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

Kanako Bessho-Uehara[a,1], Diane R. Wang[b,1], Tomoyuki Furuta[a], Anzu Minami[a], Keisuke Nagai[a], Rico Gamuyao[a], Kenji Asano[a], Rosalyn B. Angeles-Shim[a], Yoshihiro Shimizu[a], Madoka Ayano[a], Norio Komeda[a], Kazuyuki Doi[c], Kotaro Miura[d], Yosuke Toda[e], Toshinori Kinoshita[e], Satohiro Okuda[e], Tetsuya Higashiyama[e], Mika Nomoto[f], Yasuomi Tada[f], Hidefumi Shinohara[f], Yoshikatsu Matsubayashi[f], Anthony Greenberg[b], Jianzhong Wu[g], Hideshi Yasui[h], Atsushi Yoshimura[h], Hitoshi Mori[c,2], Susan R. McCouch[b,2], and Motoyuki Ashikari[a,2]

[a]Bioscience and Biotechnology Center, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan; [b]Section of Plant Breeding and Genetics, School of Integrated Plant Sciences, Cornell University, Ithaca, NY 14853-1901; [c]Graduate School of Agriculture, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8601, Japan; [d]Faculty of Biotechnology, Fukui Prefectural University, 4-1-1 Eiheiji-Town, Fukui 910-1195, Japan; [e]Institute of Transformative Bio-Molecules, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8602, Japan; [f]Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa, Nagoya, Aichi 464-8602, Japan; [g]National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634, Japan; and [h]Faculty of Agriculture, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Edited by Maarten Koornneef, Max Planck Institute for Plant Breeding Research, Cologne, Germany, and approved June 22, 2016 (received for review March 24, 2016)

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 (*Oryza glaberrima*) cultivated rice species as a result of human selection on common genes. Awnedness, in contrast, is a trait that has been lost in both cultivated species due to 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 previously unidentified EPIDERMAL PATTERNING FACTOR-LIKE (EPFL) family member named *REGULATOR OF AWN ELONGATION 2* (*RAE2*) and identify one of its requisite processing enzymes, *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. Cultivated African rice, on the other hand, retained the functional *RAE2* allele despite its awnless phenotype. Our findings illuminate the molecular function of *RAE2* in awn development and shed light on the independent domestication histories of Asian and African cultivated rice.

awn | rice | signal peptide | parallel domestication | convergent evolution

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, *Oryza sativa* derived from *Oryza rufipogon* in Asia and *Oryza glaberrima* derived from *Oryza barthii* in Africa (4, 5). Although the habitats of *O. sativa* and *O. glaberrima* are geographically isolated, the two species share a suite of domestication-related traits, such as decreased seed shattering, upright growth habit, and awnless seeds (6–9). This suggests that such characteristics significantly contribute to promoting 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 the two species of cultivated rice (15). Awns are sharp spine-like structures extending from the lemmas of rice florets (Fig. 1*A*). 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 that regulate awn development in rice and show evidence of selection during domestication: a basic helix-loop-helix transcription factor called *AWN1* (*An-1*) and a cytokinin-activating enzyme named *LONG AND BARBED AWN1* (*LABA1*), both residing on chromosome 4 (18, 19). These reports focus on allelic variation solely within Asian rice, and it remains 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)

Significance

This study investigates a previously unidentified cysteine-rich peptide (CRP). CRPs have diverse roles in plant growth and development, such as 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 previously unidentified 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.

Author contributions: K.B.-U., D.R.W., T.K., T.H., Y. Tada, Y.M., A.Y., S.R.M., and M. Ashikari designed research; K.B.-U., D.R.W., T.F., A.M., K.N., R.G., K.A., R.B.A.-S., Y.S., M. Ayano, N.K., K.D., K.M., Y. Toda, S.O., M.N., H.S., J.W., H.Y., H.M., and M. Ashikari performed research; K.B.-U. contributed new reagents/analytic tools; K.B.-U., D.R.W., A.G., S.R.M., and M. Ashikari analyzed data; and K.B.-U., D.R.W., S.R.M., and M. Ashikari wrote the paper.

Conflict of interest statement: The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The coding sequences of the four *RAE2* variants including three singletons of 5C (4C, 5C_1, 5C_2, 5C_3, 6C, 7C) reported in this paper have been deposited in the DNA Data Bank of Japan (accession nos. LC119056–LC119061). The genotype information on 67 *O. sativa* and 65 *O. rufipogon* species complex samples used for selective sweep analysis has been deposited in the National Center for Biotechnology Information with submission IDs 1062529 and 1062530, respectively.

[1]K.B.-U. and D.R.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: morihito@agr.nagoya-u.ac.jp, srm4@cornell.edu, or ashi@agr.nagoya-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1604849113/-/DCSupplemental.

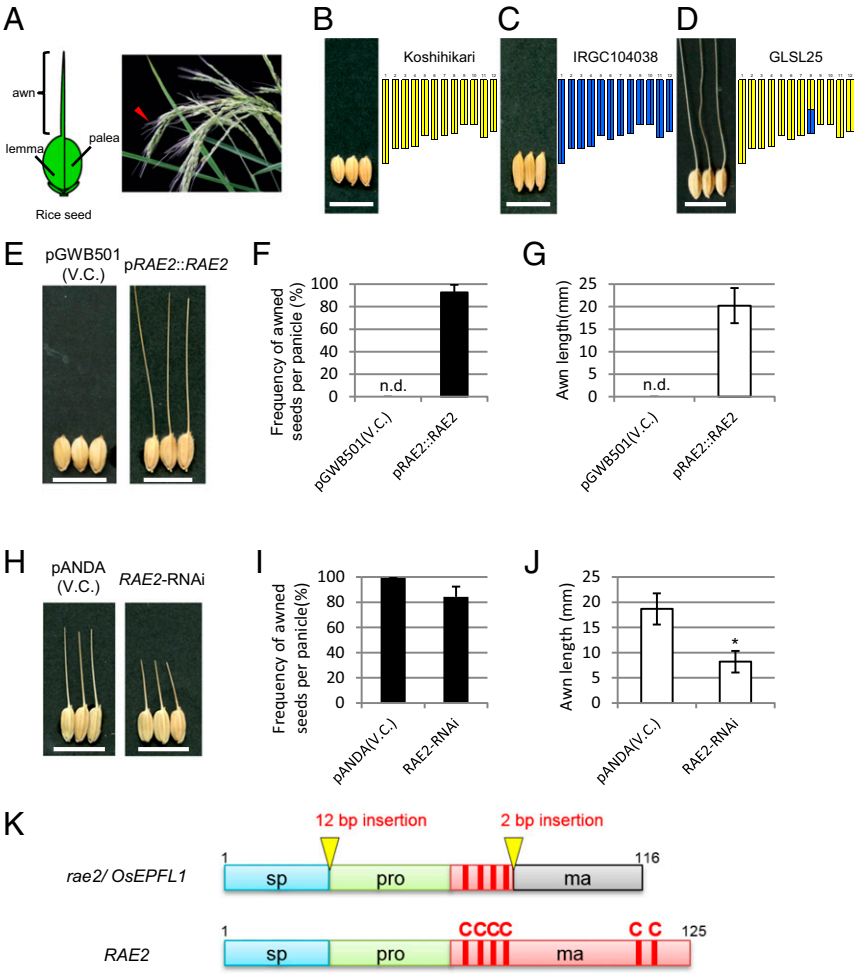

Fig. 1. Identification and functional characterization of *RAE2*. (*A*) The awn is a spine-like extension of the rice lemma. (*Left*) Awned rice seed anatomy. (*Right*) Panicles of the awned chromosome segment substitution line GLSL25. The red arrowhead points to the awn. (*B–D*) Seed phenotypes and 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*), and 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*), and awn length (*J*). (*K*) rae2/OsEPFL1 amino acid structure of Koshihikari (*O. sativa* ssp. *japonica*) and RAE2 of IRGC104038 (*O. glaberrima*). Yellow triangles indicate insertions. Each colored box represents a peptide region. ma, mature peptide (gray or red); pro, propeptide (green); sp, signal peptide (blue). Red bars indicate cysteine (C) residues. (Scale bars, 1 cm.) The statistical significance was at $*P < 0.05$ based on a two-tailed Student's *t* test. Error bars represent SD of the mean.

(Fig. 1 *B* 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 corresponds to *An-1*, and line GLSL25 carried an *O. glaberrima* introgression on chromosome 8 containing *RAE2* (Fig. 1*D*) (15). *RAE1* and *RAE2* have been shown to have redundant roles based on the fact that a functional allele at one or both loci confers awns in *O. sativa*. Although quantitative trait loci associated with awn elongation around the *RAE2* region have been reported, the responsible gene has not been identified (5, 21, 22).

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 that were selected during domestication and condition the awnless phenotype in Asian rice.

Results

***O. sativa* Has a Dysfunctional RAE2 Protein, Whereas *O. glaberrima* Carries a Functional RAE2.** To identify the functional *RAE2* allele 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 × Koshihikari (*SI Appendix*, Fig. S1*A*). Genetic linkage analysis using ~8,000 individuals from filial generation 2 (F_2) delimited the candidate region on chromosome 8 to about 80 kb, which encompassed 12 predicted genes (*SI Appendix*, Fig. S1 *B* and *C*). We screened an *O. glaberrima* bacterial artificial chromosome (BAC) library and identified a clone, Ogla0006B21, that encompassed this candidate region (*SI Appendix*, Fig. S1*C*). Five subclones of Ogla0006B21 were systematically introduced into Taichung 65 (T65), an awnless *O. sativa* ssp. *japonica* cultivar (*SI Appendix*, Fig. S1*D*). Two of the subclones (33 and 89) recovered awns in the transgenic lines (*SI Appendix*, Fig. S1 *E–G* and Table S1). A single open reading frame, *Os08g0485500*, was shared between subclones 33 and 89 (*SI Appendix*, Fig. S1 *C* 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. 1 *E–G* and *SI Appendix*, 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 (*SI Appendix*, Fig. S2), generated lines with awn lengths significantly shorter than those of the vector controls, whereas awn frequencies were not different (Fig. 1 *H–J* and *SI Appendix*, 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 1 PROTEIN in Rice. Based on amino acid sequence analysis, *RAE2* is predicted to encode an EPIDERMAL PATTERNING FACTOR-LIKE

1 (EPFL1) protein (23). This protein is a member of the EPF/EPFL family, a group of plant-specific secreted peptides that regulates a range of developmental processes (24–28). In *Arabidopsis thaliana*, the most extensively studied EPFs/EPFLs are 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 (29).

Using cysteine-rich region sequences, we evaluated the phylogenetic relationship of RAE2 with other members of the EPF/EPFL family from *A. thaliana* and several awned grass species (*Brachypodium distachyon*, *Triticum urartu*, and *Hordeum vulgare*) (*SI Appendix*, Fig. S3*A*). Phylogenetic analysis revealed that *RAE2* is classified into the AtEPFL1-3 clade, a group of genes with unknown function. Comparison of RAE2 and rae2 amino acid sequences with other EPFL relatives showed that all sequences except rae2 contained the six cysteine residues typical of EPFL peptides (*SI Appendix*, Fig. S3*B*).

Three-dimensional structure modeling revealed clear structural similarity between RAE2 and Stomagen (*SI Appendix*, Fig. S4). The loss of cysteine residues has been reported to cause dysfunction of Stomagen due to the nonformation of a critical scaffold mediated by disulfide bonds (30). 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 regions (*SI Appendix*, Fig. S5*A*). A 2-bp insertion in the second exon of *rae2* gave rise to a truncated protein through a frameshift mutation (Fig. 1*K* and *SI Appendix*, Fig. S5 *B* and *C*), and was hypothesized to be the causal mutation leading to 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 other organs (Fig. 2*A*). 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 spikelet development stage 7 (Sp7) stage (Fig. 2 *B* and *F*). The awn primordium protruded at the distal end of the lemma only in GLSL25 at the beginning of Sp8 (Fig. 2 *C*–*E* and *G*–*I*). In situ hybridization showed similar expression patterns for both *RAE2* and *rae2* from Sp4 through Sp7 (Fig. 2 *J*–*L* and *O*–*Q*). *RAE2* transcripts, however, exhibited prolonged expression compared with *rae2* transcripts in subsequent stages (Fig. 2 *M* and *R*). In GLSL25, especially high expression was observed in the vascular bundles of the awn primordium compared with Koshihikari (*SI Appendix*, Fig. S6). Together, these observations provide evidence of the importance of spatiotemporal regulation of *RAE2* expression for awn elongation.

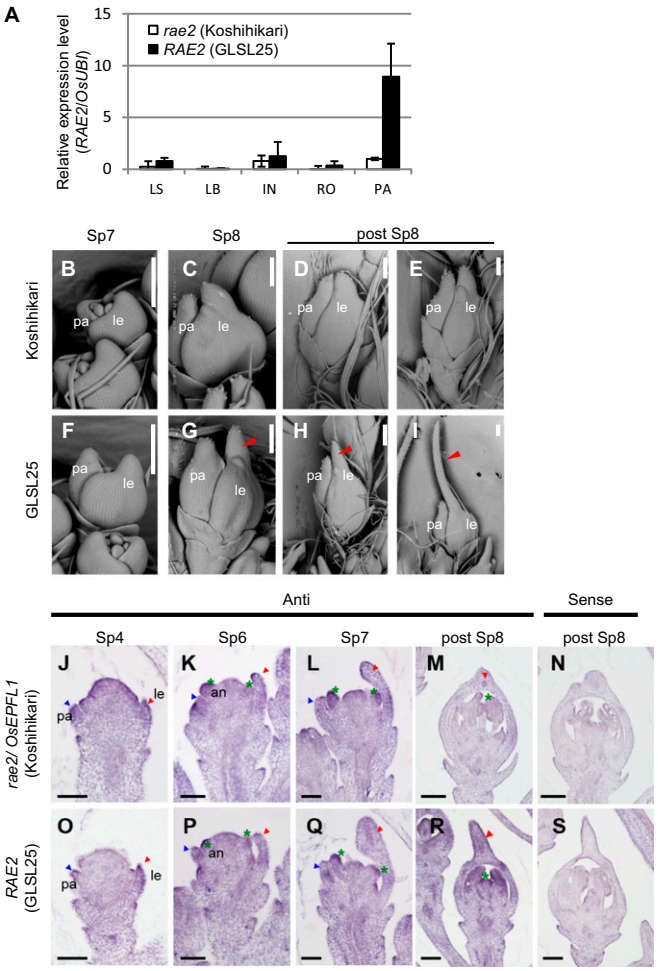

Fig. 2. *RAE2* expression pattern and correlation with awn development. (*A*) Quantitative RT-PCR showing *RAE2* mRNA levels in different organs of Koshihikari and GLSL25 (IN, internode; LB, leaf blade; LS, leaf sheath; RO, root; PA, young panicle). *OsUBI* was used as an internal control. Error bars represent SD of the mean. (*B*–*I*) SEM images of spikelets at different developmental stages in Koshihikari (*B*–*E*) and GLSL25 (*F*–*I*). Developmental stages are classified into Sp7 (*B* and *F*), Sp8 (*C* and *G*), and post Sp8 (*D*, *E*, *H*, and *I*) according to Oryzabase classification (www.shigen.nig.ac.jp/rice/oryzabase/devstageineachorgan/list). (Scale bars, 50 μm.) Red arrowheads point to the awn. (*J*–*S*) In situ hybridization using antisense probes of *rae2* (*J*–*M*), *RAE2* (*O*–*R*), and sense probes (*N* and *S*) during spikelet development in Koshihikari and GLSL25. Blue arrowheads indicate the tips of paleae; red arrowheads indicate the tips of lemmas; green asterisks show anthers (an, anther; le, lemma; pa, palea). [Scale bars, 50 μm (*J*–*L* and *O*–*Q*) and 100 μm (*M*, *N*, *R*, and *S*).]

Protein and Allelic Diversity of RAE2 and 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 *SI Appendix*, Table S2). The 2-bp insertion in the second exon of *rae2* corresponding to the functional mutation in Koshihikari (Fig. 1*K* and *SI Appendix*, Fig. S5) occurred in a highly variable GC-rich repeat region in our diversity panel (*SI Appendix*, Fig. S7*A*). We observed seven different length polymorphisms in this region (*SI Appendix*, Table S3). Among the seven variants, four translated into functional RAE2 proteins (i.e., six cysteine residues), whereas the other three gave rise to putatively dysfunctional RAE2 proteins: either a truncated, short protein in cv. Koshihikari (four cysteines) or a long protein (seven cysteines). We named these translated products according to the number of cysteine residues harbored in the cysteine-rich region and grouped them into three protein-length classes: 4C/short, 6C/medium, and 7C/long (*SI Appendix*, Fig. S7*B* and Table S3). These results demonstrated that nucleotide variations in the GC-rich repeat region are responsible for RAE2 protein length and affect protein function. In addition, three singleton variants of *RAE2* were identified that independently gave rise to translated products with five cysteine residues, resulting from polymorphisms that occurred outside the GC-rich repeat region (listed in *SI Appendix*, Table S4). All of these were predicted to cause medium-length proteins but are compositionally divergent from the 6C/medium peptide class (*SI Appendix*, Fig. S7*B*).

***RAE2* Loss-of-Function Alleles Have Been Selected in Asian but Not African Rice.** To understand the functionality of the RAE2 protein variants in relation to awn development, we evaluated overexpression lines

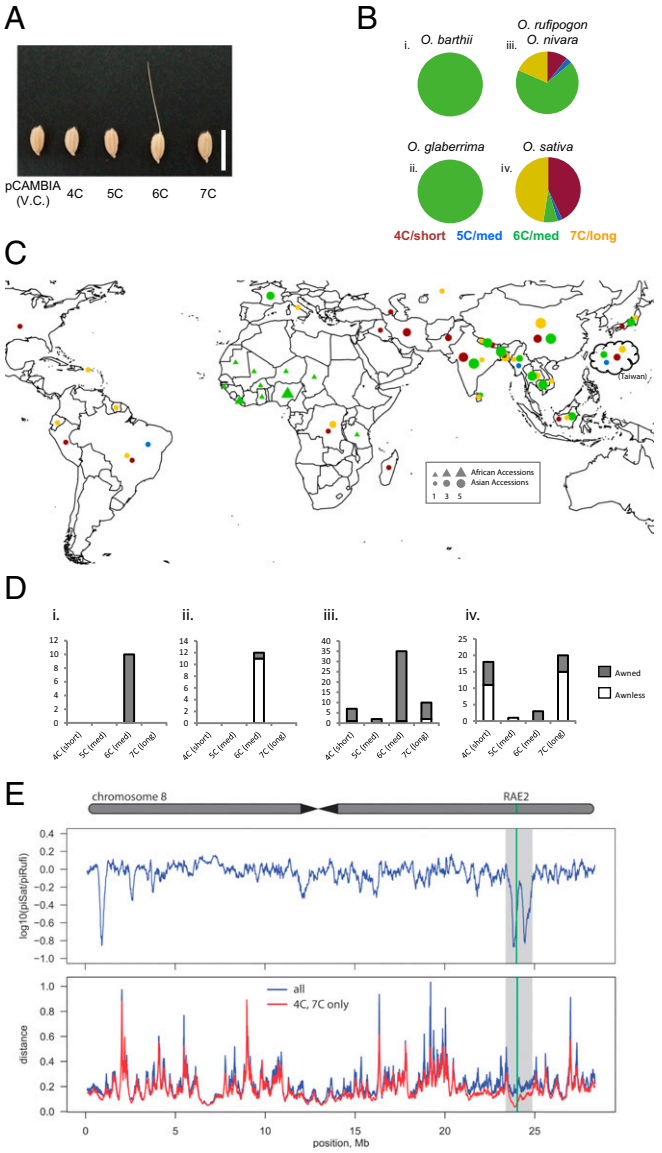

Fig. 3. Distribution of RAE2 protein variants across diverse rice accessions. (*A*) Awned phenotypes of overexpression lines carrying different RAE2 types (4C, 5C, 6C, 7C). pCAMBIA1380 was used as the vector control. (Scale bar, 1 cm.) (*B*) Distribution of the four RAE2 protein variants within *O. barthii* (*i*; $n = 11$), *O. glaberrima* (*ii*; $n = 12$), *O. rufipogon/O. nivara* (*iii*; $n = 65$), and *O. sativa* (*iv*; $n = 42$). (*C*) Geographical distribution of RAE2 protein variants found across these same 130 accession [107 Asian rice accessions (circles) and 23 African rice accession (triangles)]. Symbol sizes are proportional to the number of accessions and are indicated by the numbers in the rectangle. Colors represent each protein variant: yellow, 7C/long; green, 6C/medium; blue, 5C/medium; and red, 4C/short. (*D*) Awned phenotypes across the four RAE2 protein variants. Numbers of awned (gray bars) and awnless (white bars) accessions for *O. barthii* (*i*), *O. glaberrima* (*ii*), *O. rufipogon/O. nivara* (*iii*), and *O. sativa* (*iv*). (*E*, *Upper*) Nucleotide diversity of *O. sativa* individuals ($n = 67$) relative to nucleotide diversity of *O. rufipogon/O. nivara* individuals ($n = 65$) across chromosome 8. The gray box represents reducing relative diversity surrounding RAE2 (green line), consistent with a selective sweep. (*E*, *Lower*) Genetic distance between *O. sativa* and *O. rufipogon/O. nivara* for all RAE2 types (blue) and dysfunctional ones (red; 4C and 7C). Decreased distance in the dysfunctional class relative to the distance in the "all" class in a 1.5-Mb region surrounding RAE2 (gray box).

of each variant. Only the RAE2-6C type exhibited the awned phenotype, whereas the other three lines did not show awns, regardless of expression level (Fig. 3*A* and *SI Appendix*, Fig. S8). This indicates that the RAE2-6C type is functional for awn elongation but the other three types of protein (RAE2-4C, RAE2-5C, and RAE2-7C) are not functional. We also evaluated *RAE2* variants in two species of Asian wild rice, *Oryza nivara* (accession W0054), which is an annual type of *O. rufipogon* carrying the RAE2-6C type, and *O. rufipogon* (accession W0106), carrying the 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 (*SI Appendix*, Fig. S9*A*) (15). Phenotyping the two lines from these CSSL populations that harbored the *RAE2* locus on chromosome 8 derived from *O. nivara* (WBSL18) or *O. rufipogon* (RSL23) demonstrated that line WBSL18 formed awns but RSL23 did not (*SI Appendix*, Fig. S9*B*). Consistent with our results from overexpression analysis of RAE2 alleles, *O. nivara* harbored the functional 6C/medium RAE2, whereas *O. rufipogon* possessed the dysfunctional 7C/long RAE2 type (*SI Appendix*, Fig. S9*C*).

The geographic distribution of the 130 diverse rice accessions reflects the fact that Asian rice varieties are widely planted around the world, whereas African rice is confined to Africa (Fig. 3*B*). Interestingly, all cultivated African rice varieties (*O. glaberrima*) and the wild ancestor, *O. barthii*, were found to carry the RAE2-6C type, despite the fact that *O. glaberrima* does not have awns and *O. barthii* possesses awns (Fig. 3 *C*, *i* and *ii* and *D*, *i* and *ii*). These results support our previous report that a different gene(s), other than *RAE2*, is responsible for the awnless phenotype in *O. glaberrima* (15). The gene reported to cause awnlessness in African domesticated rice is *RAE3*.

In Asia, dysfunctional RAE2 protein types were present in 32% of *O. rufipogon* and *O. nivara* wild accessions (Fig. 3 *B*, *iii*), despite the fact that almost all individuals possessed awns (Fig. 3 *D*, *iii*). This distinguished the *RAE2* gene from previously reported awn domestication genes, such as *An-1*/*RAE1* 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 the *RAE2* locus and most accessions were awnless (Fig. 3 *C*, *iv* and *D*, *iv*), the five subpopulations of *O. sativa* harbored different frequencies of *RAE2* protein classes (*SI Appendix*, Table S5). Upon examination, it was found that Asian rice cultivars that had the awned phenotype always carried a functional allele at *An-1*/*RAE1* and/or *RAE2*.

To test for evidence of a selective sweep in the region of *RAE2*, the nucleotide diversity (π) of *O. sativa* ($n = 67$) relative to the π of *O. rufipogon* ($n = 65$) was estimated using data from 100-SNP sliding windows across chromosome 8. A drastic decrease was observed in the ratio π *O. sativa*/π *O. rufipogon* across a 1.5-Mb region flanking *RAE2* (Fig. 3*E*, *Upper*), 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 differed from the rest of the chromosome. It was indeed highly unusual ($P < 2.2 \times 10^{-16}$). 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 classes (4C and 7C) relative to d in all functional (6C) RAE2 types (Fig. 3*E*, *Lower*). This result is consistent with recent gene flow back from cultivated to wild Asian rice (31).

RAE2 Is Cleaved Specifically in the Spikelet. The EPFL family of peptides typically requires posttranslational cleavage of a propeptide to become a mature peptide (*SI Appendix*, Fig. S10*A*) (32, 33). To test whether RAE2 is cleaved, we generated transgenic plants carrying *pACT*::*RAE2-3×FLAG*. Immunoblot analysis

demonstrated that RAE2 was cleaved into an ~11-kDa peptide only in the spikelet but not in the other organs (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S10*B*). To confirm this spikelet-specific cleavage of RAE2, we conducted an in vitro processing assay by mixing the recombinant RAE2 propeptide-fused 3×FLAG tag (RAE2-pro) with each of the protein extracts from various organs (callus, stem, leaf, and spikelet). The result of immunoblot analysis using the FLAG antibody showed that only spikelet extract could cleave RAE2-pro (Fig. 4*C* and *SI Appendix*, Fig. S10*C*). In addition, we proved that the ~11-kDa band contained the C-terminal region of RAE2 by using an anti-RAE2 antibody made at the end of the RAE2 C terminus (149-RDRLFDP-125) (*SI Appendix*, Fig. S10*D*). Furthermore, the cleavage of RAE2-pro was inhibited by a protease inhibitor mixture, Complete (*SI Appendix*, Fig. S10*E*). Together, these data suggest that the 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 (SLP) family (listed in *SI Appendix*, Table S6). This protease family has been shown to cleave EPF/EPFL peptides in *A. thaliana* [*SDD1* (34); *CRSP* (35)]. Among the 63 proteases examined, we identified one promising candidate, *Os01g0702300*, which was specifically expressed around the young inflorescence according to RiceXPro (ricexpro.dna.affrc.go.jp) (*SI Appendix*, Fig. S11*A*). We confirmed this expression pattern experimentally (*SI Appendix*, Fig. S11 *B* and *C*) and named this gene *SLP1*. An analysis of phylogenetic relationships classified this gene in the same clade as CRSP (36), supporting the hypothesis that *SLP1* is the most promising candidate gene targeting RAE2. To determine whether the RAE2 protein could be cleaved by SLP1, we performed an in vitro processing assay using a synthetic RAE2 short peptide (synRAE2) spanning the predicted cleavage site (52-AGEEEKVRLGSSPPSCYSK-70) based on Stomagen (27) and an EPF2 (35) with an in vitro SLP1 protein. As a result, three cleaved peptides were detected (Fig. 4*D* and *SI Appendix*, Fig. S12*A*). Cleavage of synRAE2 was documented between amino acid positions G53 and E54, P65 and S66, or both sites by SLP1. This suggested that SLP1 cleaves two positions of this peptide; however, the shorter version would work as a ligand and, according to the Stomagen and EPF2 reports, the site between P65 and S66 is most appropriate as the cleavage site. In addition, we tested whether SLP1 digested a series of mutated synRAE2 peptides 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 (*SI Appendix*, Fig. S12*B*). Furthermore, the alanine substitution closer to the cleavage point repressed RAE2 cleavage using mutated RAE2-pro (mu-RAE2 1–4) (Fig. 4*E*). These results suggested that SLP1 cleaves RAE2 between amino acid positions P65 and S66, and that cleavage occurs specifically in the spikelet (Fig. 4*F*).

Discussion

Our study identifies *RAE2* as *Os08g0485500*, a previously unidentified EPFL gene that is preferentially expressed during early panicle development to promote awn elongation. *RAE2* harbors a conserved six-cysteine peptide region characteristic of the EPF/EPFL family, whereas *rae2* alleles, which are found in most awnless rice accessions, have alternative numbers of cysteine residues due to frameshift mutations. Ohki et al. reported that the consensus scaffold stabilized by three pairs of disulfide bonds is structurally required for EPF/EPFL activity in Stomagen (30). They showed that substituting cysteine residues with serine residues in Stomagen led to incorrect disulfide bonds, which resulted in improper conformation. It is likely that the lack of disulfide bonds impeded active conformation. According to this report, rae2 variants without two cysteine residues, or with extra cysteine residues in improper locations, also failed to achieve proper conformation of the RAE2 protein and were not able to promote awn elongation.

We also showed that SLP1 cleaves RAE2 from a propeptide to a 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 (37, 38). We hypothesize that RAE2 may mediate awn elongation through binding to an unidentified rice ERECTA kinase(s). Improper scaffolding of the mature RAE2 peptide resulting from mutations may hinder binding by this receptor(s) and lead to an awnless phenotype. At this moment, we do not know whether 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 is necessary to fully elucidate the role of

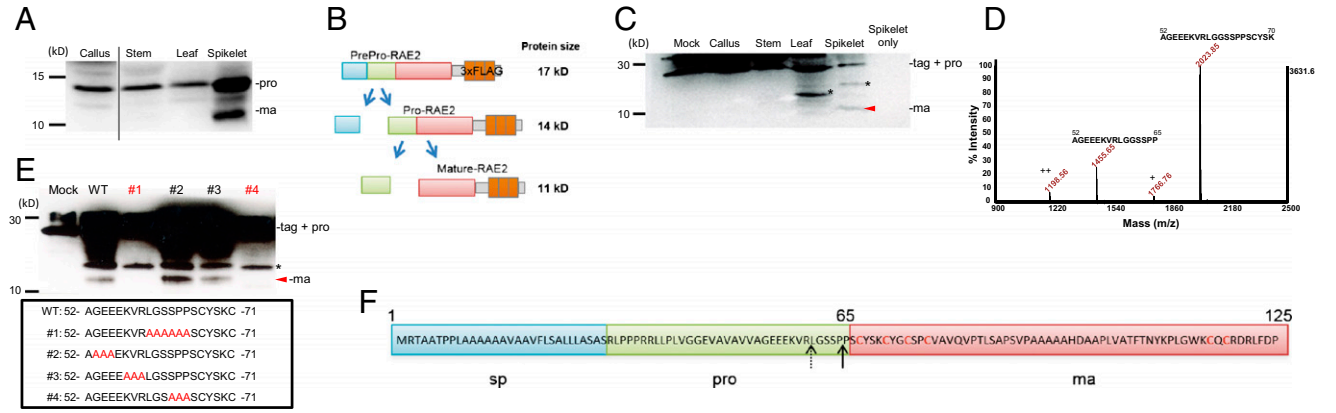

Fig. 4. RAE2 maturation caused by cleavage with SLP1 protease at the spikelet. (*A*) Immunoblot analysis of RAE2-3×FLAG in transgenic plants with an anti-FLAG antibody. The gray line indicates erased space between the callus and stem lane, although all samples were applied on the same membrane. (*B*) The expected size of the RAE2-3×FLAG peptide after cleavage in transgenic plants of overexpression constructs: signal peptide (blue), propeptide (green), and mature peptide (pink). (*C*) In vitro processing assay of recombinant RAE2 peptide incubated with plant extracts of Koshihikari or buffer (mock). The ~30-kD band is a tag-fused recombinant RAE2 propeptide (indicated by –tag+pro). Asterisks indicate nonspecific bands; the red arrowhead indicates the expected mature RAE2-3×FLAG peptide (~11 kDa; indicated by –ma). (*D*) MS ion spectrum for the synthetic peptide (52-AGEEEKVRLGSSPPSCYSK-70), which was cleaved at the position between P65 and S66, indicated by the red line. The full length of the synthetic peptide was cleaved on the other site (+, 54-EEEKVRLGSSPPSCYSK-70; ++, 54-EEEKVRLGSSPP-65). (*E*) In vitro processing assay of a series of alanine-substituted recombinant RAE2 peptides (mu-RAE2) using the spikelet extract of Koshihikari or buffer (mock). The table (*Lower*) shows the amino acid sequence around the predicted cleavage point. Other descriptions are the same as in *C*. (*F*) Predicted sequence of RAE2, which encodes a 125-amino acid peptide composed of a signal peptide (sp; blue), a propeptide (pro; green), and a mature peptide (ma; pink). The dotted and solid arrows indicate the cleavage sites of Stomagen and EPF2, respectively. See also *SI Appendix*, Figs. S10 and S12.

RAE2 cleavage in the spikelet. The observation that transgenic Nipponbare plants transduced with *O. glaberrima RAE2* alleles were able to produce awns demonstrates that cv. Nipponbare retains a functional RAE2 receptor(s). As for the cleavage of RAE2, an in vitro processing assay showed that *O. sativa* also possesses a functional SLP1, and sequence comparisons of SLP1 alleles in *O. sativa* and *O. glaberrima* identified no deleterious frameshift mutations in the coding region of either (*SI Appendix*, Fig. S13). Taken together, the conservation of these downstream components suggests that they are necessary to mediate the interactions with other EPF/EPFL peptides (other than RAE2) and may interact with important, but as yet unidentified, functions in the rice plant.

Because vascular bundles penetrate the center of awns, the *RAE2* gene may function to promote the proliferation of vasculature cells necessary for awn elongation. A similar phenomenon has been reported in *Arabidopsis*; AtEPFL4 and AtEPFL6 coordinate development of inflorescence architecture (28). In a previous mutant study, *DL* and *OsETT2* were reported to be involved in rice awn development (39). *DL* is a 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 reveal the relationship among these components.

Sequence comparison of Asian and African rice suggested that the critical mutations occurred in the 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 some 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, although African rice maintained a functional RAE2 protein, we hypothesize that a mutation(s) in a different locus, designated *RAE3*, represses the awned phenotype in *O. glaberrima* (15). 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 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.

ACKNOWLEDGMENTS. We thank Dr. H. Tsuji for donating the pANDA vector, Dr. T. Nakagawa for providing the pGWB binary vectors, Dr. Y. Sato for help with SEM, and the National BioResource Project for providing the seeds of the GLSL and *O. glaberrima* lines. This research was supported by a Japan Society for the Promotion of Science fellowship, the Integrative Graduate Education and Research Program in Green Natural Sciences of Nagoya University, a MEXT Grant-in-Aid for Scientific Research on Innovative Areas (Grant 16H01464), SATREPS by JICA and JST, Core Research for Evolutional Science and Technology by JST, Global Crop Diversity Trust Grant GS12004, and National Science Foundation Graduate Research Fellowship DGE-1144153 (to D.R.W.).

1. Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the Near East. *Nat Rev Genet* 3(6):429–441.
2. van Heerwaarden J, et al. (2011) 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 W, 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. *Nat Rev Genet* 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. (2006) An SNP caused loss of seed shattering during rice domestication. *Science* 312(5778):1392–1396.
12. Gross BL, Steffen FT, Olsen KM (2010) The molecular basis of white pericarps in African domesticated rice: Novel mutations at the *Rc* gene. *J Evol 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 (Bethesda)* 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]. *Res Rep Tokai Branch Crop Sci Soc Japan* 56(1):11–15. Japanese.
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, Shinjyo C (1996) Genetic studies on an awnness gene *An-4* on chromosome 8 in rice, *Oryza sativa* L. *Breed Sci* 46(4):321–327.
22. Fawcett JA, et al. (2013) QTL map meets population genomics: An application to rice. *PLoS One* 8(12):e83720.
23. Takata N, et al. (2013) Evolutionary relationship and structural characterization of the *EPF/EPFL* gene family. *PLoS One* 8(6):e65183.
24. 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.
25. Hunt L, Gray JE (2009) The signaling peptide EPF2 controls asymmetric cell divisions during stomatal development. *Curr Biol* 19(10):864–869.
26. Kondo T, et al. (2010) Stomatal density is controlled by a mesophyll-derived signaling molecule. *Plant Cell Physiol* 51(1):1–8.
27. Sugano SS, et al. (2010) Stomagen positively regulates stomatal density in *Arabidopsis*. *Nature* 463(7278):241–244.
28. 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.
29. 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):1677–1686.
30. Ohki S, Takeuchi M, Mori M (2011) The NMR structure of stomagen reveals the basis of stomatal density regulation by plant peptide hormones. *Nat Commun* 2:512–517.
31. 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 Phytol* 157(3):657–665.
32. Katsir L, Davies KA, Bergmann DC, Laux T (2011) Peptide signaling in plant development. *Curr Biol* 21(9):R356–R364.
33. Wheeler JI, Irving HR (2012) Plant peptide signaling: An evolutionary adaptation. *Plant Signaling Peptides*, Signaling and Communication in Plants, eds Irving HR, Gehring C (Springer, Berlin), Vol 16, pp 1–23.
34. Von Groll U, 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.
35. Engineer CB, et al. (2014) Carbonic anhydrases, EPF2 and a novel protease mediate CO_2 control of stomatal development. *Nature* 513(7517):246–250.
36. Tripathi LP, Sowdhamini R (2006) Cross genome comparisons of serine proteases in *Arabidopsis* and rice. *BMC Genomics* 7:200.
37. Lee JS, et al. (2012) Direct interaction of ligand-receptor pairs specifying stomatal patterning. *Genes Dev* 26(2):126–136.
38. Lee JS, et al. (2015) Competitive binding of antagonistic peptides fine-tunes stomatal patterning. *Nature* 522(7557):439–443.
39. Toriba T, Hirano HY (2014) The *DROOPING LEAF* and *OsETTIN2* genes promote awn development in rice. *Plant J* 77(4):616–626.