV503		*O. RUFIPOGON*		Thailand	9	15	med/6C
NSFTV505 (505_A1_2_S2)		*O. RUFIPOGON*		Thailand		21	med/6C
NSFTV508		*O. RUFIPOGON*	IRGC105890	Bangladesh	0	21	med/6C
NSFTV509		*O. RUFIPOGON*	IRGC105897	Bangladesh	3	22	long/7C
NSFTV514		*O. RUFIPOGON*	IRGC105956	Indonesia	9	24	med/6C
NSFTV549		*O. RUFIPOGON*	IRGC81881	Indonesia	9	24	med/6C
NSFTV553		*O. RUFIPOGON*	IRGC100926	Japan	9	22	long/7C
NSFTV555 (555_B1_1_S2)					9	15	med/6C
NSFTV592						24	med/6C
NSFTV600			IRGC100187		9	22	long/7C
NSFTV602			IRGC100900		9	15	med/6C
NSFTV605			IRGC100911		9	22	long/7C
NSFTV665		*O. RUFIPOGON/O. SATIVA*	IRGC100203	Taiwan	0	22	long/7C
NSFTV666		*O. RUFIPOGON*	IRGC100211	Taiwan	9	15	med/6C
NSFTV669 (669_C2_3_S2)		*O. NIVARA*		Taiwan	9	21	med/6C
NSFTV673		*O. RUFIPOGON*	IRGC100647	Taiwan	9	22	med/5C
NSFTV676 (676_A1_1_S2)		*O. RUFIPOGON*		Taiwan	0	20	short/4C
NSFTV682		*O. RUFIPOGON*	IRGC100904	Japan	9	24	med/6C
NSFTV701		*O. NIVARA/O. RUFIPOGON*	IRGC103813	China	9	20	short/4C
NSFTV704		*O. RUFIPOGON/O. NIVARA*	IRGC103818	China	9	20	short/4C
NSFTV708 (708_A1_2_S2)		*O. NIVARA*		Bangladesh	9	21	med/6C
NSFTV711		*O. NIVARA*	IRGC103841	Bangladesh	9	15	med/6C
NSFTV719 (719_A1)		*O. NIVARA*		France	9	15	med/6C
NSFTV720		*O. NIVARA*	IRGC104703	France	9	15	med/6C
NSFTV721					9	15	med/6C
NSFTV736 (736_B2_1_S2)					9	24	med/6C
NSFTV743 (743_C1_2_S2T)		*O. NIVARA*		Nepal	9	15	short/4C
NSFTV751		*O. NIVARA*	IRGC105895	Bangladesh	9	24	med/6C
NSFTV759 (759_A1_3_S2)		*O. RUFIPOGON*		Cambodia	9	22	long/7C
NSFTV760		*O. NIVARA*	IRGC106345	Myanmar	9	21	med/6C
NSFTV762		*O. NIVARA*		Myanmar	9	15	med/5C
NSFTV765			W1943			24	med/6C
NSFTV767			W1945			24	med/6C
RLS10173		*O. barthii*	IRGC103912	Tanzania	9	18	med/6C
RLS5584		*O. barthii*	IRGC104119		9	18	med/6C
RLS10188		*O. barthii*	IRGC100933		9	18	med/6C
RLS10194		*O. barthii*	IRGC106303		9	18	med/6C
RLS10128		*O. barthii*	IRGC101196	Cameroon	9	18	med/6C
RLS10123	WAB 010850	*O. barthii*	IRGC86524	Chad	9	18	med/6C
RLS10179		*O. barthii*	IRGC104983	Niger	9	18	med/6C
RLS10183	W0864	*O. barthii*	IRGC106207	Mali		18	med/6C
RLS10177		*O. barthii*	IRGC104140	Cameroon	9	18	med/6C
RLS10157		*O. barthii*	IRGC106291	Mauritania	9	18	med/6C
RLS10190		*O. barthii*	IRGC100941		9	18	med/6C
RLS10242	DAN MANU (1)	*O. glaberrima*	TOG5474	Burkina Faso	0	18	med/6C
RLS10239	YAR KARENGESHE	*O. glaberrima*	TOG5440	Nigeria	0	18	med/6C
RLS10236	TOG5286	*O. glaberrima*	TOG5286		1	18	med/6C
RLS6183	YANDEV(1)	*O. glaberrima*	TOG5949	Liberia	1	18	med/6C
RLS10233	ZAKI BIAM-YANDE(WILD)1	*O. glaberrima*	TOG6193	Nigeria	0	18	med/6C
RLS10245	SHENDAM (WEEDY)1	*O. glaberrima*	TOG5984	Nigeria	1	18	med/6C
RLS10257	YAR BUTUKA	*O. glaberrima*	TOG5467	Nigeria	1	18	med/6C
RLS10253	DAN MAIWUYA (6)	*O. glaberrima*	TOG5390	Nigeria	0	18	med/6C
RLS10262	NEW AYOMA LOCAL (2)	*O. glaberrima*	TOG7402	Ghana	5	18	med/6C
RLS10247	SHAWHON (2)	*O. glaberrima*	TOG5747	Liberia	1	18	med/6C
RLS10264	ACC 100982	*O. glaberrima*	TOG6211	Guinea Bissau	0	18	med/6C
RLS10265	QUE (2)	*O. glaberrima*	TOG5815	Liberia	0	18	med/6C

Table 3. Seven different length polymorphisms in this GC-repeat region of RAE2.

	nt length	protein length class	wild Asian	cult. Asian	wild African	cult. African
African	18	med.			10	11
Asian	15	med.	25*			
	21	med.	9			
	24	med.	9	4		
	20	short	5	17		
	23	short	1			
	22	long	12**	20		

*one accession showed short RAE2 protein length (4C) because of the 1 bp insertion in apart from GC-rich region

**one accession showed medium RAE2 protein length (5C) because of the 1 bp insertion in apart from GC-rich region

Table 4. RAE2 variants across the five subpopulations of O. sativa.

Protein type	*tej*	*trj*	*aro*	*aus*	*ind*
Short (dysfunc.)	0.33	0.33	1.00	1.00	
Medium (func.)	0.33	0.18			
Long (dysfunc.)	0.33	0.55			1.00
n	6	11	6	6	12

Table 5. Germplasm information used for RAE2 selective sweep analysis

Accession ID	Name	Species: subpopulation	Germplasm Repository	Origin	RAE2 protein length class
NSFTV1	Agostano	*O. sativa: temperate japonica*	IRGC126380		4C/short
NSFTV104	Mansaku	*O. sativa: temperate japonica*	IRGC117811		6C/med
NSFTV108	Moroberekan	*O. sativa: tropical japonica*	IRGC117621		4C/short
NSFTV110	Mudgo	*O. sativa: indica*	IRGC117818		7C/long
NSFTV154	Ta_Hung_Ku	*O. sativa: temperate japonica*	IRGC117904		6C/med
NSFTV16	Bico_Branco	*O. sativa: aromatic*	IRGC117658		4C/short
NSFTV163	Taducan	*O. sativa: indica*	IRGC117906		7C/long
NSFTV165	Trembese	*O. sativa: tropical japonica*	IRGC117921		7C/long
NSFTV173	Nipponbare	*O. sativa: temperate japonica*			4C/short
NSFTV174	Azucena	*O. sativa: tropical japonica*			7C/long
NSFTV18	BJ1	*O. sativa: aus*	IRGC117661		4C/short
NSFTV19	Black_Gora	*O. sativa: aus*	IRGC117662		4C/short
NSFTV207	Sigadis	*O. sativa: indica*	IRGC117889		7C/long
NSFTV226	IRAT_44	*O. sativa: tropical japonica*	IRGC117762		7C/long
NSFTV23	Canella_De_Ferro	*O. sativa: tropical japonica*	IRGC117675		7C/long
NSFTV248	Caucasica	*O. sativa: temperate japonica*	IRGC117677		4C/short
NSFTV268	Vavilovi	*O. sativa: temperate japonica*	IRGC117928		4C/short
NSFTV28	Champa_Tong_54	*O. sativa: aus*	IRGC117680		4C/short
NSFTV29	Chau	*O. sativa: indica*	IRGC117682		7C/long
NSFTV317	DJ123	*O. sativa: aus*	IRGC117711		4C/short
NSFTV336	Paung_Malaung	*O. sativa: aus*	IRGC117847		4C/short
NSFTV338	Sitpwa	*O. sativa: temperate japonica*	IRGC117892		4C/short
NSFTV341	Shirkati	*O. sativa: aus*	IRGC117885		4C/short
NSFTV369	Sathi	*O. sativa: aus*	IRGC117878		4C/short
NSFTV378	Kalubala_Vee	*O. sativa: aus*	IRGC117774		4C/short
NSFTV397	Cybonnet	*O. sativa: tropical japonica*	IRGC117699		4C/short
NSFTV402		*O. spontanea*	IRGC80539	India	4C/short
NSFTV410		*O. nivara*	IRGC80759	Myanmar	6C/med
NSFTV413		*O. nivara*	IRGC81850	India	6C/med
NSFTV415		*O. spontanea*	IRGC81909	India	6C/med
NSFTV416		*O. spontanea*	IRGC81970	Thailand	7C/long
NSFTV422		*O. rufipogon*		Vietnam	6C/med
NSFTV427		*O. rufipogon*		China	6C/med
NSFTV43	Dee_Geo_Woo_Gen	*O. sativa: indica*	IRGC117705		7C/long
NSFTV431		*O. rufipogon*	IRGC82992	China	6C/med
NSFTV432		*O. rufipogon*		Thailand	7C/long
NSFTV433		*O. rufipogon*	IRGC83795	India	4C/short
NSFTV435		*O. rufipogon*	IRGC86448	Thailand	6C/med
NSFTV438 (438_B2_1_S2)		*O. rufipogon*		India	7C/long
NSFTV443		*O. nivara*	IRGC93183	Nepal	6C/med
NSFTV444		*O. nivara*	IRGC93188	Nepal	6C/med
NSFTV446		*O. spontanea*	IRGC93224	Nepal	6C/med
NSFTV450		*O. nivara*	IRGC100916	China	6C/med
NSFTV453		*O. rufipogon*	IRGC103404	Bangladesh	7C/long
NSFTV457 (457_B3_1_S2)		*O. nivara*		Bangladesh	6C/med
NSFTV461 (461_A1_1_S2)		*O. rufipogon*		China	4C/short
NSFTV467		*O. RUFIPOGON*	IRGC104624	China	6C/med
NSFTV472		*O. SPONTANEA*	IRGC104636	China	7C/long
NSFTV477		*O. SPONTANEA*	IRGC104967	China	7C/long
NSFTV481		*O. NIVARA*	IRGC105343	India	6C/med
NSFTV482		*O. RUFIPOGON*	IRGC105349	India	6C/med
NSFTV483 (483_C2_1_S2)		*O. RUFIPOGON*		Thailand	6C/med
NSFTV487 (487_C2_S2)		*O. NIVARA*		Sri Lanka	6C/med
NSFTV490		*O. RUFIPOGON*		Japan	6C/med
NSFTV492		*O. RUFIPOGON*		Japan	6C/med
NSFTV493		*O. NIVARA*	IRGC105706	Nepal	6C/med
NSFTV494		*O. RUFIPOGON*	IRGC105711	India	6C/med
NSFTV495		*O. NIVARA*	IRGC105717	Cambodia	6C/med
NSFTV496		*O. RUFIPOGON*	IRGC105720	Cambodia	6C/med
NSFTV503		*O. RUFIPOGON*		Thailand	6C/med

Accession ID	Name	Species: subpopulation	Germplasm Repository	Origin	RAE2 protein length class
NSFTV505 (505_A1_2_S2)		*O. RUFIPOGON*		Thailand	6C/med
NSFTV508		*O. RUFIPOGON*	IRGC105890	Bangladesh	6C/med
NSFTV509		*O. RUFIPOGON*	IRGC105897	Bangladesh	7C/long
NSFTV51	Early_Wataribune	*O. sativa: temperate japonica*	IRGC117727		4C/short
NSFTV514		*O. RUFIPOGON*	IRGC105956	Indonesia	6C/med
NSFTV549		*O. RUFIPOGON*	IRGC81881	Indonesia	6C/med
NSFTV553		*O. RUFIPOGON*	IRGC100926	Japan	7C/long
NSFTV555 (555_B1_1_S2)					6C/med
NSFTV56	Geumobyeo	*O. sativa: temperate japonica*	IRGC117612		4C/short
NSFTV57	Gharib	*O. sativa: indica*	IRGC117739		7C/long
NSFTV592					6C/med
NSFTV600			IRGC100187		7C/long
NSFTV602			IRGC100900		6C/med
NSFTV605			IRGC100911		7C/long
NSFTV612	IR64	*O. sativa: indica*			7C/long
NSFTV620	Jasmine85	*O. sativa: indica*	IRGC125597		7C/long
NSFTV628	Jefferson	*O. sativa: tropical japonica*	IRGC126385		4C/short
NSFTV630	Saber	*O. sativa: tropical japonica*	IRGC126393		4C/short
RLS672	Minghui_63	*O. sativa: indica*	IRGC117271		7C/long
NSFTV665		*O. RUFIPOGON/O. SATIVA*	IRGC100203	Taiwan	7C/long
NSFTV666		*O. RUFIPOGON*	IRGC100211	Taiwan	6C/med
NSFTV669 (669_C2_3_S2)		*O. NIVARA*		Taiwan	6C/med
NSFTV673		*O. RUFIPOGON*	IRGC100647	Taiwan	5C/med
NSFTV676 (676_A1_1_S2)		*O. RUFIPOGON*		Taiwan	4C/short
NSFTV682		*O. RUFIPOGON*	IRGC100904	Japan	6C/med
NSFTV7	Arias	*O. sativa: tropical japonica*	IRGC126381		7C/long
NSFTV701		*O. NIVARA/O. RUFIPOGON*	IRGC103813	China	4C/short
NSFTV704		*O. RUFIPOGON/O. NIVARA*	IRGC103818	China	4C/short
NSFTV708 (708_A1_2_S2)		*O. NIVARA*		Bangladesh	6C/med
NSFTV711		*O. NIVARA*	IRGC103841	Bangladesh	6C/med
NSFTV719 (719_A1)		*O. NIVARA*		France	6C/med
NSFTV720		*O. NIVARA*	IRGC104703	France	6C/med
NSFTV721					6C/med
NSFTV736 (736_B2_1_S2)					6C/med
NSFTV743 (743_C1_2_S2T)		*O. NIVARA*		Nepal	4C/short
NSFTV751		*O. NIVARA*	IRGC105895	Bangladesh	6C/med
NSFTV759 (759_A1_3_S2)		*O. RUFIPOGON*		Cambodia	7C/long
NSFTV760		*O. NIVARA*	IRGC106345	Myanmar	6C/med
NSFTV762		*O. NIVARA*		Myanmar	5C/med
NSFTV765			W1943		6C/med
NSFTV767			W1945		6C/med
NSFTV81	Kalamkati	*O. sativa: aus*	IRGC117773		4C/short
NSFTV84	Kaniranga	*O. sativa: tropical japonica*	IRGC117776		7C/long
NSFTV85	Kasalath	*O. sativa: aus*	IRGC117617		4C/short
RLS29440	KUI_SALI	*O. sativa: aromatic*			4C/short
RLS49428	JC111	*O. sativa: aromatic*			4C/short
RLS29427	JC101	*O. sativa: aromatic*			4C/short
RLS5667	ARC_13523	*O. sativa: aromatic*			4C/short
RLS5669	Sathi_Basmati	*O. sativa: aromatic*			4C/short
RLS5665	Ambemohar	*O. sativa: aromatic*			4C/short
RLS29412	Basmati	*O. sativa: aromatic*			4C/short
RLS5670	Taraori_Basmati	*O. sativa: aromatic*			4C/short
RLS6303	Basmati_370	*O. sativa: aromatic*			4C/short
RLS29421	Sadri_Belyi	*O. sativa: aromatic*			4C/short
RLS29461	X9524	*O. sativa: aus*			4C/short
RLS29443	Khao_Gaew	*O. sativa: aus*			4C/short
RLS29426	Gie_57	*O. sativa: aus*			4C/short
RLS29437	BADAL89	*O. sativa: aus*			4C/short
RLS29464	Jhona349	*O. sativa: aus*			4C/short
RLS367	Chati_Kamma_Nangarhar	*O. sativa: aus*			4C/short

Accession ID	Name	Species: subpopulation	Germplasm Repository	Origin	RAE2 protein length class
RLS29433	TD2	*O. sativa: indica*			7C/long
RLS29419	Leung_Pratew	*O. sativa: indica*			7C/long
RLS29431	Popot_165	*O. sativa: indica*			7C/long
RLS29418	Guan.Yin.Tsan	*O. sativa: indica*			7C/long
RLS29429	JC91	*O. sativa: indica*			7C/long
RLS5364	CO39	*O. sativa: indica*			7C/long
RLS930	Short_Grain	*O. sativa: indica*			7C/long
RLS29463	X9311	*O. sativa: indica*			7C/long
RLS460	IR8	*O. sativa: indica*			7C/long
RLS5316	Taichungsien17	*O. sativa: indica*			7C/long
RLS5317	Tainungsien20	*O. sativa: indica*			7C/long

Table 6. The list of SP-8 type protease expressed in the spikelet.

No.	Locus IDs	Chromosome location	Accession	FeatureNum (Link to graph)
1	LOC_Os01g17160	chr01	AK119444	36860
2	LOC_Os01g50680	chr01	-	5682
3	LOC_Os01g52750	chr01	AK108195	17131
4	LOC_Os01g56320	chr01	AK070376	37342
5	LOC_Os01g58240	chr01	AK109067	14086
6	LOC_Os01g58260	chr01	AF200467	30347
7	LOC_Os01g58270	chr01	-	1768
8	LOC_Os01g58280	chr01	AK066488	6645
9	LOC_Os01g58290	chr01	AK100351	18445
10	LOC_Os01g6485	chr01	-	-
11	LOC_Os01g64860	chr01	AK062271	19280
12	LOC_Os02g10520	chr02	AK120287	30014
			AK106394	40115
13	LOC_Os02g16940	chr02	-	8596
14	LOC_Os02g17000	chr02	AK110825	25369
15	LOC_Os02g17060	chr02	-	-
16	LOC_Os02g17080	chr02	-	-
17	LOC_Os02g17090	chr02	AK072092	42776
18	LOC_Os02g17150	chr02	-	35042
19	LOC_Os02g44520	chr02	AK103515	33598
			AK067099	34821
			AK066478	41274
20	LOC_Os02g44590	chr02	AB037371	19329
			AK072929	20201
			AK100551	21433
21	LOC_Os02g53850	chr02	-	16087
22	LOC_Os02g53860	chr02	AK106527	5667
			AK121728	42860
23	LOC_Os02g53910	chr02	-	-
24	LOC_Os02g53970	chr02	AK070669	30949
25	LOC_Os03g02750	chr03	AK069220	12951
26	LOC_Os03g04950	chr03	-	-
27	LOC_Os03g06290	chr03	AK071242	43120
28	LOC_Os03g13930	chr03	-	39736
29	LOC_Os03g31630	chr03	-	37303
30	LOC_Os03g40830	chr03	AK105749	29457
31	LOC_Os03g55350	chr03	AK101646	22817
			AK103255	31812
32	LOC_Os04g02960	chr04	CI260116	1713
33	LOC_Os04g02980	chr04	-	4178
34	LOC_Os04g03060	chr04	-	16547
35	LOC_Os04g03100	chr04	-	26338
36	LOC_Os04g03710	chr04	-	-
37	LOC_Os04g03800	chr04	-	-
38	LOC_Os04g03810	chr04	AK062269	23506
39	LOC_Os04g03850	chr04	-	-
40	LOC_Os04g10360	chr04	CB653384	32192
41	LOC_Os04g35140	chr04	AK105112	5780
42	LOC_Os04g45960	chr04	AK106823	12015
			AY644644	15417
			AY683198	21133
43	LOC_Os04g47150	chr04	AK100861	1851
44	LOC_Os04g47160	chr04	-	23449
45	LOC_Os04g48420	chr04	-	-
46	LOC_Os05g30580	chr05	AK064686	5180
47	LOC_Os05g36010	chr05	AK067138	31360
48	LOC_Os06g06800	chr06	-	-
49	LOC_Os06g06810	chr06	AK071415	24262
50	LOC_Os06g40700	chr06	AK109185	4261
51	LOC_Os06g41880	chr06	-	39167
52	LOC_Os06g48650	chr06	AK102835	3665
53	LOC_Os07g39020	chr07	AK107610	24754
54	LOC_Os07g48650	chr07	AK119348	41308
55	LOC_Os08g23740	chr08	-	-
56	LOC_Os08g35090	chr08	CI043104	38603
57	LOC_Os09g26920	chr09	CI383807	8743
58	LOC_Os09g30250	chr09	CI269495	4777
59	LOC_Os09g36110	chr09	-	16075
60	LOC_Os10g25450	chr10	CI191448	30831
61	LOC_Os10g38080	chr10	AK069238	21727
62	LOC_Os11g15520	chr11	AK110921	33825
63	LOC_Os12g23980	chr12	-	23981

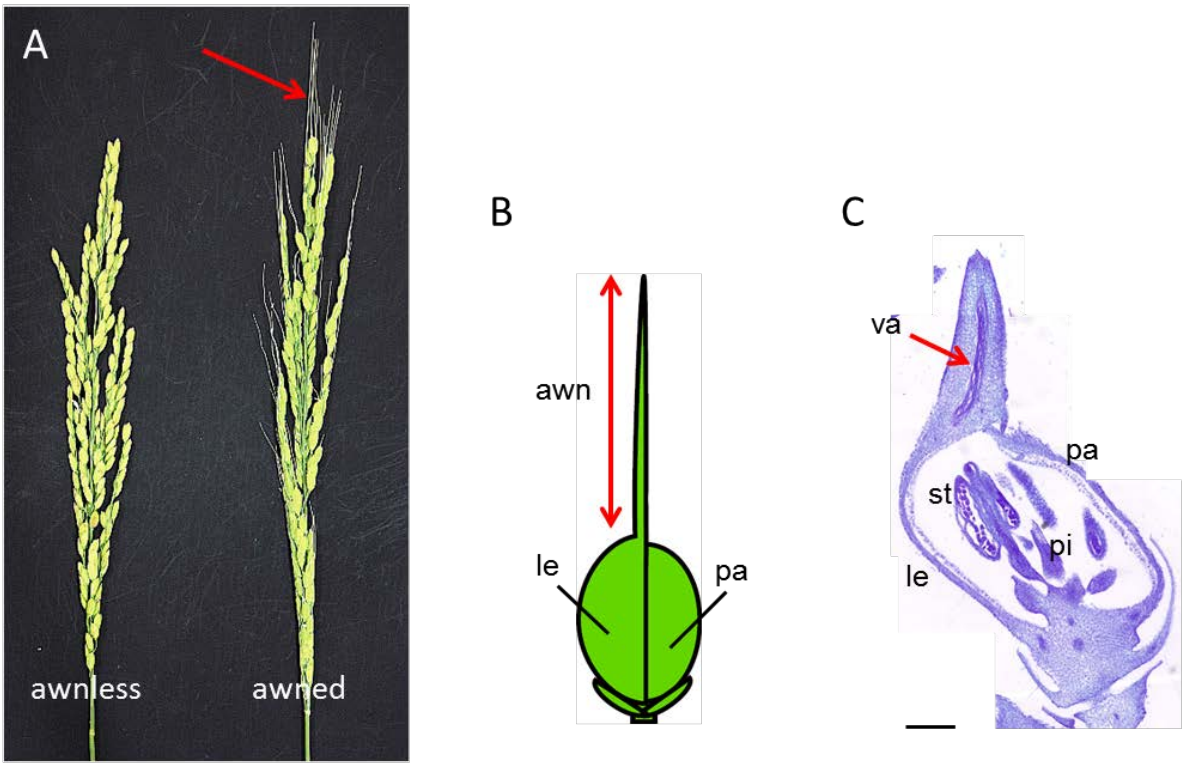
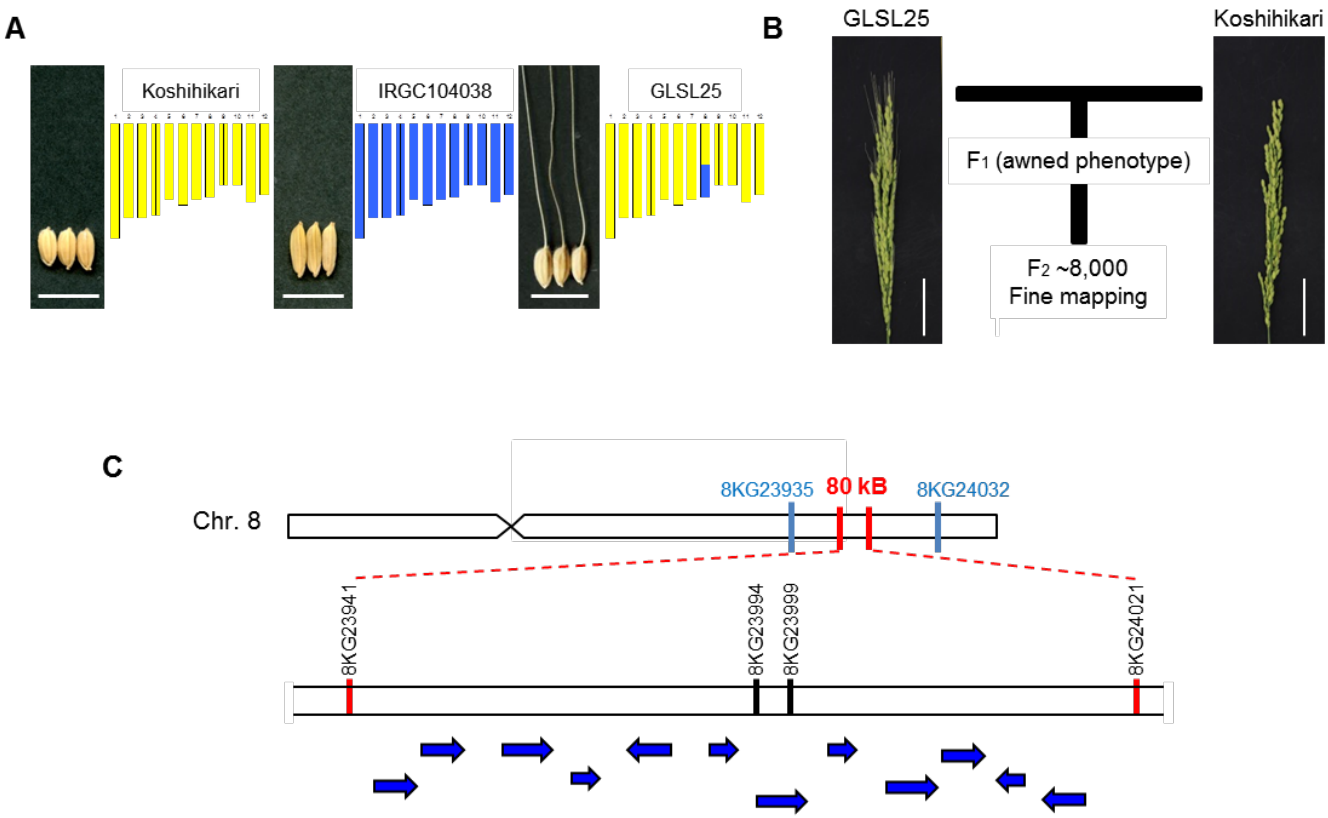


Figure 1. Awn is the spinous like structure at the tip of lemma. (A) Left is the awnless panicle and right is the awned one. Red arrow points the awn. (B) The drawing of individual spikelet with awn. (C) The longitudinal section of rice spikelet with awn. Le=lemma, pa=palea, va= vascular bundle, st=stamen, pi=pistil. Bar length is 100µm.



D

Gene ID	Description
Os08g0484600	OSK4.
Os08g0484633	Hypothetical protein.
Os08g0484700	Homeodomain-like containing protein.
Os08g0484800	Hypothetical conserved gene.
Os08g0485000	Similar to PHI-1.
LOC_Os08g37840	Transpozon-like protein.
Os08g0485300	Hypothetical conserved gene.
Os08g0485400	Similar to Oxidoreductase.
Os08g0485500	Hypothetical conserved gene.
Os08g0485600	Methyl-CpG DNA binding domain containing protein.
Os08g0485700	Methyl-CpG DNA binding domain containing protein.
Os08g0485800	Barwin-related endoglucanase domain containing protein.

Figure 2. Positional cloning of *RAE2*. (A) Seed phenotypes and graphical genotypes of Koshihikari (yellow), IRGC104038 (blue) and GLSL25. (B) The scheme of material production for fine mapping. (C)RAE2 was further delimited to an 80 kb genomic region between the markers 8KG23941 and 8KG24021. Blue arrows represent the 12 genes within the candidate region. (D) The list of 12 candidate genes.

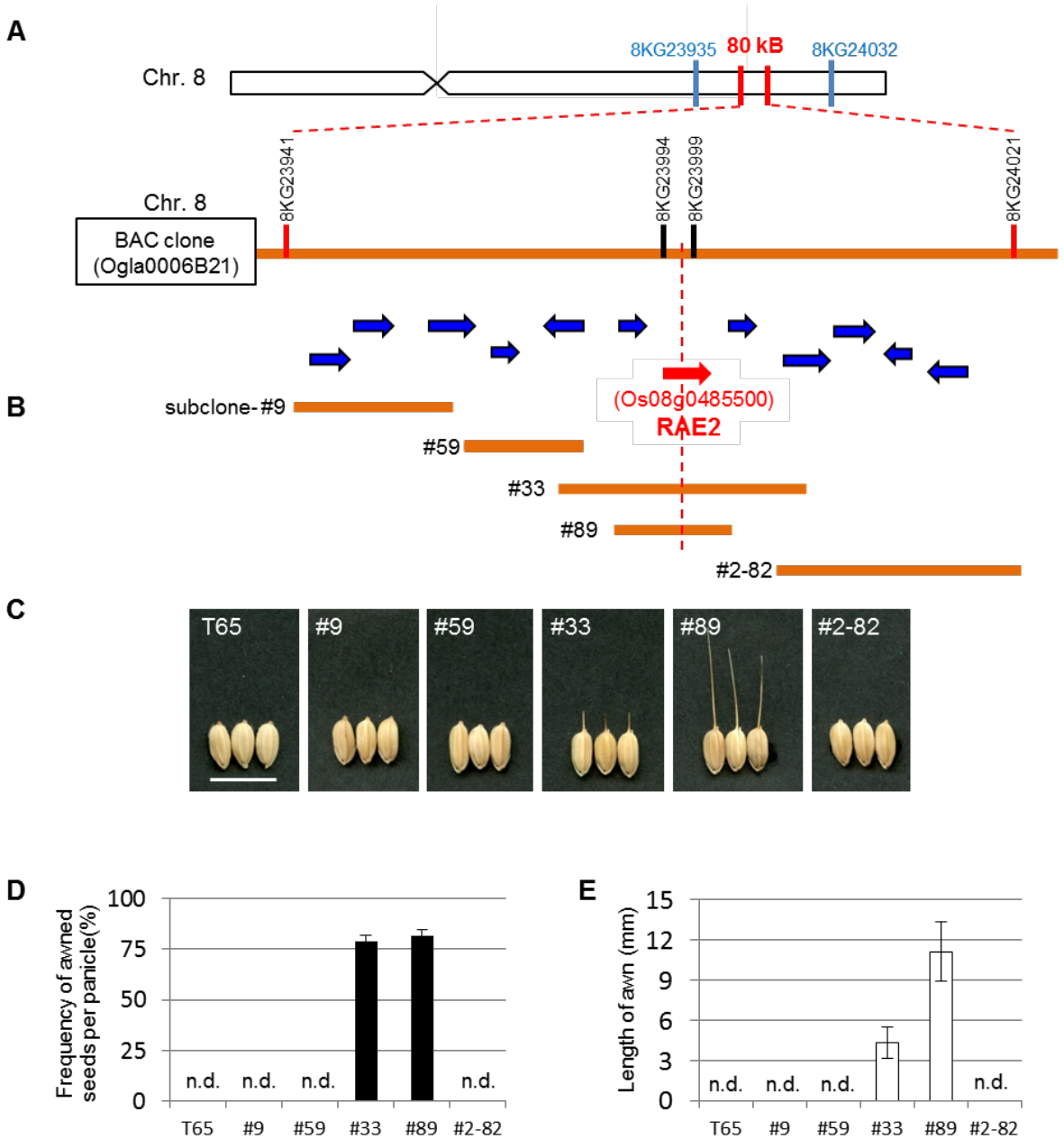


Figure 3. Complementation of *RAE2* by using *O. glaberrima* BAC subclone. (A) The RAE2 candidate region same as Fig. 2C. The orange line represents BAC clone from CG14 (*O. glaberrima*) named Ogla0006B21 overlapped entirely with the 80kb candidate region on chromosome 8. (B) Five sub-clones (#9, #59, #33, #89, #2-82) are constructed from Ogla0006B21 by partial digestion. (C-E) Evaluation of transgenic plants with each subclone in cv. Taichung65 (T65) genetic background: (C) seed phenotype, (D) frequency of awned seeds per panicle, and (E) awn length. No visible awn is observed in some lines indicated as n.d. (not detected). Bar length represents 1 cm. Values represent mean ± SE (n = 4).

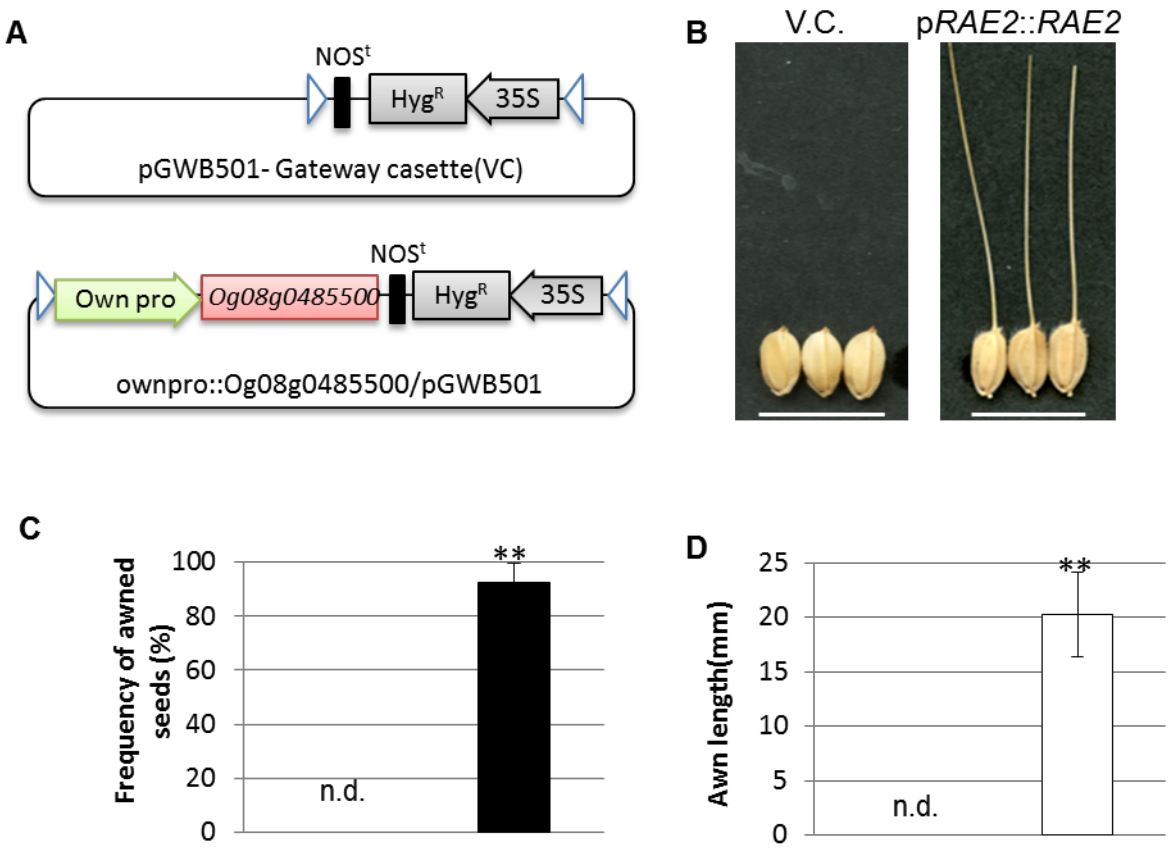


Figure 4. Identification and functional characterization of *RAE2*. (A) The construction map of plasmid vector pGWB501 (vector control; V. C.) and *RAE2* gene from *O. glaberrima* genetic DNA(pRAE2::RAE2). (B-D) Evaluation of transgenic plants with each construct showed in Fig.4A: (B) seed phenotype, (C) frequency of awned seeds per panicle, and (D) awn length. No visible awn was observed in pGWB501 (V.C.) indicated as n.d. (not detected). Bar length represents 1 cm. Values represent mean ± SE (n = 4). **P < 0.01 based on two-tailed Student's t-test.

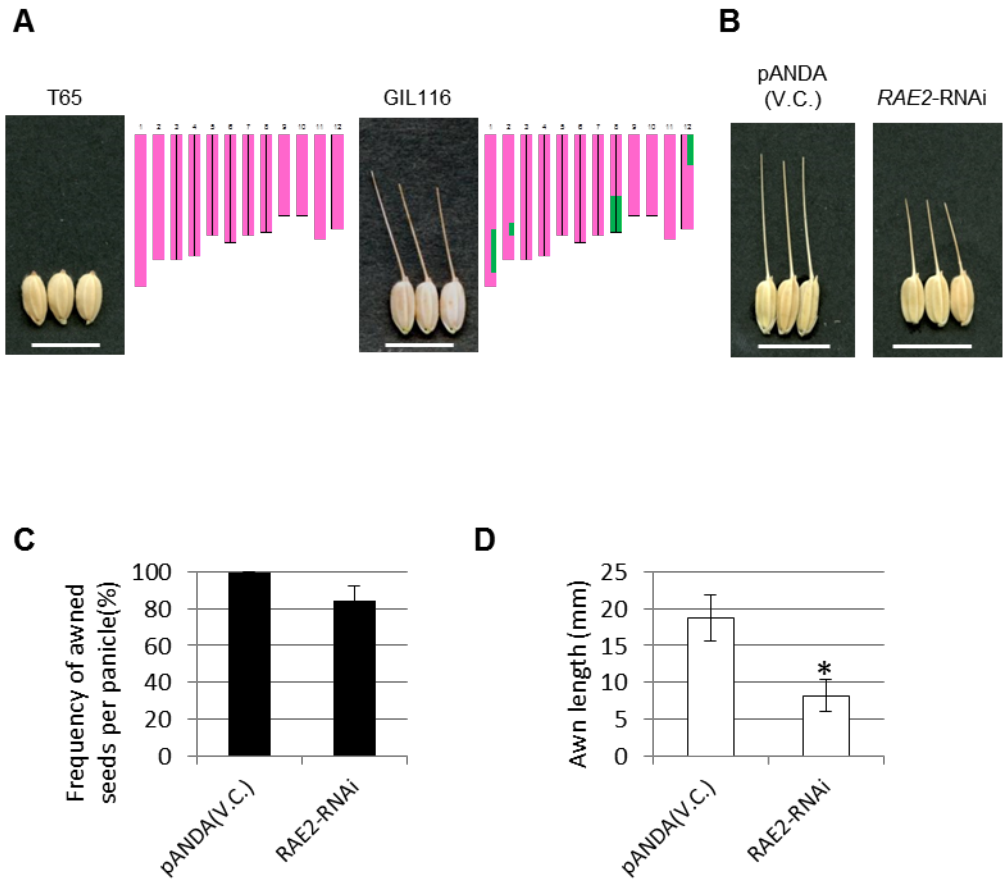


Figure 5. RNAi of RAE2 makes awn shorter. (A) Seed phenotypes and graphical genotypes of T65 (pink) and GIL 116 including IRGC104038 chromosome segment (green) Bar length is 1 cm. (B-D) Evaluation of vector control (pANDA (V.C.)) and RNAi line (RAE2-RNAi) harboring the 3' UTR region of the Os08g0485500 into GIL116: (B) seed phenotype, (C) frequency of awned seeds per panicle, and (D) awn length. Bar length represents 1 cm. Values represent mean ± SE (n = 4). The statistical significance is at *P < 0.05 based on two-tailed Student's t-test.

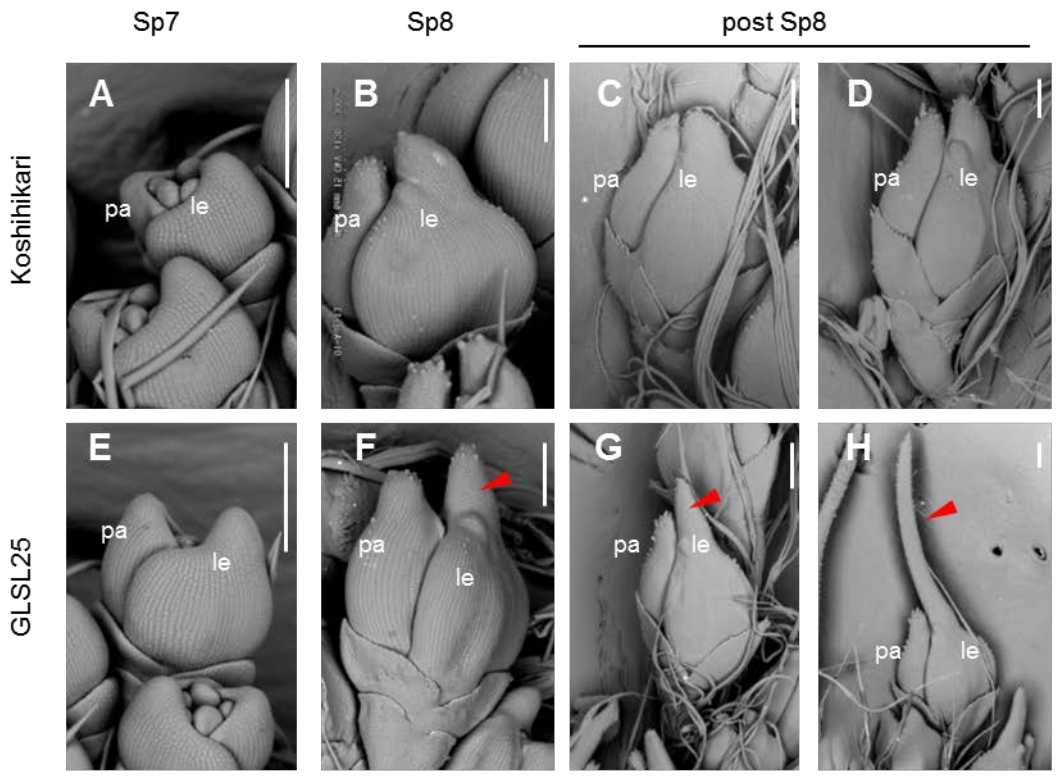


Figure 6. Scanning electron microscopy images of spikelets at different developmental stages. (A-H)The spikelets in Koshihikari (A-D) and GLSL25 (E-H). Developmental stages are classified into Sp7 (A, E), Sp8 (B, F), and post Sp8 (C, D, G, H) according to Oryzabase classification (http://www.shigen.nig.ac.jp/rice/oryzabase/devstageineachorgan/list). Scale bars represent 50μm. Red arrowhead indicates awn.

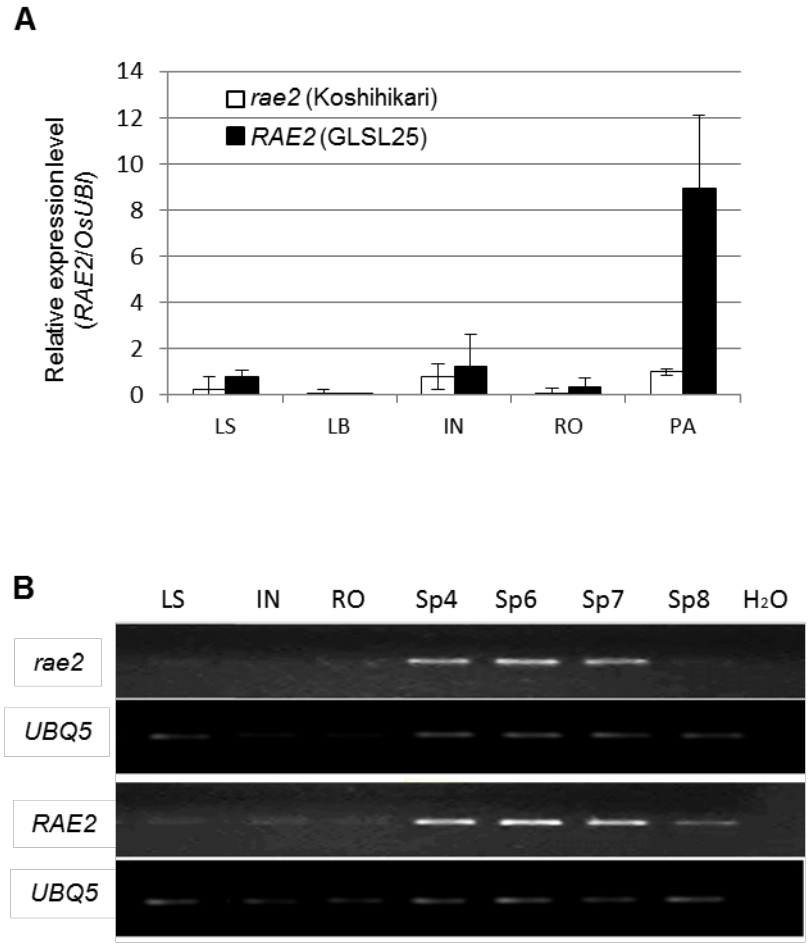


Figure 7. RAE2 expression pattern and correlation with awn development. (A) qRT-PCR showing RAE2 mRNA levels in different organs of Koshihikari and GLSL25 (LS= leaf sheath, LB= leaf blade, IN= internode, RO= root, PA= young panicle). OsUBI is used as internal control. Values represent mean ± SE (n = 3).(B) semi quantitative RT-PCR of *rae2* (upper lane) and *RAE2* (lower lane). UBQ5 is used as internal control.

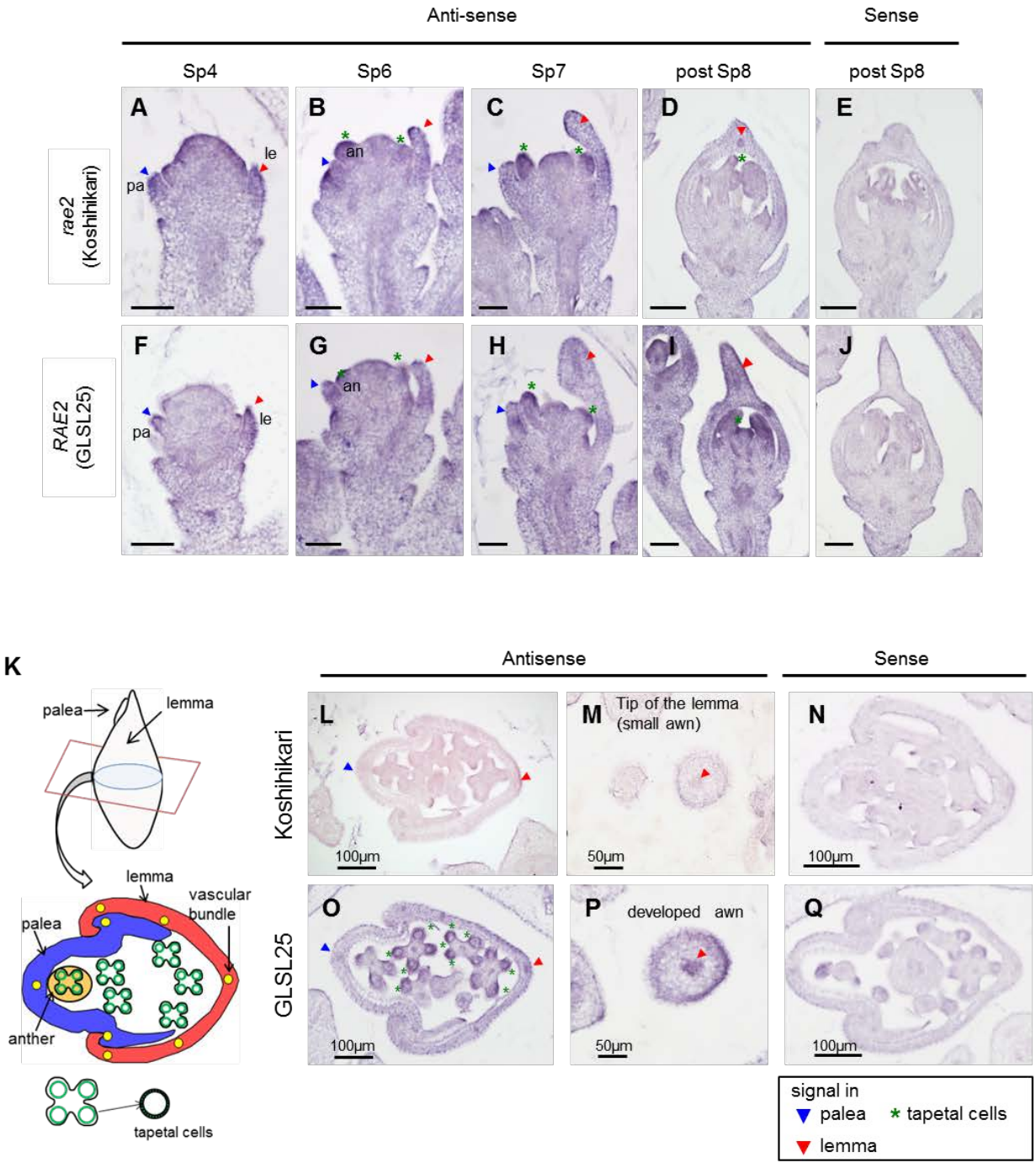
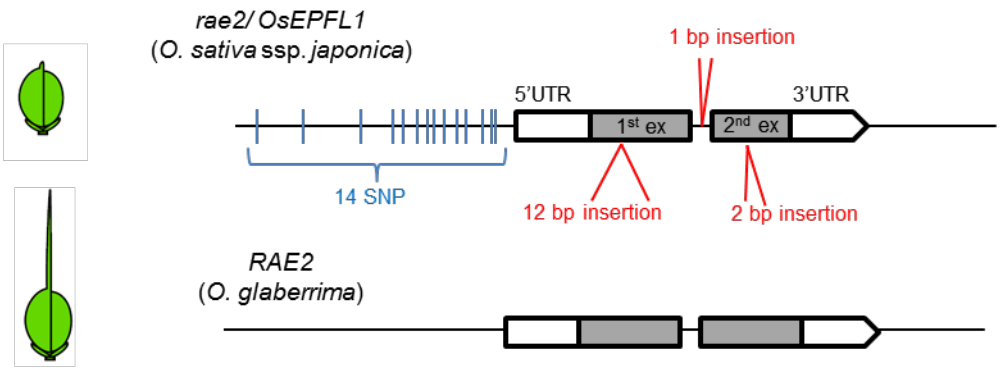


Figure 8. Tissue-specific expression of RAE2 during awn development. (A-J) The longitudinal section of *in situ* hybridization using antisense probes of rae2 (A-D), RAE2 (F-I) and sense probes (E, J) during spikelet development in Koshihikari and GLSL25. Blue arrowhead indicates the tip of palea, red arrowhead indicates the tip of lemma and green asterisks show anther.(K-Q) The cross section of *in situ* hybridization. (K) The cartoon shows one spikelet cross section. Antisense probes of rae2 (L, M), RAE2 (O, P) and sense probes (N, Q). pa= palea, le= lemma, an = anther. Scale bars indicate 50μm (A-C, M, F-H, P), 100μm (D, E, I, J, L, N, O, Q).

A



B

```
rae2/OsEPFL1    1 ATGAGGACGGCGGCCACGCCGCCTCTCGCCGCCGCCGCCGCCGCCGTCGCGGCAGTGTTCCTCTCTGCGT 70
RAE2            1 ATGAGGACGGCGGCCACGCCGCCTCTCGCCGCCGCCGCCGCCGCCGTCGCGGCAGTGTTCCTCTCTGCGT 70

rae2/OsEPFL1   71 TGCTGCTCGCCTCCGCCTCCGCCTCCGCCTCCAGGCTCCCTCCTCCTCGCCGTCTTCTTCCCCTGGTTGG 140
RAE2           71 TGCTGCTCGCCTCCGCCTCC----------AGGCTCCCTCCTCCTCGCCGTCTTCTTCCCCTGGTTGG 128

rae2/OsEPFL1  141 TGGCGAGGTGGCGGTGGCGGTGGTGGCTGGGGAGGAGGAGAAGGTGCGGCTGGGGTCGAGCCCGCCGAGC 210
RAE2          129 TGGCGAGGTGGCGGTGGCGGTGGTGGCTGGGGAGGAGGAGAAGGTGCGGCTGGGGTCGAGCCCGCCGAGC 198

rae2/OsEPFL1  211 TGCTACAGCAAGTGCTACGGGTGCAGCCCGTGCGTCGCGGTGCAGGTGCCCACCTTGTCCGCCCCGTCCG 280
RAE2          199 TGCTACAGCAAGTGCTACGGGTGCAGCCCGTGCGTCGCGGTGCAGGTGCCCACCTTGTCCGCCCCGTCCG 268

rae2/OsEPFL1  281 TCCCCGCCGCCGCCGCGCCGCGCACGACGCCGCGCCGCTCGTGGCGACGTTCACCAACTACAAGCCGCTA 350
RAE2          269 TCCCCGCCGCCGCCGC--CGCGCACGACGCCGCGCCGCTCGTGGCGACGTTCACCAACTACAAGCCGCTA 336

rae2/OsEPFL1  351 G---------------------------------------- 351
RAE2          337 GGGTGGAAGTGCCAGTGCCGCGACCGCCTGTTCGACCCCTGA 378
```

C

```
rae2/OsEPFL1    1 MRTAATPPLAAAAAAVAAVFLSALLLASASASASRLPPPRRLLPLVGGEVAVAVVAGEEEKVRLGSSPPS 70
RAE2            1 MRTAATPPLAAAAAAVAAVFLSALLLASAS----RLPPPRRLLPLVGGEVAVAVVAGEEEKVRLGSSPPS 66

rae2/OsEPFL1   71 CYSKCYGCSPCVAVQVPTLSAPSVPAAAAPRTTPRRSWRRSPTTSR                        116
RAE2           67 CYSKCYGCSPCVAVQVPTLSAPSVPAAAAAHDAAPLVATFTNYKPLGWKCQCRDRLFDP          125
```

D

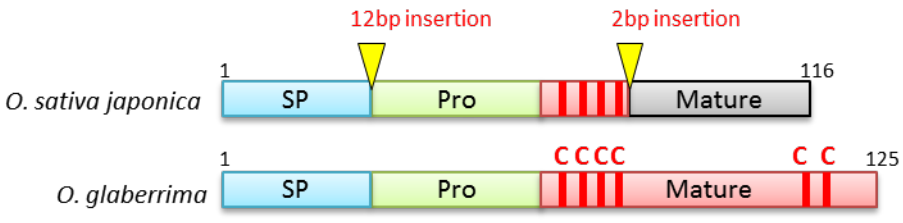
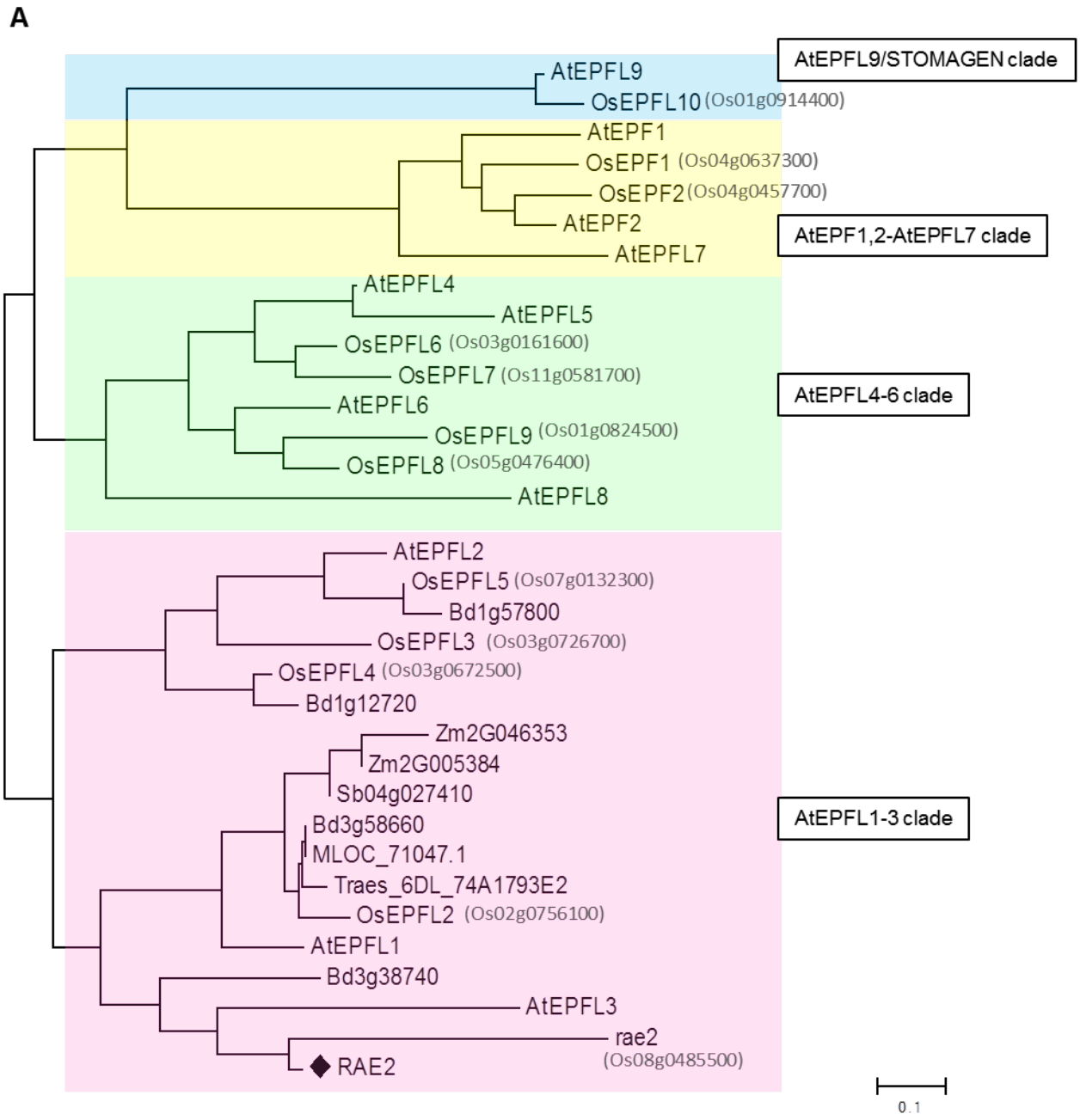


Figure 9. *RAE2* sequence comparison between *O. sativa* ssp. *japonica* cv. Koshihikari and *O. glaberrima* IRGC104038. (A) Schematic image of the *RAE2* and *rae2* gene. Gray-colored boxes in the gene model indicate exonic regions, white boxes indicate UTRs, and the line represents the promoter region, single intronic region and the terminator region of RAE2. Blue lines in the promoter region represent SNPs position and red lines represent insertion in *rae2/OsEPFL1* (*Os08g048550*) of Koshihikari (*O. sativa* ssp. *japonica*) compared with IRGC104038 (*O. glaberrima*). Koshihikari and Nipponbare have the same *rae2* sequence. (B) Comparison of the RAE2 coding sequence in Koshihikari and IRGC104038. Koshihikari has 12 bp insertion in the first exon, and 2 bp insertion in the second exon as represented by the red square. The doubled red line represent GC-rich repeat region correlated with Fig. 13A. (C) Comparison of the RAE2 amino acid sequence between Koshihikari and IRGC104038. The insertion in the second exon causes a frameshift mutation in *rae2*. (D) Schematic image of the rae2/OsEPFL1 amino acid and RAE2. Yellow triangles indicate insertion. Each colored box represents a peptide region. sp=signal peptide (blue), pro=pro-peptide (green), ma=mature peptide (red or gray). Red bar indicates cysteine (C) residues.


A
AtEPFL9/STOMAGEN clade
AtEPFL9
OsEPFL10 (Os01g0914400)
AtEPF1
OsEPF1 (Os04g0637300)
OsEPF2 (Os04g0457700)
AtEPF2
AtEPFL7
AtEPF1,2-AtEPFL7 clade
AtEPFL4
AtEPFL5
OsEPFL6 (Os03g0161600)
OsEPFL7 (Os11g0581700)
AtEPFL6
OsEPFL9 (Os01g0824500)
OsEPFL8 (Os05g0476400)
AtEPFL8
AtEPFL4-6 clade
AtEPFL2
OsEPFL5 (Os07g0132300)
Bd1g57800
OsEPFL3 (Os03g0726700)
OsEPFL4 (Os03g0672500)
Bd1g12720
Zm2G046353
Zm2G005384
Sb04g027410
Bd3g58660
MLOC_71047.1
Traes_6DL_74A1793E2
OsEPFL2 (Os02g0756100)
AtEPFL1
Bd3g38740
AtEPFL3
rae2
(Os08g0485500)
RAE2
AtEPFL1-3 clade
0.1

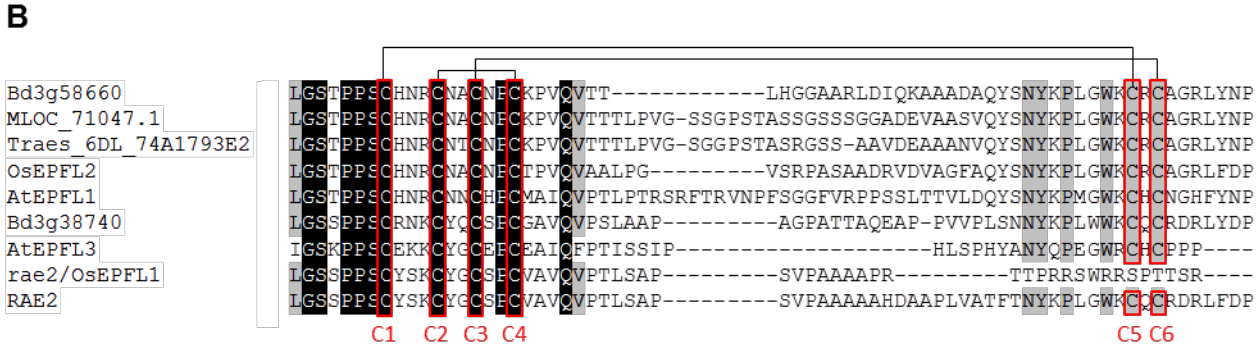

B
Bd3g58660
MLOC_71047.1
Traes_6DL_74A1793E2
OsEPFL2
AtEPFL1
Bd3g38740
AtEPFL3
rae2/OsEPFL1
RAE2
C1
C2
C3
C4
C5
C6

Figure 10. Phylogenetic tree of EPF/EPFL family genes and comparison of the sequences in mature peptide region. (A) Neighbor-joining phylogenetic tree of EPF/EPFL genes. Amino acid sequences for the predicted mature peptide region are aligned using the ClustalW program. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown next to the branches (values 50% or greater are shown). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. Evolutionary analyses are conducted in MEGA6. At: *Arabidopsis thaliana*, Os: *Oryza sativa*, Zm: *Zea mayz*, Sb: *Sorghum* bicolor, Bd: *Brachypodium distachyon*, Traes: *Triticum aestivum*. MLOC: *Hordeum vulgare*'s transcript. Each clade is consistent with previous report (Takata et al. 2013). Gray colored character indicates RAP-DB ID of *O. sativa* EPF/EPFLs. (B) Alignment of RAE2 predicted mature peptide amino acid sequences with the half member of AtEPFL1-3 clade. Pairs of cysteine residues forming disulfide bonds predicted for *A. thaliana* EPF/EPFL genes are connected by lines. Two cysteine (C5 and C6) deletions in C-terminal region could be seen only in rae2/OsEPFL1.

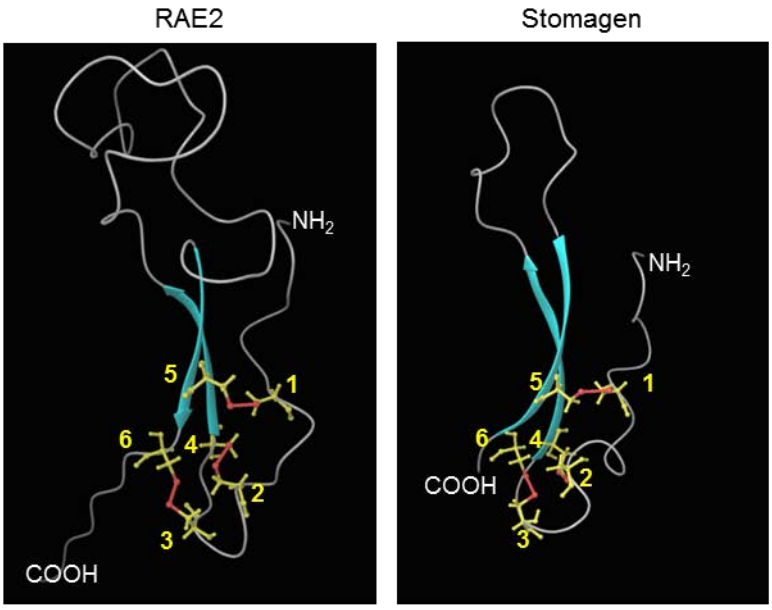


Figure 11. Predicted 3D structure of RAE2 and Stomagen. Hypothetical 3-D structure of RAE2 based on the model of Stomagen by Mäestro (Schrödinger, NY, USA) software. Yellow: cysteine residues, red: disulfide bonds and atoms, blue: antiparallel beta sheets. Yellow numbers indicate locations of the cysteine residues.

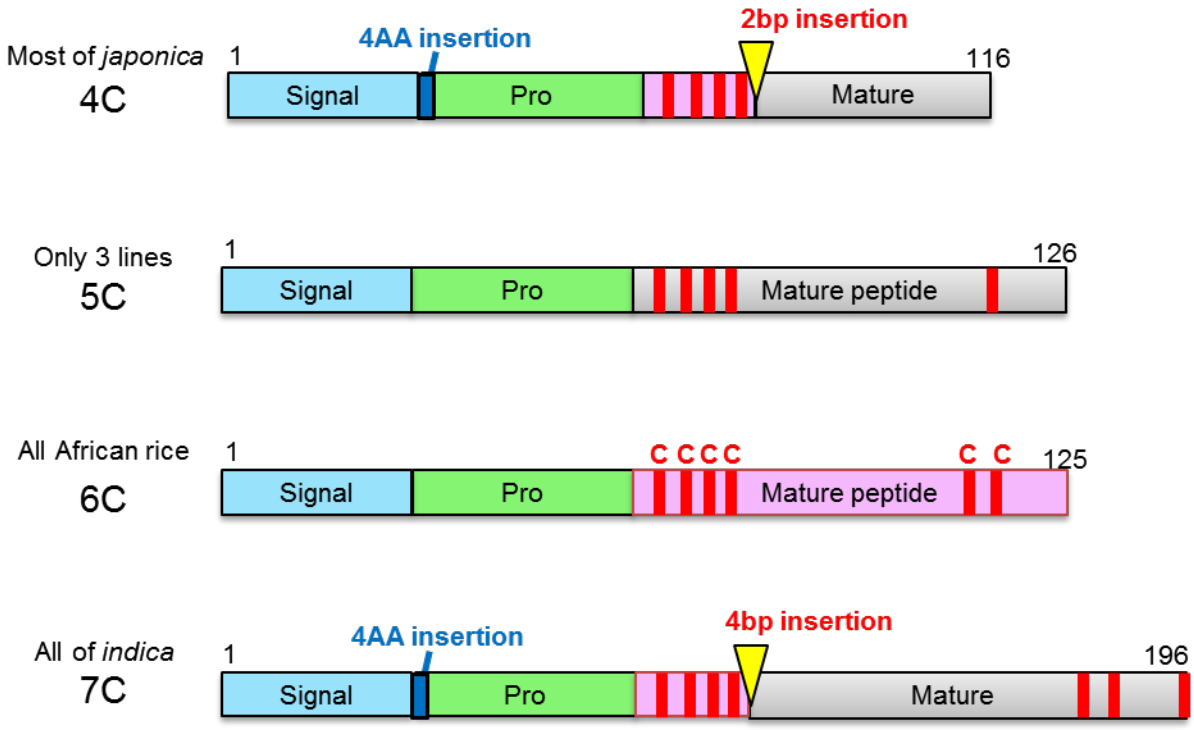
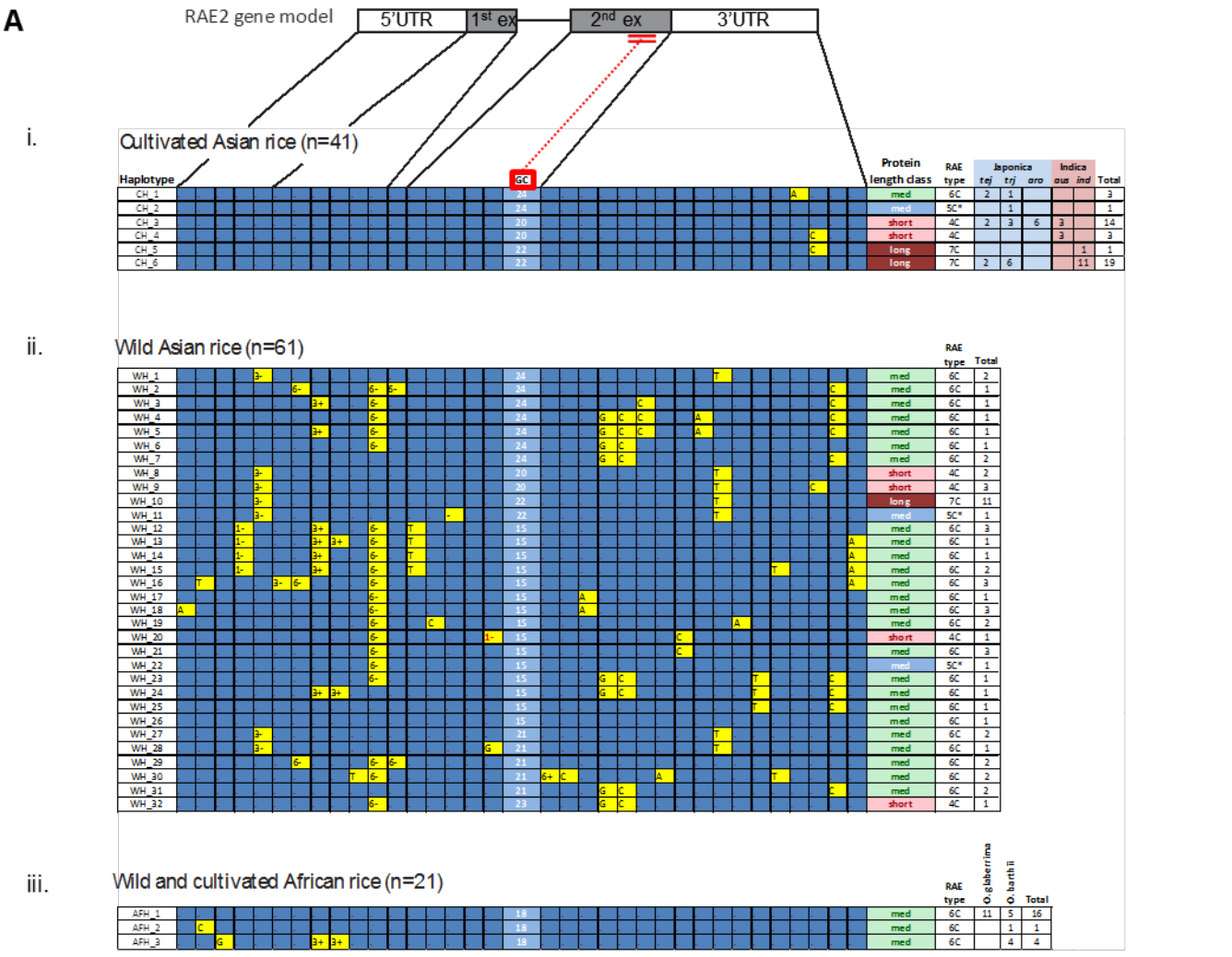
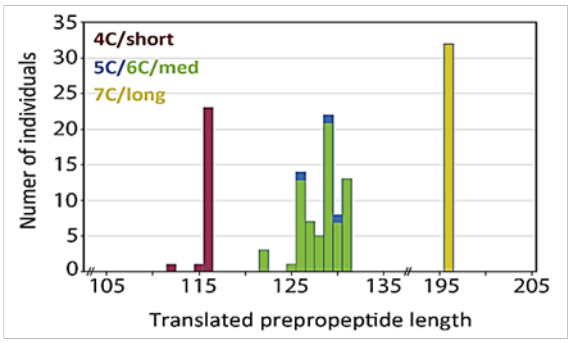


Figure 12. Four types of RAE2 classified based on cysteine number. Pale blue, green, pink and gray squares represent signal peptide region, pro-peptide region and mature peptide conserving 6 cysteine residues and mature peptide having other number of cysteine residues respectively. Red bar show cysteine residue position. Blue square and yellow triangle represent amino acid insertion and nucleotide insertion. The numbers on the C-terminal mean the representative amino acid length.


A
RAE2 gene model
5'UTR
1st ex
2nd ex
3'UTR
i.
Cultivated Asian rice (n=41)
ii.
Wild Asian rice (n=61)
iii.
Wild and cultivated African rice (n=21)


B
4C/short
5C/6C/med
7C/long
Numer of individuals
Translated prepropeptide length

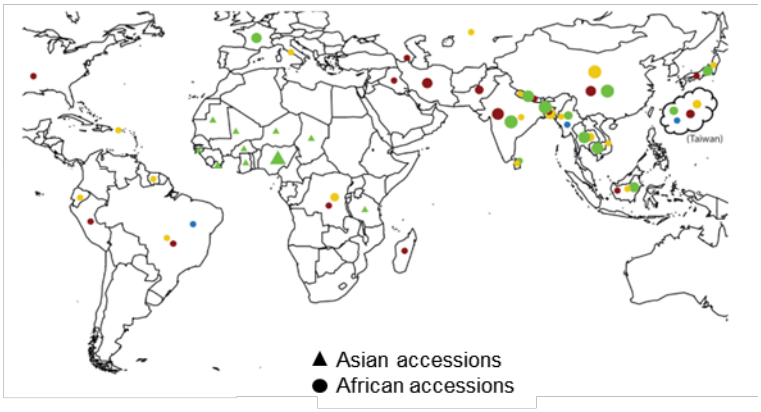

C
▲ Asian accessions
● African accessions

Figure 13. Distribution of RAE2 protein variants across diverse rice accessions. (A) RAE2 gene haplotypes across diverse Asian and African rice accessions. Polymorphic sites discovered within the coding and non-coding regions of RAE2 of 130 diverse rice accessions are shown in Supplementary Table 7. (i) Cultivated Asian rice, (ii) Wild Asian rice, (iii) Wild and cultivated African rice. Individual number in each population used for this analysis was noted in brackets (n=xx). Gray-colored boxes in the gene model shown at the top of the Fig. indicate exonic regions, white boxes indicate UTRs, and the line represents the single intronic region of RAE2. 'CH' indicates gene haplotypes present in cultivated Asian rice, 'WH' indicates gene haplotypes found in wild Asian rice while 'AFH' indicates gene haplotypes found in cultivated and wild African rice. 'RAE type' shows the number of cysteine residues predicted from translating the cDNA sequence. The value in column 'GC' represents the number of nucleotides within a highly variable GC-rich region of the RAE2 second exon (indicated in double red line as same as shown in Fig. 10B). At all other sites, blue cells represent alleles that match the reference genome (cv. Nipponbare) while yellow cells are non-reference alleles. The numbers of accessions that harbored each haplotype are indicated in the right-hand table (tej = temperate japonica, trj = tropical japonica, aro = aromatic, ind = indica, aus = aus, determined by HDRA genome-wide information). (B) Distribution of translated propeptide lengths across 107 Asian rice accessions and 23 African rice accessions (green=6C/medium, blue=5C/medium, red=4C/short, yellow=7C/long). (C) Geographical distribution of RAE2 protein variants found across these same 130 accessions (107 Asian rice (triangle) and 23 African rice accessions (circle)). The size of the marks suggests population size. Dark red sectors represent the 4C/short, blue represents the 5C/medium, green represents the 6C/medium, and yellow represents the 7C/long RAE2 variant.

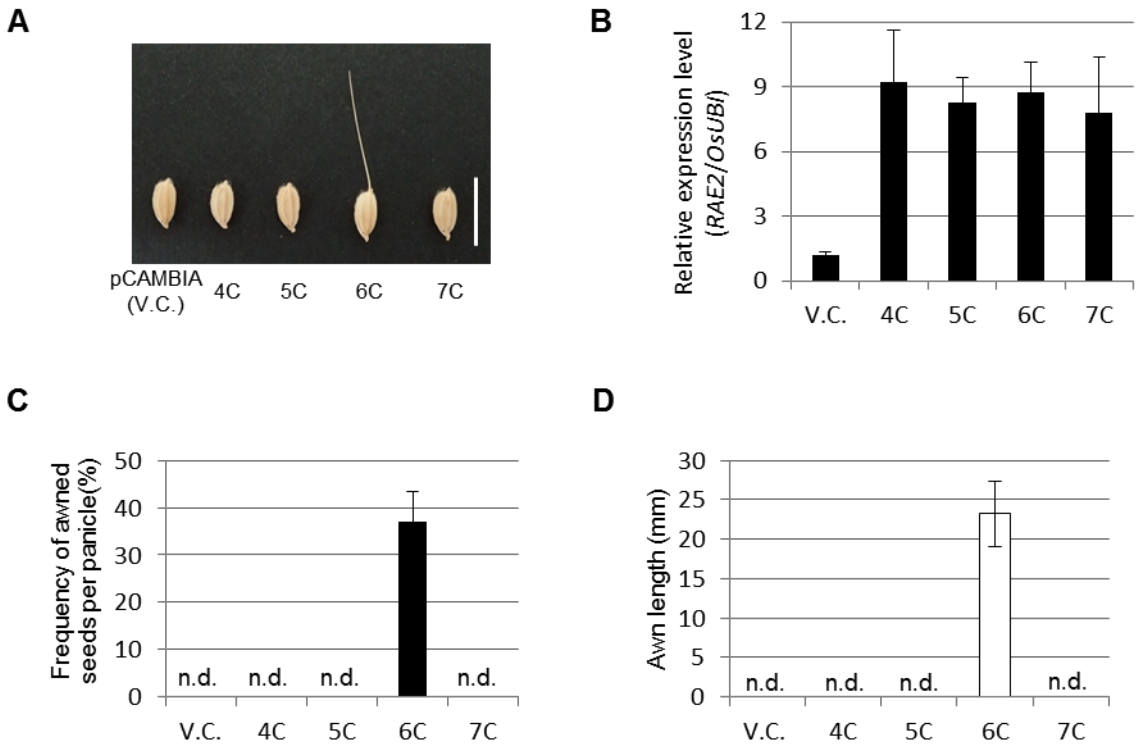


Figure 14. Different types of RAE2 and definition of each function for awn elongation. (A) The awn phenotype of overexpression lines of each RAE2 type (4C, 5C, 6C, 7C). pCAMBIA1380 was used as vector control (V.C.). Scale bar represents 1 cm. (B) Relative expression levels of RAE2 in young panicles of transgenic lines of overexpression construct; 4C, 5C, 6C, 7C. OsUBI used as housekeeping gene. (B) Frequency of awned seeds per panicle, (C) awn length in each overexprssion lines. n.d.=not detected. Error bars represent standard deviation of the mean.

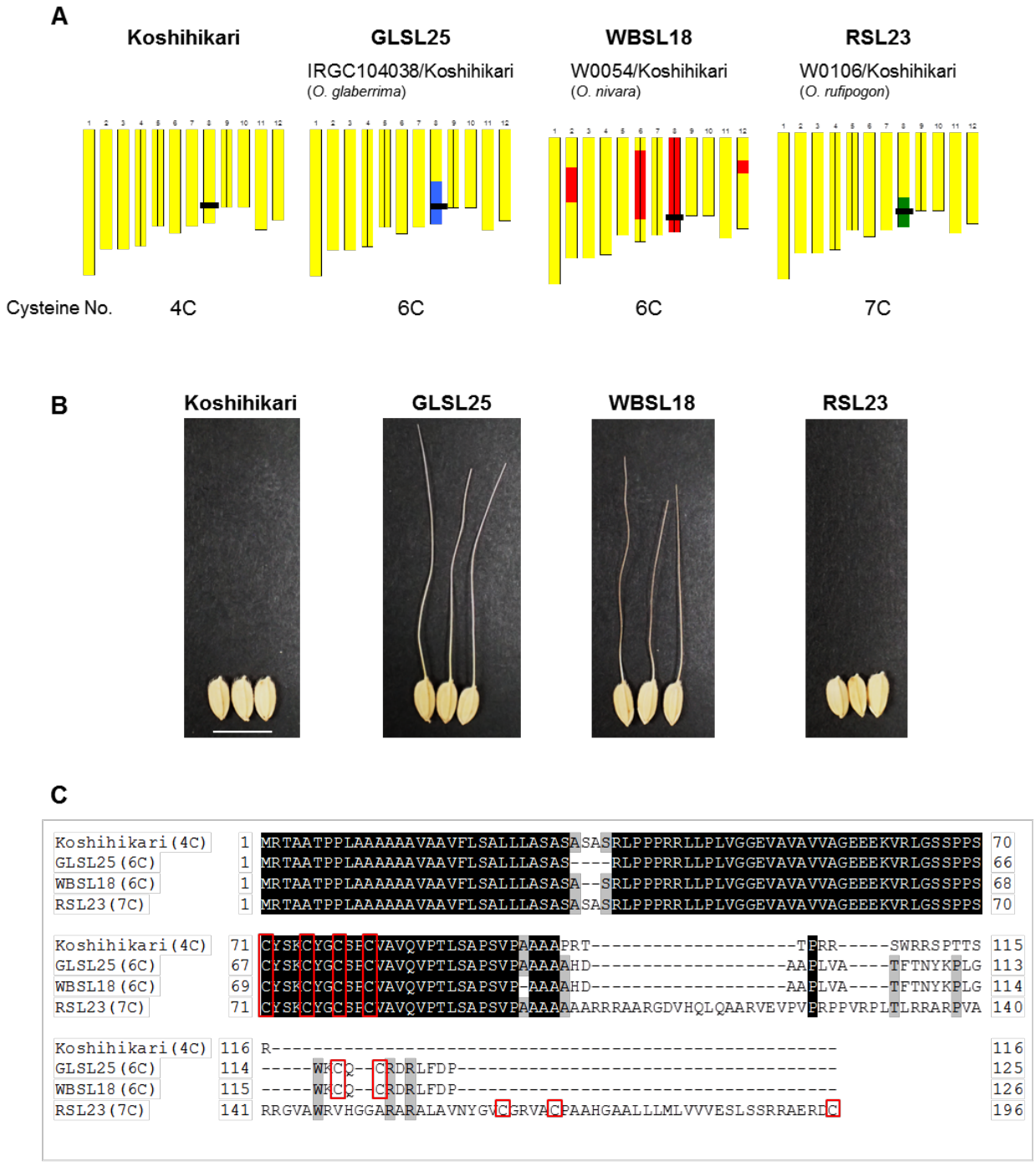


Figure 15. CSSLs awn phenotype and the number of RAE2 cysteine residues. (A) Graphical genotypes of Koshihikari and each CSSL. Yellow indicates the recurrent parent background, Koshihikari. The other colors indicate chromosome introgression from donor parents: blue=IRGC104038 (*O. glaberrima*), red=W0054 (*O. nivara*), green=W0106 (*O. rufipogon*). Black bar suggests RAE2 position. Number of cysteine residue of RAE2 set down below the each graphical genotype. (B) Seed phenotype of Koshihikari and each CSSL. Scale bar represented 1 cm. (C) The comparison of 4 types of RAE2 amino acid sequence. Red square indicates cysteine residues.

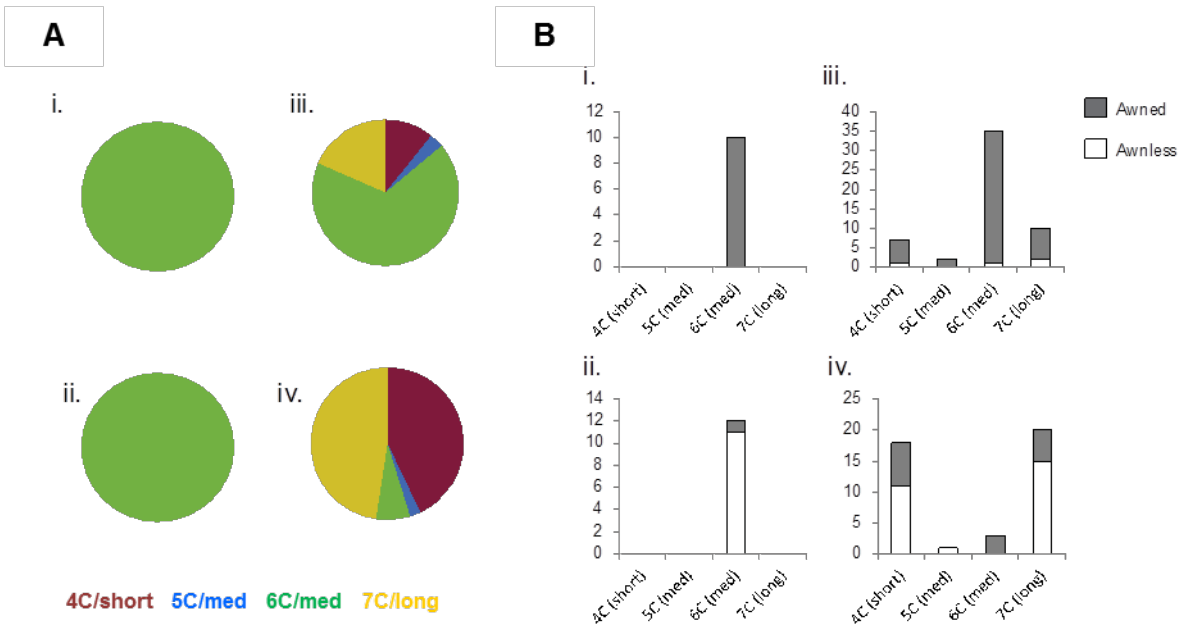


Figure 16. Diversity analysis of RAE2 across Asian and African rice. (A) Distribution of four RAE2 protein variants within *O. barthii* (i, n=11), *O. glaberimma* (ii, n=12), *O. rufipogon/O. nivara* (iii, n=65), and *O. sativa* (iv, n=42). The color is same as described in Fig. 13B. (B) Awn phenotype across four RAE2 protein variants. Numbers of awned (gray bars) and awnless (white bars) accessions for *O. barthii* (i), *O. glaberimma* (ii), *O. rufipogon/O. nivara* (iii), and *O. sativa* (iv).

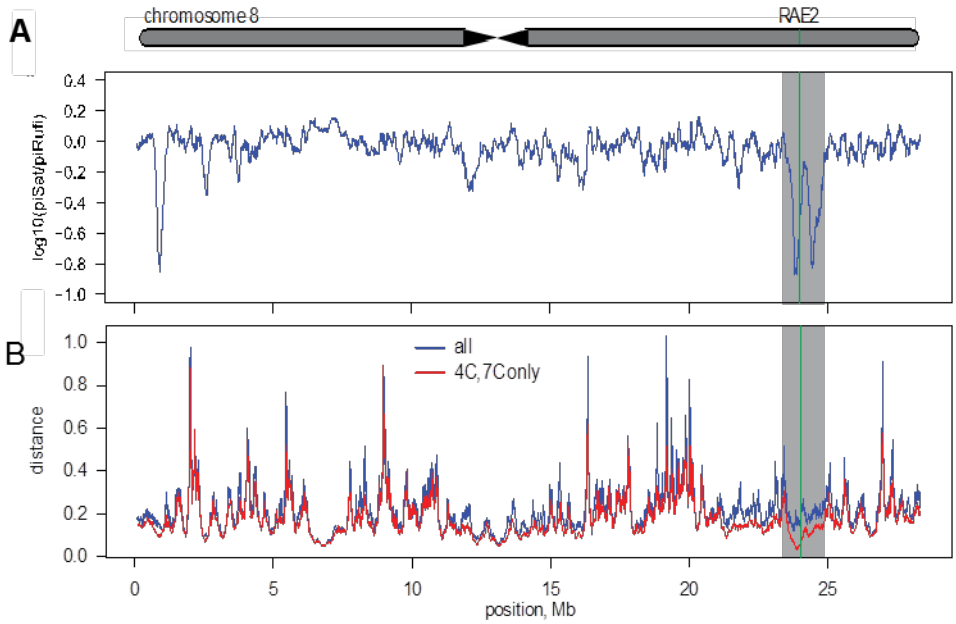


Figure 17. The decrease of the polymorphism in RAE2 region. (A) Nucleotide diversity of *O. sativa* individuals (n= 67) relative to nucleotide diversity of *O. rufipogon/O. nivara* individuals (n= 65) across chromosome 8. Gray box represented reducing relative diversity surrounding RAE2 (green line) consistent with a selective sweep. (B) Genetic distance between *O. sativa* and *O. rufipogon/O. nivara* for all RAE2 types (blue) and dysfunctional ones (red: including 4C and 7C type of RAE2). Decreased distance in the dysfunctional class relative to distance in 'all' class in a 1.5 Mb region surrounding RAE2 (gray box).

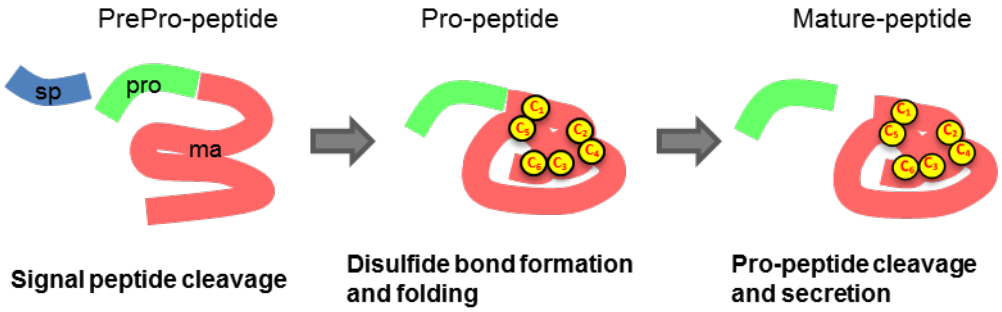


Figure 18. Cleavage process of signal peptide. The model of secretory peptide cleavage. EPFL family peptides are composed of a signal peptide region (sp, blue), pro-peptide region (pro, green) and mature peptide region (ma, pink). After transcription, the whole peptide called pre-pro-peptide, is cleaved at the border of the signal peptide sequence and then forms the pro-peptide. Disulfide bond formation and folding occurs in the endoplasmic reticulum before the pro-peptide is secreted into the extracellular matrix. Further cleavage occurs in the border between the pro-peptide and mature peptide region.

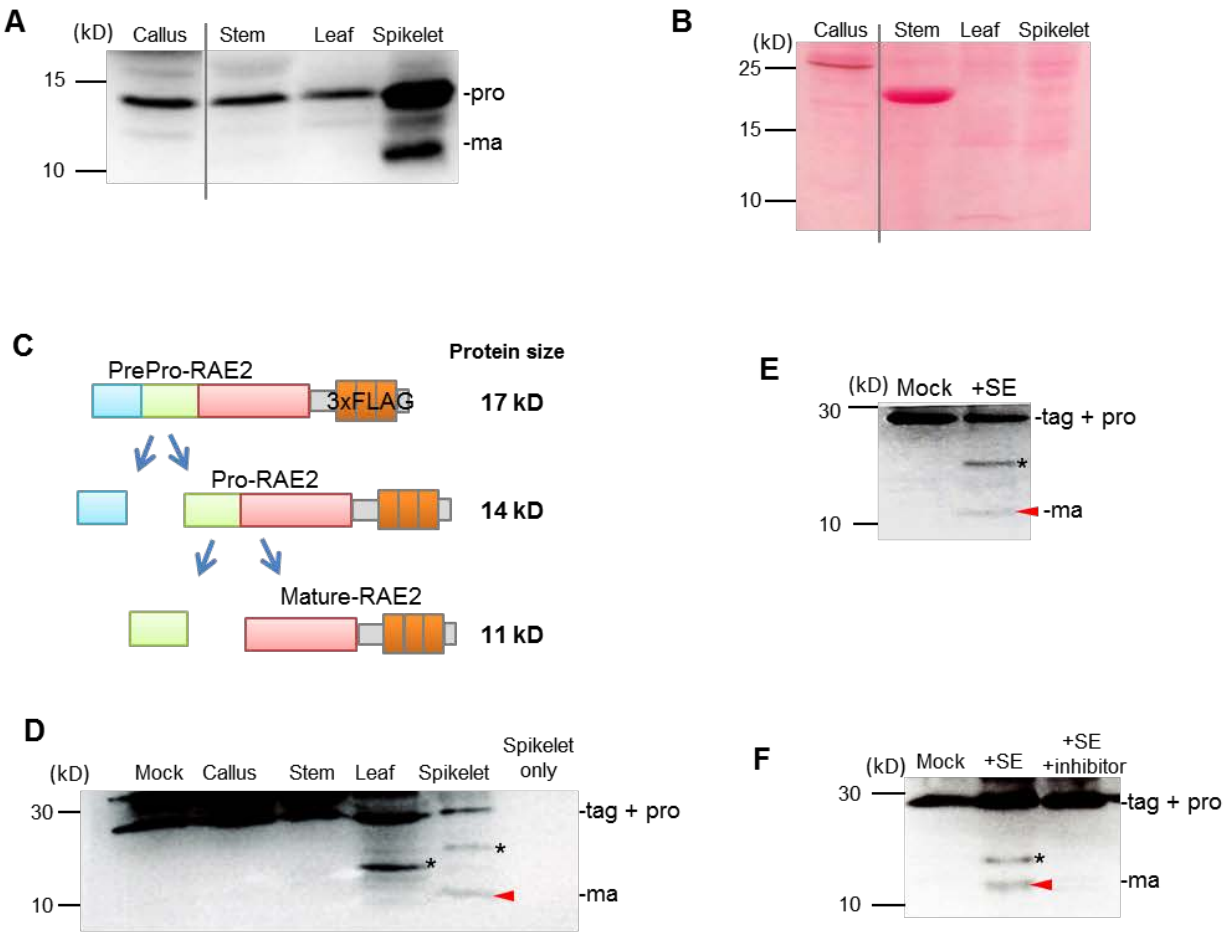
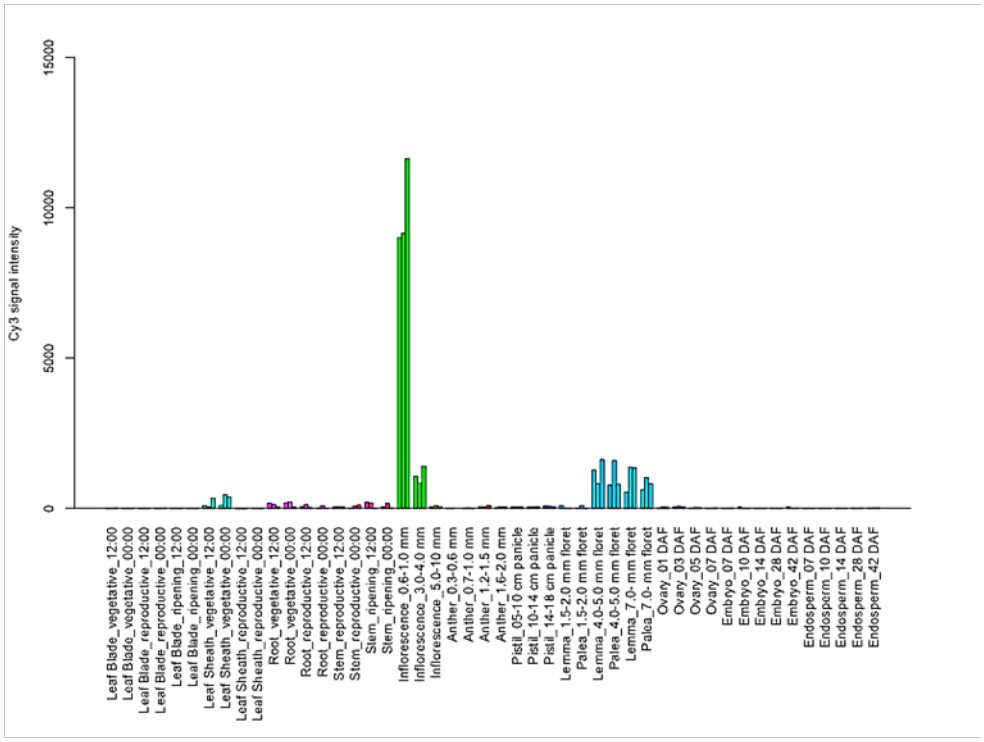
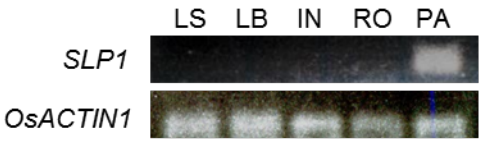


Figure 19. The RAE2 maturation process occurs specifically in the spikelet. (A) Immunoblot analysis of RAE2-3xFLAG in transgenic plant with anti-FLAG antibody. The gray line indicates erased space between callus and stem lane although all samples were applied on same membrane. (B) The membrane stained by ponceau S (Wako, Japan) as loading control of Fig. 19A. Total protein amount is 30 µg in each lane. (C) The expected size of RAE2-3xFLAG peptide after cleavage in the transgenic plant of overexpression construct: signal peptide (blue), pro-peptide (green) and mature peptide (pink). (D) *in vitro* processing assay of recombinant RAE2 peptide incubated with plant extracts of Koshihikari or buffer (mock). The ~30kD band is a tag-fused recombinant RAE2 pro-peptide (indicated by –tag+pro). Asterisk represents non-specific band, red arrowhead represents the expected mature RAE2-3xFLAG peptide (~11 kD, indicated by –ma) detected by anti FLAG-antibody. (E) *in vitro* processing assay of recombinant RAE2 pro-peptide fused with 3xFLAG tag incubated with plant extracts of Koshihikari spikelet or buffer (mock). Fusion peptide and cleaved peptides were detected by anti-RAE2 antibody. (F) *in vitro* processing assay with or without protease inhibitor cocktail, Complete (Roche, Basle). SE= spikelet extract.

A



B

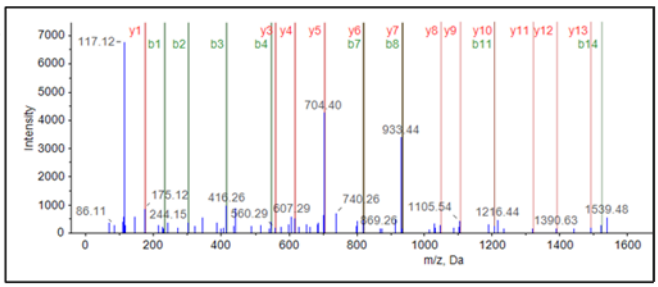


C

i.

Residue	b	y
S	232.14	2139.13
A	303.18	1908.00
I	416.26	1836.96
M	547.30	1723.88
T	648.35	1592.84
T	749.40	1491.79
A	820.44	1390.74
I	933.52	1319.70
T	1034.57	1206.62
G	1091.59	1105.57
D	1206.62	1048.55
N	1320.66	933.52
D	1435.69	819.48
S	1522.72	704.45
G	1579.74	617.42
K[IT4]	1851.94	560.40
I	1965.02	288.20
R	2121.12	175.12

ii.



iii.

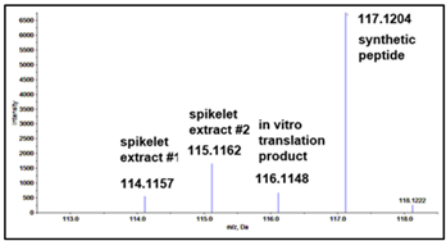


Figure 20. Specific expression pattern of SLP1. (A) Gene expression pattern of SLP1 drawn in RiceXpro. (B) semi-quantitative PCR result of SLP1 by primers KU144 and KU146 (Table 1). The template which is extracted from Koshihikari is the same as the one used in qRT-PCR in Fig. 7A. LS= leaf sheath, LB= leaf blade, IN= internode, RO= root, PA= young panicle. I used *OsACTIN1* as housekeeping gene. (C) The ion spectrum of SLP1 peptide: 543-SAIMTTAITGDNDSGKIR-560 identified after iTRAQ labeling and LC-MALDI MS/MS analysis using ProteinPilot software 5.0 AB SCIEX. (i) Identified peptide using detected b- and y- ion series, (ii) MS/MS fragmentation spectra, (iii)Quantitation of iTRAQ report ions: The peak of #1 and #2 were extracted from Koshihikari spikelets labeled with 114, 115 iTRAQ reagents respectively. Tryptic peptides from in vitro-translation SLP1, and synthetic SLP1 peptide were labeled with 116, 117 iTRAQ reagents respectively.

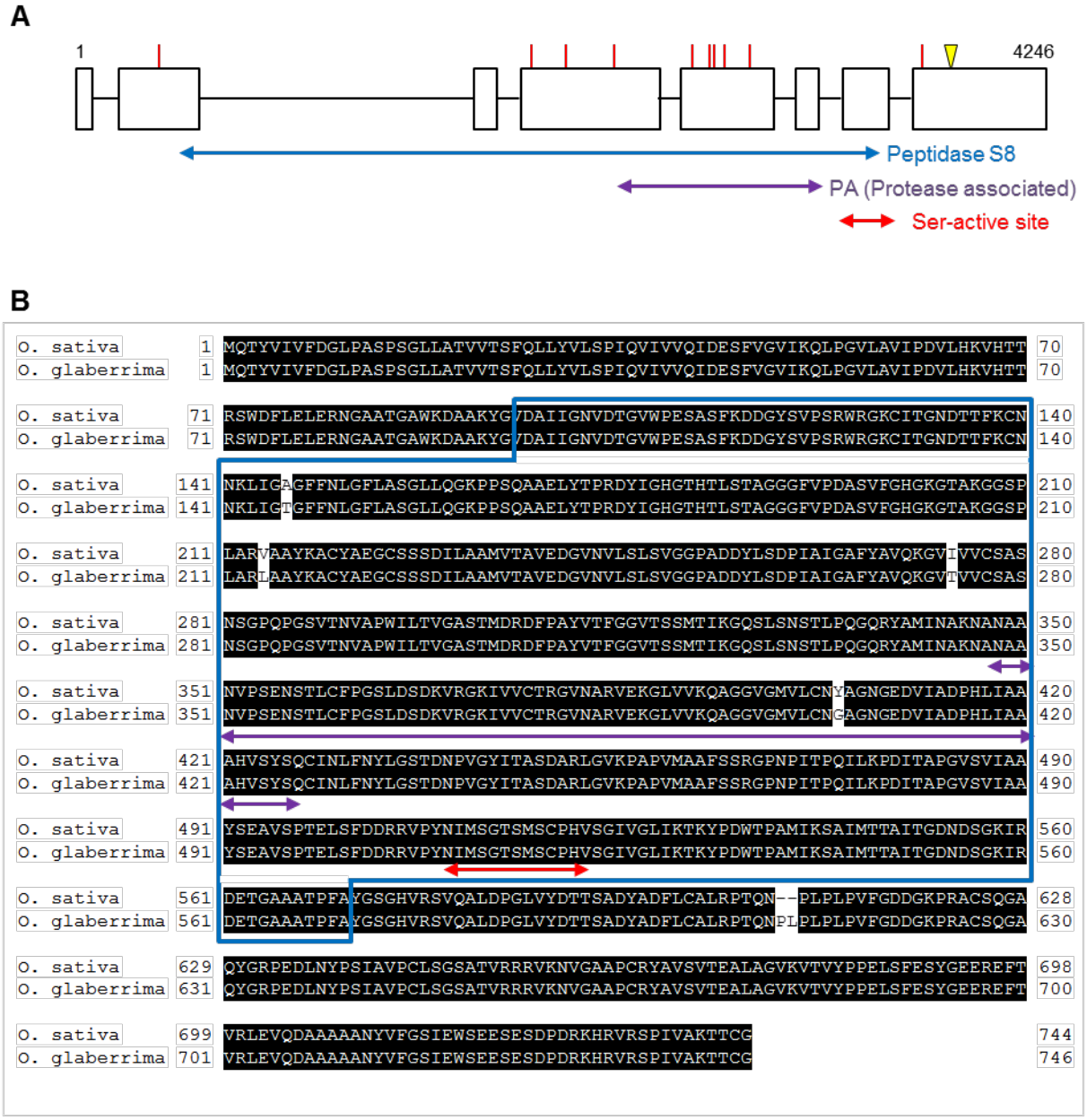


Figure 21. Sequence and amino acid structure of SLP1. (A) Schematic image of SLP1. White-colored boxes in the gene model indicate exonic regions, and the line represents the intronic regions of SLP1. Red bars represent SNPs position and yellow triangle represenst insertion in CG14 (*O. glaberrima*) compared with Nipponbare (*O. sativa* ssp. *japonica*).Blue arrow represents the peptidase S8 domain, purple arrow represents the protease-associated domain, red arrow represents the serine-active site. (B) Sequence comparison of SLP1 amino acid in Nipponbare and CG14. Blue square shows the peptidase S8 domain, the purple and red arrows represent the protease-associated domain and the serine-active site, respectively.

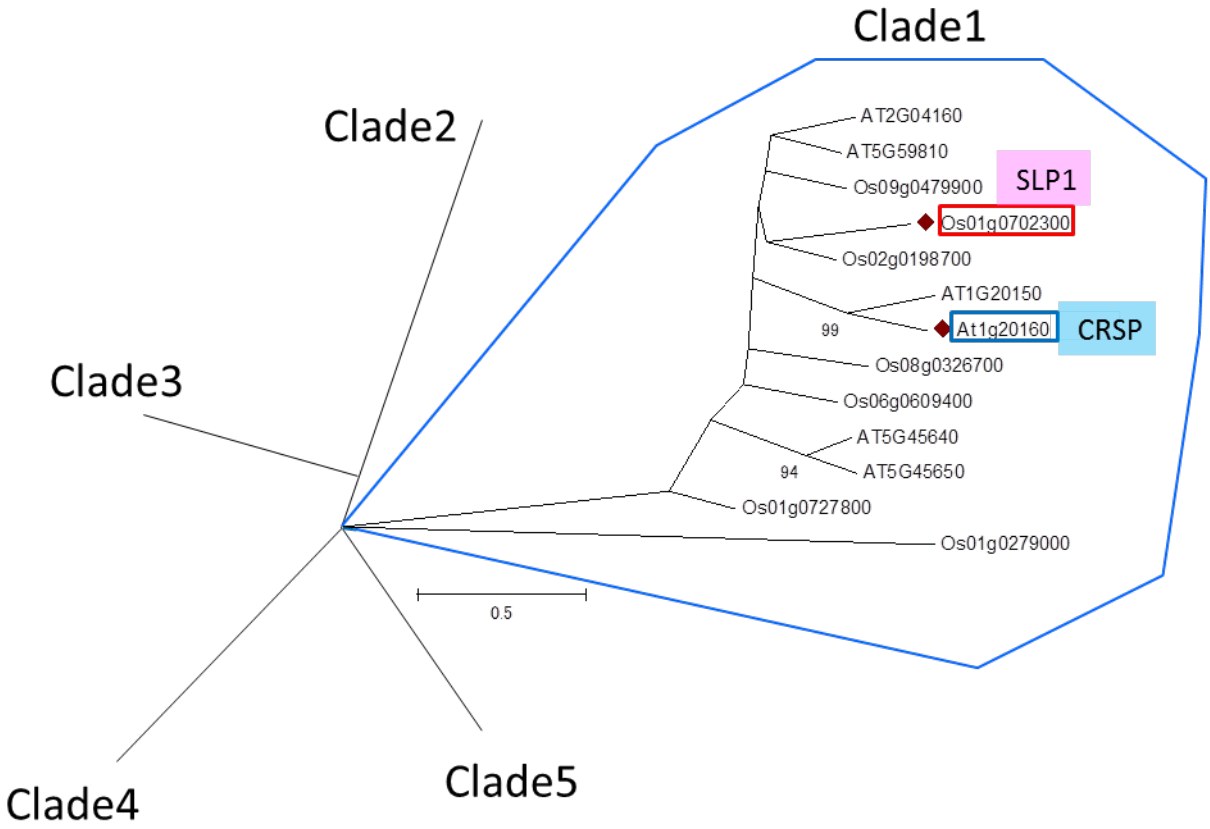


Figure 22. Phylogenetic tree of subtilisin-like protease family genes in *A. thariana* and *O. sativa*. Unrooted Neighbor-joining phylogenetic tree of subtilisin-like protease family genes computed from multiple sequence alignments of *A. thaliana* and *O. sativa* subtilisin domains. Subtilisin-like protease domains were aligned using ClustalW program and the alignments were exported to MEGA6 to generate the Neighbor-Joining tree. Bootstrap values from 100 replications are indicated at branch nodes (values 50% or greater are shown). This tree is modified the phylogenic tree which is reported previously (Tripathi & Sowdhamini 2006).

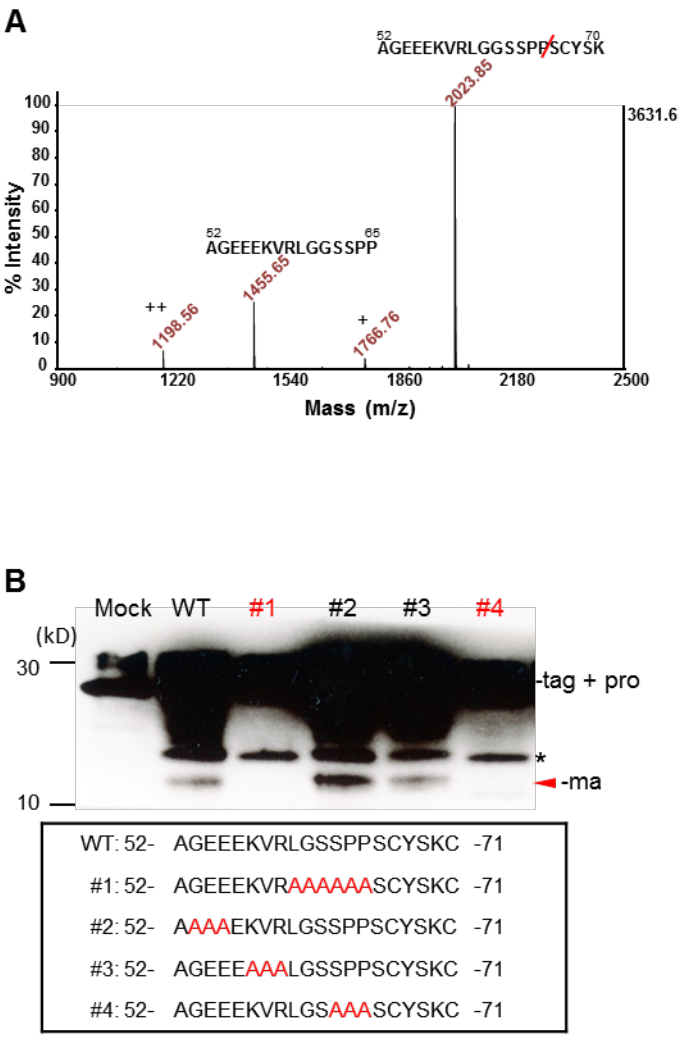
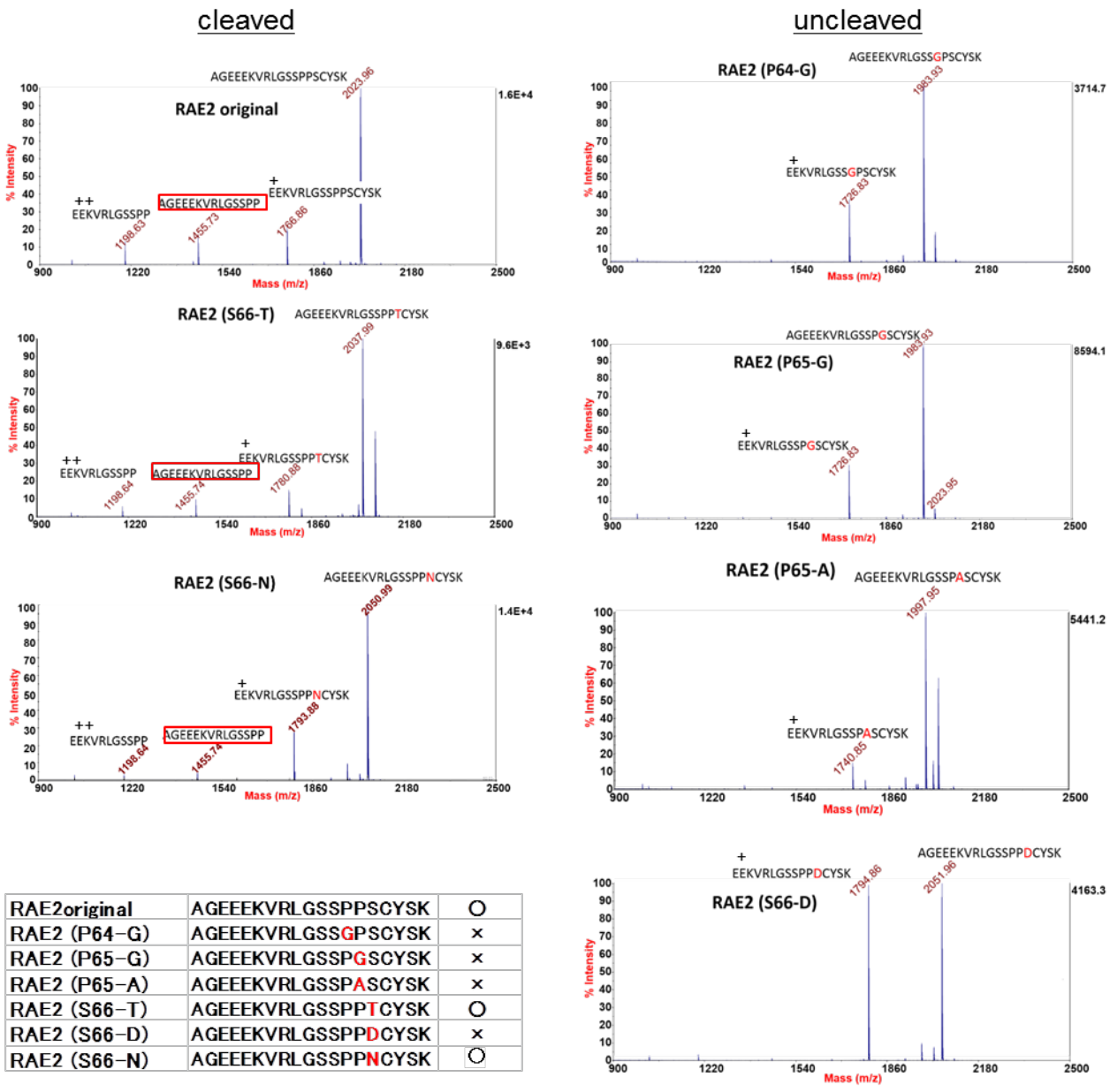


Figure 23. RAE2 maturation caused by cleavage with SLP1 protease at spikelet. (A) The MS ion spectrum for the synthetic peptide (52-AGEEEKVRLGSSPPSCYSK-70) which was cleaved at the position between P65 and S66 indicated by red line. The full length of synthetic peptide was cleaved the other site (+ = 54-EEEKVRLGSSPPSCYSK-70, ++ = 54-EEEKVRLGSSPP-65). (B) *in vitro* processing assay of a series of alanine substituted recombinant RAE2 peptide using spikelet extract of Koshihikari or buffer (mock). Table below shows the amino acid sequence around predicted cleavage point.

A



RAE2original	AGEEEKVRLGSSPPSCYSK	○
RAE2 (P64-G)	AGEEEKVRLGSSGPSCYSK	×
RAE2 (P65-G)	AGEEEKVRLGSSPGSCYSK	×
RAE2 (P65-A)	AGEEEKVRLGSSPASCYSK	×
RAE2 (S66-T)	AGEEEKVRLGSSPPTCYSK	○
RAE2 (S66-D)	AGEEEKVRLGSSPPDCYSK	×
RAE2 (S66-N)	AGEEEKVRLGSSPPNCYSK	○

B

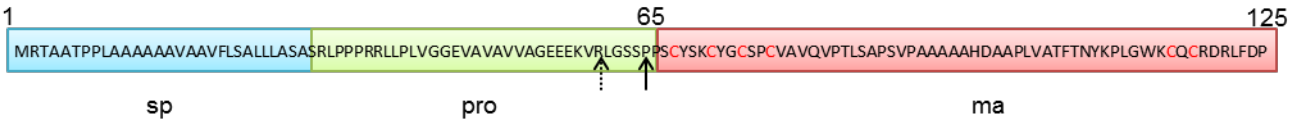


Figure 24. Detection of the cleavage site of RAE2 by *in vitro* processing assay. (A) Mass spectrometry of *in vitro* processing reactions of the series of amino acid substituted synthetic peptides of RAE2 (synRAE2) incubated with *in vitro*-synthesized SLP1. Left lane (entitled cleaved) and right lane (entitled uncleaved) showed that mu-synRAE2 which could or could not be cleaved by SLP1 respectively. Red tangle indicates the fragment cleaved between P65 and S66. The table in lower left shows the summary of the result; o and x are consistent with cleaved and uncleaved. There are two peaks which we infer the reason described in Fig. 25A (+ = 54-EEEKVRLGSSPPSCYSK-70, ++ = 54-EEEKVRLGSSPP-65. (B) Predicted sequence of RAE2 which encodes a 125 amino-acid peptide composed of a signal peptide (sp, blue), a pro-peptide (pro, green) and mature peptide (ma, pink). Dotted arrow and solid arrow indicates the cleavage site of Stomagen and EPF2 respectively.

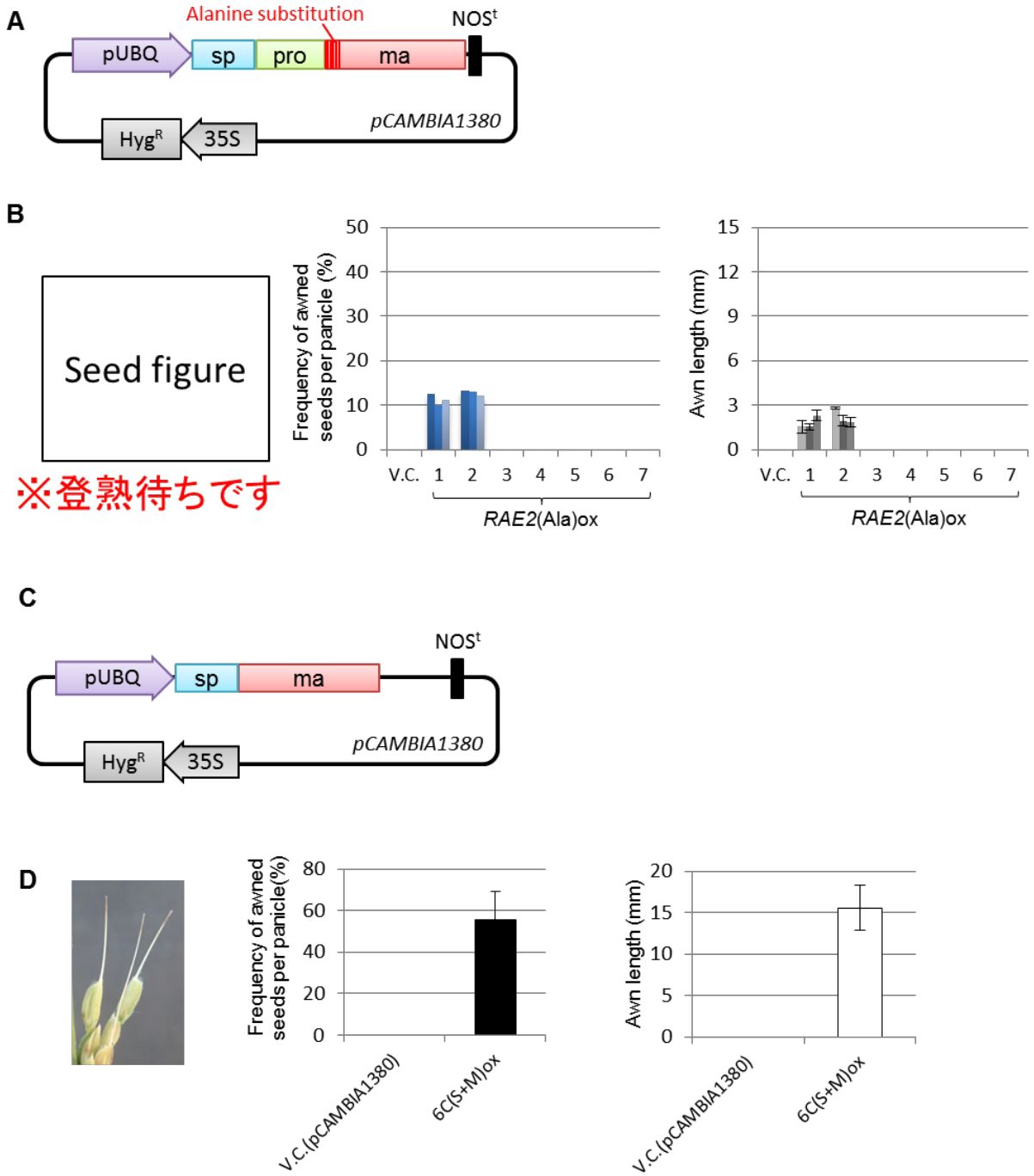


Figure 25. Detection of the cleavage site of RAE2 by *in vivo* processing assay. (A) The construction map of alanine substituted *RAE2* gene. Red bar represent the alanine substitution position between pro- and mature-peptide. (B) Evaluation of transgenic plants: seed phenotype, frequency of awned seeds per panicle, and awn length. These graphs show individual result of each line. (C) The construction map of *RAE2* gene without pro-peptide region. (D) Evaluation of transgenic plants: seed phenotype, frequency of awned seeds per panicle, and awn length. No visible awn was observed in pCAMBIA1380 (V.C.) indicated as n.d. (not detected). Bar length represents 1 cm. Values represent mean ± SE (n = 4). **P < 0.01 based on two-tailed Student's t-test.(B) (C) (D)

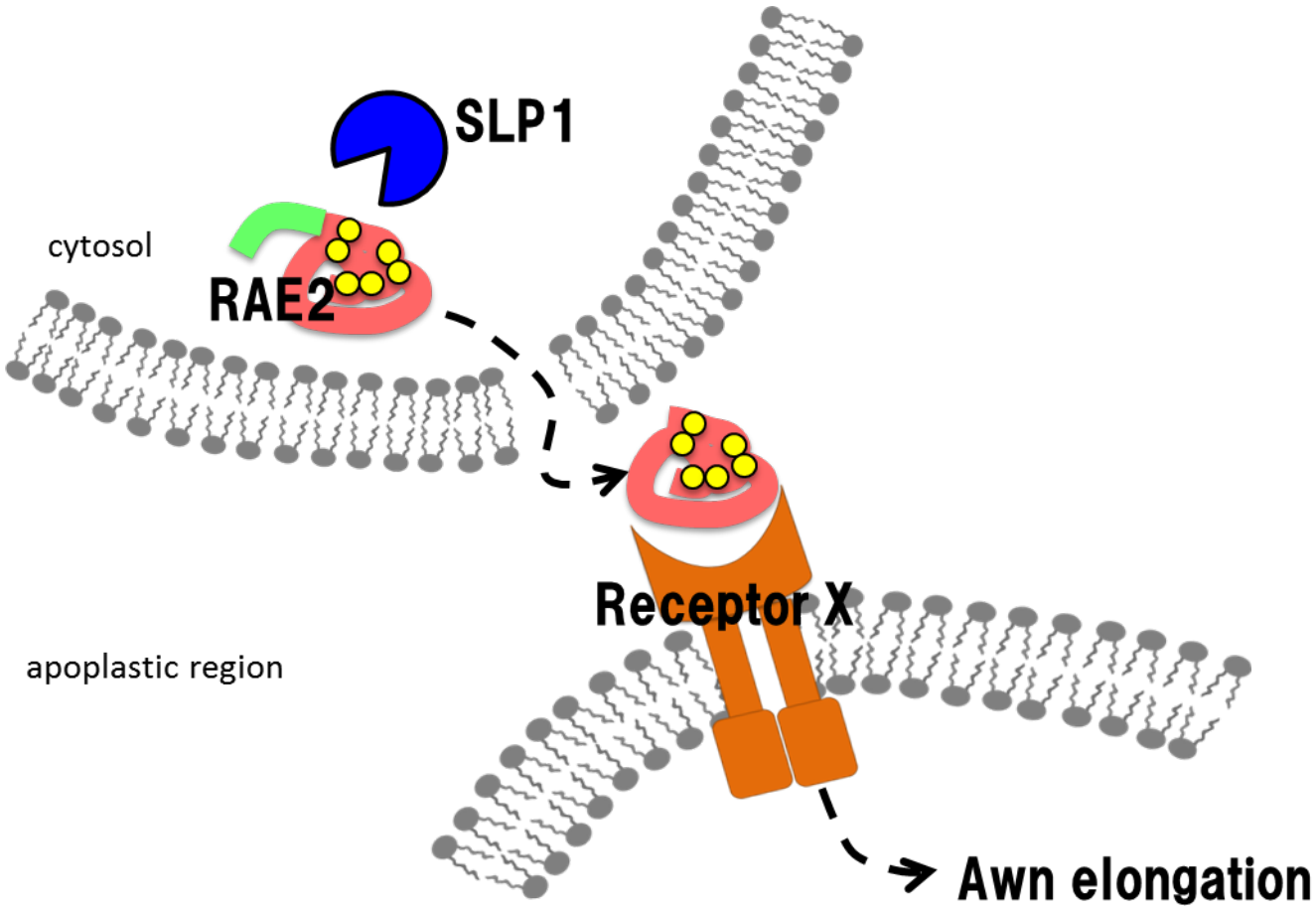


Figure 26. Schematic image of RAE2 peptide behavior. RAE2 was cleaved to mature peptide (pink colored part) by SLP1 in cytosol. Then RAE2 might be secreted to apoplastic region and recepted by receptor X.RAE2 reception induced the downstream signal transductions and promote awn elongation in rice.

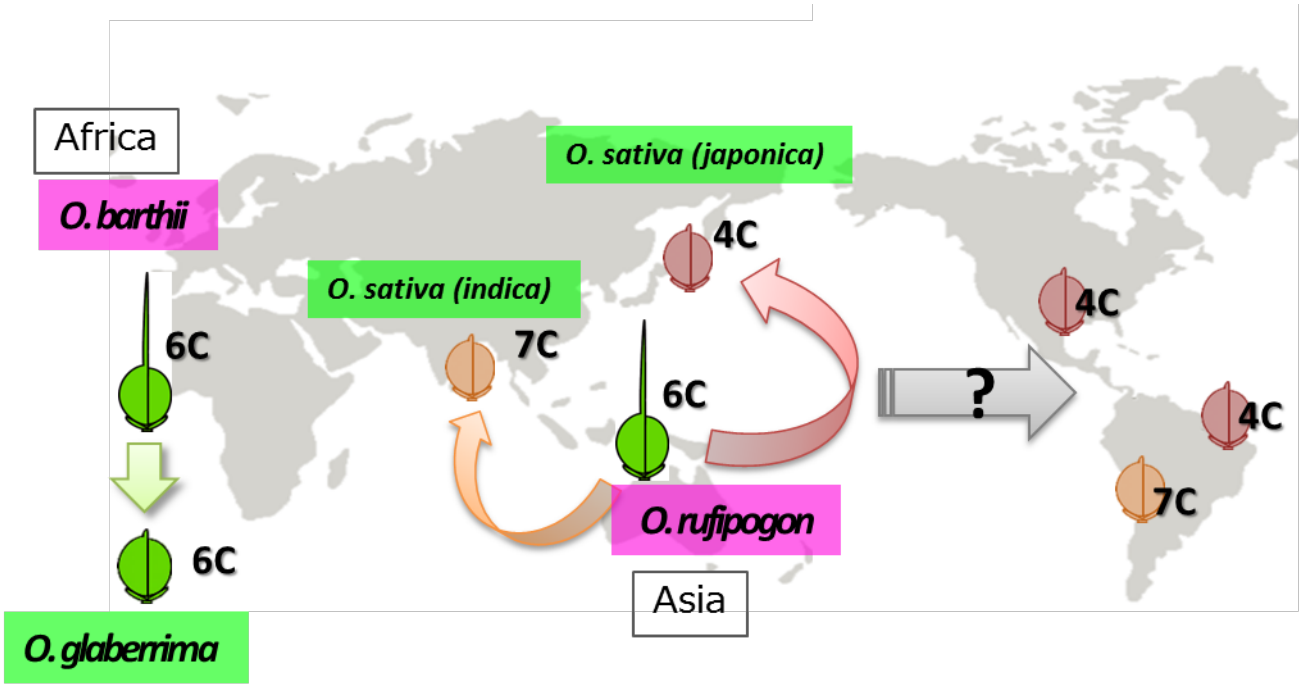


Figure 27. The model of RAE2 selection and speciation all over the world. The predicted domestication process of awnlessness in African rice and Asian rice. Pink and green squares represent the wild and cultivated rice species respectively. Green, orange and red arrows showed the RAE2 selection occurred in each area. Gray arrow suggested the speciation process of rice that Asian rice might transfer into North and South America.

Acknowledgements

This research was supported by the Japan Society for the Promotion of Science (JSPS) fellowship, the Integrative Graduate Education and Research (IGER) program in Green Natural Sciences of Nagoya University. I thank Dr. H. Tsuji for donating the pANDA vector; Dr. T. Nakagawa for providing the pGWB binary vectors; Dr. Y. Sato for help in SEM; the National Bioresource Project (NBRP) for providing the seeds of GLSL and *O. glaberrima* lines. I also thank Mr. Tomonori Takashi for the collaborative project to develop RSLs and WBSLs; Dr. Atsushi Yoshimura for their kind provision of essential materials in this work.

I wish to express my deepest gratitude to Dr. Motoyuki Ashikari, Professor of Bioscience and Biotechnology Center, Nagoya University, for accepting me as a graduate student and for his keen advice and gracious guidance during the entire course of this study. My heartfelt many thanks goes to the members in the Laboratory of Molecular Biosystem, Laboratory and o all members in Bioscience and Biotechnology Center, Nagoya University for their friendliness and help.

Finally, I express my cordial thanks to my parents, siblings and my dear husband for their warm words of encouragement and support during five years of my study in Nagoya University.

March 2017

Kanako UEHARA

List of publication

1. **Bessho-Uehara K**, Wang DR, Furuta T, Minami A, Nagai K, Gamuyao R, Asano K, Shim R, Shimizu Y, Ayano M, Komeda N, Doi K, Miura K, Greenberg A, Wu J, Yasui H, Yoshimura A, Mori H, McCouch SR and Ashikari M. (2016) Loss of function at *RAE2*, a novel EPFL, is required for awnlessness in cultivated Asian rice. *Proc Natl Acad Sci U S A* **113**: 8969-8974.

2. Furuta T, **Uehara K**, Shim RA, Shim J, Nagai K, Ashikari M and Takashi T. (2016) Development and evaluation of *Oryza nivara* chromosome segment substitution lines (CSSLs) in the background of *O. sativa* L. cv. Koshihikari. Breeding Sci. 66: 845-850.

3. Furuta T, Komeda N, Asano K, **Uehara K**, Gamuyao R, Shim RA, Nagai K, Doi K, Wang DR, Yasui H, Yoshimura A, Wu J, McCouch SR and Ashikari M. (2015) Convergent Loss of Awn in Two Cultivated Rice Species *Oryza sativa* and *Oryza glaberrima* is Caused by Mutations in Different Loci. *G3* **5**: 2267-2274.

4. Furuta T, **Uehara K**, Shim RA, Shim J, Ashikari M and Takashi T. (2014) Development and evaluation of chromosome segment substitution lines (CSSL) carrying chromosome segments derived from *Oryza rufipogon* in the genetic background of *Oryza sativa* L. *Breed. Sci.* **63**: 468-475.