**Title**

Loss of function at *RAE2*, a novel EPFL, is required for awnlessness in cultivated Asian rice.

**Authors**

Kanako Bessho- Uehara[1*], Diane R. Wang[2*], Tomoyuki Furuta[1], Anzu Minami[1], Keisuke Nagai[1], Rico Gamuyao[1], Kenji Asano[1], Rosalyn B. Angeles-Shim[1], Yoshihiro Shimizu[1], Madoka Ayano[1], Norio Komeda[1], Kazuyuki Doi[3], Kotaro Miura[4], Yosuke Toda[5], Toshinori Kinoshita[5], Satohiro Okuda[5], Tetsuya Higashiyama[5], Mika Nomoto[6], Yasuomi Tada[6], Hidefumi Shinohara[6], Yoshikatsu Matsubayashi[6], Anthony Greenberg[2], Jian Zhong Wu[7], Hideshi Yasui[8], Atsushi Yoshimura[8], Hitoshi Mori[3]†, Susan R. McCouch[2]† & Motoyuki Ashikari[1]†

**Affiliations**

[1] Bioscience and Biotechnology Center, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan

[2] Section of Plant Breeding and Genetics, School of Integrated Plant Sciences, Cornell University, 162 Emerson Hall, Ithaca, NY 14853-1901, USA

[3] Graduate School of Agriculture, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan

[4] Faculty of Biotechnology, Fukui Prefectural University, 4-1-1 Eiheiji-Town, Fukui 910-1195, Japan

[5] Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8602, Japan

[6] Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8602, Japan

[7] National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634, Japan

[8] Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

[*] These authors contributed equally to this paper

† Corresponding authors

E-mail: morihito@agr.nagoya-u.ac.jp, srm4@cornell.edu, ashi@agr.nagoya-u.ac.jp

**Abstract**

Domestication of crops based on artificial selection has contributed numerous beneficial traits for agriculture. Wild characteristics such as red pericarp and seed shattering were lost in both Asian, *Oryza sativa,* and African, *O. glaberrima*, cultivated rice species as a result of human selection on common genes. Awnedness, however, is a trait that has been lost in both cultivated species, caused by selection on different sets of genes. In a previous report, we revealed that at least three loci regulate awn development in rice, however the molecular mechanism underlying awnlessness remains unknown. Here, we isolate and characterize a novel EPFL family member named *REGULATOR OF AWN ELONGATION 2* (*RAE2*) and identify one of its requisite processing enzyme, *SUBTILISIN-LIKE PROTEASE 1* (*SLP1*). The RAE2 precursor is specifically cleaved by SLP1 in the rice spikelet where the mature RAE2 peptide subsequently induces awn elongation. Analysis of *RAE2* sequence diversity identified a highly variable GC-rich region harboring multiple, independent mutations underlying protein length variation that disrupt the function of the RAE2 protein and condition the awnless phenotype in Asian rice. In contrast, cultivated African rice retained the functional *RAE2* allele despite its awnless phenotype. Our findings illuminate the molecular function of *RAE2* in awn development and shed new light on the independent domestication histories of Asian and African cultivated rice.

**Significance statement**

This study investigates a novel Cysteine Rich Peptide (CRP). CRPs have diverse roles in plant growth and development, e.g. control of stomata density and guidance of pollen-tube elongation. Despite numerous studies on CRPs in *Arabidopsis thaliana*, there are still many peptides with unknown function. We identify a novel rice CRP named *Regulator of Awn Elongation 2* (*RAE2*) and show that it is cleaved specifically in the spikelet to promote awn elongation. We demonstrate that *RAE2* was a target of selection during domestication, contributing to loss of awns in Asian but not African rice. The discovery of *RAE2* simultaneously deepens our understanding of plant developmental pathways and lends insight into the complex processes underlying cereal domestication.

**Keywords**: awn, rice, signal peptide, parallel domestication

**Data deposition**

The accession numbers for the coding sequences of four RAE2 variants including three singletons of 5C (4C, 5C_1, 5C_2, 5C_3, 6C, 7C) are registered in DDBJ as LC119056- LC119061.

**\body**

## Introduction

Since the dawn of agriculture, cultivated plants have contributed to human health and prosperity. Among the three major cereals, wheat (*Triticum aestivum*), maize (*Zea mays*) and rice, only rice has been independently domesticated twice (1-3). There are two cultivated rice species, one is *O. sativa* derived from *O. rufipogon* in Asia, and the other, *O. glaberrima* derived from *O. barthii* in Africa (4, 5). While the habitats of *O. sativa* and *O. glaberrima* are geographically isolated, the two species share a suite of domestication-related traits, e.g., decreased seed shattering, upright growth habit, and awnless seeds (6-9). This suggests that such characteristics significantly contribute to promote agriculture. There is evidence for artificial selection at a genome-wide scale in *O. sativa* and *O. glaberrima* (10), and several domestication traits are regulated by the same genes in both species (11-14). However, convergent evolution via different genes is another scenario by which shared phenotypes may arise from geographically independent domestication processes. Recently, we reported that the genetic regulation of awnlessness differs in both cultivated rice (15). Awns are sharp spine-like structures extending from the lemmas of rice florets (Fig. 1A). Their bristle-like architecture enhances seed dispersal and provides protection against seed predation (16). Thus, long awns are considered important for habitat expansion and survival of wild rice. Under domestication, the awnless phenotype has been selected to facilitate planting, harvesting and storage of seeds (17).

To date, two genes have been identified to regulate awn development in rice and show evidence of selection during domestication: a bHLH transcription factor called

*An-1* and a cytokinin-activating enzyme named *LABA1,* both residing on chromosome 4 (18, 19). These reports focused on allelic variation solely within Asian rice, and it remained unclear whether mutations in these same genes might be responsible for the awnless phenotype in African cultivated rice. Using a library of chromosome segment substitution lines (CSSLs) derived from a cross between two awnless parents, *O. sativa* ssp. *japonica* and *O. glaberrima* (20) (Fig. 1B and C), two of the lines were observed to have long awns; line GLSL13 carried an *O. glaberrima* introgression on chromosome 4 that included *Regulator of Awn Elongation 1* (*RAE1*) which is correspond to *An-1*, and line GLSL25 carried an *O. glaberrima* introgression on chromosome 8 containing *RAE2* (Fig. 1D) (15). Although QTLs associated with awn elongation around the *RAE2* region have been reported, the responsible gene has not been identified (21-23).

Here we show that *RAE2* is an *EPIDERMAL PATTERNING FACTOR-LIKE 1* (*EPFL1*) gene and demonstrate that the *RAE2* allele found in *O. glaberrima* induces awn development in *O. sativa*. We also identify *SUBTILISIN-LIKE PROTEASE 1* (*SLP1*) as a protease required for RAE2 processing in the rice spikelet. Finally, through *RAE2* sequence analysis in a diverse collection of Asian and African rice, we reveal a highly variable GC-rich region that harbors multiple, independent mutations underlying protein length variation. We show that the array of differences found in this GC-rich region gives rise to a suite of dysfunctional RAE2 variants selected upon during domestication that conditions the awnless phenotype in Asian rice.

**Results**

## *O. sativa* has a dysfunctional RAE2 protein while *O. glaberrima* carries a functional RAE2

To identify the causal gene *RAE2* on chromosome 8 from *O. glaberrima* that induces awn elongation in the genetic background of *O. sativa* ssp. *japonica*, cv Koshihikari, we undertook positional cloning in a mapping population derived from GLSL25 x Koshihikari (Fig. S1A). Genetic linkage analysis using about ~8,000 $F_2$ individuals delimited the candidate region to about 80 kb which encompassed twelve predicted genes (Fig. S1B and C). We screened an *O. glaberrima* BAC library and identified a clone Ogla0006B21 that encompassed this candidate region (Fig. S1C). Five sub-clones of Ogla0006B21 were systematically introduced into Taichung 65 (T65), an awnless *O. sativa* ssp. *japonica* cultivar (Fig. S1D). Two of the sub-clones (#33 and #89) recovered awns in the transgenic lines (Fig. S1E-G and Table S1). A single ORF, *Os08g0485500,* was shared between subclones #33 and #89 (Fig. S1C and D). When a construct containing only *Os08g0485500* (*O. glaberrima* allele) was introduced into Nipponbare (awnless *O. sativa* ssp. *japonica*), the resultant transgenic plants produced awns of comparable frequency and length to the awns of GLSL25 (Fig. 1E-G and Table S1). RNA interference (RNAi) experiments using a construct harboring the *Os08g0485500* 3'UTR transformed into GIL116, an awned introgression line carrying the chromosome 8 segment derived from *O. glaberrima* in the T65 genetic background (Fig. S2), generated lines with awn lengths significantly shorter than those of the vector controls, while awn frequencies were not different (Fig. 1H-J and Table S1). Together, these results indicate that *Os08g0485500* is *RAE2,* and that this gene acts to regulate awn elongation.

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

Based on amino acid sequence analysis, *RAE2* is predicted to encode an *EPIDERMAL PATTERNING FACTOR-LIKE1* (*EPFL1*) protein (24). This protein is a member of the EPF/EPFL family, a group of plant-specific secreted peptides that regulates a range of developmental processes (25-29). In *Arabidopsis thaliana,* the most extensively studied EPF/EPFL is Stomagen (also known as AtEPFL9) and its competitive factors, EPF1 and EPF2. In contrast, there are no reports about this family 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 (30).

Using cysteine-rich region sequences, we evaluated the phylogenetic relationship of RAE2 with other members of the EPF/EPFL family from *Arabidopsis thaliana* and several awned grass species (*Brachypodium distachyon*, *Triticum urartu* and *Hordeum vulgare*) (Fig. S3A). Phylogenetic analysis revealed that *RAE2* is classified into the AtEPFL1-3 clade, a group of unknown function. 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. S3B).

Three-dimensional structure modeling revealed clear structural similarity between RAE2 and Stomagen (Fig. S4). The loss of cysteine residues has been reported to cause dysfunctional activity of Stomagen due to the non-formation of a critical scaffold

mediated by disulfide bonds (31). Comparative sequence analysis of *RAE2* and *rae2* from IRGC104038 (*O. glaberrima*) and Koshihikari (*O. sativa* ssp. *japonica*), respectively, revealed several SNPs and insertions in the promoter and coding region (Fig. S5A). A 2-bp insertion in the second exon of *rae2* leads to a truncated protein through a frame-shift mutation (Fig. 1K, S5B and C), therefore hypothesized to be the causal mutation for dysfunctional conformation.

**Coordination of *RAE2* expression with awn development**

We next analyzed the expression of *RAE2* throughout the rice plant to determine its correlation with awn development. Plant-wide, *RAE2* expression was about 10-fold higher in young panicles than in the other organs (Fig. 2A). Within the young panicle, expression of *RAE2* was significantly greater than that of *rae2*. We additionally observed awn development in Koshihikari and GLSL25 by scanning electron microscopy (SEM). SEM observations showed that lemma and palea morphology did not differ between the two lines until the Sp7 stage (Fig. 2B and F). The awn primordium protruded at the distal end of the lemma only in GLSL25 at the beginning of Sp8 (Fig. 2C-E and 2G-I). *In situ* hybridization showed similar expression patterns for both *RAE2* and *rae2* from Sp4 through Sp7 (Fig. 2J-L and 2O-Q). *RAE2* transcripts, however, exhibited prolonged expression compared with *rae2* transcripts in the subsequent stages (Fig. 2M and R). GLSL25 showed especially high expression in the vascular bundles of the awn primordium compared to Koshihikari (Fig. S6A-G). Together, these observations provide evidence of the importance of the spatio-temporal

regulation of *RAE2* expression for awn elongation.

**Protein and allelic diversity 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, 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 S2). The 2 bp insertion in the second exon of *rae2* corresponding to the functional mutation in Koshihikari (Fig. 1K and S5), occurred in a highly variable GC-rich repeat region in our diversity panel (Fig. S7A (i) - (iii)). We observed seven different length-polymorphisms in this region (Table S3). Among seven variants, four translated into functional RAE2 proteins (i.e. 6 cysteine residues), while the other three gave rise to putatively dysfunctional RAE2 proteins; either a truncated protein, as short protein in cv. Koshihikari (4 cysteines), or a long protein (7 cysteines). We named these translated products according to the number of cysteine residues they harbor in cysteine rich region and grouped them into three protein-length classes: 4C/short, 6C/medium, and 7C/long (Fig. S7B and Table S3). In addition, three singleton variants of *RAE2* were identified that independently gave rise to translated products with 5 cysteine residues, resulting from polymorphisms that occurred outside the GC-rich repeat region (listed in Table S4). All of these were predicted to cause medium length proteins but are compositionally divergent from the 6C/medium peptide class (Fig. S7B). Based on these results, nucleotide variations in the

GC-rich repeat region are responsible for RAE2 protein length and its function.

**RAE2 loss of function alleles have been selected in Asian but not in African rice**

To understand the functionality of the RAE2 protein variants in relation to awn development we evaluated overexpression lines of each variant. Only the RAE2-6C type exhibited the awn phenotype, while the other 3 lines did not show awns, regardless of their equal expression levels (Fig 3A; S8A-C). This indicates RAE2-6C type is functional for awn elongation, but other 3 types (RAE2-4C, RAE2-5C and RAE2-7C) are not functional. We also evaluated *RAE2* variants in two species of Asian wild rice, *O. nivara* (Acc. W0054) which is annual type of *O. rufipogon* having RAE2-6C type, and *O. rufipogon* (Acc. W0106) carrying RAE2-7C type. These wild accessions were used as donors for two previously developed CSSL populations (WBSL and RSL, respectively) in the background of cv. Koshihikari (Fig. S9A) (15). Phenotyping the two lines from these CSSL populations that harbored the *RAE2* locus on chromosome 8 derived from *O. nivara* (WBSL18) and *O. rufipogon* (RSL23) demonstrated that line WBSL18 formed awns, but RSL23 did not (Fig. S9B). Consistent with our results from overexpression analysis of RAE2 alleles, *O. nivara* harbored the functional 6C/medium RAE2, while *O. rufipogon* possessed the dysfunctional 7C/long RAE2 type (Fig. S9C).

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. 3B). Interestingly, all cultivated African rice varieties (*O. glaberrima*) and its ancestor wild African rice, *O. barthii* were found to carry the RAE2-6C type,

despite the fact that *O. glaberrima* does not have awn and *O. barthii* possessed awns (Fig. 3C(i, ii), 3D(i, ii)). These results support our previous report that a different gene(s), other than *RAE2*, is responsible for the awnless phenotype of *O. glaberrima* in Africa (15). Namely, loss of function of *RAE3* causes awnless in domestication in Africa.

In Asia, dysfunctional RAE2 protein types were present in 32% of *O. rufipogon* and *O. nivara* wild accessions (Fig. 3C(iii)), while almost all individuals possessed awns (Fig. 3D(iii)). This distinguished *RAE2* from previously reported awn domestication genes, *An-1*/*RAE1* and/or *LABA1*, which were documented to persist as functional alleles in the vast majority of wild Asian rice populations (18, 19). Although *O. sativa* was nearly fixed (93%) for dysfunctional alleles at *RAE2* and most accessions were awnless (Fig. 3C(iv) and 3D(iv)), the five subpopulations harbored different frequencies of *RAE2* protein classes (Table S5). Awned cultivated Asian rice would have functional *An-1*/*RAE1* and/or *LABA1* still.

To test for evidence of a selective sweep in the region of *RAE2*, nucleotide diversity ($\pi$) of *O. sativa* (n=67) relative to nucleotide diversity ($\pi$) of *O. rufipogon* (n=65) was estimated using data in 100-SNP sliding windows across chromosome 8. A drastic decrease was observed in the ratio $\pi$ *O. sativa*/ $\pi$ *O. rufipogon* across a 1.5 Mb region flanking *RAE2* (Fig. 3E(i)), 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 another target of selection nearby. Using the Kolmogorov-Smirnov test, we tested whether the distribution of this diversity ratio from the RAE2 region is different from the rest of the chromosome. It is indeed highly unusual ($P < 2.2\times10^{-16}$).

Analyzing the genetic distance (*d*) between *O. sativa* and *O. rufipogon* using the same 100-SNP windows across chromosome 8 revealed that the dysfunctional protein types 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. 3E(ii)). This result is consistent with recent gene flow back from cultivated to wild Asian rice (32).

**RAE2 is cleaved specifically in the spikelet**

The EPFL family of peptides typically requires post-translational cleavage of a pro-peptide to become a mature peptide (Fig. S10A) (33, 34). 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. 4A, B and S10B). To confirm this spikelet-specific cleavage of RAE2, we conducted *in vitro* processing assay by mixing the recombinant RAE2 pro-peptide fused 3xFLAG tag (RAE2-pro) with each of the protein extracts derived from various organs (callus, stem, leaf and spikelet). The result of immunoblot analysis by FLAG antibody showed that only spikelet extract could cleave RAE2-pro (Fig. 4C and S10C). In addition we 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 (149-RDRLFDP-125) (Fig. S10D). Furthermore, the cleavage of RAE2-pro was inhibited by a protease inhibitor cocktail, Complete (Fig. S10E).

Together, these data suggest that RAE2-targeted protease(s) specifically functions in rice spikelets.

**Identification of RAE2-targeted protease**

To find the candidate protease(s), we compared expression patterns of 63 members of the rice subtilisin-like protease family (listed in Table S6). This protease family has been shown to cleave EPF/EPFL peptides in *A. thaliana* (*SDD1*(35), *CRSP* (36)). Among the 63 proteases examined, we identified one promising candidate, *Os01g0702300*, which was specifically expressed around the young inflorescence according to Rice-Xpro (http://ricexpro.dna.affrc.go.jp/) (S11A). We confirmed this expression pattern experimentally (Fig. S11B and C) and named this gene *SUBTILISIN-LIKE PROTEASE 1* (*SLP1*). Analysis of phylogenetic relationships classified this gene in the same clade as CRSP (37), supporting the hypothesis that *SLP1* is the most promising candidate gene targeting RAE2. To determine whether the RAE2 could be cleaved by SLP1, we performed *in vitro* processing assay using a synthetic RAE2 short peptide (synRAE2) spanning the predicted cleavage site (52-AGEEEKVRLGSSPPSCYSK-70) according to Stomagen (28) and EPF2 (36) with *in vitro*-synthesized SLP1 protein. As a result three cleaved peptide were detected (Fig. 4D and S12A). 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 the shorter version would work as ligand and according to the Stomagen and EPF2 reports the site between P65 and S66 is more appropriate as

cleavage site. In addition, we tested whether SLP1 digested a series of mutated synRAE2 with a single amino acid substitution near the predicted cleavage site (mu-synRAE2). Cleavage of mu-synRAE2 (P64G, P65G P65A and S66D) by *in vitro*-synthesized SLP1 was inhibited (Fig. S12B). Further the alanine substitution closer to the cleavage point repressed RAE2 cleavage using mutated RAE2-pro (muRAE2 #1-#4) (Fig. 4E). These results suggested SLP1 cleaves RAE2 between the amino acid positions P65 and S66 in the spikelet specifically (Fig. 4F).

**Discussion**

Our study identifies *RAE2* as *Os08g0485500,* a novel EPFL gene that is preferentially expressed during early panicle developmental stage to promote awn elongation. RAE2 harbors a conserved six cysteine peptide region characteristic of the EPF/EPFL family while rae2 in most awnless rice accessions have alternative numbers of cysteine residues due to frame shift mutations. Ohki *et al.* (2011) reported that the consensus scaffold stabilized by three pairs of disulfide bonds is structurally required for the EPF/EPFL activity in Stomagen (31). They showed that substituting cysteine residues with serine residues in Stomagen caused incorrect disulfide bonds, which led to improper conformation. It is likely that the lack of disulfide bonds made it difficult to cause an active conformation. According to this report, rae2 variants without two cysteine residues or had extra cysteine residues in improper locations lost proper conformation of RAE2 and were not able to promote awn elongation.

We also showed that SLP1 cleaves RAE2 from pro-peptide to mature-peptide. It is

known that EPF/EPFL peptides are processed into mature peptides, which then bind to receptors such as the ERECTA family of receptor kinases (38, 39). We hypothesize that RAE2 may mediate awn elongation through binding with unidentified rice ERECTA kinase(s). Improper scaffolding of mature RAE2 peptide due to mutations may hinder binding by these receptor(s) and lead to an awnless phenotpey. At this moment, we do not know if the cleavage of RAE2 by SLP1 is essential for awn elongation. Further evaluation of transgenic plants transduced with RAE2 constructs with and without mutations at the cleavage site are necessary to elucidate the role of RAE2 cleavage in spikelet. The observation that transgenic Nipponbare plants transduced with *O. glaberrima RAE2* were able to produce awns demonstrates that cv. Nipponbare retains functional RAE2 receptor(s). As for the cleavage of RAE2, *in vitro* processing assay showed that *O. sativa* also possesses a functional SLP1 and supporting sequence comparison of *O. sativa* and *O. glaberrima* SLP1 demonstrated that there are no deleterious frame shift mutations in coding region (Fig. S13A and B). Taken together, these results suggest that EPF/EPFL peptides other than RAE2 may interact with these downstream components to mediate unidentified, conserved and important functions in the rice plant.

Since vascular bundles penetrate the center of awns, RAE2 may function in promoting proliferation of vasculature cells for awn elongation. A similar phenomenon has been reported in Arabidopsis; AtEPFL4 and AtEPFL6 coordinate development of inflorescence architecture (29). In a previous mutant study, DL and OsETT2 were reported to be involved in rice awn development (40). *DL* is YABBY gene essential for forming leaf midribs by promoting cell proliferation, and *OsETT2* is an auxin

responsive factor. Future identification of the receptor(s) of RAE2 and downstream factors affecting awn development will be able to reveal the relationship among these components.

Sequence comparison of Asian and African rice suggested that the various mutations in the critical GC-rich repeat region, and that speciation led to RAE2 diversification. We infer 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 genes (e.g. RAE1/An-1 or LABA1) can mask RAE2's loss-of-function, this raises the possibility that RAE2 is conserved in the wild due to pleiotropic effects on other fitness phenotypes. On the other hand, while African rice maintained a functional RAE2 protein, we hypothesize that a mutation(s) in different locus, designated *RAE3*, represses the awn phenotype in *O. glaberrima* (15). With its hyper variable GC-rich region occuring protein length diversity, and its multiple protein classes contributing to variation in awn elongation. 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 divergent domestication history of Asian and African rice.

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

**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, MEXT Grant-in-Aid for Scientific Research on Innovative Areas Grant Number 16H01464, SATREPS by JICA and JST, and the Core Research for Evolutional Science and Technology by JST. We 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.

**References**

1. Salamini F, Özkan H, Brandolini A, Schäfer-pregl R (2002) Genetics and Geography of Wild Cereal Domestication in the Near East. *Genetics* 3(6):429-441.

2. Heerwaarden JV, et al. (2010) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci USA* 108(3):1088–1092.

3. Flint-garcia SA (2013) Genetics and Consequences of Crop Domestication. *J. Agric. Food Chem.* 61(35):8267−8276.

4. Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol.* 35(1-2):25-34.

5. Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor Appl Genet.* 104(8):1217-1228.

6. Khush GS (2001) Green revolution : the way forward. *Genetics* 2(10):815-822.

7. Jin J, et al. (2008) Genetic control of rice plant architecture under domestication. *Nat Genet* 40(11):1365-1369.

8. Huang X, et al. (2009) Natural variation at the DEP1 locus enhances grain yield in rice. *Nat Genet* 41(4):494-497.

9. Huang X, et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490(7421):497-501.

10. Wang M, et al. (2014) The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet* 46(9):982-988.

11. Konishi S, et al. (1999) An SNP Caused Loss of Seed Shattering During Rice Domestication. *Science* 70(1999):1392-1396.

12. Gross BL, Steffen FT, Olsen KM (2011) The molecular basis of white pericarps in African domesticated rice: Novel mutations at the Rc gene. *J Evo Biol.* 23(12):2747-2753.

13. Vigueira CC, Li W, Olsen KM (2013) The role of Bh4 in parallel evolution of hull colour in domesticated and weedy rice. *J Evol Biol.* 26(8):1738-1749.

14. Purugganan MD (2014) An evolutionary genomic tale of two rice species. *Nat Genet.* 46(9):931-932.

15. Furuta T, 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* 5(11):2267-2274.

16. Grundbacher FJ. (1963) The physiological function of the cereal awn. *Bot. Rev.* 29(3):366-381.

17. Tatsumi J, Kawano K (1972) Suitou no noge ni tsuite (about the awn in rice). *Research reports of Tokai branch of crop science society of japan* 56(1):11-15.

18. Luo J, 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(9):3360-3376.

19. Hua L, et al. (2015) LABA1, a Domestication Gene Associated with Long , Barbed Awns in Wild Rice. *Plant Cell* 27(7):1875-1888.

20. Angeles-Shim RB, Angeles ER, Ashikari M, 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.* 63(5):613-619.

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

22. Cai HW, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theor Appl Genet.* 104(8):1217-1228.

23. Fawcett JA, et al. (2013) QTL Map Meets Population Genomics: An Application to Rice. *Plos One* 8(12): e83720

24. Takata N, et al. (2013) Evolutionary Relationship and Structural Characterization of the EPF / EPFL Gene Family. *Plos One* 8(6):4-9.

25. Hara K, Kajita R, Torii KU, Bergmann DC, Kakimoto T (2007) The secretory peptide gene EPF1 enforces the stomatal one-cell-spacing rule. *Genes Dev.*

21(14):1720-1725.

26. Hunt L, Gray JE, Court F, Bank W, Sheffield S (2009) Report The Signaling Peptide EPF2 Controls Asymmetric Cell Divisions during Stomatal Development. *Curr Biol.* 19(10):864-869.

27. Kondo 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):1-8.

28. Sugano SS, et al. (2010) Stomagen positively regulates stomatal density in Arabidopsis. *Nature* 463(7278):241-244.

29. Uchida N, et al. (2012) Regulation of inflorescence architecture by intertissue layer ligand–receptor communication between endodermis and phloem. *Proc Natl Acad Sci USA* 109(16):6337-6342.

30. Marshall E, Costa LM, Gutierrez-Marcos J (2011) Cysteine-Rich Peptides (CRPs) mediate diverse aspects of cell–cell communication in plant reproduction and development. *J Exp Bot.* 62(5):1677-86.

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

32. Song ZP, Lu BR, Zhu YG, Chen JK (2003) Gene flow from cultivated rice to the wild species Oryza rufipogon under experimental field conditions. *New Phytologist* 157(3): 657–665.

33. Katsir L, Davies KA, Bergmann DC, Laux T (2011) Peptide Signaling in Plant Development. *Curr Biol.* 21(9):R356-R364.

34. Wheeler JI, Irving HR. Plant Signaling Peptides. 2011.

35. Groll UV, Berger D, Altmann T (2002) The Subtilisin-Like Serine Protease SDD1 Mediates Cell-to-Cell Signaling during Arabidopsis Stomatal Development. *Plant cell* 14(7):1527-1539.

36. Engineer CB, et al. (2014) Carbonic anhydrases, EPF2 and a novel protease mediate CO2 control of stomatal development. *Nature* 513(7517):246-250.

37. Tripathi PL, Sowdhamini R (2006) Cross genome comparisons of serine proteases in Arabidopsis and rice. *BMC Genomics.* 7(200):1-31.

38. Lee JS, et al. (2012) Direct interaction of ligand – receptor pairs specifying stomatal patterning. *Genes Dev.* 26(2):126-136.

39. Lee JS, et al. (2015) Competitive binding of antagonistic peptides fine-tunes

stomatal patterning. *Nature* 522(7557):439-443.

40. Toriba T, Hirano HY. (2014) The DROOPING LEAF and OsETTIN2 genes promote awn development in rice. *Plant J.* 77(4):616-626.

**Figure Legend**

**Fig. 1. Identification and functional characterization of *RAE2*.**

(**A**) Awn is a spine-like extension of the rice lemma. Left panel: awned rice seed anatomy. Right panel: panicles of awned chromosome segment substitution line, GLSL25. Red arrowhead represents awn. (**B-D**) Seed phenotypes and graphical genotypes of Koshihikari (*O. sativa* ssp. *japonica*, yellow) (**B**), IRGC104038 (*O. glaberrima*, blue) (**C**) and GLSL25 (**D**). (**E-G**) Evaluation of transgenic plants with plasmid vector pGWB501 (vector control; V. C.) and *RAE2* gene (p*RAE2*::*RAE2*): seed phenotype (**E**), frequency of awned seeds per panicle (**F**), awn length (**G**). No visible awn was observed in pGWB501 (V.C.) indicated as n.d. (not detected). (**H-J**) Evaluation of vector control (pANDA (V.C.) and RNAi line (*RAE2*-RNAi); seed phenotype (**H**), frequency of awned seeds per panicle (**I**), awn length (**J**). (**K**) rae2/ OsEPFL1 amino acid structure of Koshihikari (*O. sativa* ssp. *japonica*) and RAE2 of IRGC104038 (*O. glaberrima*). Yellow triangle indicates insertion. Each colored box represents a peptide region. sp=signal peptide (blue), pro=pro-peptide (green), ma=mature peptide (gray or red). Red bar indicates cysteine (C) residues. Scale bar length represents 1 cm. The statistical significance was at * $P < 0.05$ based on two-tailed Student's t-test. Error bars represent standard deviation of the mean.

**Fig. 2. 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* was used as internal control. Error bar indicates standard deviation of the mean. (**B-I**) Scanning electron microscopy images of spikelets at different developmental stages in Koshihikari (**B-E**) and GLSL25 (**F-I**). Developmental stages are classified into Sp7 (**B, F**), Sp8 (**C, G**), and post Sp8 (**D, E, H, I**) according to Oryzabase classification

(http://www.shigen.nig.ac.jp/rice/oryzabase/devstageineachorgan/list). Scale bars represent 50μm. Red arrowhead indicates awn. (**J-S**) i*n situ* hybridization using antisense probes of *rae2* (**J-M**), *RAE2* (**O-R**) and sense probes (**N, S**) during spikelet development in Koshihikari and GLSL25. Blue arrowhead indicates the tip of palea, red indicates the tip of lemma and green asterisks show anther. pa= palea, le= lemma, an = anther. Scale bars indicate 50μm (**J-L, O-Q**), 100μm (**M-N, R-S**).

**Fig. 3. Distribution of RAE2 protein variants across diverse rice accessions.**

(**A**) Awn phenotypes 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**) Geographical distribution of RAE2 protein variants found across these same 130 accessions (107 Asian rice accessions (circle) and 23 African rice accessions (triangle)). Symbol sizes are proportional to number of accessions and are indicated by the numbers in the rectangle box. The color represented each protein variant: green=6C/medium, blue=5C/medium, red=4C/short, yellow=7C/long. (**C**) 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. 4B. (**D**) 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). (**E**) (i) 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 represents reducing relative diversity surrounding RAE2 (green line) consistent with a selective sweep. (ii) Genetic distance between *O. sativa* and *O. rufipogon/O. nivara* for all RAE2 types (blue) and dysfunctional ones (red: 4C, 7C). Decreased distance in the dysfunctional class relative to distance in 'all' class in a 1.5 Mb region surrounding RAE2 (gray box).

**Fig. 4. RAE2 maturation caused by cleavage with SLP1 protease at 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 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). (**C**) *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). (**D**) 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). (**E**) *in vitro* processing assay of a series of alanine substituted recombinant RAE2 peptide (muRAE2) using spikelet extract of Koshihikari or buffer (mock). Table below shows the amino acid sequence around predicted cleavage point. Other description is same as Fig. 4C. (**F**) 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. See also Figures S10 and S12.