Supporting Information

Kesari et al. 10.1073/pnas.1203433109

SI Materials and Methods

Quantitative Trait Loci Mapping. Quantitative trait loci (QTL) mapping used the previously described Landsberg *erecta* (L*er*) × Shakdara (Sha) recombinant inbred line (RIL) mapping population and a genetic map composed of 89 genetic markers (1). The 117 RILS were divided into three randomized blocks. Each RIL was plated on two agar plates for low water potential treatment and four samples (two from each plate) collected for proline analysis. The entire experiment was then repeated with the population divided into different randomized blocks giving a total of eight proline measurements per line. QTL associated with measured proline abundances were mapped using a combination of interval and composite interval mapping (CIM) (2) as implemented in the R statistics package R/QTL (3). Our approach centered on initial genome scans using an interval mapping algorithm to identify candidate QTL regions as expressed in each environmental condition. These scans were subsequently followed by more detailed multiple QTL models as implemented in CIM. CIM tests the hypothesis that an interval flanked by two adjacent markers contains a QTL affecting a phenotype, but statistically accounts for the effects of additional QTL outside the test interval. CIM is expected to improve power and localization of QTL for traits segregating multiple QTL. In each case, a genome-wide critical threshold value for the experiment-wise type 1 error rate was be set for each phenotype by randomly permuting the line means among genotypes 1,000 times and using empirical false-positive rates (4–6). In each case, mapping was conducted on a 1-cM grid using a Haley–Knott regression approximation and assuming normally distributed phenotypes. Given the relatively simple genetic architecture observed from interval mapping, we implemented conservative CIM models fitting the number of markers detected in interval mapping as background cofactors for each phenotype and using a window size of 15 cM. All mapping was conducted on recombinant inbred line least-squares means after controlling for blocking in the experimental designs using the Proc Mixed procedure in SAS (7). We searched for pair-wise epistatic interactions between QTL using the stepwise scan algorithm in R/QTL (8). However, these scans revealed no strong support for interactions and so we present only our simpler additive QTL results.

We calculated the additive effect ($2a$) of each QTL indentified from our final CIM models as the difference between the L*er* and Sha homozygotes at a particular QTL position for each trait in each environment. Positive-additive effects indicate that the Sha parental genotype has the higher mean. We estimated confidence intervals for the location of QTL using a standard 1.5-LOD score drop on the results from our CIM models. Scripts detailing our QTL mapping approach are available upon request.

Analysis of *P5CS1* alternative splicing by 5′-RACE. Total RNA isolated from L*er* and Sha seedlings after 10 h of −1.2 MPa water stress treatment using RNeasy Plant mini kit (Qiagen) was used for 5′-RACE. The 5′-RACE-Ready cDNA were synthesized by using the SMARTer RACE cDNA Amplification Kit (Clontech). Briefly, the synthesis of cDNA for 5′-RACE was performed in a final volume of 10 μL containing 1 μg of total RNA, 1.2 μM lock-docking oligo(dT) primer [5′-RACE CDS Primer A: 5′-$(T)_{25}$V N-3′, where n = A, C, G, or T and V = A, G, or C], 1.2 μM SMARTer IIA oligo (5′-AAG CAG TGG TAA CAA CGC AGA GTA CXX XXX-3′, where X = undisclosed base in the proprietary SMARTer oligo sequence), 1× first-strand buffer, 2 mM DTT, 1 mM each of dNTPs, 10 Units of RNase inhibitor and 10 units of SMARTScribe reverse transcriptase. After the reaction, the 5′-RACE-Ready first strand cDNAs were diluted six times in 10 mM Tricine-1 mM EDTA buffer and used as template for 5′-RACE. The 5′-RACE PCR was carried out in a final volume of 20 μL containing 1× Advantage 2 PCR buffer, 0.25 mM each of dNTPs, 1 μL of 5′-RACE-Ready cDNA, 1× universal primer A mix (long: 5′-CTA ATA CGA CTC ACT ATA GGG C AA GCA GTG GTA TCA ACG CAG AGT-3′; short: 5′-CTA ATA CGA CTC ACT ATA GGG C-3′), 0.5 μM *P5CS1* gene-specific primer (P5CS1 5′-RACE GSP1) (Table S3) and 1× Advantage 2 Polymerase mix. The PCR products were then cloned into the pJET1.2 vector (Fermentas) using CloneJet PCR cloning kit and clones were sequenced using vector and gene specific primers. Sequence analysis of clones based on the position of SMARTer IIA oligo suggested that transcriptional start site of *P5CS1* differed from the TAIR annotation and instead was consistent with the RIKEN RAFL cDNA clone (RAFL05-20–023, National Center for Biotechnology Information accession no. AY080771). Besides exon 3-skip *P5CS1*, the only other *P5CS1* alternative splicing detected was a single clone with retention of intron 17 in Col-0. This form of *P5CS1* could not be detected by RT-PCR, suggesting that it was a very low-frequency event.

Semiquantitative RT-PCR. Total RNA was extracted from seedlings of L*er* and Sha ecotypes after 0, 10, and 96 h of −1.2 MPa water stress treatment using RNeasy Plant mini kit (Qiagen). Semiquantitative RT-PCR experiments were carried out with cDNAs generated from 1 μg of total RNA using the SuperScript III Reverse Transcriptase (Invitrogen). The PCR was carried out within the exponential range of the amplification to allow comparison of cDNA amounts. *Actin2* was used as a reference gene. The primers used in the semiquantitative PCR are shown in the primer table (Table S3).

Quantification of Exon 3-Skip *P5CS1* Transcript. Total RNA were extracted from seedlings of various ecotypes or *p5cs1-4* lines complimented with L*er* or Sha *P5CS1* genomic fragment (see below) after 96 h of −1.2 MPa water stress treatment using RNeasy Plant mini kit (Qiagen). One microgram of total RNA was used for first strand cDNA synthesis using the SuperScript III Reverse Transcriptase (Invitrogen). The resulting cDNA was used as DNA template for amplification of exon 3-skipped/exon 3-retained *P5CS1* fragments using a forward primer in exon 2 and reverse primer in exon 6 (Table S3). Twenty-four cycles of the PCR was used which was within exponential range of the amplification. Amplified products were run on the agarose gel and the unsaturated gel images were scanned using a Bio-Rad Gel Doc EQ system and Quantity One v4.5.0 software. The band density of the exon 3-skip and full length *P5CS1* fragments were determined using ImageJ software (National Institutes of Health v1.44J).

Genotyping of Intron 2 Insertion Size. For genotyping, DNA were isolated from seedlings of *Arabidopsis* ecotypes using C-TAB extraction and used to amplify the region flanking the 3′ end of *P5CS1* intron 2 using primer LSPDP2 F and LSPDP2 R (Table S3). The resultant fragments were run on a 4% agarose gel and band shift in relation to the L*er*/Sha amplicon was scored and use to determine the size of DNA fragments of other ecotypes. For sequencing to confirm genotyping results of some ecotypes, fragments corresponding to *P5CS1* intron 2, exon 3, and intron 3 were amplified using *P5CS1* exon 2 F and LSPDP R primer and phusion DNA polymerase (Finnzymes). The resulted fragments

were cloned in pJET1.2 (Fermentas) Vector using the CloneJet PCR cloning kit and sequenced.

Cloning, Construct Preparation, and Plant Transformation. For preparation of *35S:P5CS1*$_{exon3skip}$*-eYFP* and *35S:P5CS1-eYFP* plant transformation construct as well as C-terminal FLAG fusions of both cDNAs, full-length or exon 3-skip *P5CS1* were amplified using L*er* or Sha cDNA and moved to gateway entry vector pDONR207 (Invitrogen) by BP reaction. The DNA fragments were then moved to gateway destination vector by LR reaction. The binary vectors pGWB441 and pGWB411 (9) were used as a gateway destination vectors for the expression of recombinant protein in fusion with C-terminal eYFP or FLAG tags, respectively, under the control of 35S promoter. Plant transformation was done by *Agrobacterium* mediated floral-dip method. Wild-type Col or *p5cs1-4* plants were used for the preparation of transgenic plants with eYFP tag. For L*er*/Sha *P5CS1* genomic construct preparation, a ~5.9-kb fragment containing the *P5CS1* proximal promoter, 5′ UTR, ORF, and partial 3′ UTR were amplified using L*er*/Sha genomic DNA as template using phusion DNA polymerase (Finnzymes) and cloned in to pENTR/D-TOPO (Invitrogen). The L*er*/Sha *P5CS1* genomic fragments were moved to plant gateway binary vector pGWB401 (9) using LR reaction. L*er*, Sha, or *p5cs1-4* plants were transformed by *Agrobacterium* mediated floral-dip method. T_3 homozygous lines were used for all measurements with the exception of observation of P5CS1:eYFP fusion protein (Fig. S3), which was done in the T_2 generation.

Expression and Purification of P5CS1 Protein, Antibody Preparation, and Western Blots. *P5CS1* ORF (cloned from Col-0) in pDONR207 was moved to gateway vector pET-301/CT-DEST (Invitrogen) using LR reaction for expression of the P5CS1-6× His recombinant protein in *Escherichia coli*. Rosetta (Merck) *E. coli* cells were transformed with the resulting construct and induced with 0.4 mM isfopropyl-β-D-thiogalactopyranoside for production of recombinant P5CS1-6× His protein. The recombinant protein was expressed primarily as a truncated from of ~55 kDa and was present in inclusion bodies. The recombinant protein was solubilized and then purified using Ni-NTA column and used for antibody generation in rabbit (performed by LTK BioLaboratories).

For Western blotting, proteins were extracted from seedlings as described by Martínez-García et al. (10) with minor modification. Briefly, 50- to 100-mg seedlings were ground to fine powder with liquid N_2 using a microfuge pestle and 100 μL of extraction buffer [125 mm Tris–HCl pH 8.8, 1% (wt/vol) SDS, 10% (vol/vol) glycerol, 1 mM PMSF] added with continued mixing. After centrifugation, the protein content of supernatant was determined using Pierce BCA protein assay kit (Thermo Scientific). Proteins (generally 50 μg of total protein) were separated by SDS-PAGE using 12% gels and wet-transferred to PVDF membrane (Bio-Rad). Membranes were probed with primary antibodies (Anti-P5CS1 at 1:3,000 dilution). Anti-rabbit HRP-conjugated antibody was used as a secondary antibody (Jackson Laboratories). Immunoblots were developed by using Pierce ECL Western blotting substrate (Thermo Scientific). Stripping and reprobing blots of with α-tubulin (T5168; Sigma-Aldrich) was used as a loading control.

Association of SNPs Near *P5CS1* with Proline and Exon 3-Skip *P5CS1* Percentage. We used a mixed model association mapping framework to study whether SNPs within 10 kb of *P5CS1* were associated with proline abundance (log) and exon 3-skip *P5CS1* percentage (log + 1). Haplotypes were constructed at 29 5-SNP intervals generated by sliding a window across SNPs within 10 kb of *P5CS1*. Each unique combination of allele states at the 5 SNPs comprising an interval was considered a haplotype. We compared phenotypes of accessions sharing the L*er* haplotype to those sharing the Sha haplotype in each 5-SNP interval. We again used the kinship method as a technique of controlling for population structure. In this model, proline or exon 3-skip *P5CS1* percentages were response variables, haplotypes were treated as fixed-effect factors, and the kinship matrix was a random effect. We used the R package "kinship" to implement this model.

Of 24 5-SNP intervals around *P5CS1* where L*er* vs. Sha haplotypes were polymorphic, haplotype variation at seven intervals was significantly associated with alternative splicing, while controlling for kinship (linear mixed model on log-percent exon 3-skip *P5CS1* +1, $\alpha = 0.05$). In the interval where accessions with L*er* vs. Sha haplotypes were most divergent, Sha, KZ-9, N4, Ts-5, and four other accessions shared the haplotype associated the greatest exon 3-skip *P5CS1* percentage. With the exception of Ts-5, these named accessions also had relatively long insertions in intron 2 and had identical sequences at intron 3 (Fig. 3*C*). L*er*, Bch-1, Tsu-1, Si-0, Wt-5, Ob-1, Jm-2, and 46 other accessions shared a haplotype associated with significantly lower exon 3-skip *P5CS1* percentage. These accessions had relatively short or no insertions in intron 2, and all had identical sequences in intron 3 except for Tsu-1 (Fig. 3*C*).

The 5-SNP interval where L*er* vs. Sha haplotypes were most divergent in proline abundance was the same as the interval most strongly associated with exon 3-skip *P5CS1* percentage. Haplotype variation at 6 of 24 *P5CS1* 5-SNP intervals where L*er* and Sha were polymorphic was significantly associated with proline abundance, (linear mixed model on log-proline abundance, $\alpha = 0.05$). In addition to comparing L*er* vs. Sha haplotypes, we compared phenotypes of all haplotypes represented by more than one accession in each 5-SNP interval, while accounting for kinship. Haplotypes had significantly different exon 3-skip *P5CS1* percentage at 9 of 29 total 5-SNP intervals, and had significantly different proline abundance at 10 of the 29 intervals (Wald test of fixed effects, linear mixed model, $\alpha = 0.05$).

Correlation of Exon 3-Skip *P5CS1* or Proline with Climate Factors. We compiled 101 climate variables on precipitation, temperature, vapor pressure deficit, interannual precipitation variation, growing season conditions, and climate extremes (11–13). We joined climate data to accessions based on published collection locations (14). We focused on accessions in the native Eurasian range, discarding those from North America that are more recently introduced (14). Accessions with unreliable collection location were also discarded (15), as were accessions for which 250 K SNP data were missing (14). We calculated principal components axes of climatic variation among accessions, and we kept the first 10 axes that cumulatively explained 97% of climatic variation.

We then studied the association between proline abundance and exon 3-skip *P5CS1* percentage with climate variables. Phenotypic variation can be caused by population structure, and climate-phenotype associations can be observed if population structure is correlated to climate. Thus, to discern climate-phenotype relationships that are putatively adaptive we attempted to correct for population structure (16). A kinship matrix among accessions was calculated using SNP identity-in-state (17). The kinship matrix was then included as a random effect in a linear model (EMMA) where a climate variable was the predictor and a molecular phenotype the dependent variable (16). We tested all climate variables for associations with proline abundance (log) and exon 3-skip *P5CS1* percentage (log + 1). We also tested for associations using less conservative nonparametric Spearman's rank correlations between climate and molecular phenotypes.

Population Genetics of *P5CS1* and Evidence for Selection. Population genetic analyses were completed on the aligned coding sequences of *P5CS1* from 80 accessions of *Arabidopsis* published as part of the 1001 Genomes Project (18) (www.1001genomes.org) using

the *Arabidopsis lyrata P5CS1* coding sequence as an outgroup. We performed a McDonald–Kreitman test (19) using DNASP (20). In addition, genome selection scans of the pair-wise haplotype sharing (PHS) and F_{st} statistics were used to ask whether *P5CS1* variation was extreme in its values relative to the genome-wide distribution of values using the published scans from Hancock et al. (14) (http://bergelson.uchicago.edu/regmap-data/climate-genome-scan). The PHS test asks whether the haplotype associated with a given SNP extends further than is expected given a genome-wide distribution of haplotype lengths, and F_{ST} asks whether there is spatially structured variation at a locus. In both cases, extreme test statistics at a candidate gene relative to the genome-wide pattern give support for rejecting the null hypothesis of neutral evolution at the target locus. Finally, using analyses presented by Hancock et al. (14), we asked whether SNPs in *P5CS1* are associated with particular climate variables.

1. 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.
2. Zeng Z-B (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457–1468.
3. Broman KW, Sen S (2009) A guide to QTL mapping with R/qtl. *Statistics for Biology and Health* (Springer Science+Business Media, New York).
4. Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971.
5. Doerge RW, Churchill GA (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142:285–294.
6. Sen S, Churchill GA (2001) A statistical framework for quantitative trait mapping. *Genetics* 159:371–387.
7. Littel RC, Milliken GA, Stroup WW, Wolfinger RD (1996) *SAS System for Mixed Models* (The SAS Institute, Cary, NC).
8. Manichaikul A, Moon JY, Sen S, Yandell BS, Broman KW (2009) A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. *Genetics* 181:1077–1086.
9. Nakagawa T, et al. (2007) Improved Gateway binary vectors: High-performance vectors for creation of fusion constructs in transgenic analysis of plants. *Biosci Biotechnol Biochem* 71:2095–2100.
10. Martínez-García JF, Monte E, Quail PH (1999) A simple, rapid and quantitative method for preparing *Arabidopsis* protein extracts for immunoblot analysis. *Plant J* 20:251–257.
11. Kalnay, et al. (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471.
12. New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21:1–25.
13. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978.
14. Hancock AM, et al. (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334:83–86.
15. Anastasio AE, et al. (2011) Source verification of mis-identified *Arabidopsis thaliana* accessions. *Plant J* 67:554–566.
16. Kang HM, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723.
17. Atwell S, et al. (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631.
18. Cao J, et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43:956–963.
19. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
20. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

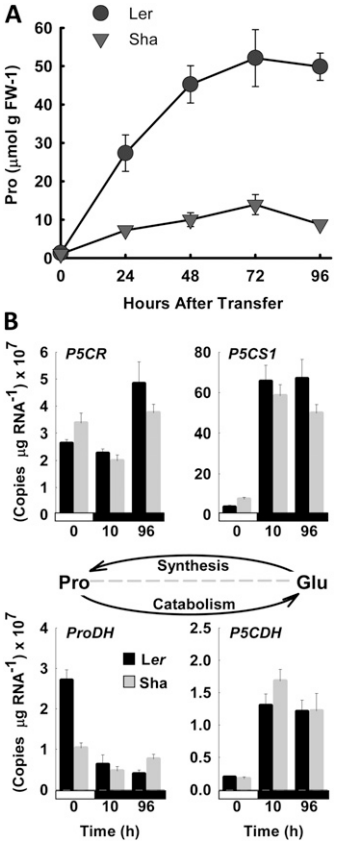

Fig. S1. Time course of proline content and proline metabolism gene expression in L*er* and Sha. (*A*) Time course of proline content after transfer of 7-d-old L*er* and Sha seedlings from control (−0.25 MPa) to low water potential (−0.7 MPa) media. Data are means ± SE (n = 4). (*B*) Proline metabolism gene expression measured before transfer (Time 0) or at 10 and 96 h after transfer of 7-d-old seedlings to low ψ_w (−1.2 MPa). Data are means ± SE (n = 3). Quantitative PCR was conducted using Taqman primer/probe sets as described in Sharma and Verslues (1). The Taqman primer/probe set recognizes the 3′ end of *P5CS1*, downstream of the alternative splicing event described in the main text. *P5CS1*: *Δ^1-pyrroline-5-carboxylate synthetase1*; *P5CR*: *Δ^1-pyrroline-5-carboxylate reductase*; *ProDH*: *proline dehydrogenase1*; *P5CDH*: *Δ^1-pyrroline-5-carboxylate dehydrogenase*.

1. 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.

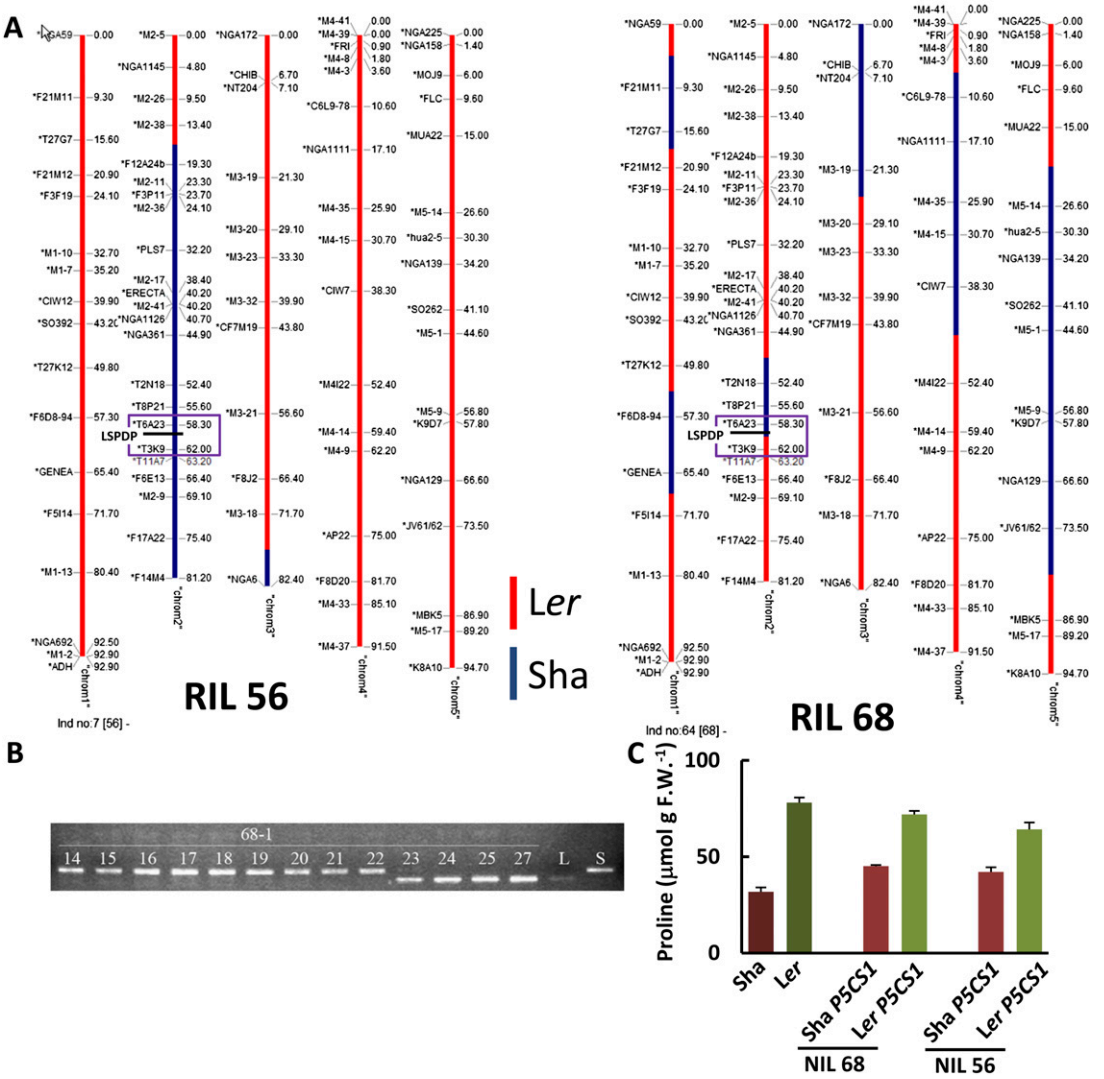

Fig. S2. Near Isogenic Lines (NILs) confirm the *Pro-W2* QTL effect. (*A*) Genome composition of the two RIL lines used to generate NILs. Purple boxes indicate the region containing the *PRO-W2* QTL. (*B*) Example of genotyping individual NIL lines with the LSPDP marker after the second backcross to L*er*. RILs 68 and 56 were twice backcrossed to L*er* with the T6A23, T3K9 markers used for genotyping after the first backcross and LSPDP (32-bp insertion in *P5CS1* intron 3 of Sha) for genotyping after second backcross. (*C*) Proline content of NILs after −1.2 MPa stress treatment. "L*er P5CS1*" refers to NILs that had the Sha genotype in the QTL interval after the first backcross but recombined to L*er* genotype at *P5CS1* after the second backcross. NILs having the Sha genotype at the *P5CS1* region consistently had low proline contents similar to Sha and those with L*er* genotype had high proline contents similar to L*er*. Forty-eight total NIL plants were genotyped and proline phenotyped after second backcross for NIL 68 and for NIL 56 six plants was scored.

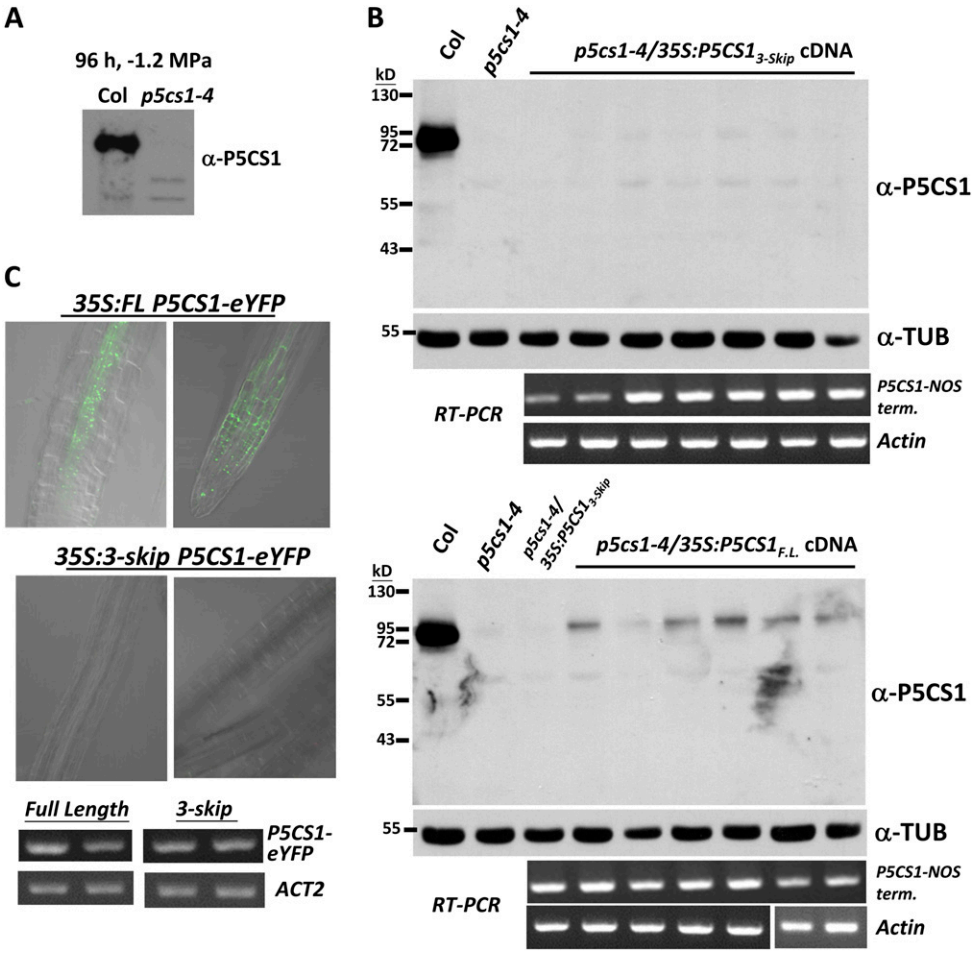

Fig. S3. Exon 3-skip *P5CS1* transcript is not translated to P5CS1 protein. (*A*) Characterization of α-P5CS1 polyclonal antisera. α-P5CS1 recognizes a single major band in wild-type and this band is absent in *p5cs1-4* (Salk_063517). Occasionally, small amounts of truncated P5CS1 protein could be detected in *p5cs1-4*; consistent with the position of the T-DNA insertion near the 3′ end (exon 14) of *P5CS1*. (*B*) RT-PCR and Western blotting of several transgenic lines expressing either the exon 3-skip *P5CS1* cDNA (*Upper*) or full-length (F.L.) *P5CS1* cDNA (*Lower*). Western blotting consistently detected P5CS1 protein in transgenic lines expressing the full-length *P5CS1* cDNA. *P5CS1* protein was not detected in any transgenic lines expressing the exon 3-skip *P5CS1* cDNA. Western blots were reprobed with α-tubulin (TUB) as a loading control. RT-PCR was conducted with primers recognizing *P5CS1* and the NOS terminator to specifically detect transgene-encoded *P5CS1* RNA. (*C*) Example confocal microscope images of root tissue from plants expressing a C-terminal fusion of eYFP to either the full-length *P5CS1* cDNA or the exon 3-skip *P5CS1* cDNA. YFP flourescence could be detected in transgenic lines expressing the full-length *P5CS1* cDNA-eYFP fusion but was not detected in transgenic lines containing the exon 3-skip *P5CS1-eYFP* (more than 10 independent transgenic lines were analyzed). Transcript of the *P5CS1-eYFP* transgene was detected in all lines using *P5CS1-eYFP*–specific primers.

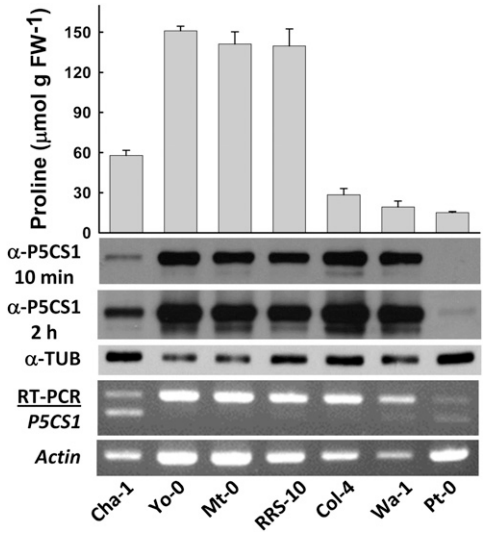

Fig. S4. Additional variation in proline content not associated with P5CS1 protein level and identification of Pt-0 as a naturally occurring P5CS1-deficient accession. Proline contents are from Dataset S1. For Western blotting, 50 μg of total protein from stressed seedlings (−1.2 MPa) was loaded and a long exposure of the P5CS1 blot is shown to demonstrate the extreme low level of P5CS1 in Pt-0. RT-PCR was performed for 27 cycles using primers A and C shown in Fig. 2*A*.

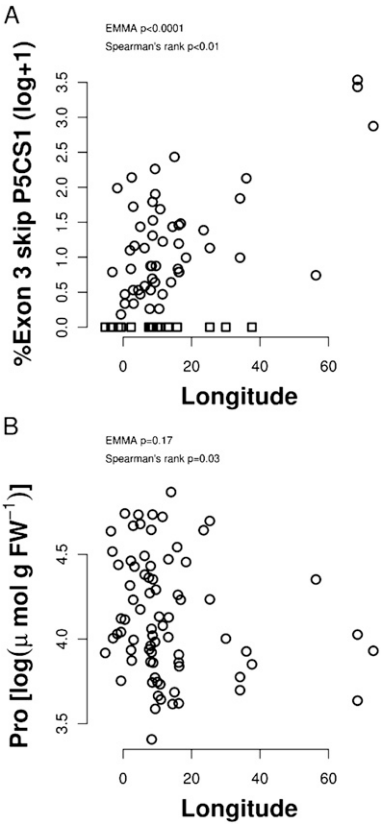

Fig. S5. (*A* and *B*) Change in exon 3-skip *P5CS1* percentage and proline abundance across Eurasia (measured with longitude). Accessions lacking exon 3-skip *P5CS1* are shown with squares. Associations were tested while accounting for population structure (EMMA) and nonparametrically (Spearman's rank correlation).

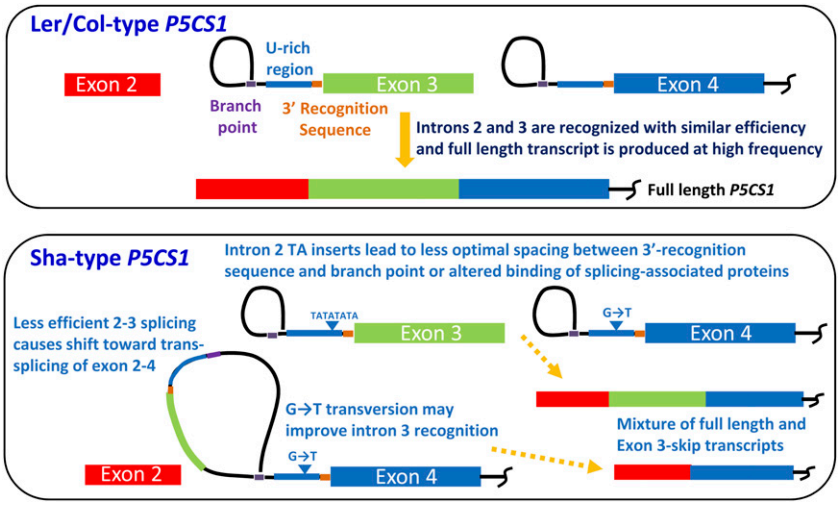

Fig. S6. Model of splicing differences between Ler/Col-type *P5CS1* vs. Sha type-*P5CS1* containing intron 2 TA insertions and intron 3 G-to-T transversion. Optimal spacing between the 3′ intron recognition sequence and the branch point is needed for efficient splicing. The TA insertions in Sha and other high exon 3-skip *P5CS1* accessions are likely to disrupt this spacing or alter the binding of splicing-associated proteins. This decreases the efficiency of intron 2 recognition and exon 2–3 splicing. Because the presence of TA-repeat sequence in intron 2 rather than a poly-T tract already suggest that this may be a relatively weakly recognized intron, the altered spacing or binding of splicing proteins is enough to cause substantial exon 2–4 splicing. The G-to-T tranversion in intron 3 may also contribute by making intron 3 more efficiently recognized and further promoting exon 2–4 splicing.

Table S1. QTL mapping results

Treatment	Marker	Chromosome	cM	Confidence interval	LOD score	*P* value	Additive effect ($2a \pm 1SE$)
PEG (−0.7)							
PRO-W2	T3K9	2	61	(54–65)	7.16	0.000	−8.89 ± 1.49
PEG (−1.2)							
PRO-W2	T3K9	2	62	(58–64)	7.02	0.001	−12.35 ± 2.12
PRO-W5	MUA22	5	13	(0–14)	4.75	0.005	9.08 ± 2.12

QTL mapping of proline contents at either −0.7 or −1.2 MPa detected 2 QTL (*Pro-W2* and *Pro-W5*). (See *SI Materials and Methods* for details of QTL mapping procedure.)

Table S2. QTL × PEG-treatment interaction analysis

Fixed effect	*df*	*F*-value	*P* value
PRO-W2	1,111	41.3	<0.0001
PRO-W5	1,111	17.26	<0.0001
PEG-Treatement	1,111	592.97	<0.0001
*PRO-W2**PEG-Treatment	1,111	5.76	0.0181
*PRO-W5**PEG-Treatment	1,111	8.18	0.0051
*PRO-W2***PRO-W5*	1,111	2.61	0.1091
*PRO-W2***PRO-W5**PEG-Treatment	1,111	0.24	0.6283

No evidence of interaction of the two QTL (*Pro-W2***Pro-W5*) was found. (See *SI Materials and Methods* for details of QTL mapping procedure.)

Table S3. Primers used for cloning, 5′ RACE, RT-PCR, and genotyping

Primer name and purpose	Sequence
5′-RACE, cloning and RT-PCR of *P5CS1*	
5′-RACE CDS Primer A	(T)25V *N*-3, where *n* = A, C, G, or T and V = A, G, or C
SMARTer IIA oligo	AAGCAGTGGTAACAACGCAGAGTACXXXXX, where X = undisclosed base in the proprietary SMARTer oligo sequence
Universal primer A mix long	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
Universal primer A mix short	CTAATACGACTCACTATAGGGC
P5CS1 Seq1 F	AAGTCAAAGCTGCAGTCAATG
P5CS1 5′-RACE GSP1 (3′ UTR primer, Fig. 2)	TGGTGCAATCCTAACCCTAAATCAAAACC
P5CS1 Exon2 F (primer a, Fig. 2)	TGGCTCTTGGTCGTTTAGGAG
P5CS1 Exon3 F (primer b, Fig. 2)	ACTCGGATGGATTTGAGGTG
P5CS1 Exon 6 (primer c, Fig. 2)	TTGCAGTCATACCCCCTCTC
P5CS1 Exon 18 F (Fig. 2)	GCACAAGATTCTCAGATGGTTTC
ACT2 RT F	GCCATCCAAGCTGTTCTCTC
ACT2 RT R	GAACCACCGATCCAGACACT
p5cs1 GW F no stop (cDNA Gateway cloning)	GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGAGGAGCTAGATCGT
p5cs1 GW R no stop (cDNA Gateway cloning)	GGGGACCACTTTGTACAAGAAAGCTGGGTCAGCTTGGATGGGAATGTC
P5CS1 F (RT-PCR of Nos-T or eYFP fusion trans.)	AAGTTGCAGAGCTATTCCTTC
eYFP R	GATGAACTTCAGGGTCAGCTT
NosT R	AGACCGGCAACAGGATTCAATC
AtP5CS1 genomic D-TOPO 1 F	CACCAGAAATTGGATTGGCTAATCACAC
AtP5CS1 genomic D-TOPO 1 R	GTGTAGGTAGCTTACAATGACAA
Markers used for screening of NILs	
T6A23 F	GGTATTGATATCCAGTCACA
T6A23 R	TCTAATGACTGAAGCAAAGC
T11A7 F	GGAGATCAAAATAAGTCTGA
T11A7 R	ACGTATTACGTAGCATAGATC
P5CS1 markers for second backcross NIL selection and intron 2 insertion scoring	
LSPDP F (NIL genotyping)	CAGTGACATGTGTTTATGTGTG
LSPDP R (NIL genotyping)	CCTTCCCATCAAGTTCAGT
LSPDP2 F (scoring of intron 2 insertion size)	GACATGTGAAATTATTCATGCGTAG
LSPDP2 R (scoring of intron 2 insertion size)	CAAGACCAACCGCACCAGAT

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)
Dataset S3 (XLS)
Dataset S4 (XLS)
Dataset S5 (XLS)