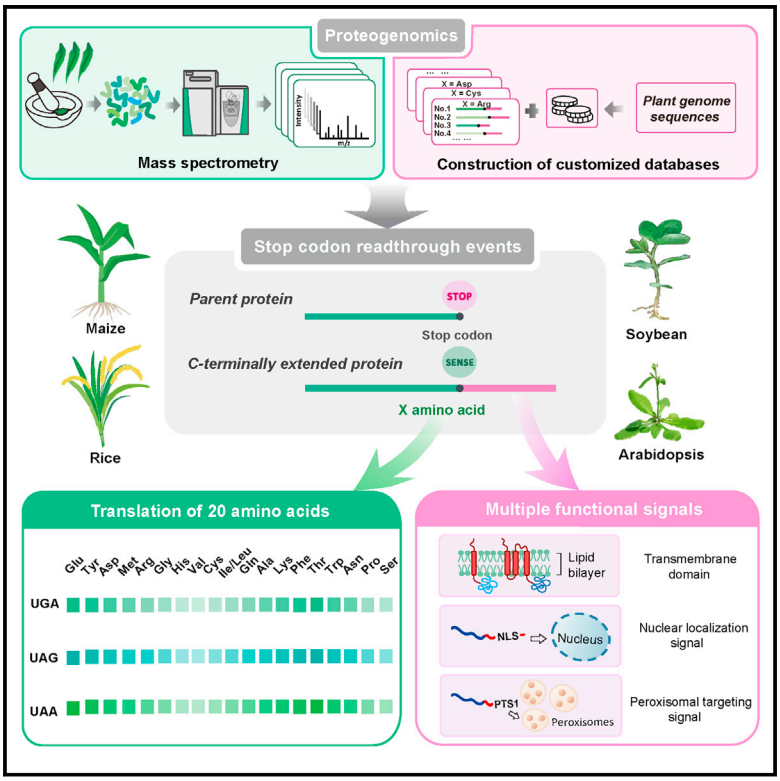
Cell Reports

Readthrough events in plants reveal plasticity of stop codons

Graphical abstract



Highlights

- Stop codon readthrough (SCR) events are commonly present in monocots and dicots
- UGA, UAG, and UAA in plant SCR events are recoded as all 20 standard amino acids
- Transcripts in plant SCR events tend to be shorter and have fewer exons and splice variants
- Plant SCR events are not conserved, and SCR proteins have functional signals

Authors

Yuqian Zhang, Hehuan Li, Yanting Shen, ..., Zhixi Tian, Shu-Biao Wu, Liuji Wu

Correspondence

zxtian@genetics.ac.cn (Z.T.), shubiao.wu@une.edu.au (S.-B.W.), wlj200120@163.com (L.W.)

In brief

Zhang et al. identify 1,009 stop codon readthrough events in four plants using a proteogenomic strategy, which generate C-terminally extended protein isoforms with multiple functional signals. Notably, the finding that three stop codons can be recoded as 20 standard amino acids provides insights into the flexibility of genetic decoding.

Zhang et al., 2024, Cell Reports *43*, 113723
February 27, 2024
https://doi.org/10.1016/j.celrep.2024.113723

CellPress

Cell Reports

CellPress
OPEN ACCESS

Report

Readthrough events in plants reveal plasticity of stop codons

Yuqian Zhang,[1,2,6] Hehuan Li,[1,6] Yanting Shen,[3,6] Shunxi Wang,[1] Lei Tian,[1] Haoqiang Yin,[1] Jiawei Shi,[1] Anqi Xing,[4] Jinghua Zhang,[1] Usman Ali,[1] Abdul Sami,[1] Xueyan Chen,[1] Chenxuan Gao,[1] Yangtao Zhao,[1] Yajing Lyu,[1] Xiaoxu Wang,[1] Yanhui Chen,[1] Zhixi Tian,[3,5,*] Shu-Biao Wu,[2,*] and Liuji Wu[1,2,7,*]

[1]National Key Laboratory of Wheat and Maize Crop Science, Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou 450046, Henan, China

[2]School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

[3]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[4]Department of Plant and Environmental Sciences, Clemson University, Clemson, SC 29634, USA

[5]College of Advanced Agriculture Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: zxtian@genetics.ac.cn (Z.T.), shubiao.wu@une.edu.au (S.-B.W.), wlj200120@163.com (L.W.)

https://doi.org/10.1016/j.celrep.2024.113723

SUMMARY

Stop codon readthrough (SCR) has important biological implications but remains largely uncharacterized. Here, we identify 1,009 SCR events in plants using a proteogenomic strategy. Plant SCR candidates tend to have shorter transcript lengths and fewer exons and splice variants than non-SCR transcripts. Mass spectrometry evidence shows that stop codons involved in SCR events can be recoded as 20 standard amino acids, some of which are also supported by suppressor tRNA analysis. We also observe multiple functional signals in 34 maize extended proteins and characterize the structural and subcellular localization changes in the extended protein of basic transcription factor 3. Furthermore, the SCR events exhibit non-conserved signature, and the extensions likely undergo protein-coding selection. Overall, our study not only characterizes that SCR events are commonly present in plants but also identifies the recoding plasticity of stop codons, which provides important insights into the flexibility of genetic decoding.

INTRODUCTION

The rules of the genetic code were long considered universal dogma in all organisms after the standard genetic code was deciphered decades ago. However, exceptions to the rules have been increasingly observed in different species over the years, substantially altering the perception of the decoding rules from static to dynamic.[1,2] Living organisms are now widely understood to employ various modes to expand the coding potential of their genome, such as alternative translation initiation,[3] ribosome frameshifting,[4] translational bypassing,[5] and downstream open reading frame translation in 3′ untranslated regions (UTRs).[6] Stop codon readthrough (SCR) is another important means of recoding that is also classified as a non-canonical translational event.[7,8] SCR occurs when the ribosome reads a stop codon (UGA, UAG, or UAA) as a sense codon rather than the stop signal with the translation terminating at a later in-frame stop codon.[8,9] This process can generate a C-terminally extended protein isoform that may acquire novel biological functions, since the amino acid sequence appended at the C terminus possibly alters the activity,[10] stability,[11] and/or subcellular localization of the protein.[2,12] Indeed, SCR proteins have been implicated in a variety of biological processes, such as the regulation of intercellular cytoplasm flow,[13] neuronal integrity and reproductive health,[14] cellular redox homeostasis,[2,15] and angiogenesis.[16,17]

Initially discovered in viruses as a strategy to expand protein-coding potential,[18–20] SCR was subsequently identified in other organisms, such as plant pathogens,[2] yeast,[21,22] fruit fly,[23,24] and mammals.[16,25–27] In plants, there was one study that reported 144 readthrough genes in *Arabidopsis* by ribosome profiling analysis.[28] In SCR events, it is crucial to identify the amino acids decoded at the stop codon sites, as they can strongly affect the functionality or stability of readthrough protein isoforms.[12,29,30] Multiple factors have been identified to influence SCR and/or amino acid incorporation at stop codon sites. In addition to *cis* and *trans* elements such as proximal nucleotides, motifs, and particular proteins,[1,9,26] the presence of specific transfer RNAs (tRNAs) constitutes a fundamental factor in facilitating readthrough and amino acid incorporation. Currently, two tRNA species have been identified to participate in SCR: suppressor tRNAs that form conventional base pairing with the

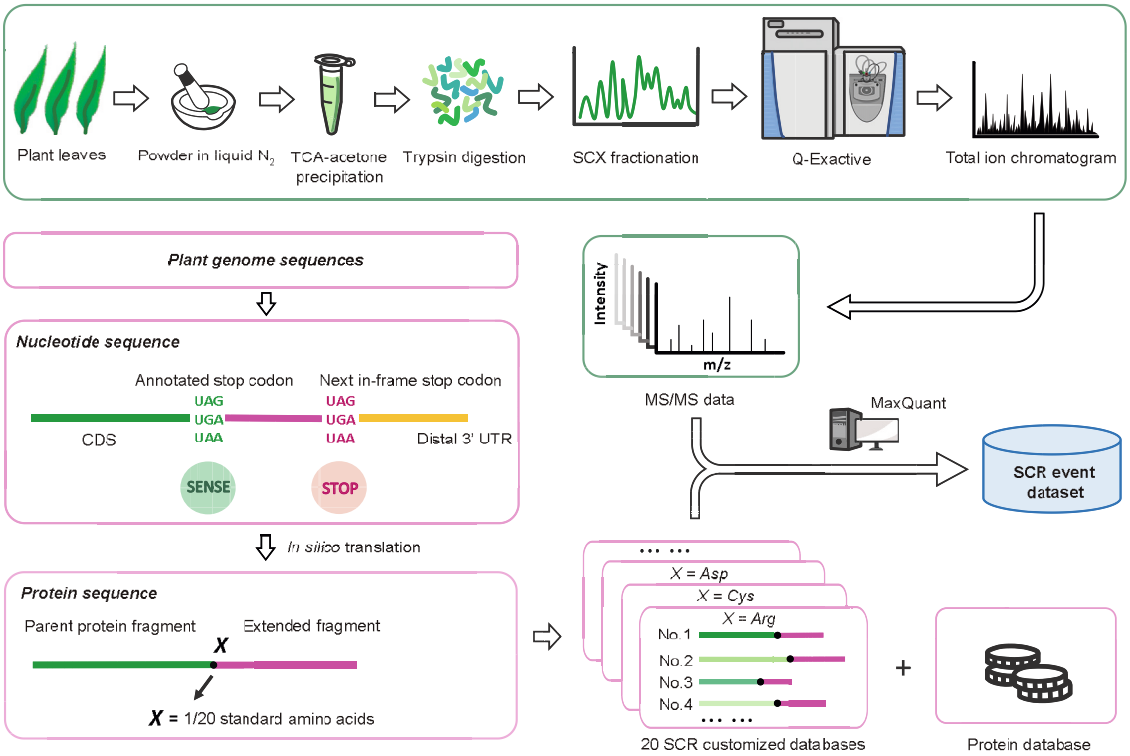


Figure 1. Proteogenomic workflow for identifying plant stop codon readthrough (SCR) events

The strategy consists of two procedures: shotgun proteomics experiment (green box) and customized database construction (pink boxes). Total protein was extracted from plant leaves and subjected to digestion, fragmentation, and LC-MS/MS analysis for identification of spectra datasets. For customized database construction, plant genome sequences were subjected to *in silico* translation. The annotated stop codon sites were replaced with "X," which represented one of the 20 standard amino acids during translation each time. The translation was assigned to terminate at the next in-frame stop codon, producing 20 customized databases containing theoretical readthrough proteins. The distal 3′ UTR represents the portion of the 3′ UTR downstream of the putative readthrough extension. MS/MS spectra datasets were searched against the 20 customized databases merged with the public plant protein database by MaxQuant. We defined the MS/MS-identified peptides that exclusively spanned the annotated stop codon sites as "through" peptides. A merged SCR protein group that contained proteins sharing one common stop codon site spanned by a detected through peptide was defined as an SCR event.

stop codon at all three nucleotide positions of the anticodon and the near-cognate tRNAs with unconventional base pairing at a single position of the anticodon with the stop codon.[31–34] Previous studies have reported a portion of amino acids that could be specified by the stop codons (see Table S1 for details). However, knowledge about the amino acids translated at stop codon sites in SCR events remains unknown in Plantae. In addition, the potential of stop codons to be recoded as amino acids other than those reported has not been investigated yet.

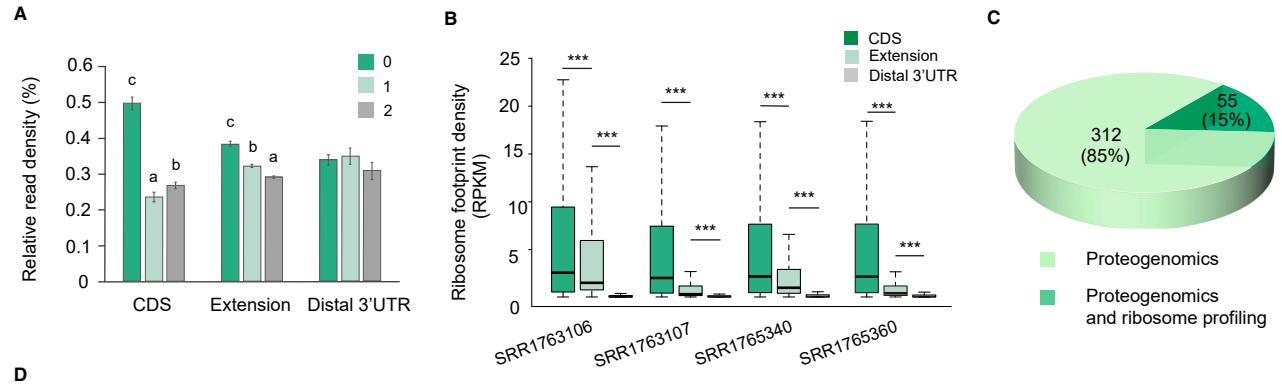
Here, we employed a mass spectrometry (MS)-based proteogenomic strategy to study SCR events in different plants and identify the amino acids at the readthrough stop codon sites. We detected a total of 1,009 SCR events in the monocots maize and rice and dicots soybean and *Arabidopsis* and characterized their features. Interestingly, we discovered that the three stop codons could be recoded as all 20 standard amino acids, which reveals the unprecedented recoding plasticity of stop codons and the dynamic nature of genetic decoding. We also delineated a comprehensive landscape of SCR events in plants at the protein level, uncovering a collection of previously unknown readthrough proteins to further explore their biological functions.

RESULTS

A proteogenomic strategy to identify SCR events in plants

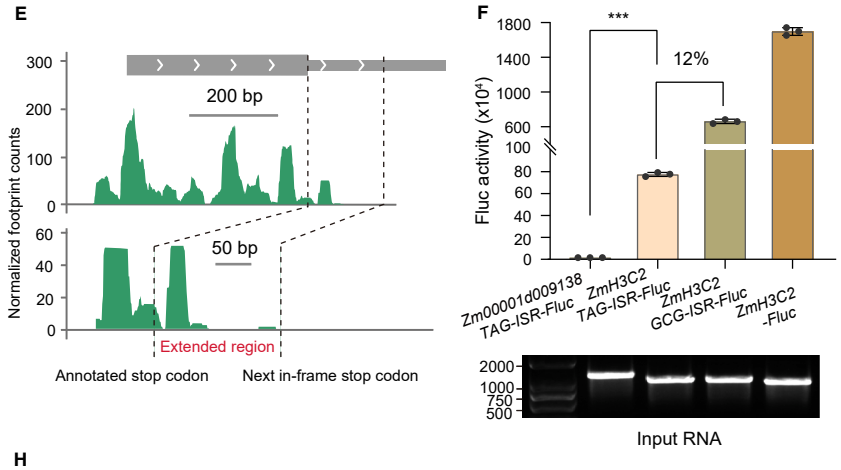
The SCR event is a means of alternative translation by redefining UGA, UAG, or UAA to specify amino acids, thus extending the translation at the C terminus. The most definitive evidence for SCR occurrence is the direct detection of translation at the stop codon sites. However, SCR events are non-canonical and cannot be identified by a typical proteomic strategy. Therefore, we implemented a proteogenomic strategy integrating MS-based proteomics with 20 customized databases to identify SCR events in maize at the protein level (Figure 1). First, total protein was extracted from maize leaves, followed by digestion, fractionation, and liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis to obtain MS/MS datasets of peptides for subsequent identification. We then constructed 20 customized databases based on the assumption that the annotated stop codons of genes were readable as sense codons and could be decoded as one of the 20 standard amino acids. Each of the 20 customized databases therefore contained the

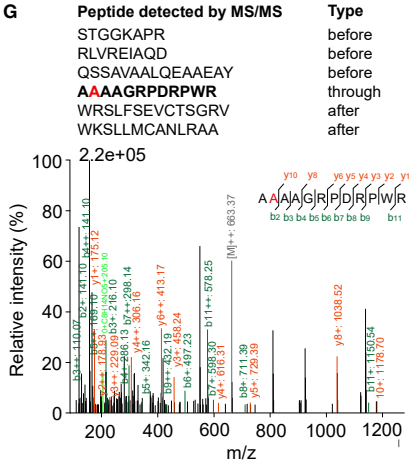
CellPress
OPEN ACCESS



D

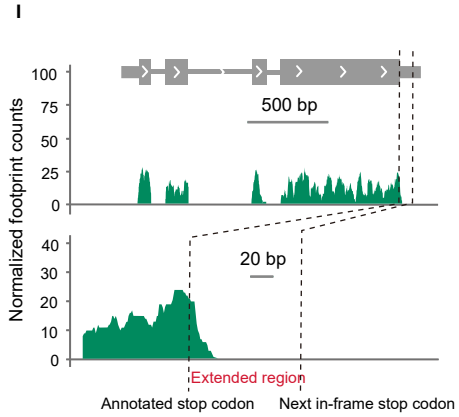
M A R T K Q T A R K S T G G K A P R K Q L A T K A A R K S A P A T G G V K K P H R F R P G T V A L R
E I R K Y Q K S T E L L I R K L P F Q R L V R E I A Q D F K T D L R F Q S S A V A A L Q E A A E A Y
L V G L F E D T N L C A I H A K R V T I M P K D I Q L A R R I R G E R **A A A A G R P D R P W R** S L F
S E V C T S G R V V W R E W C T N Q Q C L K P C Q P L C C H Q W K S L L M C A N L R A A F V S

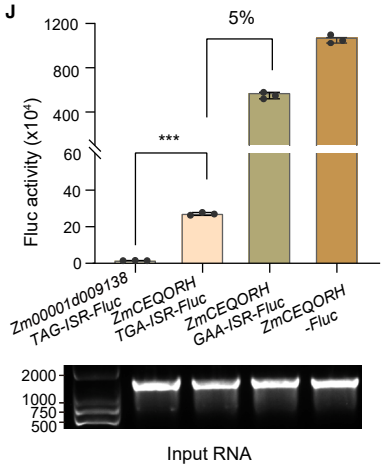


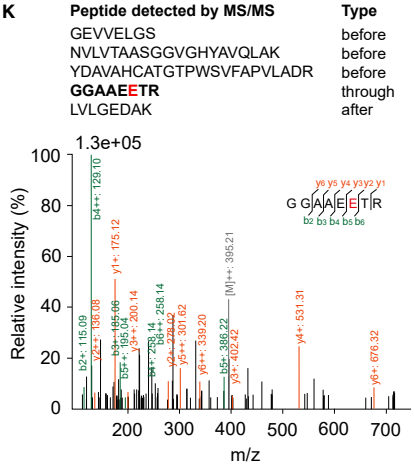


H

M A A P K K T M R A L R Y D K Y G G G A E G L K H V E V P V P S P S K G E V L L K L
E A A S I N P I D W K I Q K G M V R P F L P R K F P F I P V G D M S G E V V E L G S
G V T D F K P G D K V I S I S F P N G G G L A E Y A V A P A S L T V A R P P E V S A A
E G A S L P T A A S T A L Q Q L K A A G V S S F D G G S G S G S G S G S G A D S D G P K
N V L V T A A S G G V G H Y A V Q L A K L A G L R V T A T C G A R N L G L V A G L G
A D E V L D Y K T P E G A R L L S P S G T R Y D A V A H C A T G T P W S V F A P V L
A D R A T V V D V T P G V A A A A K S F L Q K V T F S R K R L V P L V L I P K K E E M
E W L V E M T R Q G K L R T V V D S R Y S L S R **A Q E V W A K S I D G H A T G K I V**
V E M G G A A E E T R Q A G M R A H C V G L F M D C S L C S I I A T A I H N K P S I







(legend on next page)

in silico-translated protein isoforms whose annotated stop codon sites were replaced by one of the 20 standard amino acids, and the protein sequences were extended to the subsequent in-frame stop codon.

Next, we used the obtained MS/MS data to search against the 20 SCR customized databases with the MaxQuant search engine. We defined a peptide that spanned the position of the stop codon site in MS identification as a "through" peptide (i.e., "SCR" peptide) and those proteins that contained a unique through peptide with high confidence as SCR proteins. A merged SCR protein group that contained proteins sharing one common stop codon site spanned by a detected through peptide was defined as an SCR event. Accordingly, one SCR event contains one or more SCR proteins based on the isoform annotation of the gene. The strategy identified 708 unique through peptides. For further quality control of MS data, we next applied more stringent criteria for through peptides to identify SCR events and proteins with high confidence (see STAR Methods). The screening steps resulted in much more reliable "through" peptides, which corresponded to 367 high-confidence SCR events involving 1,063 SCR proteins (Table S2).

The overall SCR frequency at the genomic level in maize was 3.2%, calculated by dividing the number of SCR events by the total number of SCR events and non-SCR events. All identified peptides for SCR proteins are listed in Table S3. Previous studies have shown that SCR is mRNA context dependent, which can be influenced by multiple *cis* elements, such as the stop codon identity, nucleotides immediately surrounding the stop codon (typically from −2 to +9 position, and the first nucleotide of stop codon is assigned +1), and downstream secondary structures.[1,8,9,23] Based on these factors, those different mRNA isoforms that share one common readthrough stop codon site and have a similar sequence context around the stop codon may undergo SCR events. Collectively, these results not only demonstrate the abundance of readthrough events naturally occurring in maize but also uncover a previously unknown source of protein isoforms produced by SCR.

Verifying SCR events in maize

To verify the SCR events in maize identified above, we retrieved the available ribosome profiling (Ribo-seq) datasets of 56 maize samples from NCBI to analyze the SCR events at the translation level. Four Ribo-seq datasets were finally retained after further screening for their high quality (see STAR Methods), as was also indicated by the clear three-nucleotide (3-nt) periodicity observed in the coding sequences (CDSs) (Figure 2A). In addition to using the methods applied in the studies on SCR events,[7,12] we had attempted to use another method for Ribo-seq analysis, and no visible 3-nt periodicity was observed in the CDSs, suggesting that the datasets may apply to specific methods. RNA-seq analyses in the four datasets showed that the average expression level of transcripts in SCR events was significantly lower than that of non-SCR events ($p < 0.001$, Figure S1), revealing a different expression pattern of SCR and non-SCR genes. The expression levels of transcripts involved in SCR events are listed in Table S4. We then analyzed the 3-nt periodicity of the Ribo-seq reads (30-mers) in putative extensions and distal 3′ UTRs. The results showed that the putative extensions exhibited a visible hallmark of 3-nt periodicity ($p < 0.05$), which is consistent with that of the CDSs ($p < 0.05$, Figure 2A). By contrast, such difference was not observed in the distal 3′ UTRs. More importantly, the major component of the 3-nt periodicity in putative extensions occurs in the same reading frame as that in the coding regions (Figure 2A). In addition, the average ribosome footprint density (in RPKM) of putative extensions was significantly higher than that of distal 3′ UTRs in all four datasets ($p < 0.001$, Figure 2B), although it was still lower than that of the CDSs ($p < 0.001$). These results collectively suggest the presence of coding potential within the putative extensions, and readthrough (as opposed to, for example, frameshifting) is the major contributor to the accumulation of ribosome footprints in the putative extensions. After applying the conditions described in the STAR Methods for the four Ribo-seq datasets to verify SCR, we found that 55 out of the 367 SCR events identified by

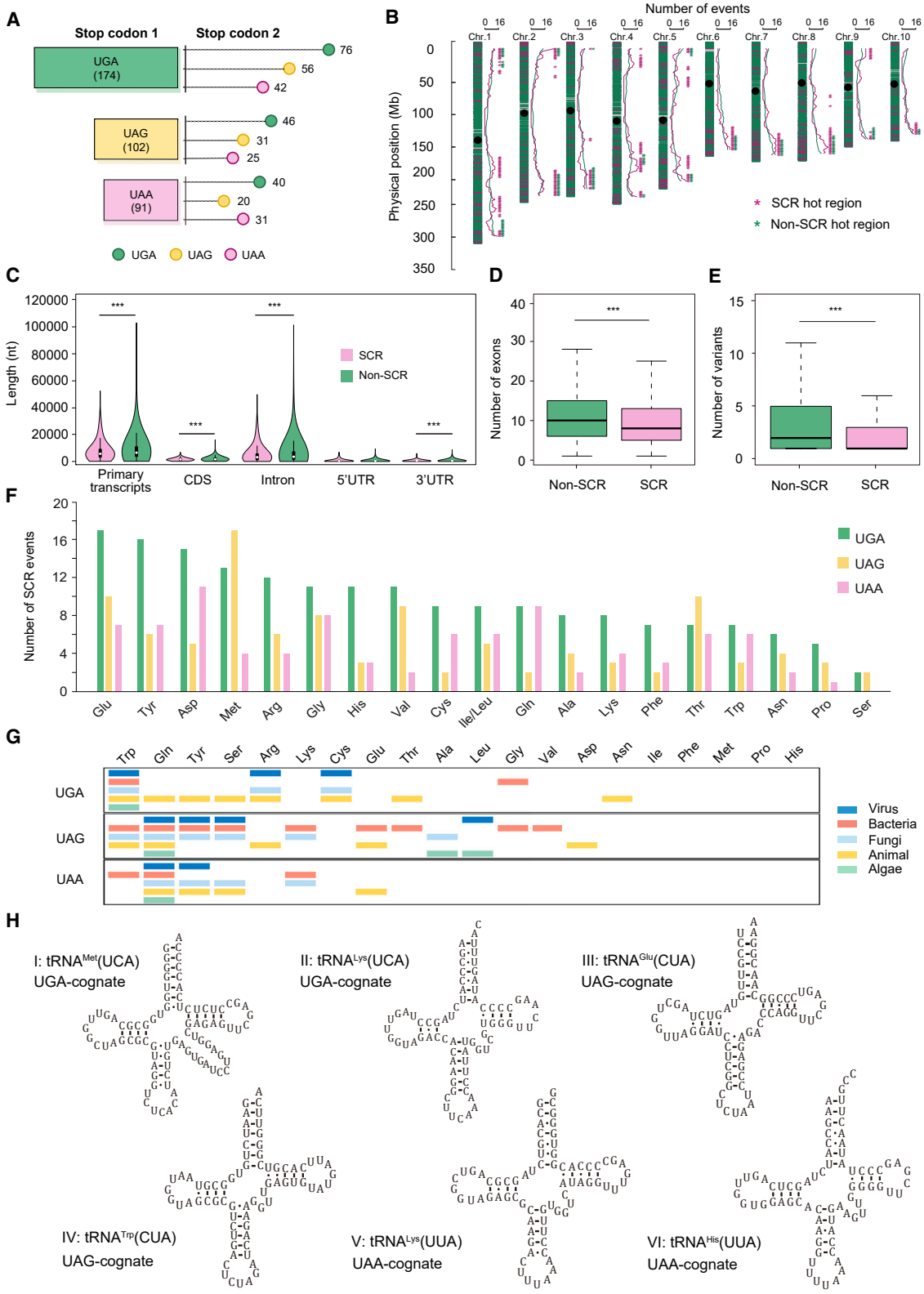
Figure 2. Verifying SCR events in maize

(A) Three-nucleotide periodicity of 30-mers in the CDSs, putative extensions, and the distal 3′ UTRs in the Ribo-seq datasets of the four samples. Data are mean ± SD. The error bar is the standard deviation of the average read ratio among samples of the four Ribo-seq datasets. Different letters indicate significant differences at $p < 0.05$ by one-way ANOVA with Tukey's multiple comparisons test. Green, light green, and gray bars represent frame 0, frame 1, and frame 2, respectively.

(B) Distribution of ribosome footprint density in CDSs, putative extensions, and the distal 3′ UTRs in the four Ribo-seq datasets. Distal 3′ UTR was defined as the region 40 codons after the next stop codon or the region from the next stop codon to the third stop codon within 40 codons. Box edges represent the 0.25 and 0.75 quantiles. The bold lines within boxes represent the median, and whiskers represent the 1.5 × interquartile range (IQR). Statistical significance was determined using a Wilcoxon signed rank test, ***$p < 0.001$.

(C) The proportion of SCR events confirmed by ribosome profiling among those identified by proteogenomics.

(D–K) Validation of two readthrough examples, *ZmH3C2* (D–G) and *ZmCEQORH* (H–K). (D and H) The complete protein sequences with the extended region to the next in-frame stop codon of the two genes. The underlined peptides were identified by mass spectrometry analysis, and the through peptides are in bold. The amino acids translated at the stop codon sites in through peptides are highlighted in red, which are alanine [A] at UAG in (D) and glutamate [E] at UGA in (H). (E and I) Footprint signals in the C-terminal extensions of the two transcripts revealed by Ribo-seq analysis. Dashed lines represent the annotated stop codon and the next in-frame stop codon. (F and J) Readthrough activities of *ZmH3C2* and *ZmCEQORH* measured by the *in vitro* translation experiment. Plasmids containing test cDNA-stop codon-ISR-*Fluc* were firstly transcribed *in vitro*, followed by *in vitro* translation using wheat germ extract. Luciferase activities were used to represent the readthrough activities. Constructs with the stop codon mutated to a sense codon were used as a positive control to measure the efficiency of SCR. The non-SCR gene *Zm00001d009138* with its TAG and ISR sequence fused with *Fluc* (*Zm00001d009138-TAG-ISR-Fluc*) was used as the SCR-negative control, as it showed no SCR sign in either proteogenomics or Ribo-seq analysis. All samples were tested in biological triplicates. Data are mean ± SEM (n = 3). Asterisks indicate statistical significance by two-sided Student's t test (***$p < 0.001$). ISR, inter-stop codon region (extended region). (G and K) Different types of peptides labeled as "before," through (bold), and "after" identified by MS/MS (top) and the MS/MS spectra (bottom) of the identified through peptides AAAAGRPDRPWR for *ZmH3C2* (G) and GGAAEETR for *ZmCEQORH* (K) readthrough products expressed from (F) and (J), respectively. See also Figure S1 and Tables S2, S3, S4, S5, and S12.



(legend on next page)

proteogenomic approach were also corroborated by Ribo-seq analysis (Figure 2C; Table S5).

We selected two SCR representative genes, *ZmH3C2* and *ZmCEQORH*, which exerted functional roles in distinct biological processes related to chromatin assembly and stress response, respectively, for further analysis. For *ZmH3C2*, which encodes HISTONE H3.2, MS analysis identified the readthrough peptide AAAAGRPDRPWR with its second alanine (A) translated at the UAG site (Figure 2D). The readthrough of this gene was also evidenced by ribosome footprints within the extended region downstream of the annotated stop codon (Figure 2E). For *ZmCEQORH*, encoding CHLOROPLAST ENVELOPE QUINONE OXIDOREDUCTASE, MS analysis identified the through peptide AQEVWAKSIDGHATGKIVVEMGGAAEETR, in which the antepenultimate glutamate (E) was incorporated at the UGA site (Figure 2H). We also observed ribosome footprints beyond the annotated stop codon of the corresponding transcript (Figure 2I).

Ribo-seq analysis provided readthrough clues for the two genes. However, considering the complexity of ribosomal footprint interpretation and low abundance of readthrough,[12,35] we next performed *in vitro* translation experiments using wheat germ extract to further demonstrate the readthrough of the above two genes. The results showed that *ZmH3C2* and *ZmCEQORH* had significantly higher readthrough activities compared to the SCR-negative gene *Zm00001d009138* ($p < 0.001$), as demonstrated by the detection of luciferase activities. The *in vitro* SCR rates of *ZmH3C2* and *ZmCEQORH* were 12% and 5%, respectively (Figures 2F and 2J). The much higher readthrough rates of the two genes than the basal translation error (<0.1%)[9] evidenced they underwent programmed SCR events. Further MS analysis identified different types of peptides, including STGGKAPR, and RLVREIAQD, QSSAVAALQEAAEAY labeled as before, AAAAGRPDRPWR as through, as well as WRSLFSEVCTSGRV and WKSLLMCANLRAA as after, were detected in the readthrough protein of *ZmH3C2* (Figure 2G, top). In addition, peptides including GEVVELGS, NVLVTAASGGVGHYAVQLAK, and YDAVAHCATGTPWSVFAPVLADR labeled as before, GGAAEETR as through, and LVLGEDAK as after were identified in the readthrough protein of *ZmCEQORH* (Figure 2K, top). Among them, the through peptides AAAAGRPDRPWR and GGAAEETR for the two genes were detected, where the UAG and UGA were recoded as alanine and glutamate (Figures 2G and 2K, bottom), respectively. Taken together, the Ribo-seq analysis, *in vitro* translation experiments, and MS analysis validated the SCR events identified using proteogenomics in maize and demonstrated the feasibility of the proteogenomic method for large-scale identification of SCR events in plants.

Characteristics of SCR events and recoding plasticity of stop codons in maize

To understand the characteristics of SCR events in maize, we first analyzed the stop codon preferences. All three stop codons were readthrough among the 367 SCR events: UGA was involved in 174 events (47.4%), followed by UAG in 102 events (27.8%) and UAA in the remaining 91 events (24.8%) (Figure 3A). We found that UGA was the most frequent next in-frame stop codon following these readthrough stop codons, accounting for 76 cases after the readthrough UGA, 46 cases after the readthrough UAG, and 40 cases after the readthrough UAA (Figure 3A). The genome-wide distribution pattern showed that these SCR events were distributed heterogeneously across each chromosome of maize, and many of the SCR events were closer to the telomeres (Figure 3B). In addition, a collection of 379 hotspot regions were obtained, 244 of which were SCR hotspot regions (Figure 3B). These results revealed the wide distribution of SCR events in the maize genome.

Furthermore, the length and complexity of transcripts in SCR and non-SCR events were compared. We designated transcripts involved in SCR events as SCR candidates, and transcripts whose proteins were solely detected with before peptides (peptides mapped within the protein sequence of the coding regions) were defined as non-SCR transcripts. Interestingly, we observed that the lengths of primary transcripts for SCR candidates were significantly shorter than the non-SCR transcripts (Figure 3C). The average lengths of the elements from the primary transcripts, including CDSs, introns, and 3′ UTRs, were also markedly shorter in SCR events than those in non-SCR events (Figure 3C). Moreover, SCR candidates had significantly fewer exons (Figure 3D) and splice variants (Figure 3E) than those of non-SCR counterparts. This consistency is expected, as there was a positive correlation

Figure 3. Characteristics of SCR events and recoding plasticity of stop codons in maize

(A) Frequency of the readthrough stop codon and the next in-frame stop codon in SCR events. Stop codon 1, readthrough stop codon; Stop codon 2, the next in-frame stop codon.

(B) Genome-wide distribution of SCR events (pink) and non-SCR events (green) in the chromosomes of maize. For each chromosome, the distribution contains three columns. Left: SCR events (pink) and non-SCR events (green) across the chromosomes. Black dots: centromeres. Middle: curves showing the distribution patterns of SCR events (pink) and non-SCR events (green), using a 20-Mb sliding window with a 2-Mb step based on the B73 reference genome. Right: distribution of hotspot regions for SCR events (pink) and non-SCR events (green) shown with asterisks. We defined hotspot regions as $\geq$5 SCR events and $\geq$335 non-SCR events in a 20-Mb window.

(C) Violin plot describing the length distribution of primary transcripts and transcript elements in SCR and non-SCR events. Internal boxplots show the median values (white dots in the center) and the interquartile range (IQR; 0.25 and 0.75 quantiles) of the length distribution.

(D) Number of exons in SCR and non-SCR primary transcripts.

(E) Number of splice variants of the primary transcripts in SCR and non-SCR events. For (C)–(E), ***$p < 0.001$, two-sided Student's t test. For all boxplots, box edges represent the 0.25 and 0.75 quantiles. The bold lines represent the median values, and the whiskers represent 1.5 × IQR.

(F) Frequency of amino acids specified by the three stop codons.

(G) Previously reported amino acids translated at UGA, UAG, and UAA in different species. Blue, virus; orange, bacteria; light blue, fungi; yellow, animal; green, algae.

(H) Predicted secondary structures of example suppressor tRNAs for UGA, UAG, and UAA. I–VI are $tRNA^{Met}$ (UCA) and $tRNA^{Lys}$ (UCA) for UGA, $tRNA^{Glu}$ (CUA) and $tRNA^{Trp}$ (CUA) for UAG, and $tRNA^{Lys}$ (UUA) and $tRNA^{His}$ (UUA) for UAA, respectively. See also Figures S2, S3, S4, S5, S6, S7, and S8, Tables S1, S2, and S6, and Data S1.

between exon numbers and the number of splice variants in maize (Pearson's $r = 0.944$, $p < 0.01$). In addition, we found that the factor exon number (regression coefficient = −1.710e–02, $p < 0.05$) made a more significant contribution to the correlation with SCR compared to the primary transcript length (regression coefficient = −1.204e–05, $p < 0.05$) by using the logistic regression analysis. Furthermore, we explored the motifs in the flanking sequences downstream of the readthrough stop codons. The results showed that the extended regions downstream of the annotated stop codons in the SCR candidates were enriched with featured sequence motifs (Data S1). These motifs enriched in extended regions likely exert effects on the particular facets of SCR by interacting with specific proteins.

Since translation of the stop codon is crucial for SCR events, we examined which amino acids were translated at every readthrough stop codon site based on the MS data. We found that with the exception of UAA, which was recoded to 19 amino acids (except serine), UGA and UAG were translated to all 20 standard amino acids in maize SCR events (Figure 3F). Among all the recoded amino acids, glutamate was the most recoded amino acid for UGA (in 17 SCR events, 8.9%), methionine for UAG (in 17 cases, 15.6%), and aspartate for UAA (in 11 cases, 11.6%). The result is quite unexpected since previous literature has reported only several amino acids that were able to be translated at stop codon sites in different species (Figure 3G; Table S1). The remarkable diversity in amino acid translation at stop codon sites has not been documented in other species, prompting us to explore the reasons behind the recoding plasticity of the stop codons.

Suppressor tRNAs, the specific tRNA species that can form conventional base pairing with the stop codon in all three nucleotide positions of the anticodon, have been identified in different eukaryotes leading to readthrough.[31,36] Suppressor tRNAs can be predicted through genome-wide searches followed by sequence alignments with canonical tRNAs from databases to confirm their identity.[37] We thus conducted a genome-wide search for potential suppressor tRNAs capable of reading stop codons in maize by running tRNAscan-SE and ARAGORN.[38,39] As a result, we acquired a collection of UGA-, UAG-, and UAA-recognizing tRNAs. Multiple sequence alignments of the predicted suppressor tRNAs against the NCBI Nucleotide (NT) database resulted in different groups of suppressor tRNAs well aligned with the canonical tRNAs of a range of amino acids (Table S6). We identified six classes of UGA suppressor tRNAs (Tables S1 and S6): $tRNA^{Met}$(UCA), $tRNA^{Lys}$(UCA), $tRNA^{Val}$(UCA), $tRNA^{Gly}$(UCA), $tRNA^{Ile}$(UCA), and $tRNA^{Thr}$(UCA). The predicted secondary structures of the first two UGA suppressor tRNAs are shown in Figures 3H-I and 3H-II, and their nucleotide sequence alignments are presented in Figures S2A and S2B. We also detected six classes of UAG suppressor tRNAs (Tables S1 and S6): $tRNA^{Glu}$(CUA) (Figures 3H-III and S3A), $tRNA^{Trp}$(CUA) (Figures 3H-IV and S3B), $tRNA^{Arg}$(CUA), $tRNA^{Asp}$(CUA), $tRNA^{Gly}$(CUA), and $tRNA^{Lys}$(CUA). Furthermore, we identified eight classes of UAA suppressor tRNAs (Tables S1 and S6): $tRNA^{Lys}$(UUA) (Figures 3H-V and S2C), $tRNA^{His}$(UUA) (Figures 3H-VI and S2D), $tRNA^{Ala}$(UUA), $tRNA^{Asn}$(UUA), $tRNA^{Gly}$(UUA), $tRNA^{Phe}$(UUA), $tRNA^{Met}$(UUA), and $tRNA^{Thr}$(UUA). Collectively, the suppressor tRNAs identified in maize provide potential reasons for the translation of at least some amino acids at stop codon sites.

SCR events and recoding plasticity of stop codons in other plants

In light of the abundant SCR events identified in maize, we questioned if SCR also occurred in other plant species. We thus followed the same proteogenomic method with the respective 20 custom databases to explore SCR events in rice, soybean, and *Arabidopsis*. Indeed, we detected abundant SCR events in these species with similar features to those in maize. We identified 198 SCR events in rice, 179 SCR events in soybean, and 265 SCR events in *Arabidopsis* (Table S7), corresponding to 269, 385, and 468 proteins, respectively. The overall SCR frequency at the genomic level was 2.1% in rice, 1.7% in soybean, and 2.9% in *Arabidopsis*. We also analyzed the available Ribo-seq datasets for rice and *Arabidopsis* using the same pipeline as in maize to further verify SCR events at the translation level. The results showed that of the 198 SCR events identified by MS in rice, 45 were further corroborated by the Ribo-seq analysis (Table S8). In the case of *Arabidopsis*, 65 out of the 265 MS-detected SCR events were verified by the Ribo-seq evidence (Table S8). In addition, seven of the MS-identified SCR events in *Arabidopsis* were reported by the previous study of Sahoo et al. (Table S9).[28]

In rice, we observed that the stop codon usage in SCR events was in the order of UGA > UAA > UAG, and it was UGA > UAA > UAG in both soybean and *Arabidopsis* (Figures 4A–4C). The next in-frame stop codon following the readthrough stop codons also displayed a preference for UGA in rice and soybean, and UAA for *Arabidopsis* (Figures 4A–4C). Genome-wide distribution analysis showed that SCR events were distributed across all chromosomes in their respective plant species (Figures 4D–4F). We also analyzed the features of SCR candidates in the three plants. In rice, the lengths of primary transcripts and transcript elements including CDSs, introns, and the 5′ UTR of SCR candidates were shorter than those of non-SCRs ($p < 0.05$, Figures S4A and S4B). In addition, the SCR candidates tended to possess fewer exons (Figure S4C) and had significantly fewer splice variants ($p < 0.001$, Figure S4D) than non-SCR transcripts. In the dicots soybean and *Arabidopsis*, these transcript features of SCR candidates showed a similar trend, though they were not as significant as that in the monocots (Figures S4E–S4L).

Analysis of the recoded amino acids at stop codon sites revealed that UGA exhibited great plasticity to be recoded as the 20 standard amino acids in all three plant species (Figures 4G–4I). For UAG, it was recoded to 18 amino acids (except serine and tryptophan) in rice (Figure 4G), 17 amino acids (except cystine, aspartate, and phenylalanine) in soybean (Figure 4H), and 18 amino acids (except aspartate and arginine) in *Arabidopsis* (Figure 4I). Similarly, UAA was translated to 17, 16, and 20 amino acids in rice, soybean, and *Arabidopsis*, respectively (Figures 4G–4I). These results further uncover the high recoding plasticity of stop codons in plant SCR events.

Likewise, we conducted the prediction of suppressor tRNAs in these three plant species. The results showed that five classes of UGA suppressor tRNAs were predicted in rice (Table S10), including $tRNA^{Gly}$(UCA), $tRNA^{Ile}$(UCA), $tRNA^{Lys}$(UCA), $tRNA^{Met}$(UCA), and $tRNA^{Thr}$(UCA). In soybean, we predicted six classes of UGA suppressor tRNAs (Table S10), namely, $tRNA^{Asp}$(UCA), $tRNA^{Glu}$(UCA), $tRNA^{Lys}$(UCA), $tRNA^{Met}$(UCA), $tRNA^{Trp}$(UCA),

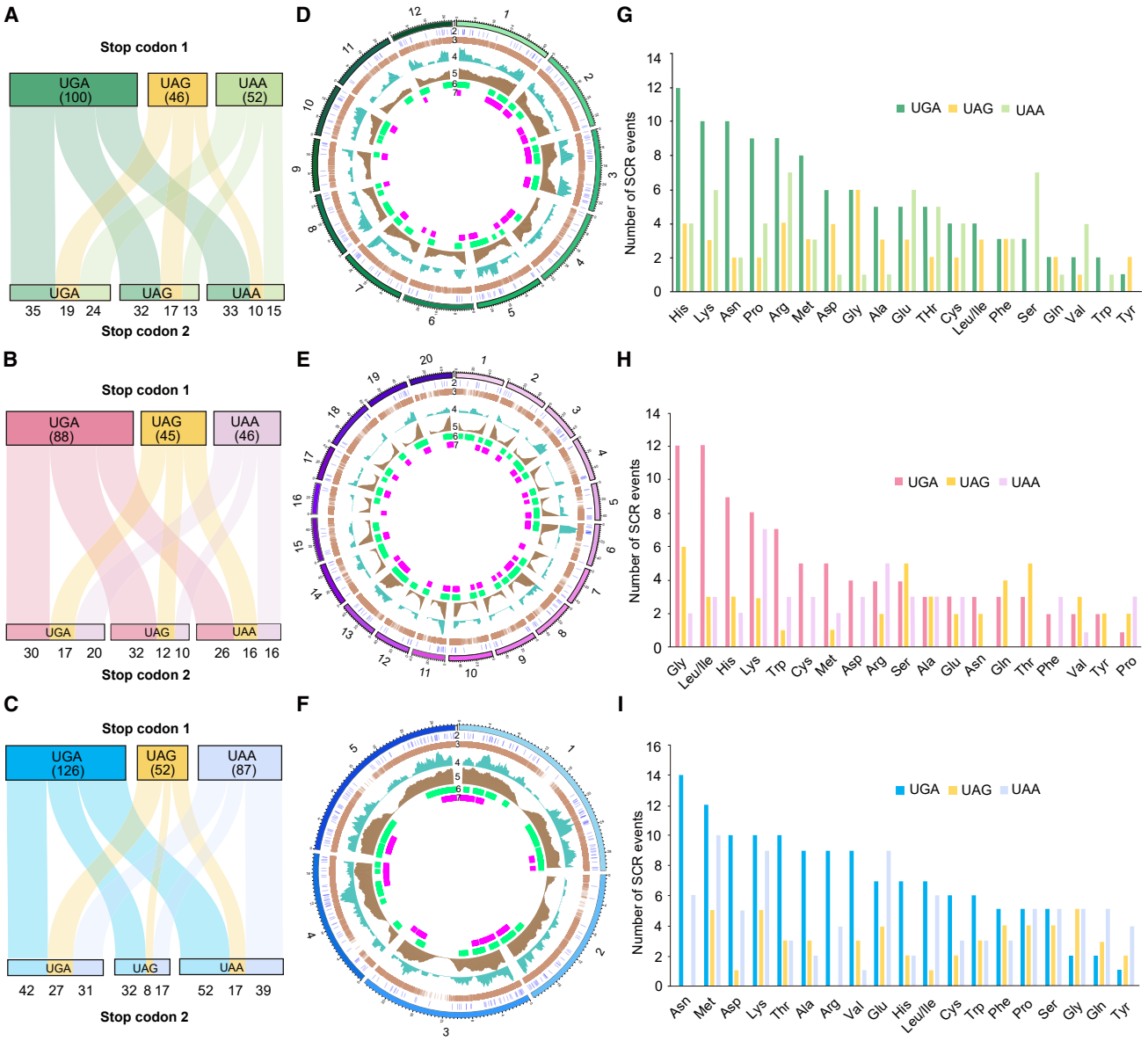


Figure 4. SCR events and recoding plasticity of stop codons in other plants

(A–C) Frequency of the readthrough stop codon and the next in-frame stop codon of SCR events in rice (A), soybean (B), and *Arabidopsis* (C). Stop codon 1, the readthrough stop codon; Stop codon 2, the next in-frame stop codon.

(D–F) Circos plots showing the genome-wide distribution of SCR and non-SCR events in rice (D), soybean (E), and *Arabidopsis* (F). Each Circos plot consists of seven layers. 1, chromosomes; 2–3, distribution patterns of SCR events and non-SCR events across the chromosomes; 4–5, distribution patterns of SCR events and non-SCR events determined using a 4-Mb window size with a step of 400 kb based on the Nipponbare reference genome for rice (D), a 10-Mb window with a 1-Mb step based on the ZH13 reference genome (E), and a 1.2-Mb window with a 100-kb step based on the Columbia-0 reference genome (F); 6–7, distribution of hotspot regions for SCR and non-SCR events. Hotspot regions were defined as $\geq$3 SCR events and $\geq$178 non-SCR events in a 4-Mb window for rice, $\geq$2 SCR events and $\geq$253 non-SCR events in a 10-Mb window for soybean, and $\geq$4 SCR events and $\geq$157 for non-SCR events in a 1.2-Mb window for *Arabidopsis*.

(G–I) Frequency of amino acids specified by the three stop codons in rice (G), soybean (H), and *Arabidopsis* (I) SCR events. The x axis represents the 20 standard amino acids. The y axis is the number of SCR events. See also Figures S4, S5, S6, S7, and S8 and Tables S7, S8, S9, and S10.

and $tRNA^{Val}$(UCA). Five classes of UAA suppressor tRNAs including $tRNA^{Arg}$(UUA), $tRNA^{Asn}$(UUA), $tRNA^{Leu}$(UUA), $tRNA^{Lys}$(UUA), and $tRNA^{Trp}$(UUA) were also predicted. In *Arabidopsis*, four UGA suppressor tRNA species, containing $tRNA^{Gly}$(UCA), $tRNA^{Lys}$(UCA), $tRNA^{Met}$(UCA), and $tRNA^{Thr}$(UCA), as well as two classes of UAA suppressor tRNAs, including $tRNA^{Met}$(UUA) and $tRNA^{Asn}$(UUA), were identified (Table S10). By comparing the nucleotide sequences of two classes of suppressor tRNA species that were predicted in all four plant species, including UGA suppressor tRNAs for methionine and lysine, we found that they had a high degree of similarity within the alignments (Figure S5). However, due to the very low abundance of

CellPress
OPEN ACCESS

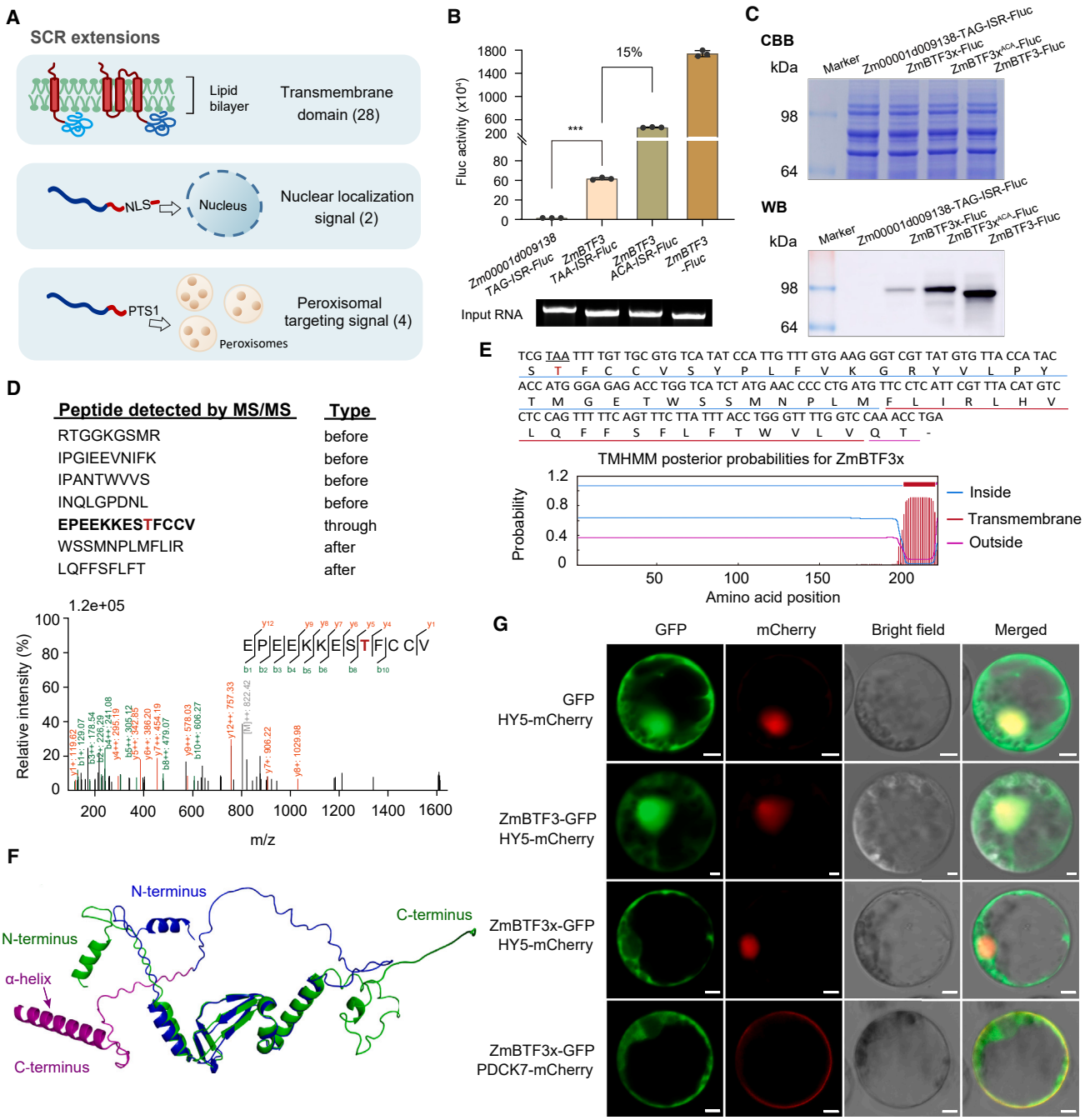


Figure 5. C-terminal extensions in maize SCR events contain multiple functional signals

(A) Multiple functional signals predicted in the extensions of SCR events.

(B) Readthrough validation of *ZmBTF3* using the *in vitro* translation experiment. The plasmid containing *ZmBTF3-TAA-ISR-Fluc* was firstly transcribed *in vitro*, followed by *in vitro* translation using wheat germ extract. *ZmBTF3* with TAA mutated to ACA was used as the positive control to measure the efficiency of SCR. The *Zm00001d009138-TAG-ISR-Fluc* was used as the SCR-negative control. Luciferase activity reflects the readthrough activity. All experiments were performed in triplicates. Data are mean ± SEM (n = 3). ***$p < 0.001$, two-sided Student's t test. ISR, inter-stop codon region.

(C) SDS-PAGE and immunoblot analysis of the ZmBTF3x-Fluc (ZmBTF3-TAA-ISR-Fluc), ZmBTF3x^{ACA}-Fluc (ZmBTF3-ACA-ISR-Fluc), and ZmBTF3-Fluc fusion proteins. Top: Coomassie brilliant blue (CBB)-stained gel of expressed products from (B). Bottom: immunoblot of the ZmBTF3x-Fluc, ZmBTF3x^{ACA}-Fluc, and ZmBTF3-Fluc fusion proteins using an anti-Fluc antibody. The ZmBTF3x^{ACA}-Fluc was used as the positive control. The Zm00001d009138-TAG-ISR-Fluc was used as the negative control.

(D) MS/MS analysis of the digested peptides of ZmBTF3x-Fluc fusion protein excised from the CBB gel in (C). Top: sequences of MS-detected peptides labeled as before, through, and after. Bottom: the MS/MS spectrum of the through peptide EPEEKKESTFCCV. UAA was recoded as threonine [T], shown in red.

(legend continued on next page)

suppressor tRNA species, the experimental validation of suppressor tRNA expression is still challenging in the current technical situation.

In addition to tRNAs, we further explored other potential factors associated with amino acid translation at stop codon sites in these plant species. Recent studies have revealed that the nucleotide immediately following the stop codon can exert an influence on the amino acid incorporation at this site.[40–42] We calculated the frequencies of the subsequent nucleotide of those readthrough stop codons with the most frequently translated amino acids in the four species. The results showed a prevailing presence of uracil or adenine following these stop codons across all four species. Notably, in the monocots, more uracil was observed as the subsequent nucleotide after the readthrough stop codons, whereas adenine was most frequent at this position in dicots (Figure S6).

C-terminal extensions in maize SCR events contain multiple functional signals

To examine the biological implications of SCR events, we searched the functional signals in all C-terminal extensions of readthrough proteins in maize. As a result, multiple functional signals were detected in 34 extensions, including the transmembrane domains in 28 extensions, nuclear localization signals in two extensions, and peroxisomal targeting signals (PTS1) in four extensions (Figure 5A; Table S11). Among them, we selected basic transcription factor 3 (ZmBTF3), which was predicted to possess a transmembrane domain in its C-terminal extension (54 amino acids), for further analysis. The *in vivo* SCR rate of this gene was 7%, as calculated by dividing the average intensity of the through peptide by the sum of the average intensity of detected before peptides and the through peptide. Next, luminescence-based SCR assays were performed as previously described[26] to validate the readthrough of this gene. We cloned the cDNA of *ZmBTF3* with its stop codon TAA and inter-stop codon regions (ISRs) upstream of and in-frame with the cDNA of firefly luciferase without its start codon. A non-SCR gene (*Zm00001d009138*) was constructed as the background level of luciferase activity. *ZmBTF3* with TAA mutated to ACA was used as the positive control to measure the efficiency of SCR. The results of *in vitro* transcription followed by *in vitro* translation using wheat germ extract showed that *ZmBTF3-TAA-ISR* had a significantly higher luminescence activity compared to the background level, with an *in vitro* SCR rate calculated as 15% (Figure 5B). The results revealed that *ZmBTF3* was readthrough and generated a novel isoform (hereafter termed ZmBTF3x). Then, we confirmed the expression of the fusion protein ZmBTF3x-Fluc by immunoblotting using an anti-Fluc antibody. The results showed that the molecular weight of the readthrough protein ZmBTF3x-Fluc was the same as its positive control ZmBTF3x^{ACA}-Fluc and higher than that of ZmBTF3-Fluc (Figure 5C). By contrast, no visible band of readthrough products expressed from the SCR-negative control *Zm00001d009138-TAG-ISR-Fluc* was observed in the immunoblot (Figure 5C). The fusion protein ZmBTF3x-Fluc was then subjected to mass spectrometric analysis, resulting in the detection of different types of peptides labeled as before, through, and after, which further validated the production of the readthrough protein isoform ZmBTF3x (Figure 5D, top). The spectrum for the through peptide "EPEEKKESTFCCV" revealed that the stop codon UAA was translated as threonine (T) (Figure 5D, bottom).

The single transmembrane domain predicted in the C terminus of ZmBTF3x was found to contain 20 amino acid residues crossing the membrane (Figure 5E). The predicted 3D structures of ZmBTF3 and ZmBTF3x revealed an α helix appended at the ZmBTF3x C terminus that was not present in ZmBTF3 (Figure 5F). Furthermore, the α helix exactly corresponded to the region of the predicted transmembrane domain (Figure 5F). To examine the biological significance of the C-terminal extension derived from SCR of *ZmBTF3*, we constructed *35S:ZmBTF3-GFP* and *35S:ZmBTF3x-GFP* vectors and individually transfected them in maize protoplasts for subcellular localization analysis using confocal microscopy. The results showed that ZmBTF3 localized to the nucleus and cytosol (Figure 5G), consistent with the previous study reporting that *Arabidopsis* BTF3 acted as the transcription factor and the β subunit of nascent polypeptide-associated complex.[43] By contrast, the readthrough protein ZmBTF3x localized to the plasma membrane and cytosol (Figure 5G), confirming the localization change between the isoforms ZmBTF3 and ZmBTF3x. In summary, these results indicate that SCR can lead to the structural and subcellular localization change of the readthrough protein isoform, highlighting the significance of SCR events in plants.

Extensions of SCR events in plants are likely undergoing protein-coding selection

Next, we explored the evolutionary signatures of the identified SCR events based on their abundance and common traits among monocot and dicot plants. We first determined the orthologous relationships of the SCR genes among the four species. For each plant species, fewer than 10% of SCR genes had orthologs that also underwent readthrough in at least one other plant (Figure S7A). In total, we identified 75 SCR genes with SCR orthologs among the four species, with 72 orthologous in two species and three orthologous in three species (Figures S7A and S7B). These results suggest that SCR events are not conserved.

Secondly, we used the ratio of nonsynonymous to synonymous substitutions (d*N*/d*S*) to assess the potential selective pressure acting on the extensions of the SCR genes. We identified a total of 19 SCR single-copy orthogroups among the four

(E) Top: the amino acid sequence from the annotated stop codon to the next in-frame stop codon of ZmBTF3x. Bottom: transmembrane domain predicted in the C terminus of ZmBTF3x. The blue, red, and magenta lines represent residues 1–201 inside the membrane, 202–221 as a transmembrane helix, and 222–223 outside the membrane, respectively.

(F) AlphaFold structural models for ZmBTF3 and its readthrough protein ZmBTF3x. Green: ZmBTF3; dark blue: parent protein region of the readthrough protein ZmBTF3x; magenta: C-terminal extended region of ZmBTF3x. The predicted α helix is shown at the C terminus.

(G) Subcellular localization of ZmBTF3 and ZmBTF3x. GFP-vector transfected protoplasts were used as the control. The experiments were performed three times with similar results. HY5-mCherry was the nucleus marker, and PDCK7-mCherry was the plasma membrane marker. Scale bars, 5 μm. See also Tables S2, S11, and S12.

plant species for subsequent d*N*/d*S* calculation. Each of these orthogroups comprised four single-copy genes from maize, rice, soybean, and *Arabidopsis*, with at least one of them being identified as an SCR gene. Then, we compared the d*N*/d*S* ratios in different regions within these 19 orthogroups. The analysis revealed that the d*N*/d*S* ratio of the extensions in the 19 SCR single-copy orthogroups was significantly higher than that of the main ORF (false discovery rate [FDR] <0.01, Figure S7C) but lower than the distal 3′ UTRs (FDR <0.01, Figure S7C). These results suggest that the extensions are more conserved compared to the distal 3′ UTRs, and they are likely undergoing protein-coding selection. A similar tendency was also observed in the GLYCERALDEHYDE-3 DEHYDROGENASE (GAPDH) orthogroup, which contained four GAPDH orthologs with three of them being detected as SCR genes from maize, rice, and *Arabidopsis*. The extensions of GAPDH orthologs showed intermediate d*N*/d*S* values between the main ORFs and distal 3′ UTRs (Figure S7D). Collectively, these results suggest that SCR events in plants are not conserved, which is consistent with those reported in previous studies,[12,28,44] and the extensions are likely undergoing selection for protein coding.

DISCUSSION

In the current study, we employed a high-resolution MS-based proteogenomic strategy to analyze SCR events in different plants and to identify the amino acids translated at stop codon sites. Using this strategy, we detected 1,009 SCR events corresponding to a collection of 2,185 readthrough proteins in four plant species. Our systematic analysis of SCR events in these plant species revealed their similar characteristics among the four plant species.

Firstly, SCR events occur in all chromosomes of the four species, indicating no chromosomal preference; and SCR events are prone to be dispersively distributed in the chromosome, except in maize where the higher distribution of SCR was closer to telomeres. Previous studies have identified a lot of non-canonical peptides derived from the 3′ UTR in plants,[45,46] and it is likely that some of them may originate from the extended proteins generated by SCR events. Secondly, consistent with previous studies,[7,23] UGA is the most preferential readthrough stop codon, and it also turns out to be the most preferred next in-frame stop codon. We suggest that this preference could be attributed to the higher prevalence of TGA in the genome of the two monocots (Figures S8A and S8B). However, it is less likely due to the same reason for the two dicots, as TAA is the most frequent annotated stop codon and the next in-frame stop codon in their genomes (Figures S8C and S8D). In addition, SCR candidates in plants tended to have shorter transcript lengths and fewer exons and splice variants than non-SCR transcripts. As reported previously, SCR in organisms with small genomes such as virus and yeast allows them to expand the functional versatility of genes to survive in fluctuating environments.[19,21] Plants may employ a similar strategy to enrich the functional diversity of those short genes, which may function in adapting to different environments. However, in animals like *D. melanogaster* and *Anopheles gambiae*, SCR candidates were found to have much longer transcript lengths than the non-SCR transcripts, possibly because the occurrence of SCR requires more regulatory elements embedded in their longer transcripts.[23,24] Our findings suggest that the regulation of SCR in plants might differ from that in animals.

In addition, we found that the stop codons can be recoded as all 20 standard amino acids in the SCR events of these plant species, which has not been reported elsewhere. This result suggests a greater recoding potential of stop codons underestimated in the past and reveals a more flexible mode of decoding. Since tRNA is the adapter molecule in decoding a codon to an amino acid, we explored the reasons for the plasticity of stop codons by tRNA analysis. Suppressor tRNAs, of which the anticodon could fully form canonical base pairing with stop codons at all three nucleotide positions, have been reported to contribute to the translation of stop codons.[24,36,37,47] Here, we detected an array of suppressor tRNAs in different plant species, which may be the underlying reason for the translation of different amino acids at UGA, UAG, and UAA sites. In addition to suppressor tRNAs, it has been shown that near-cognate tRNAs can also lead to SCR due to non-canonical base pairing with the stop codon at a single nucleotide position of the anticodon. Therefore, near-cognate tRNAs also explain a subset of amino acids translated at stop codons.[33,34,48,49] For maize, suppressor tRNAs respectively constitute 30%, 30%, and 42.1% of the total tRNA usage for UGA, UAG, and UAA recoding. Near-cognate tRNAs represented 30%, 35%, and 31.6% of the total tRNA usage for the recoding of the three stop codons in SCR events. However, there remains a portion of the 20 amino acids recoded at stop codons that cannot be explained by either the suppressor tRNAs or near-cognate tRNAs in our results. Indeed, these amino acids are not exclusively observed in this study, as previous studies have also reported some of their translations at stop codon sites.[16,30,50,51] These cases suggest that our present understanding of stop codon reading by suppressor or near-cognate tRNAs is still incomplete.

In summary, we implemented a proteogenomic strategy and identified abundant SCR events across the genomes of four plant species. Stop codons in these SCR events could be recoded as all 20 standard amino acids. We demonstrated that SCR events are commonly present in plant species and uncovered the unprecedented plasticity of stop codons. Our findings provide evidence that genetic decoding is far more flexible than has been recognized. We also identified a large volume of novel protein isoforms derived from SCR, which not only expands the plant proteome diversity but also lays a foundation for further exploring the functionality of SCR proteins.

Limitations of the study

In this study, we discovered that three stop codons involved in plant SCR events could be recoded as all 20 standard amino acids, which unveils the unprecedented recoding plasticity of stop codons. Although we performed suppressor tRNA analysis to explore potential reasons behind the recoding plasticity of the stop codons, the mechanisms underlying SCR have not been thoroughly elucidated. In addition, we have demonstrated that SCR can lead to the structural and subcellular localization changes of the readthrough protein ZmBTF3x in the study, which revealed the significance of plant SCR. In future studies, more in-depth investigations

of ZmBTF3x and other SCR events are needed to fully elucidate the biological roles and mechanisms of SCR in the plant kingdom.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant material and growth conditions
- METHOD DETAILS
  - Protein extraction
  - Protein digestion and peptide fractionation
  - Liquid chromatography-tandem mass spectrometry analysis
  - Customized database construction
  - Identification of SCR events by MaxQuant
  - Ribosome profiling analysis
  - Distribution of readthrough events at the genomic level
  - Motif analysis of extended regions within SCR events in maize
  - Logistic regression analysis
  - Identification and alignment of suppressor tRNAs
  - Functional signal and structure prediction of C-terminal extended proteins in maize
  - Plasmid construction
  - *In vitro* translation for SCR assay
  - Immunoblotting and mass spectrometric analysis
  - Subcellular localization of ZmBTF3-GFP and ZmBTF3x-GFP fusions
  - Evolutionary analysis of SCR events
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.113723.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (nos. U22A20474 and 32172073), the National Key Research and Development Program of China (no. 2022YFD1201802), and the Seed-Industrialized Development Program of Shandong Province (no. 2021LZGC003). We are grateful to Steven P. Briggs and David Backhouse for their constructive comments and suggestions. We also thank Dr. Anguo Sun and Dr. Yanwen Xiang for the technical assistance and Yingying Sha and Mengxia Niu for their assistance with the microscopic observations.

AUTHOR CONTRIBUTIONS

L.W., S.-B.W., and Z.T. designed the project. Y. Zhang, L.T.,Y.S., H.Y., A.X., J.Z., U.A., A.S., X.C., and Y.C. were involved in bioinformatics analysis. Y. Zhang, H.L., S.W., and J.S. conducted the experiments. C.G., Y. Zhao, Y.L., and X.W. collected the materials for mass spectrometry detection and subsequent experiments. Y. Zhang, H.L., and Y.S. wrote the manuscript. L.W., S.-B.W., and Z.T. contributed substantially to revisions. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 20, 2023
Revised: October 2, 2023
Accepted: January 15, 2024
Published: January 31, 2024

REFERENCES

1. Baranov, P.V., Atkins, J.F., and Yordanova, M.M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. Nat. Rev. Genet. *16*, 517–529. https://doi.org/10.1038/nrg3963.
2. Freitag, J., Ast, J., and Bölker, M. (2012). Cryptic peroxisomal targeting via alternative splicing and stop codon read-through in fungi. Nature *485*, 522–525. https://doi.org/10.1038/nature11051.
3. Meydan, S., Marks, J., Klepacki, D., Sharma, V., Baranov, P.V., Firth, A.E., Margus, T., Kefi, A., Vázquez-Laslop, N., and Mankin, A.S. (2019). Retapamulin-Assisted Ribosome Profiling Reveals the Alternative Bacterial Proteome. Mol. Cell *74*, 481–493.e6. https://doi.org/10.1016/j.molcel.2019.02.017.
4. Jacks, T., Varmus, H.E., Masiarz, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science *230*, 1237–1242. https://doi.org/10.1038/331280a0.
5. Lang, B.F., Jakubkova, M., Hegedusova, E., Daoud, R., Forget, L., Brejova, B., Vinar, T., Kosa, P., Fricova, D., Nebohacova, M., et al. (2014). Massive programmed translational jumping in mitochondria. Proc. Natl. Acad. Sci. USA *111*, 5926–5931. https://doi.org/10.1073/pnas.1322190111.
6. Wu, Q., Wright, M., Gogol, M.M., Bradford, W.D., Zhang, N., and Bazzini, A.A. (2020). Translation of small downstream ORFs enhances translation of canonical main open reading frames. EMBO J. *39*, e104763. https://doi.org/10.15252/embj.2020104763.
7. Sapkota, D., Lake, A.M., Yang, W., Yang, C., Wesseling, H., Guise, A., Uncu, C., Dalal, J.S., Kraft, A.W., Lee, J.M., et al. (2019). Cell-type-specific profiling of alternative translation identifies regulated protein isoform variation in the mouse brain. Cell Rep. *26*, 594–607.e7. https://doi.org/10.1016/j.celrep.2018.12.077.
8. Rodnina, M.V., Korniy, N., Klimova, M., Karki, P., Peng, B.Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E., and Peske, F. (2020). Translational recoding: canonical translation mechanisms reinterpreted. Nucleic Acids Res. *48*, 1056–1067. https://doi.org/10.1093/nar/gkz783.
9. Palma, M., and Lejeune, F. (2021). Deciphering the molecular mechanism of stop codon readthrough. Biol. Rev. Camb. Philos. Soc. *96*, 310–329. https://doi.org/10.1111/brv.12657.
10. Torabi, N., and Kruglyak, L. (2012). Genetic basis of hidden phenotypic variation revealed by increased translational readthrough in yeast. PLoS Genet. *8*, e1002546. https://doi.org/10.1371/journal.pgen.1002546.
11. Namy, O., Duchateau-Nguyen, G., and Rousset, J.P. (2002). Translational readthrough of the *PDE2* stop codon modulates cAMP levels in *Saccharomyces cerevisiae*. Mol. Microbiol. *43*, 641–652. https://doi.org/10.1046/j.1365-2958.2002.02770.x.
12. Dunn, J.G., Foo, C.K., Belletier, N.G., Gavis, E.R., and Weissman, J.S. (2013). Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. Elife *2*, e01179. https://doi.org/10.7554/eLife.01179.
13. Xue, F., and Cooley, L. (1993). *kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers. Cell *72*, 681–693. https://doi.org/10.1016/0092-8674(93)90397-9.

14. Karki, P., Carney, T.D., Maracci, C., Yatsenko, A.S., Shcherbata, H.R., and Rodnina, M.V. (2022). Tissue-specific regulation of translational readthrough tunes functions of the traffic jam transcription factor. Nucleic Acids Res. *50*, 6001–6019. https://doi.org/10.1093/nar/gkab1189.

15. Schueren, F., Lingner, T., George, R., Hofhuis, J., Dickel, C., Gärtner, J., and Thoms, S. (2014). Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. Elife *3*, e03640. https://doi.org/10.7554/eLife.03640.

16. Eswarappa, S.M., Potdar, A.A., Koch, W.J., Fan, Y., Vasu, K., Lindner, D., Willard, B., Graham, L.M., DiCorleto, P.E., and Fox, P.L. (2014). Programmed translational readthrough generates antiangiogenic VEGF-Ax. Cell *157*, 1605–1618. https://doi.org/10.1016/j.cell.2014.04.033.

17. Xin, H., Zhong, C., Nudleman, E., and Ferrara, N. (2016). Evidence for pro-angiogenic functions of VEGF-Ax. Cell *167*, 275–284.e6. https://doi.org/10.1016/j.cell.2016.08.054.

18. Weiner, A.M., and Weber, K. (1971). Natural read-through at the UGA termination signal of Q-beta coat protein cistron. Nat. New Biol. *234*, 206–209. https://doi.org/10.1038/newbio234206a0.

19. Pelham, H.R. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. Nature *272*, 469–471. https://doi.org/10.1038/272469a0.

20. Yoshinaka, Y., Katoh, I., Copeland, T.D., and Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA *82*, 1618–1622. https://doi.org/10.1073/pnas.82.6.1618.

21. True, H.L., and Lindquist, S.L. (2000). A yeast prion provides a mechanism for genetic variation and phenotypic diversity. Nature *407*, 477–483. https://doi.org/10.1038/35035005.

22. Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret, D., and Namy, O. (2014). Genome-wide translational changes induced by the prion [PSI+]. Cell Rep. *8*, 439–448. https://doi.org/10.1016/j.celrep.2014.06.036.

23. Jungreis, I., Lin, M.F., Spokony, R., Chan, C.S., Negre, N., Victorsen, A., White, K.P., and Kellis, M. (2011). Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. Genome Res. *21*, 2096–2113. https://doi.org/10.1101/gr.119974.110.

24. Jungreis, I., Chan, C.S., Waterhouse, R.M., Fields, G., Lin, M.F., and Kellis, M. (2016). Evolutionary dynamics of abundant stop codon readthrough. Mol. Biol. Evol. *33*, 3108–3132. https://doi.org/10.1093/molbev/msw189.

25. Geller, A.I., and Rich, A. (1980). A UGA termination suppression tRNATrp active in rabbit reticulocytes. Nature *283*, 41–46. https://doi.org/10.1038/283041a0.

26. Singh, A., Manjunath, L.E., Kundu, P., Sahoo, S., Das, A., Suma, H.R., Fox, P.L., and Eswarappa, S.M. (2019). Let-7a-regulated translational readthrough of mammalian *AGO1* generates a microRNA pathway inhibitor. EMBO J. *38*, e100727. https://doi.org/10.15252/embj.2018100727.

27. Yordanova, M.M., Loughran, G., Zhdanov, A.V., Mariotti, M., Kiniry, S.J., O'Connor, P.B.F., Andreev, D.E., Tzani, I., Saffert, P., Michel, A.M., et al. (2018). *AMD1* mRNA employs ribosome stalling as a mechanism for molecular memory formation. Nature *553*, 356–360. https://doi.org/10.1038/nature25174.

28. Sahoo, S., Singh, D., Singh, A., Pandit, M., Vasu, K., Som, S., Pullagurla, N.J., Laha, D., and Eswarappa, S.M. (2022). Identification and functional characterization of mRNAs that exhibit stop codon readthrough in *Arabidopsis thaliana*. J. Biol. Chem. *298*, 102173. https://doi.org/10.1016/j.jbc.2022.102173.

29. Namy, O., Duchateau-Nguyen, G., Hatin, I., Hermann-Le Denmat, S., Termier, M., and Rousset, J.P. (2003). Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. Nucleic Acids Res. *31*, 2289–2296. https://doi.org/10.1093/nar/gkg330.

30. Blanchet, S., Cornu, D., Argentini, M., and Namy, O. (2014). New insights into the incorporation of natural suppressor tRNAs at stop codons in *Saccharomyces cerevisiae*. Nucleic Acids Res. *42*, 10061–10072. https://doi.org/10.1093/nar/gku663.

31. Hatfield, D.L., Dudock, B.S., and Eden, F.C. (1983). Characterization and nucleotide sequence of a chicken gene encoding an opal suppressor tRNA and its flanking DNA segments. Proc. Natl. Acad. Sci. USA *80*, 4940–4944. https://doi.org/10.1073/pnas.80.16.4940.

32. Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T., Söll, D., and Podar, M. (2013). UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc. Natl. Acad. Sci. USA *110*, 5540–5545. https://doi.org/10.1073/pnas.1303090110.

33. Roy, B., Leszyk, J.D., Mangus, D.A., and Jacobson, A. (2015). Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. Proc. Natl. Acad. Sci. USA *112*, 3038–3043. https://doi.org/10.1073/pnas.1424127112.

34. Blanchet, S., Cornu, D., Hatin, I., Grosjean, H., Bertin, P., and Namy, O. (2018). Deciphering the reading of the genetic code by near-cognate tRNA. Proc. Natl. Acad. Sci. USA *115*, 3018–3023. https://doi.org/10.1073/pnas.1715578115.

35. Artieri, C.G., and Fraser, H.B. (2014). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res. *24*, 2011–2021. https://doi.org/10.1101/gr.175893.114.

36. Hottinger, H., Stadelmann, B., Pearson, D., Frendewey, D., Kohli, J., and Söll, D. (1984). The *Schizosaccharomyces pombe sup3-i* suppressor recognizes ochre, but not amber codons *in vitro* and *in vivo*. EMBO J. *3*, 423–428. https://doi.org/10.1002/j.1460-2075.1984.tb01823.x.

37. Ivanova, N.N., Schwientek, P., Tripp, H.J., Rinke, C., Pati, A., Huntemann, M., Visel, A., Woyke, T., Kyrpides, N.C., and Rubin, E.M. (2014). Stop codon reassignments in the wild. Science *344*, 909–913. https://doi.org/10.1126/science.1250691.

38. Chan, P.P., Lin, B.Y., Mak, A.J., and Lowe, T.M. (2021). tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. *49*, 9077–9096. https://doi.org/10.1093/nar/gkab688.

39. Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. *32*, 11–16. https://doi.org/10.1093/nar/gkh152.

40. Beznosková, P., Gunišová, S., and Valášek, L.S. (2016). Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. RNA *22*, 456–466. https://doi.org/10.1261/rna.054452.115.

41. Bartoschek, M.D., Ugur, E., Nguyen, T.A., Rodschinka, G., Wierer, M., Lang, K., and Bultmann, S. (2021). Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells. Nucleic Acids Res. *49*, e62. https://doi.org/10.1093/nar/gkab13.

42. Chen, Y., He, X., Ma, B., Liu, K., Gao, T., Niu, W., and Guo, J. (2022). Non-canonical amino acid mutagenesis in response to recoding signal-enhanced quadruplet codons. Nucleic Acids Res. *50*, e94. https://doi.org/10.1093/nar/gkac474.

43. Ding, Y., Jia, Y., Shi, Y., Zhang, X., Song, C., Gong, Z., and Yang, S. (2018). OST1-mediated BTF3L phosphorylation positively regulates CBFs during plant cold responses. EMBO J. *37*, e98228. https://doi.org/10.15252/embj.201798228.

44. Artieri, C.G., and Fraser, H.B. (2014). Evolution at two levels of gene expression in yeast. Genome Res. *24*, 411–421. https://doi.org/10.1101/gr.165522.113.

45. Wang, S., Tian, L., Liu, H., Li, X., Zhang, J., Chen, X., Jia, X., Zheng, X., Wu, S., Chen, Y., et al. (2020). Large-scale discovery of non-conventional peptides in maize and *Arabidopsis* through an integrated peptidogenomic pipeline. Mol. Plant *13*, 1078–1093. https://doi.org/10.1016/j.molp.2020.05.012.

46. Peng, Z., Yao, S., Zhang, B., Huang, K., and Wan, C. (2022). Identification and analysis of smORFs in *Chlamydomonas reinhardtii*. Genomics *114*, 110444. https://doi.org/10.1016/j.ygeno.2022.110444.

47. Swart, E.C., Serra, V., Petroni, G., and Nowacki, M. (2016). Genetic codes with no dedicated stop codon: context-dependent translation termination. Cell *166*, 691–702. https://doi.org/10.1016/j.cell.2016.06.020.

48. Beier, H., and Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res. *29*, 4767–4782. https://doi.org/10.1093/nar/29.23.4767.

49. Dabrowski, M., Bukowy-Bieryllo, Z., and Zietkiewicz, E. (2015). Translational readthrough potential of natural termination codons in eucaryotes–The impact of RNA sequence. RNA Biol. *12*, 950–958. https://doi.org/10.1080/15476286.2015.1068497.

50. Manjunath, L.E., Singh, A., Sahoo, S., Mishra, A., Padmarajan, J., Basavaraju, C.G., and Eswarappa, S.M. (2020). Stop codon read-through of mammalian *MTCH2* leading to an unstable isoform regulates mitochondrial membrane potential. J. Biol. Chem. *295*, 17009–17026. https://doi.org/10.1074/jbc.RA120.014253.

51. Aerni, H.R., Shifman, M.A., Rogulina, S., O'Donoghue, P., and Rinehart, J. (2015). Revealing the amino acid composition of proteins within an expanded genetic code. Nucleic Acids Res. *43*, e8. https://doi.org/10.1093/nar/gku1087.

52. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. *11*, 2301–2319. https://doi.org/10.1038/nprot.2016.136.

53. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal *17*, 3. https://doi.org/10.14806/ej.17.1.200.

54. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962. https://doi.org/10.1371/journal.pone.0163962.

55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

56. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697. https://doi.org/10.1093/bioinformatics/btr189.

57. Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. Genome Biol. *8*, R24. https://doi.org/10.1186/gb-2007-8-2-r24.

58. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

59. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

60. Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. USA *106*, 10171–10176. https://doi.org/10.1073/pnas.0900604106.

61. Neuberger, G., Maurer-Stroh, S., Eisenhaber, B., Hartig, A., and Eisenhaber, F. (2003). Prediction of peroxisomal targeting signal 1 containing proteins from amino acid sequence. J. Mol. Biol. *328*, 581–592. https://doi.org/10.1016/s0022-2836(03)00319-x.

62. Maurer-Stroh, S., and Eisenhaber, F. (2005). Refinement and prediction of protein prenylation motifs. Genome Biol. *6*, R55. https://doi.org/10.1186/gb-2005-6-6-r55.

63. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

64. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

65. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

66. Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., and Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. *47*, W636-w641. https://doi.org/10.1093/nar/gkz268.

67. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612. https://doi.org/10.1093/nar/gkl315.

68. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

69. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362. https://doi.org/10.1038/nmeth.1322.

70. Lei, L., Shi, J., Chen, J., Zhang, M., Sun, S., Xie, S., Li, X., Zeng, B., Peng, L., Hauck, A., et al. (2015). Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. Plant J. *84*, 1206–1218. https://doi.org/10.1111/tpj.13073.

71. Chotewutmontri, P., and Barkan, A. (2016). Dynamics of chloroplast translation during chloroplast differentiation in maize. PLoS Genet. *12*, e1006106. https://doi.org/10.1371/journal.pgen.1006106.

72. Chotewutmontri, P., and Barkan, A. (2018). Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. PLoS Genet. *14*, e1007555. https://doi.org/10.1371/journal.pgen.1007555.

73. Zoschke, R., Chotewutmontri, P., and Barkan, A. (2017). Translation and co-translational membrane engagement of plastid-encoded chlorophyll-binding proteins are not influenced by chlorophyll availability in maize. Front. Plant Sci. *8*, 385. https://doi.org/10.3389/fpls.2017.00385.

74. Jiang, J., Chai, X., Manavski, N., Williams-Carrier, R., He, B., Brachmann, A., Ji, D., Ouyang, M., Liu, Y., Barkan, A., et al. (2019). An RNA chaperone-like protein plays critical roles in chloroplast mRNA stability and translation in *Arabidopsis* and maize. Plant Cell *31*, 1308–1327. https://doi.org/10.1105/tpc.18.00946.

75. Zhu, W., Xu, J., Chen, S., Chen, J., Liang, Y., Zhang, C., Li, Q., Lai, J., and Li, L. (2021). Large-scale translatome profiling annotates the functional genome and reveals the key role of genic 3' untranslated regions in translatomic variation in plants. Plant Commun. *2*, 100181. https://doi.org/10.1016/j.xplc.2021.100181.

76. Hsu, P.Y., Calviello, L., Wu, H.Y.L., Li, F.W., Rothfels, C.J., Ohler, U., and Benfey, P.N. (2016). Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. Proc. Natl. Acad. Sci. USA *113*, E7126–e7135. https://doi.org/10.1073/pnas.1614788113.

77. Yoo, S.D., Cho, Y.H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. *2*, 1565–1572. https://doi.org/10.1038/nprot.2007.199.

78. Zhao, Y., Du, H., Wang, Y., Wang, H., Yang, S., Li, C., Chen, N., Yang, H., Zhang, Y., Zhu, Y., et al. (2021). The calcium-dependent protein kinase ZmCDPK7 functions in heat-stress tolerance in maize. J. Integr. Plant Biol. *63*, 510–527. https://doi.org/10.1111/jipb.13056.

79. Li, Y., Xue, J., Wang, F.Z., Huang, X., Gong, B.Q., Tao, Y., Shen, W., Tao, K., Yao, N., Xiao, S., et al. (2022). Plasma membrane-nucleo-cytoplasmic coordination of a receptor-like cytoplasmic kinase promotes EDS1-dependent plant immunity. Nat. Plants *8*, 802–816. https://doi.org/10.1038/s41477-022-01195-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Fluc	Abcam	Cat#ab185924; RRID: AB_2938620
Bacterial and virus strains		
Escherichia coli strain DH5α	WEIDI	Cat#DL1001
Chemicals, peptides, and recombinant proteins		
Trypsin	Promega Corporation	Cat#V5113
BCA Protein Assay Reagent	Sigma-Aldrich	Cat#QPBCA-1KT
T7 RNA polymerase	Promega Corporation	Cat#P2075
DTT	Promega Corporation	Cat#P1171
Recombinant RNasin	Promega Corporation	Cat#N2511
Ribonuclease Inhibitor	Promega Corporation	Cat#N2515
rATP	Promega Corporation	Cat#E6011
rUTP	Promega Corporation	Cat#E6021
rGTP	Promega Corporation	Cat#E6031
rCTP	Promega Corporation	Cat#E6041
PVDF membranes (0.45 μm)	Biosharp	Cat#BS-PVDF-45
Chymotrypsin	Promega Corporation	Cat#V1061
Pepsin	Promega Corporation	Cat#V1959
Critical commercial assays		
GeneJET RNA Purification kit	Thermo Fisher Scientific	Cat#K0731
Wheat Germ Extract	Promega Corporation	Cat#L4380
Deposited data		
Proteomics raw datasets	ProteomeXchange	PXD035958; PXD038762
Experimental models: Organisms/strains		
Zea mays: B73 inbred line	Widely distributed	N/A
Oryza sativa L. ssp. *japonica* cv. Nipponbare	Widely distributed	N/A
Glycine max L., ZH13	Widely distributed	N/A
Arabidopsis thaliana: Columbia-0	Widely distributed	N/A
Oligonucleotides		
See Table S12 for primers used in this study	This work	N/A
Recombinant DNA		
35S:ZmBTF3-GFP	This work	N/A
35S:ZmBTF3x-GFP	This work	N/A
35S:ZmPDCK7-mCherry	This work	N/A
35S:AtHY5-mCherry	This work	N/A
Software and algorithms		
EMBOSS 6.6.0	The European Molecular Biology Open Software Suite	https://emboss.sourceforge.net/index.html
MaxQuant v.1.6.1.10	Tyanova et al.[52]	https://www.maxquant.org/maxquant/
fastax v.36.06	Cold Spring Harbor Laboratory	http://hannonlab.cshl.edu/fastx_toolkit/
cutadapt v.1.13	Martin et al.[53]	https://cutadapt.readthedocs.io/en/stable/
seqkit v.0.5.4	Shen et al.[54]	https://bioinf.shenwei.me/seqkit/
STAR v.2.7.3a	Dobin et al.[55]	https://code.google.com/archive/p/rna-star/
R package v.3.2.5	R Development Core Team	https://www.r-project.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ribosomeProfilingQC v.1.10.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/ribosomeProfilingQC.html
MEME-ChIP v.5.4.1	Machanick et al.[56]	https://meme-suite.org/meme/doc/meme-chip.html?man_type=web
Tomtom v.5.0.5	Gupta et al.[57]	https://meme-suite.org/meme/tools/tomtom
tRNAscan-SE v.2.0	Chan et al.[38]	http://trna.ucsc.edu/tRNAscan-SE/
ARAGORN	Laslett et al.[39]	http://www.ansikte.se/ARAGORN/
BLASTn	Camacho et al.[58]	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+
TMHMM v.2.0	Krogh et al.[59]	https://services.healthtech.dtu.dk/services/TMHMM-2.0/
cNLS Mapper	Kosugi et al.[60]	http://nls-mapper.iab.keio.ac.jp/
PTS1 Predictor	Neuberger et al.[61]	https://mendel.imp.ac.at/pts1/
PrePS	Maurer-Stroh et al.[62]	https://mendel.imp.ac.at/PrePS/index.html
Microsoft Office	Microsoft 365	https://www.microsoft.com/zh-cn/microsoft-365
Alphafold v.2.0	Jumper et al.[63]	https://alphafold.ebi.ac.uk/
Zen	Zeiss	https://www.zeiss.com/microscopy/en/products/software/zeiss-zen.html
OrthoFinder v.2.3.11	Emms et al.[64]	http://www.stevekellylab.com/software/orthofinder
Muscle	Edgar et al.[65] Madeira et al.[66]	https://www.ebi.ac.uk/Tools/msa/muscle/
PAL2NAL v. 14	Suyama et al.[67]	http://www.bork.embl.de/pal2nal/
PAML package v.4.8	Yang et al.[68]	http://abacus.gene.ucl.ac.uk/software/paml.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Liuji Wu (wlj200120@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All data reported in this paper will be shared by the lead contact upon request.
- The raw datasets of proteomics have been deposited at the ProteomeXchange under the accession ID PXD035958 and PXD038762. This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant material and growth conditions

Maize (*Zea mays,* B73 inbred line), rice (*Oryza sativa* L. ssp. *japonica* cv. Nipponbare), *Arabidopsis thaliana* (Columbia-0), and soybean (*Glycine max* L., ZH13) plants were grown in a temperature- and light-controlled greenhouse. The growth conditions for maize were a 15-h light/9-h dark photoperiod and a 28/25°C day/night air temperature, and samples were collected at the three-leaf stage. For rice, the growth conditions were a 12-h light/12-h dark photoperiod and a 28/25°C day/night air temperature, and samples were collected on day 56. *Arabidopsis* plants were grown in a greenhouse under a 16-h light/8-h dark photoperiod and a 22/21°C day/night air temperature, and samples were collected on day 21. Soybean plants were grown in a greenhouse under a 16-h light/8-h dark photoperiod and a 22/21°C day/night air temperature until the trifoliate leaves were fully expanded. Three replicates of leaf samples for each of the plant species were harvested, frozen in liquid nitrogen, and stored at −80°C until analysis.

METHOD DETAILS

Protein extraction

Total proteins from each sample were extracted as previously reported.[69] Briefly, leaf tissues (1 g) for each plant species were pulverized in liquid nitrogen, and the powder was precipitated in 10% (w/v) trichloroacetic acid/acetone solution containing 65 mM

dithiothreitol (DTT) at −20°C for 1 h. Then, the solution was centrifuged at 10,000×g at 4°C for 45 min and the supernatant was removed. The precipitate was then vacuum-dried and solubilized in 1/10 volumes of SDT buffer (4% [w/v] SDS, 100 mM DTT, and 150 mM Tris-HCl, pH 8.0). After a 3-min incubation in boiling water, the suspensions were ultrasonicated (80 w, ultrasonic 10 s/time, every 15 s, repeated 10 times) and incubated at 100°C for 3 min. The extract was centrifuged at 12,000×g at 25°C for 10 min, and the total protein contents were quantified using BCA Protein Assay Reagent (Sigma-Aldrich, USA). The supernatants were stored at −80°C until subsequent analysis.

Protein digestion and peptide fractionation

The total proteins (250 μg for each sample) were digested by FASP procedure using trypsin (Promega Corporation, USA) as described previously.[69] Then, the digested peptide filtrate was fractionated by strong cation exchange (SCX) chromatography using an AKTA Purifier system (GE Healthcare, USA). The dried peptide mixture was reconstituted and acidified with 2 mL buffer A (10 mM KH_2PO_4 in 25%[v/v] of ACN, pH = 2.7) and loaded onto a PolySULFOETHYL 4.6 × 100 mm column (5 μm, 200 Å, PolyLC Inc, USA). The peptides were eluted at a flow rate of 1 mL/min buffer B (500 mM KCl, 10 mM KH_2PO_4 in 25% of ACN, pH 2.7) in a gradient of 0–10% (v/v) for 2 min, 10–20% (v/v) for 25 min, 20–45% (v/v) for 5 min, and 50–100% (v/v) for 5 min. The elution was monitored by absorbance at 214 nm, and fractions were collected every 1 min. The 30 collected fractions were combined into 15 pools and desalted on C18 Cartridges (Empore SPE Cartridges C18 standard density, bed I.D. 7 mm, volume 3 mL, Sigma-Aldrich, Germany). Each fraction was concentrated by vacuum centrifugation and reconstituted in 40 μL of 0.1% (v/v) trifluoroacetic acid. All samples were stored at −80°C before LC-MS/MS analysis.

Liquid chromatography-tandem mass spectrometry analysis

The samples were analyzed on a Q-Exactive HF-X mass spectrometer coupled to the Easy nLC 1200 (Thermo Fisher Scientific, USA). Peptide (2 μg) was loaded onto a 25 cm, 75 μm inner diameter C18-reversed phase column packed with 3 μm resin (EASY-Column Capillary Columns, Thermo Fisher Scientific, USA) in buffer A (2% [v/v] acetonitrile and 0.1% [v/v] formic acid) and separated with a linear gradient of buffer B (80% [v/v] acetonitrile and 0.1% formic acid) at a flow rate of 250 nL/min controlled by IntelliFlow technology over 90 min. MS data were acquired using a data-dependent top 10 method dynamically choosing the most abundant precursor ions from the survey scan (m/z 300-1,800) for higher-energy collision dissociation fragmentation (HCD). Determination of the target value was based on predictive Automatic Gain Control (pAGC). The dynamic exclusion duration was 25 s. Survey scans were acquired at a resolution of 70,000 (m/z 200), and the resolution for HCD spectra was set to 17,500 (m/z 200). Normalized collision energy was 30 eV, and the underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.1%. The instrument was run with peptide recognition mode enabled. MS experiments were conducted in triplicate for each sample.

Customized database construction

The complete genome sequences of the four plants were downloaded from Ensembl Plants (http://ftp.ensemblgenomes.org/pub/plants/release-41/fasta/zea_mays/dna/) for maize, MSU Rice Genome Annotation Project (http://rice.uga.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/) for rice, the National Genomics Data Center (http://download.big.ac.cn/gwh/Plants/Glycine_max_Gmax_ZH13_v2.0_GWHAAEV00000000.1/GWHAAEV00000000.1.genome.fasta.gz) for soybean, and TAIR 10 (http://ftp.ensemblgenomes.org/pub/plants/release-28/fasta/arabidopsis_thaliana/dna/Arabidopsis_thaliana.TAIR10.28.dna.toplevel.fa.gz) for *Arabidopsis* in FASTA format. The tailored databases for SCR event identification were generated from the computational translation of plant genome sequences to amino acid sequences by EMBOSS 6.6.0, as previously described with modifications.[51] *In silico* translation followed the principle that the annotated stop codon (UGA, UAG, or UAA) of a given gene was considered a translatable codon to be designated as the symbol "X", representing any one of the 20 standard amino acids, and the elongation terminated at the next in-frame stop codon in the 3′ UTR. The *in silico* translation was conducted based on Shell under a Linux environment to produce 20 customized databases for each plant species. Each database has been in conjunction with a public plant protein database for every plant species as follows: Ensembl Plants (http://ftp.ebi.ac.uk/ensemblgenomes/pub/release-41/plants/fasta/zea_mays/pep/Zea_mays.B73_RefGen_v4.pep.all.fa.gz) for maize, MSU Rice Genome Annotation Project protein database (http://rice.uga.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_7.0/all.dir/all.pep) for rice, the National Genomics Data Center (https://download.cncb.ac.cn/gwh/Plants/Glycine_max_Gmax_ZH13_v2.0_GWHAAEV00000000.1/GWHAAEV00000000.1.Protein.faa.gz) for soybean, and Ensembl Plants (http://ftp.ensemblgenomes.org/pub/plants/release-51/fasta/arabidopsis_thaliana/pep/Arabidopsis_thaliana.TAIR10.pep.all.fa.gz) for *Arabidopsis*. The conjunction of the public protein database ensured the searching precision and calculation of the false discovery rate (FDR) in the subsequent searching stage.

Identification of SCR events by MaxQuant

The raw data were analyzed using MaxQuant software (v.1.6.1.10).[52] The spectra were searched against the 20 customized databases for each of the four plant species. The initial search was set at a precursor mass window of 6 ppm. The search followed an enzymatic cleavage rule of trypsin and allowed a maximum of two missed cleavage sites and a mass tolerance of 20 ppm for

fragment ions. Carbamidomethylation of cysteines was defined as fixed modification, while methionine oxidation was defined as variable modification for database searching. The cutoff of global FDR for peptide and protein identification was set to 0.01. Protein quantification was calculated by normalized spectral protein intensity.

All matched peptides were then relocated to the corresponding positions of protein sequences in the reference databases and mapped to their genomic locations. Every stop codon site was given an ID for genomic position recognition. Peptides that were exclusively mapped to a specific location of a given gene across protein databases generated from all transcript isoforms were defined as unique peptides. The non-redundant peptides uniquely mapped to the regions spanning the annotated stop codon were labeled as through, the peptides mapped to the coding sequence (CDS) region were labeled as before, and the peptides mapped to the 3′ UTR were labeled as after. We further implemented more stringent criteria to determine the through peptides to identify SCR events. Only the unique through peptides that met the following conditions were retained: (1) presented in at least two of the three replicates in the MS experiment; (2) at least two of the quality scores were higher than 20. We merged proteins sharing one readthrough stop codon site spanned by a unique through peptide into one group and the group was defined as one SCR event.

Ribosome profiling analysis

We obtained ribosome profiling (Ribo-seq) datasets of 56 maize samples from previous reports for SCR event analyzing.[70–74] The Ribo-seq data was analyzed according to the methods described by Dunn et al.[12] and Sapkota et al.[7] with modifications. Briefly, the ribosomal sequencing data were quality-controlled using fastax (v.36.06) or cutadapt (v.1.13)[53] to remove low-quality bases, adaptor sequences, and random primers. Reads of 25–35 nucleotides (nts) were retained by seqkit (v.0.5.4)[54] for subsequent analysis. The remaining reads were aligned to the maize genome by STAR (v.2.7.3a).[55] Ribosome-protected footprint alignments were mapped to their estimated P-sites as follows: 12 nt were cut from each end of the alignment, resulting in a fragment of n×nt in length (n = l-12 × 2). Each genomic position covered by a nucleotide remaining in the truncated alignment was then incremented by 1/n (count number). Then, the reads were mapped to the following positions: 12 nts after the start codon, 15 nts before and 9 nts after the first stop codon, and 15 nts after the second stop codon. This resulted in four regions remaining in the transcripts: the 5′ UTR, CDS, readthrough region, and the distal 3′ UTR.

We then analyzed the length distribution of Ribo-seq reads in each dataset. Four Ribo-seq datasets[70] in which the read length was mainly distributed in 30-nt (30-mer) and prepared with RNase I (little cutting bias) were finally retained for further SCR validation. For analysis of 3-nt periodicity (phasing) of footprints, R package ribosomeProfilingQC (v.1.10.0) was used and only 30-nt RPFs were counted, as shown in maize with good-phased footprint population.[70] Reads were then counted in all putative extended regions. Mean value of normalized RPKM (reads per kilobase per million mapped reads) was calculated in R (v.3.2.5) in the four datasets. The ribosome footprint density of putative extensions, the distal 3′ UTRs, and the coding regions were calculated in RPKM. Distal 3′ UTR was defined as the region 40 codons after the next stop-codon or the region from the next stop-codon to the third stop codon within 40 codons. Only extensions with RPKM >1, read coverage >10%, and no overlap with any annotated 5′ UTRs, CDSs, and other non-coding RNAs were retained and regarded as readthrough. Full data of the expression levels of transcripts involved in the SCR events in the four maize sample datasets are shown in Table S4. The transcript expression levels of SCR candidates and non-SCR genes were calculated and compared in RPKM using RNA-seq data of the four sample datasets.[70]

For the Ribo-seq analyses of SCR events in rice and *Arabidopsis*, we retrieved their respective Ribo-seq data from the previous studies[75,76] and applied the same pipeline as that used in maize (the analysis was not conducted for soybean due to a lack of available Ribo-seq datasets for the cultivar ZH13).

Distribution of readthrough events at the genomic level

The genome-wide density of SCR and non-SCR events in maize chromosomes was plotted using a sliding window of 20 Mb with 2-Mb steps according to the annotated maize genome in Ensembl (https://plants.ensembl.org/index.html). We used "hotspot region" to reflect the distribution density of SCR and non-SCR events. Hotspot regions were defined as ≥ 5 SCR events for SCR hotspots and ≥335 for non-SCR events in a window size of 20 Mb. The chromosomal distributions of rice SCR and non-SCR events were drawn using a window size of 4 Mb with a step of 400 kb based on the annotated rice genome from MSU (http://rice.uga.edu/). The same distribution pattern was plotted with a window size of 10 Mb with a 1-Mb step based on the ZH13 reference genome (https://ngdc.cncb.ac.cn/gwh/Assembly/652/show) for soybean. For *Arabidopsis*, the SCR and non-SCR genomic localizations were obtained by a 1.2-Mb sliding window with a step size of 100 kb according to the annotated *Arabidopsis* genome in TAIR (https://www.Arabidopsis.org/). Hotspot regions were defined as ≥3 for SCR events and ≥178 for non-SCR events in a window size of 4 Mb for rice, ≥2 for SCR events and ≥253 for non-SCR events in a window size of 10 Mb for soybean, and ≥4 for SCR events and ≥157 for non-SCR events in a window size of 1.2 Mb for *Arabidopsis*.

Motif analysis of extended regions within SCR events in maize

Transcription factor-binding motifs in readthrough regions of SCR candidates in maize were identified using the MEME-ChIP webserver (v.5.4.1),[56] and the annotations of the obtained motifs were searched against the available databases for comparison using Tomtom (v.5.0.5) with default parameters.[57]

Logistic regression analysis

The logistic regression analysis using R was performed to assess the contribution of the exon number and the primary transcript length to the correlation with SCR. The exon number and the primary transcript length were set as the two independent variables, and the occurrence of SCR was the dependent variable with the SCR event set as 1 and the non-SCR event as 0. The absolute values of regression coefficients for the two factors were compared at a significance level of 0.05.

Identification and alignment of suppressor tRNAs

The complete genome sequences of the four plant species were downloaded as described above. A genome-wide search for suppressor tRNA genes was conducted using tRNAscan-SE (v.2.0)[38] with default settings and ARAGORN[39] with both default settings and 0.95 default settings as a lower scoring threshold. This analysis generated a collection of suppressor/suppressor-like tRNAs recognizing the three stop codons from each search. Comparison of the canonical tRNAs from different species with these predicted suppressor tRNAs was performed using BLASTn (v.2.5.0+)[58] against the NCBI GenBank nucleotide (NCBI NT) database as previously described.[37,47] The potential identity of the suppressor tRNAs corresponding to different amino acids could be predicted from the high similarity of the sequence alignment between the suppressor tRNAs and the canonical tRNAs of the standard amino acids.

Functional signal and structure prediction of C-terminal extended proteins in maize

C-terminal extensions at least 20 amino acids in length in maize SCR events were scanned for transmembrane helices using TMHMM (v.2.0)[59] with default settings and for nuclear localization signals using cNLS Mapper.[6]°C-terminal extensions at least 12 amino acids in length were searched for peroxisomal targeting signals using PTS1 Predictor[61] and prenylation signals using PrePS.[62] The 3D structures of ZmBTF3 and ZmBTF3x were predicted using Alphafold (v2.0).[63]

Plasmid construction

Luciferase constructs for the luminescence-based SCR assay were generated in the reconstructive pFN19K T7 SP6 flexi vector (Promega Corporation, USA). The coding sequences of the tested genes, *ZmH3C2*, *ZmCEQORH*, *ZmBTF3*, and *Zm00001d009138*, along with their canonical stop codons and inter-stop codon regions (ISRs, the extension regions) were cloned upstream of the coding sequence of the firefly luciferase gene (*Fluc*) without ATG between *EcoRI* and *KpnI* sites. A linker sequence (5′ GGCGGCTCCGGCGGCTCCCTCGTGCTCGGG 3′) was added upstream of the Fluc sequence to avoid potential distortions between the tested protein and Fluc in the fusion protein.[16,26,28] Primer sequences are listed in Table S12.

For subcellular localization analysis, the ZmBTF3 and ZmBTF3x coding sequences were introduced into the pAN580 vector containing the 35S promoter and the coding sequence for GFP. GFP was fused to the C terminus of ZmBTF3 and ZmBTF3x. Primer sequences are listed in Table S12.

In vitro translation for SCR assay

For SCR assay, we used the *in vitro* translation system based on Wheat Germ Extract (Promega Corporation, USA), which has been used for SCR validation as reported previously.[26,28] The constructed plasmid DNA including the test gene and *Fluc* (no ATG) was linearized using BamH I or XbaI enzyme, and 2 μg of the linearized DNA was transcribed *in vitro* using T7 RNA polymerase (Promega Corporation, USA). The obtained RNA was then purified using GeneJET RNA Purification kit (Thermo Fisher Scientific, USA). The concentration and purity of the RNA were measured using Nanodrop 2000 (Thermo Fisher Scientific, USA). A total of 2 μg purified RNA was translated *in vitro* using the Wheat Germ Extract at 25°C for 2 h according to the manufacturer's instructions. Luciferase activity was then measured at room temperature (below 25°C) using the Luciferase Assay System (Promega Corporation, USA) in a Glomax luminometer (Promega Corporation, USA) with a 10-s measurement read for luciferase activity. The *in vitro* SCR rate was calculated by dividing the average luciferase value of the SCR protein by the average luciferase value of the corresponding extended protein with the stop codon mutated to a sense codon.[16,26,28]

Immunoblotting and mass spectrometric analysis

The proteins extracted from the *in vitro* protein expression system were separated by 10% (w/v) SDS-polyacrylamide gel electrophoresis (SDS-PAGE). One PAGE gel was stained with Coomassie Brilliant Blue (CBB), while another was transferred to a polyvinylidene difluoride membrane (0.45 μm, Biosharp, China). Proteins were detected by immunoblotting with anti-Fluc antibodies (1:10,000, ab185924, Abcam, UK). The regions of interest in the CBB staining gel were excised manually according to the immunoblotting band, and then destained with 30% (v/v) ACN in 100 mM NH_4HCO_3 and dried in a vacuum centrifuge. The extracted proteins were reduced with dithiothreitol (DTT) (10 mM DTT in 100 mM NH_4HCO_3) for 30 min at 56°C, then alkylated with iodoacetamide (200 mM IAA in 100 mM NH_4HCO_3) in the dark at room temperature for 30 min. Gel pieces were briefly rinsed with 100 mM NH_4HCO_3 and ACN, respectively. The gel pieces were then divided equally into three parts and separately digested overnight in 12.5 ng/μL trypsin, overnight in chymotrypsin in 25 mM NH_4HCO_3, and for 30 min in 12.5 ng/μL pepsin in PH3 HCl. The peptides were extracted using 60% (v/v) ACN in 0.1% (v/v) trifluoroacetic acid (TFA) three times, and the extracts were pooled and dried completely by a vacuum centrifuge.

LC-MS/MS was performed on a Q-Exactive HF-X mass spectrometer coupled to Easy nLC 1200 (Thermo Fisher Scientific, USA). The peptide mixture was loaded onto a C18-reversed phase column (15 cm, 75 μm inner diameter) packed in-house with RP-C18 (5 μm resin, Thermo Fisher Scientific) in buffer A (0.1% [v/v] formic acid in HPLC-grade water) and separated with a linear gradient of buffer B (0.1% formic acid in 84% [v/v] ACN) at a flow rate of 250 nL/min controlled by IntelliFlow technology over 60 min. MS data were acquired using a data-dependent top-10 method dynamically choosing the most abundant precursor ions from the survey scan (m/z 300-1,800) for HCD fragmentation. The target value was determined based on pAGC. The dynamic exclusion duration was 20 s. Survey scans were acquired at a resolution of 70,000 (m/z 200) and the resolution for HCD spectra was set to 17,500 (m/z 200). Normalized collision energy was 27 eV and the underfill ratio, which specifies the minimum percentage of the target value likely to be reached at maximum fill time, was defined as 0.1%.

The MS data were then separately searched against the target protein sequences (readthrough protein sequences of *H3C2*, *ZmCEQORH*, and *ZmBTF3*) using MaxQuant software (v.1.6.1.10).[52] An initial search was set with a precursor mass window of 6 ppm. The search followed an enzymatic cleavage rule of "unspecified" and a mass tolerance of 20 ppm for fragment ions. Carbamidomethylation of cysteines was defined as fixed modification, while methionine oxidation was defined as variable modification for database searching. The cutoff of the global FDR for peptide and protein identification was set to 0.01.

Subcellular localization of ZmBTF3-GFP and ZmBTF3x-GFP fusions

The GFP fusion plasmids were transformed into maize mesophyll protoplasts by polyethylene glycol-mediated transformation. Protoplast preparation and transfection were conducted as previously described.[77] A plasma membrane-localized marker *35S:ZmPDCK7-mCherry* and a nucleus-localized marker *35S:AtHY5-mCherry* were used for colocalization analyses.[78,79] Fluorescence signals were observed on an LSM800 confocal laser scanning microscope (Zeiss, Germany).

Evolutionary analysis of SCR events

The CDSs and protein sequences of maize, rice, soybean, and *Arabidopsis* genes and proteins, respectively, were downloaded from the public databases described earlier. Orthogroups among the four species were predicted based on the protein sequences by OrthoFinder (v.2.3.11)[63] with default parameters. A total of 19 SCR single-copy orthogroups were detected, and each group contained at least one SCR gene. Next, the ratio of nonsynonymous to synonymous substitutions (d*N*/d*S*) of the main ORF, extensions, and the distal 3′ UTR of the 19 single-copy orthogroups were calculated. Protein sequences were aligned by Muscle[65,66]; the aligned protein sequences were then used to guide the codon-wise alignment of the annotated transcript using PAL2NAL (v. 14)[67]; and after removing gaps and stop codons, the multiple sequence alignment results were uploaded to codeml embedded in the PAML package (v.4.8),[68] and the d*N*/d*S* ratio was computed for each ortholog pair. The parameters of codeml were set as: runmode = 2, model = 2, NSsites = 2, CodonFreq = 2, fix_kappa = 0, and fix_omega = 0. Sequence alignment and d*N*/d*S* calculation were performed for the main ORF, extension, and distal 3′ UTR (120 bp downstream of the second stop codon), respectively. A permutation test was used for d*N*/d*S* comparisons. The number of permutations was set to 1,000.

To calculate the distribution of the d*N*/d*S* ratio and sequence identity for the orthogroup containing GADPH orthologs that underwent SCR simultaneously in three plant species, one representative paralog that harbored multiple gene copies in one orthologous group was selected in each species. The paralog that showed the shortest branch length among all the paralogs was selected to minimize the effect of species-specific changes in the d*N*/d*S* calculation. The protein sequence, including the main ORF, extension, and the distal 3′ UTR (120 bp downstream of the second stop codon), were aligned by Muscle (v.3.5) and edited by GeneDoc (v.2.7.0). The codon-wise and non-gap alignment of the annotated transcripts were obtained using PAL2NAL (v.14) as mentioned above. Finally, the d*N*/d*S* ratio was computed in sliding windows throughout the alignment, with a width of 30 codons and a step size of one codon using codeml embedded in PAML package (v.4.8) with the parameters described above. Sequence identity at nucleotide level was calculated on the same alignment in a sliding window using BLASTn (v.2.6.0)[58] with the following settings: -task blastn-short, evalue = 0.001.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were conducted using R Software (version 3.2.5) or Microsoft Office Excel 2019. The number of samples and the statistical details relating to specific experiments can be found in Figure legends. Adobe Illustrator was used for final adjustment of fonts, colors, and positions of figures. Drawings were also created in Adobe Illustrator.