Intron-mediated alternative splicing of *Arabidopsis* *P5CS1* and its association with natural variation in proline and climate adaptation

Ravi Kesari[a], Jesse R. Lasky[b], Joji Grace Villamor[a,1], David L. Des Marais[b], Ying-Jiun C. Chen[a,2,3], Tzu-Wen Liu[a,2,4], Wendar Lin[a], Thomas E. Juenger[b,c], and Paul E. Verslues[a,5]

[a]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan; and [b]Section of Integrative Biology, and [c]Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712

Drought-induced proline accumulation is widely observed in plants but its regulation and adaptive value are not as well understood. Proline accumulation of the *Arabidopsis* accession Shakdara (Sha) was threefold less than that of Landsberg *erecta* (L*er*) and quantitative trait loci mapping identified a reduced function allele of the proline synthesis enzyme Δ^1-pyrroline-5-carboxylate synthetase1 (*P5CS1*) as a basis for the lower proline of Sha. Sha *P5CS1* had additional TA repeats in intron 2 and a G-to-T transversion in intron 3 that were sufficient to promote alternative splicing and production of a nonfunctional transcript lacking exon 3 (exon 3-skip *P5CS1*). In Sha, and additional accessions with the same intron polymorphisms, the nonfunctional exon 3-skip *P5CS1* splice variant constituted as much as half of the total *P5CS1* transcript. In a larger panel of *Arabidopsis* accessions, low water potential-induced proline accumulation varied by 10-fold and variable production of exon 3-skip *P5CS1* among accessions was an important, but not the sole, factor underlying variation in proline accumulation. Population genetic analyses suggest that *P5CS1* may have evolved under positive selection, and more extensive correlation of exon 3-skip *P5CS1* production than proline abundance with climate conditions of natural accessions also suggest a role of *P5CS1* in local adaptation to the environment. These data identify a unique source of alternative splicing in plants, demonstrate a role of exon 3-skip *P5CS1* in natural variation of proline metabolism, and suggest an association of *P5CS1* and its alternative splicing with environmental adaptation.

amino acid metabolism | drought adaptation | stress gene expression | osmoprotectant | compatible solute

Proline acts as an osmoprotectant and cryoprotectant in organisms as diverse as bacteria, plants, and insects (1, 2). Many plants accumulate proline in response to low water potential (ψ_w) and dehydration caused by drought or freezing. The mechanisms by which proline may promote drought resistance include osmoprotectant functions, as well as newly emerging functions of proline metabolism in NADP/NADPH balance and transfer or storage of energy and reducing potential (1–3). However, the overall importance of proline in drought adaptation is not as firmly established. In *Arabidopsis thaliana*, transcriptional up-regulation of Δ^1-pyrroline-5-carboxylate synthetase1 (*P5CS1*) is essential for low ψ_w-induced proline accumulation, and proline accumulation of *p5cs1* mutants is only 15–20% of the wild-type level (4–6). Additional regulation of proline metabolism is likely but not understood. The timing and duration of water limitation, as well as other environmental variables, all influence drought-adaptation strategies used by plants. Thus, we may expect substantial natural variation of these traits within a widely distributed species, such as *Arabidopsis* (7–10), with many of these differences related to the local conditions to which a particular accession has adapted (11–13). Several studies have identified quantative trait loci (QTL) controlling differences in metabolite content (14–18), but such studies have yet to be extended to drought-induced metabolic changes, such as proline accumulation.

Here we used natural variation and QTL mapping of low ψ_w-induced proline accumulation to reveal a unique mechanism whereby intron sequence variation in *P5CS1* was sufficient to drive alternative splicing and production of a nonfunctional *P5CS1* transcript variant, and thus affect the level of P5CS1 protein. This alternative splicing was an important factor underlying variation in proline accumulation across a panel of *Arabidopsis* accessions. Correlation with climate variation, as well as population genetic analysis of *P5CS1* and its alternative splicing, further support *P5CS1* as a gene under selection for environmental adaptation.

Results

Large Effect QTL Controls Differential Proline Accumulation Between *Arabidopsis* Accessions Landsberg *erecta* and Shakdara. Landsberg *erecta* (L*er*) and Shakdara (Sha; also known as Shahdara) differed by three- to fourfold in low ψ_w-induced proline accumulation (Fig. 1A and Fig. S1A). To find the causal genetic polymorphisms, we scored an existing L*er* × Sha recombinant inbred line (RIL) population (19) for proline content at low ψ_w (−0.7 MPa and −1.2 MPa) and identified major QTL on chromosomes 2 and 5 (Fig. 1B, and Tables S1 and S2). We focused further efforts on the chromosome 2 QTL (*Pro-W2*) and narrowed its location to ∼4 cM. Interestingly, this interval contained *P5CS1*, which encodes the first enzyme of proline synthesis. *Pro-W2* may correspond to a proline QTL found under unstressed conditions in Col × C24 (14).

To determine the extent of proline variation controlled by *Pro-W2*, two RILs (56 and 68) (Fig. S2 A and B) were twice backcrossed to L*er* and scored with markers flanking the *Pro-W2* interval or with a marker within *P5CS1* itself (see below) to recover near isogenic lines (NILs), which captured the *Pro-W2* QTL region from Sha in a predominantly L*er* genomic background. These NILs had reduced proline across a range of ψ_w (Fig. 1A), indicating that the *Pro-W2* QTL was responsible for as much as 70% of the variation between L*er* and Sha. Conversely NILs that

PLANT BIOLOGY

Fig. 1. A chromosome 2 QTL causes reduced proline accumulation in Sha. (*A*) Proline accumulation across a range of ψ_w for Ler and Sha seedlings as well as NILs generated by twice backcrossing RILs 56 and 68 (Fig. S2) to Ler. Data are means ± SE, n = 6–10. (*B*) Plot of LOD scores for ψ_w-induced proline of the Ler × Sha RIL population. The gray line is for proline at −0.7 MPa and the darker line is for −1.2 MPa. The dashed line indicates the significance threshold (LOD = 2.5).

were selected for the Ler genotype at *P5CS1* after the second backcross had high proline (Fig. S2C).

Alternative Splicing Leading to Reduced P5CS1 Protein Level Is the Basis for the *Pro-W2* QTL and Low Proline Accumulation of Sha. *P5CS1* was the main candidate gene for *Pro-W2* yet no substantial differences in gene expression of *P5CS1*, or other proline metabolism enzymes, were found between Ler and Sha (Fig. S1B). Sequencing of full-length *P5CS1* cDNA from Ler and Sha found only a silent change of a GAT to GAC codon at Asp-644. However, 5′-RACE of Ler, Sha, and Col-0 identified deletion of exon 3 as an alternative splicing event occurring at substantial frequency in Sha *P5CS1*, but much less frequently in Ler (Fig. 2A) or Col-0. RT-PCR using primers flanking *P5CS1* exon 3 (Fig. 2A, primers a and c) confirmed that Sha had more of the shorter fragment generated from the exon 3-skip transcript than Ler (Fig. 2B). Primers b and c that only amplified the full-length *P5CS1* transcript (Fig. 2A), showed less full-length transcript in Sha (Fig. 2B), but downstream primers indicated similar amounts of total *P5CS1* transcript in both accessions (exon 18 and 3′ UTR) (Fig. 2B), consistent with quantitative PCR results (Fig. S1B). Quantification of the relative intensity of the two bands produced by primers a and c showed that 30% of *P5CS1* transcript in Sha was the exon 3-skip variant compared with 1% in Ler and Col-0.

The exon 3-skip *P5CS1* transcript cannot produce protein starting from the same start codon because deletion of exon 3 changes the reading frame and introduces a stop codon after 53 amino acids. Instead, the exon 3-skip transcript is annotated (TAIR10) as producing a truncated P5CS1, starting from an alternative downstream start codon (Fig. 2A). However, Western blotting with P5CS1-specific polyclonal antisera (Fig. S3A) detected no evidence of this smaller form of P5CS1 and showed reduced level of P5CS1 in Sha compared with Ler (Fig. 2B). A single P5CS1 band was detected, which consistently ran 8–10 kDa above the expected molecular weight of full-length P5CS1 (~87 kDa compared with an expected molecular weight of 77.7 kDa), possibly because of posttranslational modification.

To test whether difference in *P5CS1* was a basis of the *Pro-W2* QTL, we first crossed both Ler and Sha to the null mutant *p5cs1-4* (3, 5). F₁ seedlings of Ler × *p5cs1-4* had proline content only slightly less than that of Ler and significantly higher than Sha or F₁ seedlings of Sha × *p5cs1-4* (Fig. 2C). Western blotting confirmed that both Sha and Sha × *p5cs1-4* had reduced levels of

P5CS1 compared with Ler or Ler × *p5cs1-4* (Fig. 2C). These results indicated that the Sha allele of *P5CS1* had reduced function.

We also transformed Sha with a Ler *P5CS1* genomic clone including 1 kb of promoter (gP5CS1) and transformed *p5cs1-4* with both Sha and Ler gP5CS1. Sha transformed with Ler gP5CS1 had increased proline (Fig. 2D), indicating that *P5CS1* expression was limiting for proline accumulation in Sha. Higher proline and P5CS1 protein levels were observed in *p5cs1-4* transformed with Ler gP5CS1 than with Sha gP5CS1 (Fig. 2D), again indicating reduced function of the Sha allele. In the Sha and *p5cs1-4* complemented lines, the transgene produced much less P5CS1 protein than the endogenous gene (Fig. 2D), consistent with the partial complementation of proline accumulation. Such data were also consistent with experiments in our laboratory where Col-0 *P5CS1* constructs were unable to fully complement *p5cs1* mutants or where transgenic *P5CS1*promoter:reporter fusions were not stress-induced in the same manner as endogenous *P5CS1*. Despite this limitation, the combined data were consistent with variation in *P5CS1* as the main factor underlying the low proline effect of the *Pro-W2* QTL. We cannot, however, exclude the possibility that the *Pro-W2* QTL interval also contains an additional variation that has smaller effects on proline accumulation.

P5CS1 Intron Sequence Variation Is Sufficient to Increase the Frequency of a Nonfunctional Exon 3-Skip Splice Variant. These data indicated that Sha had a reduced function *P5CS1* allele, likely related to alternative splicing. As the *Pro-W2* QTL mapped to *P5CS1* rather than a component of the splicing machinery, the basis for the increased exon 3-skip transcript should be intrinsic to *P5CS1*. Sequencing of the *P5CS1* genomic region found an insertion of four TA-repeats near the 3′ end of intron 2 in Sha (Fig. 2A). Both Ler and Sha had a stretch of TA-repeats, with Sha having more of these repeats than Ler. Sha also had a G-to-T substitution in the 3′ end of intron 3, as well as a 32-bp intron 3 insertion (Dataset S2) (this larger insertion was used to score NILs after the second backcross; see above).

We quantified exon 3-skip *P5CS1* percentage in F₁ seedlings of Sha crossed to *p5cs1-1* or *p5cs1-4* and found that the F₁ seedlings had a similar amount of exon 3-skip *P5CS1* transcript as Sha itself (Fig. 2E). In Sha × *p5cs1-4* F₁ seedlings the percentage of exon 3-skip *P5CS1* was half of the Sha level (13% vs. 27%) (Fig. 2E). This finding was consistent with the *P5CS1* transcript in those F₁ seedlings being an approximately equal mixture of the Sha allele and truncated *P5CS1* transcript containing exon 3 produced by the endogenous Col-0 *P5CS1* [the T-DNA insertion in *p5cs1-4* is in exon 14 (5)]. A higher exon 3-skip percentage was seen in Sha × *p5cs1-1*, in which expression of Col-0 *P5CS1* was reduced because of a promoter T-DNA insertion (5). These data were consistent with variation within *P5CS1*, rather than Sha alleles of other genes that would be heterozygous in the F₁ seedlings, as the basis of increased exon 3-skip *P5CS1* production. Transgenic expression of Ler or Sha gP5CS1 in *p5cs1-4* also demonstrated that the Sha *P5CS1* allele produced more exon 3-skip *P5CS1* transcript than the Ler allele (Fig. 2F). The difference between the observed (7%) and expected (13–15%) exon 3-skip *P5CS1* in the *p5cs1-4* transgenic plants was likely because the transgenic Sha *P5CS1* was expressed at lower level than the endogenous Col-0 allele.

The lower level of P5CS1 protein produced by Sha gP5CS1 compared with Ler gP5CS1 (Fig. 2D) and our inability to see a shorter P5CS1 isoform by Western blotting suggested that exon 3-skip *P5CS1* transcript does not produce protein. To confirm this finding, we transformed *p5cs1-4* with the full-length *P5CS1* cDNA or the exon 3-skip cDNA. Transgenic lines expressing the full-length cDNA produced P5CS1 protein of the expected size and had increased proline content, but the exon 3-skip cDNA did not produce detectable P5CS1 nor increase proline content (Fig. 2G and Fig. S3B). The exon 3-skip cDNA was also unable to produce YFP-tagged fusion protein (Fig. S3C). These results demonstrated that *P5CS1* of Sha was itself sufficient to increase

Fig. 2. Differences in *P5CS1* intron sequence are the basis of the *Pro-W2* QTL and are sufficient to promote production of nonfunctional exon 3-skip *P5CS1* transcript. (*A*) Diagram of the 5' portion of the two alternative *P5CS1* transcripts detected by 5'-RACE analysis. Boxes indicate exons with dark portions indicating the coding region. Gray boxes indicate possible coding region of the exon 3-skip transcript starting from an alternative downstream ATG. The number of clones of each variant found is indicated along with sequence of the 3' ends of *P5CS1* introns 2 and 3; see Dataset S2 for complete intron 2 and 3 sequence. (*B*) Semi-quantitative RT-PCR using primers indicated in *A* to detect *P5CS1* splice variants before low ψ_w treatment (0 h) or after 10 and 96 h at −1.2 MPa. *Actin2* was used as a control gene. P5CS1 protein in samples from the same treatments was detected by Western blot (50 µg protein loaded). Tubulin antibody was used as a loading control. (*C*) Proline content of Ler × *p5cs1-4* or Sha x *p5cs1-4* F₁ seedlings after 96 h at −1.2 MPa. For both Ler and Sha, reciprocal crosses gave identical results and combined data are shown. Data are means ± SE (*n* = 6–9) with *P* value of the comparison of the two sets of F₁ seedlings indicated. Western blot detection of P5CS1 and tubulin (*Lower*) are also shown. (*D*) Proline and P5CS1 level of plants transformed with *P5CS1* genomic fragments (gP5CS1). Sha was transformed with the Ler genomic *P5CS1* clone (Sha/Ler-gP5CS1) and *p5cs1-4* transformed with either Ler or Sha gP5CS1(*p5cs1-4/gP5CS1*). Note that *p5cs1-4* is in the Col-0 genetic background. Proline data are combined from three-to-five independent T₃ homozygous lines and are means ± SE (*n* = 6–12); Western blot is for one representative transgenic line. (*E*) Exon 3-skip *P5CS1* after 96 h at -1.2 MPa in F₁ seedlings of Ler and Sha crossed to *p5cs1-1* (*1–1*) or *p5cs1-4* (*1–4*). RT-PCR was performed for 24 cycles when both *P5CS1* transcripts were in the linear range of amplification. Quantitation of exon 3-skip percentage is indicated by numbers along bottom of the gel. Reciprocal crosses gave similar results and representative data are shown. (*F*) Percentage of exon 3-skip *P5CS1* transcript produced by Ler or Sha gP5CS1 in *p5cs1-4*. For both gP5CS1 constructs, data shown are means ± SE of three to four independent T₃ homozygous lines. (*G*) Proline content and P5CS1 level for *p5cs1-4* transformed with either the exon 3-skip or full-length (F.L.) *P5CS1* cDNA under control of the 35S promoter (*p5cs1-4/35S:P5CS1*). Proline data are means ± SE (*n* = 6–12) of three-to-four independent T₃ homozygous lines and Western blot is from one representative transgenic line. Western blot of additional transgenic lines is shown in Fig. S3.
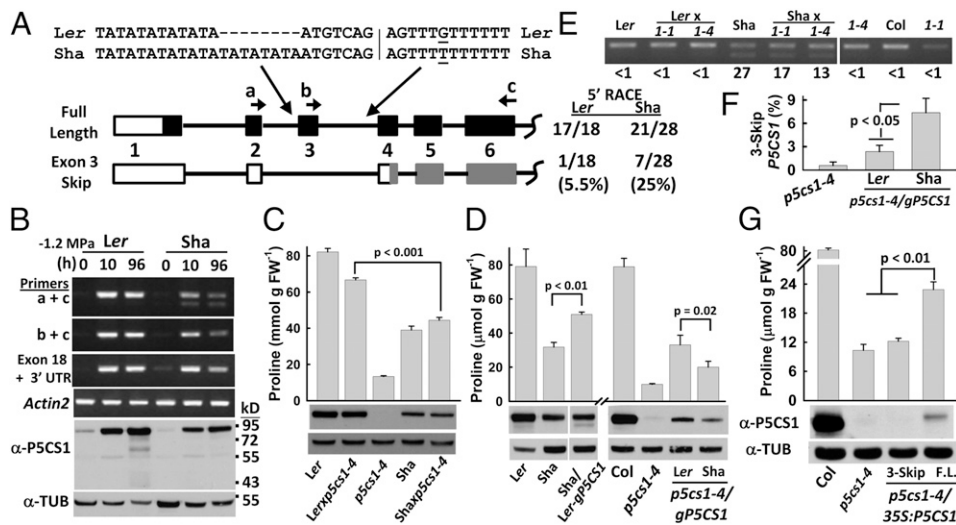
production of a nonfunctional exon-3 skip transcript and thereby decrease P5CS1 protein and proline accumulation.

P5CS1 Alternative Splicing Contributes to Proline Variation Among *Arabidopsis* Accessions.

We found a negative relationship between proline and exon 3-skip *P5CS1* across a large set of accessions. Proline content at low ψ_w (−1.2 MPa) varied nearly 10-fold between the highest and lowest accessions with a mean and median of ~65 micromoles per gram fresh weight (µmol·g·F. W.⁻¹) (Fig. 3*A* and Dataset S1). For 140 accessions, we isolated RNA from seedlings after 96 h at −1.2 MPa to quantify the percentage of exon 3-skip *P5CS1* transcript and also used PCR to determine the size of their intron 2 insertion relative to Ler. Most (129 accessions) had either no or a small intron 2 insertion (−2 to +4) (Fig. 3*B*) and a low or nondetectable level of exon 3-skip *P5CS1*. A smaller number (nine accessions) had intron 2 insertions of similar size as Sha (+6 to +8) (Fig. 3*B*) and a high percentage of exon 3-skip *P5CS1*. In between these two groups were 14 accessions that had an intermediate level of exon 3-skip *P5CS1* but where an intron 2 insertion could not be detected by gel-based scoring (outliers in the −2 to +4 category of Fig. 3*B*). The differences in both percent exon 3-skip *P5CS1* and proline accumulation between the −2 to +4 and +6 to +8 groups of accessions were highly significant (ANOVA on log-percent exon 3-skip *P5CS1*+1; R^2 = 0.46, P < 10⁻¹⁶; log proline abundance; R^2 = 0.06, P = 0.004).

Sequencing of *P5CS1* introns 2 and 3 from a subsample of accessions showed that most low exon 3-skip accessions had the same intron 2 and 3 sequence as Ler (Fig. 3*C* and Dataset S2). Some accessions having an intermediate level of exon 3-skip *P5CS1* (outliers in the −2 to +4 category of Fig. 3*B*) also had the same sequence as Ler. Other accessions had insertion of one or two TA repeats in intron 2 but lacked the intron 3 G-to-T transition. A few accessions (Be-0, Ts-5, Tsu-1, and N4) (Fig. 3*C*) had no or small intron 2 insertion coupled with the intron 3 G-to-T transversion found in Sha. These accessions also had a relatively low level of exon 3-skip *P5CS1*. The data indicated that neither

a small intron 2 TA-insertion nor the intron 3 polymorphism by itself was sufficient to cause frequent exon 3-skip transcript formation. However, the high exon 3-skip *P5CS1* accessions (with exception of Pt-0) all had both the intron 3 G-to-T transition and insertion of three or four TA repeats in intron 2 (Fig. 3*C*). Thus, it was the combination of these intron 2 and 3 changes that was associated with the highest levels of exon 3-skip *P5CS1* formation. These data also suggested how an accession could transition from low to high levels of exon 3-skip *P5CS1* transcript by adding TA repeats in intron 2 and acquiring the intron 3 G-to-T transversion. The 32-bp insertion in Sha intron 3 was not found in any other accession, indicating that it was not associated with alternative splicing. Western blotting of several accessions indicated that increased frequency of nonfunctional exon 3-skip *P5CS1* transcript was associated with decreased levels of P5CS1 protein (Fig. 3*D*).

Consistent with increased exon 3-skip *P5CS1* leading to reduced P5CS1 protein, there was a significant negative correlation between the frequency of exon 3-skip *P5CS1* transcript and proline content in seedlings at low ψ_w (Fig. 3*E*), both when analyzing raw data (ANOVA, R^2 = 0.09, P = 0.0004) and log-proline content vs. log exon 3-skip *P5CS1* percentage + 1 (ANOVA, R^2 = 0.11, P = 0.00006). In accessions where the proportion of exon 3-skip *P5CS1* was 6–8% percent or higher, it was the dominating factor as all of these accessions had low or moderate levels of proline accumulation.

Although variation in percentage exon 3-skip *P5CS1* had a substantial influence on proline accumulation, it was not the only factor. Accessions with low or nondetectable exon 3-skip *P5CS1* covered the range from high to low proline accumulation (Fig. 3*E*). Some of the accessions with the highest and lowest levels of proline did not differ in P5CS1 (Fig. S4). In addition, Pt-0 was an interesting outlier accession in that it had a high level of exon 3-skip *P5CS1* without the intron 2 insertion (Fig. 3*B*). Pt-0 also had a low level of total *P5CS1* transcript and a nearly undetectable level of P5CS1 protein (Fig. S4). Pt-0 had the lowest proline content of any accession in our panel, comparable to *p5cs1-4*. Thus, Pt-0 is essentially a naturally occurring *p5cs1*-deficient mutant.

PLANT BIOLOGY
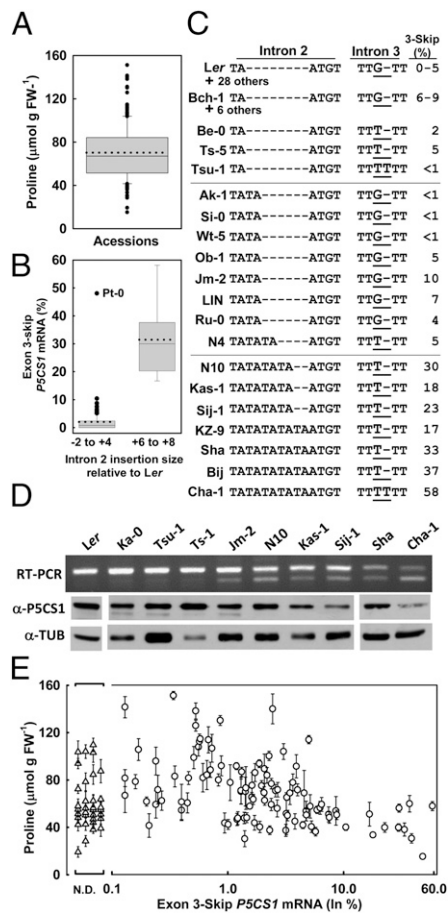
Fig. 3. Alternative splicing of *P5CS1* is a major factor underlying variation in low ψ_w-induced proline accumulation between *Arabidopsis* accessions. (*A*) Proline contents of 180 *Arabidopsis* accessions after 96-h low ψ_w (−1.2 MPa) treatment. Box encompasses the 25–75 percentiles with the solid line indicating the median and dashed line the mean proline content of all accessions. Whiskers show the 10–90 percentiles and dots indicate outliers. Proline data for each accession is listed in Dataset S1. (*B*) Percentage of exon 3-skip *P5CS1* mRNA as related to size of *P5CS1* intron 2 insertion. Percentage of exon 3-skip *P5CS1* was determined for seedlings exposed to −1.2 MPa for 96 h. The intron 2 insertion size was estimated by PCR and gel electrophoresis. Presentation of data are as described in *A*. *n* = 129 for the −2 to +4 insertion size and *n* = 9 for +6 to +8. Exon 3-skip percentage and intron 2 insertion size for each accession are listed in Dataset S1. (*C*) *P5CS1* intron 2 and 3 sequences of selected accessions having varying intron 2 insertion size and exon 3-skip *P5CS1* percentage. Complete intron 2 and 3 sequence alignments can be seen in Dataset S2. (*D*) Exon 3-skip *P5CS1* and P5CS1 protein level of selected accessions. (*Upper*) RT-PCR using primers a and c (Fig. 2A). (*Lower*) Western blot detection of P5CS1 and tubulin (loading control). (*E*) Relationship of Pro accumulation at −1.2 MPa to percentage of exon 3-skip *P5CS1* mRNA. Triangles indicate accessions where the exon 3-skip *P5CS1* mRNA could not be detected (N.D., not detected).

We also analyzed the association of *P5CS1* sequence variation with exon 3-skip percentage and proline content using publicly available SNP datasets (13) as a way to control for population structure and minimize false-positive associations. Average genome-wide SNP similarities between accessions were used to infer population structure. A sliding window across *P5CS1* and 10 kb on either side was used to characterize haplotype variation in 5-SNP intervals. Accessions sharing L*er* and Sha 5-SNP haplotypes differed significantly in exon 3-skip *P5CS1* and proline at seven haplotype windows while accounting for kinship (linear mixed model on log-percent exon 3-skip *P5CS1* +1, α = 0.05). Sha haplotypes (*n* = 8) had 16% more exon 3-skip *P5CS1* than L*er* (*n* = 53) haplotypes at the most divergent 5-SNP interval (*P* < 10⁻⁵). The

same 5-SNP interval was also where accessions with L*er* (*n* = 67) vs. Sha (*n* = 8) haplotypes differed most in proline, with L*er* haplotypes averaging 20 μmol·g F.W.⁻¹ more proline than Sha haplotypes (linear mixed model on log proline abundance, *P* = 0.01). More generally, we found that within this interval, all unique 5-SNP haplotypes with more than one accession were significantly different in percent exon 3-skip *P5CS1* (*n* = 100 accessions, *P* < 10⁻⁶) and proline abundance (*n* = 129 accessions, *P* = 0.004).

Exon 3-Skip *P5CS1* Abundance Is Associated with Climate Variation. To understand the importance of proline and *P5CS1* splicing in a broader context, we asked whether variation in exon 3-skip *P5CS1* or proline was associated with environmental variation while statistically controlling for population structure (12, 20). Spatial variation in the exon 3-skip *P5CS1* percentage was strongly nonrandom across the native Eurasian range of *Arabidopsis*; the percentage of exon 3-skip *P5CS1* increased in eastern Eurasia, with longitude explaining 28% of (log-percent + 1) variation independently of a SNP-based kinship matrix (a proxy for population structure; EMMA, *P* < 0.0001) (Fig. S5). Spatial variation in exon 3-skip *P5CS1* percentage was strongly associated with climate and the second principal component (PC2) of climate across Eurasia (explaining 21% of climatic variation) (Dataset S3 and *SI Materials and Methods*). The percent of exon 3-skip *P5CS1* decreased along PC2 (Fig. 4*A*) as climates became wetter and had less temporal variation in temperature (explaining 16% of variation in exon 3-skip *P5CS1*; EMMA, *P* = 0.03). Conversely, proline accumulation increased with PC2 (explaining 7% of proline variation), indicating greater proline accumulation in locations with less variable temperature and more precipitation (Fig. 4*B*).

When specific climate variables were tested, significant associations (α = 0.05) with exon 3-skip *P5CS1* percentage, independent of kinship, were found for 39 of 101 variables tested (Dataset S4). Diurnal temperature range had the strongest loading on PC2 and one of the strongest associations with exon 3-skip *P5CS1* percentage, explaining 32% of its variation (Fig. 4*C*). Proline content was also significantly correlated to some climatic variables, although associations were not as strong as for exon 3-skip *P5CS1* percentage (11 of 101 climate variables had *P* < 0.05) (Dataset S5). For example, mean diurnal temperature range showed the opposite trend for proline compared with exon 3-skip *P5CS1* percentage but with a weaker association (explaining only 6% of the proline variation) (Fig. 4*D*). Although log transformation reduced the skewness of exon 3-skip *P5CS1* percentage, residuals were not normal and thus the associated *P* values are approximate. The opposite relationships of exon 3-skip *P5CS1* percentage and proline accumulation with variables such as climate PC2 is consistent with the molecular data, showing that increasing exon 3-skip *P5CS1* transcript leads to reduced P5CS1 protein and reduced proline accumulation.

***P5CS1* Shows a Signature of Natural Selection for Local Adaptation to Climate.** *P5CS1* also had a population genetic signature consistent with an involvement in local adaptation to the environment. Two lines of evidence suggest that neutral evolution at *P5CS1* can be rejected and that adaptation to climate factors may drive patterns of genetic variation at this locus. First, we found a slight excess of segregating nonsynonymous substitutions in the protein-coding domain of *P5CS1* (McDonald–Kreitman test marginally nonsignificant; G = 2.915, *P* = 0.087). Excess nonsynonymous substitutions suggest either that *P5CS1* is a recurrent target of selection across the range of *Arabidopsis* or that selection on the intron 2 and 3 polymorphisms associated with high exon 3-skip *P5CS1* has reduced the strength of purifying selection on tightly linked variants. Because tests based on models of sequence evolution can be sensitive to the effects of population structure, we also used the test of Toomajian et al. (21) that compares the haplotype structure at a given locus to a genome-wide distribution of haplotype lengths. This test provides evidence that selection has acted to increase the frequency of a
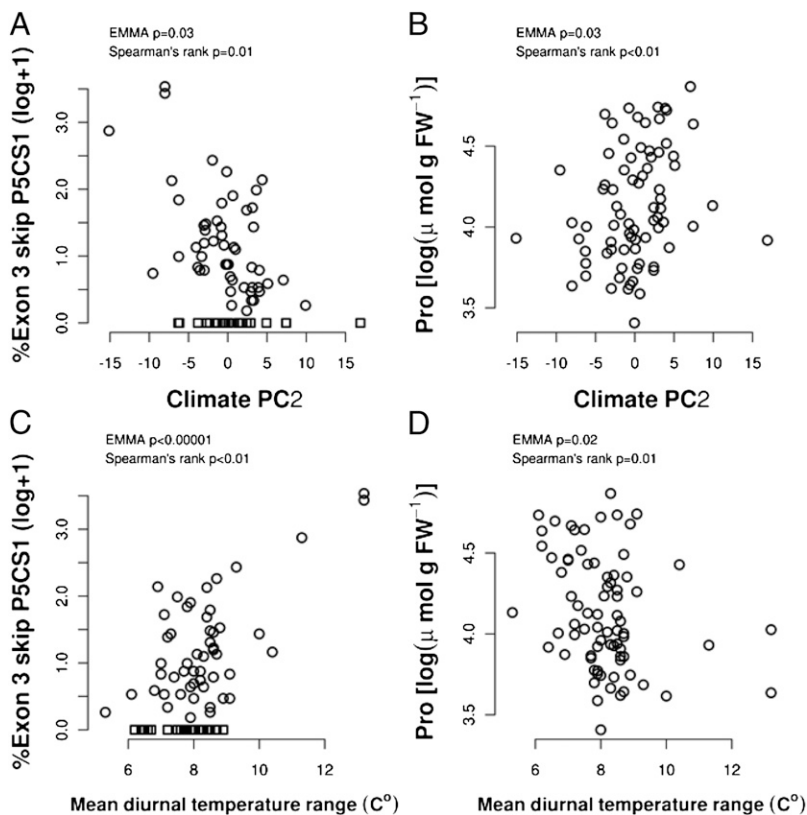
Fig. 4. Association of exon 3-skip *P5CS1* percentage and proline content with climatic factors. (*A*) Negative association of exon 3-skip *P5CS1* percentage with the PC2 of climate (*SI Materials and Methods* and Dataset S3). Accessions where exon 3-skip *P5CS1* could not be detected are shown as squares. Only accessions with SNP data and reliable collection locations (*SI Materials and Methods*) are included (*n* = 76). (*B*) Positive association of proline content at low ψ_w with climate PC2. (*C*) Increased frequency of exon 3-skip *P5CS1* in accessions from more variable temperature environments. Mean diurnal temperature range was the climate variable most strongly associated with exon 3-skip *P5CS1* percentage (Dataset S4). (*D*) Proline accumulation had an opposite, but weaker, correlation to mean diurnal temperature range and an overall weaker association with climate variables (Dataset S5).

P5CS1 haplotype in the recent past (pair-wise haplotype sharing score = 1.834, *P* = 0.059). Moreover, this SNP has a significantly elevated value of Wright's fixation index for subdivided populations (F_{ST} = 0.265, *P* = 0.024), suggestive of local adaptation at the *P5CS1* locus, and also shows significant associations with measures of aridity (ρ = 0.143, *P* = 0.0041) and light intensity (measured as photosynthetically available radiation; ρ = 0.137, *P* = 0.0027). These latter analyses are derived from previously reported genome-wide scans of European samples (12). These patterns may be because of the intron polymorphism we characterized or additional polymorphisms. In either case, the combined results suggest that *P5CS1* is evolving nonneutrally.

Discussion

Alternative splicing has been shown to be a factor in plant response to the environment (22–25). In such reports, it has been variation in the expression or activity of splicing-related proteins that leads to alternative splicing of target genes. In contrast, sequence variation in *P5CS1* introns 2 and 3 was sufficient to change the *P5CS1* splicing pattern. Efficient splicing depends on both intron sequence, with properly spaced U-rich and UA-rich regions promoting intron recognition, and on optimal spacing between the 3′ end of the intron and the branch point (23, 26, 27). For *P5CS1*, extra intron 2 TA repeats may lead to less efficient recognition of intron 2, while the intron 3 G-to-T transition enhances recognition of that intron. The combined effect is to shift the balance toward splicing of exons 2 and 4 rather than exons 2 and 3. This hypothesized mechanism is illustrated in Fig. S6.

We are not aware of reports of similar intron polymorphisms causing alternative splicing in plants. In humans, TG and T insertions in the 3′ end of intron 8 of cystic fibrosis transmembrane regulator (CFTR) lead to skipping of exon 9 and production of a nonfunctional transcript associated with cystic fibrosis (28–30). Specific splicing factors promote exon skipping and splicing factor variation affects penetrance of disease-associated CFTR alleles (31, 32). This finding is qualitatively similar to our data where as much as 10% variation in exon 3-skip *P5CS1*

percentage can be seen for accessions with the same intron 2 and 3 sequence. This finding suggests that, in addition to the specific polymorphisms identified here, recognition and splicing of *P5CS1* intron 2 may be influenced by natural variation in splicing machinery or RNA processing. However, because plant-splicing mechanisms are not well understood, how similar *P5CS1* is to the well-studied *CFTR* example is not certain.

The value of proline in drought adaptation is a question of both interest and uncertainty in plant stress biology. The climate associations and population genetic analysis presented here suggest an association of proline metabolism, and *P5CS1* specifically, with climate adaptation. Although these analyses are correlational and cannot by themselves establish cause and effect, they do raise questions about the role of proline metabolism. Particularly interesting is the overall stronger association of exon 3-skip *P5CS1* percentage than proline level itself with climate variables, and the observation that higher proline was associated with wetter environments and more stable temperatures. One possible explanation for the latter observation is that accessions adapted to dryer or more variable climates have acquired additional stress-adaptive metabolic mechanisms to supplement proline accumulation. An example of this type of adaptation was observed in the *Plumbaginaceae* family, where species adapted to chronically dry environments convert proline to the more potent osmoprotectant proline-betaine (33). It has often been assumed that more proline accumulation is better in drought resistance; however, the stronger relationship of exon 3-skip *P5CS1* percentage than proline content with climate variation is consistent with the amount of proline synthesis, rather than just the amount of proline accumulation, as a factor in environmental adaptation.

The overall importance of proline metabolism is also indicated by the P5CS1-deficient accession Pt-0. Pt-0 shows that P5CS1 can be reduced or lost without serious consequence for normal growth and development; however, the vast majority of accessions surveyed had little exon 3-skip *P5CS1* and maintained the ability to induce high levels of P5CS1. Further study of this type of natural variation, such as transgenic introgression of naturally occurring

PLANT BIOLOGY

P5CS1 alleles into different genetic backgrounds, promises to shed light on drought-adaptive mechanisms, as well as the novel alternative splicing mechanism exemplified by *P5CS1*.

Materials and Methods

Plant Material and QTL Mapping. *Arabidopsis* accessions and the L*er* × Sha RIL population (19) were obtained from the *Arabidopsis* Biological Resource Center. Details of low ψ_w treatment and QTL mapping can be found in *SI Materials and Methods*.

P5CS1 Alternative Splicing and Protein Detection. RT-PCR and Western blot assays were performed using standard methods with primers and antisera described in *SI Materials and Methods* and Table S3.

Transgenic Plants. Constructs and transgenic lines were generated as described in *SI Materials and Methods*. Data presented is from three to five homozygous T$_3$ lines derived from independent transformation events, unless otherwise noted.

Climate Associations, Population Genetics, and P5CS1 Haplotypes. We tested 101 climate variables and the principal components of those variables for association with either exon 3-skip *P5CS1* percentage or proline accumulation (*SI Materials and Methods*). Associations were tested while accounting for population structure among accessions (EMMA) (20) or nonparametrically, ignoring population structure (Spearman's rank correlation), and only included accessions for which 250 K SNP data are available (12) to estimate population structure and reliable collection location information are available (34).

Population genetic analyses were completed on the aligned coding sequences of *P5CS1* from 80 accessions from the 1001 Genomes Project (35) using *Arabidopsis lyrata P5CS1* coding sequence as an outgroup (*SI Materials and Methods*). Pair-wise haplotype sharing, F_{st}, and aridity/PAR patterns come from previously reported SNP datasets (12).

Identification of *P5CS1* haplotypes and accessions representing different haplotypes are described in *SI Materials and Methods*.

1. Szabados L, Savouré A (2010) Proline: A multifunctional amino acid. *Trends Plant Sci* 15:89–97.
2. Verslues PE, Sharma S (2010) Proline metabolism and its implications for plant-environment interaction. *Arabidopsis Book* 8:e0140.
3. Sharma S, Villamor JG, Verslues PE (2011) Essential role of tissue-specific proline synthesis and catabolism in growth and redox balance at low water potential. *Plant Physiol* 157:292–304.
4. Sharma S, Verslues PE (2010) Mechanisms independent of abscisic acid (ABA) or proline feedback have a predominant role in transcriptional regulation of proline metabolism during low water potential and stress recovery. *Plant Cell Environ* 33:1838–1851.
5. Székely G, et al. (2008) Duplicated *P5CS* genes of *Arabidopsis* play distinct roles in stress regulation and developmental control of proline biosynthesis. *Plant J* 53:11–28.
6. Yoshiba Y, et al. (1995) Correlation between the induction of a gene for delta 1-pyrroline-5-carboxylate synthetase and the accumulation of proline in *Arabidopsis thaliana* under osmotic stress. *Plant J* 7:751–760.
7. Verslues PE, Juenger TE (2011) Drought, metabolites, and *Arabidopsis* natural variation: A promising combination for understanding adaptation to water-limited environments. *Curr Opin Plant Biol* 14:240–245.
8. Bouchabke O, et al. (2008) Natural variation in *Arabidopsis thaliana* as a tool for highlighting differential drought responses. *Plos One* 3:e1705.
9. Des Marais DL, Juenger TE (2010) Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Ann N Y Acad Sci* 1206:56–79.
10. McKay JK, et al. (2008) Genetics of drought adaptation in *Arabidopsis thaliana* II. QTL analysis of a new mapping population, KAS-1 x TSU-1. *Evolution* 62:3014–3026.
11. Fournier-Level A, et al. (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89.
12. Hancock AM, et al. (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334:83–86.
13. Horton MW, et al. (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nat Genet* 44:212–216.
14. Lisec J, et al. (2008) Identification of metabolic and biomass QTL in *Arabidopsis thaliana* in a parallel analysis of RIL and IL populations. *Plant J* 53:960–972.
15. Lisec J, et al. (2009) Identification of heterotic metabolite QTL in *Arabidopsis thaliana* RIL and IL populations. *Plant J* 59:777–788.
16. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* 20:1199–1216.
17. Wentzell AM, et al. (2007) Linking metabolic QTLs with network and *cis*-eQTLs controlling biosynthetic pathways. *PLoS Genet* 3:1687–1701.
18. Fu J, et al. (2009) System-wide molecular evidence for phenotypic buffering in *Arabidopsis*. *Nat Genet* 41:166–167.
19. Clerkx EJM, et al. (2004) Analysis of natural allelic variation of *Arabidopsis* seed germination and seed longevity traits between the accessions Landsberg *erecta* and Shakdara, using a new recombinant inbred line population. *Plant Physiol* 135:432–443.
20. Kang HM, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723.
21. Toomajian C, et al. (2006) A nonparametric test reveals selection for rapid flowering in the *Arabidopsis* genome. *PLoS Biol* 4:e137.
22. Matsukura S, et al. (2010) Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Mol Genet Genomics* 283:185–196.
23. Reddy ASN, Day IS, Göhring J, Barta A (2012) Localization and dynamics of nuclear speckles in plants. *Plant Physiol* 158:67–77.
24. Sanchez SE, et al. (2010) A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* 468:112–116.
25. Sugliani M, Brambilla V, Clerkx EJM, Koornneef M, Soppe WJJ (2010) The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in *Arabidopsis*. *Plant Cell* 22:1936–1946.
26. Cellini A, Felder E, Rossi JJ (1986) Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3′ splice site. *EMBO J* 5:1023–1030.
27. Reddy ASN (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* 58:267–294.
28. Hefferon TW, Groman JD, Yurk CE, Cutting GR (2004) A variable dinucleotide repeat in the CFTR gene contributes to phenotype diversity by forming RNA secondary structures that alter splicing. *Proc Natl Acad Sci USA* 101:3504–3509.
29. Niksic M, Romano M, Buratti E, Pagani F, Baralle FE (1999) Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum Mol Genet* 8:2339–2349.
30. Pagani F, et al. (2000) Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem* 275:21041–21047.
31. Buratti E, Brindisi A, Pagani F, Baralle FE (2004) Nuclear factor TDP-43 binds to the polymorphic TG repeats in CFTR intron 8 and causes skipping of exon 9: A functional link with disease penetrance. *Am J Hum Genet* 74:1322–1325.
32. Buratti E, et al. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* 20:1774–1784.
33. Hanson AD, et al. (1994) Osmoprotective compounds in the Plumbaginaceae: A natural experiment in metabolic engineering of stress tolerance. *Proc Natl Acad Sci USA* 91:306–310.
34. Anastasio AE, et al. (2011) Source verification of mis-identified *Arabidopsis thaliana* accessions. *Plant J* 67:554–566.
35. Cao J, et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43:956–963.