Nucleic Acids Research, 2026, **54**, gkag400
https://doi.org/10.1093/nar/gkag400
Genomics

Impact of alternative splicing on *Arabidopsis* proteome

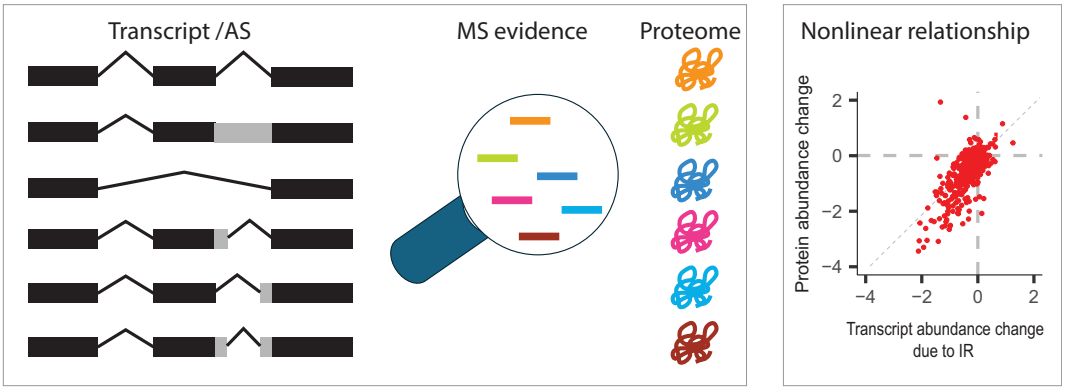
Andres V. Reyes[1,2], Christopher Zhang[1], Sumudu S. Karunadasa[1], Ruben Shrestha[1], TaraBryn S. Grismer[1,2], Danbi Byun[1], Shou-Ling Xu[1,2,*]

[1]Department of Plant Biology, Division of Biosphere Science and Engineering, Carnegie Institution for Science, Stanford, CA 94305, United States
[2]Carnegie Mass Spectrometry Facility, Carnegie Institution for Science, Stanford, CA 94305, United States
*To whom correspondence should be addressed. Email: slxu@stanford.edu

Abstract

Limited proteomic evidence makes it unclear to what extent alternative splicing (AS) isoforms are translated and functionally relevant in eukaryotes. Here, we present a comprehensive proteomic analysis in plants using large-scale data mining, extensive fractionation of AspN- and trypsin-digested proteomes, and both label-free and TMT labeling. In total, we identified 471 196 peptides from 22 479 proteins by searching against Araport11, revealing 32 110 isoform-specific peptides. Using an integrated proteogenomic workflow coupled with SUPPA, we classified these peptides into 2442 AS events, 879 of which involved intron retention (IR). Further analysis of unannotated events revealed 91 additional IRs that are translated, supporting that retained introns can give rise to peptides. AlphaFold modeling predicted the structural and functional impacts of these isoforms. Our dataset improved existing gene model annotations. By comparing wild-type plants with the AS mutant *acinus pinin*, we found that IR regulates transcript and protein abundance nonlinearly. Phenotypic assays revealed the functional consequences, including reduced chlorophyll, impaired growth, and increased anthocyanin. Overall, our results support widespread translation of AS isoforms in plants and suggest that AS contributes to proteome diversification, protein abundance regulation, and growth and developmental outcomes.

Graphical abstract



Introduction

Alternative splicing (AS) is a widespread and conserved regulatory mechanism, affecting over 95% of multi-exon genes in animals and 70% in plants [1–3]. AS is essential for development, tissue identity, immunity, and stress responses across species, and its dysregulation is linked to diseases such as cancer [4–10]. Functional studies have highlighted AS's impact on growth and development across species [11–14]. For example, thousands of Dscam1 isoforms are essential for normal neural circuit development in *Drosophila* [11, 12]. Other examples include plant Rubisco activase and *HYPERSENSITIVE TO ABA1* (*HAB1*), which generate isoforms with distinct regulatory roles [15–17], while retained introns in animals can confer growth advantages to cancer cells [18]. These examples underscore AS as a key mechanism shaping gene regulation and adaptation across evolution.

Although plants and animals share the same major types of AS events, the frequency and predominant forms differ significantly between the two kingdoms [19, 20]. Exon skipping is the most common type in metazoans, whereas intron retention (IR) is the dominant form in plants, accounting for over 60% of events in plants [3, 21, 22]. The exact reasons for this divergence remain unclear but likely involve a combination of *cis*-regulatory elements, such as intron length, and differences in splicing factors [20, 23].

These differences may have various consequences for the proteome [7, 19, 22, 24, 25]. Exon skipping and the use of alternative 5′ or 3′ splice sites are thought to increase

Received: October 22, 2025. **Revised:** April 2, 2026. **Accepted:** April 9, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

proteome diversity by creating different protein isoforms or by modulating gene expression through mechanisms such as nonsense-mediated decay (NMD) and uORFs [19, 26]. IR is also hypothesized to increase diversity by introducing novel sequences or truncated isoforms. Emerging studies also propose that retained introns may play additional regulatory roles. For example, they may be exported to the cytoplasm for degradation by NMD, sequestered in the nucleus for delayed processing into fully spliced messenger RNAs (mRNAs) [13, 27], or degraded by the nuclear exosome. Most of these outcomes, particularly in plants, remain speculative and require experimental validation to determine their actual impact on proteome regulation.

A key unresolved question is the extent to which splice variants are translated and functionally relevant [22, 28, 29]. Evidence from animal studies suggests that AS events often preserve frame and exhibit evolutionary conservation [30, 31]. Ribosome profiling studies further indicate that many transcript variants associate with polysomes, suggesting that they are translated [32–36]. In addition, many splicing events and their regulatory elements are conserved across species; for example, splicing signals in retained introns are found to be more conserved than those in constitutive introns [37, 38], implying selective pressure to preserve their functions. However, many of these AS variants may be expressed at low levels, leading to debate about their overall functional significance within the proteome.

Direct evidence of splice variant translation at the proteome-wide level has come from mass spectrometry-based analyses. However, the limited number of available studies has yielded conflicting conclusions [39–41], highlighting persistent technical challenges associated with proteomics approaches. These challenges include limited proteomic coverage and sensitivity, the tissue- or condition-specific nature of many AS events, and the difficulty of detecting splice junctions using standard tryptic digestion. This difficulty arises because evolutionarily conserved nucleotide usage at exon boundaries increases the frequency of lysine- and arginine-coding triplets at exon ends, making detection of exon–exon and exon–intron junctions challenging [42]. Several efforts have begun to address these barriers in human systems. For example, some studies have leveraged splicing mutants [43], while others have applied deep proteome sequencing strategies [44]. However, comparable large-scale efforts remain lacking in plants.

To address this knowledge gap, we conducted a thorough investigation into how AS influences the *Arabidopsis thaliana* proteome. By analyzing over 900 liquid chromatography–tandem mass spectrometry (LC-MS/MS) datasets—720 from public repositories and 180 generated in-house—we achieved the first large-scale identification of isoform-specific peptides in plants. Using our in-house proteogenomic approach, coupled with SUPPA [45] for event classification and additional library annotations, we detected a total of 2533 splice events. Our results support that diverse AS events are translated and detectable at the proteome level, substantially expanding the diversity of the *Arabidopsis* proteome. Notably, not only are canonical events such as exon skipping and 5′ or 3′ AS translated, but many IR events are also translated. This proteomic evidence further provides a better basis for gene model annotation.

We further incorporated the *acinus pinin* double mutant, which has previously been shown to exhibit 2446 altered AS events at the transcript level, including 1106 IR, 206 exon skipping, 371 alternative donor, 498 alternative acceptor, and 265 combined alternative donor-acceptor events [23, 46]. Many of these regulated IR events can be recapitulated in wild-type (WT) plants subjected to biotic or abiotic stress. Our approach allowed us to quantitatively analyze the impact of retained introns on transcripts and proteins, revealing a nonlinear relationship between IR and transcript/protein abundance. Our data further show that these impacts, driven by both AS and transcriptional regulation, are functionally significant, correlating with the phenotypic defects observed in the *acinus pinin* mutant.

Taken together, our findings provide the first large-scale demonstration directly linking AS variation to proteomic outcomes in plants, offering new insights into the regulatory impact of AS on plant development and physiology.

Materials and methods

Plant materials and growth conditions

Arabidopsis thaliana Col-0 (WT), *acinus-2 pinin-1* (*ap*) mutants, and complementation line (*ap* + ACINUS) [23, 46] were used. Seeds were surface-sterilized, stratified for 3 days, and germinated on ½ MS agar plates at 22°C under constant light. Seedlings were harvested at specific stages. All plates were grown vertically except for Fig. 6C.

Chlorophyll measurement and root length assay

Chlorophyll was extracted from 30 mg of 7-day-old seedling tissue, following protocol described in [47]. Samples were flash-frozen, ground, and extracted with 80% acetone overnight at 4°C. After centrifugation (5544 × *g*, 15 min, RT), absorbance was measured at 663 nm and 645 nm (Tecan plate reader). Chlorophyll content was calculated as follows: Chlorophyll = 8.02 × (chlorophyll a absorbance A663) + 20.20 × (chlorophyll b absorbance A645). For root assays, seedlings grown vertically for 7 days on ½ MS were photographed. Images were analyzed in ImageJ, and root lengths were measured from the hypocotyl-root junction to the root tip (WT: $n = 60$; *ap*: $n = 82$; *ap* + ACINUS: $n = 60$). Statistical significance was assessed using Student's t-test.

Anthocyanin analysis

Anthocyanin was extracted from 30 mg of 11-day-old vertical plates, frozen, and ground tissue in 1 ml of 1% HCl-methanol. Samples were vortexed and incubated overnight at 4°C in the dark, then centrifuged (16 000 × *g*, 5 min). Absorbance of the supernatant was measured at 530 nm and 657 nm (Tecan plate reader), and anthocyanin content was calculated as [absorbance at 530 (A530) − 0.25 × absorbance at 657 (A657)].

Sample preparation for AspN and TMT11 experiments

Tissue (100 mg) from 9-day-old WT and *acinus pinin* seedlings grown on vertical plates was harvested, flash-frozen, and ground. Proteins were extracted as described in [48], with a slight modification to extraction buffer Y (100 mM Tris–HCl, pH 8.0; 2% SDS w/v; 5 mM EGTA; 10 mM ethylenediaminetetraacetic acid; 1 mM PMSF; 2 × protease inhibitor, Roche). Extracts were reduced with TCEP, alkylated, and digested with trypsin or AspN, followed by desalt-

ing using Sep-Pak columns (Waters). For TMT experiments, five biological replicates per group were prepared.

TMT 11-plex labeling

TMT 11-plex reagents (Thermo Fisher, Cat# A34808) were reconstituted in anhydrous acetonitrile (10 μg/μl). Approximately 200 μg of peptides from each sample were dissolved in 66 μl of 50 mM HEPES buffer. Of this, 60 μl (182 μg) was labeled with 40 μl of TMT reagents (TMT126-130N), while the remaining 6 μl (18.2 μg) was pooled to 60 μl and labeled with 40 μl TMT131C. Reactions proceeded for 1 h at 25°C with shaking at 700 rpm and were quenched with 5 μl of 1 M Tris (pH 8). Peptides were then acidified with 45 μl of 10% FA in 10% acetonitrile. The final reaction contained TMT at 4 μg/μl, peptides at 2 μg/μl, and 40% acetonitrile (v/v), similar to formular described in [49]. The labeling scheme is shown in Supplementary Fig. S3. One-eighth of the 11 labeled samples was combined for high-pH reversed-phase HPLC fractionation [48].

Fractionation of peptide samples for AspN and TMT samples

High pH reverse-phase chromatography was performed on the Vanquish Flex system (Thermo Fisher) equipped with a 4.6 × 150-mm Gemini 5 μm C18 column for TMT-labeled peptides and AspN digested samples. Peptides were loaded onto the column in 25 μl of buffer A (20 mM ammonium formate, pH 10). Buffer B consisted of buffer A with 90% (vol/vol) acetonitrile. Peptides were separated at 0.5 ml/min using high-pH reversed-phase gradients. Common steps included column equilibration at 1% B and sequential increases to 50%–70% B, 70%–100% B, and 100% B hold. Fractions were collected every 1–2 min. TMT batch 1: 5 min equilibration; 1%–5% B in 2 min, 5%–50% B in 58 min; fractions every 1 min, with 39 fractions selected for subsequent analysis. TMT batch 2: 10 min equilibration; 1%–10% B in 10 min, 10%–50% B in 40 min; fractions every 2 min, with 10 fractions selected for subsequent analysis. TMT Batch 3: 10 min equilibration; 1%–10% B in 20 min, 10%–50% B in 80 min; fractions every 1 min, with 50 fractions selected for subsequent analysis. AspN sample: 5 min equilibration; 1%–5% B in 2 min, 5%–50% B in 58 min; fractions every 1 min, with 45 fractions selected for further analysis. All collected fractions were vacuum-dried prior to LC-MS/MS analysis.

LC/MS/MS analysis of AspN and TMT samples

Peptides were analyzed on a Q-Exactive HF or Orbitrap Eclipse hybrid quadrupole–Orbitrap mass spectrometer (Thermo Fisher) coupled to an Easy LC 1200 UPLC system. Peptides were trapped on an Acclaim PepMap 100 C18 column (75 μm × 2 cm) and separated on a 25 cm × 75 μm, 1.7 μm C18 Aurora column (IonOpticks) at 300 nl/min using a 120 min gradient: 3%–28% solvent B (80% ACN, 0.1% FA) over 106 min, 28%–44% B over 15 min, followed by a 9 min wash at 90% B. Batch 1 TMT samples were analyzed on the Q-Exactive HF with MS1 scans over m/z 375–1600 and were acquired at m/z 375–1600, resolution 120 000, AGC 3E6, and max injection 100 ms. MS2 scans select the top 20 multiply charged precursors for HCD fragmentation (NCE 30, isolation 1.0 m/z) with a scan range of 200–2000 m/z, resolution 30 000, AGC 5E4, max injection 60 ms, and dynamic exclusion 24 s. Batch 2 and 3 TMT and AspN samples were analyzed on the Eclipse; AspN MS1 scans were acquired at m/z 375–1600, resolution 120 000, AGC 2E5, max injection 50 ms, and MS2 targeted top multiply charged precursors (charge 2–8) with HCD 27%, isolation 1.4 m/z, resolution 15 000, AGC 5E4, max injection 22 ms, cycle time 3 s, dynamic exclusion 30 s. TMT samples were measured using real-time library search [50], with MS1 m/z 400–1600, resolution 120 000, intensity threshold 5E3, charge 2–6, and dynamic exclusion 45 s; MS2 in the ion trap (isolation 0.7 m/z, CID 35%, max injection 35 ms); MS3 using synchronous precursor selection of 10 precursors (isolation 0.7 m/z, HCD 65%, Orbitrap resolution 50 000, scan range 100–500 m/z, max injection 200 ms, cycle time 2.5 s). Real-time data searches [51] were performed against the TAIR10 database with carbamidomethyl and TMT11plex (static) and oxidation (variable) modifications.

Overlap of intron retention analysis by PlanIntrontDB

Introns exhibiting increased retention in *acinus pinin* mutants (single-end RNA-seq) [46] were retrieved from the Plant Intron Splicing Efficiency Database (PlantIntronDB; plantintron.com) [3], prioritizing datasets generated under comparable conditions by independent groups, and used for statistical analysis in Supplementary Fig. S1.

Data analysis of overlapping differentially expressed genes and altered retained intron transcripts

Pair-end RNA-seq data for WT and *acinus pinin* mutants were processed as described [23]. Additional RNA-seq datasets for selected biotic and abiotic stresses—flg22 (PRJNA1124505) [52], cold (12 h, PRJNA1062790) [53], and heat (PRJNA1262915)—were downloaded using SRA-Tools (v2.10.0). Reads were aligned with STAR (v2.7.10b) using quantMode = GeneCounts, outFilterMultimapNmax = 20, and outSAMattributes = NH HI AS nM MD. Transcriptomic and AS analyses followed [23, 46]. Dataset overlaps were evaluated with a hypergeometric test. Pairwise overlaps were classified as concordant (shared upregulated or downregulated events) or discordant (oppositely regulated), and statistical significance was assessed via Fisher's exact test on a collapsed 2 × 2 contingency table. Intron numbers from RackJ analysis (Supplementary Table S1) are assigned relative to the plus strand, from 5′ to 3′; for genes on the negative strand, introns are numbered in descending order. In figures, intron numbers have been adjusted to ascending order for clarity.

Data search and detection of isoform-specific peptides

Isoform-specific peptides were identified by searching all 909 runs (Fig. 2) using Protein Prospector and MSFragger with similar parameters. Previous studies showed that these two search engines yield complementary peptide identification results [54]. For all HCD data, the precursor mass tolerance was set to 10 ppm and MS/MS tolerance to 20 ppm. For CID TMT data, low resolution, precursor mass tolerance was set to 10 ppm and MS/MS tolerance to 0.6 Da. Carbamidomethylation of cysteine was set as a fixed modification, while variable modifications included protein N-terminal acetylation, peptide N-terminal Gln conversion to pyroglutamate, and methionine oxidation. Additionally, for TMT data, a mass of the TMT11

tag (229.16293) was set as a variable modification on lysine or on the N-terminal of a peptide. False discovery rates were set at 5% for proteins and 1% for peptides. Cleavage specificity was defined according to the protease used (trypsin or AspN), allowing one missed cleavage and up to two modifications per peptide. Raw data were searched against the Araport11 protein database, which contains 48 354 entries.

Peptide-level processing and isoform assignment

Peptide-level identification results were imported into R and processed as follows. Decoy matches were removed, and sequences with N-terminal methionine loss were adjusted accordingly. Peptides from both Protein Prospector and MSFragger searches were combined, retaining only one peptide per distinct sequence. Peptides mapping to single-isoform genes or to multiple genes were excluded. Isoform-specific peptides were defined as those mapping to isoforms disjoint from at least one set of isoforms represented by other peptides from the same gene locus (see algorithm breakdown, Supplementary Fig. S7).

Proteome-genomic mapping and classification of splicing events

Peptides identified from the database were filtered to retain only those corresponding to genes with multiple distinct protein isoforms. For each peptide, an in-house R script (available on GitHub and Zenodo) was used to retrieve its position on the protein and the transcript to which it maps. Genomic coordinates of each transcript exon were then obtained. Coding sequence (CDS) exon coordinates were enumerated, and the base pairs corresponding to each amino acid of the peptide were calculated (first amino acid: position $\times$ 3 $-$ 2; last amino acid: position $\times$ 3). Peptides spanning exon junctions were identified by comparing these coordinates with exon boundaries.

In parallel, Araport11 annotations were provided to SUPPA (v2.3) [45] to generate a comprehensive set of annotated AS events with genomic coordinates. SUPPA was run under default settings, with the variable option enabled for IR events. Peptide-genome coordinates were integrated with SUPPA-derived AS coordinates to identify splicing events. For retained introns, alternative 5′ and 3′ splice sites, and alternative first and last exons, events were called based on coordinate overlap. For exon skipping and mutually exclusive exon events, junction-spanning peptides were used to support splicing patterns. All analyses, except for those using SUPPA, were performed with custom R scripts, available in the project repository.

Intron depth ratio and creation of custom databases and data search

Intron depth ratios were calculated according to the protocol described by [23]. Briefly, STAR alignment files were processed by RACKJ to retrieve both read counts and read depth for each annotated intron and exon. To analyze IR, the intron read depth ratio (IDratio) was determined by dividing the read depth of a target intron by the average read depth of its two flanking exons. For the first and last introns, the IDratio was calculated using only the second (3′) or the first (5′) flanking exon, respectively. This adjustment was made because read depths for terminal exons can be inaccurate due to inherent sequencing biases.

A custom database of unannotated IR transcript events was generated by first filtering introns with high IR depth ratios ($\geq$0.15) in either WT or mutant samples from our high-coverage RNA-seq dataset. Genomic coordinates of the selected introns were retrieved, and introns already annotated in Araport11 (as determined by SUPPA) were excluded to avoid redundancy. Each intron of interest was mapped onto the genomic coordinates of all transcripts for its corresponding gene. Introns fitting without gaps—either immediately upstream of the first exon, downstream of the last exon, or between two exons—were used to assemble the genomic sequences, sorted by strand, using the IRanges R package. The underlying genomic sequences were translated into protein sequences using the BSgenome R package. Newly constructed proteins were appended to the Araport11 protein database for downstream peptide searches. Subsequent searches were performed as described earlier, retaining only peptides that mapped to the custom proteins.

Quantitative TMT11-plex data analysis

For Eclipse MS data, raw files were searched using MaxQuant v2.4.2.0. In the global parameters, the type was set to MS3 ion reporter mode using the correction factors provided by the TMT reagents. No normalization setting was set. The first peptide mass tolerance was set to 20 ppm and the main search peptide mass tolerance was set to 4.6 ppm. The MS2 match tolerance was set to 20 ppm. Reporter mass tolerance was set to 0.003, and the isobaric weight exponent was set to 0.75. The reference protein database used was a modified version of the *Arabidopsis thaliana* TAIR 10 database, downloaded from TAIR: "TAIR10_pep_20101214.fa". All other parameters remained at their default. *For protein level analysis*: MaxQuant outputs proteinGroups.txt and evidence.txt were used with R package MStatsTMT [55]. The parameters for MSstatsTMT's proteinSummarization function: method was set to msstats, global normalization was set to true, reference normalization was set to true, and remove normalization channel was set to true. Protein level quantifications for each sample were analyzed to impute protein quantifications that were only completely missing in one condition while being completely quantified in all 5 replicates in the other condition. The imputation was done from a normal distribution with a width of 0.3 and downshifted 1.8 from the values of the total matrix. Following imputation, a moderated statistical test was done using MSstatsTMT's groupComparisonTMT function. Proteins with missing *P*-values were filtered out. If an ambiguous protein group existed with different gene loci, they were discarded. If multiple protein groups existed for a single gene locus, the protein group with the most peptides quantified was kept (also see Supplementary Fig. 16). Protein groups with a *P*-value <.05 and a fold change greater than 1 or less than -1 were considered as differentially abundant proteins. Peptide level quantifications shown in Fig. 5G were generated from Maxquants's peptides.txt output. To determine each peptide's position relative to intronic regions, we developed a custom R script that maps intron coordinates onto specific isoforms to retrieve the amino acid position of the intron insertion.

GO analysis

Gene Ontology (GO) analysis was performed using the topGO package in R. For both transcriptomic and proteomic datasets, the ontology was restricted to biological pro-

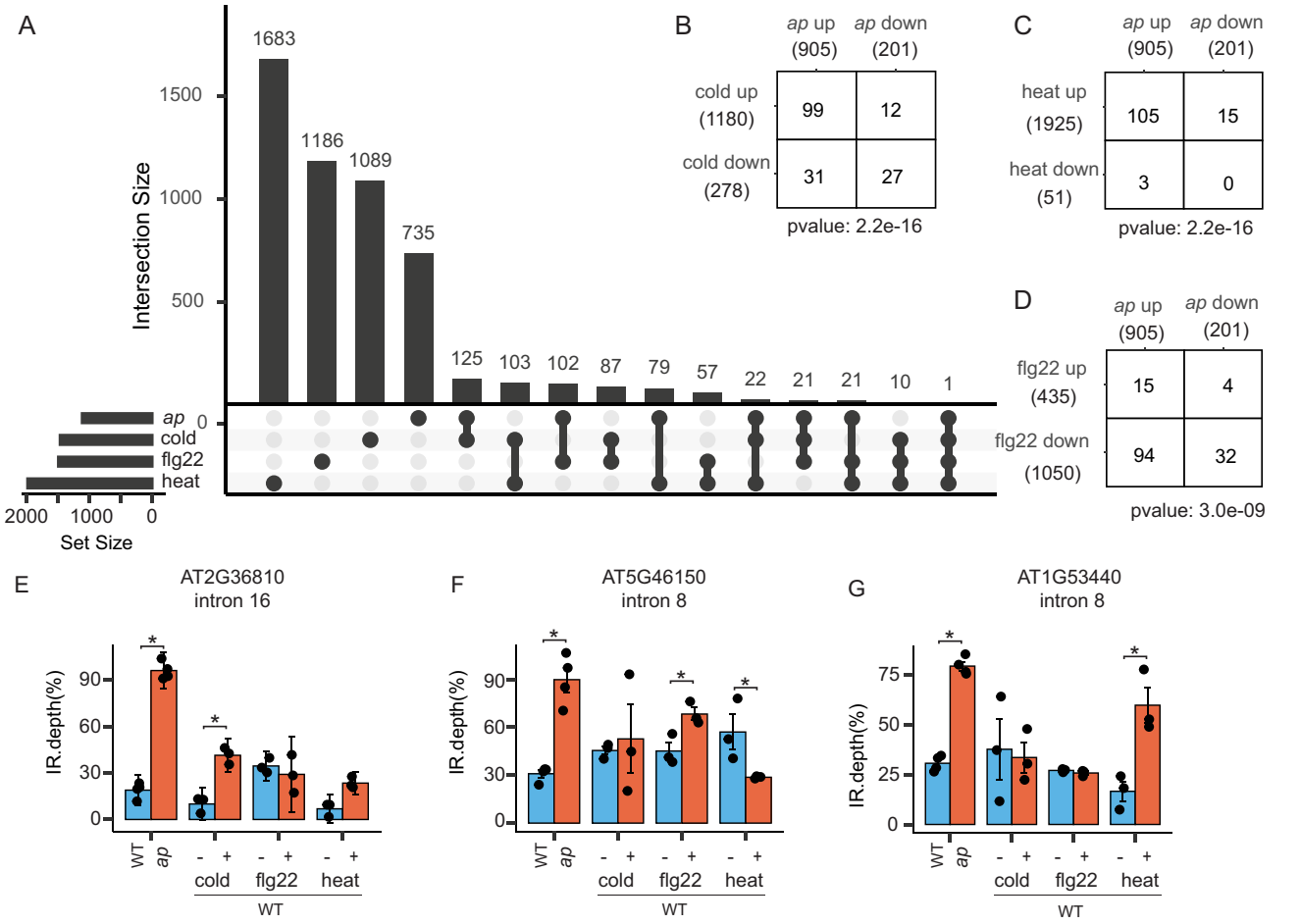


Figure 1. IR events in *acinus pinin* mutants overlap with stress-induced IR events in WT plants. (**A**) UpSet diagram showing the overlap of differentially retained introns between the *acinus pinin* mutant and WT plants subjected to cold, heat, or flg22 treatment. (**B**–**D**) Pairwise comparisons of upregulated and downregulated retained introns in the WT upon various stress conditions and in the *acinus pinin* mutant. (**E**–**G**) Representative examples of retained introns. Statistically significant differences (t-test) ($P \leq .05$) are indicated by an asterisk (*). Error bars represent the mean ± standard deviation. Sample sizes indicate the number of biological replicates ($n = 4$ for WT and the *ap* mutant; $n = 3$ for all other genotypes).

cesses, and differentially significant genes were tested against all genes detected in the respective dataset. Two statistical approaches were applied: a classic Fisher's exact test and topGO's weight01 algorithm. GO categories were considered statistically significant if the weight01 Fisher's test yielded a *P*-value ≤.05.

Structure display

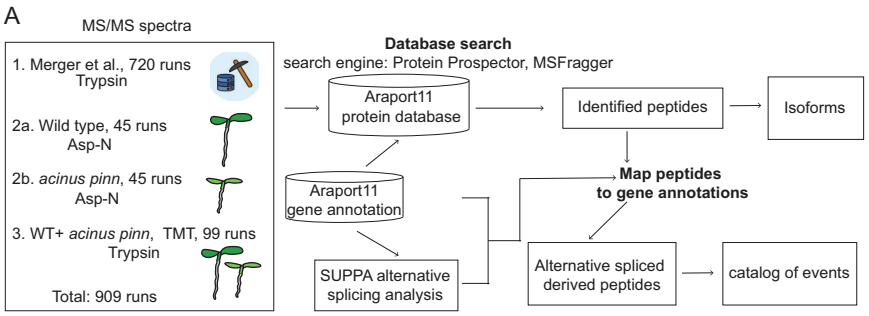
Protein structures were obtained from the Uniprot database as AlphaFold-predicted models and visualized using PyMOL (https://www.pymol.org/).

Results

Intron retention events detected in *acinus pinin* mutants are recapitulated in WT plants under abiotic or biotic stress

IR is the most prevalent type of AS in plants and is often responsive to stress [19, 25]. Therefore, we examined whether the altered IR events at the transcript level observed in the *acinus pinin* mutant overlap with stress-responsive events in WT plants. To accomplish this, we searched for altered IR events detected in the mutant using *PlantIntronDB* [56] and prioritized those reported in independent studies under similar stress conditions. Several altered IR events present in *acinus pinin* mutants occurred in response to cold, osmotic stress, dehydration, flagellin, fungal infection, or heat (Supplementary Fig. S1A–E and Supplementary Table S1).

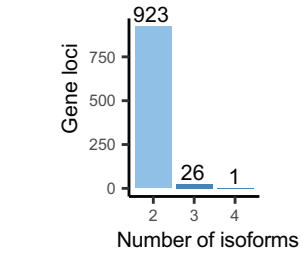
To evaluate the statistical significance of the observed overlaps, we reanalyzed a few publicly available RNA-seq datasets that met the following stringent criteria: at least three biological replicates; paired-end reads; sufficient sequencing depth; and stress treatments applied at the seedling stage (Supplementary Table S1). We observed significant overlap in IR events between *acinus pinin* mutants and WT plants subjected to cold (PRJNA1062790) [53], heat (PRJNA1262915), and flagellin (PRJNA1124505) [52] treatments (Fig. 1A–D and Supplementary Fig. S2A–C). While some IR events were shared across stress conditions, others were stress-specific (Fig. 1A, E–G), highlighting a complex and condition-dependent splicing landscape. Notably, IR events shared between the *acinus pinin* mutant and cold or heat treatments were predominantly upregulated (Fig. 1B and C). Conversely, IR events associated with the flagellin response exhibited an opposite pattern of regulation (Fig. 1D). However, this trend was not evident at the level of differentially expressed genes (Supplementary Fig. S2D–I).



B

Summary

Dataset	Proteins	Peptides	Isoform-specific peptides
1	22,165	377,990	26,008
2	9,917	41,728	3,222
3	15,658	188,152	12,099
Total unique	22,479	471,196	32,110



D

Name	Scheme	Annotated Events in CDS	Peptides Detected	Events Detected	Percentage
Retained intron		5,350	1,919	879	16.4%
Exon Skipping		706	255	192	27.2%
Alt. 5' Splice Site		1,505	813	437	31.4%
Alt. 3' Splice Site		2,688	1,474	852	31.7%
Alt. First Exon		103	167	58	56.3%
Alt Last Exon		36	39	13	36.1%
Mutually Exclusive Exon		26	19	11	42.3%
Total				2,442	

Figure 2. AS makes a substantial contribution to the diversity of proteins validated by proteomic analysis. (**A**) Computational workflow for identifying splice isoform-specific peptides and mapping AS events. A total of 909 nanoscale LC-MS/MS runs, including both trypsin- and AspN-digested samples, were analyzed to maximize proteome coverage, enable isoform quantification, and catalog AS events. Proteome-to-genome mapping strategy was performed and integrated with SUPPA-derived splicing coordinates for comprehensive AS annotation. (**B**) Summary table showing the number of proteins, total peptides, and isoform-specific peptides identified across three datasets, as well as the combined unique totals, highlighting large-scale peptide identification achieved in the *Arabidopsis* proteome. (**C**) Distribution of gene loci with proteomic evidence supporting 2, 3, or 4 isoforms based on isoform-specific peptides. (**D**) Summary of AS events annotated in Araport11 and detected by proteomics. Events are grouped by AS type within coding regions (CDS), with peptide identifications supporting each type. Detection percentages represent detected versus annotated events.

Together, these results imply that many IR events in the mutant mirror the physiological responses to both abiotic and biotic stress observed in WT plants. These findings imply ACINUS and PININ as key integrators of environmental signals in the regulation of AS. We included the *acinus pinin* mutant in our qualitative proteomic analysis to identify the splice variants, and in our quantitative analysis to evaluate the impact of IR on transcript and protein abundance.

Alternative splicing makes a substantial contribution to the diversity of proteins validated by proteomic analysis

To investigate whether splice variants are translated, we analyzed large-scale proteomics datasets, including data from the *Arabidopsis* draft proteome project [57]. This reference dataset, generated from 30 distinct tissues, involved tryptic digestion combined with extensive fractionation and comprised 720 nLC-MS/MS runs. In addition, we generated two complementary datasets: (i) 90 runs from AspN-digested, fractionated proteomes of WT and *acinus pinin* mutant seedlings, and (ii) 99 trypsin-digested TMT-labeled, fractionated proteome runs derived from pooled peptide samples of WT and *acinus pinin* mutants (Supplementary Fig. S3) (see the "Materials and methods" section). Together, these datasets encompass 909 nanoscale liquid chromatography-tandem mass spectrometry (nLC-MS/MS) runs (Fig. 2A).

All MS data were acquired using Orbitrap mass spectrometers, which provides high-resolution MS1 acquisition for more accurate identifications. The resulting MS data were searched against the Araport11 protein database [58], which provides a reannotation of the *Arabidopsis thaliana* genome. This more comprehensive database, which is larger than the canonical TAIR annotation, contains 48 354 protein-coding transcript isoforms encoded by 27 650 genes, representing a total of 40 779 non-redundant protein forms. Importantly, 31% of these protein-coding genes in this database encode more than one distinct isoform (Supplementary Fig. S4A), with the majority producing two (Supplementary Fig. S4B). To maximize peptide and protein identification, database searching was performed using two complementary search engines: Protein Prospector and MSFragger [54, 59, 60].

Using the *Arabidopsis* draft proteome and our in-house data, we identified a total of 471 196 unique peptides from 22 479 gene loci (Fig. 2B). We systematically refined this dataset through a multi-stage filtering pipeline (Supplementary Fig. S5A and B) to isolate peptides specific to particular protein isoforms. The filtering removed 280 916 peptides for single-isoform genes, 12 412 that were not locus-specific, and 145 758 peptides representing a cumulative total across many annotated proteins, where each peptide is shared among all isoforms of its corresponding gene locus. The final high-confidence dataset comprised 32 110 peptides that map either uniquely to a single protein isoform or to a subset of isoforms within the same gene locus (Fig. 2B and Supplementary Table S2). Subsequent mapping of these isoform-specific peptides (Supplementary Fig. S6A–C) suggested the expression of multiple protein variants at 950 gene loci (Fig. 2C). This anal-

ysis revealed 923 loci with two isoforms, 26 with three, and one with four, supporting a greater number of multi-isoform loci than was previously documented [57].

The classification of AS events and the extent of their translation in plants have remained unknown, despite previous large-scale proteomic studies in *Arabidopsis* [57, 61]. To investigate this, we developed and employed an integrated proteogenomics workflow with SUPPA event classification (Fig. 2A). Our robust pipeline directly mapped peptides derived from mass spectrometry to their genomic locations in the Araport11 genome, achieving high resolution by accounting for intron–exon junctions and codon ambiguity. Using SUPPA's splicing coordinates, we accurately classified isoform-specific peptides to identify AS event types (see the "Materials and methods" section) (Supplementary Table S2). This analysis identified 2442 AS events (Fig. 2D), categorized as follows: 879 IRs, 192 exon skippings, 437 alternative 5′ splice sites, 852 alternative 3′ splice sites, 58 alternative first exons, 13 alternative last exons, and 11 mutually exclusive exons. Importantly, this dataset provides the first direct proteomic evidence supporting the translation of numerous IR events in plants.

Representative alternative splicing events validated by proteomic evidence

Figure 3 presents representative cases of AS captured at the proteomic level. AspN digestion provided complementary sequence coverage by generating unique peptides that supported the presence and translation of distinct spliced isoforms.

In Auxilin-like 1, frame-preserving IR of intron 2 adds 27 amino acids to the N-terminal intrinsically disordered regions (Fig. 3A and Supplementary Fig. S8A and B). Peptide evidence supports expression of both the intron-retained (exon 2–intron 2 and intron 2–exon 3) and spliced out (exon 2–exon 3) isoforms, with five unique junction peptides identified. Notably, this specific intron has been previously annotated as an exitron, with the corresponding retained isoform characterized as the predominant (major) form [62].

Rubisco activase (RCA; Fig. 3B and C) has been re-examined, revealing a third isoform and revising our understanding of the previously defined RCA-β. While earlier work identified only the full-length RCA-α and a shorter RCA-β [15, 16], our proteomic data suggests the translation of three distinct isoforms. Junction peptides provided evidence for the full-length isoform (AT2G39730.1, 474 aa), derived from the exon 6–exon 7 junction, as well as a 441 aa isoform (AT2G39730.3), which results from intron 6 retention (exon 6–intron 6 junction) (Supplementary Fig. S9A–D). Furthermore, five peptides supported an intermediate 446 aa isoform (AT2G39730.2), generated by a frameshift from an alternative 5′ splice site. Our findings show that the previously considered single RCA-β isoform consists of these two shorter variants, both of which lack the redox-sensitive cysteines and are thus insensitive to redox regulation (Supplementary Fig. S10A) [15, 16]. Alternative explanations, such as a model involving transcription initiation at exon 6 of RCA combined with IR, cannot be formally excluded. However, both events occur at very low frequency, making their co-occurrence unlikely. We therefore consider the retained intron model to be the most parsimonious explanation for the observed peptides.

In LEAF RUST 10 DISEASE-RESISTANCE LOCUS RECEPTOR-LIKE PROTEIN KINASE-LIKE 3 (LRK10L3) [63] (Fig. 3D), multiple peptides supported the use of an alternative first exon. This suggests alternative transcription start site (TSS) usage, which generates isoforms with distinct N-terminal extracellular domains but identical transmembrane and C-terminal kinase regions (Supplementary Fig. S10B). This modular arrangement may enable recognition of diverse extracellular signals while retaining a common signaling core.

We also detected an alternative 3′ splice site in EIF2 BETA (Fig. 3E and Supplementary Fig. S11A) that inserts an additional glutamine residue and exon skipping in ribosomal protein P2Z (Fig. 3F and Supplementary Fig. S11B), further illustrating the diversity of AS events detectable by proteomics.

Collectively, these examples provide direct proteomic evidence that a wide spectrum of AS isoforms are translated *in vivo*. In addition to expanding proteome diversity, many of these isoforms likely alter protein structure and function, as shown by AlphaFold structure predictions, with potential biological consequences.

Detection of translated retained introns improves annotation of plant proteome

We show that IR, the most prevalent form of AS in plants, contributes to proteome diversity (Figs 2D and 3A and B). While many IR protein events are captured in the Araport11 annotation (Fig. 2D), a substantial proportion is missing. To systematically assess these unannotated events, we analyzed transcripts with an IR ratio greater than or equal to 15% (i.e. ≥15% of transcripts retaining the intron) in WT or *acinus pinin* seedlings using paired-end RNA-seq data from seedling tissues [23]. This analysis identified 4931 IR protein events, of which 2458 (49.8%) were annotated in Araport11 and 2473 (50.2%) were absent. Because proteomic searches depend on reference protein databases, the absence of unannotated IR protein sequences limits the ability to detect AS-derived peptides by mass spectrometry.

To address this limitation, we created an expanded protein database that incorporates unannotated IR protein events within coding regions (CDS) (Fig. 4A and B and Supplementary Fig. S12). Of the 1901 unannotated CDS-IR events, 1742 (91.6%) were predicted to produce truncated proteins, compared with only 509 (40.2%) of the 1265 annotated IR protein events in Araport11. Such truncated proteins typically result from premature stop codons or frameshifts introduced by retained introns, producing shorter polypeptides than the fully spliced isoforms.

We next searched the proteomic data against the combined Araport11 + IR-inclusive database. As noted previously, IR-derived peptides are often challenging to detect because exon–intron junctions are frequently enriched in K/R residues, which can obscure the identification of AS-specific peptides [42, 44]. Nevertheless, our analysis identified 106 unique peptides supporting 91 previously unannotated IR-derived proteins (Fig. 4B). For example, we detected peptides supporting translation of a truncated isoform (500 aa) of AT1G15280 (CASC3/Barentsz, an eIF4AIII-binding protein, BTZ2) compared with its fully spliced version (584 aa) (Fig. 4C and Supplementary Fig. S13A and B). BTZ2 protein is intrinsically disordered, which allows it to interact with various binding partners [64] (Supplementary Fig. S14A). Similarly, peptide evidence supported a truncated isoform (878 aa) of AT4G28650 (PXL2) compared with the full-length protein (1 013 aa) (Supplementary Fig. S15).

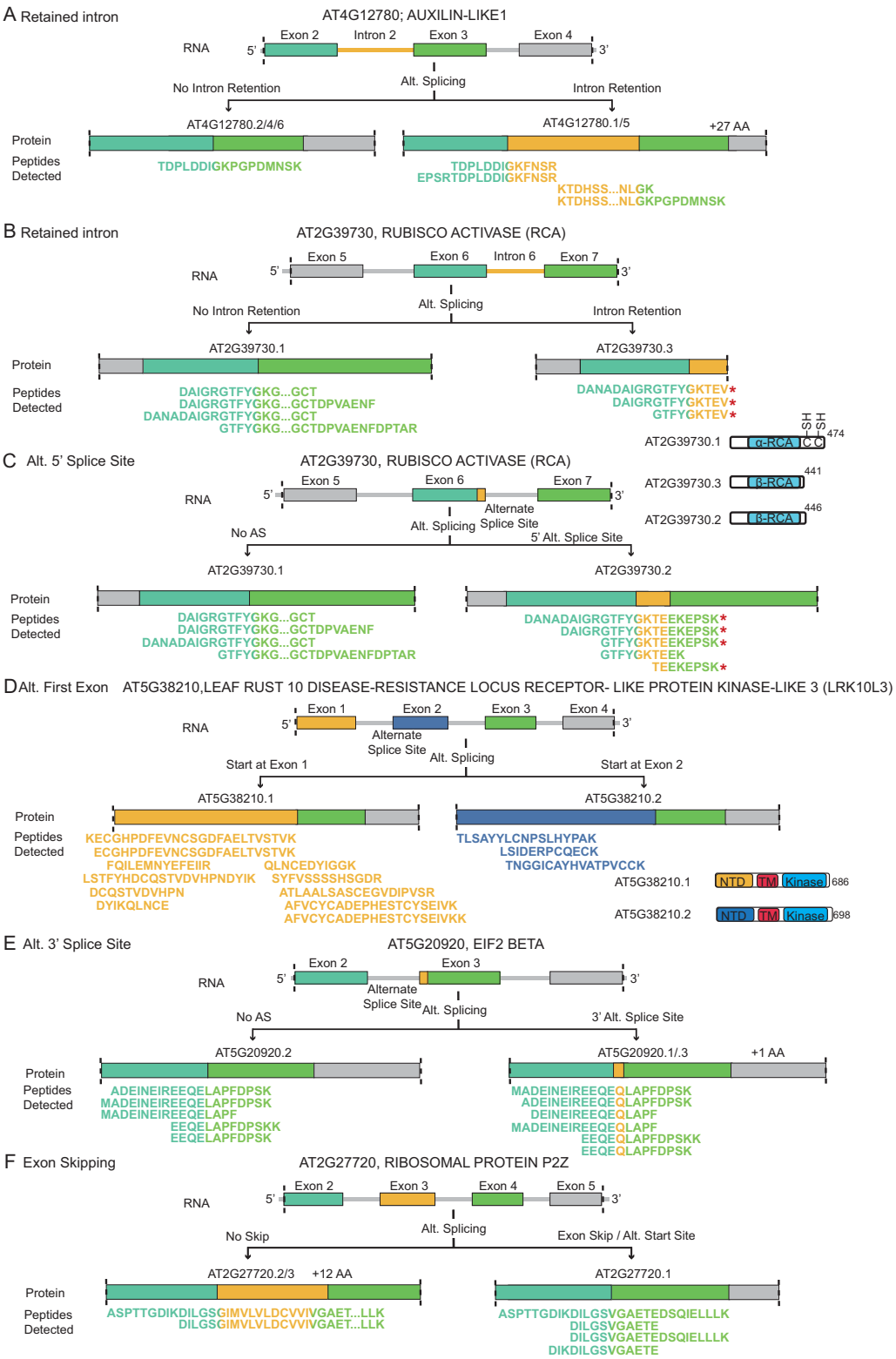


Figure 3. Representative AS events validated by proteomic evidence. Junction and isoform-specific peptides identified by deep proteomics provide direct evidence for the translation of both isoforms across multiple types of AS events: IR (**A**, **B**), alternative 5′ splice site (**C**), alternative first exon (**D**), alternative 3′ splice site (**E**), and exon skipping (**F**). (A) IR in *AUXILIN-LIKE1* generates two isoforms, one of which has a 27 amino acid (aa) extension translated from the retained intron. Both isoforms are supported by peptide evidence. (B, C) Peptide evidence supports the translation of three Rubisco activase (RCA) isoforms: a full-length isoform and two truncated isoforms. IR produces a 441 aa isoform, while the use of an alternative 5′ splice site introduces a frameshift, yielding a 446 aa isoform. Both truncated isoforms lack redox-sensitive cysteines and are not subject to redox regulation. (D) Proteomic support for both isoforms of LRK10L3 supports alternative first exon usage, producing isoforms with distinct N-terminal extracellular domains but identical transmembrane and C-terminal kinase domains. (E) Isoform-specific peptides support that an alternative 3′ splice site in EIF2 BETA generates two isoforms, one of which carries an additional glutamine residue (Q). (F) In ribosomal protein P2Z, peptide evidence validates the translation of both isoforms, with exon skipping producing a shorter isoform lacking 12 aa.

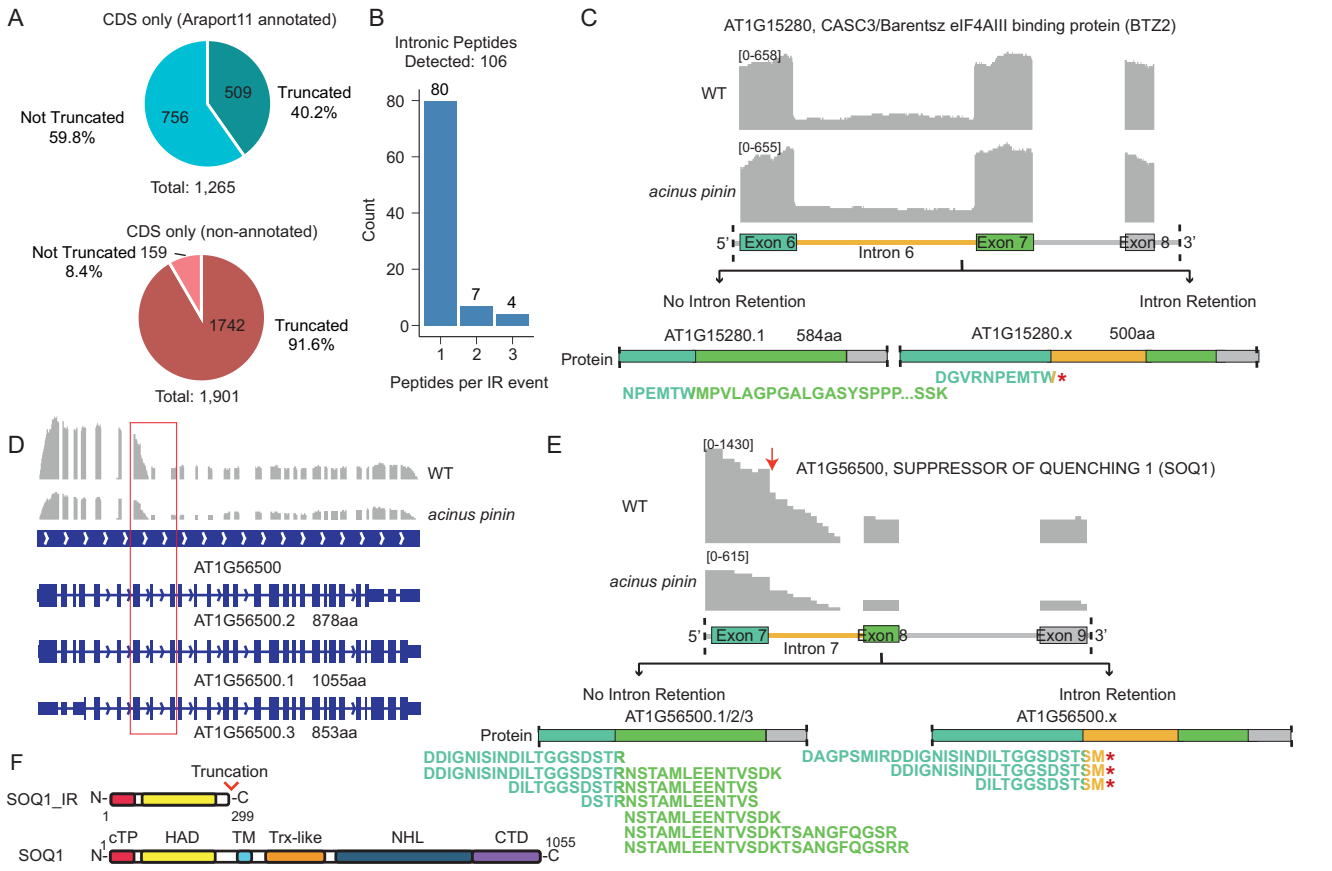


Figure 4. Detection of translated retained introns improves annotation of plant proteome. (**A**) Comparison of IR effects in annotated versus non-annotated CDS events. In Araport11 annotations, 40.2% of IR events are predicted to produce truncated proteins, compared with 91.6% of non-annotated events newly identified by RNA-seq. RNA-seq data were obtained from paired-end sequencing of WT and *acinus pinin* mutants [22]; events with a retention ratio ≥15% in either genotype were considered. (**B**) Expanded database in peptide searches identified 106 additional peptides, providing direct proteomic evidence for 91 previously unannotated IR events. (**C**) Example of truncated protein detection caused by retention of intron 6. Proteomics identified peptides from both the fully spliced and intron-retained isoforms in WT and *acinus pinin* mutants, supporting translation of both. A peptide spanning the exon–intron junction was detected in both WT and mutant backgrounds. (**D**) RNA-seq and Araport11 annotation of AT1G56500 (SOQ1). While all three annotated isoforms encode transcripts longer than 850 aa, RNA-seq shows a sharp drop in signal across Intron 7 and downstream exons, contradicting the annotation. (**E**) Proteomic evidence supports both the intron 7–retained truncated isoform and the fully spliced isoform of *SOQ1*. Three junction peptides spanning exon 7–intron 7 support the truncated isoform, while four junction peptides spanning exons 7–8 and three peptides from exon 8 support the fully spliced isoform. Retention of intron 7 introduces a stop codon that truncates the protein. (**F**) The truncated *SOQ1* isoform (SOQ1-IR) lacks the transmembrane, Trx-like, NHL, and CTD domains; the Trx and CTD domains confer redox sensitivity to SOQ1.

Integrating proteomics and RNA-seq also allowed us to refine existing genome annotations. For example, although Araport11 predicts three isoforms for AT1G56500 (SOQ1, *SUPPRESSOR OF QUENCHING1*), our data revealed an unannotated IR protein event in intron 7 that generates a truncated protein of 299 amino acids (Fig. 4D–F and Supplementary Figs S14B and S16A and B). RNA-seq coverage shows a sharp reduction in read density across intron 7 and downstream exons, while proteomics supported translation of the retained intron through three unique peptides, which introduce a stop codon. The data further suggests an alternative 3′ end processing within the retained intron. The truncated protein retains the chloroplast transit peptide and HAD domain of the full-length isoform but lacks the transmembrane, Trx-like, NHL, and CTD domains. Notably, the Trx and CTD domains confer redox sensitivity to SOQ1 (Supplementary Fig. S14B) [65, 66]. In addition, peptide evidence revealed novel isoforms of AT4G31980 (PPPDE thiol peptidase), AT3G50030 (ARM-repeat/TPR-like protein), and AT5G02360 (DC1 domain-containing protein) (Supplementary Figs S17–S19).

In sum, these results show that combining deep proteomics with RNA-seq uncovers previously hidden IR-derived proteins and refines gene model annotations, thereby expanding the known plant proteome.

Quantitative TMT proteomics reveals the non-linear impact of intron retention on the proteome

While we detected translation of many retained introns, the overall identification rate remains low, partly because most IR transcripts are only a small proportion of the total transcripts. To more comprehensively assess the effects of IR, we leveraged the AS mutant *acinus pinin* (*ap*). We performed genome-wide proteomic and transcriptomic profiling of WT and mutants (Fig. 5A and Supplementary Table S3), enabling a direct evaluation of how retained IR transcript influences protein abundance.

TMT quantified 11 565 proteins (13 850 protein groups) across five biological replicates (Supplementary Fig. S20), while RNA-seq measured 24 500 transcripts. Differential expression analysis identified hundreds of upregulated and

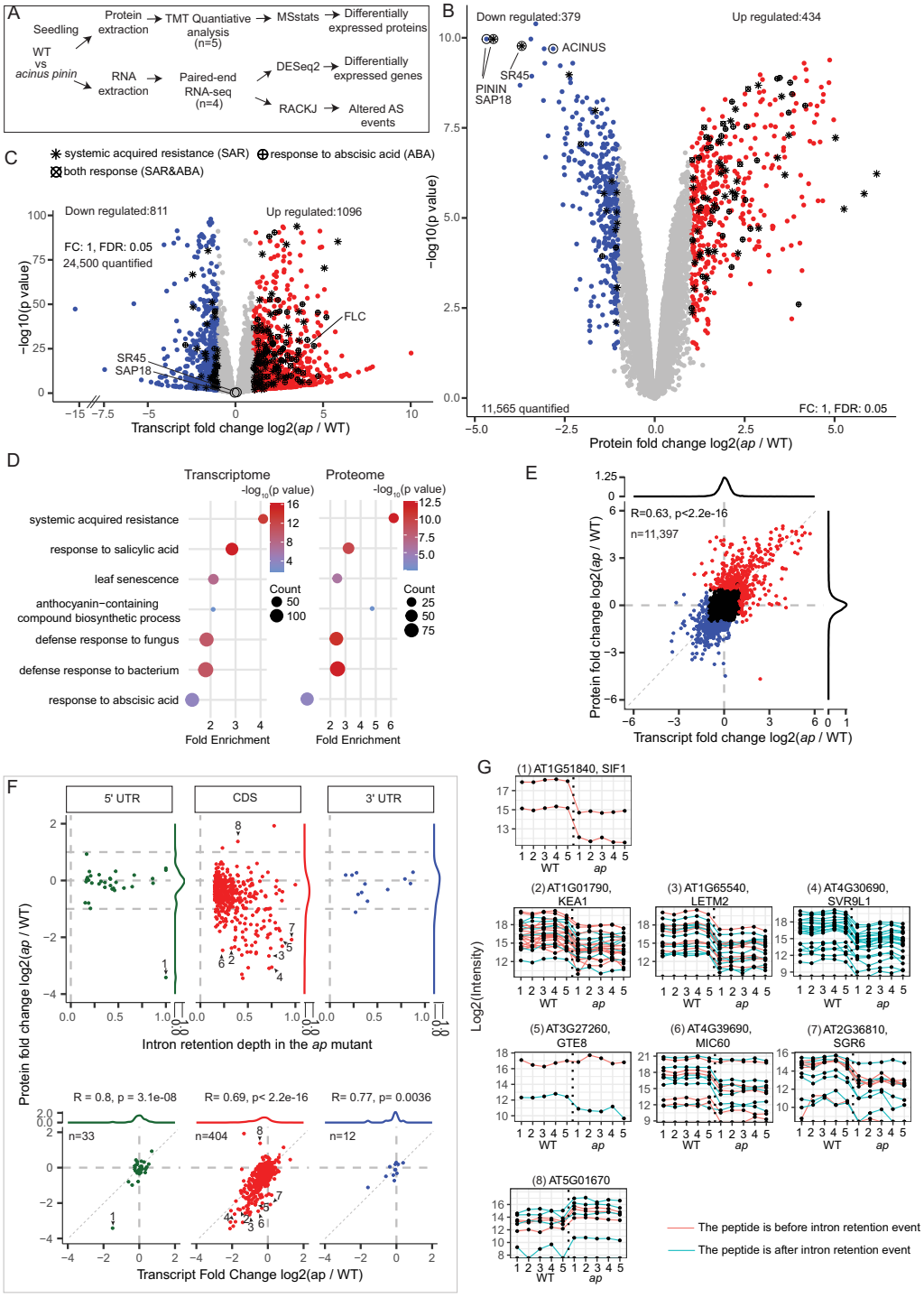


Figure 5. Quantitative TMT proteomics reveals the non-linear impact of IR on the proteome. (**A**) Genome-wide proteomic and transcriptomic profiling of WT and *acinus pinin* mutants. Proteomics was performed with five biological replicates, providing quantitative measurement of protein abundance, while transcriptomics used four biological replicates. (**B**, **C**) Differential expression analysis using TMT-based proteomics and RNA-seq identified hundreds of proteins and transcripts that were upregulated and downregulated, including known targets. Proteins and genes associated with systemic acquired resistance and/or ABA signaling are highlighted. Positive controls quantified by TMT (ACINUS, PININ, SR45, and SAP18) and by RNA-seq (FLC) are labeled. (**D**) GO analysis of both proteomic and transcriptomic data revealed additional pathways affected in *acinus pinin* mutants, beyond those reported in our prior study of ABA signaling. (**E**) Transcript–protein changes show a modest correlation (R = 0.63). (**F**) Nonlinear effects of retained introns that are increased in *acinus pinin* mutants on transcript and protein abundance. Protein abundance changes (*y* axis) were plotted against RNA changes (*x*-axis, bottom) and IR depth in the mutant (*x*-axis, top panel), with density plots also generated. Density plots show the distribution of changes in RNA (bottom panel) and protein (top panel) levels in the *acinus pinin* mutants, both shifted toward negative values for IR events in CDS. IR in untranslated regions (UTRs) had minimal impact on protein levels, except for *AT1G51840* (1). Retained introns in CDS regions typically reduced transcript abundance and caused stronger reductions at the protein level, as illustrated by *AT1G01790/KEA1*, *AT1G65540/LETM2*, and *AT4G30690/SVR9L1* (2–4). Some genes, e.g. *AT3G27260/GTE8*, *AT4G39690/MIC60*, and *AT2G36810/SGR6* (5–7), showed minimal transcript changes but substantial protein decreases. Conversely, IR occasionally increased protein levels without affecting transcripts, as observed for *AT5G01670* (8). (**G**) Peptide quantifications for the proteins highlighted in panel (F). Peptides upstream and downstream of retained introns are colored green and red, respectively.

A

Symbol	IIR	Avg mut IRdepth	Phenotype	Reference
SVR9L1	yes	0.75	*svr9-1 svr9l1-1/+* mutants exhibit pale green color	Zheng et al., 2016
LETM2	yes	0.72	*letm1 letm2/+* mutants exhibit slow growth	Zhang et al., 2012
ATNPC1-2	yes	0.58	*atnpc1-2atnpc1-1/+* mutants exhibit severe dwarfism	Feldman et al., 2015
KEA1	yes	0.33	*kea1-1 kea2-1* mutants exhibit pale green color and growth retardation	Aranda-Sicilia et al., 2016
ATLDA	yes	0.36	*atlda-2 atisa3-2* mutants exhibit delayed growth	Delatte et al., 2006
NHX5	yes	0.54	*nhx5 nhx6* mutants exhibit reduced growth	Wang et al., 2015
BE3	no		*be2-1be3-2* mutants exhibit pale green color and reduced size	Dumez et al., 2006
SAMT1	no		*samt1* mutants exhibit pale green leaf color and severe growth retardation	Bouvier et al., 2006

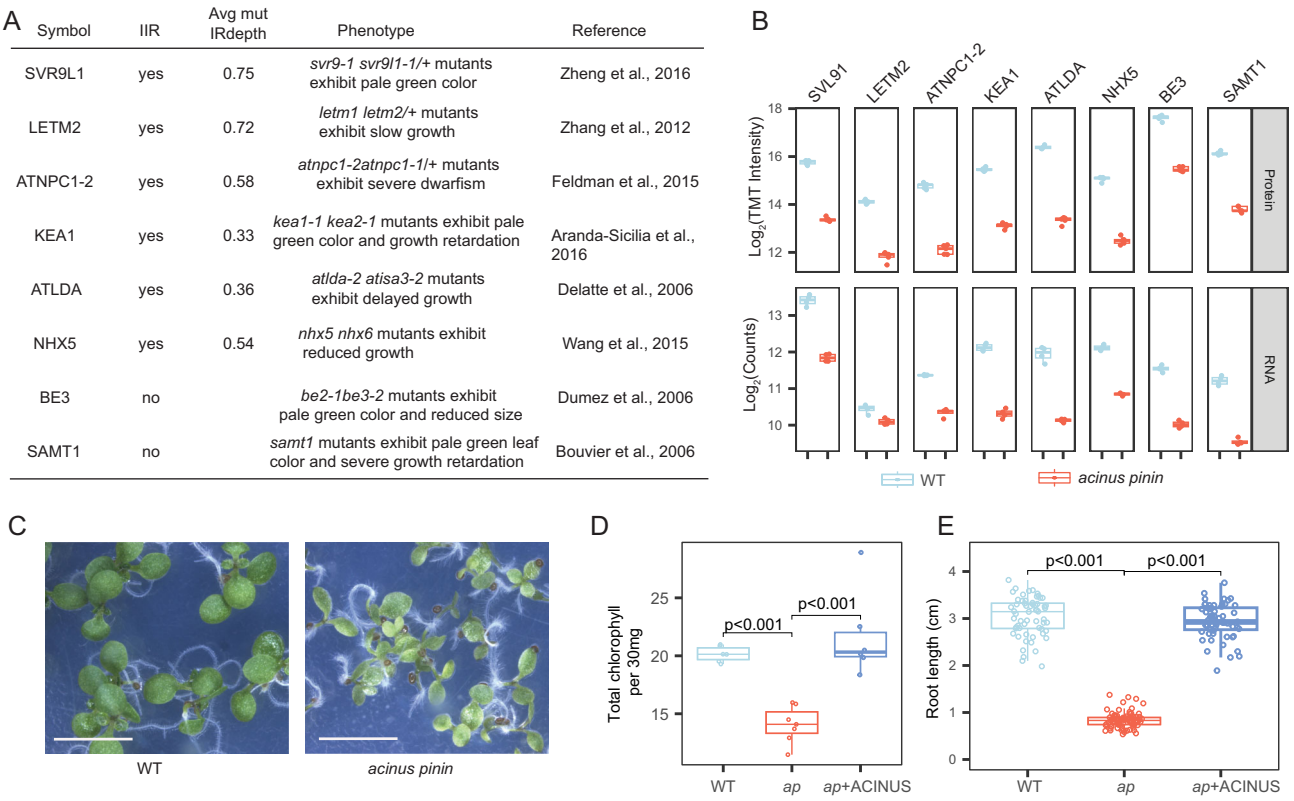


Figure 6. Cumulative effects of IR on protein abundance drive mutant phenotypes. (**A**) Proteins involved in regulating leaf color and plant growth. For genes with increased IR (IIR), the average IR depth in the *acinus pinin* mutant is included. The number of biological replicates was $n = 5$ for proteomic analyses and $n = 4$ for RNA analyses. (**B**) Corresponding reductions in protein and transcript levels for these genes, as revealed by quantitative proteomics ($\log_2$ TMT intensity) and RNA-seq of ($\log_2$ counts). (**C**) Phenotype of *acinus pinin* mutants showing reduced growth and pale-green leaves. Scale bar = 2 mm. Quantitative analysis of chlorophyll content (**D**) and root length (**E**) in WT, *acinus pinin* mutants, and the complemented line (+ACINUS). More than 60 biological samples were analyzed for each genotype.

downregulated proteins and transcripts, including known targets such as ACINUS, PININ, SR45, and SAP18 at the protein level and FLC at the transcript level (Fig. 5B and C). Integrative pathway analysis further showed that ACINUS and PININ regulate broad biological processes—including systemic acquired resistance, salicylic acid signaling, senescence, and anthocyanin biosynthesis—beyond their established role in ABA signaling (Fig. 5D and Supplementary Table S3). Transcript and protein changes were modestly correlated (Pearson R = 0.63, $P < 2.2 \times 10^{-16}$), indicating widespread post-transcriptional regulation (Fig. 5E).

We then quantified increased IR (IIR) transcript events in *acinus pinin* mutants, defined as ≥15% IR depth and ≥1.5-fold change, and assessed their impact on RNA and protein abundance (Fig. 5F). These IIRs occurred in 5′ UTRs (33 genes), CDS regions (404 genes), and 3′ UTRs (12 genes), in contrast to animal systems where IR is enriched in UTRs and non-coding RNAs [24, 38]. To visualize their impact, protein abundance changes (*y* axis) were plotted against RNA changes (*x*-axis, bottom) and IR depth in the mutant (*x*-axis, top panel), with density plots also generated. IR in UTRs generally did not alter protein levels, except for SIF1 (At1G51840), where IR caused a marked reduction (Fig. 5F and G). In contrast, CDS IR typically led to decreased transcript abundance and an even stronger reduction in protein abundance (e.g. KEA1, LETM2, SVR9L1; numbered 2–4), consistent with NMD-mediated degradation (Fig. 5F and G).

The analysis also revealed non-linear behaviors (Fig. 5F and G): some genes showed little change in RNA but strong reductions in protein abundance (e.g. GTE8, MIC60, SGR6; numbered 5–7), while others exhibited protein increases (e.g. AT5G01670; numbered 8). These patterns suggest that IR can alter proteome output through mechanisms beyond simple transcript degradation.

To further investigate, we annotated peptides relative to retained introns, reasoning that truncated proteins would generate differential peptide abundance upstream versus downstream of the IR site. For most genes, peptide levels changed concordantly in the mutant, suggesting that truncated proteins are rarely produced, unstable, or that current TMT methods lack the resolution or precision to detect such differences. In one isolated case (AT3G27260/GTE8), the detected upstream peptide remained unchanged while the downstream peptide was reduced, consistent with the production of a truncated protein, although such instances appear to be rare.

Overall, the effects of IR were non-linear: CDS IR strongly reduced mRNA and protein levels, UTR IR was largely neutral, and a subset of cases showed disproportionate effects at the protein level. These results indicate that IR exerts complex, context-dependent effects on gene expression and proteome output in *Arabidopsis*. For events with decreased IR, conclusions remain limited, likely due to the smaller number of detectable cases (Supplementary Fig. S21).

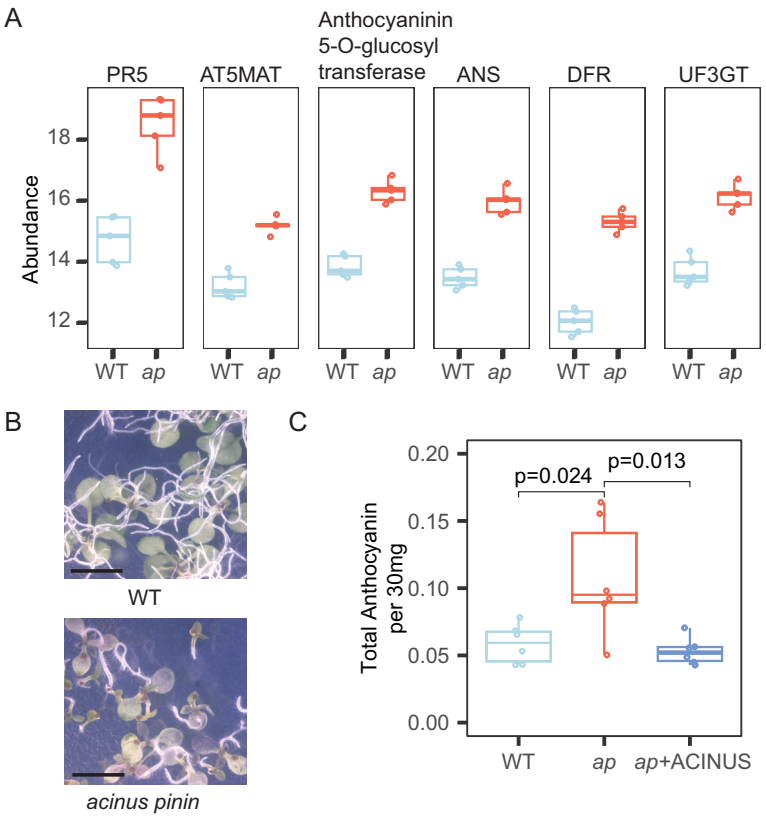


Figure 7. Upregulated proteins in anthocyanin biosynthesis correlate with increased anthocyanin accumulation in *acinus pinin* mutants. (**A**) Box plot showing the abundance of proteins associated with anthocyanin biosynthesis, as defined by GO analysis, in *acinus pinin* mutants (*ap*, red) and wild type (WT, blue). Protein levels are significantly higher in the mutant. The number of biological replicates was $n = 5$ for proteomic analyses. (**B**) Abaxial cotyledon images showing elevated anthocyanin accumulation in *acinus pinin* mutants. Scale bar, 2 mm. (**C**) Quantification of total anthocyanin content in WT, *acinus pinin* mutants, and the complemented line (+ACINUS), showing increased levels in the double mutant. Statistical significance was assessed using student's t-test. Six biological replicates were analyzed for each genotype.

Cumulative effects of intron retention on protein abundance drive mutant phenotypes

The functional impact of AS on individual proteins is often difficult to assess, partly due to redundancy within protein families and the challenge of knocking out a single isoform without affecting others. However, the striking phenotype of the *acinus pinin* mutant suggests that the altered AS and transcript change have functional consequences. We therefore tested whether the cumulative downregulation of multiple proteins contributes to *acinus pinin* mutant phenotypes. Analysis of proteomic and transcriptomic data (Fig. 6A and B) revealed that STRESS-RESPONSIVE PROTEIN 9-LIKE 1 (SVR9L1) [67], LEUCINE ZIPPER-EF-HAND-CONTAINING TRANSMEMBRANE PROTEIN 2 (LETM2) [68], NIEMANN-PICK DISEASE TYPE C1-2 (ATNPC1-2) [69], POTASSIUM EFFLUX ANTIPORTER 1 (KEA1), LIMIT DEXTRINASE/ATLDA [70], and SODIUM/HYDROGEN ANTIPORTER 5 (NHX5) [71] were downregulated at both transcript and protein levels due to significant IR, whereas BRANCHING ENZYME 3 (BE3) [72] and S-ADENOSYLMETHIONINE TRANSPORTER 1 (SAMT1) [73] were downregulated at transcript and protein levels without affecting the splicing. Notably, these genes have been previously reported to influence chlorophyll content and root growth when combined with additional mutations.

Guided by these proteomic insights, we measured chlorophyll content and root length, finding significant reductions in both traits in *acinus pinin* mutants that were rescued by ACINUS expression (Fig. 6C–E). These results indicate that the pleiotropic defects may arise from the combined reduction of multiple proteins through IR and other regulatory mechanisms, linking AS directly to functional outcomes.

Upregulated proteins in anthocyanin biosynthesis correlate with increased anthocyanin accumulation in *acinus pinin* mutants

Many splicing factors have been reported to influence transcription, although the underlying mechanisms remain unclear [43, 74]. Guided by transcriptomic and proteomic data, we found that ACINUS and PININ regulate the expression of numerous genes (Supplementary Fig. S2 and Supplementary Table S1) without affecting their AS, indicating regulation at the transcriptional level. Notably, several genes encoding proteins involved in anthocyanin-containing compound biosynthesis, as cataloged by GO analysis, were strongly upregulated in the mutants at both the transcript and protein levels (Figs 5D and 7A and Supplementary Table S3).

Based on these molecular changes, we predicted altered anthocyanin accumulation in the double mutants and examined their anthocyanin phenotype. Indeed, when grown on ½ MS plates, the mutants exhibited pronounced anthocyanin accumulation beneath the cotyledons, resulting in a distinct purple coloration and quantitatively higher anthocyanin content, both of which were fully rescued by ACINUS expression

(Fig. 7B and C). Because both transcript and protein levels are altered, we conclude that the anthocyanin phenotype in *acinus pinin* mutants is driven by transcriptional regulation. Whether ACINUS and PININ directly bind these loci to control expression remains unknown and will be addressed in future studies.

Discussion

For decades, researchers have debated whether AS represents a key evolutionary innovation or merely stochastic splicing noise. A central question is whether splice variants are translated and functionally relevant [28, 29]. Ribosome profiling studies suggest that many transcript variants are associated with polysomes [32–34, 36, 75], yet it remains unclear whether these translation events yield stable, functional proteins. Proteomics offers more direct evidence; however, technical constraints in coverage and sensitivity have historically limited conclusions [42], leading to conflicting reports [39–41].

In this study, we analyzed over 900 LC-MS/MS datasets—720 from public repositories and 189 generated in-house—using a custom-expanded library, dual protease digestion (trypsin and AspN), both label-free and TMT labeling, and two complementary search engines (Protein Prospector and MSFragger) (Figs 2 and 3). This strategy enabled the first large-scale identification of isoform-specific peptides in plants and the identification of 2533 splice events. Our findings support that diverse AS isoforms, particularly those involving IR, are translated and detectable at the proteome level. Although our dataset is smaller in scope than comparable human studies [44], it provides strong evidence that AS substantially expands the *Arabidopsis* proteome. Additionally, cases of IR followed by polyadenylation (e.g. SOQ1) highlight the complexity of transcript regulation. These findings suggest that coordinated control among transcription, AS, and polyadenylation collectively shapes mRNA and protein isoform outcomes. Notably, 58 detected events involved alternative first exons, which may arise from alternative TSS usage. Because tools like SUPPA classify these broadly as AS, caution is warranted when interpreting their mechanistic origin. While our results confirm the presence of isoform-specific peptides, the underlying regulatory processes may not always stem from classical AS alone. Ultimately, improved proteome annotation and the integration of long-read RNA sequencing will be essential for further resolving the complexity of plant AS.

Future efforts incorporating enzymes such as chymotrypsin and Lys-C, as suggested by the Ruan group's *in silico* analysis for junction peptide detection [76], will further enhance sequence coverage and deepen our understanding of how AS shapes the proteome. Notably, our large-scale analysis identified 32 110 trypsin/AspN isoform-unique peptides—an extensive resource that substantially expands experimental evidence for AS-derived proteoforms. Although this represents ~6% of peptides predicted *in silico*, the difference likely reflects inherent technical and biological constraints. Many theoretical peptides may be underrepresented due to losses during sample preparation (e.g. extreme hydrophilicity or hydrophobicity) or suboptimal ionization efficiency. In addition, numerous splice isoforms are dynamically regulated by environmental cues or restricted to specific cell types, which likely limit their detectability under standard growth conditions and bulk proteomic analyses.

The identification of protein isoforms containing retained introns is noteworthy, as many of these transcripts harbor premature termination codons (PTCs) typically targeted by NMD [25, 77–79]. Although protein annotation in this context remains incomplete and requires further refinement, their detection at the protein level indicates that a subset of these transcripts can bypass, evade, or be translated prior to canonical mRNA surveillance and degradation pathways. To determine whether these events correspond to exitrons—internal protein-CDS—we cross-referenced our findings with the 1002 exitrons previously defined in *Arabidopsis* [62]. Among the 879 IR-derived protein events detected in Fig. 2D, 56 overlap with annotated exitrons, whereas none of the 91 IR protein events in Fig. 4B show such overlap. Importantly, the identification of peptides mapping to retained intron sequences provides direct experimental evidence that at least a subset of these isoforms (i) are exported from the nucleus, (ii) evade—or are translated before—NMD-mediated degradation, and (iii) give rise to detectable protein products.

Beyond expanding proteomic diversity, our results support that AS, specifically IR, serves as a potent regulator of gene expression (Fig. 5F). We provide experimental evidence that retained introns not only reduce transcript abundance but frequently lead to a disproportionately severe reduction in protein levels. Notably, for a subset of genes, protein levels decreased substantially while mRNA levels remained stable, highlighting a non-linear relationship between IR and proteomic output.

This control mechanism suggests that AS performs a dual function: simultaneously broadening the repertoire of protein isoforms and fine-tuning gene expression in response to developmental and environmental signals [13]. While proving the relevance of individual isoforms remains difficult, established examples—such as dominant-negative truncated forms or regulatory "escape" variants like RCA proteins [15–17, 80]—support our findings. Our data also suggests that the cumulative effect of multiple AS events drives the observed growth and chloroplast defects in our mutants. By utilizing AS to manage a cost-to-benefit balance, plants can adjust their physiological responses through either transcript downregulation or the synthesis of novel protein variants. Our work thus provides a critical link between AS and proteome complexity in plants.

Funding

This work was funded by the National Institutes of Health grants (R01GM135706) and (S10OD030441) to S.-L.X. and diversity supplement to support A.V.R., and by the Carnegie Endowment Fund to the Carnegie Mass Spectrometry Facility. Funding to pay the Open Access publication charges for this article was provided by the National Institutes of Health (R01GM135706) and Carnegie Endowment Fund.

Acknowledgements

We thank Dr Jixian Zhai and Dr Jinbu Jia for providing the *Arabidopsis thaliana* IIR datasets for various biotic and abiotic conditions. We also thank Timothy Pham for critically reading the manuscript.

Author contributions: Andres V. Reyes (Data curation [lead], Formal analysis [lead], Methodology [lead], Project administration [lead], Validation [lead], Visualization [lead],

Writing—original draft [lead]), Christopher Zhang (Formal analysis [supporting]), Sumudu S. Karunadasa (Data curation [supporting]), Ruben Shrestha (Data curation [supporting]), Tarabryn S. Grismer (Data curation [supporting]), Danbi Byun (Resources [supporting]), and Shou-Ling Xu (Conceptualization [lead], Formal analysis [supporting], Funding acquisition [lead], Investigation [lead], Supervision [lead], Writing—original draft [lead], Writing—review & editing [lead])

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with accession numbers (PXD068586, PXD068589, and PXD068593). All other data is available from the corresponding author on reasonable request. The RNA-seq data have been deposited at NCBI Gene Expression Omnibus (GEO) database under the access number GSE286870. The custom scripts and library have been deposited to GitHub (https://github.com/shoulingxu-lab) and Zenodo as follows: IDratio: https://doi.org/10.5281/zenodo.19500582, Proteo-Genomic R Scripts: https://doi.org/10.5281/zenodo.19500534, TMT_Processing_MSstatsTMT: https://doi.org/10.5281/zenodo.19500303.

References

1. Pan Q, Shai O, Lee LJ *et al.* Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008;40:1413–5. https://doi.org/10.1038/ng.259
2. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 2010;463:457–63. https://doi.org/10.1038/nature08909
3. Zhang H, Jia J, Zhai J. Plant intron-splicing efficiency database (PISE): exploring splicing of approximately 1,650,000 introns in *Arabidopsis*, maize, rice, and soybean from approximately 57,000 public RNA-seq libraries. *Sci China Life Sci* 2023;66:602–11. https://doi.org/10.1007/s11427-022-2193-3
4. Salz HK. Sex determination in insects: a binary decision based on alternative splicing. *Curr Opin Genet Dev* 2011;21:395–400. https://doi.org/10.1016/j.gde.2011.03.001
5. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet* 2016;17:19–32. https://doi.org/10.1038/nrg.2015.3
6. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* 2017;18:437–51. https://doi.org/10.1038/nrm.2017.27
7. Chaudhary S, Jabre I, Reddy ASN *et al.* Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci* 2019;24:496–506. https://doi.org/10.1016/j.tplants.2019.02.006
8. Bonnal SC, Lopez-Oreja I, Valcarcel J. Roles and mechanisms of alternative splicing in cancer—implications for care. *Nat Rev Clin Oncol* 2020;17:457–74. https://doi.org/10.1038/s41571-020-0350-x
9. Godinho DP, Yanez RJR, Duque P. Pathogen-responsive alternative splicing in plant immunity. *Trends Plant Sci* 2024;30:615–28.
10. Bradley RK, Anczukow O. RNA splicing dysregulation and the hallmarks of cancer. *Nat Rev Cancer* 2023;23:135–55. https://doi.org/10.1038/s41568-022-00541-7
11. Wojtowicz WM, Flanagan JJ, Millard SS *et al.* Alternative splicing of drosophila dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 2004;118:619–33. https://doi.org/10.1016/j.cell.2004.08.021
12. Hattori D, Chen Y, Matthews BJ *et al.* Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. *Nature* 2009;461:644–8. https://doi.org/10.1038/nature08431
13. Yan Y, Luo H, Qin Y *et al.* Light controls mesophyll-specific post-transcriptional splicing of photoregulatory genes by AtPRMT5. *Proc Natl Acad Sci USA* 2024;121:e2317408121. https://doi.org/10.1073/pnas.2317408121
14. Cai YD, Liu X, Chow GK *et al.* Alternative splicing of Clock transcript mediates the response of circadian clocks to temperature changes. *Proc Natl Acad Sci USA* 2024;121:e2410680121. https://doi.org/10.1073/pnas.2410680121
15. Werneke JM, Chatfield JM, Ogren WL. Alternative mRNA splicing generates the two ribulosebisphosphate carboxylase/oxygenase activase polypeptides in spinach and *Arabidopsis*. *Plant Cell* 1989;1:815–25.
16. Kim SY, Stessman DJ, Wright DA *et al.* Arabidopsis plants expressing only the redox-regulated Rca-alpha isoform have constrained photosynthesis and plant growth. *Plant J* 2020;103:2250–62. https://doi.org/10.1111/tpj.14897
17. Wang Z, Ji H, Yuan B *et al.* ABA signalling is fine-tuned by antagonistic HAB1 variants. *Nat Commun* 2015;6:8138. https://doi.org/10.1038/ncomms9138
18. Braun CJ, Stanciu M, Boutz PL *et al.* Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell* 2017;32:411–26.e11. https://doi.org/10.1016/j.ccell.2017.08.018
19. Chaudhary S, Khokhar W, Jabre I *et al.* Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci* 2019;10:708. https://doi.org/10.3389/fpls.2019.00708
20. Ule J, Blencowe BJ. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell* 2019;76:329–45. https://doi.org/10.1016/j.molcel.2019.09.017
21. Marquez Y, Brown JW, Simpson C *et al.* Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 2012;22:1184–95. https://doi.org/10.1101/gr.134106.111
22. Petrillo E. Do not panic: an intron-centric guide to alternative splicing. *Plant Cell* 2023;35:1752–61. https://doi.org/10.1093/plcell/koad009
23. Shrestha R, Reyes AV, Carey S *et al.* Next-generation mapping of the ACINUS-mediated alternative splicing machinery and its regulation by o-glycosylation in *Arabidopsis*. bioRxiv, https://doi.org/10.1101/2025.01.04.631329, 26 March 2025, preprint: not peer reviewed.
24. Jacob AG, Smith CWJ. Intron retention as a component of regulated gene expression programs. *Hum Genet* 2017;136:1043–57. https://doi.org/10.1007/s00439-017-1791-x
25. Martin G, Marquez Y, Mantica F *et al.* Alternative splicing landscapes in arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. *Genome Biol* 2021;22:35. https://doi.org/10.1186/s13059-020-02258-y
26. Zhang H, Wang Y, Lu J. Function and evolution of upstream ORFs in eukaryotes. *Trends Biochem Sci* 2019;44:782–94. https://doi.org/10.1016/j.tibs.2019.03.002
27. Jia J, Long Y, Zhang H *et al.* Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants.

Nat. Plants 2020;6:780–8. https://doi.org/10.1038/s41477-020-0688-1
28. Tress ML, Abascal F, Valencia A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem Sci* 2017;42:98–110. https://doi.org/10.1016/j.tibs.2016.08.008
29. Blencowe BJ. The relationship between alternative splicing and proteomic complexity. *Trends Biochem Sci* 2017;42:407–8. https://doi.org/10.1016/j.tibs.2017.04.001
30. Resch A, Xing Y, Alekseyenko A *et al.* Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. *Nucleic Acids Res* 2004;32:1261–9. https://doi.org/10.1093/nar/gkh284
31. Irimia M, Weatheritt RJ, Ellis JD *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 2014;159:1511–23. https://doi.org/10.1016/j.cell.2014.11.035
32. Floor SN, Doudna JA. Tunable protein synthesis by transcript isoforms in human cells. *eLife* 2016;5:e10921. https://doi.org/10.7554/eLife.10921
33. Sterne-Weiler T, Martinez-Nunez RT, Howard JM *et al.* Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res* 2013;23:1615–23. https://doi.org/10.1101/gr.148585.112
34. Weatheritt RJ, Sterne-Weiler T, Blencowe BJ. The ribosome-engaged landscape of alternative splicing. *Nat Struct Mol Biol* 2016;23:1117–23. https://doi.org/10.1038/nsmb.3317
35. Yu H, Tian C, Yu Y *et al.* Transcriptome survey of the contribution of alternative splicing to proteome diversity in *Arabidopsis thaliana*. *Mol Plant* 2016;9:749–52. https://doi.org/10.1016/j.molp.2015.12.018
36. Tian C, Wang Y, Yu H *et al.* A gene expression map of shoot domains reveals regulatory mechanisms. *Nat Commun* 2019;10:141. https://doi.org/10.1038/s41467-018-08083-z
37. Boutz PL, Bhutkar A, Sharp PA. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* 2015;29:63–80. https://doi.org/10.1101/gad.247361.114
38. Braunschweig U, Barbosa-Morais NL, Pan Q *et al.* Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* 2014;24:1774–86. https://doi.org/10.1101/gr.177790.114
39. Severing EI, van Dijk AD, van Ham RC. Assessing the contribution of alternative splicing to proteome diversity in *Arabidopsis thaliana* using proteomics data. *BMC Plant Biol* 2011;11:82. https://doi.org/10.1186/1471-2229-11-82
40. Abascal F, Ezkurdia I, Rodriguez-Rivas J *et al.* Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput Biol* 2015;11:e1004325. https://doi.org/10.1371/journal.pcbi.1004325
41. Fesenko I, Khazigaleeva R, Kirov I *et al.* Alternative splicing shapes transcriptome but not proteome diversity in physcomitrella patens. *Sci Rep* 2017;7:2698. https://doi.org/10.1038/s41598-017-02970-z
42. Wang X, Codreanu SG, Wen B *et al.* Detection of proteome diversity resulted from alternative splicing is limited by trypsin cleavage specificity. *Mol Cell Proteomics* 2018;17:422–30. https://doi.org/10.1074/mcp.RA117.000155
43. Liu Y, Gonzalez-Porta M, Santos S *et al.* Impact of alternative splicing on the human proteome. *Cell Rep* 2017;20:1229–41. https://doi.org/10.1016/j.celrep.2017.07.025
44. Sinitcyn P, Richards AL, Weatheritt RJ *et al.* Global detection of human variants and isoforms by deep proteome sequencing. *Nat Biotechnol* 2023;41:1776–86. https://doi.org/10.1038/s41587-023-01714-x
45. Trincado JL, Entizne JC, Hysenaj G *et al.* SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* 2018;19:40. https://doi.org/10.1186/s13059-018-1417-1
46. Bi Y, Deng Z, Ni W *et al.* Arabidopsis ACINUS is O-glycosylated and regulates transcription and alternative splicing of regulators of reproductive transitions. *Nat Commun* 2021;12:945. https://doi.org/10.1038/s41467-021-20929-7
47. Trinh CS, Jeong CY, Lee WJ *et al.* Paenibacillus pabuli strain P7S promotes plant growth and induces anthocyanin accumulation in *Arabidopsis thaliana*. *Plant Physiol Biochem* 2018;129:264–72. https://doi.org/10.1016/j.plaphy.2018.06.001
48. Xu S-L, Chalkley RJ, Maynard JC *et al.* Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in *Arabidopsis*. *Proc Natl Acad Sci USA* 2017;114:E1536–43.
49. Zecha J, Satpathy S, Kanashova T *et al.* TMT labeling for the masses: a robust and cost-efficient, in-solution labeling approach. *Mol Cell Proteomics* 2019;18:1468–78. https://doi.org/10.1074/mcp.TIR119.001385
50. Furtwangler B, Uresin N, Motamedchaboki K *et al.* Real-time search-assisted acquisition on a tribrid mass spectrometer improves coverage in multiplexed single-cell proteomics. *Mol Cell Proteomics* 2022;21:100219. https://doi.org/10.1016/j.mcpro.2022.100219
51. Schweppe DK, Eng JK, Yu Q *et al.* Full-featured, real-time database searching platform enables fast and accurate multiplexed quantitative proteomics. *J Proteome Res* 2020;19:2026–34. https://doi.org/10.1021/acs.jproteome.9b00860
52. Kim SI, Ma X, Kong L *et al.* Global profiling of CPL3-mediated alternative splicing reveals regulatory mechanisms of DGK5 in plant immunity and phosphatidic acid homeostasis. *Genome Biol* 2025;26:65. https://doi.org/10.1186/s13059-025-03529-2
53. Meena SK, Quevedo M, Nardeli SM *et al.* Antisense transcription from stress-responsive transcription factors fine-tunes the cold response in *Arabidopsis*. *Plant Cell* 2024;36:3467–82. https://doi.org/10.1093/plcell/koae160
54. Grismer T, Karunadasa S, Shrestha R *et al.* Workflow enhancement of TurboID-mediated proximity labeling for SPY signaling network mapping. bioRxiv, https://doi.org/10.1101/2024.02.17.580820, 18 February 2024, preprint: not peer reviewed.
55. Huang T, Choi M, Tzouros M *et al.* MSstatsTMT: statistical detection of differentially abundant proteins in experiments with isobaric labeling and multiple mixtures. *Mol Cell Proteomics* 2020;19:1706–23. https://doi.org/10.1074/mcp.RA120.002105
56. Bi Y, Shrestha R, Zhang Z *et al.* SPINDLY mediates O-fucosylation of hundreds of proteins and sugar-dependent growth in *Arabidopsis*. *Plant Cell* 2023;35:1318–33. https://doi.org/10.1093/plcell/koad023
57. Mergner J, Frejno M, List M *et al.* Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature* 2020;579:409–14. https://doi.org/10.1038/s41586-020-2094-2
58. Cheng CY, Krishnakumar V, Chan AP *et al.* Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* 2017;89:789–804. https://doi.org/10.1111/tpj.13415
59. Shrestha R, Reyes AV, Baker PR *et al.* 15N metabolic labeling quantification workflow in *Arabidopsis* using protein prospector. *Front Plant Sci* 2022;13:832562. https://doi.org/10.3389/fpls.2022.832562
60. Kong AT, Leprevost FV, Avtonomov DM *et al.* MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* 2017;14:513–20. https://doi.org/10.1038/nmeth.4256
61. van Wijk KJ, Leppert T, Sun Q *et al.* The arabidopsis peptideatlas: harnessing worldwide proteomics data to create a comprehensive community proteomics resource. *Plant Cell* 2021;33:3421–53. https://doi.org/10.1093/plcell/koab211
62. Marquez Y, Hopfler M, Ayatollahi Z *et al.* Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res* 2015;25:995–1007. https://doi.org/10.1101/gr.186585.114
63. Dong X, Xu J, Fu Y *et al.* LRK10L3 and BAK1 are collaboratively involved in extracellular ATP-regulated seedling growth of arabidopsis thaliana. *New Phytol* 2025;247:2709–25 https://doi.org/10.1111/nph.70352
64. Rademacher DJ, Bello AI, May JP. CASC3 biomolecular condensates restrict turnip crinkle virus by limiting host factor

availability. *J Mol Biol* 2023;435:167956. https://doi.org/10.1016/j.jmb.2023.167956
65. Amstutz CL, Fristedt R, Schultink A *et al.* An atypical short-chain dehydrogenase-reductase functions in the relaxation of photoprotective qH in *Arabidopsis*. *Nat Plants* 2020;6:154–66. https://doi.org/10.1038/s41477-020-0591-9
66. Hao J, Johansson A, Svensson Fall J *et al.* Suppressor of quenching 1 functions as a methionine sulfoxide reductase in the chloroplast lumen for regulation of photoprotective qH in *Arabidopsis*. bioRxiv, https://doi.org/10.1101/2024.11.01.621559, 2 November 2024, preprint: not peer reviewed.
67. Zheng M, Liu X, Liang S *et al.* Chloroplast translation initiation factors regulate leaf variegation and development. *Plant Physiol* 2016;172:1117–30.
68. Zhang B, Carrie C, Ivanova A *et al.* LETM proteins play a role in the accumulation of mitochondrially encoded proteins in *Arabidopsis thaliana* and AtLETM2 displays parent of origin effects. *J Biol Chem* 2012;287:41757–73. https://doi.org/10.1074/jbc.M112.383836
69. Feldman MJ, Poirier BC, Lange BM. Misexpression of the niemann-pick disease type C1 (NPC1)-like protein in arabidopsis causes sphingolipid accumulation and reproductive defects. *Planta* 2015;242:921–33. https://doi.org/10.1007/s00425-015-2322-4
70. Delatte T, Umhang M, Trevisan M *et al.* Evidence for distinct mechanisms of starch granule breakdown in plants. *J Biol Chem* 2006;281:12050–9. https://doi.org/10.1074/jbc.M513661200
71. Wang L, Wu X, Liu Y *et al.* AtNHX5 and AtNHX6 control cellular K^+ and pH homeostasis in *Arabidopsis*: three conserved acidic residues are essential for K^+ transport. *PLoS One* 2015;10:e0144716. https://doi.org/10.1371/journal.pone.0144716
72. Dumez S, Wattebled F, Dauvillee D *et al.* Mutants of arabidopsis lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. *Plant Cell* 2006;18:2694–709. https://doi.org/10.1105/tpc.105.037671
73. Bouvier F, Linka N, Isner JC *et al.* Arabidopsis SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *Plant Cell* 2006;18:3088–105. https://doi.org/10.1105/tpc.105.040741
74. Wang Y, Liu J, Huang BO *et al.* Mechanism of alternative splicing and its regulation. *Biomed Rep* 2015;3:152–8. https://doi.org/10.3892/br.2014.407
75. Xu Y, Gristwood T, Hodgson B *et al.* Archaeal orthologs of Cdc45 and GINS form a stable complex that stimulates the helicase activity of MCM. *Proc Natl Acad Sci USA* 2016;113:13390–5. https://doi.org/10.1073/pnas.1613825113
76. Puliasis SS, Lewandowska D, Hemsley PA *et al.* ProtView: a versatile tool for in silico protease evaluation and selection in a proteomic and proteogenomic context. *J Proteome Res* 2023;22:2400–10. https://doi.org/10.1021/acs.jproteome.3c00135
77. Kalyna M, Simpson CG, Syed NH *et al.* Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res* 2012;40:2454–69. https://doi.org/10.1093/nar/gkr932
78. Drechsel G, Kahles A, Kesarwani AK *et al.* Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the arabidopsis steady state transcriptome. *Plant Cell* 2013;25:3726–42. https://doi.org/10.1105/tpc.113.115485
79. Wang BB, Brendel V. Genomewide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci USA* 2006;103:7175–80. https://doi.org/10.1073/pnas.0602039103
80. Kelemen O, Convertini P, Zhang Z *et al.* Function of alternative splicing. *Gene* 2013;514:1–30. https://doi.org/10.1016/j.gene.2012.07.083

Received: October 22, 2025. **Revised:** April 2, 2026. **Accepted:** April 9, 2026
© The Author(s) 2026. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.