




HOT3/eIF5B1 confers Kozak motif-dependent translational control of photosynthesis-associated nuclear genes for chloroplast biogenesis

Received: 14 February 2024

Accepted: 31 October 2024

Published online: 14 November 2024

Runlai Hang ¹, Hao Li¹, Wenjing Liu¹, Runyu Wang¹, Hao Hu ¹, Meng Chen ²,
Chenjiang You ^{3,4} & Xuemei Chen ¹ ✉

Photosynthesis requires chloroplasts, in which most proteins are nucleus-encoded and produced via cytoplasmic translation. The translation initiation factor eIF5B gates the transition from initiation (I) to elongation (E), and the Kozak motif is associated with translation efficiency, but their relationship is previously unknown. Here, with ribosome profiling, we determined the genome-wide I-E transition efficiencies. We discovered that the most prevalent Kozak motif is associated with high I-E transition efficiency in Arabidopsis, rice, and wheat, thus implicating the potential of the Kozak motif in facilitating the I-E transition. Indeed, the effects of Kozak motifs in promoting translation depend on HOT3/eIF5B1 in Arabidopsis. HOT3 preferentially promotes the translation of photosynthesis-associated nuclear genes in a Kozak motif-dependent manner, which explains the chloroplast defects and reduced photosynthesis activity of *hot3* mutants. Our study linked the Kozak motif to eIF5B-mediated I-E transition during translation and uncovered the function of HOT3 in the cytoplasmic translational control of chloroplast biogenesis and photosynthesis.

Cytoplasmic translation of messenger RNA (mRNA) is a process under immense regulation for cellular homeostasis and reprogramming in all eukaryotes¹. Translation comprises initiation, elongation, termination, and ribosome recycling steps, among which initiation is rate-limiting and coordinated by both *cis*-elements and *trans*-factors^{2–7}. *Cis* regulatory elements include sequence motifs⁸, RNA modifications⁹, and structural features¹⁰. In eukaryotes, the initiator AUG triplet codon is recognized by the methionyl-initiator tRNA (Met-tRNA_i^{Met}) in the ribosome and this is promoted by the Kozak consensus sequence around the AUG, which features a purine at position -3 and a guanine at

position +4 (the A of AUG is designated as +1) in animals¹¹. Although consensus nucleotides at the -3 and +4 positions were also observed in plants^{12–16}, their nucleotide identities vary between different plant species^{4,8,17–20}. Optimal Kozak consensus sequences are thought to confer higher translational efficiency^{8,20,21}, however, the underpinning mechanism has not been well examined in any organism. Moreover, a systematic evaluation of translation initiation efficiencies mediated by various Kozak motifs has been lacking in any organism, hampering our understanding of evolutionarily conserved regulatory mechanisms of translation via *cis* regulatory elements. A molecular understanding of

¹State Key Laboratory for Protein and Plant Gene Research, Peking-Tsinghua Joint Center for Life Sciences, Beijing Advanced Center of RNA Biology (BEACON), School of Life Sciences, Peking University, 100871 Beijing, China. ²Department of Botany and Plant Sciences, Institute of Integrative Genome Biology, University of California, Riverside, Riverside, CA 92521, USA. ³College of Life Sciences, Guangdong Provincial Key Laboratory for the Development Biology and Environmental Adaptation of Agricultural Organisms, South China Agricultural University, Guangzhou, Guangdong 510642, China. ⁴Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou, Guangdong 510642, China. ✉e-mail: hangr1200@163.com; cjyou@scau.edu.cn; xuemei.chen@pku.edu.cn

Kozak motif-mediated translational control in plants will ultimately benefit the development of new technologies for sustaining crop yield in a changing climate.

A series of eukaryotic initiation factors (eIFs) works synergistically with *cis* regulatory elements to enable translation initiation^{3,6}. In cap-dependent mRNA translation initiation, the eIF4F complex first recognizes the 5' 7-methylguanosine 5'-5' triphosphate cap. Then the 43S pre-initiation complex, which comprises the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, Met-tRNA_i^{Met} and eIF2-GTP, is recruited by eIF4F to form the 48S pre-initiation complex. The 48S pre-initiation complex scans the mRNA 5' untranslated region (5' UTR) in the 5'-to-3' direction. When the AUG initiation codon is recognized, Met-tRNA_i^{Met} is secured in the Peptidyl-site (P-site) and eIF1 is then displaced, triggering the eIF5-mediated hydrolysis of eIF2-bound GTP and the release of phosphate and eIF2-GDP³. Subsequently, as a multifunctional initiation factor, eIF5B stabilizes initiating ribosomes to maintain reading frame fidelity during start codon recognition²², and to mediate the transition from translation initiation to elongation²³. eIF5B-GTP first binds to the 48S pre-initiation complex and recruits the 60S large ribosomal subunit, along with the release of most eIFs. Finally, the hydrolysis of the eIF5B-bound GTP facilitates the release of eIF5B-GDP and eIF1A to yield the 80S ribosome competent for translation elongation^{24,25}. IF2/eIF5B is one of the two universally conserved initiation factors across Bacteria, Archaea, and Eukarya^{25,26}. Knocking out eIF5B led to severe developmental defects in *Saccharomyces cerevisiae*²⁷ and *Drosophila melanogaster*²⁸. Plants have similar eIFs but evolved multiple paralogs for each eIF^{2,29}, which may represent functional redundancy or diversification. In Arabidopsis, HOT TEMPERATURE 3 (HOT3)/eIF5B1 (AT1G76810), is one of a four-member eIF5B family. Loss-of-function *hot3* mutants such as *hot3-2* and *hot3-3* show pleiotropic phenotypes such as pale-green leaves, dwarfism, and defective acclimation to heat stress, while mutants in the other three genes have no developmentally abnormal phenotypes³⁰. HOT3 preferentially affects a subset of genes. Polysome association was reduced for ~1000 transcripts and enhanced for a similar number of other transcripts in *hot3* mutants³⁰. Recently, we reported a new role of HOT3 in 18S rRNA 3' end maturation to suppress the biogenesis of rRNA-derived small interfering RNAs (risiRNAs)³¹. However, the function of HOT3 in translation initiation is independent of its role in regulating risiRNAs³¹. The mechanism by which HOT3 preferentially affects a subset of transcripts remains elusive.

Chloroplasts are photosynthetically active plastids essential for the photoautotrophic growth in plants. As such, chloroplast biogenesis and maintenance are pivotal to cellular reprogramming and homeostasis during plant development. Chloroplasts are either derived from proplastids in undifferentiated stem cells during the development of aerial organs, such as leaves and flowers, or converted from an intermediate type such as etioplasts during seedling de-etiolation³². Plastids are semi-autonomous organelles as over 90% of plastid-localized proteins are encoded by the nuclear genome³³. The biogenesis of photosynthetically active chloroplasts requires intricate control of photosynthesis-associated nuclear genes (PhANGs) for producing the photosynthetic machinery in chloroplasts^{34,35}. Extensive studies have demonstrated that the transcriptional regulation of PhANGs exerts a pivotal role in chloroplast biogenesis^{36,37}. Accumulating evidence suggests that chloroplast biogenesis also entails dynamic control of PhANGs at the translational level. During de-etiolation, when seedlings emerge from the soil and encounter light for the first time, light promotes the translation of PhANGs to initiate the conversion from etioplasts to chloroplasts^{12,38}. The light-inducible translation of PhANGs is mediated by the photoreceptor phytochrome A through derepressing CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1)-mediated inhibition of the activity of TARGET OF RAPAMYCIN (TOR) and ribosomal protein S6 (RPS6)³⁹. Translation of PhANGs is also dynamically regulated under diurnal day-night cycles and can be

rapidly repressed in response to a darkness treatment⁴⁰. The plant-specific isoforms of the eIF4G subunits, eIFiso4G1 and eIFiso4G2, were shown to be required for the translation of PhANGs^{41,42}. A glycine-rich RNA-binding protein named SIRBP1 from tomato interacts with the eukaryotic translation initiation factor SleIF4A2 to promote the translation of PhANG transcripts⁴³. Despite these advances, the mechanism underlying the selective translational control of PhANGs as a group remains elusive.

Here, by comparing the ribosome-occupancy profiles between wild-type and *hot3-2* plants during vegetative seedling establishment and reproductive inflorescence development – two developmental stages involving active chloroplast biogenesis – we uncovered the function of HOT3 in promoting the transition from translation initiation (I) to elongation (E) at the genome-wide scale. Loss-of-function of *HOT3* caused prevalent ribosomal stalling at the translation initiation site (TIS), suggesting a defect in the transition from initiation to elongation (I-E). Using ribosome footprints at the TIS vs. those in the gene body, we defined the relative Pausing at the Start codon (rPS) index that quantitatively assesses the I-E transition efficiency. Leveraging the rPS index, we conducted a comprehensive analysis of the I-E transition efficiencies of 64 simplified Kozak motifs defined by the -3/+4/+5 nucleotides surrounding the TIS. Our findings indicate that the most prevalent Kozak motifs exhibit high I-E transition efficiency with a low rPS index value, a phenomenon that is conserved across *Arabidopsis thaliana*, *Oryza sativa* (rice) and *Triticum aestivum* (wheat). Notably, transcripts that require HOT3 for efficient I-E transition and translation possess the optimal A/GC Kozak consensus and are enriched in PhANGs, particularly those encoding proteins of the photosynthetic apparatus on the thylakoid membrane. Dysfunctional HOT3 caused a dramatic defect in the I-E transition, decreased translational efficiency (TE), and reduced protein abundance for these genes. This, in turn, results in chloroplast defects represented by reduced thylakoid layers and abnormal photosynthesis. Collectively, our study revealed a genetic link between the Kozak motif and eIF5B-mediated transition from translation initiation to elongation and uncovered the mechanism whereby HOT3/eIF5B1-mediated cytoplasmic translation initiation enables chloroplast biogenesis and optimized photosynthesis.

Results

Establishing the rPS index to quantify I-E transition efficiencies

Loss-of-function of *HOT3* resulted in chlorosis in both vegetative seedling development and reproductive inflorescence development (Supplementary Fig. 1a). Consistent with these observations, the amount of chlorophyll (Chl) a and Chl b, and total chlorophyll content (Chl a + Chl b) were significantly reduced in *hot3-2* seedlings and inflorescences (Supplementary Fig. 1b). These phenotypes suggested a necessary role of HOT3-mediated cytoplasmic translation initiation in chloroplast biogenesis. As HOT3 is a conserved cytoplasmic translation initiation factor and its loss of function results in global ribosomal stalling at TISs in inflorescences³¹, we hypothesized that HOT3 plays a pivotal role in the cytoplasmic translation of nuclear genes encoding proteins targeted to chloroplasts including PhANGs.

To evaluate whether HOT3 promotes the translation of transcripts encoding chloroplast proteins, we performed ribosome profiling (Ribo-seq) and transcriptome profiling (RNA-seq) using Col-0 (WT) and *hot3-2*. We had previously performed Ribo-seq using inflorescences³¹. Here we optimized the library preparation method to enhance the depth of Ribo-seq and included samples of both inflorescences and seedlings—two developmental stages involving active chloroplast biogenesis (Supplementary Dataset 1)^{44,45}. Independent replicates showed high reproducibility for both the Ribo-seq and RNA-seq samples (Supplementary Fig. 2a, b). The ribosome protected fragments (RPFs) in all libraries peaked at 28 nucleotides (nt) in length for inflorescences and 28 nt or 29 nt for seedlings (Supplementary

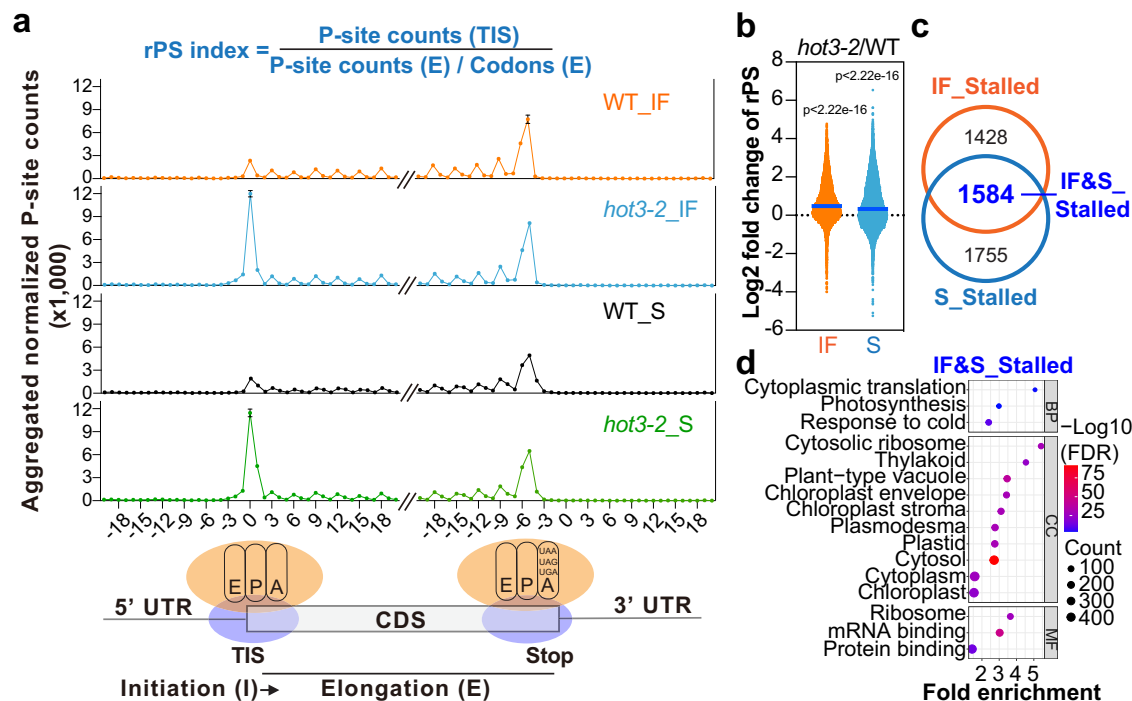


Fig. 1 | Ribosome profiling revealed defects in the transition from translation initiation to elongation in *hot3-2*. **a** A metagene analysis was conducted to determine normalized P-site counts that were mapped to regions close to the translation initiation site (TIS) and stop codon of coding sequence (CDS). A 41-nt region located near the TIS and stop codon is displayed. A diagram representing the A-site (entry point for aminoacyl-tRNA), P-site (where peptide bond formation occurs), and E-site (exit site for uncharged tRNA) within ribosomes is provided below. The P-site counts for the initiating ribosome (P-site position on TIS) and terminating ribosome (A-site position on the stop codon) are shown. Three biological replicates were performed to generate the average value and standard deviation shown for each nucleotide. The relative Pausing at Start codon of CDS (rPS) represents the inability in translational I-E transition by calculating the ratio of P-site counts mapped to the TIS vs. the average P-site counts on each codon within the E region (CDS downstream of the TIS). Transcripts with at least 10 P-site counts mapped to the E region were included for further analysis. **b** A scatter plot showing

the global increase in rPS values in inflorescence (IF) and seedling (S) tissues of *hot3-2* mutants. A total of 21,425 and 20,805 transcripts with valid *p* values were utilized to calculate the fold change in rPS in *hot3-2* mutants as compared to wild type. The median log2-transformed fold-change in rPS values was 0.48 and 0.33 for inflorescence and seedling samples, respectively. The Wilcoxon signed-rank test was employed to assess significance, resulting in two-tailed *p* value presented. **c** A Venn diagram showing the overlap between transcripts in IF_Stalled and S_Stalled, which were defined as those with significantly increased rPS values in *hot3-2* vs. WT in inflorescences and seedlings, respectively. The 1584 transcripts in common are designated as IF&S_Stalled. **d** A Gene Ontology (GO) term enrichment plot for IF&S_Stalled. The terms “BP”, “CC”, and “MF” stand for biological process, cellular component, and molecular function, respectively. The GO term analysis was conducted using the Database for Annotation, Visualization, and Integrated Discovery (DAVID)^{76,77}. The enrichment plot was generated at <https://www.bioinformatics.com.cn>, an online platform for data analysis and visualization.

Fig. 2c), consistent with the canonical RPF length observed in plants^{13,44}. Most importantly, the RPFs displayed a strong 3 nt periodicity (Supplementary Fig. 2d), indicative of high-quality Ribo-seq libraries.

To identify and quantify the codons being translated, we subjected our datasets to RiboTaper analysis⁴⁶, which defines translated ORFs based on the 3 nt periodicity of P-site signals. We mapped the P-site signals to the Arabidopsis genome Araport11⁴⁷ and identified a total of 27,232 ORFs, 98% of which were annotated ORFs (Supplementary Fig. 2e; Supplementary Dataset 2). As expected, most P-site nucleotides were in the coding regions in both WT and *hot3-2* (Supplementary Fig. 2f). A higher proportion of P-site nucleotides was found at TISs in both inflorescences and seedlings of the *hot3-2* mutant as compared to WT (Supplementary Fig. 2g). This pattern of the gene-body distribution of P-sites suggested that loss of function of *HOT3* triggered genome-wide ribosome stalling at TISs, corroborating our previous conclusion³¹ and the function of eIF5B in translation initiation defined by biochemical and genome-wide studies in yeast^{23,48} and human cells²², respectively.

Aggregated normalized P-site signals increased particularly at the TIS but no other codons from the gene body in both inflorescences and seedlings of *hot3-2* (Fig. 1a). To quantify the translation stalling defect at the genome-wide level and to identify the most-affected

genes in *hot3-2*, we defined the rPS index to reflect the inability in translational I-E transition by calculating the ratio of I/E, in which “I” represents P-site counts mapped to the TIS of each gene, while “E” represents the averaged P-site counts on each codon of the ORF excluding the TIS (Fig. 1a). We found that rPS values were globally and significantly increased in *hot3-2* compared with those in WT; the median log2-transformed fold change in the rPS values of *hot3-2* vs. WT were 0.48 and 0.33 for inflorescences and seedlings, respectively (Fig. 1b; Supplementary Dataset 3). Using *p* value less than 0.05 and log2-transformed fold change (*hot3-2* vs. WT) of rPS values greater than 1 as the thresholds, we identified 3012 translationally stalled transcripts in inflorescences (referred to as IF_Stalled) and 3339 in seedlings (referred to as S_Stalled) in *hot3-2* (Fig. 1c and Supplementary Fig. 3a). Comparing IF_Stalled and S_Stalled yielded 1584 common *HOT3*-dependent transcripts, which are hereafter called IF&S_Stalled (Fig. 1c). Interestingly, IF&S_Stalled genes were highly enriched in the cellular functions of cytoplasmic translation and photosynthesis (Fig. 1d).

The predominant Kozak motif confers a high efficiency in the I-E transition

Because the sequence context surrounding the TIS directly impacts translation^{21,49–51}, we examined whether the translationally stalled

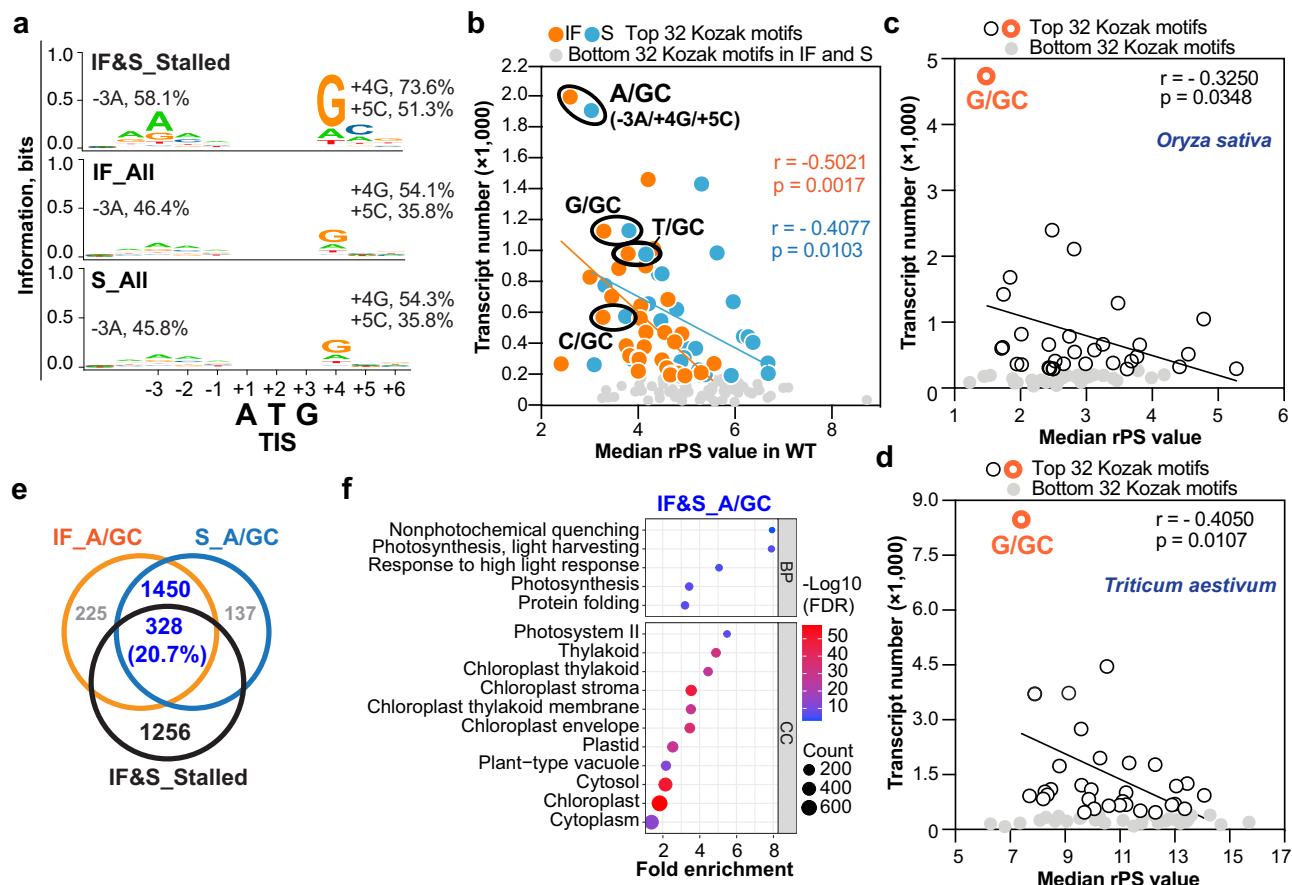


Fig. 2 | Transcripts with the predominant Kozak motif possess high efficiency in the transition from translation initiation to elongation. **a** Seqlogo analysis of Kozak consensus sequences for nucleotides in the vicinity of the translation initiation site (TIS) in transcripts from three groups: IF&S_Stalled, IF_All, and S_All. The nucleotide frequencies of -3A, +4G, and +5C are displayed for each group. The values of bits indicate the information of each nucleotide at each position. The TIS codon ATG is indicated below. Seqlogo analysis was conducted using TBtools⁷⁸. **b** Scatter plot showing the inverse correlation between median rPS values and transcript numbers for each of the simplified Kozak motifs (-3/+4/+5) in inflorescences (IF) and seedlings (S) in WT. The 32 Kozak motifs (orange dots for IF and cyan dots for S) with transcript numbers above the median value were used for the correlation analysis. Gray dots represent the bottom 32 Kozak motifs in IF and S. The 32 Kozak motifs including A/GC, G/GC, T/GC, and C/GC are encircled in black ovals. Scatter plots to show the anti-correlation between the median rPS index and transcript number for simplified Kozak motifs in *Oryza sativa* Japonica (**c**) and *Triticum aestivum* (**d**). The 32 Kozak motifs (black and orange circles) with

transcript numbers above the median value were used for the correlation analysis. Gray dots represent the bottom 32 Kozak motifs. The orange-colored G/GC Kozak motif is the most prevalent one in *Oryza sativa* Japonica (**c**) and *Triticum aestivum* (**d**). **e** A Venn diagram to show that 20.7% of the transcripts in the IF&S_Stalled group utilize the A/GC-type Kozak motif. IF_A/GC and S_A/GC represent 2003 and 1915 transcripts, respectively, that utilize the A/GC-type Kozak motif, detected in inflorescences (IF) and seedlings (S). **f** Gene ontology (GO) term enrichment plot for 1778 transcripts that utilize the A/GC-type Kozak motif (termed IF&S_A/GC), which are shared by both the IF_A/GC and S_A/GC groups. The Pearson correlation coefficient r and one-tailed p value were generated by the Prism version 8.4.0 software and displayed in (b–d). For (b), the correlation coefficient r and p value are displayed in orange and blue colors to represent Kozak motifs in inflorescences (IF) and seedlings (S), respectively. GO term analysis was conducted using the DAVID database^{76,77}. Enrichment plots were generated at <https://www.bioinformatics.com.cn>, an online platform for data analysis and visualization.

transcripts in *hot3-2* possess any biases in their Kozak consensus motifs. With the A of the initiator AUG as the +1 nucleotide, we found that the predominant nucleotides at each of the conserved Kozak motif positions in the annotated transcripts in Arabidopsis are -3A, +4G, and +5C (Supplementary Fig. 3b). The frequencies of -3A, +4G, and +5C in all translating inflorescence transcripts (referred to as IF_All) were 46.3%, 54.1%, and 35.8%, respectively. Similarly, the frequencies of -3A, +4G, and +5C in all translating seedling transcripts (referred to as S_All) were 45.8%, 54.3%, and 35.8%, respectively (Fig. 2a; Supplementary Dataset 3). Intriguingly, the occurrence frequencies of these prevalent nucleotides increased significantly in the Kozak motifs of IF&S_Stalled transcripts in *hot3-2*; the frequency of -3A, +4G, and +5C in these genes increased to 58.1%, 73.6%, and 51.3% (Fig. 2a), respectively.

The findings of the prevalent Kozak motif in Arabidopsis and its enrichment in HOT3-dependent transcripts prompted us to evaluate the I-E transition efficiencies associated with individual Kozak motifs.

The relationship between the endogenous Kozak motif and the I-E transition efficiency has not been systemically evaluated at the genome-wide level in any species previously⁸. To that end, all translating transcripts from inflorescences and seedlings were divided into 64 groups based on their simplified Kozak motifs (nucleotides at positions of -3/+4/+5). Then, the detected transcript number of each Kozak motif was plotted against the median rPS index value in WT (Fig. 2b). Interestingly, we found that the transcript number of individual Kozak motifs negatively correlated with the median rPS value in WT (Fig. 2b), indicating that a large number of transcripts contained Kozak motifs associated with efficient I-E transition. The most prevalent Kozak motif was the A/GC motif, which exhibited the highest I-E transition efficiency in both inflorescences and seedlings. Kozak motifs G/GC, T/GC and C/GC also exhibited high I-E transition efficiencies with relatively low median rPS values in WT, but transcripts containing these Kozak motifs were relatively less (Fig. 2b).

To test whether A/GC is also the most abundant Kozak motif in other plant species and whether the most abundant Kozak motif is always associated with highly efficient I-E transition, we re-analyzed publicly available Ribo-seq datasets on rice^{52,53} and wheat⁵⁴ (Supplementary Dataset 1). Interestingly, unlike the dicot Arabidopsis, the predominant Kozak sequence context in monocots rice and wheat was the simplified G/GC-type motif (Fig. 2c, d, and Supplementary Fig. 3b), which was consistent with earlier *in silico* studies of Kozak sequence context in monocots^{17,20}. In Arabidopsis, the translating transcripts containing the A/GC-type Kozak motif in inflorescences (referred to as IF_A/GC) and seedlings (referred to as S_A/GC) largely overlapped (Fig. 2e), with 1778 transcripts (referred to as IF&S_A/GC) shared by both groups and enriched in genes associated with “photosynthesis” (Fig. 2f). Likewise, the predominant G/GC-Kozak motif-containing transcripts in rice and wheat were also enriched in photosynthesis-associated genes (Supplementary Fig. 3c, d). Interestingly, homologous transcripts of Arabidopsis IF&S_A/GC in rice and wheat harbored G/GC as the predominant Kozak motif (Supplementary Fig. 3e, f). It suggested that the preferred Kozak motifs can vary among plant species^{18–20}, possibly due to higher GC contents in monocot genomes and 5' UTRs^{4,17}. Meanwhile, like Arabidopsis, the most prevalent G/GC-type Kozak motif in rice and wheat was also associated with a high I-E transition efficiency (Fig. 2c, d; Supplementary Fig. 3g, h), suggesting that the association of the predominant Kozak motif with a high translation initiation efficiency is likely a general principle. Moreover, there was an overall tendency for transcripts to have efficient I-E transition, as transcript numbers of individual Kozak motifs exhibited significant anti-correlation with their median rPS values (Fig. 2c, d).

To interpretate the anti-correlation between rPS value and the number of transcripts from regulatory or biological perspectives, we reviewed the top Kozak motifs based on their transcript numbers (Fig. 2b–d). Unexpectedly, the A/GA-type Kozak motif was in the second place in Arabidopsis (Supplementary Fig. 4a). Likewise, the G/GA-type Kozak motif was in the second place in rice and wheat (Supplementary Fig. 4b). Comparing with the N/GA-type (A/GA, G/GA, T/GA, and C/GA) Kozak motifs, the corresponding N/GC-type (A/GC, G/GC, T/GC, and C/GC) Kozak motifs had lower rPS values and larger transcript numbers in all three species (Supplementary Fig. 4a–c). Notably, the anti-correlation between median rPS values and transcript numbers was also true for the N/GC- and N/GA-type Kozak motifs, as a common feature shared by Arabidopsis, rice, and wheat (Supplementary Fig. 4a, b). Different from the N/GC-type Kozak motif-containing transcripts enriched in photosynthesis-related genes (Supplementary Fig. 4d–f), genes associated with DNA transcription in the nucleus were enriched in the N/GA group in all three species (Supplementary Fig. 4g–i). Taken together, the “high number” of genes with low I-E transition indicates that (1) during evolution more transcripts prefer to use Kozak sequence contexts with less ribosome pausing at the start codon and better I-E transition; (2) adopting the Kozak motif with the highest translational initiation efficiency is likely an evolutionarily-conserved feature for photosynthesis-associated genes – despite variations of the predominant Kozak consensus sequence among plant species, whereas genes in DNA transcription tend to harbor the N/GA-type Kozak motif with less efficient I-E transition. The Kozak motif types probably contribute to protein abundances *in vivo*, such that the highly abundant photosynthesis-related proteins harbor the Kozak motif with the best translation efficiency.

Defective I-E transition in *hot3-2* lowers the translation efficiency of PhANGs encoding the plastid photosynthetic apparatus

Because as high as 20.7% of IF&S_A/GC transcripts were translationally stalled in *hot3* (Fig. 2e), this raised the possibility that the A/GC-type Kozak motif is preferentially controlled by HOT3. Indeed, we observed a Kozak-motif bias in HOT3-dependent I-E transition. This conclusion

was supported by the following evidence. Firstly, the median log₂-transformed fold change of rPS values (*hot3-2* vs. WT) for the 64 Kozak motifs exhibited a significantly positive correlation between inflorescences and seedlings (Fig. 3a), indicative of a general and non-tissue-specific association between HOT3 and its preferred Kozak motifs on the I-E transition. Secondly, HOT3 is required for efficient I-E transition of most Kozak motifs, especially the N/GC type including A/GC, C/GC, G/GC, and T/GC Kozak sequence contexts (Supplementary Fig. 4c). Lastly and notably, the A/GC-type motif, in particular among the 64 Kozak motifs, was the most affected in *hot3-2* with the median log₂-transformed rPS fold change (*hot3-2* vs. WT) greater than 1 (Fig. 3a). Thus, HOT3 exerts a particularly important role in conferring the high translation I-E transition efficiency of transcripts containing the A/GC Kozak motif. In agreement with this conclusion, the percentages of -3A, +4G, and +5C gradually increased as the fold change of rPS values (*hot3-2* vs. WT) increased in both inflorescences and seedlings (Fig. 3b).

We next investigated the connection between HOT3-mediated I-E transition and translation efficiency (TE). Conventionally, the TE of individual genes has been calculated as the total RPF read counts over the entire ORF divided by the transcript abundance from RNA-seq data⁵⁵. But when global ribosome stalling occurs at the TIS (Fig. 1b), including the RPF reads at the TIS would arbitrarily inflate the TE of the translationally stalled transcripts, especially in *hot3-2*. To avoid this issue, we calculated TE using the RPFs from the gene body excluding that of the TIS (Fig. 1a). Totally, 6208 and 9466 transcripts were found to have rPS and TE values suitable for comparative analyses (see Methods) between WT and *hot3-2* in inflorescence and seedling samples, respectively, and referred to as IF_TE_All and S_TE_All groups (Fig. 3c; Supplementary Dataset 3). We then assessed the relationship between the TE and I-E transition efficiency. At the global level, both inflorescences ($r = -0.3587$, $p < 0.0001$) and seedlings ($r = -0.2725$, $p < 0.0001$) exhibited significant anti-correlation in the log₂-transformed fold changes of rPS and TE values in *hot3-2* vs. WT (Fig. 3c; Supplementary Dataset S3), indicating a negative effect of ribosome stalling at TIS on translation efficiency. For IF_TE_All, there were 463 transcripts showing a significant upregulation of rPS values (adjusted p value < 0.05 and log₂-transformed fold change of rPS > 1) in *hot3-2* vs. WT. 359 transcripts, accounting for around 78% of these 463 transcripts, showed a significant decrease in TE values (adjusted p value < 0.05 and log₂-transformed fold change of TE < 0) in *hot3-2* inflorescences, and were designated as the IF_TE_Down group (Fig. 3c; Supplementary Dataset 3). Similarly, in the S_TE_All group of seedlings, 622 transcripts in *hot3-2* showed a significant increase in rPS values. Among these, 319 transcripts, accounting for approximately 51% of the 622 transcripts, exhibited a significant decrease in TE values and were labeled as S_TE_Down (Fig. 3c; Supplementary Dataset 3).

To further determine the relationship between the Kozak motif and HOT3-mediated translation I-E transition, we next examined the frequencies of -3A, +4G, and +5C in the translationally stalled transcripts in *hot3-2*. Compared to IF_TE_All and S_TE_All, transcripts in IF_TE_Down and S_TE_Down showed higher frequencies of -3A, +4G, and +5C surrounding the AUG start codon (Fig. 3d). Even higher nucleotide frequencies of -3A (87.7%), +4G (93.8%), and +5C (72.8%) were observed in the 81 common transcripts (referred to as IF&S_TE_Down) shared by IF_TE_Down and S_TE_Down groups (Fig. 3d, e). These 81 genes were highly enriched in chloroplast-related GO terms (Fig. 3f). To evaluate whether having Kozak motifs with higher frequencies of -3A, +4G, and +5C is a general principle for chloroplast biogenesis genes, we further examined transcripts encoding proteins targeted to plastids (Fig. 3d), which were referred to as All_Plastid based on the PPDB database⁵⁶. Surprisingly, All_Plastid transcripts exhibited significantly lower frequencies of -3A (46.4%), +4G (65.4%), or +5C (60.2%) compared to IF_TE_Down, S_TE_Down, or IF&S_TE_Down (Fig. 3d), indicating that not all genes encoding plastid proteins possess the optimal Kozak motif. In total, we found that 753 inflorescence

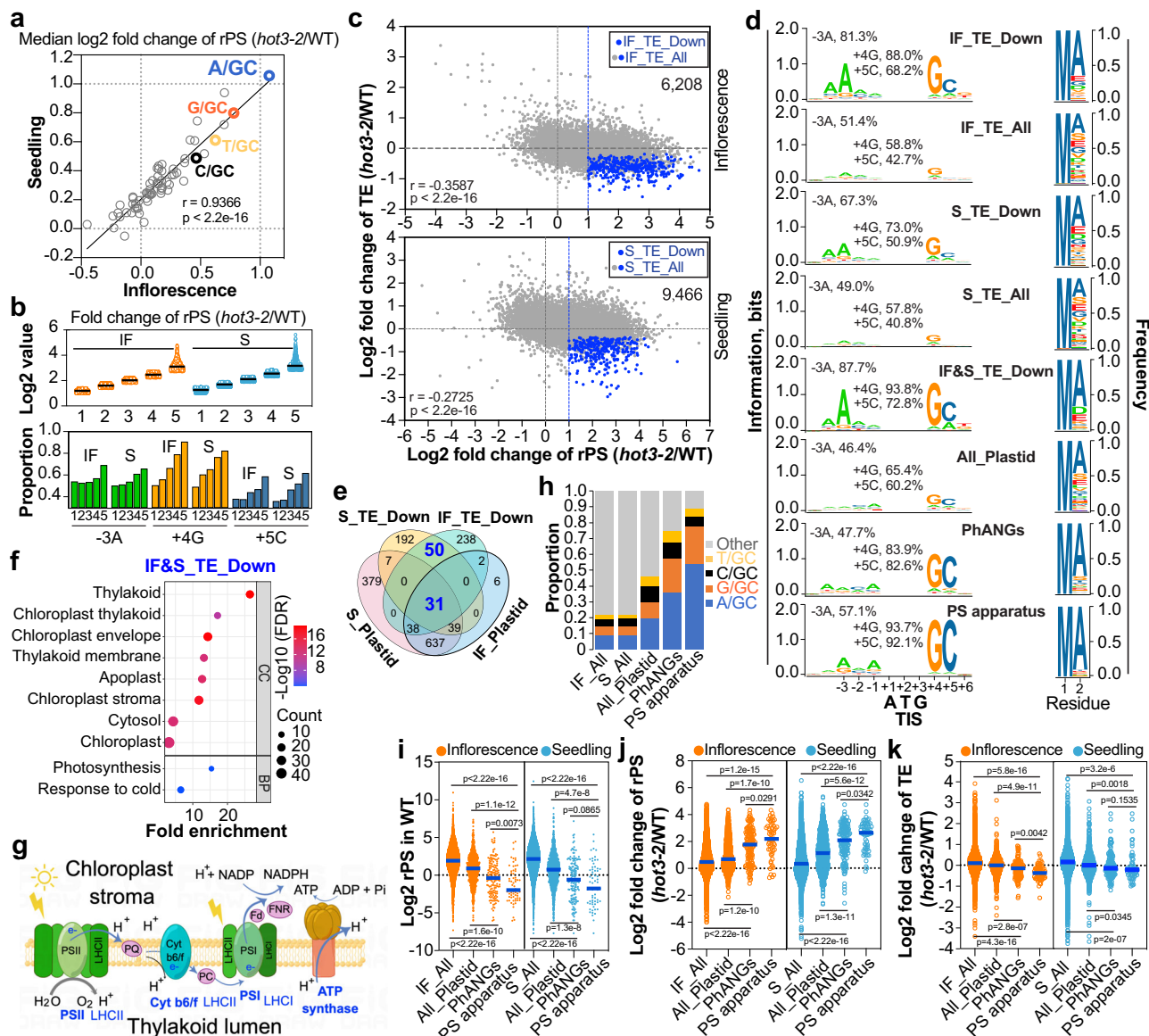


Fig. 3 | Defective I-E transition in *hot3-2* lowers the translational efficiency of transcripts with the A/GC-type Kozak motif. **a** Scatter plot to show the median log₂-transformed rPS fold change (*hot3-2* vs. WT) in inflorescences (x-axis) and seedlings (y-axis). Pearson correlation coefficient *r* and the two-tailed *p* value are presented. The N/GC-type (A/GC, G/GC, T/GC, and C/GC) Kozak motifs are depicted in different colors. The A/GC-type Kozak motif is the top one exhibiting a median log₂-transformed rPS fold change greater than 1 in both inflorescences and seedlings. **b** The correlation between log₂-transformed fold change (*hot3-2* vs. WT) of rPS and proportions of -3A, +4G, and +5C. Transcripts with a fold change of rPS values (*hot3-2* vs. WT) more than 2 were categorized into 5 groups based on their fold change. The frequencies of -3A, +4G, and +5C in these transcripts are shown for each group. IF, inflorescences; S, seedlings. **c** Scatter plots to illustrate the negative correlation between log₂-transformed rPS fold change (*hot3-2* vs. WT) and the log₂-transformed TE fold change (*hot3-2* vs. WT) in inflorescences (IF) and seedlings (S). The Pearson correlation coefficient *r* and two-tailed *p* value are shown. 6208 (inflorescences) and 9466 (seedlings) transcripts were included. Transcripts marked by IF_TE_Down in inflorescences and S_TE_Down in seedlings (highlighted in blue) exhibit significant downregulation in TE (*t* test, *p* < 0.05) and an rPS fold change greater than 2 in *hot3-2* vs. WT. **d** Seqlogo analysis of the Kozak consensus sequence for nucleotides near the translation initiation site (TIS) of transcripts and the first two amino acids at the N-termini of protein products in various categories: IF_TE_Down, IF_TE_All, S_TE_Down, S_TE_All, IF&S_TE_Down, All_Placid, PhANGs, and the photosynthetic (PS) apparatus. The frequencies of -3A, +4G, and +5C nucleotides are displayed in each group. Seqlogo analysis was performed using TBtools⁷⁸. **e** A Venn diagram illustrating the relationship among four categories of transcripts. S_Placid (seedlings) and IF_Placid (inflorescences) are nucleus-transcribed transcripts encoding proteins in the plastid, having valid values in log₂-transformed fold change in TE and rPS in our datasets. Among the 81 transcripts shared by IF_TE_Down and S_TE_Down, marked in blue and referred to as IF&S_TE_Down, 31 transcripts encode proteins in the plastid/chloroplast. **f** Gene ontology (GO) term enrichment plot for IF&S_TE_Down transcripts. **g** A diagram depicting the photosynthetic apparatus on the thylakoid membrane, which includes PSII, PSI, LHCI, LHCII, and ATP synthase. It was created by Figdraw (www.figdraw.com). Transcripts of PhANGs encoding the photosynthetic apparatus are denoted as “PS apparatus”. **h** Histogram showing the proportion of -3N/GC Kozak motifs, including A/GC, G/GC, T/GC, and C/GC types, in transcripts of IF_All, S_All, PhANGs, and PS apparatus in inflorescences (IF) and seedlings (S). Scatter plots displaying the log₂-transformed rPS values in WT (i), log₂-transformed fold change (*hot3-2*/WT) of rPS (j), and log₂-transformed fold change (*hot3-2*/WT) of TE (k) for transcripts from IF_All, S_All, All_Placid, PhANGs, and PS apparatus. Orange, inflorescences; Blue, seedlings. Statistical significance was determined using the Mann-Whitney test with two-tailed *p* values presented.

IF_TE_Down, IF_TE_All, S_TE_Down, S_TE_All, IF&S_TE_Down, All_Placid, PhANGs, and the photosynthetic (PS) apparatus. The frequencies of -3A, +4G, and +5C nucleotides are displayed in each group. Seqlogo analysis was performed using TBtools⁷⁸. **e** A Venn diagram illustrating the relationship among four categories of transcripts. S_Placid (seedlings) and IF_Placid (inflorescences) are nucleus-transcribed transcripts encoding proteins in the plastid, having valid values in log₂-transformed fold change in TE and rPS in our datasets. Among the 81 transcripts shared by IF_TE_Down and S_TE_Down, marked in blue and referred to as IF&S_TE_Down, 31 transcripts encode proteins in the plastid/chloroplast. **f** Gene ontology (GO) term enrichment plot for IF&S_TE_Down transcripts. **g** A diagram depicting the photosynthetic apparatus on the thylakoid membrane, which includes PSII, PSI, LHCI, LHCII, and ATP synthase. It was created by Figdraw (www.figdraw.com). Transcripts of PhANGs encoding the photosynthetic apparatus are denoted as “PS apparatus”. **h** Histogram showing the proportion of -3N/GC Kozak motifs, including A/GC, G/GC, T/GC, and C/GC types, in transcripts of IF_All, S_All, PhANGs, and PS apparatus in inflorescences (IF) and seedlings (S). Scatter plots displaying the log₂-transformed rPS values in WT (i), log₂-transformed fold change (*hot3-2*/WT) of rPS (j), and log₂-transformed fold change (*hot3-2*/WT) of TE (k) for transcripts from IF_All, S_All, All_Placid, PhANGs, and PS apparatus. Orange, inflorescences; Blue, seedlings. Statistical significance was determined using the Mann-Whitney test with two-tailed *p* values presented.

transcripts from IF_TE_All and 1131 seedling transcripts from S_TE_All belonged to All_Plastid and referred to them as IF_Plastid and S_Plastid, respectively (Fig. 3e). Around 20% (71 transcripts) of IF_TE_Down and 24% (77 transcripts) of S_TE_Down encode proteins targeted to plastids (Fig. 3e). Moreover, these 71 and 77 transcripts shared 31 common transcripts with the IF&S_TE_Down group (Fig. 3e). It suggested that as high as 38% transcripts of the IF&S_TE_Down group encode proteins targeted to plastids, which probably accounted for the higher enrichment of thylakoid-related terms in IF&S_TE_Down (Fig. 3f).

Chloroplast thylakoid membranes are rich in the light reactions of photosynthesis including photosystems I and II, photosynthetic electron transport, and ATP synthase (Fig. 3g), which are encoded by the subset of PhANGs for the photosynthetic apparatus^{35,57,58}. We found that the frequencies of N(A,C,G,T)/GC, especially that of the A/GC-type Kozak motif, increased from All_Plastid to PhANGs and to the photosynthetic apparatus (Fig. 3h). Consistent with the performance of N/GC-type Kozak motifs (Fig. 2b), transcripts of PhANGs and the photosynthetic apparatus exhibited higher translational I-E transition efficiencies, with significantly lower rPS values in inflorescences and seedlings of WT (Fig. 3i). Accordingly, both PhANGs and the photosynthetic apparatus showed significant and concomitant upregulation in their rPS values and downregulation in their TE values in *hot3-2* (Fig. 3j, k), which suggested that the N/GC-type Kozak motifs, especially the A/GC-type, promote both the I-E transition and TEs of transcripts in a HOT3-dependent manner (Fig. 3a, b, d).

The A/GC-type Kozak motif relies on HOT3 to achieve high translation efficiency

As the predominant Kozak sequence context in Arabidopsis, the A/GC-type motif was found to be present in transcripts exhibiting high I-E transition efficiencies (Fig. 2b). To further demonstrate the importance of the A/GC type Kozak motif in translation, we used the A/GC-Kozak-motif-containing *LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5* (*LHCBS*, AT4G10340) gene as an example to determine the role of the Kozak motif in translational initiation. We fused the DNA fragment encoding its intact 5' UTR and the first two codons, which includes the A/GC Kozak motif, to the 5' end of the Firefly luciferase (LUC) reporter gene (*LHCBS^{A/GC}-LUC* as the WT) (Fig. 4a). We then introduced mutations into the Kozak motif in *LHCBS^{A/GC}-LUC* to generate two constructs harboring two Kozak motif variants C/TG and C/AT, named *LHCBS^{C/TG}-LUC* and *LHCBS^{C/AT}-LUC*, respectively (Fig. 4a). Transcripts with C/TG and C/AT Kozak motifs exhibited low I-E transition efficiency with relatively higher median rPS values in WT (Supplementary Dataset 3). We then performed dual LUC assays in protoplasts to examine the translation efficiency of these constructs. Consistent with the observed I-E transition efficiency, *LHCBS^{A/GC}-LUC* produced more proteins compared with *LHCBS^{C/TG}-LUC* and *LHCBS^{C/AT}-LUC* (Fig. 4b). Because the *LUC* transcript levels were similar for all constructs (Fig. 4c), the differences in LUC proteins between the constructs were most likely attributed to variations in translation efficiency. Thus, the results validate the role of the A/GC-type Kozak motif in promoting translation efficiency of the *LHCBS* transcript in WT.

On the other hand, relative to all genes or all plastid genes, PhANGs contained a higher proportion of genes with the A/GC Kozak motif and exhibited higher I-E transition efficiencies in a HOT3-dependent manner (Fig. 3h, j). To test whether HOT3 is required for their expression, we examined *LHCBS* and two more PhANGs with the A/GC-type Kozak motif – AT3G54890 encoding LHCA1 and AT1G03130 encoding Psd2. Dual-LUC assays showed that loss of function in *HOT3* significantly reduced the protein levels of A/GC-containing transcripts of LHCA1, Psd2, and *LHCBS* (Fig. 4d), without significant changes in the corresponding RNA levels (Fig. 4e). To elucidate the regulatory relationship between the Kozak motif and HOT3-mediated I-E transition, we further performed Dual-LUC assays by comparing *LHCBS^{A/GC}-LUC* and *LHCBS^{C/AT}-LUC* products in WT and *hot3-2* protoplasts. The

results showed that products of *LHCBS^{A/GC}-LUC* with an optimal Kozak motif were significantly reduced in *hot3-2* vs. WT (Fig. 4f) whereas those of *LHCBS^{C/AT}-LUC* with a poor Kozak motif were unchanged, suggesting that the translation activity of the optimal A/GC-type Kozak motif requires HOT3. It could also be concluded that HOT3's effect in translation depends on the optimal Kozak motif. Therefore, the results are consistent with a role of Kozak motifs in HOT3-mediated I-E transition in translation, although it remains unknown whether the role is direct or indirect at the molecular level.

Ribo-seq genome browser views revealed strong ribosome stalling at their TISs in *hot3-2* in both inflorescence and seedling samples (Fig. 4g and Supplementary Fig. 5a). If there is a causal relationship between HOT3-mediated I-E transition and translation efficiency, one would expect that translation-stalled transcripts in *hot3-2* should exhibit low ribosome density and yield low protein products. Next, we used the A/GC-type Kozak motif-containing PhANGs like *LHCBS*, LHCA1, and Psd2 to test these hypotheses. To determine whether the stalling at TISs reduced the number of ribosomes on these transcripts, we examined the distribution of these transcripts in polysome fractions. Total lysates from inflorescences or seedlings from WT and *hot3-2* were fractionated in a 15%–55% sucrose gradient by ultracentrifugation, and RNA was isolated from each of the 12 fractions. Good RNA quality in each fraction was reflected by largely intact 25S and 18S ribosomal RNAs (rRNAs) (Fig. 4h). As the internal control, the 25S rRNA showed similar distribution patterns between inflorescence and seedling samples for each genotype, attesting to reproducibility. For both WT and *hot3-2* samples, the main peak fraction was 6 although stronger signals in 5 were observed in *hot3-2* (Fig. 4i and Supplementary Fig. 5b), which probably reflected the presence of more 80S monosomes on RNA, or the accumulation of free 60S ribosomal subunit due to defective maturation of the 40S ribosomal subunit in *hot3-2* as reported before³¹. Next, we quantified the relative distributions of *LHCBS*, LHCA1, and Psd2 RNAs in these fractions and found that the main peaks of both *LHCBS* and LHCA1 RNAs shifted from 10 to 8 in both inflorescences and seedlings of *hot3-2* (Fig. 4i and Supplementary Fig. 5c). For the Psd2 RNA, its peaks shifted from 9 and 10 to 8 in both samples of *hot3-2* (Supplementary Fig. 5d). As the negative control, AT1G25580 without significant changes in rPS values (*hot3-2* vs. WT) showed similar distribution patterns between WT and *hot3-2* in either inflorescences or seedlings (Supplementary Fig. 5e, f).

To better quantify the shifts, we combined 6–8 as light polysomes (LP) and 10 to 12 as heavy polysomes (HP) and measured relative transcript abundances (Fig. 4h). The ratio of LP/HP represents the relative amount of RNA transcripts undergoing less active translation (LP) vs. active translation (HP). Compared to WT, the LP/HP ratios of *LHCBS*, LHCA1, and Psd2 in *hot3-2* were much higher in both inflorescences and seedlings, while this was not observed for AT1G25580 (Fig. 4j). In addition, as an opposite case, AT1G21160 exhibited an increase in RNA level and a decrease in rPS fold change in *hot3-2* relative to WT (Supplementary Fig. 5g). Loss of function in *HOT3* resulted in an obvious shift of AT1G21160 RNA from LP to HP with a strong reduction in LP/HP values in *hot3-2* (Fig. 4j and Supplementary Fig. 5h). In agreement with the above results, the protein levels of LHCA1 and Psd2 were largely reduced in seedlings (Fig. 4k) and inflorescences (Supplementary Fig. 5i).

HOT3/eIF5B1-mediated translation initiation is required for chloroplast development and photosynthesis

Thylakoids in chloroplasts are rich in the photosynthetic apparatus encoded by PhANGs for the light reactions in photosynthesis (Fig. 3g)^{35,57,58}. As our results indicate that HOT3 facilitates the translation of PhANGs particularly those encoding the photosynthetic apparatus in the thylakoid membranes, we applied transmission electron microscopy to decipher the ultrastructure of chloroplasts in seedlings of WT and *hot3-2*. We found that the size of the grana stacks

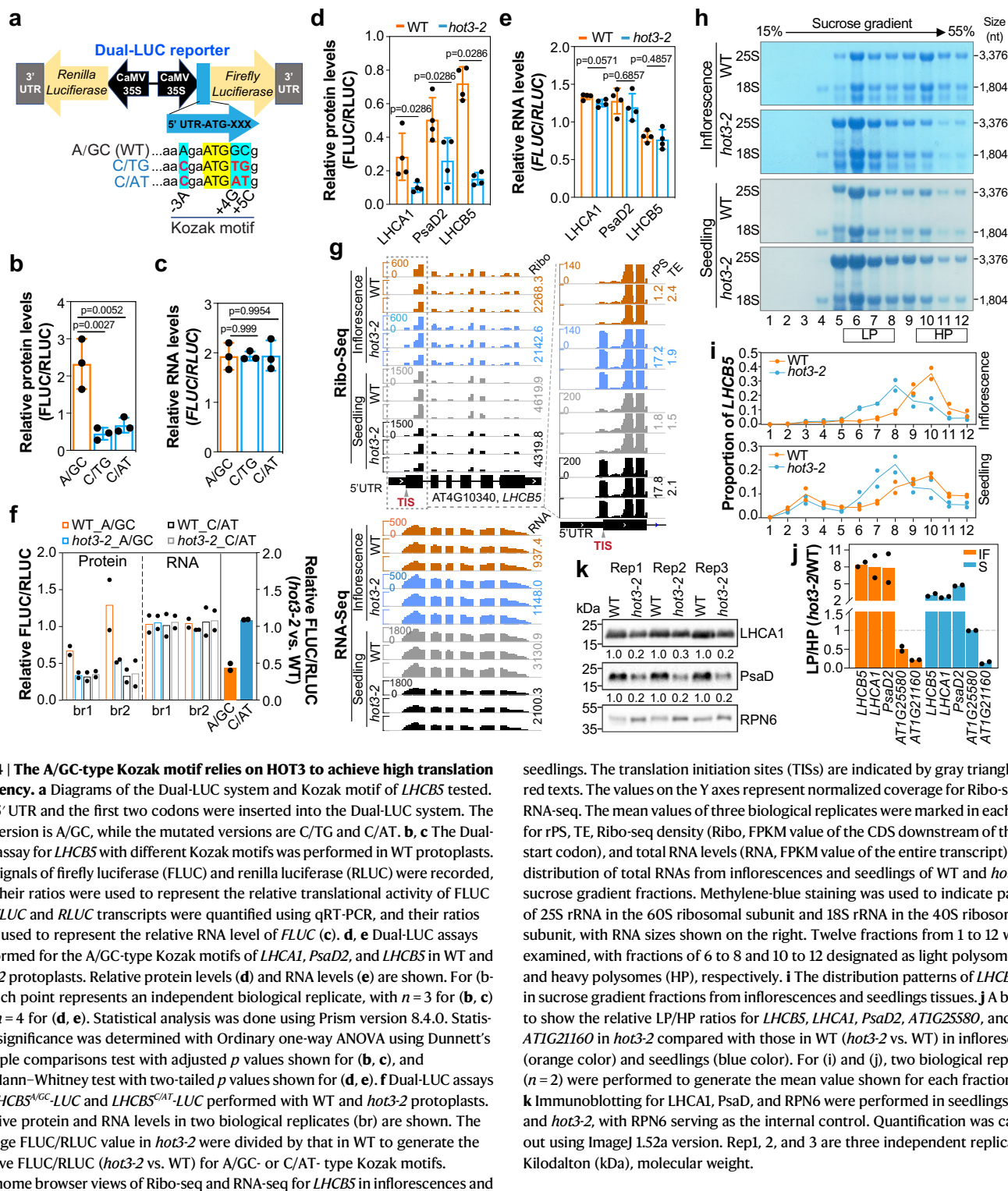


Fig. 4 | The A/GC-type Kozak motif relies on HOT3 to achieve high translation efficiency. **a** Diagrams of the Dual-LUC system and Kozak motif of *LHCBS* tested. The 5' UTR and the first two codons were inserted into the Dual-LUC system. The WT version is A/GC, while the mutated versions are C/TG and C/AT. **b, c** The Dual-LUC assay for *LHCBS* with different Kozak motifs was performed in WT protoplasts. The signals of firefly luciferase (FLUC) and renilla luciferase (RLUC) were recorded, and their ratios were used to represent the relative translational activity of FLUC (**b**). FLUC and RLUC transcripts were quantified using qRT-PCR, and their ratios were used to represent the relative RNA level of FLUC (**c**). **d, e** Dual-LUC assays were performed for the A/GC-type Kozak motifs of *LHCA1*, *Psad2*, and *LHCBS5* in WT and *hot3-2* protoplasts. Relative protein levels (**d**) and RNA levels (**e**) are shown. For (**b, c**), each point represents an independent biological replicate, with $n=3$ for (**b, c**) and $n=4$ for (**d, e**). Statistical analysis was done using Prism version 8.4.0. Statistical significance was determined with Ordinary one-way ANOVA using Dunnett's multiple comparisons test with adjusted p values shown for (**b, c**), and the Mann-Whitney test with two-tailed p values shown for (**d, e**). **f** Dual-LUC assays for *LHCBS*^{A/GC}-LUC and *LHCBS*^{C/AT}-LUC were performed with WT and *hot3-2* protoplasts. Relative protein and RNA levels in two biological replicates (br) are shown. The average FLUC/RLUC value in *hot3-2* were divided by that in WT to generate the relative FLUC/RLUC (*hot3-2* vs. WT) for A/GC- or C/AT- type Kozak motifs. **g** Genome browser views of Ribo-seq and RNA-seq for *LHCBS* in inflorescences and

seedlings. The translation initiation sites (TISs) are indicated by gray triangles and red texts. The values on the Y axes represent normalized coverage for Ribo-seq and RNA-seq. The mean values of three biological replicates were marked in each panel for rPS, TE, Ribo-seq density (Ribo, FPKM value of the CDS downstream of the ATG start codon), and total RNA levels (RNA, FPKM value of the entire transcript). **h** The distribution of total RNAs from inflorescences and seedlings of WT and *hot3-2* in sucrose gradient fractions. Methylene-blue staining was used to indicate patterns of 25S rRNA in the 60S ribosomal subunit and 18S rRNA in the 40S ribosomal subunit, with RNA sizes shown on the right. Twelve fractions from 1 to 12 were examined, with fractions of 6 to 8 and 10 to 12 designated as light polysomes (LP) and heavy polysomes (HP), respectively. **i** The distribution patterns of *LHCBS* RNA in sucrose gradient fractions from inflorescences and seedlings tissues. **j** A bar plot to show the relative LP/HP ratios for *LHCBS*, *LHCA1*, *Psad2*, *ATIG25580*, and *ATIG21160* in *hot3-2* compared with those in WT (*hot3-2* vs. WT) in inflorescences (orange color) and seedlings (blue color). For (i) and (j), two biological replicates ($n=2$) were performed to generate the mean value shown for each fraction. **k** Immunoblotting for *LHCA1*, *Psad*, and *RPN6* were performed in seedlings of WT and *hot3-2*, with *RPN6* serving as the internal control. Quantification was carried out using ImageJ 1.52a version. Rep1, 2, and 3 are three independent replicates. Kilodalton (kDa), molecular weight.

in *hot3-2* decreased significantly compared to that in WT (Fig. 5a). The number of thylakoids per grana was reduced from 7 in the WT to 3 in *hot3-2* (Fig. 5b). Measurements showed that the net photosynthetic rate of *hot3-2* leaves was significantly lower than WT under two different photon flux densities (Fig. 5c). In agreement with the impaired thylakoid structure (Fig. 5a, b) and reduced photosynthesis activity (Fig. 5c), the number of starch granules also significantly decreased in *hot3-2* (Fig. 5d). These chloroplast defects likely accounted for the chlorosis phenotypes of *hot3-2* during both seedling and inflorescence development (Supplementary Fig. 1a, b). These phenotypes attested to

a positive effect of HOT3 on chloroplast biogenesis and photosynthesis.

Discussion

The Kozak sequence context surrounding the TIS greatly impacts translational activity^{8,11,21,50,51}. Although the exact Kozak consensus sequence varies in different taxa, the -3 purine (A and G) and +4G are universally conserved among vertebrates and green plants^{8,17,18}. In this study, we developed an rPS index to quantitatively measure TIS stalling for a given transcript and further comprehensively evaluated the

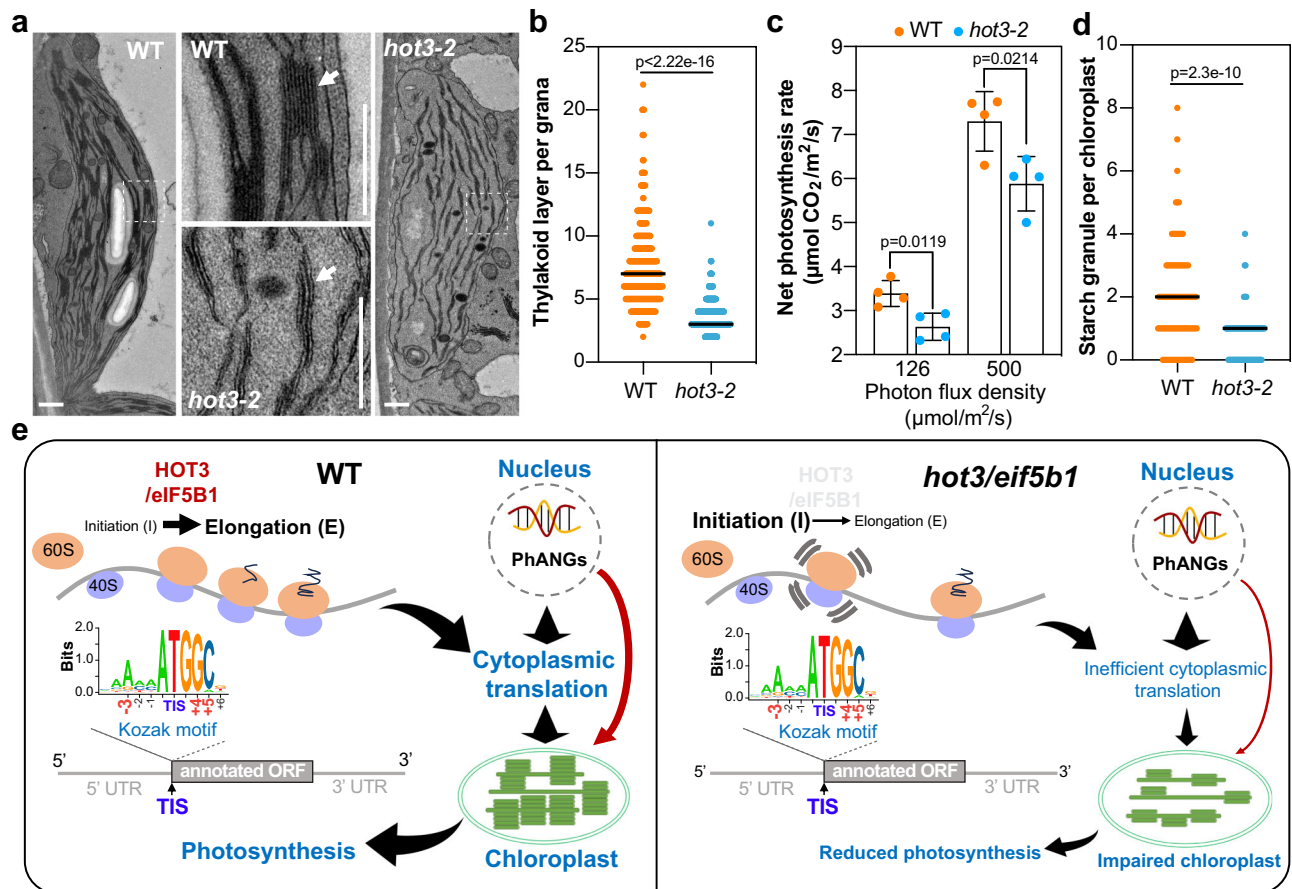


Fig. 5 | HOT3/eIF5B1-mediated translation initiation is required for chloroplast development and photosynthesis. **a** Representative transmission electron microscopy images of chloroplasts in WT and *hot3-2*. Thylakoid membranes are indicated with white arrows. Scale bars, 0.5 μm . 20 chloroplasts each from WT and *hot3-2* were observed with similar results. **b** Quantification of thylakoid layers per grana in WT and *hot3-2*. 555 and 862 grana from WT and *hot3-2*, respectively, were recorded. **c** Net photosynthetic rate of WT and *hot3-2* under two photon flux densities. Four biological replicates were performed for each group, with error bars representing the standard deviation. **d** Quantification of starch granules per chloroplast in WT and *hot3-2*. 100 chloroplasts each from WT and *hot3-2* were used to record starch granules. In (**b–d**), statistical significance was determined using the unpaired *t* test with two-tailed *p* values presented. The black lines in (**b**) and (**d**)

represent the median values. **e** A model of HOT3/eIF5B1 in cytoplasmic translation initiation and chloroplast biogenesis. HOT3/eIF5B1 facilitates mRNA translation by promoting the transition of the 80S ribosome from initiation (I) to elongation (E) at TISs. Transcripts with different Kozak sequences surrounding the TISs exhibit varied translational activity, with the A/GC type Kozak motif being the predominant type correlated with the highest efficiency in I-E transition in a HOT3-dependent manner in Arabidopsis. PhANGs transcripts, especially those encoding proteins for the photosynthesis apparatus on the thylakoid, tend to harbor the A/GC-type Kozak motif and preferentially require HOT3 for translation. Depletion of HOT3 compromises the I-E transition, as evidenced by severe ribosomal stalling at the AUG in *hot3-2*, leading to decreased mRNA translation, particularly for PhANGs encoding proteins essential for photosynthesis.

translational I-E transition efficiencies for different Kozak sequence contexts in *Arabidopsis thaliana*, *Oryza sativa*, and *Triticum aestivum*. Our findings revealed that translational I-E transition efficiencies vary for different Kozak motifs. The predominant Kozak motif used by the most transcripts show good I-E transition efficiency in all three species, suggesting that adopting a highly translationally efficient Kozak motif as the predominant *cis*-regulatory element may represent a common principle in evolution. We identified the optimal Kozak sequence in Arabidopsis as the A/GC-type Kozak motif, which is very similar to the consensus sequence for the AUG context in dicots proposed previously based on in silico analysis^{18,20,49}, as well as that from translational analysis under mild dehydration stress in Arabidopsis¹⁴. Furthermore, we found that the I-E transition efficiencies of most Kozak motifs are dependent on HOT3/eIF5B1. In particular, the A/GC type Kozak motif requires HOT3/eIF5B1 for its highest efficiency in I-E transition. Earlier genetic, biochemical, and structural studies in yeast, *Tetrahymena thermophila*, rabbit, and human cells have revealed that *cis*- and *trans*-factors of the translational machinery including the 40S ribosome P-site, 18S rRNA and eIF factors synergistically recognize

nucleotides at -3, +4, and +5 positions to promote initiation from AUG in an appropriate context^{4,8,59–64}. In addition, single-molecule fluorescence assays revealed that the residence time of the eIF5B-80S ribosome initiation complex (IC) at the TIS was strongly dependent on the Kozak sequence context surrounding the start codon²³. The non-optimal Kozak motifs led to longer residence times of the eIF5B-80S IC and delayed formation of an elongation-competent 80S complex²³. Most recently, eIF5B in human cells was reported to stabilize initiating ribosomes to maintain reading frame fidelity, which was tightly coupled with start codon recognition for nutrient stress adaptation²². In fact, Met-tRNA^{Met} could be the mediator linking the Kozak motif and eIF5B physically, through interacting with the Kozak motif at the TIS start codon and eIF5B at Domain IV⁴⁸. Together, the previous studies and our findings suggest that an optimal Kozak sequence context and HOT3/eIF5B1 may coordinately function in start codon recognition, and an optimal Kozak motif might directly or indirectly promote the GTP-hydrolysis by eIF5B and/or eIF5B-GDP dissociation to reduce the residence time of the 80S ribosome at the TIS to facilitate the I-E transition. Nevertheless, more studies are required to elucidate how

Kozak motifs are mechanistically involved in HOT3-mediated I-E transition.

As the four 3-nt codons starting with GC encode Alanine (Ala), it is conceivable that Methionine-Alanine (Met-Ala) would be the predominant N-terminal di-peptide encoded by transcripts with high levels of +4G and +5C (Fig. 3d). Indeed, Met-Ala is the most prevalent N-terminal residues of proteins from both vertebrates⁶⁵ and higher plants^{14,18}. Also, Met-Ala is the most frequently occurring N-terminal residues of the chloroplast transit peptide (cTP)^{66,67}, with up to 57% of Arabidopsis and 59% of rice nuclear genes encoding plastid proteins containing the two N-terminal residues⁵⁶. Our study uncovered higher levels of N-terminal Met-Ala in the IF_TE_Down, S_TE_Down, IF&S_TE_Down, and PhANGs peptides relative to IF_TE_All, S_TE_All, or All_Plastid (Fig. 3d). Although the exact molecular function of Met-Ala in cTPs is largely unknown, the 2nd Ala may be beneficial for protein products of PhANGs encoding the photosynthetic apparatus in the thylakoid in different ways. In addition to enhancing the efficiency of translation initiation by the Ala codons at the +4 and +5 positions of the Kozak consensus sequence, Ala may assist the entrance of the nascent peptide into the ribosome tunnel⁶⁸. In *E. coli* and *S. cerevisiae*, after removal of the initiator Met, the N-terminal Ala is a stable residue favorable for protein half-life⁶⁹. Similarly, adopting Ala as the second N-terminal residue in chloroplast-targeted proteins may maintain their stability in the cytosol before being imported into chloroplasts. Interestingly, the second N-terminal residue is almost invariably a Leucine in mitochondria transit signal peptides⁶⁶, which may account for the insensitivity of nuclear genes encoding mitochondrial proteins to *hot3-2* mutations in terms of rPS and TE based on GO term analysis. Moreover, it is interesting that distinct Kozak motifs (e.g., N/GC vs. N/GA) were enriched in genes involved in different biological processes, cellular components, or molecular functions (Supplementary Fig. 4d-i). For example, the suboptimal N/GA-type Kozak motif is enriched in proteins in DNA transcription, which may represent a new layer of regulation in cellular processes.

Accumulating evidence suggests that cytoplasmic translation of PhANGs may play a pivotal role in chloroplast biogenesis⁷⁰. However, the mechanism that confers the selective translational control of the PhANGs as a group remained unclear. Here, we demonstrate that PhANGs, particularly those encoding the thylakoid-localized photosynthetic apparatus, harbor the optimal Kozak motif of -3A, +4G, and +5C surrounding the TIS and show high translation initiation efficiency. More importantly, the translation activity of the optimal A/GC Kozak motif-containing PhANGs is mediated via a HOT3/eIF5B1-dependent mechanism (Fig. 5e). Using three photosynthesis genes as models, we demonstrate that the A/GC-type Kozak motif plays a critical role in translation efficiency and the accumulation of photosynthesis proteins. Consistent with the important role of HOT3 in translation of PhANGs, disruption of *HOT3* compromises the translational I-E transition causing ribosome stalling on the TIS at these RNA transcripts, resulting in reduced protein production and chloroplast defects represented by less thylakoid layers per grana (Fig. 5e). The stalled and inefficient translation of PhANGs likely accounts for the impaired chloroplasts, reduced photosynthesis activity, and pale-green leaves of the *hot3* mutants (Fig. 5e)^{30,31}. Together, these results elucidate a HOT3-mediated translational control of nuclear photosynthesis genes for chloroplast biogenesis.

Moreover, since HOT3 is required for plant survival under heat stress³⁰, eIF5B1-mediated translational I-E transition is probably particularly important for plants' response to heat. We observed that *eIF5B2* (AT1G21160) was upregulated in *hot3-2* mutants in terms of RNA abundance (Supplementary Fig. 5g), translation I-E transition with reduced rPS fold change, and translational efficiency (Supplementary Fig. 5h), perhaps through an unknown negative feedback mechanism. However, eIF5B2 cannot completely replace eIF5B1 in *hot3-2*. It is possible that the eIF5B proteins differ in their biochemical properties

such as the GTPase activity or eIF5B-GDP dissociation, or that they act on different sets of transcripts.

Methods

Plant materials and growth conditions

The *hot3-2* (SALK_124251) allele was previously reported^{30,31}. The wild-type used in this study was the Arabidopsis Columbia (Col-0) ecotype. Plants were grown either in soil in growth rooms or on 1/2 Murashige & Skoog growth media (PhytoTech Labs, M524) containing 1% sucrose (Fisher Scientific, BP220-10) and 0.8% Agar (Sigma-Aldrich, A1296) in a growth chamber (JIUPO, BPC500-2H) from FUJIAN JIUPO BIOTECHNOLOGY CO., LTD, at 22 °C under full-spectrum white light (LED light, JIUPO, JIUPO-5050TLED-300-S), with a long-day photoperiod (16 h light/8 h dark). Inflorescence and 14-day-old seedlings from the wild-type and *hot3-2* plants were collected, frozen in liquid nitrogen, and stored in -80 °C until used.

Ribo-seq and RNA-seq library construction

Ribosome profiling libraries were generated according to a published protocol⁴⁴ with modifications detailed below. 0.2 g of inflorescence or seedling tissues were ground into fine powder in liquid nitrogen and lysed in 0.45 ml Buffer D containing the following components: 100 mM Tris-HCl pH8.0, 40 mM KCl, 20 mM MgCl₂, 2% (v/v) polyoxyethylene (10) tridecyl ether (Sigma-Aldrich, P2393), 0.5% (w/v) sodium deoxycholate (Sigma-Aldrich, D6750), 1 mM dithiothreitol (Thermo Fisher Scientific, R0861), 100 µg/ml cycloheximide (Sigma-Aldrich, C1988) and 10 U/ml DNase I (Roche, 4716728001). After a 10-min incubation on ice, two rounds of centrifugation at 19,283 g for 10 min at 4 °C were performed to generate cleared supernatant labeled as total lysate. The OD₂₆₀ of the cleared total lysate was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, ND-2000) with Buffer D as the blank. Around 2,000U OD₂₆₀ of the total lysate was incubated with 75U RNase I (Thermo Fisher Scientific, AM2294) at 23 °C for one hour. The reaction was terminated by adding 300U SUPERase-In RNase Inhibitor (Thermo Fisher Scientific, AM2694) and further processed with Amersham MicroSpin S-400 HR columns (Cytiva, 27514001). Ribosome-protected footprints (RPFs) were isolated and purified using the RNA Clean & Concentrator-5 kit (Zymo Research, R1016). Subsequently, rRNA depletion was carried out using the rRNA removal mix in TruSeq® Stranded Total RNA Library Prep Plant (Illumina, 20020610) according to the manufacturer's instructions. The rRNA-depleted RPFs were concentrated by the RNA Clean & Concentrator-5 kit, and fragments of around 25-nt to 32-nt were selected in 15% urea-PAGE. The sliced gels were ground by RNase-free 1 ml Olympus Premium Reach Pipet Tips (Genesee Scientific, 23-165RL). The extraction buffer (0.5 M ammonium acetate, 10% SDS) was added to extract RPFs at room temperature for 2 h. Gel debris was removed by Corning Costar Spin-X centrifuge tube filters (Sigma-Aldrich, CLS9301). The RPFs were precipitated by incubation with 1 µl glycogen (Thermo Fisher Scientific, R0551) and 3 volumes of 100% ethanol at -20 °C overnight. End repair was carried out for the RPFs using T4 PNK without ATP for 30 min at 37 °C, followed by 5' phosphorylation with ATP for additional 30 min at 37 °C. The RPFs recovered from the reactions were dissolved in 3 µl RNase-free ddH₂O and subjected to small RNA library construction using NEBNext Multiplex Small RNA Library Prep Set for Illumina (NEB, E7300). The barcoded libraries were pooled for paired-end 150-bp sequencing in HiSeq X Ten platform.

Total RNA was isolated from the tissue powder using TRI-reagent (MRC, TRI18). The polyadenylated RNAs, isolated from total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490L), were subjected to RNA-seq library construction. The NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, E7420L) was employed for library construction. The barcoded libraries were pooled for paired-end 150-bp sequencing on the Novaseq 6000 platform.

RNA-seq data analysis

RNA-seq reads were processed by the pRNASeqTools pipeline (<https://github.com/grubbybio/pRNASeqTools>). Briefly, paired raw reads were trimmed with cutadapt v4.1⁷¹, and then mapped to the Arabidopsis genome Araport11⁴⁷ (<https://www.arabidopsis.org>), rice genome LSU v7, or wheat genome IWGSC v2.1 by STAR v2.7.9a⁷² with parameters “-outSAMmultNmax 1 -outFilterMultimapNmax 50 -outFilterMismatchNoverLmax 0.1”. Only properly and uniquely mapped read pairs were counted by featureCounts v2.0.3⁷³ for downstream analyses.

P-site and ORF identification

Ribo-seq reads were trimmed by cutadapt v4.1⁷¹ and rRNA, tRNA, snRNA, snoRNA and other repeat sequences were removed. Remaining reads were then mapped to Araport11 by STAR with parameters “-alignIntronMax 5000 -alignIntronMin 15 -outFilterMismatchNmax 2 -outFilterMultimapNmax 20 -outFilterType BySJout -alignSJoverhangMin 4 -alignSJDBoverhangMin 1 -outSAMmultNmax 1”. Mapped RPFs were then processed using the RiboTaper pipeline⁴⁶. Briefly, metaplots for all RPFs from 12 samples around the start and stop codons were created by ‘create_metaplots.bash’, and the P-site offset for RPFs in each length was determined accordingly. Next, ‘RiboTaper.sh’ was called to identify ORFs with options ‘21,22,24,25,26,27,28,29,30,31,32,33 12,13,8,9,10,11,12,13,15,16,17’.

For Metagene analysis, the P sites obtained from RiboTaper were split into individual files corresponding to each biological replicate, and their closest start and stop codons were searched using start_stops_FAR.bed generated by RiboTaper with closestBed from Bedtools. Only P sites with distances to annotated start or stop codons ranging from -20 to 20 nt were retained. Normalized P site counts in RPM were calculated for plotting.

rPS and TE calculations and comparison between WT and *hot3-2*

P-sites mapped to the CDS of each gene (N_{CDS}) for rPS and TE calculations were obtained from RiboTaper⁴⁶. Specifically, P-sites coinciding with TISs (N_{TIS}) were counted by featureCounts⁷³ separately. Additionally, the number of codons in each CDS (N_{codon}) were used for length normalization. Transcripts with ≥ 10 reads mapped to the CDS downstream of the TIS were used for analysis. Therefore, the rPS index for each gene was calculated using the following equation:

$$rPS = \frac{N_{TIS}}{(N_{CDS} - N_{TIS}) / (N_{codon} - 1)}$$

The calculation of TE was following a previous report⁵³, using DESeq2 for read normalization⁷⁴, but a revised equation with N_{ORF} being replaced by $N_{CDS} - N_{TIS}$ was used.

$$TE = \frac{\text{Normalized}(N_{CDS} - N_{TIS})}{\text{Normalized RNA read counts}}$$

To compare the TE of each transcript between WT and *hot3-2*, an in-house R script (http://github.com/grubbybio/pRNASeqTools/scripts/tf_mrna.R) was employed based on the “interaction” design in DESeq2⁷⁴. Briefly, raw counts of ($N_{ORF} - N_{TIS}$) and RNA read counts for each transcript were imported and results for the interaction term “genotype.condition” were extracted as fold change of TE. Only transcripts encoded by nuclear genes with adjusted *p* value less than 1 were retained for later analyses. For inflorescence and seedling samples, 6208 and 9466 transcripts with valid fold-change values (*hot3-2* vs. WT) of rPS and TE, respectively, were retained.

Homologous transcripts of Arabidopsis IF&S_A/GC in rice and wheat were obtained using the Homologs module at the TGT platform (<http://wheat.cau.edu.cn/TGT/>). For rice, *Oryza sativa* Japonica (IRGSP1.0) was used as the searching reference. For wheat,

homologous genes were first searched from *Triticum aestivum* (IWGSC RefSeqv1.1) and then given the standard gene names based on the *Triticum aestivum* (IWGSC RefSeqv2.1) reference at the TGT platform.

Polyribosome isolation and quantitative RT-PCR

Polyribosome isolation and fractionation were conducted following the previously described method³¹, with minor adjustments. One milliliter of polysome extraction buffer (PEB) containing 0.2 M Tris-HCl pH8.0, 0.2 M KCl, 35 mM MgCl₂, 25 mM EGTA, 1% TritonX-100, 2% polyoxyethylene-10-tridecyl ether (PTE), 100 mM 2-Mercaptoethanol, 25 μg/ml cycloheximide (CHX), 100 μg/ml chloramphenicol (CHL) and 40 U/ml SUPERase-In™ RNase Inhibitor (Thermo Fisher Scientific, AM2694) was added and mixed thoroughly with the tissue powder. After 10 min incubation on ice, cell debris was removed by centrifugation in a microcentrifuge at 4 °C for 5 min at 14,167 g. Sodium deoxycholate (DOC) was added to the supernatant to a final concentration of 0.5% and the supernatant was left on ice for 5 min. Following centrifugation in a microcentrifuge for 15 min at 14,167 g, the supernatant was transferred to a new 1.5 ml RNase-free tube and labeled as PEB-input. 0.5 ml of the supernatant was loaded onto a 4.4-ml sucrose gradient (15%-55% sucrose in 40 mM Tris-HCl pH8.0, 20 mM KCl, 10 mM MgCl₂, 25 μg/ml CHX, 4 U/ml SUPERase-In™ RNase Inhibitor). After ultracentrifugation in a Beckman SW-55Ti rotor at 191,643 g for 65 min at 4 °C, 0.41 ml fractions from top to bottom were manually transferred to 12 tubes. Ethylenediamine tetraacetic acid (EDTA) and sodium dodecyl sulfate (SDS) were added to each fraction at room temperature to final concentrations of 20 mM and 0.5%, respectively. RNA was immediately extracted from each fraction using 0.5 ml TRI-reagent (MRC, TR118) according to the manufacturer's instructions, and pelleted RNA was resuspended in 25 μl RNase-free ddH₂O. Twenty percent (5 μl) of the RNA in each fraction was separated in a 1.2% (w/v) agarose/formaldehyde gel and transferred to an Hybond N+ hybridization membrane (Cytiva, RPN303B) by capillary elution. Methylene blue (MB) staining was used to assess RNA quality and relative quantity. Subsequently, forty percent (10 μl) of the RNA in each fraction was used for reverse transcription with RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, K1621), employing oligo(dT) as the RT primer following the manufacturer's instructions. Real-time PCR analysis was performed by using SYBR-Green qPCR Supermix (BioRad, 170-8882) in a CFX96 Real-Time system (BioRad, C1000 Touch) according to the manufacturer's instructions. Each sample consisted of two independent biological replicates, and RT-qPCR reactions were conducted at least three times. A list of primers is provided in Supplementary Table 1.

Antibodies used in immunoblotting

The commercial antibodies used in this study were anti-PsaD (Agrisera, AS09461), anti-LHCA1 (Agrisera, AS01005), anti-RPN6 (Enzo Life Sciences, BML-PW8370-0100), anti-ACTIN (Sigma-Aldrich, A0480), anti-rabbit secondary antibodies conjugated with horseradish peroxidase (Bio-Rad, 1706515), and anti-mouse secondary antibodies conjugated with horseradish peroxidase (Bio-Rad, 1706516). The primary and secondary antibodies were used at 1:1500 and 1:5000 dilutions, respectively.

Quantification of chlorophyll content

The chlorophyll content was quantified using the published methods with some modifications⁷⁵. 0.05 g of 18-day-old seedlings were submerged in 1 ml 95% ethanol and incubated at 80 °C in darkness for 20 min with occasional shaking. After centrifugation at 2460 × g for 1 min, the cleared supernatant was transferred to a new tube and measured at wavelengths of 664 nm and 648 nm to obtain absorbance values, using SmartSpec Plus Spectrophotometer (Bio-Rad). Six biological replicates were performed for each genotype. The total chlorophyll content was calculated using the equations: Chlorophyll a

(mg/L) = $13.36 \times A_{664} - 5.19 \times A_{648}$, Chlorophyll b (mg/L) = $27.43 \times A_{648} - 8.12 \times A_{664}$, Total chlorophyll = chlorophyll a + chlorophyll b. The chlorophyll value (mg/L) was then divided by 0.05 g/0.001 L to obtain the chlorophyll value (mg/g FW).

Measurements of net photosynthetic rate

The net photosynthetic rates of leaves were measured with a standard system (Heinz Walz GmbH, Effeltrich, Germany), which combines a GFS-3000 gas analyzer and a DUAL-PAM-100 PAM-fluorometer adapted with a DUAL-PAM Gas-Exchange Cuvette 3010-DUAL common measuring head. The measurement conditions were as follows: 21 °C, 50% relative humidity, 126 $\mu\text{mol}/\text{m}^2/\text{s}$ or 500 $\mu\text{mol}/\text{m}^2/\text{s}$ photon flux density, and 400 ppm CO_2 . Expanded leaves from 45-day-old WT or *hot3-2* seedlings grown in soil under short-day conditions (8-h light and 16-h dark cycle at 22 °C) were used for measurements.

Vector construction and dual luciferase assay

The DNA fragments containing the intact 5' UTR plus the first two codons of each gene were cloned into XF2428 at NcoI and EcoRI sites by using NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621L). LHCb5_LP and LHCb5_A/GC_RP were used for LHCb5 (A/GC). LHCb5_LP and LHCb5_C/AT_RP were used for LHCb5 (C/AT). LHCA1_LP and LHCA1_A/GC_RP were used for LHCA1 (A/GC). Psd2_LP and Psd2_A/GC_RP were used for Psd2 (A/GC) (Supplementary Table 1). After confirmation by DNA sequencing using primer R380 (Supplementary Table 1), plasmids were purified for transformation. Arabidopsis protoplasts were isolated from fully expanded rosette leaves of 45-day-old WT or *hot3-2* grown in soil under short-day conditions (8-h light and 16-h dark cycle at 22 °C). The preparation and transformation of Arabidopsis protoplast were performed as reported before⁴⁵. After transformation and overnight incubation, protoplasts in one tube were separated into two tubes and harvested by centrifugation at $100 \times g$ for 3 min at room temperature. After discarding the supernatant, protoplasts in one tube were quickly frozen in liquid nitrogen and stored in -80 °C for RNA extraction. The lysates were generated from protoplasts in the other tube by adding 50 μL 1 \times Cell Lysis Buffer included in the Dual Luciferase Reporter Assay Kit (Vazyme, DL101-01) to the protoplasts and vigorously shaking at room temperature for 15 min. The lysates were cleared by centrifugation at 12,000 g for 3 min, and 20 μL of the supernatant was used for the dual-luciferase assay measured in a multimode Reader (Berthold, LB 942) as specified by the manufacturer. Firefly luminescence (FLUC) was normalized to their corresponding Renilla luminescence (RLUC) to represent relative expression of FLUC. RNAs extracted from the protoplasts were used for RT-qPCR using FLUC_qLP and FLUC_qRP for FLUC, RLUC_qLP and RLUC_qRP for RLUC. FLUC was normalized to their corresponding RLUC to represent relative expression of the FLUC transcript.

Transmission electron microscopy (TEM)

To conduct TEM analysis, cotyledons from 14-day-old seedling were fixed in a solution of 2.5% (v/v) glutaraldehyde and 2% (v/v) paraformaldehyde diluted in 0.1 M PBS buffer (pH 7.4) at 4 °C overnight. After being washed three times with 0.1 M PBS buffer, the samples were stained with 1% osmium tetroxide (w/v) and 0.8% potassium ferricyanide (w/v) in the same buffer for 2 h at 4 °C, followed by three washes with deionized water. Subsequently, 1% uranyl acetate (w/v) was used for en-bloc staining at 4 °C overnight. Sample dehydrations were achieved through a series of graded ethanol, and fresh resin was used for embedding. After polymerization, ultrathin sections (70 nm) were obtained at 65 °C using a Leica UC7 ultramicrotome equipped with a Dito diamond knife. Post-staining was carried out using uranyl acetate and lead citrate. The grids with samples were observed at 80 kV in a JEOL JEM-1400Flash 120 kV TEM and imaged with a CMOS camera (XAROSA, EMSIS).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus under the accession number [GSE212857](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE212857). Source data are provided with this paper.

References

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
2. Browning, K. S. & Bailey-Serres, J. Mechanism of cytoplasmic mRNA translation. *Arabidopsis Book* **13**, e0176 (2015).
3. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.* **83**, 779–812 (2014).
4. Fang, J. C. & Liu, M. J. Translation initiation at AUG and non-AUG triplets in plants. *Plant Sci.* **335**, 111822 (2023).
5. Merchante, C., Stepanova, A. N. & Alonso, J. M. Translation regulation in plants: an interesting past, an exciting present and a promising future. *Plant J.* **90**, 628–653 (2017).
6. Wu, H. L., Jen, J. & Hsu, P. Y. What, where, and how: Regulation of translation and the translational landscape in plants. *Plant Cell* **36**, 1540–1564 (2024).
7. Brito Querido, J., Diaz-Lopez, I. & Ramakrishnan, V. The molecular basis of translation initiation and its regulation in eukaryotes. *Nat. Rev. Mol. Cell Biol.* **25**, 168–186 (2024).
8. Hernandez, G., Osnaya, V. G. & Perez-Martinez, X. Conservation and variability of the AUG initiation codon context in eukaryotes. *Trends Biochem. Sci.* **44**, 1009–1021 (2019).
9. Roy, B. Effects of mRNA modifications on translation: An overview. *Methods Mol. Biol.* **2298**, 327–356 (2021).
10. Xiang, Y. et al. Pervasive downstream RNA hairpins dynamically dictate start-codon selection. *Nature* **621**, 423–430 (2023).
11. Kozak, M. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**, 5233–5252 (1981).
12. Liu, M. J. et al. Translational landscape of photomorphogenic Arabidopsis. *Plant Cell* **25**, 3699–3710 (2013).
13. Wu, H. L., Song, G., Walley, J. W. & Hsu, P. Y. The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiol.* **181**, 367–380 (2019).
14. Kawaguchi, R. & Bailey-Serres, J. mRNA sequence features that contribute to translational regulation in Arabidopsis. *Nucleic Acids Res.* **33**, 955–965 (2005).
15. Sotta, N. et al. Translational landscape of a C4 plant, sorghum bicolor, under normal and sulfur-deficient conditions. *Plant Cell Physiol.* **63**, 592–604 (2022).
16. Lei, L. et al. Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.* **84**, 1206–1218 (2015).
17. Nakagawa, S., Niimura, Y., Gojibori, T., Tanaka, H. & Miura, K. Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic Acids Res.* **36**, 861–871 (2008).
18. Joshi, C. P., Zhou, H., Huang, X. & Chiang, V. L. Context sequences of translation initiation codon in plants. *Plant Mol. Biol.* **35**, 993–1001 (1997).
19. Rangan, L., Vogel, C. & Srivastava, A. Analysis of context sequence surrounding translation initiation site from complete genome of model plants. *Mol. Biotechnol.* **39**, 207–213 (2008).
20. Gupta, P., Rangan, L., Ramesh, T. V. & Gupta, M. Comparative analysis of contextual bias around the translation initiation sites in plant genomes. *J. Theor. Biol.* **404**, 303–311 (2016).

21. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
22. Mao, Y., Jia, L., Dong, L., Shu, X. E. & Qian, S. B. Start codon-associated ribosomal frameshifting mediates nutrient stress adaptation. *Nat. Struct. Mol. Biol.* **30**, 1816–1825 (2023).
23. Wang, J. et al. eIF5B gates the transition from translation initiation to elongation. *Nature* **573**, 605–608 (2019).
24. Lapointe, C. P. et al. eIF5B and eIF1A reorient initiator tRNA to allow ribosomal subunit joining. *Nature* **607**, 185–190 (2022).
25. Chukka, P. A. R., Wetmore, S. D. & Thakor, N. Established and emerging regulatory roles of eukaryotic translation initiation factor 5B (eIF5B). *Front Genet.* **12**, 737433 (2021).
26. Roll-Mecak, A., Shin, B. S., Dever, T. E. & Burley, S. K. Engaging the ribosome: universal IFs of translation. *Trends Biochem. Sci.* **26**, 705–709 (2001).
27. Choi, S. K., Lee, J. H., Zoll, W. L., Merrick, W. C. & Dever, T. E. Promotion of met-tRNA^{Met} binding to ribosomes by yIF2, a bacterial IF2 homolog in yeast. *Science* **280**, 1757–1760 (1998).
28. Carrera, P. et al. VASA mediates translation through interaction with a Drosophila yIF2 homolog. *Mol. Cell* **5**, 181–187 (2000).
29. Castellano, M. M. & Merchante, C. Peculiarities of the regulation of translation initiation in plants. *Curr. Opin. Plant Biol.* **63**, 102073 (2021).
30. Zhang, L. et al. Mutations in eIF5B Confer Thermosensitive and Pleiotropic Phenotypes via Translation Defects in Arabidopsis thaliana. *Plant Cell* **29**, 1952–1969 (2017).
31. Hang, R. et al. Arabidopsis HOT3/eIF5B1 constrains rRNA RNAi by facilitating 18S rRNA maturation. *Proc. Natl Acad. Sci. USA* **120**, e2301081120 (2023).
32. Taylor, W. C. Regulatory interactions between nuclear and plastid genomes. *Annu. Rev. Plant Physiol.* **40**, 211–233 (1989).
33. Woodson, J. D. & Chory, J. Coordination of gene expression between organellar and nuclear genomes. *Nat. Rev. Genet.* **9**, 383–395 (2008).
34. Yoo, C. Y., Han, S. & Chen, M. Nucleus-to-Plastid phytochrome signalling in controlling chloroplast biogenesis. *Annu. Plant Rev.* **3**, 251–280 (2020).
35. Hwang, Y. et al. Anterograde signaling controls plastid transcription via sigma factors separately from nuclear photosynthesis genes. *Nat. Commun.* **13**, 7440 (2022).
36. Jiao, Y., Lau, O. S. & Deng, X. W. Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genet.* **8**, 217–230 (2007).
37. Berry, J. O., Yerramsetty, P., Zielinski, A. M. & Mure, C. M. Photosynthetic gene expression in higher plants. *Photosynth Res.* **117**, 91–120 (2013).
38. Liu, M. J., Wu, S. H., Chen, H. M. & Wu, S. H. Widespread translational control contributes to the regulation of Arabidopsis photomorphogenesis. *Mol. Syst. Biol.* **8**, 566 (2012).
39. Chen, G. H., Liu, M. J., Xiong, Y., Sheen, J. & Wu, S. H. TOR and RPS6 transmit light signals to enhance protein translation in deetioliating Arabidopsis seedlings. *Proc. Natl Acad. Sci. USA* **115**, 12823–12828 (2018).
40. Juntawong, P. & Bailey-Serres, J. Dynamic light regulation of translation status in Arabidopsis thaliana. *Front Plant Sci.* **3**, 66 (2012).
41. Lellis, A. D. et al. Deletion of the eIFiso4G subunit of the Arabidopsis eIFiso4F translation initiation complex impairs health and viability. *Plant Mol. Biol.* **74**, 249–263 (2010).
42. Lellis, A. D. et al. eIFiso4G augments the synthesis of specific plant proteins involved in normal chloroplast function. *Plant Physiol.* **181**, 85–96 (2019).
43. Ma, L. et al. SlRBP1 promotes translational efficiency via SlEIF4A2 to maintain chloroplast function in tomato. *Plant Cell* **34**, 2747–2764 (2022).
44. Hsu, P. Y. et al. Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **113**, E7126–E7135 (2016).
45. Wu, H. L. et al. Improved super-resolution ribosome profiling reveals prevalent translation of upstream ORFs and small ORFs in Arabidopsis. *Plant Cell* **36**, 510–539 (2024).
46. Calviello, L. et al. Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* **13**, 165–170 (2016).
47. Cheng, C. Y. et al. Araport11: A complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.* **89**, 789–804 (2017).
48. Wang, J. et al. Structural basis for the transition from translation initiation to elongation by an 80S-eIF5B complex. *Nat. Commun.* **11**, 5003 (2020).
49. Lukaszewicz, M., Feuermann, M., Jerouville, B., Stas, A. & Boutry, M. In vivo evaluation of the context sequence of the translation initiation codon in plants. *Plant Sci.* **154**, 89–98 (2000).
50. Diaz de Arce, A. J., Noderer, W. L. & Wang, C. L. Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons. *Nucleic Acids Res.* **46**, 985–994 (2018).
51. Noderer, W. L. et al. Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Mol. Syst. Biol.* **10**, 748 (2014).
52. Zhu, X. T. et al. Ribosome profiling reveals the translational landscape and allele-specific translational efficiency in rice. *Plant Commun.* **4**, 100457 (2023).
53. Yang, X. et al. Comparative ribosome profiling reveals distinct translational landscapes of salt-sensitive and -tolerant rice. *BMC Genomics* **22**, 612 (2021).
54. Guo, Y. et al. The translational landscape of bread wheat during grain development. *Plant Cell* **35**, 1848–1867 (2023).
55. Ingolia, N. T., Ghaemmhami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
56. Zybailov, B. et al. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **3**, e1994 (2008).
57. Eberhard, S., Finazzi, G. & Wollman, F. A. The dynamics of photosynthesis. *Annu. Rev. Genet.* **42**, 463–515 (2008).
58. Leister, D. & Schneider, A. From genes to photosynthesis in Arabidopsis thaliana. *Int. Rev. Cytol.* **228**, 31–83 (2003).
59. Pisarev, A. V. et al. Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* **20**, 624–636 (2006).
60. Pelletier, J. & Sonenberg, N. The Organizing Principles of Eukaryotic Ribosome Recruitment. *Annu. Rev. Biochem.* **88**, 307–335 (2019).
61. Hussain, T. et al. Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597–607 (2014).
62. Dresios, J., Chappell, S. A., Zhou, W. & Mauro, V. P. An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* **13**, 30–34 (2006).
63. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature* **500**, 307–311 (2013).
64. Simonetti, A., Guca, E., Bochler, A., Kuhn, L. & Hashem, Y. Structural Insights into the Mammalian late-stage initiation complexes. *Cell Rep.* **31**, 107497 (2020).
65. Sanchez, J. Alanine is the main second amino acid in vertebrate proteins and its coding entails increased use of the rare codon GCG. *Biochem Biophys. Res. Commun.* **373**, 589–592 (2008).
66. von Heijne, G., Steppuhn, J. & Herrmann, R. G. Domain structure of mitochondrial and chloroplast targeting peptides. *Eur. J. Biochem.* **180**, 535–545 (1989).
67. Chotewutmontri, P. & Bruce, B. D. Non-native, N-terminal Hsp70 molecular motor recognition elements in transit peptides support plastid protein translocation. *J. Biol. Chem.* **290**, 7602–7621 (2015).

68. Tenson, T. & Ehrenberg, M. Regulatory nascent peptides in the ribosomal tunnel. *Cell* **108**, 591–594 (2002).
69. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149 (1996).
70. Jarvis, P. & Lopez-Juez, E. Biogenesis and homeostasis of chloroplasts and other plastids. *Nat. Rev. Mol. Cell Biol.* **14**, 787–802 (2013).
71. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 3 (2011).
72. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
73. 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).
74. 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).
75. Lichtenthaler, H. K. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. In *Methods in Enzymology*, Vol. 148, 350–382 (Academic Press, 1987).
76. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
77. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
78. Chen, C. et al. TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* **16**, 1733–1742 (2023).

Acknowledgements

We would like to express our gratitude to Dr. Xiaofeng Cao at the Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, for providing the Dual-LUC construct XF2428. Additionally, we extend our thanks to Dr. Yuling Jiao and Dr. Danmeng Zhu at Peking University for sharing their short-day growth room and Dual-LUC analytic instrument, respectively. We also acknowledge the assistance of Nora Flynn, Dr. Ye Xu and Dr. Youra Hwang at the University of California, Riverside, in the project. Furthermore, we appreciate the support from the Core Facilities of the School of Life Sciences and National Center for Protein Sciences (Beijing) at Peking University for assistance with TEM analysis, particularly Dr. Yiqun Liu and Mr. Xinpeng He, for technical help with preparing EM samples and capturing images. Lastly, we thank Dr. Chunyan Zhang and Dr. Yan Yin from the Plant Science Facility of the Institute of Botany, Chinese Academy of Sciences for their technical assistance in measurements of photosynthesis-related parameters. The Chen laboratory is supported by the State Key Laboratory for Protein and Plant Gene Research, Peking-Tsinghua Joint Center for Life Sciences, and Beijing Advanced Center of RNA Biology (BEACON). This work was supported by an NIH grant (GM061146) to X.C. and a grant from National

Natural Science Foundation of China (32170334) to C.Y. and a grant from Guangdong Laboratory for Lingnan Modern Agriculture (NG2022002) to C.Y.

Author contributions

R.H. and X.C. designed the research. R.H. conducted the wet-lab experiments. C.Y., R.H., and H.L. analyzed the sequencing data. R.H., C.Y., M.C., and X.C. interpreted the data. W.L., R.W., and H.H. provided assistance in wet-lab experiments. R.H. wrote the draft. R.H., C.Y., M.C., and X.C. revised the paper. All authors read and revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54194-1>.

Correspondence and requests for materials should be addressed to Runlai Hang, Chenjiang You or Xuemei Chen.

Peer review information *Nature Communications* thanks Pavel Baranov and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024