# Cytoplasmic U5 snRNA promotes the heterotrophic-to-autotrophic transition via selective translational activation

Yuntong Jiao[1,2,3#], Yangyang Fan[1,10#], Ruikai Zhang[1,4#], Kai He[1,4#], Bei Liu[5#], Yubo Zhang[2#], Yuqiu Wang[1,11], Jiahao Fan[6,7], Yanyan Gong[1], Mengfei Song[2,12], Lihua Zhao[4], Jie Huang[3], Han Wen[3,7], Wei Zhang[2,8,9], Chuan He[5], Xing Wang Deng[1,4], Danmeng Zhu[*2,3]

Small nuclear RNAs (snRNAs) are best known as essential components of the nuclear spliceosome, where they catalyze pre-mRNA splicing, yet their broader regulatory potential has remained unclear. Here, we show that U5 snRNA and its cognate Sm protein partners are evolutionarily co-opted to regulate cytoplasmic translation. Light promotes U5 snRNA accumulation on polysomes, enhancing translation of nuclear-encoded mRNAs for chloroplast proteins during the heterotrophic-to-autotrophic transition, largely independent of changes in mRNA abundance. U5 snRNA base-pairs with 3′UTRs of target mRNAs via the conserved Sm site; Brassicaceae-derived cytoplasmic SmD3a couples U5 snRNA–Sm protein assemblies to eEF2, the spliceosomal SNU114 homolog, to enable translational activation. Building on this mechanism, we adapted the U5 snRNA/Sm module for the non-Brassicaceae plant *Nicotiana benthamiana*, which lacks an SmD3a homolog, by engineering an SmD3 variant with enhanced cytoplasmic localization through artificial intelligence-assisted rational design and optimizing U5 snRNA for target mRNA base-pairing. Transient expression of this module elevated endogenous chloroplast protein levels and increased the maximum photochemical efficiency of photosystem II by over 20%. This work reveals that spliceosomal RNAs act beyond splicing to modulate translation and suggests that manipulating the U5 snRNA/Sm module could be a viable strategy for improving photosynthetic performance.

https://doi.org/10.15302/vita.2026.10.0081

## INTRODUCTION

Eukaryotic U-type small nuclear RNAs (snRNAs), including U1, U2, U4, U5 and U6 snRNAs, are highly conserved non-coding RNAs best known for their role in the spliceosome, where they catalyze the splicing of pre-mRNAs[1,2]. These snRNAs undergo a key maturation step in the cytoplasm, where they assemble with Smith antigen (Sm) proteins through the single-stranded Sm site element to form the small nuclear ribonucleoprotein (snRNP) complexes[3]. The Sm proteins, initially identified as nuclear antigens in systemic lupus erythematosus patients[4], form a heptameric ring that is essential for the structural integrity and function of snRNPs. After assembly, the snRNPs return to the nucleus to catalyze pre-mRNA splicing[5,6].

Observations of snRNA accumulation in the cytoplasm suggest potential noncanonical roles for these RNAs beyond the nucleus. For instance, cytoplasmic accumulation of snRNAs has been reported in *Drosophila* and mammalian cells[7,8], and viral snRNA-like RNAs have been shown to inhibit translation to suppress host cell apoptosis[9]. Additionally, mRNA–RNP complexes containing Sm proteins have been identified in the cytoplasm of European larch cells[10]. However, the precise cytoplasmic functions of snRNAs remain largely undefined.

The transition from heterotrophic to autotrophic growth in angiosperms is critical for early seedling survival. Light is the most important environmental signal and the energy source for this process. After germination under the soil, seedlings lack functional chloroplasts and depend on seed reserves to fuel heterotrophic growth until exposure to light that initiates chloroplast biogenesis[11,12]. Within 24 h, etioplasts in the cotyledons develop into functional chloroplasts, enabling photosynthesis. The establishment of functional chloroplasts depends on protein synthesis from both nuclear and chloroplast genomes. More than 90% of the required proteins are encoded by nuclear genes, and their expression needs to occur at the appropriate time during this process[13,14]. Yet how the translation of these nuclear-encoded plastid mRNAs is regulated during this developmental transition remains not fully understood.

Here, we identify a cytoplasmic function for the spliceosomal U5 snRNA in the regulation of mRNA translation. We

1. National Key Laboratory of Wheat Improvement, School of Advanced Agricultural Sciences and School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China. 2. State Key Laboratory of Gene Function and Modulation Research, School of Life Sciences, Peking University, Beijing, China. 3. Beijing Advanced Center of RNA Biology (BEACON), School of Life Sciences, Peking University, Beijing, China. 4. National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong, China. 5. Department of Chemistry, The University of Chicago, Chicago, IL, USA. 6. School of Mathematical Science, Peking University, Beijing, China. 7. AI for Science Institute, Beijing, China. 8. Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. 9. Medog Biodiversity Observation and Research Station of Xizang Autonomous Region, Nyingchi, Xizang, China. 10. Present address: Institute of Plant Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. 11. Present address: Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. 12. Present address: Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang, China. #These authors contributed equally. *Correspondence: Danmeng Zhu (zhudanmeng@pku.edu.cn)

**Received: March 26, 2026; Accepted: October 15, 2026; Published: November 3, 2026**

show that U5 snRNA associates with polysomes and promotes the translation of mRNAs encoding nuclear-encoded plastid proteins required for the heterotrophic-to-autotrophic transition. This activity requires direct interactions between U5 snRNA and its target mRNAs at their 3′UTRs and depends on the Sm proteins associated with polysomes. These findings reveal an unexpected layer of regulation whereby spliceosomal RNAs associate with the translational machinery, suggesting a broader principle that conserved housekeeping RNA–protein complexes can be repurposed to regulate selective mRNA translation beyond their canonical roles.

## RESULTS

### Light promotes polysome association of specific spliceosomal snRNAs

To identify light-induced changes in the RNA landscape, we profiled transcriptome-wide RNA abundance across different subcellular compartments: the nucleus, ribosome-free cytoplasm, monosomes, and polysomes. This unbiased analysis revealed the presence of U-type spliceosomal snRNAs not only in the nucleus but also in the ribosome-free cytoplasmic, monosomal and polysomal fractions (Supplementary Fig. S1a). Among the five U-type snRNA species (U1, U2, U4, U5 and U6), over 90% were more abundant in the ribosome-free cytoplasm than in the nucleus in white light-grown seedlings compared with dark-grown seedlings (Fig. 1a, b; Supplementary Table S1). Northern blot analysis confirmed these light-dependent changes of U5 snRNAs, as elevated cytoplasmic levels were evident in light-grown but not in dark-grown seedlings, or in mature plants (Supplementary Fig. S1b). Strikingly, among U type snRNAs, only all U5 and most U6 snRNAs became associated with polysomes in the light, in contrast to their monosomal association in darkness (Fig. 1b). U5 snRNAs were more abundant than U6 snRNAs in polysomes (Supplementary Fig. S1c) and were therefore selected for further functional analysis. U5 snRNAs comprised 14 variants of different abundance (Fig. 1c; Supplementary Table S2), all of which showed increased association with polysomes relative to monosomes in light-exposed seedlings; the five most abundant are shown as representatives (Fig. 1d). Among these, U5-3 and U5-4 were selected for functional analysis because they could be disrupted without affecting adjacent protein-coding genes. Besides nuclear-cytoplasmic fractionation, this light-dependent polysome association of U5 snRNAs was also confirmed by sucrose gradient-based polysome profiling. This association was time-dependent, as shown for U5-3 and U5-4 snRNAs, becoming evident within 12 h of illumination without changes in total abundance (Fig. 1e, f), and required phyA, phyB, and CRY1 photoreceptors[15,16] (Fig. 1g, h). Polysome profiling and RNA pull-down assays also confirmed the detection of specific association of U5 snRNAs with ribosomal proteins in the light (Supplementary Fig. S1d–g). Together, these results establish a light-mediated pathway that promotes U5 snRNA association with polysomes, suggesting a functional expansion beyond their nuclear role.

### A translational role for U5 snRNAs during seedling greening

To examine the physiological role of U5 snRNAs, we characterized *Arabidopsis* mutants during de-etiolation, the transition from heterotrophic to autotrophic growth upon light exposure. While single *u5-3* or *u5-4* mutants were phenotypically wild type (WT), the *u5-3 u5-4* double mutant exhibited a pronounced defect in seedling greening, seen as yellow-green cotyledons and reduced greening rates (Fig. 2a–c). This phenotype was pronounced at 24 h of light illumination following continuous dark growth (DTL-24), coincided with maximal U5 snRNA polysome association, and was fully rescued by U5-3 or U5-4 overexpression (Fig. 2a–c). Thus, U5 snRNAs contribute to efficient seedling greening during the transition to autotrophic growth.

We next investigated the role of U5 snRNA association with polysomes during seedling greening. Integrated transcriptome, translatome, and proteome analyses in *u5-3 u5-4* mutants grown under DTL-24 condition revealed a translational defect affecting 242 genes, with reduced translation efficiency (TE). Changes in TE in *u5-3 u5-4* mutant relative to WT were largely uncoupled from changes in mRNA levels (R = −0.04; Fig. 2d, e; Supplementary Tables S3, S4). Notably, the vast majority of genes with reduced TE showed no significant change in mRNA levels (Fig. 2f). Genes with differential alternative splicing events showed mild changes of TE (Fig. 2g; Supplementary Table S5). These results suggest that the observed changes in TE in *u5-3 u5-4* mutants are not simply explained by changes in mRNA abundance or detectable alternative splicing changes, although subtle splicing-dependent indirect effects cannot be fully excluded. Together, these findings support an additional role of U5 snRNA in translational regulation.

To explore the specific pathways regulated by U5 snRNAs at the translational level, we conducted gene ontology (GO) enrichment analysis. Strikingly, this analysis revealed that transcripts with reduced TE were enriched for nuclear-encoded plastid proteins, including key genes essential for chloroplast development and photosynthetic function such as *LIGHT-HARVESTING CHLOROPHYLL A/B BINDING 3* (*LHCB3*), *PHOTOSYNTHETIC NDH SUBCOMPLEX B2* (*PnsB2*), and *CHLOROPLAST RIBOSOMAL PROTEIN S1* (*RPS1*) (Fig. 2h). Integrated mRNA-seq and proteomic analyses revealed reduced protein abundance relative to mRNA levels for representative genes in the *u5-3 u5-4* mutant (Fig. 2i), despite unchanged or increased mRNA levels and no detectable changes in splicing (Supplementary Fig. S2a, b). In WT, the accumulation of proteins such as RPS1, LHCB3, and PnsB2 from DTL-12 and DTL-24 exceeded the corresponding mRNA increase (Supplementary Fig. S2c, d), this enhanced translation was reduced in *u5-3 u5-4* mutants (Supplementary Fig. S2e), indicating that U5 snRNAs enhance protein biosynthesis during this period.

To validate the translational regulation by U5 snRNAs, we selected six representative genes encoding chloroplast ribosomal and photosynthetic proteins. Polysome profiling assay showed significantly decreased polysome association for all six mRNAs in the *u5-3 u5-4* mutant (Fig. 2j). For three of these genes with available antibodies, western blots and bioorthogonal noncanonical amino acid tagging (BONCAT) assays, which detect newly synthesized proteins, confirmed reduced protein abundance and protein synthesis rates (Fig. 2k, l). We further performed ribosome sequencing (ribo-seq) to examine ribosome occupancy in WT and *u5-3 u5-4* mutant. For the 242 candidate transcripts showing reduced polysome association and decreased protein abundance in *u5-3 u5-4*, meta-

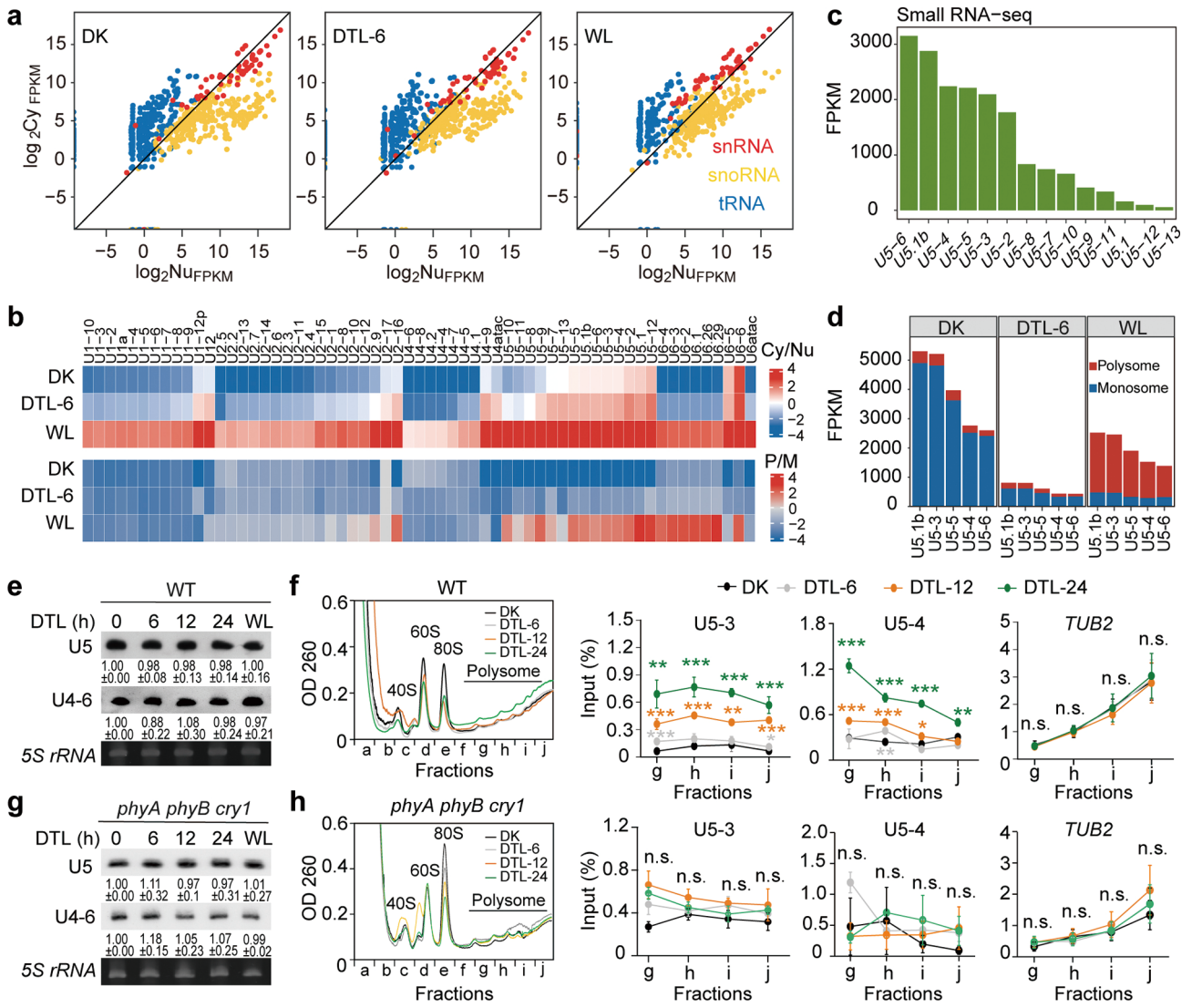



**Fig. 1 Light induces the enrichment of U5 snRNAs in polysomes. a** Scatterplot of relative snRNAs (red), snoRNAs (yellow) and tRNAs (blue) levels in nucleus ($Nu_{FPKM}$) vs ribosome-free cytoplasm ($Cy_{FPKM}$) from dark-grown seedlings (DK), DK seedlings exposed to light for 6 h (DTL-6), and light-grown (WL) seedlings. **b** Heatmaps of $log_2$(Cy/Nu) (top) and $log_2$(P/M; polysome/monosome) (bottom) values for U-type snRNAs under indicated light conditions. **c** Abundance of individual U5 snRNA variants in WL-grown WT seedlings determined by small RNA-seq. **d** Expression levels of the five most abundant U5 snRNA variants in polysomes (red) and monosomes (blue) under indicated conditions. **e**, **g** Northern blots of U5 snRNAs in WT (**e**) and *phyA phyB cry1* (**g**) seedlings under indicated light conditions. **f**, **h** Polysome profiles (left) and polysome association of U5 snRNAs (right), represented by U5-3 and U5-4, as determined by RT-qPCR in dark-grown WT (**f**) and *phyA phyB cry1* (**h**) seedlings following transfer to light for the indicated times. *TUB2* was used as a negative control. Data are shown as mean ± SD. Statistics by two-tailed Student's *t*-test. In **f**, **h**, data at each time point after light exposure were compared with those under DK conditions for the corresponding polysome fraction g–j. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s., not significant.

gene analysis of P-site-corrected ribo-seq reads revealed reduced P-site density at the start codon, with moderate decreases at several nearby positions (Fig. 2m, left). This alteration was more evident in a subset of 16 representative transcripts relevant to the heterotrophic-to-autotrophic transition, which showed reduced start codon P-site density and attenuated downstream P-site peaks (Fig. 2m, right). Together, these results indicate that U5 snRNAs contribute to the efficient translation of nuclear-encoded plastid proteins during de-etiolation.

### U5 snRNA targets mRNAs via the Sm site

Canonically, U5 snRNA facilitates exon ligation via loop I base-pairing with splice sites[17]. Given this sequence recognition capability, we hypothesized that U5 snRNA regulates translation on polysomes through base-pairing with target mRNAs. To test this, we employed the kethoxal assisted RNA–RNA interaction sequencing (KARR-seq)[18], a transcriptome-wide RNA chemical labeling and crosslinking approach, to map U5 snRNA–mRNA interactions in polysome fractions of seedlings grown under DK and DTL-24 conditions. Antisense oligonucleotides (ASOs) targeting U5 snRNA efficiently enriched U5 snRNA-containing chimeric reads, yielding high reproducibility between replicates (r > 0.85) (Supplementary Fig. S3a, b). Motif analysis of mRNA targets with reduced translation in *u5-3 u5-4* double mutants revealed an 8-nt consensus

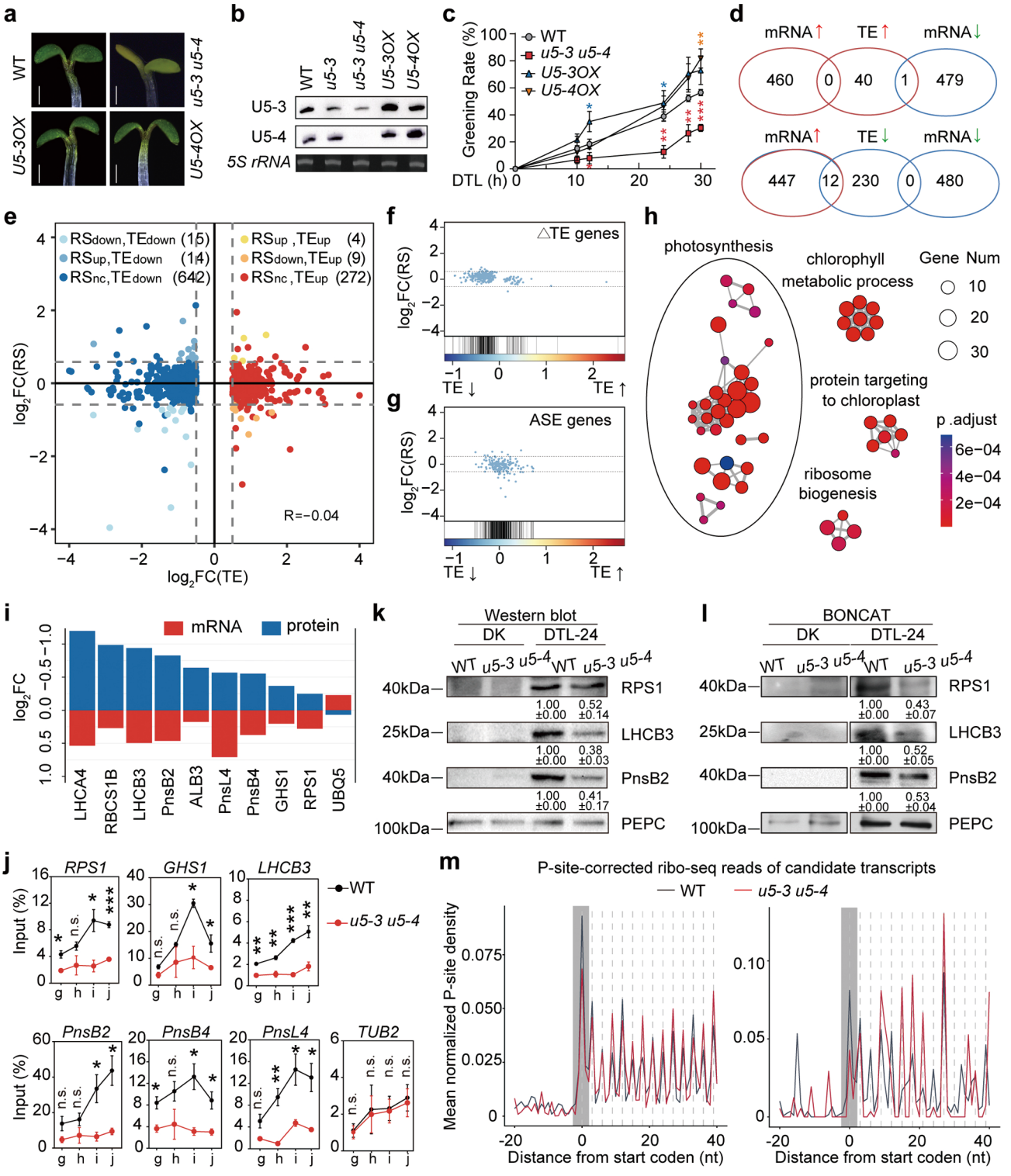



**Fig. 2 U5 snRNAs promote translation of nuclear-encoded plastid mRNAs. a** Representative image of WT, *u5-3 u5-4*, and *u5-3 u5-4* lines overexpressing U5-3 (*U5-3OX*), or U5-4 (*U5-4OX*) seedlings grown under DTL-24 condition. Scale bar = 0.5 mm. **b** Northern blots of U5-3 and U5-4 in the indicated seedlings grown under DTL-24. *5S rRNA* was used as loading control. **c** Greening rates of seedlings for the indicated times. **d** Venn diagrams showing the overlap between genes with altered translation efficiency (TE) and genes with altered mRNA abundance in *u5-3 u5-4* mutants vs WT. In each diagram, the center circle represents genes with altered TE, while the left and right circles represent genes with increased and decreased mRNA abundance, respectively. Red and green arrows indicate increased and decreased changes, respectively. **e** Scatter plot of $log_2$FC (RS) vs $log_2$FC (TE). Numbers indicate gene counts for each group. up, upregulated; down, downregulated; nc, no change. RS: RNA-seq; TE: translational efficiency. **f**, **g** $Log_2$FC (TE) and $log_2$FC (RS) values for genes with altered TE (**f**) and alternative splicing events (ASE) (**g**) in *u5-3 u5-4*. **h** GO terms enriched among genes with reduced translation in *u5-3 u5-4* mutants. Circle color indicates *P* value and circle size reflects gene number. **i** Representative genes exhibiting reduced translation in *u5-3 u5-4* mutants. **j** RT-qPCR showing decreased polysome association of mRNAs in *u5-3 u5-4*. **k**, **l** Western blots showing decreased levels of total (**k**) and newly synthesized (**l**) RPS1, LHCB3 and PnsB2 proteins in *u5-3 u5-4* under DTL-24. BONCAT was used to detect newly synthesized proteins. Data are shown

as mean ± SD. Statistics by two-tailed Student's *t*-test (**c** and **i**). ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$, n.s., not significant. **m** Metagene analysis of normalized ribo-seq P-site density around start codons for the 242 candidate transcripts (left) and a subset of 16 representative chloroplast/photosynthesis-related transcripts relevant to the heterotrophic-to-autotrophic transition (right). P-site occupancy was assigned based on ribosome footprint positions.

sequence 5'-G/UGAGAAGU-3' (Fig. 3a), defined as the U5 snRNA binding element (UBE) (Fig. 3b; Supplementary Fig. S3c), which forms a duplex with the 5'-AUUUUUG-3' core sequence of the U5 snRNA Sm site[3]. This duplex can extend beyond the core sequence, forming stable duplexes longer than 8 bp with free energy ($\Delta G_{duplex}$) lower than –8 kcal/mol (Supplementary Fig. S3d and Table S6). Two representative U5 snRNA targets, *RPS1* and *GHS1* mRNAs, were shown forming a 10–11 bp duplex via their Sm site within the 3′UTR (Fig. 3b), indicating Sm site–UBE complementarity. Importantly, the mRNA-binding motif is distinct from the canonical loop I sequence used by U5 snRNA during splicing, indicating a non-canonical function of U5 snRNA.

We next asked whether this U5 snRNA–mRNA interaction is regulated by light *in vivo*. KARR-seq showed enhanced chimeric interactions between the Sm site of U5 snRNA and the UBE-containing 3′UTRs of *RPS1* and *GHS1* under DTL-24 compared with DK conditions (Fig. 3c, d). Consistent with this light-enhanced interaction, *in situ* RNA immunoelectron microscopy revealed close proximity (around 20 nm) between U5 snRNA and *RPS1* mRNA signals in the cytoplasm of cotyledon cells from seedlings grown under DTL-24 condition, but not in those grown under DK condition (Fig. 3e). Together, these data suggest that light enhances U5 snRNA engagement with target mRNAs on polysomes.

### U5 snRNAs mediate mRNA translation via Sm site–UBE interaction

To assess the functional significance of U5 snRNA–*RPS1* mRNA base-pairing *in vivo*, we generated two *rps1*-mutant plants with altered UBE sequences by CRISPR/Cas9 editing (Fig. 3f). U5 snRNA pull-down failed to enrich *RPS1* mRNAs in either mutant (Fig. 3g), and both mutant plants exhibited reduced RPS1 protein levels without affecting *RPS1* mRNA levels (Fig. 3h, i). Both *rps1* mutants also exhibited reduced chlorophyll accumulation (Fig. 3j). Consistently, U5 snRNA pull-down enriched full-length *RPS1* mRNA but not the UBE-deleted *RPS1* in transgenic lines (Supplementary Fig. S4a). Transgenic lines expressing UBE-deleted *RPS1* also showed reduced RPS1 protein levels and chlorophyll accumulation without changes in *RPS1* mRNA levels (Supplementary Fig. S4b–d). We next asked whether restoring base-pairing between U5 snRNA and the mutated *RPS1* UBE could increase RPS1 protein accumulation in *rps1* mutants. To this end, we engineered U5 snRNAs carrying a 13-bp sequence complementary to the mutated *RPS1* UBE, inserted either downstream of the Sm site or into the 3′ loop. Both variants restored U5 snRNA–mRNA base-pairing and increased RPS1 protein expression in CRISPR/Cas9-generated *rps1* mutant protoplasts, without changes in U5 snRNA abundance (Fig. 3k–m). Together, these results confirm that precise base-pairing between the Sm site of U5 snRNA and *RPS1* mRNA is essential for translational activation of *RPS1*.

A genome-wide search revealed 4,100 mRNAs containing UBE motifs complementary to U5 snRNA (Supplementary Fig. S4e). Among 1,089 proteins upregulated from DTL-12 to DTL-24, 202 genes overlapped with UBE-containing genes complementary to U5 snRNAs, with 107 genes encoding plastid proteins (Supplementary Fig. S4e, f and Table S7). Consistent with the KARR-seq results, *in vitro* RNA pull-down assays showed that deletion of UBEs from the 3′UTRs markedly reduced U5 snRNA binding to *RPS1* and five additional targets (Supplementary Figs. S4g–i). Endogenous U5 snRNA pull-downs further confirmed light-dependent enrichment of these six target mRNAs in WT seedlings (Supplementary Fig. S4j). Together, these results demonstrate U5 snRNA–mRNA interactions, which are enhanced by light *in vivo*.

We next asked whether these interactions enhance mRNA translation using dual-LUC reporter assays. For the six representative U5 target mRNAs, mutation of either the UBE or the Sm site of U5 snRNA significantly reduced LUC reporter translation without changing mRNA levels, while reciprocal swapping of the Sm site and UBE sequences preserved their complementarity but failed to restore LUC reporter translation (Fig. 3n; Supplementary Fig. S5a–c). Consistently, deletion of UBEs reduced LUC reporter translation without affecting mRNA levels (Supplementary Fig. S5c–e). Collectively, these findings reveal that U5 snRNAs enhance translation of a substantial subset of mRNAs encoding nuclear-encoded plastid proteins through Sm site–UBE base-pairing, with translational activation requiring the native sequence configuration rather than base-pairing alone.

### Polysome association of Sm protein homologs

We next assessed whether spliceosomal U5 snRNP components are present on polysomes. U5 small nuclear ribonucleoprotein 114 kDa (SNU114) and pre-mRNA processing factor 8 (PRP8)[19-21] were not detected in polysome-associated fractions (Supplementary Fig. S6a–c). In contrast, all seven Sm proteins co-immunoprecipitated with His-FLAG-tagged RPL18 from transgenic seedlings grown under DTL-24 condition (Supplementary Fig. S6d). Their polysome association was further confirmed by immunoblotting using antibodies against representative Sm proteins, including SmB, SmG and SmF (Supplementary Fig. S6e). To characterize polysome-associated Sm proteins, we selected His-FLAG-tagged SmD3a and HA-tagged SmFa as representative components. We performed co-immunoprecipitation followed by mass spectrometry using polysome fractions from the corresponding transgenic seedlings grown under DTL-24 condition. These analyses identified at least one representative homolog from each of the seven Sm protein families, including SmD2a, SmBb, and SmGa (Fig. 4a; Supplementary Fig. S6f and Table S8). These results suggest that Sm proteins are present in polysome-associated fractions.

### Sm proteins are essential for U5 snRNA-mediated translational activation

To examine whether Sm proteins associate with U5 snRNAs on polysomes, we performed RNA immunoprecipitation (RIP)

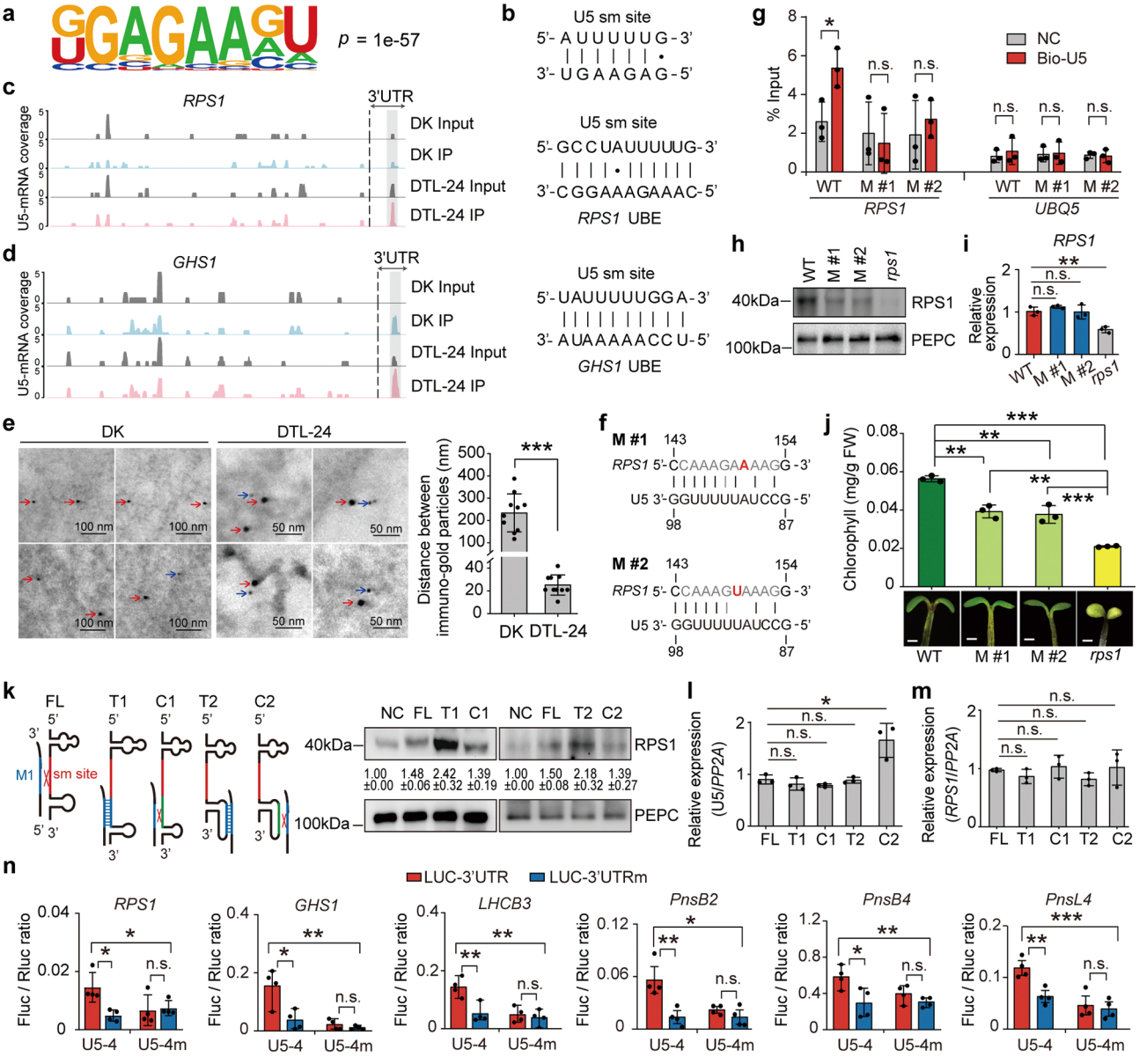



**Fig. 3 U5 snRNA targets mRNAs through Sm site and facilitates their translation.** **a** Consensus motif of U5 snRNA target mRNAs. **b** The core U5 snRNA::mRNA duplex structure. **c**, **d** U5 snRNA binding at *RPS1* (**c**) and *GHS1* (**d**) 3′UTRs, with corresponding Sm site::UBE structures. Gray shade: UBE region. Input/IP: KARR-seq without/with U5 snRNA ASO. **e** TEM images showing U5 snRNA co-localizes with *RPS1* mRNA in WT cotyledons at DTL-24. Red: *RPS1*; Blue: U5 snRNA. Mean inter-particle distances were shown ($n$ = 10). **f** CRISPR/Cas9-generated UBE mutations in *RPS1* 3'UTR (M#1, M#2 in red). **g** U5 snRNA fails to pull down *RPS1* mRNA in M#1/M#2 mutants ($n$ = 3). **h**, **i** Protein (**h**) and mRNA (**i**) levels of RPS1 in mutants at DTL-24. **j** Chlorophyll levels and phenotypes of WT, M#1 and M#2 plants. Scale bar = 0.5 mm. **k** Modified U5 snRNAs (T1, T2) enhance RPS1 protein in M#1 mutant. T1/T2 contain 13-bp insertions base-pairing with the edited *RPS1* UBE sequence in M#1 plants downstream of the Sm site (T1) or in the 3′ loop (T2); C1/C2: reverse-sequence controls. Protein levels were quantified (mean ± SD, $n$ = 3). **l**, **m** Expression of WT or modified U5 snRNAs (**l**) and *RPS1* (**m**) mRNA. **n** Dual-luciferase assays with WT or UBE-mutant 3'UTR reporters and WT and mutant U5-4 snRNA effectors. Each dot: one biological replicate. Two-tailed Student's $t$-test in **e**, **g**, **i**, **j**, **l**–**n**. $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s., not significant. Data are shown as mean ± SD.

using anti-SmB and anti-SmD3 antibodies. RIP assays showed that both SmB and SmD3 co-immunoprecipitated with U5 snRNA in the polysome fractions of WT seedlings at DTL-24 (Fig. 4b, c). However, in the *smd2a* and *smd3a* mutants, U5 snRNA was nearly undetectable in the immunoprecipitates of both SmB and SmD3 (Fig. 4b, c). These findings suggest that U5 snRNA interacts with Sm proteins on polysomes, and that this association depends on the presence of individual Sm proteins.

We therefore used *smd2a* and *smd3a* mutants to test whether polysome-associated Sm proteins are required for U5 snRNA-mediated translational activation. LUC reporters carrying WT or UBE-deleted 3′UTRs of *RPS1*, *PnsB2*, and *LHCB3* were transiently expressed in protoplasts from WT, *smd3a*, and *smd2a* mutants. We found that U5 snRNA-promoted translation of WT reporters was compromised in the *smd3a* and *smd2a* mutants, similar to UBE deletion reporters (Fig. 4d; Supplementary Fig. S7a). We next examined the polysome association, mRNA and protein levels of *RPS1*, *PnsB2*, and *LHCB3* in WT, *smd2a*, and *smd3a* seedlings.

**a**

| Sm | Proteins | PSMs | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | NC-1 | IP-1 | Total-1 | NC-2 | IP-2 | Total-2 | NC-3 | IP-3 | Total-3 |
| SmD3 | SmD3a | 0 | 73 | 100 | 2 | 55 | 79 | 3 | 73 | 97 |
| | SmD3b | 1 | 11 | | 4 | 9 | | 5 | 11 | |
| | Common | 1 | 16 | | 2 | 15 | | 2 | 13 | |
| SmB | SmBa | 0 | 0 | 105 | 0 | 0 | 104 | 0 | 0 | 93 |
| | SmBb | 8 | 105 | | 11 | 104 | | 10 | 93 | |
| | Common | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| SmD1 | SmD1a | 0 | 14 | 56 | 0 | 14 | 60 | 1 | 14 | 53 |
| | SmD1b | 4 | 31 | | 5 | 33 | | 4 | 30 | |
| | Common | 1 | 11 | | 2 | 13 | | 1 | 9 | |
| SmE | SmEa | 0 | 11 | 50 | 0 | 13 | 51 | 1 | 8 | 40 |
| | SmEb | 1 | 19 | | 0 | 18 | | 0 | 14 | |
| | Common | 0 | 20 | | 0 | 20 | | 0 | 18 | |
| SmD2 | SmD2a | 7 | 34 | 34 | 9 | 35 | 35 | 3 | 29 | 29 |
| | SmD2b | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| | Common | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| SmG | SmGa | 4 | 30 | 30 | 6 | 32 | 32 | 4 | 24 | 24 |
| | SmGb | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| | Common | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| SmF | SmFa | 0 | 0 | 12 | 0 | 0 | 10 | 0 | 0 | 4 |
| | SmFb | 0 | 0 | | 0 | 0 | | 0 | 0 | |
| | Common | 1 | 12 | | 0 | 10 | | 0 | 4 | |

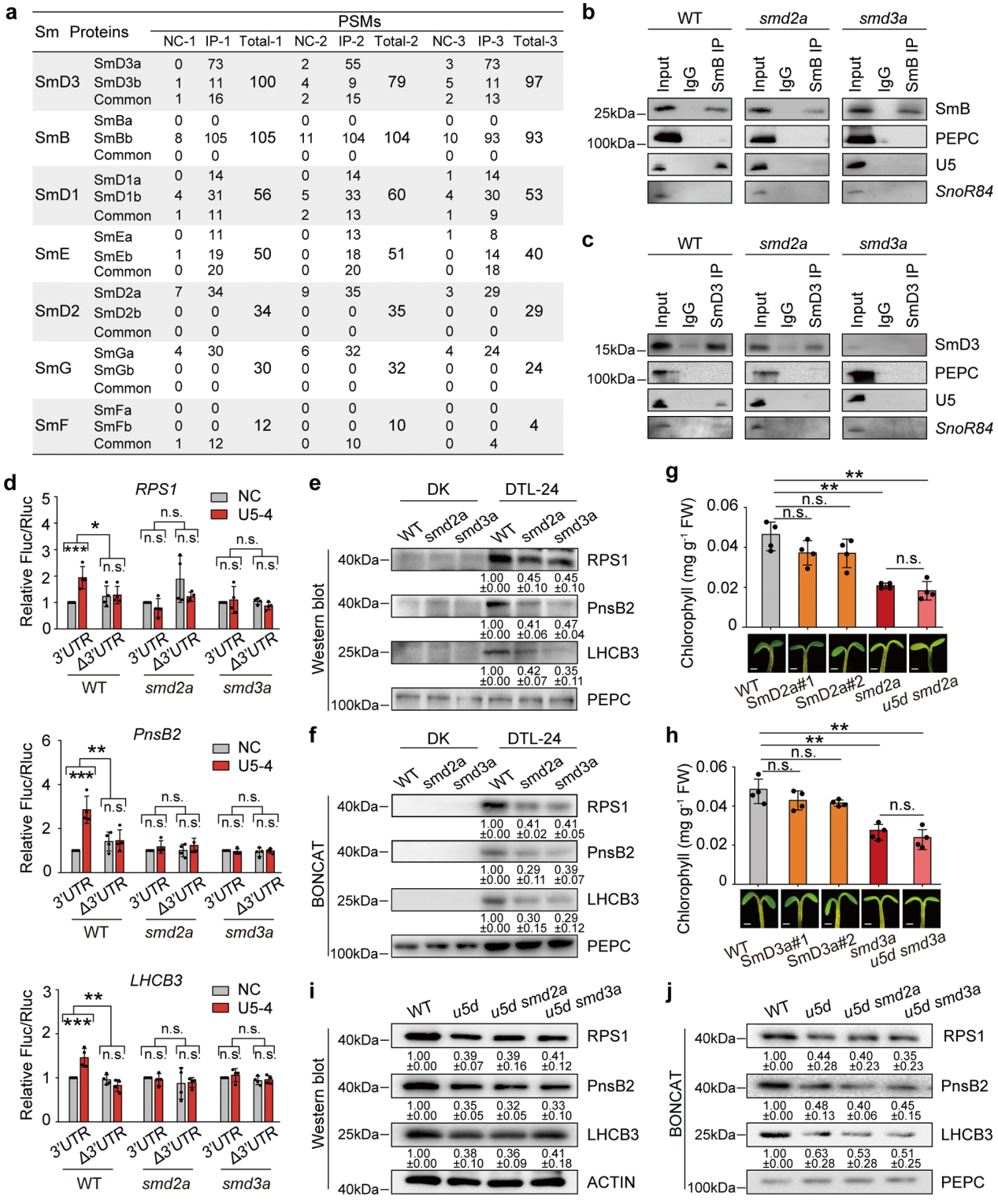



**Fig. 4 Polysome-associated Sm proteins are required for U5 snRNA-promoted mRNA translation.** **a** Sm proteins detected in polysome fractions of *Pro35S:SmD3a-HIS-FLAG*/WT transgenic seedlings grown under DTL-24 condition by IP-MS. NC: WT plant control. Common: peptides that do not distinguish between the two homologous Sm proteins. **b**, **c** U5 snRNA is nearly undetectable in SmB (**b**) and SmD3 (**c**) immunoprecipitates in *smd2a* and *smd3a* mutants. **d** Dual-luciferase assays showing the reduced U5 snRNA-mediated translation of *RPS1* (upper), *PnsB2* (middle), and *LHCB3* (lower) reporters in WT, *smd2a* and *smd3a* protoplasts. **e**, **f** Western blots (**e**) and BONCAT assays (**f**) showing reduced total and newly synthesized protein levels of RPS1, PnsB2, and LHCB3 in WT, *smd2a* and *smd3a*. **g** Chlorophyll contents (upper) and phenotypes (lower) of WT, *smd2a*, *u5-3 u5-4 smd2a*, and two *Pro35S:SmD2a-3HA/smd2a* transgenic lines (SmD2a#1 and SmD2a#2). Scale bar = 0.5 mm. **h** Chlorophyll contents (upper) and phenotypes (lower) of WT, *smd3a*, *u5-3 u5-4 smd3a*, and two *Pro35S:SmD3a-HIS-FLAG/smd3a* transgenic seedlings (SmD3a#1 and SmD3a#2). Scale bar = 0.5 mm. **i**, **j** Western blots showing the protein levels (**i**) and BONCAT assays showing newly synthesized protein levels (**j**) of three representative proteins in WT, *u5-3 u5-4* (*u5d*), *u5-3 u5-4 smd2a* (*u5d smd2a*) and *u5-3 u5-4 smd3a* (*u5d smd3a*) under DTL-24 condition. Data are shown as mean ± SD. Data were analyzed by two-way ANOVA (**d**), and two-tailed Student's *t*-test (**g, h**), $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s., not significant.

Compared with WT, both mutants showed reduced polysome association of all three representative mRNAs (Supplementary Fig. S7b), whereas total mRNA levels were unchanged (Supplementary Fig. S7c). Consistently, the protein abundance and synthesis rates of RPS1, PnsB2, and LHCB3 were markedly reduced (Fig. 4e, f). Together, these data suggest that Sm proteins, as shown by SmD3a and SmD2a, are required for U5 snRNA-mediated activation of translation.

To assess the impact of impaired translation of these plastid proteins in *smd* mutants, we examined the cotyledon greening phenotype. Loss of either *SmD2a* or *SmD3a* reduced chlorophyll accumulation and impaired cotyledon greening (Fig. 4g, h). Moreover, the *u5-3 u5-4 smd2a* and *u5-3 u5-4 smd3a* triple mutants exhibited reduced chlorophyll levels (Fig. 4g, h), decreased protein abundance and synthesis rates of RPS1, PnsB2 and LHCB3 (Fig. 4i, j), comparable to those observed in the *u5-3 u5-4* double mutant. These results suggest that U5 snRNA and SmD2a/SmD3a act together to mediate light-dependent cotyledon greening.

### SmD3a interacts with EF2a on polysomes

To explore the molecular mechanism by which U5 snRNA/Sm proteins promote mRNA translation, we reanalyzed the SmD3a interactome using Co-IP and mass spectrometry data. This reanalysis identified several high-confidence SmD3a interactors related to translation, including four translation initiation factors and five translation elongation factors (Supplementary Table S8). Among these, eukaryotic translation elongation factor 2a (EF2a)[22,23], the *Arabidopsis* homolog of the eukaryotic EF2 (eEF2) that is essential for protein synthesis, was identified as a direct interactor of SmD3a. This interaction was confirmed by Co-IP using polysome fractions from His-FLAG-tagged SmD3a transgenic seedlings grown under DK and DTL-24 conditions, with stronger polysome-associated interactions observed under DTL-24 condition, and by *in vitro* pull-down assays (Fig. 5a, b).

### Mammalian SmD3 associates with eEF2 in the cytoplasm

Because Sm proteins and eEF2 are conserved across eukaryotes, we performed protein–protein docking, which predicted a similar interaction interface between SmD3 and eEF2 in *Arabidopsis* and human (Fig. 5c). We therefore asked whether SmD3–eEF2 interaction also occurs in mammalian cells. Co-expression of human SmD3-HA and EF2-FLAG in HEK293T and HeLa cells, followed by cytoplasmic Co-IP, confirmed their association (Fig. 5d, e). These findings indicate that SmD3–eEF2 interaction may be conserved across eukaryotes.

### EF2a is essential for U5 snRNA-mediated translation

To examine the role of EF2a in U5 snRNA-mediated activation of translation, we transiently expressed LUC reporters with either WT or UBE-deleted 3′UTRs of *RPS1*, *PnsB2*, and *LHCB3* in protoplasts from WT and *ef2a* mutant seedlings (Supplementary Fig. S8a–c). LUC reporter assays revealed that U5 snRNA expression enhanced LUC reporter translation in WT but not in *ef2a* mutants (Fig. 5f; Supplementary Fig. S8d–f). Using MS2 tethering assays, we found that recruiting EF2a to the 3′UTR of a LUC reporter mRNA enhanced its translation (Supplementary Fig. S8g–i). Accordingly, *ef2a* mutants grown under DTL-24 showed defective polysome association and reduced protein synthesis of U5 snRNA targets (Fig. 5g, h; Supplementary Fig. S8j–o), indicating that EF2a is essential for U5 snRNA-promoted mRNA translation.

### Blocking EF2a–SmD3a interaction inhibits U5 snRNA-mediated translation

To determine whether the EF2a–SmD3a interaction is necessary for U5 snRNA-promoted mRNA translation, we identified key amino acid residues on EF2a mediating the binding between EF2a and SmD3a. Protein–protein docking and mutational analysis revealed that H754 and E803 on EF2a are critical for the interaction with SmD3a. *In vitro* pull-down and luciferase complementation imaging (LCI) assays showed that substitutions of both H754 and E803 with alanines in EF2a significantly reduced its interaction with SmD3a (Fig. 5b; Supplementary Fig. S9b). These results were further confirmed by Co-IP in transgenic plants expressing EF2a or EF2a$^{H754A/E803A}$ (Fig. 5i).

The identification of H754 and E803 as key residues for EF2a–SmD3a interaction allowed us to test the functional importance of this interaction in U5 snRNA target mRNA translation. We co-expressed LUC reporters carrying WT or UBE-deleted 3′UTRs of *RPS1*, *PnsB2*, and *LHCB3* with WT EF2a or EF2a$^{H754A/E803A}$ as effectors in *N. benthamiana* leaves. The analysis revealed that expression of EF2a, but not EF2a$^{H754A/E803A}$, enhanced translation of WT reporters. However, neither of them enhanced translation of the UBE-deleted reporters (Fig. 5j; Supplementary Fig. S9c–e). In contrast, translation of the *UBQ5* control reporter was unaffected by EF2a expression (Fig. 5j; Supplementary Fig. S9f). Additionally, transient expression of SNU114, the homolog of EF2a and a U5 snRNP component of the spliceosome[19,24], did not enhance LUC reporter's translation (Fig. 5j; Supplementary Fig. S9c–e). Together, these data suggest that EF2a promotes U5 snRNA-mediated target mRNA translation through its interaction with SmD3a.

Next, we examined whether the reduced translation of U5 snRNA target mRNAs in *ef2a* mutants affects cotyledon greening. Knockdown of *EF2a* led to delayed cotyledon greening and chlorophyll accumulation (Fig. 5k). Expression of EF2a in *ef2a*, but not EF2a with H754A/E803A mutations, increased cotyledon chlorophyll levels and the expression levels of RPS1, PnsB2 and LHCB3 proteins at DTL-24 (Fig. 5k, l). These results further demonstrate that the EF2a–SmD3a interaction is required for U5 snRNA-mediated translational activation during seedling greening.

### A designed SmD3 variant for translation enhancement

Having identified the translational regulatory function of the U5 snRNA/Sm module, we next explored whether it could be engineered to increase the production of nuclear-encoded plastid proteins in other plant species. We focused on SmD3 based on our findings that SmD3a directly interacts with translation elongation factor eEF2 (Figs. 4, 5), which suggests that it functions as the essential molecular link in U5 snRNA-mediated translation. In *Arabidopsis*, SmD3a was detected in both polysome-associated fractions and nuclear fractions (Supplementary Fig. S10a, b). Loss of *SmD3a* impaired seedling greening and thermotolerance (Supplementary Fig. S10c–h). Together with our translational analyses, these

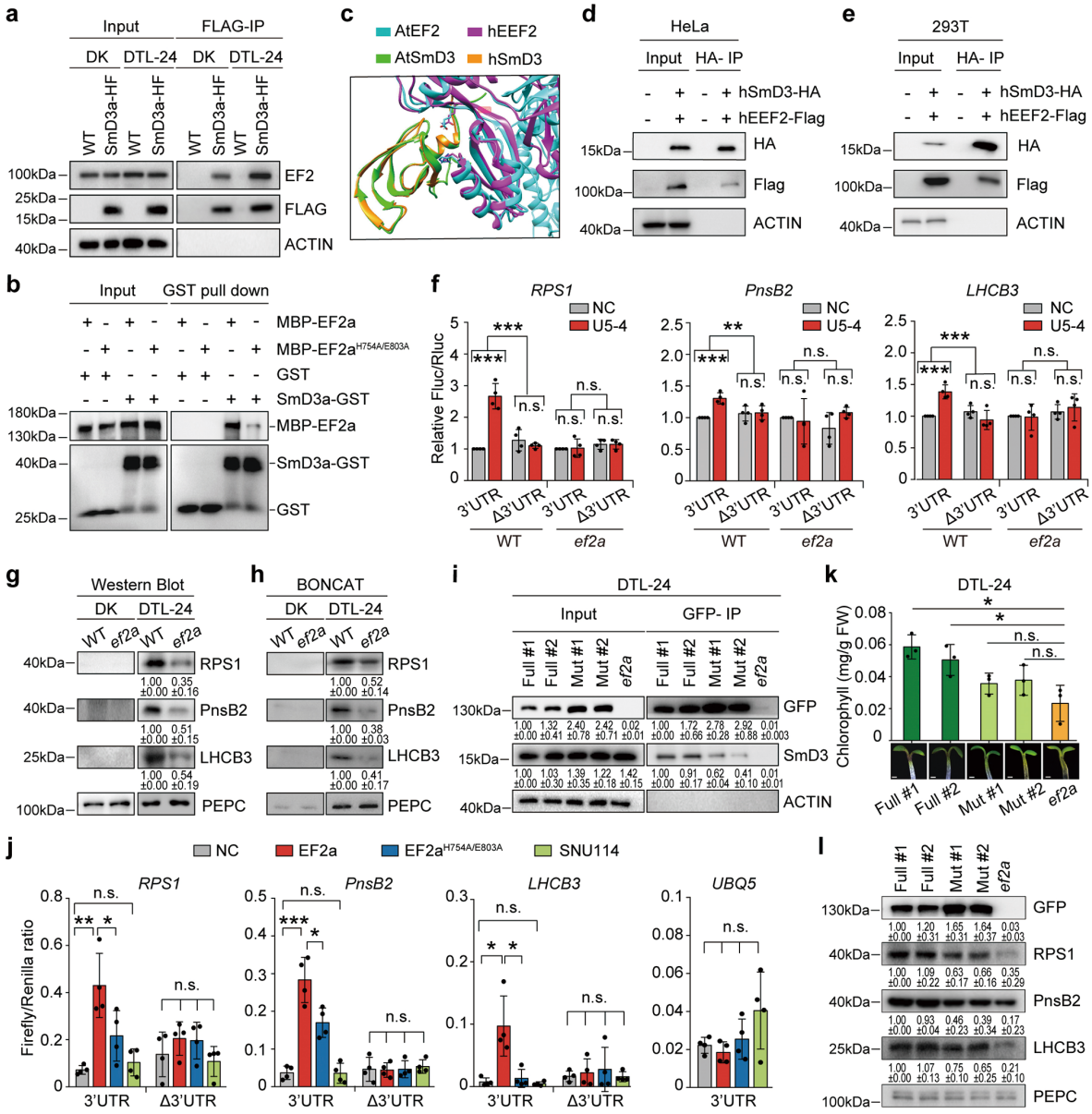



**Fig. 5 U5 snRNA/Sm proteins facilitate target mRNAs translation via EF2a interaction. a** Co-IP showing that EF2a associates with SmD3a in seedlings grown under DK and DTL-24 conditions. **b** Pull-down assays showing that EF2a residues H754/E803 are critical for SmD3a binding. **c** Predicted EF2–SmD3 interaction interfaces in *Arabidopsis* (cyan/green) and human (purple/orange). **d**, **e** Co-IP of hSmD3-HA with FLAG-hEEF2 in HeLa (**d**) and 293T (**e**) cells. **f** Dual-luciferase assays showing effects of EF2a knockdown on U5 snRNA-mediated activation of *RPS1*, *PnsB2*, and *LHCB3* reporters in WT and *ef2a* mutant protoplasts. **g**, **h** Western blot (**g**) and BONCAT (**h**) showing total and newly synthesized levels of RPS1, PnsB2, and LHCB3 in WT and *ef2a* under DTL-24 condition. PEPC: negative control. **i** Co-IP showing weakened EF2a–SmD3a interaction by H754A/E803A mutations in *Pro35S:EF2a-GFP/ef2a* (Full) and *Pro35S:EF2a*$^{H754A/E803A}$*-GFP/ef2a* (Mut) transgenic seedlings. ACTIN: loading control. **j** Effects of EF2a, EF2a$^{H754A/E803A}$, and SNU114 on *RPS1*, *PnsB2*, and *LHCB3* reporters with/without UBE. *UBQ5*: negative control. **k** Chlorophyll content (top) and phenotypes (bottom) of *ef2a* and transgenic seedlings under DTL-24 condition. Scale bar = 0.5 mm. **l** Western blot of RPS1, PnsB2, and LHCB3 in transgenic seedlings under DTL-24. Full#1 and Full#2 represent two independent *Pro35S:EF2a-GFP/ef2a* transgenic lines, Mut#1 and Mut#2 represent two independent *Pro35S:EF2a*$^{H754A/E803A}$*-GFP/ef2a* transgenic lines. *ef2a* seedlings were used as a control. PEPC: negative control. Data are shown as mean ± SD. Data were analyzed by two-way ANOVA (**f**), and two-tailed Student's *t*-test (**j**, **k**), $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.05$, n.s., not significant.

results support a cytoplasmic role of SmD3a in translational regulation, without excluding a possible role in nuclear splicing. By contrast, overexpression of *SmD3b* failed to rescue the greening and thermotolerance defects of *smd3a* mutants (Supplementary Fig. S10c–h), consistent with its reported role in pre-mRNA splicing. Phylogenetic and comparative genomic analyses suggested that SmD3a and SmD3b arose from the crucifer-specific At-α whole genome duplication, with SmD3a undergoing positive selection in some Brassicaceae lineages (Fig. 6a; Supplementary Figs. S11, S12). Together, these evolutionary and functional analyses suggest that SmD3a and SmD3b have undergone functional divergence, with SmD3a playing a distinct cytoplasmic role in translational regulation in *Arabidopsis*. We therefore hypothesized that designing an SmD3 variant with enhanced cytoplasmic localization could enhance U5 snRNA-mediated translation in other plant

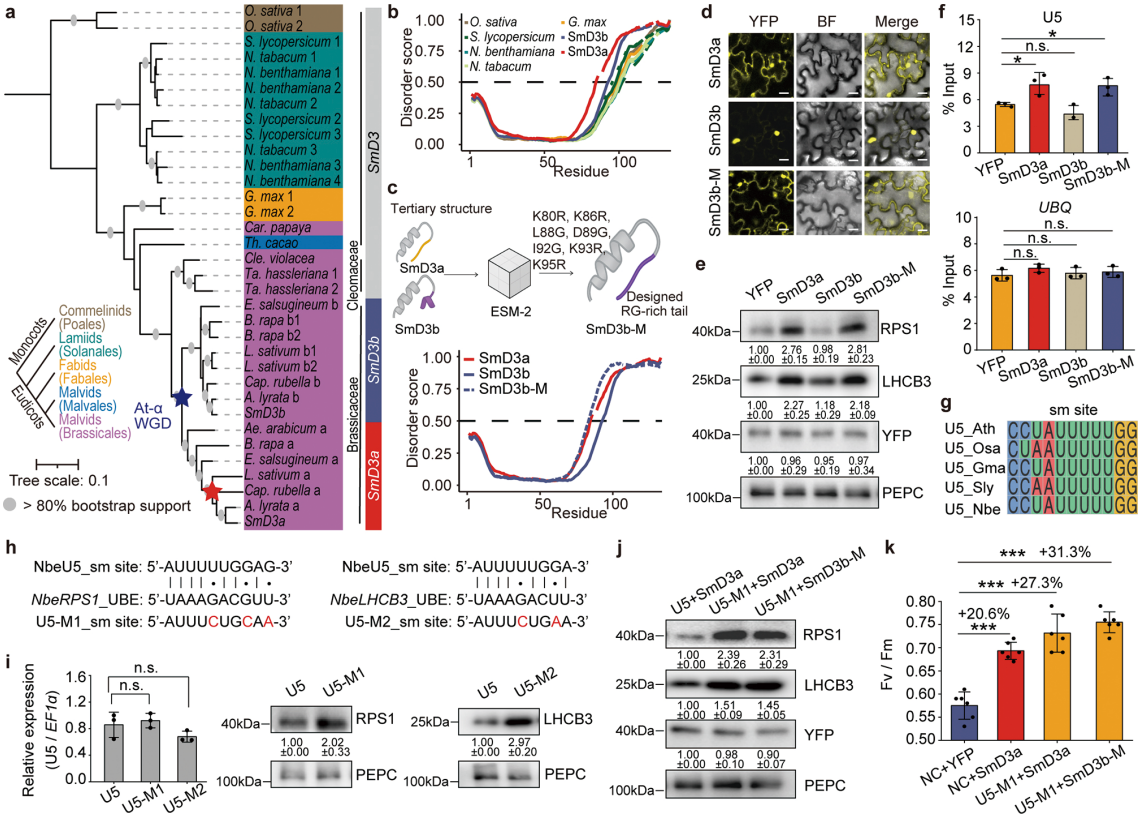


**Fig. 6 Engineered U5 snRNA/Sm protein module enhances PSII photochemical efficiency in *N. benthamiana*.** **a** Phylogeny of SmD3 coding sequences in angiosperms showing Brassicaceae-specific At-α WGD (blue star) gave rise to SmD3a and SmD3b clades. SmD3a underwent positive selection in lineage I (red star; PAML, $P < 0.01$). The left diagram: the phylogenetic relationships among the major angiosperm lineages. **b** Predicted disorder scores of SmD3 residues in *Arabidopsis* and major crops. **c** The schematic diagram shows artificially designed SmD3b variant (SmD3b-M) based on the protein language model ESM-2. **d** Confocal images showing the subcellular localization of transiently expressed YFP-tagged SmD3a, SmD3b and SmD3b-M in *N. benthamiana* leaves. Scale bar = 20 μm. **e** Western blots of RPS1 and LHCB3 in *N. benthamiana* leaves expressing YFP, SmD3a, SmD3b, or SmD3b-M. **f** RT-qPCR showing U5 levels in polysomal fractions from corresponding samples. *UBQ5*: control. **g** Conservation of U5 snRNA Sm site across five plant species. **h** Schematic of U5 snRNA variants (U5-M1 and U5-M2) designed to base-pair with *NbeRPS1* (left) and *NbeLHCB3* (right). **i** U5 snRNA levels (left), RPS1 (middle) and LHCB3 (right) protein levels in tobacco leaves expressing U5, U5-M1, or U5-M2. **j** Western blot showing increased RPS1 and LHCB3 levels in tobacco leaves co-expressing U5-M1 with SmD3a or SmD3b-M, relative to co-expression of U5 snRNA with SmD3a. **k** Fv/Fm values in tobacco leaves co-expressing U5-M1 with SmD3a or SmD3b-M, or empty vector (NC) with YFP or SmD3a. Data are shown as mean ± SD. Each dot: one biological replicate. Two-tailed Student's *t*-test in **f**, **i**, **k**. $^*P < 0.05$; n.s., not significant. Protein levels (**e**, **i**, **j**): $n = 3$; Fv/Fm (**k**): $n = 6$.

species, thereby increasing the accumulation of nuclear-encoded plastid proteins.

Comparison of predicted SmD3 structures revealed that the C-terminal regions of SmD3 homologs from major crops, including *Oryza sativa*, *Glycine max*, *Solanum lycopersicum*, and *Nicotiana benthamiana*, resemble that of *Arabidopsis* SmD3b, while *Arabidopsis* SmD3a possesses a more extensive loop/coil-rich C-terminal region (Fig. 6b). Evolutionary analysis identified several SmD3a-specific substitutions and deletions that reduce local electrostatic potential and likely enhance C-terminal flexibility (Supplementary Fig. S11). To test whether SmD3b-like proteins could be engineered for cytoplasmic translational regulation, we applied an artificial intelligence-assisted rational design strategy to modify *Arabidopsis* SmD3b (Fig. 6c, upper, also see Materials and Methods). We introduced seven substitutions designed to increase RG-richness and disrupt helix-stabilizing interactions. Structural prediction suggested that the redesigned protein (SmD3b-M) possesses a more flexible, SmD3a-like C-terminal region (Fig. 6c, lower).

## Engineered U5 snRNA/Sm module enhances chloroplast protein accumulation

We next compared the cytoplasmic localization and translational activity of SmD3b-M with those of SmD3b and SmD3a in *N. benthamiana*. Transient expression of codon-optimized *Arabidopsis* SmD3a, SmD3b, and SmD3b-M in *N. benthamiana* leaves revealed similar cytoplasmic localization of SmD3a and SmD3b-M, while SmD3b was mainly localized in the nucleus (Fig. 6d). Expression of either SmD3a or SmD3b-M increased endogenous RPS1 and LHCB3 protein levels in *N. benthamiana*, whereas expression of SmD3b had no significant effect compared with the YFP controls (Fig. 6e). Consistently, expression of either SmD3a or SmD3b-M increased polysome association of U5 snRNAs (Fig. 6f), supporting their ability to promote U5 snRNA-mediated translation in *N. benthamiana*.

Given the suboptimal yet conserved base-pairing between U5 snRNA and its targets, e.g., *RPS1* and *LHCB3*, in *N. benthamiana* (Fig. 6g, h), we hypothesized that designing U5 snRNA variants for perfect complementarity would enhance

protein production. Indeed, expression of U5 snRNA mutants (U5-M1 and U5-M2), designed to perfectly match the UBE of *RPS1* and *LHCB3* respectively, enhanced endogenous protein levels in tobacco, compared to expression of the WT U5 snRNA control at similar RNA levels (Fig. 6i).

Building on these results, co-expression of U5-M1 with codon-optimized *Arabidopsis* SmD3a or SmD3b-M further increased RPS1 and LHCB3 protein levels compared with co-expression of SmD3a and WT U5 snRNA (Fig. 6j). In both cases, the maximum photochemical efficiency of PS II (Fv/Fm) was also higher than that observed with SmD3a alone and exceeded the YFP control by over 20% (Fig. 6k). Together, these findings demonstrate the potential of engineering U5 snRNA and SmD3 to enhance chloroplast protein accumulation and photosynthetic performance in non-Brassicaceae plant species.

## DISCUSSION

Our findings uncover a previously unrecognized role of U5 snRNA in mediating selective translational upregulation during the transition from heterotrophic to autotrophic growth in *Arabidopsis*. In darkness, the abundance of nuclear-encoded plastid proteins largely tracks their transcript levels. Upon light illumination, U5 snRNA associates with polysomes and is directly linked to the enhanced translation of these proteins, an effect that exceeds what would be expected from the changes in their mRNA levels (Supplementary Fig. S13). This light-responsive engagement suggests that U5 snRNA can be redeployed outside the nucleus to support the rapid proteome remodeling required for seedling greening. Although previous studies have described light-dependent translational regulation of individual mRNAs[25-28], our results reveal that spliceosomal RNAs can associate with the translational apparatus to coordinate selective changes in protein expression during developmental transitions.

Our results show that light promotes U5 snRNA association with polysomes in a manner that requires phyA, phyB, and CRY1 (Fig. 1). We further show that interactions between the Sm site of U5 snRNA and target 3′UTR UBEs are enhanced under DTL-24 condition (Fig. 3), suggesting that light-promoted U5 snRNA–target mRNA interaction may facilitate U5 snRNA association with polysomes. Our data also indicate that U5 snRNA associates with Sm proteins on polysomes and that Sm proteins are required for U5 snRNA-mediated translational regulation (Fig. 4). These findings raise the possibility that Sm protein engagement influences U5 snRNA subcellular distribution and polysome-associated function, potentially affecting whether U5 snRNAs enter the canonical nuclear spliceosomal pathway or participate in cytoplasmic translational regulation through UBE-dependent 3′UTR recognition. Defining the photoreceptor-mediated regulation of U5 snRNA association with polysomes will require further study. Unlike Argonaute- or PIWI-associated small RNA pathways[29-33], this mechanism uses a spliceosomal RNA-associated Sm module to engage the translational machinery. This finding expands the known regulatory capacity of spliceosomal RNAs and points to a mechanism by which light signals can modulate selective protein synthesis.

The mechanistic basis of this process involves the heptameric Sm complex, in which the Brassicaceae-derived paralog SmD3a acts as a representative Sm component that engages eEF2 and links U5 snRNA-mediated target recognition to translational control on polysomes. Although our experiments were performed in *Arabidopsis*, the presence of cytoplasmic snRNAs and Sm proteins in other eukaryotes[7,8,10], together with the association between human SmD3 and eEF2 detected by Co-IP (Fig. 5), raises the possibility that spliceosomal components may associate with translational regulation more broadly. Whether similar interactions have regulatory functions beyond plants will require further investigation. Because U5 snRNA recognizes target mRNAs through 3′UTR elements, the SmD3a–EF2a interaction may provide a way for a 3′UTR-associated U5 snRNA/Sm protein module to communicate with the translation machinery on selected transcripts. However, the precise translation step affected by SmD3a–EF2a interaction remains unresolved. Future studies that monitor ribosome dynamics will be needed to determine which step of the translation cycle is regulated by this pathway.

The emergence of the eukaryotic nucleus established a functional separation between nuclear pre-mRNA splicing and cytoplasmic translation, enabling increased fidelity and diversity of the proteome[34]. Eukaryotic Sm proteins, which evolved from Hfq-like chaperones, subsequently diversified to support spliceosomal function in the nucleus[35,36]. In *Arabidopsis*, the Brassicaceae-derived paralog SmD3a, estimated to have diverged approximately 47 million years ago[37], engages eEF2, a translation elongation GTPase homologous to SNU114, the GTPase component of the spliceosomal U5 snRNP, to support U5 snRNA-directed translational regulation. U5 snRNA-directed enhancement of nuclear-encoded plastid protein synthesis includes regulators of chloroplast biogenesis such as the translation initiation factor RPS1[38-40]. Consistent with this role, *u5-3 u5-4* and *smd3a* mutants exhibit defects in chloroplast biogenesis and increased sensitivity under heat stress (Supplementary Fig. S10), indicating that this regulatory mechanism supports plastid function under physiologically demanding conditions. Given that chloroplast development engages extensive retrograde signaling, some translational changes may coincide with broader organellar signaling dynamics. However, the sequence-specific association of U5 snRNA with defined mRNAs on polysomes, together with its dependence on Sm-associated factors and eEF2 engagement, supports a direct regulatory contribution rather than a secondary consequence of retrograde signaling.

Together, our findings reveal an additional cytoplasmic function of the U5 snRNA/Sm protein module in selective mRNA translation. The evolutionary divergence of SmD3a and SmD3b in Brassicaceae, together with the successful engineering of U5 snRNA and SmD3b to enhance chloroplast protein accumulation and PSII photochemical efficiency in tobacco, illustrates how conserved spliceosomal components can be harnessed for cytoplasmic translational regulation.

## MATERIALS AND METHODS

### Plant materials and growth conditions

The ecotype of *Arabidopsis thaliana* used in this study was Columbia-0 (Col-0). WT *Nicotiana benthamiana* plants were grown in soil at 25 °C in a 16/8 h light/dark photoperiod for 4–5 weeks before being used for transient assays. Previously

published lines are *rps1*[38], *phyA phyB cry1*[41], *Pro35S: HIS-FLAG-RPL18*/WT[42]. The T-DNA insertion mutants *smd2a* (SALK_210028C), *smd3a* (SALK_025193), *smd3b* (SALK_006410), *ef2a* (SALK_004229C) were obtained from AraShare (arashare.cn/); *u5-3* (SALK_206352C) was obtained from the European Arabidopsis Stock Centre (NASC); *u5-4* mutant was generated by CRISPR/Cas9 system. The *ProUBQ: U5-3* (*U5-3OX*) and *ProUBQ: U5-4* (*U5-4OX*) transgenic lines were generated in the *u5-3 u5-4* double mutant background; the *Pro35S: SmD3a-HIS-FLAG*, *Pro35S: SmD3b-HIS-FLAG* and *Pro35S: SmFa-3HA* lines in the WT background; the *ProRPS1: RPS1-3′UTR*$_{RPS1}$ and *ProRPS1: RPS1-Δ3′UTR*$_{RPS1}$ lines in *rps1* mutant background; the *Pro35S: SmD3a-HIS-FLAG*, *ProSmD3a: SmD3a-YFP* and *ProSmD3a: SmD3b-YFP* lines in *smd3a*; the *Pro35S: SmD2a-3HA* line in *smd2a*; the *Pro35S: EF2a-GFP and Pro35S: EF2a*$^{H754A/E803A}$*-GFP* lines in *ef2a*. The mutants *u5-3 u5-4*, *u5-3 u5-4 smd2a*, *u5-3 u5-4 smd3a* were generated by genetic crossing. The specific PCR primers used for genotyping are detailed in Supplementary Table S9.

Seeds were surface-sterilized with 15% (w/v) bleach, then stratified at 4 °C in darkness for 2 days. The stratified seeds were placed on Murashige and Skoog (MS) medium, supplemented with 1% sucrose and 2.5% phytagel (Sigma-Aldrich), and adjusted to pH 5.7 with 1 M KOH. The plated seeds were subjected to white light for 12 h to induce germination before incubation under the specified light conditions. For dark incubation, seeds were grown in continuous darkness at 22 °C for 5 days; for the dark-to-light treatment, 5-day-old etiolated seedlings were exposed to white light (100 μmol $m^{-2}$ $s^{-1}$) at 22 °C for the indicated times; for white light treatment, seeds were grown under continuous white light (100 μmol $m^{-2}$ $s^{-1}$) at 22 °C for 5 days. For the thermotolerance assay, 5-day-old etiolated seedlings were illuminated with white light (100 μmol $m^{-2}$ $s^{-1}$) at 22 °C for 24 h, subsequently transferred to 40 °C for 3 h, and then returned to 22 °C. Chlorophyll content was measured after 2 days of recovery growth.

### Plasmid Construction

The *CRISPR-u5-4* and *CRISPR-rps1* mutants were generated by YAO promoter-controlled CRISPR/Cas9 system[43]. To generate transgenic plants overexpressing U5-3 and U5-4, we first constructed a modified vector named as pCAMBIA1300-UBQ. A fragment containing the *Arabidopsis UBQ10* promoter (636 bp), multiple cloning site (MCS) with restriction sites *Kpn*I, *Eco*RI, *Nco*I and *Bam*HI, and the OCS (octopine synthase) terminator was obtained by the overlap extension PCR, and then was inserted into the *Bam*HI site of the pCAMBIA1300 vector (Biovector) using in-fusion cloning (Takara), resulting in pCAMBIA1300-UBQ. Subsequently, to generate ProUBQ: U5-3 and ProUBQ: U5-4 constructs, sequences containing 432 bp upstream of U5-3 and 383 bp upstream of U5-4, along with their full-length coding sequences plus downstream sequences (430 bp for U5-3 and 530 bp for U5-4) were amplified, and then were inserted into the *Eco*RI site of pCAMBIA1300-UBQ vector respectively by in-fusion cloning (Takara).

For the dual luciferase assay, we generated a modified dual-luciferase vector named as pYQ000-dual-LUC. First, the CaMV 35S promoter and NOS terminator were cloned and inserted into the *Eco*RI and *Hind*III sites of pCAMBIA1300 vector respectively by in-fusion cloning to make the pCAMBIA1300-YQ000 vector; then the fragments of *Pro35S: Renilla luciferase* (REN) and *Pro35S: Firefly luciferase* (LUC) were cloned into the *Sac*I/*Kpn*I and *Sal*I/*Pst*I sites of pCAMBIA1300-YQ000, resulting in Pro35S: REN-Pro35S: LUC-NOS (pYQ000-dual-LUC). Subsequently, each of the 3′UTRs — *RPS1* (262 bp), *GHS1* (168 bp), *LHCB3* (185 bp), *PnsB2* (671 bp), *PnsB4* (760 bp), *PnsL4* (492 bp) and *UBQ5* (398 bp) — was cloned downstream of the *LUC* gene in the pYQ000-dual-LUC vector by in-fusion cloning. To generate the UBE-deleted (LUC-Δ3′UTR) and UBE-mutated 3′UTR (LUC-3′UTRm) reporters, UBE in each 3′UTR of *RPS1*, *GHS1*, *LHCB3*, *PnsB2*, *PnsB4* and *PnsL4* was either deleted or substituted with the sequences complementary to UBE in the 3′UTR, and then cloned into the *Pst*I site of pYQ000-dual-LUC by in-fusion cloning (Supplementary Fig. S4g). To generate ProUBQ: U5-4m, the Sm site of U5-4 in ProUBQ: U5-4 was substituted with UBE using primer-based site-directed mutagenesis (Supplementary Fig. S5c). LUC-3′UTRm reporters were generated with nucleotides complementary to the UBE sequences in each of the 3′UTRs (Supplementary Fig. S5c). *ProUBQ: SNU114*, *ProUBQ: EF2a* and *ProUBQ: EF2a*$^{H754A/E803A}$ were generated by inserting their coding sequences into the *Eco*RI site of pCAMBIA1300-UBQ vector by in-fusion cloning. The mutated sequences were created using PCR-based mutagenesis. For the MS2 tethering assay, the 6× MS2 fragment[44] was synthesized by Genscript (Nanjing, China) and cloned into the *Pst*I site of pYQ000-dual-LUC by in-fusion cloning. To generate the Pro35S: MCP-YFP construct, the coding sequence of eYFP was inserted into the *Pst*I site of pCAMBIA1300-YQ000 by the in-fusion cloning to create the pCAMBIA1300-YQ-eYFP vector. Subsequently, coding sequence of MCP was inserted into the *Bam*HI and *Xba*I sites of pCAMBIA1300-YQ-eYFP. The Pro35S: MCP-EF2a-YFP construct was generated by inserting the coding sequence of *EF2a* without stop codon into the *Bam*HI site of Pro35S: MCP-YFP vector (Supplementary Fig. S8g–i).

To generate the *Pro35S: SmD3a-His-FLAG*/WT (SmD3a-HF) and *Pro35S: SmD3b-His-FLAG*/WT (SmD3b-HF) transgenic plants, the SmD3a- and SmD3b-coding sequences without their stop codons along with 3× FLAG and 6× His were obtained by overlap extension PCR and then cloned into the *Bam*HI site of pCAMBIA1300-YQ000 vector respectively by in-fusion cloning. To generate *Pro35S: SmFa-3HA* transgenic plants in the WT background, the coding sequence of 3× HA was inserted into the *Pst*I site of pCAMBIA1300-YQ000 to create the pCAMBIA1300-YQ-3HA vector. Then the SmFa-coding sequence without stop codon was PCR-amplified from WT cDNA and inserted into the *Kpn*I site of pCAMBIA1300-YQ-3HA vector by in-fusion cloning. Similarly, to generate *Pro35S: SmD2a-3HA/smd2a* transgenic plants, the SmD2a-coding sequence without stop codon was PCR-amplified and inserted into the *Kpn*I site of pCAMBIA1300-YQ-3HA vector by in-fusion cloning. To generate ProSmD3a: SmD3a-YFP and ProSmD3b: SmD3b-YFP constructs, the 1,548 bp upstream sequence from the *SmD3a* start codon and its coding sequence without stop codon were cloned into the *Sac*I and *Xba*I sites of pCAMBIA1300-YQ-eYFP vector, respectively. Similarly, the 1,458 bp upstream sequence from *SmD3b* start codon along with its coding sequence without stop codon was inserted into the *Sac*I and *Xba*I sites of pCAMBIA1300-YQ-

eYFP, respectively. ProSmD3a: SmD3b-YFP construct was generated by cloning the 1,548 bp upstream sequence from the *SmD3a* start codon along with SmD3b-coding sequence without stop codon into the *Sac*I and *Xba*I sites of pCAMBIA1300-YQ-eYFP, respectively. To generate *Pro35S: EF2a-GFP* and *Pro35S: EF2a*$^{H754A/E803A}$*-GFP* transgenic plants in the *ef2a* mutant background, mGFP5-coding sequence was inserted into the *Pst*I site of pCAMBIA1300-YQ000 to generate the pCAMBIA1300-YQ-mGFP5 vector. The EF2a-coding sequence without stop codon was then amplified and inserted into the *Sal*I site of pCAMBIA1300-YQ-GFP using the in-fusion method. The EF2a$^{H754A/E803A}$ sequence was generated by PCR-based mutagenesis using the Pro35S: EF2a-GFP construct as the template. To generate the *ProRPS1: RPS1-3′UTR*$_{RPS1}$ and *ProRPS1: RPS1-Δ3′UTR*$_{RPS1}$ (with the UBE deleted in the 3′UTR) transgenic plants in the *rps1* mutant background, the 2,139 bp upstream sequence along with the 1,962 bp genomic sequence of *RPS1* was amplified and inserted into the *Kpn*I site of pCAMBIA1300 by in-fusion cloning. The *RPS1-Δ3′UTR*$_{RPS1}$ sequence was generated by PCR-based mutagenesis using the ProRPS1: RPS1-3′UTR$_{RPS1}$ construct as the template.

For the split-luciferase assay, fragment containing the SmD3a-coding sequence without stop codon was amplified and inserted into the *Kpn*I site of pCAMBIA1300-nLUC vector[45] by the in-fusion cloning. Fragments containing coding sequences of *EF2a* and *EF2a*$^{H754A/E803A}$ were amplified and inserted into the *Kpn*I site of pCAMBIA1300-cLUC vector by the in-fusion cloning, respectively. The EF2a$^{H754A/E803A}$ sequence was generated through PCR-based mutagenesis using a Pro35S: EF2a-GFP plasmid as the template. To purify the MBP-EF2a and MBP-EF2a$^{H754A/E803A}$ recombinant proteins, the coding sequences of EF2a and EF2a$^{H754A/E803A}$ were amplified and inserted into the *Eco*RI site of pMAL-C2X by in-fusion cloning, respectively. For purification of the GST-SmD3a recombinant protein, the coding sequence of *SmD3a* was amplified and inserted into the *Eco*RI site of pGEX4T-1 by the in-fusion cloning.

To generate modified U5 snRNA constructs (Fig. 3k) for the U5 snRNA functional assay on RPS1 expression, ProUBQ:U5-4 T1 and T2 were generated by inserting a 13-bp sequence complementary to the mutated UBE (M1, Fig. 3f) either downstream of the Sm site (T1) or within the 3′ loop (T2) into ProUBQ:U5-4 using primer-based mutagenesis. ProUBQ:U5-4 C1 and C2 contained reversed insertion sequences of T1 and T2, respectively.

### Isolation of nuclear and cytoplasmic RNA and library preparation for RNA deep-sequencing

Total seedlings were harvested and homogenized to a fine powder in liquid nitrogen, and then suspended in cell lysis buffer (20 mM Tris-HCl, pH 8.0, 50 mM KCl, 5 mM $MgCl_2$, 0.1% NP-40, 250 mM sucrose, 5 mM DTT, 1× protease inhibitor cocktail and 100 U/mL RNase inhibitor). The mixture was then centrifuged at 1,500× *g* for 15 min at 4 °C to separate the supernatant (cytoplasmic fraction) for preservation. The pellet was washed with nuclear resuspension buffer (10 mM Tris-HCl, pH 8.0, 2 mM $MgCl_2$, 0.1% NP-40, 250 mM sucrose, 5 mM DTT, 1× protease inhibitor cocktail and 100 U/mL RNase inhibitor) and kept for the subsequent RNA extraction and library construction for RNA-seq.

The cytoplasmic fraction was filtered through two layers of Miracloth and then mixed with cytosol supplementary buffer to the following final concentration of each reagent: 100 mM Tris-HCl, pH 8.0, 50 mM KCl, 25 mM $MgCl_2$, 1% PTE, 1% Sodium DOC, 5 mM DTT, 1× protease inhibitor cocktail and 100 U/mL RNase inhibitor, and centrifuged at 16,000× *g* for 10 min at 4 °C to remove any possible debris. The supernatant was loaded onto a linear 10–60% sucrose gradient to collect the free cytoplasmic fractions (cytosolic without ribosomes), 80S ribosome (monosome) fractions and translation-activated ribosome (polysome) fractions as previously described[46].

Fractionated RNA was immediately resuspended in TRIzol-LS (for cytoplasmic RNA) or TRIzol (for nuclear RNA), followed by DNase I treatment to remove genomic DNA. All the RNA in the subcellular fraction was subjected to rRNA depletion via a Ribo-Zero rRNA removal kit. Then, rRNA-depleted RNA was used to construct strand-specific RNA-seq libraries with the NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina (NEB). All libraries were pooled for paired-end 150 bp sequencing via an Illumina HiSeq 4000 system.

### Subcellular RNA-seq data analysis

For the high-throughput sequencing data analysis, the adapter sequences in the raw data were first removed with Cutadapt[47] using the parameters "-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA-A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -m 20". Reads of low quality were removed via fastp[48]. The reads mapped to ribosomal RNAs were subsequently filtered via Bowtie2[49]. Next, we mapped them to *Arabidopsis* genome (TAIR 10) using HISAT2[50] with the parameters "--dta --min-intronlen 50 --max-intronlen 6000" and assembled the transcripts via StringTie[51] with the parameters "--rf -m 100 -j 5 -c 5", respectively. Only unique mapping reads were retained. The gene annotations were obtained from Araport11[52], while the new assembled transcripts were subjected to coding potential analysis via CPC2[53] and COME[54]. Read counts were obtained using featureCounts v1.6.4[55] and used to calculate gene expression levels, represented as fragments per kilobase of transcript per million mapped reads (FPKM) and transcripts per million (TPM).

To identify RNAs with altered subcellular localization under different light conditions, differential expression analysis was performed using DESeq2[56]. Fold enrichment ratios were calculated for cytoplasm without ribosomes (Cy) vs nucleus (Nu) (Cy/Nu score) and polysome (P) vs monosome (M) fold enrichment (P/M score) for each RNA. RNAs were considered relatively enriched in the cytoplasm and polysome fractions if they met the criteria Cy/Nu score > 2 and P/M score > 2 with $P < 0.05$. The DESeq2-normalized snRNA expression matrix was visualized as a heatmap using the ComplexHeatmap package[57].

### U5 snRNA variant alignment and quantification

Reads used for U5 snRNA quantification were aligned to the *Arabidopsis* reference genome supplemented with the 14 previously annotated U5 snRNA variants[58]. These variants range from 123 to 128 nt in length and are highly conserved

in loop I, stem I, and the Sm-site regions, but contain variant-specific sequence differences in other regions. Only perfectly matched reads with no mismatches were retained for quantification. Reads matching more than one U5 variant were excluded from variant-specific quantification. Under these criteria, retained reads could be assigned to individual U5 variants.

### Circular RT-PCR

To determine the size of matured U5-3 and U5-4 snRNAs after processing, a circular RT-PCR assay was performed following a previously described protocol[46]. The purified RNA was circularized using T4 RNA ligase I (NEB) and subjected to reverse transcription with specific primers. The resulting DNA products were then amplified via PCR and subsequently cloned and inserted into the pEASY-T vector for DNA sequencing and analysis. Primers used are listed in Supplementary Table S9.

### Northern blot analysis

Northern blot analysis was performed as described previously[46]. The oligonucleotide probes used in this study are listed in Supplementary Table S9.

### mRNA-seq and data analysis

Total RNA from WT, *u5-3 u5-4* was used for mRNA purification with Dynabeads Oligo $(dT)_{25}$ (Invitrogen). The purified mRNA was used to construct RNA-seq libraries with the NEBNext Ultra™ II RNA Library Prep Kit for Illumina (NEB). Libraries were barcoded, pooled and sequenced on an Illumina NovaSeq 6000 machine (paired-end, 150 bp). Quality control was performed using fastp[48] to remove low-quality reads. Clean reads were aligned to the TAIR10 reference genome using HISAT2[50] with the parameters: --dta --min-intronlen 50 --max-intronlen 6000. Read assignment was performed using featureCounts v1.6.4[55] based on Araport11 annotations. Transcriptome changes were analyzed with DESeq2[56]. Differentially expressed genes were initially defined using a fold change cutoff of > 1.5 and *P* value < 0.05. FDR-adjusted *q* value was also calculated to assess the robustness of the analysis.

For the analysis of alternative splicing (AS), rMATS v4.0.2[59] was used to detect differential AS events, including alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), retained intron (RI), and skipped exon (SE). The following parameters were applied: -t paired --readLength 150 --libType fr-unstranded --tstat 15 --anchorLength 6. Events with a false discovery rate (FDR) < 0.05 and a difference in the percent spliced in index (ΔPSI) > 0.1 were considered significantly differential AS events.

GO term analysis was performed using the clusterProfiler package[60]. All the libraries were constructed with three biological replicates. Data analyses were performed using the R software package.

### Polysome-seq and data analysis

Polysome profiling was carried out as previously described with minor modifications[46]. One gram of *Arabidopsis* seedlings was ground in liquid nitrogen and resuspended in 1 mL of precooled polysome extraction buffer (200 mM Tris-HCl, pH 9.0, 200 mM KCl, 35 mM $MgCl_2$, 25 mM EGTA pH 8.0, 1% detergent mix, 1% sodium deoxycholate, 1% PTE, 5 mM dithiothreitol, 1 mM PMSF, 50 μg/mL cycloheximide, 50 μg/mL chloramphenicol and 100 U/mL RNase inhibitor). The mixture was centrifuged at 4 °C and 16,000× *g* for 35 min, and a total of 6,000 $A_{260}$ units of the supernatant were layered onto a 5–50% sucrose gradient. After ultracentrifugation in a Hitachi P40ST rotor at 35,000 rpm for 2.5 h at 4 °C, the sucrose gradients were fractionated into 30 fractions with a Piston gradient fractionator (Biocomp Instruments), and the ultraviolet absorbance at 260 nm was continuously monitored via a Model EM-1 Econo UV Monitor (Bio-Rad). Additionally, 10% of the supernatant volume was reserved as input for total RNA or protein isolation. The fractions numbered 24 to 27 were utilized for the subsequent polysome-seq analysis. For ribosome dissociation analysis, 6,000 $A_{260}$ units of the supernatant were treated with 50 mM EDTA for 30 min on ice before ultracentrifugation.

For polysome-seq, libraries were constructed using the NEBNext Ultra™ II Directional RNA Library Prep Kit for Illumina, and sequenced on the Illumina HiSeq X platform (paired-end, 150 bp). The sequencing data obtained from polysome profiling were designated as polysome-bound fragments, while those from RNA input were denoted as mRNA input. Translation efficiency was defined as the ratio of polysome-protected fragments to mRNA input, reflecting the relative occupancy of active polysomes per mRNA species[61]. A one-sided Student's *t*-test (significance level $\alpha = 0.05$) was applied for statistical analysis. $|\log_2 FC| > 0.5$, $P < 0.05$ were considered indicative of differential TE.

### TMT-based quantitative proteomic analysis

Proteomics analysis of WT and *u5-3 u5-4* mutant was conducted as previously described[62]. Briefly, the plant powder was lysed in a buffer containing 4% SDS and 100 mM Tris-HCl (pH 7.6). After centrifugation, total proteins were subjected to chloroform-methanol precipitation, and the concentrations were quantified with a BCA protein assay kit (Thermo Scientific, Rockford, IL). Proteins were digested using the filter-aided proteome preparation (FASP) protocol as previously described[63], and labeled with acetonitrile-dissolved TMT reagents (Thermo Scientific, Rockford, IL). The labeled samples were then pre-fractionated via reversed-phase (RP)-high performance liquid chromatography (HPLC). For LC-MS/MS/MS analysis, an Orbitrap Fusion™ Lumos™ Tribrid™ mass spectrometer (Thermo Scientific, Rockford, IL Waltham, MA) was used. Database searches for all the raw files were conducted using the software MaxQuant (version 1.6.3.4), with default values applied for all the other parameters. In the quantitative proteomics data, proteins with a fold change > 1.2, or < 0.88, with $P < 0.05$ as determined by a two-tailed Student's *t*-test in the *u5-3 u5-4* mutant compared to the WT were considered differentially expressed.

### Ribosome profiling sequencing and data analysis

Ribosome profiling was performed as previously described, with minor modifications[64]. Briefly, polysomes were extracted from WT and *u5-3 u5-4* mutant *Arabidopsis* seedlings grown under DTL-24 conditions with buffer D (100 mM Tris-HCl (pH 8.0), 40 mM KCl, 20 mM $MgCl_2$, 2% (v/v) polyoxyethylene (10) tridecyl ether, 1% (w/v) deoxycholic acid, 1 mM DTT, 100

μg/mL cycloheximide, and 10 units/mL DNase I). The lysate was then treated with RNase I (NEB) for 1 h at 23 °C. Ribosomes were isolated using size-exclusion columns (illustra MicroSpin S-400 HR Columns; GE Healthcare). Following RNA isolation with the RNA Clean & Concentrator-25 kit (Zymo Research), ribosome footprints ranging from 25 to 35 nt were recovered after separation on a denaturing gel. The purified RNA fragments were then used for library construction with the NEBNext® Multiplex Small RNA Library Prep Set for Illumina (NEB). The resulting cDNA libraries were sequenced on an Illumina NovaSeq platform (150-bp paired-end reads).

Ribosome-protected fragments (RPFs) were quality filtered and adaptor trimmed prior to analysis using fastp[48]. Reads mapping to rRNA and other contaminating RNA species were removed by Bowtie2[49]. The remaining reads were aligned to the *Arabidopsis thaliana* reference genome using STAR[65] in two-pass mode, and uniquely mapped reads were retained for downstream analyses.

RPF length distributions were evaluated using Ribotricer[66], and read lengths exhibiting robust three-nucleotide periodicity (phase score > 0.9) were selected for downstream analyses. Based on this criterion, RPFs of 25–32 nt were retained. riboWaltz[67] was then used to estimate read-length-specific P-site offsets, and the corresponding offsets were applied to assign single-nucleotide P-site positions for each RPF. For metagene analysis, P-site positions assigned by riboWaltz were analyzed separately for each biological replicate. P-site densities were calculated relative to annotated translation start codons using windows of −20 to +40 nt around start codons. Transcripts with fewer than five P-site counts within the analysis window were excluded. To minimize the influence of transcript abundance, P-site counts were normalized within each gene before averaging across all genes, generating per-gene normalized metagene profiles that were visualized in R.

### U5 snRNA KARR-seq

U5 snRNA KARR-seq was performed as previously described[18] with some modifications. In brief, seedlings were treated with 0.1% formaldehyde, then cytosolic ribosome was extracted by sucrose cushion and treated with N3-kethoxal, subsequently RNA in proximity was crosslinked by dendrimers. The sample was then treated with proteinase K and DNase. RNA was extracted using phenol chloroform. 10–50 ng of ligated RNA was used for reverse transcription (RT) using the RT reagents in the AccuNext Stranded RNA-seq Library Kit for Illumina (AG12504). After RT, 2 μL of RNase H (Thermo, EN0201) was added to each sample, which was then incubated for 20 min at 37 °C. The reaction was quenched by heating inactivation for 5 min at 70 °C. Subsequently, cDNA was purified using the DNA Clean & Concentrator-5 kit. 10% of the cDNA was saved as an input sample, while the remaining cDNA was used for U5 snRNA enrichment. After streptavidin beads pulldown, the resulting cDNA was subjected to library construction using the AccuNext Stranded RNA-seq Library Kit for Illumina (AG12504).

### U5 snRNA-enriched KARR-seq data processing, analysis, and visualization

KARR-seq data analysis follows a published pipeline (github.com/Bei-1205/snoKARR-seq)[18]. Adaptors from raw reads were trimmed, and the overlapping read pairs (2 × 150 bp) were merged into single-end reads and de-duplicated using fastp[48]. To assess U5 enrichment, the resulting single-end FASTQ files were aligned to *Arabidopsis* genome (Araport 11) using STAR[65]. Primary chimeric reads containing different snRNAs were extracted and counted to assess U5 chimeric-reads enrichment. The coordinate pairs (U5 and targets) were then subjected to duplex group (DG) clustering using a spectral clustering algorithm[68]. To assess the reproducibility of KARR-seq, multiBamSummary from deepTools[69] was used to calculate the genomic coverage correlation, and unique chimeric read counts from each DG group in two replicates were used to evaluate the correlation of chimeric reads. Unique chimeric read counts for each DG group were obtained by intersecting U5 and target coordinates in each DG group with raw chimeric read coordinates.

To detect U5–mRNA interactions, FASTQ files were aligned to customized reference genome containing sequences of U5 variants (Supplementary Table S2) and all mRNA transcripts from TAIR10. Chimeric reads containing U5 snRNA sequences were extracted from the resulting BAM files. The U5 snRNA target sequences from reproducible chimeric reads from two replicates were used to perform sequence motif analysis using findMotifs.pl in the HOMER suite[70]. The structure and free energy of the duplex formed between U5 Sm sites and their targets identified were predicted using RNAplex (with option "-c 5, -l 40")[71]. Duplex structures with duplex length > 8 and free energy < −8 kcal/mol were considered as stable interactions. To visualize the binding regions of U5, the bed file containing U5 binding sequences was converted to a BigWig file using bedGraphToBigWig[72]. Genome tracks displaying U5 target coverage were plotted using CoolBox[73]. The U5 secondary structure was visualized using VARNA[74]. For motif enrichment analysis, random background sequences were generated using bedtools shuffle from the BEDTools suite. Specifically, snRNA target intervals were shuffled across *Arabidopsis* genome while preserving their lengths, restricting the shuffled regions to TAIR10-annotated loci, and preventing overlaps. The resulting random intervals were then used as a matched background set for HOMER motif enrichment against the observed snRNA target sequences.

### Electron microscopy for detection of U5 snRNA-*RPS1* mRNA co-localization

The sample preparation for electron microscopy was performed as described previously[75], with slight modification. In brief, for RNA double immunogold labeling, biotin-labeled *RPS1* mRNA probes and digoxigenin-labeled U5 probes were hybridized with the samples overnight at room temperature. Anti-biotin antibody (rabbit) and anti-digoxigenin antibody (mouse) were added, along with anti-rabbit IgG conjugated to 12 nm gold-nanoparticles and anti-mouse IgG conjugated to 6 nm gold-nanoparticles. The specimens were examined in Tecnai G2 Spirit BioTWIN transmission electron microscope (FEI company), and the distance between the two types of immuno-gold particles in the images of samples under different growth conditions was measured using ImageJ software.

### U5 snRNA functional assay on RPS1 expression

The plasmids proUBQ10:U5-T1, C1, T2, and C2 were individu-

ally transformed into protoplasts extracted from 10-day-old *Arabidopsis* M1 (Fig. 3k) seedling mesophyll cells and incubated under weak light for 14 h. Protoplasts transformed with an empty vector or a vector expressing unmodified U5 snRNA served as controls. The protoplast transfection was performed as described[76].

For protein extraction, the protoplasts were lysed in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, and 1× protease inhibitor cocktail). Immunoblot analysis was performed using anti-RPS1 antibody. PEPC was used as a loading control.

**RNA Immunoprecipitation (RIP)**

RIP analysis of sucrose gradient fractions was performed as previously described[30], with some modifications. In brief, the indicated polysomes fractions derived from sucrose gradients were collected and subsequently centrifuged at 35,000 rpm overnight at 4 °C. The resulting pellets were resuspended in 200 µL of ice-cold polysome extraction buffer before being added to 1 mL of IP buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT, 0.1% NP-40, 1× protease inhibitor cocktail and 100 U/mL RNase inhibitor). The mixture was incubated overnight at 4 °C with 2 µL anti-SmB (Proteintech), anti-SmD2 (Proteintech) or anti-SmD3 antibody (Proteintech). Subsequently, prewashed protein A/G magnetic beads (Invitrogen™) were added to the mixture and incubated for 2 h at 4 °C, followed by three washes with IP buffer. After the final wash, one-third of the beads were boiled in 1× SDS loading buffer for western blot, while the remaining beads were treated with TRIzol reagent for RNA extraction and analyzed by Northern blot or RT-qPCR.

**RNA pull-down**

RNA pull-down analyses of sucrose gradient fractions were performed as previously described[46] with some modifications. Briefly, the indicated polysomes fractions derived from sucrose gradients were collected and then centrifuged at 35,000 rpm overnight at 4 °C. The resulting pellets were resuspended in 200 µL of ice-cold polysome extraction buffer before being added to 1 mL hybridization buffer (750 mM NaCl, 1% SDS, 50 mM Tris-HCl, pH 7.0, 1 mM EDTA, 15% formamide, 1× protease inhibitor cocktail and 100 U/mL RNase inhibitor). Next, 100 pmol of RNA oligonucleotide probes were added to the lysate and incubated with continuous rotation at 37 °C for 4 h. Then, prewashed Dynabeads™ MyOne™ Streptavidin C1 Beads (Invitrogen) were added, and the mixture was rotated at 37 °C for 30 min. The beads were then washed twice with low-salt wash buffer, and the bead-bound complexes were eluted using either 1× SDS loading buffer for western blot or RNA elution buffer (100 mM NaCl, 50 mM Tris-HCl, pH 7.0, 1 mM EDTA and 1% SDS) for Northern blot or RT-qPCR. Biotinylated and nonbiotinylated RNA oligonucleotide probes targeting U5 snRNAs are listed in Supplementary Table S9.

For *in vitro* RNA interaction analyses, templates for *in vitro* transcription were amplified by PCR and purified using a NucleoSpin® Gel and PCR Clean-up Kit (MACHEREY-NAGEL). Biotin-labeled RNA was transcribed using Biotin-16-UTP (Roche) and the AmpliScribe™ T7-Flash™ Transcription Kit (Lucigen). The transcribed RNA was then treated with RNase-free DNase I and subsequently purified using the RNA Clean & Concentrator-25 (Zymo Research). A mixture containing equimolecular amounts of *in vitro* transcribed U5 snRNA and target mRNA in 1 mL of IP buffer was incubated at 4 °C for 3 h. Prewashed streptavidin C1 beads were then added, and the incubation continued at 4 °C for an additional 2 h. After the beads were washed for three times with IP buffer, the RNA was extracted with TRIzol reagent and subjected to northern blot analysis.

**Phenotypic analyses**

Chlorophyll levels were measured as previously described[77]. In brief, 50 mg of seedlings were homogenized in liquid nitrogen and immersed in 95% (v/v) ethanol to extract the chlorophyll. The absorbances of the samples were measured at 665 nm and 649 nm via a UV-1800 spectrophotometer (SHIMADZU). The chlorophyll content was calculated using the equation $(6.63A_{665}+18.08A_{649})$/g fresh weight. At least three independent biological replicates were performed. To assess cotyledon greening rates, at least 50 seedlings were evaluated for one biological replicate as previously reported[77]. Seedlings with expanded green cotyledons were counted. Triplicates of biological replicates were performed, and the data are presented as the mean ± SD, n ≥ 50. Chlorophyll fluorescence was measured by a Closed FluorCam FC 800 (Photon Systems Instruments (PSI), Czech Republic). We measured photosynthetic parameters when the plants were dark adapted for 30 min before measuring the basal fluorescence (F0), which means initial fluorescence intensity. This was followed by measuring maximum dark-adapted fluorescence (Fm), which means maximal fluorescence yield. Dark adapted quantum efficiency of photosystem II was calculated as Fv/Fm (Fv = Fm−F0)[78].

**RT-qPCR**

*Arabidopsis* seedlings and tobacco leaves were harvested in triplicate. Total RNA was extracted using the Spectrum™ Plant Total RNA Kit (Sigma) or TRIzol reagents (Invitrogen™). Subsequently, DNase I treatment was performed, and the extracted RNA was reverse-transcribed using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific™). The subsequent quantitative PCR (qPCR) was conducted using SYBR Premix ExTaq Kits (TaKaRa) on a 7500 Fast Real-Time PCR System (Applied Biosystems) following the manufacturer's instructions. The primers used are listed in Supplementary Table S9.

**IP-MS analysis**

IP analysis was performed as described previously[45] with minor modifications. Briefly, to identify SmD3a- and SmFa-associated proteins on polysomes, 30 µL prewashed anti-Flag nanobody agarose beads or anti-HA nanobody agarose beads were applied to the polysome fractions extracted from *Pro35S: SmD3a-His-FLAG*/WT or *Pro35S: SmFa-3HA*/WT seedlings for 4 h at 4 °C, respectively. After incubation, the beads were washed twice with IP buffer and heated for 10 min in 1× SDS loading buffer at 100 °C. The protein samples were purified from 12% SDS-PAGE gels and digested overnight with LysC-trypsin at 37 °C. After quenching with 5% formic acid, the samples were desalted using a Monospin C18 column (GL Science, Tokyo, Japan). Liquid chromatography

and mass spectrometry analyses were performed using the nanoElute liquid chromatography system (Bruker Daltonics) and a hybrid TIMS quadrupole time-of-flight mass spectrometer (Bruker tims TOF Pro). The MS/MS raw data were analyzed using MaxQuant (version 1.5.3.30) software (Max Planck Institute of Biochemistry).

### BONCAT assay

The BONCAT assay was performed as described previously[79] with some modifications, to assess nascent protein synthesis of selected proteins. In brief, five-day-old seedlings were fully submerged in AHA labeling medium (1/2 MS liquid medium, 5% sucrose, 0.025 ‰ Silwet L-77, 1 mM AHA and 10 mM citrate buffer) for 2 min, while the labeling medium without AHA served as the control. After the liquid medium was removed, the seedlings were exposed to constant white light for 4 h at 22 ℃. All the samples were subsequently harvested immediately, pulverized with liquid nitrogen and resuspended in PBS buffer. The lysates were subjected to sonication at 4 ℃ for 40 min and then subjected to click reactions for 2 h. Total protein was extracted by methanol/chloroform/water precipitation, and approximately 300 μg protein samples were added to prewashed streptavidin C1 beads and shaken overnight. The beads were washed three times with PBS buffer and then boiled in 1× SDS loading buffer. The newly synthesized proteins were analyzed by western blot.

### Luciferase reporter assay

Dual-luciferase assays in *N. benthamiana* plants were conducted as previously described[80] with minor modifications. After a 3-h incubation at 28 ℃ in an infiltration buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$ and 100 μM acetosyringone), *Agrobacterium tumefaciens* strain GV3101 containing the indicated effector and reporter constructs were infiltrated into fully expanded *N. benthamiana* leaves at an $OD_{600}$ of 0.5. After infiltration, the *N. benthamiana* plants were kept in darkness for 12 h, then exposed to white light for 36 h before analysis. Luminescent signals of LUC and REN were detected using the Dual-Luciferase Reporter Assay System Kit (Promega), and LUC expression levels were quantified as the ratio of LUC to REN enzyme activities.

Luciferase assays in *Arabidopsis* protoplasts were performed as previously described[81]. In brief, 10 μg of each indicated effector and reporter construct were transformed into one million protoplasts via PEG-mediated transfection, followed by overnight incubation under weak light. Luminescent signals of LUC and REN were measured with the Dual-Luciferase Reporter Assay System Kit (Promega), and LUC expression was quantified by calculating the LUC/REN enzyme activity ratio.

For the MS2 tethering assay, *Agrobacterium tumefaciens* strain GV3101 carrying either the LUC or LUC-6× MS2 reporter construct was mixed at 1:1 ratio with strains containing MCP-EF2a-YFP or MCP-YFP constructs, then infiltrated into *N. benthamiana* leaves. After 48 h of incubation under white light, dual luciferase assays were performed. Relative LUC activity was determined by calculating the LUC/REN activity ratio, while relative *LUC* mRNA was calculated as the ratio of *LUC* mRNA to *REN* mRNA.

The firefly LCI assay was performed as previously described[45], and the constructed plasmids used in this study are listed in Supplementary Table S9.

### *In vitro* pull-down assays

The *in vitro* pull-down assays were performed as described previously[45]. Briefly, plasmids expressing the indicated recombinant proteins were transformed into *E. coli* strain BL21 competent cells, with protein expression induced by IPTG (TransGen Biotech). The recombinant proteins tagged with GST or MBP were purified using GLUTATHIONE SEPHAROSE 4B (Cytiva) or Amylose Resin (NEB), respectively. Then, equal amounts of bait and prey recombinant proteins, or the GST/MBP control were incubated in 1 mL binding buffer at 4 ℃ for 2 h. Prewashed beads were added to the mixture and incubated for an additional 2 h. The beads were subsequently washed five times before being boiled in 1× SDS loading buffer. The eluted proteins were subjected to western blot analysis.

### Prediction of the interaction surface between EF2a and SmD3a

The structural models of individual proteins of EF2 and SmD3 from *Arabidopsis thaliana* and *Homo sapiens* were obtained via AlphaFold2-multimer[82]. Protein–protein docking was performed utilizing Schrödinger package (version 2021-04)[83]. A total of 100 models were initially generated and submitted for clustering. Ten diverse binding poses were then selected for further MD simulation refinement, with Desmond module[84] in Schrödinger. The MD simulation was performed utilizing the OPLS force field[85] and the TIP3P model for 200 ns for each complex model. Finally, the top-ranked pose with minimal root-mean-square-fluctuation (RMSF) and interaction energy was determined as the binding pose of EF2 and SmD3 for further analysis.

### Phylogenetic analyses

The homolog proteins of *A. thaliana SmD3a* and *SmD3b* were identified via blastp[86] against the Phytozome v14[87] database. For the species without annotation, the CDS sequences were determined based on results of tblastn[86] against the genome assembly. CDS sequences were aligned using the translated alignment model of PRANK[88]. Maximum-likelihood phylogenetic trees were constructed using RAxML[89] with the GTRGAMMA nucleotide substitution model and 100 rapid bootstrap replicates. The signatures of positive selection were detected using the PAML[90] branch-site model, and the sites under selection were determined by the Bayes Empirical Bayes method.

The disorder scores of SmD3 proteins were predicted using metapredict V3 (metapredict.net)[91].

### Comparative genomic analyses

The genome annotations of *Arabidopsis thaliana* TAIR10[92], *Arabidopsis lyrata* v2.1[93], *Eutrema salsugineum* v1.0[94], *Cleome violacea* v2.1, and *Theobroma cacao* v2.1[95] were downloaded from Phytozome v14[87]. The homologous gene pairs were identified using the function "compara.catalog ortholog" in JCVI[96], and the local synteny blocks were determined using the function "compara.synteny mcscan". The blocks containing target genes were visualized using "graph-

ics.synteny".

### Rational design of U5 snRNA/Sm protein translational pathway

Secondary structure predictions of *Arabidopsis* SmD3a and SmD3b, and their homologs from *Nicotiana benthamiana*, *Solanum lycopersicum*, *Oryza sativa*, and *Glycine max* were predicted using PSIPRED v4.0[97]. Tertiary structures were obtained from the Alphafold database[98]. To engineer a SmD3b variant that mimics the C-terminal structure of SmD3a, we employed a rational design strategy based on the ESM-2 protein language model[99] and evolutionary conservation patterns. Seven targeted substitutions (K80R, K86R, L88G, D89G, I92G, K93R, and K95R) were introduced to (i) enhance RG/RGG motif density via K-to-R substitutions and (ii) disrupt helix-stabilizing interactions via G substitutions, thereby promoting a loop-rich C-terminal configuration. Designed variants were re-evaluated by PSIPRED to ensure a predominant loop/coil conformation. Predicted nuclear localization signals were scanned using cNLS Mapper (cutoff 5.0) and NucPred, and candidates containing disrupted NLS motifs were selected for further experimental validation.

*Arabidopsis* SmD3a and SmD3b and designed variant (SmD3b-M) were codon-optimized for *Nicotiana benthamiana* and cloned into the pCAMBIA1300-YQ-eYFP vector. Transient transformation in *N. benthamiana* leaves was followed by confocal microscope (Zeiss LSM710) analysis for subcellular localization. U5 snRNA mutants (ProUBQ:U5-M1 and ProUBQ:U5-M2) were generated by modifying nucleotides in Sm site to match the UBE sequences of *NbeRPS1* and *NbeLHCB3*, respectively (Fig. 6h). Endogenous RPS1 and LHCB3 protein levels were assayed relative to the empty-vector control. All plasmids and primers in this study can be found in Supplementary Table S9.

## DATA AVAILABILITY

The data supporting the findings of this study are available within the paper and its Supplementary Materials. The RNA sequencing data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession codes GSE266658 and GSE283246 (Arabidopsis seedlings mRNA-seq), GSE266659 (subcellular profiling), GSE266868 (Polysome-seq), GSE289876 (KARR-seq), GSE338488 (Ribo-seq). The mass spectrometry proteomics data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD058958 (TMT-proteome) and PXD059033 (SmD3a/SmFa IP-MS). Additionally, previously published proteomics data regarding *Arabidopsis* seedlings during de-etiolation were analyzed and are available under accession code PXD021518 (Pipitone *et al.*). Source data are provided with this paper.

## ACKNOWLEDGMENTS

We thank the National Center for Protein Sciences and the Core Facilities of Life Sciences at Peking University. We are grateful to Dr. Peiyong Xin and Dr. Jinfang Chu for assistance with tandem mass tag (TMT)-based quantitative proteomic data collection and analysis. We thank Prof. Catherine C.L. Wong and Shuaixin Gao for assistance with IP-MS experiments, Dr. Wei Wang and Dr. Ying Li for sharing the BONCAT method. We thank Drs. Xuemei Chen, Yijun Qi, Chengqi Yi, Ning Gao, Yangming Wang, Peng Du, Ningning Li, Jizong Wang, Zhengdong Wang, Hao Hu, Chuyun Gao, Chunyan Liu, Zhipeng Zhou, Rongcheng Lin, William Terzaghi, and Pan Zhu for helpful comments and technical assistance. This work was supported by Beijing Advanced Center of RNA Biology (to D.Z.), and by grants from the National Natural Science Foundation of China (91940301 and 32370252 to D.Z., 32230006 to X.W.D., 32325009 and 32170420 to W.Z.), the National Key R&D Program of China (2024YFA1306701 to X.W.D., 2023YFC3403200 to H.W.), and the key R&D Program of Shandong Province, China (ZR202211070163 to X.W.D.), Shanghai Action Plan for Science, Technology and Innovation (24JS2820200 to H.W.), Shennong Young Talent Program to D.Z., Benyuan Charity Young Investigator Exploration Fellowship in Life Science (to W.Z.), and the Feng Foundation of Biomedical Research and the Peking-Tsinghua Center for Life Sciences to W.Z.

## AUTHOR CONTRIBUTIONS

D.Z. and X.W.D. conceived and supervised the study. Y.J., R.Z., K.H., Y.G., Y.W., and L.Z. performed most of the experiments. Y.F., M.S., and J.H. performed the bioinformatic analysis. B.L. and C.H. provided important technical guidance and data analysis for KARR-seq. Y.Z. and W.Z. conducted the phylogenetic analysis of Sm proteins. J.F. and H.W. performed the rational design of Sm proteins. The manuscript was written by Y.J., Y.F., R.Z., K.H., B.L., Y.Z., C.H., X.W.D., and D.Z., with input from all authors.

## COMPETING INTERESTS

D.Z. and Y.J. are listed as inventors on a granted Chinese patent (No. CN122081331B), held by Peking University, related to the engineered U5 snRNA/Sm module described in this work for enhancing chloroplast protein accumulation and the maximum photochemical efficiency of photosystem II. The other authors declare no competing interests.

## REFERENCES

1. Will, C.L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
2. Wilkinson, M.E., Charenton, C. & Nagai, K. RNA splicing by the spliceosome. *Annu. Rev. Biochem.* **89**, 359–388 (2020).
3. Jones, M.H. & Guthrie, C. Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site. *EMBO J.* **9**, 2555–2561 (1990).
4. Lerner, M.R. & Steitz, J.A. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. USA* **76**, 5495–5499 (1979).
5. Patel, S.B. & Bellini, M. The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* **36**, 6482–6493 (2008).
6. Fischer, U., Englbrecht, C. & Chari, A. Biogenesis of spliceosomal small nuclear ribonucleoproteins. *WIREs RNA* **2**, 718–731 (2011).
7. Liu, J.L. & Gall, J.G. U bodies are cytoplasmic structures that contain uridine-rich small nuclear ribonucleoproteins and associate with P bodies. *Proc. Natl. Acad. Sci. USA* **104**, 11655–11659 (2007).
8. Benoit Bouvrette, L.P. et al. CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in *Drosophila* and human cells. *RNA* **24**,

98–113 (2018).
9. Gorbea, C., Mosbruger, T. & Cazalla, D. A viral Sm-class RNA base-pairs with mRNAs and recruits microRNAs to inhibit apoptosis. *Nature* **550**, 275–279 (2017).
10. Hyjek-Składanowska, M. et al. Core spliceosomal Sm proteins as constituents of cytoplasmic mRNPs in plants. *Plant J.* **103**, 1155–1173 (2020).
11. von Arnim, A. & Deng, X.W. Light control of seedling development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 215–243 (1996).
12. Fankhauser, C. & Chory, J. Light control of plant development. *Annu. Rev. Cell Dev. Biol.* **13**, 203–229 (1997).
13. Armarego-Marriott, T., Sandoval-Ibañez, O. & Kowalewska, Ł. Beyond the darkness: recent lessons from etiolation and de-etiolation studies. *J. Exp. Bot.* **71**, 1215–1225 (2020).
14. Pipitone, R. et al. A multifaceted analysis reveals two distinct phases of chloroplast biogenesis during de-etiolation in *Arabidopsis*. *eLife* **10**, e62709 (2021).
15. Cheng, M.C., Kathare, P.K., Paik, I. & Huq, E. Phytochrome signaling networks. *Annu. Rev. Plant Biol.* **72**, 217–244 (2021).
16. Yang, Z.H. et al. Cryptochromes orchestrate transcription regulation of diverse blue light responses in plants. *Photochem. Photobiol.* **93**, 112–127 (2017).
17. Sontheimer, E.J. & Steitz, J.A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
18. Liu, B. et al. snoRNA-facilitated protein secretion revealed by transcriptome-wide snoRNA target identification. *Cell* **188**, 465–483.e22 (2025).
19. Moll, C. et al. *CLO/GFA1* and *ATO* are novel regulators of gametic cell fate in plants. *Plant J.* **56**, 913–921 (2008).
20. Deng, X. et al. Recruitment of the NineTeen complex to the activated spliceosome requires AtPRMT5. *Proc. Natl. Acad. Sci. USA* **113**, 5447–5452 (2016).
21. Brenner, T.J. & Guthrie, C. Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. *RNA* **12**, 862–871 (2006).
22. Guo, Y., Xiong, L.M., Ishitani, M. & Zhu, J.K. An *Arabidopsis* mutation in translation elongation factor 2 causes superinduction of *CBF/DREB1* transcription factor genes but blocks the induction of their downstream targets under low temperatures. *Proc. Natl. Acad. Sci. USA* **99**, 7786–7791 (2002).
23. Milicevic, N., Jenner, L., Myasnikov, A., Yusupov, M. & Yusupova, G. mRNA reading frame maintenance during eukaryotic ribosome translocation. *Nature* **625**, 393–400 (2024).
24. Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S. & Lührmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092–4106 (1997).
25. Paik, I., Yang, S. & Choi, G. Phytochrome regulates translation of mRNA in the cytosol. *Proc. Natl. Acad. Sci. USA* **109**, 1335–1340 (2012).
26. Chen, G.H., Liu, M.J., Xiong, Y., Sheen, J. & Wu, S.H. TOR and RPS6 transmit light signals to enhance protein translation in deetiolating *Arabidopsis* seedlings. *Proc. Natl. Acad. Sci. USA* **115**, 12823–12828 (2018).
27. Kurihara, Y. et al. Transcripts from downstream alternative transcription start sites evade uORF-mediated inhibition of gene expression in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **115**, 7831–7836 (2018).
28. Jang, G.J., Yang, J.Y., Hsieh, H.L. & Wu, S.H. Processing bodies control the selective translation for optimal development of *Arabidopsis* young seedlings. *Proc. Natl. Acad. Sci. USA* **116**, 6451–6456 (2019).
29. Nemeth, K., Bayraktar, R., Ferracin, M. & Calin, G.A. Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat. Rev. Genet.* **25**, 211–232 (2024).
30. Dai, P. et al. A translation-activating function of MIWI/piRNA during mouse spermiogenesis. *Cell* **179**, 1566–1581.e16 (2019).
31. Meijer, H.A. et al. Translational repression and eIF4A2 activity are critical for microRNA-mediated gene regulation. *Science* **340**, 82–85 (2013).
32. Bukhari, S.I.A. et al. A specialized mechanism of translation mediated by FXR1a-associated MicroRNP in cellular quiescence. *Mol. Cell* **61**, 760–773 (2016).
33. Muthukumar, S., Li, C.T., Liu, R.J. & Bellodi, C. Roles and regulation of tRNA-derived small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **25**, 359–378 (2024).
34. Martin, W. & Koonin, E.V. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
35. Vogel, J. & Luisi, B.F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
36. Scofield, D.G. & Lynch, M. Evolutionary diversification of the Sm family of RNA-associated proteins. *Mol. Biol. Evol.* **25**, 2255–2267 (2008).
37. Kagale, S. et al. Polyploid evolution of the Brassicaceae during the Cenozoic era. *Plant Cell* **26**, 2777–2791 (2014).
38. Yu, H.D. et al. Downregulation of chloroplast RPS1 negatively modulates nuclear heat-responsive expression of *HsfA2* and its target genes in *Arabidopsis*. *PLoS Genet.* **8**, e1002669 (2012).
39. Zhou, K.N. et al. *Albino seedling lethality 4*; Chloroplast 30S ribosomal protein S1 is required for chloroplast ribosome biogenesis and early chloroplast development in rice. *Rice (N Y)* **14**, 47 (2021).
40. Robles, P. & Quesada, V. Unveiling the functions of plastid ribosomal proteins in plant development and abiotic stress tolerance. *Plant Physiol. Biochem.* **189**, 35–45 (2022).
41. Wang, F.F., Lian, H.L., Kang, C.Y. & Yang, H.Q. Phytochrome B is involved in mediating red light-induced stomatal opening in *Arabidopsis thaliana*. *Mol. Plant* **3**, 246–259 (2010).
42. Tian, C.H. et al. A gene expression map of shoot domains reveals regulatory mechanisms. *Nat. Commun.* **10**, 141 (2019).
43. Yan, L.H. et al. High-efficiency genome editing in *Arabidopsis* using *YAO* promoter-driven CRISPR/Cas9 system. *Mol. Plant* **8**, 1820–1823 (2015).
44. Li, W.Y. et al. EIN2-directed translational regulation of ethylene signaling in *Arabidopsis*. *Cell* **163**, 670–683 (2015).
45. Li, J. et al. Modulation of BIN2 kinase activity by HY5 controls hypocotyl elongation in the light. *Nat. Commun.* **11**, 1592 (2020).
46. Zhu, P. et al. *Arabidopsis* small nucleolar RNA monitors the efficient pre-rRNA processing during ribosome biogenesis. *Proc. Natl. Acad. Sci. USA* **113**, 11967–11972 (2016).
47. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
48. Chen, S.F., Zhou, Y.Q., Chen, Y.R. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
49. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
50. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
51. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. & Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667 (2016).
52. Cheng, C.Y. et al. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804 (2017).
53. Kang, Y.J. et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **45**, W12–W16 (2017).
54. Hu, L., Xu, Z.Y., Hu, B.Q. & Lu, Z.J. COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res.* **45**, e2 (2017).
55. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
56. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
57. Gu, Z.G., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
58. Wang, Y.Q. et al. Genomic features and regulatory roles of intermediate-sized non-coding RNAs in *Arabidopsis*. *Mol. Plant* **7**, 514–527 (2014).
59. Shen, S.H. et al. rMATS: robust and flexible detection of differential alternative

splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **111**, E5593–E5601 (2014).
60. Yu, G.C., Wang, L.G., Han, Y.Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
61. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
62. Hong, H.L. et al. Comparative proteome and *cis*-regulatory element analysis reveals specific molecular pathways conserved in dog and human brains. *Mol. Cell. Proteomics* **21**, 100261 (2022).
63. Wiśniewski, J.R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
64. Hsu, P.Y. et al. Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **113**, E7126–E7135 (2016).
65. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
66. Choudhary, S., Li, W.Z. & Smith, A.D. Accurate detection of short and long active ORFs using Ribo-seq data. *Bioinformatics* **36**, 2053–2059 (2020).
67. Lauria, F. et al. riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* **14**, e1006169 (2018).
68. Zhang, M.J. et al. Classification and clustering of RNA crosslink-ligation data reveal complex structures and homodimers. *Genome Res.* **32**, 968–985 (2022).
69. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
70. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
71. Lorenz, R. et al. ViennaRNA package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
72. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204–2207 (2010).
73. Xu, W.Z. et al. CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC Bioinformatics* **22**, 489 (2021).
74. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
75. Jiang, A.L. et al. The PIF1-miR408-PLANTACYANIN repression cascade regulates light-dependent seed germination. *Plant Cell* **33**, 1506–1529 (2021).
76. Zhou, Y. et al. Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in *Arabidopsis*. *Nat. Genet.* **50**, 638–644 (2018).
77. Wang, Y.Q., Li, J., Deng, X.W. & Zhu, D.M. *Arabidopsis* noncoding RNA modulates seedling greening during deetiolation. *Sci. China Life Sci.* **61**, 199–203 (2018).
78. Baker, N.R. Chlorophyll fluorescence: a probe of photosynthesis in vivo. *Annu. Rev. Plant Biol.* **59**, 89–113 (2008).
79. Glenn, W.S. et al. Bioorthogonal Noncanonical Amino Acid Tagging (BONCAT) enables time-resolved analysis of protein synthesis in native plant tissue. *Plant Physiol.* **173**, 1543–1553 (2017).
80. Sparkes, I.A., Runions, J., Kearns, A. & Hawes, C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* **1**, 2019–2025 (2006).
81. Niu, Y.J. & Sheen, J. Transient expression assays for quantifying signaling output. In: Wang, Z.Y. & Yang, Z.B. eds. Plant signalling networks: methods and protocols. Totowa: Humana, 195–206 (2012).
82. Yin, R., Feng, B.Y., Varshney, A. & Pierce, B.G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **31**, e4379 (2022).
83. Chuang, G.Y., Kozakov, D., Brenke, R., Comeau, S.R. & Vajda, S. DARS (Decoys As the Reference State) potentials for protein-protein docking. *Biophys. J.* **95**, 4217–4227 (2008).
84. Bowers, K.J. et al. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE conference on supercomputing. Tampa: ACM, 84–es (2006).
85. Jorgensen, W.L. & Tirado-Rives, J. The OPLS potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).
86. Altschul, S.F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
87. Goodstein, D.M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).
88. Löytynoja, A. Phylogeny-aware alignment with PRANK. In: Russell, D.J. ed. Multiple sequence alignment methods. Totowa: Humana, 155–170 (2014).
89. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
90. Zhang, J.Z., Nielsen, R. & Yang, Z.H. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479 (2005).
91. Emenecker, R.J., Griffith, D. & Holehouse, A.S. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021).
92. Lamesch, P. et al. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
93. Hu, T.T. et al. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43**, 476–481 (2011).
94. Yang, R.L. et al. The reference genome of the halophytic plant *Eutrema salsugineum*. *Front. Plant Sci.* **4**, 46 (2013).
95. Motamayor, J.C. et al. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* **14**, r53 (2013).
96. Tang, H.B. et al. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* **3**, e211 (2024).
97. McGuffin, L.J., Bryson, K. & Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
98. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
99. Lin, Z.M. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.15302/vita.2026.10.0081.

**Correspondence** and requests for materials should be addressed to Danmeng Zhu.

**Reprints and permission information** is available at https://www.vita-journal.com/.

© The Author(s) 2026. Published by Higher Education Press. This is an Open Access article distributed under the terms of the CC BY license (https://creativecommons.org/licenses/by/4.0/).