



Genome-Wide Identification and Characterization of Small Peptides in Maize

Yan Liang[†], Wanchao Zhu[†], Sijia Chen, Jia Qian and Lin Li*

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China

OPEN ACCESS

Edited by:

Qingyu Wu,
Chinese Academy of Agricultural
Sciences (CAAS), China

Reviewed by:

Liuji Wu,
Henan Agricultural University, China
Xu Fang,
Shandong University, China

*Correspondence:

Lin Li
hzauillin@mail.hzau.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 15 April 2021

Accepted: 20 May 2021

Published: 16 June 2021

Citation:

Liang Y, Zhu W, Chen S, Qian J and
Li L (2021) Genome-Wide
Identification and Characterization of
Small Peptides in Maize.
Front. Plant Sci. 12:695439.
doi: 10.3389/fpls.2021.695439

Small peptides (sPeptides), <100 amino acids (aa) long, are encoded by small open reading frames (sORFs) often found in the 5' and 3' untranslated regions (or other parts) of mRNAs, in long non-coding RNAs, or transcripts from introns and intergenic regions; various sPeptides play important roles in multiple biological processes. In this study, we conducted a comprehensive study of maize (*Zea mays*) sPeptides using mRNA sequencing, ribosome profiling (Ribo-seq), and mass spectrometry (MS) on six tissues (each with at least two replicates). To identify maize sORFs and sPeptides from these data, we set up a robust bioinformatics pipeline and performed a genome-wide scan. This scan uncovered 9,388 sORFs encoding peptides of 2–100 aa. These sORFs showed distinct genomic features, such as different Kozak region sequences, higher specificity of translation, and high translational efficiency, compared with the canonical protein-coding genes. Furthermore, the MS data verified 2,695 sPeptides. These sPeptides perfectly discriminated all the tissues and were highly associated with their parental genes. Interestingly, the parental genes of sPeptides were significantly enriched in multiple functional gene ontology terms related to abiotic stress and development, suggesting the potential roles of sPeptides in the regulation of their parental genes. Overall, this study lays out the guidelines for genome-wide scans of sORFs and sPeptides in plants by integrating Ribo-seq and MS data and provides a more comprehensive resource of functional sPeptides in maize and gives a new perspective on the complex biological systems of plants.

Keywords: maize, small open reading frame, small peptide, Ribo-seq, mass spectrometry

INTRODUCTION

Small peptides (sPeptides), which are defined as those peptides shorter than 100 amino acids (aa), represent a class of small molecules with important roles in various biological processes and are translated from small open reading frames (sORFs) shorter than 300 nucleotides (Wang S. et al., 2020). sORFs are widely distributed in the genome and are likely to be located at 5' and 3' ends in the untranslated regions of mRNAs [upstream ORFs (uORFs) and downstream ORFs (dORFs)], in the internal regions of annotated ORFs of mRNAs (mORFs) but in a different reading frame, in the short isoforms in spliced mRNAs, and in RNAs produced by transcribed loci in the introns or intergenic regions (Couso and Patraquim, 2017). Although long non-coding RNAs (lncRNAs) are not defined as encoding proteins, some lncRNAs contain sORFs that are engaged by the ribosome, potentially encoding sPeptides (Ruiz-Orera et al., 2018; Ruiz-Orera and Albà, 2019).

sPeptides play important regulatory roles in multiple physiological processes, including growth, development, reproduction, and stress responses (De Coninck and De Smet, 2016). For example, the artificial synthesis and application of the sPeptide hormone insulin were one of the greatest achievements of the twentieth century. Small signaling peptides or peptide hormones in plants, such as cystein-rich peptides, are 5–75 aa long and function as signaling molecules in cell-to-cell communication in defense responses, development, reproduction, and plant–bacteria symbiosis (Marshall et al., 2011; Wang S. et al., 2020). A conserved sORF, *TAS3a*, is associated with the biogenesis of *trans*-acting small-interfering RNAs (tasiRNAs) in *Arabidopsis thaliana*, and a small signaling peptide, *IMMUNE RESPONSE PEPTIDE*, regulates the expression of some defense genes and responds to bacterial or fungal infection in rice (*Oryza sativa*) (Bazin et al., 2017; Wang P. et al., 2020). Multiple sPeptides show high expression in root tissues and a tight association with root growth and absorption. Overexpression of the *C-TERMINALLY ENCODED PEPTIDE1* gene inhibited root growth in *Arabidopsis* (Ohyama et al., 2008). In rice, overexpression of the sPeptide genes, *DROUGHT AND SALT STRESS RESPONSE1*, enhanced drought and stress tolerance, and the knockdown of the sPeptide gene, *OscADMIUM TOLERANT 3* resulted in a decreased tolerance to aluminum (Xia et al., 2013; Cui et al., 2018).

sORFs and sPeptides have been identified through multiple methods, and their functions have been explored in plants. However, to date, only a few sPeptides have been studied through forward genetic screens because the small size of their encoding sORFs makes them difficult to be targeted by the common mutagenesis methods. Thus, the limited availability of verified sPeptides that can be used as training sets limits the ability of machine learning algorithms to predict sPeptides (Zhou et al., 2013). Further complicating sPeptide identification, many algorithms, such as *de novo* genome annotation, exclude putative proteins of <100 aa in length (Basrai et al., 1997; Claverie, 1997). However, 606,285 potential sORFs (25–250 codons) have been identified in the *A. thaliana* genome (Lease and Walker, 2006).

The bioinformatics and experimental methods have been developed to scan sPeptides encoded by sORFs on a genome-wide level. Bioinformatics approaches based on sequence conservation, functional domains or motifs, gene family clustering, and expression support have been used to search for homologs of known peptides and to predict novel peptides, such as the ROT-FOUR-LIKE/DEVIL (RTFL/DVL) family in *Arabidopsis* (Wen et al., 2004; Guillén et al., 2013; Guo et al., 2015). The transcriptome analysis can reveal the expression of transcripts containing candidate sORFs, but it cannot validate the presence of translational products of these sORFs. The ribosome profiling [also called ribosome sequencing (Ribo-seq)] reveals ribosome footprints by extracting and sequencing RNA that is protected by ribosomes; this can resolve three-nucleotide periodicity, enabling precise definition of translated regions within individual transcripts (Wu et al., 2019). However, several factors introduce contamination in Ribo-seq reads, including structured RNAs and RNAs embedded in protein complexes like rRNA and small nucleolar RNAs (snoRNA), as well as scanning or stalled ribosomes that do not engage in translation

(Guttman et al., 2013; Guydosh and Green, 2014; Brar and Weissman, 2015; Archer et al., 2016). Additionally, sPeptides can be directly detected globally using mass spectrometry (MS). For example, an integrated peptidogenomic pipeline using high-throughput MS to probe a customized six-frame translation database was generated and applied to identify non-conventional peptides in maize (*Zea mays*) and *Arabidopsis* (Wang S. et al., 2020). However, since peptides with a bigger size and higher abundance have a better chance of being detected by using MS, the identification of sPeptides solely through MS might miss some of the peptides.

In this study, we collected a large-scale dataset including mRNA sequencing (mRNA-seq), Ribo-seq, and MS data from six tissues (each with two replicates) of maize and performed a *de novo* translome annotation using the RiboCode software (Xiao et al., 2018). We extracted ORFs of 3–300 nucleotides from the dataset and identified 9,388 sORFs potentially encoding sPeptides. These sORFs showed different Kozak region sequences, higher specificity of translation, and high translational efficiency compared with the canonical protein-coding genes. Furthermore, we searched all sORF sequences in the MS data of the corresponding tissues and verified 2,695 sPeptides. These verified sPeptides clustered perfectly with tissues/replicates, and the verified sPeptides showed higher expression and were longer than the unverified sPeptides. Importantly, the expression in translome of some annotated sPeptides was positively correlated with that of parental genes, which showed the enrichment of multiple functional gene ontology (GO) terms related to abiotic stress and development. Taken together, the results of this study provide a more comprehensive resource for functional analysis of sPeptides and give helpful information for functional genomics analysis.

MATERIALS AND METHODS

Plant Materials

The seeds of maize (*Z. mays* L.) inbred line B73 were planted in a greenhouse under a temperature and a photoperiod of 30°C for 16 h of light and 25°C for 8 h of darkness. The stem, root, leaf, and whole seedling tissues were collected 14-days after planting, and the ear and tassel tissues were collected in the V12 stage with two biological replicates for each tissue.

Analysis of RNA-Seq and Ribo-Seq Raw Data

The RNA-seq and Ribo-seq data of these plant samples were collected from our previous study (Zhu et al., 2021). A non-coding RNA data set including rRNA, tRNA, and snoRNA sequences were downloaded from the database Rfam (<http://rfam.xfam.org/>). After removing the adaptors, the collected data were mapped to this dataset using bowtie2 v2.4.1 with default parameters (Langmead and Salzberg, 2012; Kalvari et al., 2018), and the unmapped reads were kept for downstream analysis. The unaligned mRNA-seq and Ribo-seq reads were mapped to the exon sequences and coding sequences (CDSs) of the B73 reference genome (AGPv4), respectively, using STAR v2.7.3 with default parameters (Schnable et al., 2009; Dobin et al., 2013).

Identification of sORFs and Calculation of Transcriptional and Translational Abundance

By comparing the alignment results in the BAM format with the B73 reference genome (AGPv4), we identified ORFs in different tissues using the RiboCode software (<https://github.com/xryanglab/RiboCode>) with default parameters (Xiao et al., 2018). ORFs identified using the RiboCode were classified into seven types as follows:

1. “annotated”: ORFs, overlapping with annotated CDSs, have the same start and stop codon with annotated CDSs.
2. “uORF”: ORFs located upstream of annotated CDSs, not overlapping with annotated CDSs.
3. “dORF”: ORFs located downstream of annotated CDSs, not overlapping with annotated CDSs.
4. “Overlap_uORF”: ORFs located upstream of annotated CDSs and overlapping with annotated CDSs.
5. “Overlap_dORF”: ORFs located downstream of annotated CDSs and overlapping with annotated CDSs.
6. “Internal”: ORFs located internal regions of annotated CDSs, but in a different reading frame.
7. “novel”: ORFs derived from non-coding genes or non-coding transcripts of the coding genes.

The ORFs shorter than 300 nucleotides were considered to be sORFs potentially encoding peptides. Then, the coordinate information of all these sORFs was extracted and merged with the B73 reference genome annotation in the GTF format (https://ftp.ensemblgenomes.org/pub/plants/release-50/gtf/zea_mays/Zea_mays.B73_RefGen_v4.50.gtf.gz) using in-house shell scripts and Stringtie v2.1.4 (parameter: `-merge -m 0`) (Pertea et al., 2015). The aa sequences of all sORFs were generated by RiboCode and merged together for further analysis.

Using the merged genome annotation, transcriptional and translational abundance was calculated by fragments per kilobase of exon model per million mapped reads (FPKM) using Cufflinks v2.2.1 with parameters: `-p 5 -G` (Trapnell et al., 2012). Only the unique reads of RNA-seq and Ribo-seq mapped to exons and CDSs were used for the calculation of abundance.

Protein Preparation and MS

The tissue from maize was grounded using liquid nitrogen and then transferred into a 5-mL centrifuge tube with a suitable volume of lysis buffer [1% TritonX-100, 10 mM dithiothreitol, 1% Protease Inhibitor Cocktail, 50 μ M 2,6-diamino-3,5-dithiocyanopyridin, 3 μ M trichostatin A, 50 mM *N*-arachidonyl maleimide, and 2 mM ethylenediaminetetraacetic acid (EDTA)]. The mixture was sonicated three times on ice using a high-intensity ultrasonic processor (Scientz), and an equal volume of Tris-saturated phenol (pH 8.0) was added and further vortexed for 5 min. The upper phenol phase was transferred to a new centrifuge tube after centrifugation. Then, at least four volumes of ammonium sulfate-saturated methanol were added to precipitate the proteins at -20°C for at least 6 h. After centrifugation, the precipitate with proteins was collected and washed with ice-cold methanol once, followed by ice-cold

acetone three times. The protein was redissolved in 8 M urea, and the concentration was determined using a BCA kit based on the instructions of the manufacturer. After digestion with trypsin, the samples were submitted for the MS detection on the Thermo Scientific Q Exactive platform using a label-free method. The resulting MS data were processed with the Maxquant search engine (v.1.5.2.8). Tandem mass spectra were searched against the aa sequences of sORFs identified by RiboCode in the translome data (Xiao et al., 2018). The mass tolerance for precursor ions was set as 20 ppm in the first search and 5 ppm in the main search, and the mass tolerance for fragment ions was set as 0.02 Da. The false discovery rate (FDR) was adjusted to $<1\%$, and the minimum score for peptides was set to >40 .

The MS data of maize leaves sampled from the 14-day-old seedlings were collected from our previous study (Zhu et al., 2021).

Calculation of Shannon Entropy and Translation Efficiency (TE)

To compare the TE of sORFs and other canonical transcripts in different tissues, we selected transcripts expressing in both the transcriptome and translome (FPKM ≥ 0.5). RPKM for single-end sequencing is the unit to quantify gene's expression level, equivalent to FPKM for pair-end sequencing. The RPKM is defined as follows:

$$RPKM = \frac{\text{Exon Mapper Reads} * 1,000,000,000}{\text{Total Mapped Reads} * \text{Exon Length}}$$

In the pair-end sequencing two paired reads is a fragment. The FPKM is defined as follows:

$$FPKM = \frac{\text{Exon Mapper Fragments} * 1,000,000,000}{\text{Total Mapped Fragments} * \text{Exon Length}}$$

For each transcript, we calculated TE by the following equation:

$$TE = \frac{RPKM \text{ of Ribo} - seq}{FPKM \text{ of RNA} - seq}$$

We compared the tissue specificity of expression between verified and non-verified peptides by analyzing the Shannon entropy of peptides. For each sORF, we defined the Shannon entropy of expression-level across different tissues as follows: Given expression levels of a sORF in N tissues, the proportion of expression in tissue i out of the sum of all expression-levels in all tissues:

$$P(i) = \frac{FPKM(T_i)}{\sum_{i=1}^N FPKM(T_i)}$$

Shannon entropy:

$$SE = - \sum_{i=1}^N P(i) \log_2 [P(i)]$$

Clustering, Gene Ontology Enrichment Analysis, and Statistical Analysis

The verified and previously annotated 501 peptides were clustered into three groups using the *gplots* package in R with default parameters. For each group of peptides, the GO enrichment analysis was performed on their parental genes using the AgriGO v 2.0 webserver (Tian et al., 2017). The GO terms with an FDR threshold of 0.05 were considered as significant terms.

All of the statistical analyses in this study were performed using R version 4.0.4.

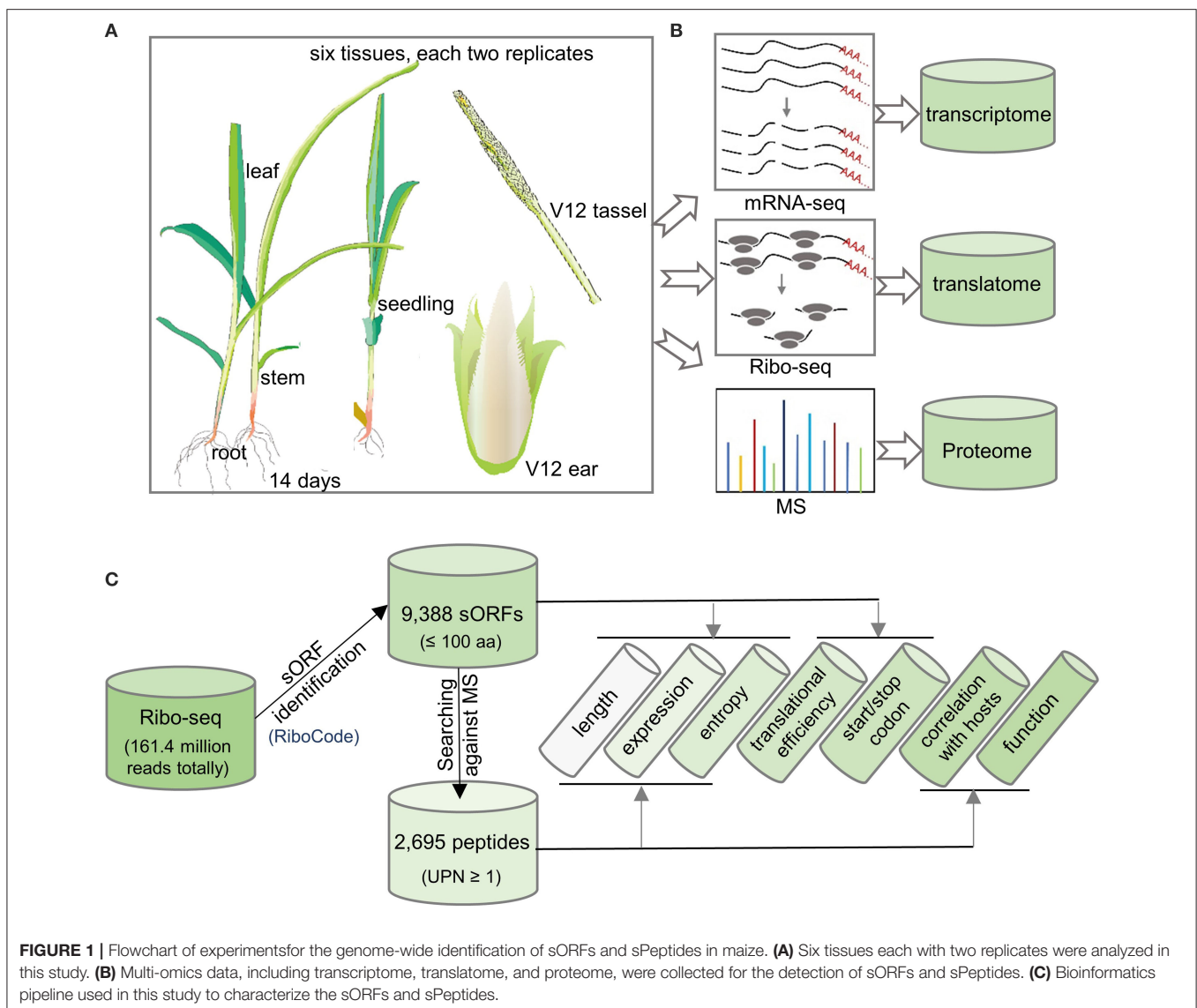
RESULTS

The Widespread Existence of sORFs in Maize

In a previous study by authors, we sampled whole seedlings, roots, stems, and leaves of 14-day-old maize plants, as well

as ears and tassels at the V12 stage with two replicates each and used these to perform RNA sequencing (RNA-seq), Ribo-seq, and MS analysis (MS only for leaf samples) (Figures 1A,B and Supplementary Table 1) (Zhu et al., 2021). We obtained a total of 236 million RNA-seq reads to quantify transcript abundance. We also collected 161.4 million Ribo-seq reads to map ribosome occupancy on genome-wide transcripts (Brar and Weissman, 2015). Additionally, MS was performed to detect and quantify protein abundance for all samples (except the leaf samples) in this study. Taking these results together, we collected a comprehensive transcriptome, translome, and proteome dataset for the genome-wide identification of sORFs and sPeptides in maize (Figure 1B).

After filtering low-quality reads and contaminant reads in the Ribo-seq data (as shown in the “Materials and Methods” section), we mapped the remaining reads to the maize B73 reference genome (Schnable et al., 2009). By applying stringent filters using the RiboCode software, we identified 9,388 sORFs potentially



encoding putative sPeptides (Xiao et al., 2018). Comparison of sORF annotation with the latest maize reference annotation B73 RefGen v4.50 (ensembl) demonstrated 7,589 (80.84%) newly identified sORFs, which complements the functional annotation of the maize genome. Finally, we verified the presence of sPeptides using the MS analysis and identified 2,695 sPeptides, which are supported by the presence of one or more unique peptides ($UPN \geq 1$) in the MS analysis (Figure 1C). These results indicate the widespread existence of sORFs and sPeptides in maize.

Maize sORFs Are a Class of Distinct Translational Elements

By comparing data from this study to the B73 reference genome annotation, we uncovered a total of 2,907 (30.97%) upstream sORFs (uORFs), 485 (5.17%) overlap upstream sORFs (overlap uORFs), 300 (3.2%) internal sORFs, 301 (4.27%) overlap downstream sORFs (overlap dORFs), 3,445 (36.7%) downstream sORFs (dORFs), 1,799 (19.16%) annotated sORFs, and 49 (0.52%) novel sORFs (as shown in the “Materials and methods” section). Similar to genes, sORFs are more likely to be enriched toward the telomeres and depleted in the pericentromeric regions (Figure 2A). All sORFs encoded sPeptides shorter than 100 aa in length. However, the different types of sORFs showed distinct length distribution. Annotated sORFs were significantly longer than other types of sORFs ($p < 2.2e-16$, Mann-Whitney U test), while novel and internal sORFs were significantly shorter than the others ($p < 2.445e-09$, $p < 2.2e-16$, respectively, Mann-Whitney U test) (Figure 2B). Interestingly, the three shortest uORFs potentially encode peptides with only 5 aa.

The upstream and downstream sequences of start and stop codon sites might be associated with the translational efficiency of ORFs (Hinnebusch et al., 2016). Thus, we extracted the upstream and downstream 5 bp sequences of start and stop codons of the sORFs as well as canonical transcripts and further performed a motif enrichment using R package gseqlogo (Wagih, 2017). We found that the sequences of the start and stop codons of sORFs were roughly similar to those of conventional genes. However, the frequency of upstream and downstream sequences of start and stop codons was significantly different between sORFs and other canonical transcripts (Figures 2C,D). These results suggest that sORFs exhibit different features (e.g., the frequency of translational start sites) compared with conventional genes.

Moreover, we quantified and compared the expression levels of sORFs and conventional transcripts at the transcriptome level. Although the transcriptome abundance of both conventional genes and sORFs could discriminate different tissues, the expression patterns differed dramatically among the different tissues (Figures 2E,F). Furthermore, we calculated the Shannon entropy of expression levels across all 12 samples for sORFs and other protein-coding transcripts and demonstrated that sORFs are significantly more likely to be tissue-specific than the canonical transcripts ($p < 2.2e-16$, Mann-Whitney U test) (Figure 2G).

Additionally, we checked the difference in translational efficiency between sORFs and canonical protein-coding genes

across all tissues. Unexpectedly, sORFs exhibited significantly higher translational efficiency than canonical protein-coding genes for all 11 samples (the sequencing library of ear1 for RNA sequencing was constructed unsuccessfully) (Figure 2H), which is likely associated with the different base frequency upstream and downstream of the start and stop codons. All these results indicate that sORFs exhibit distinct genomic and expression features compared with canonical protein-coding genes.

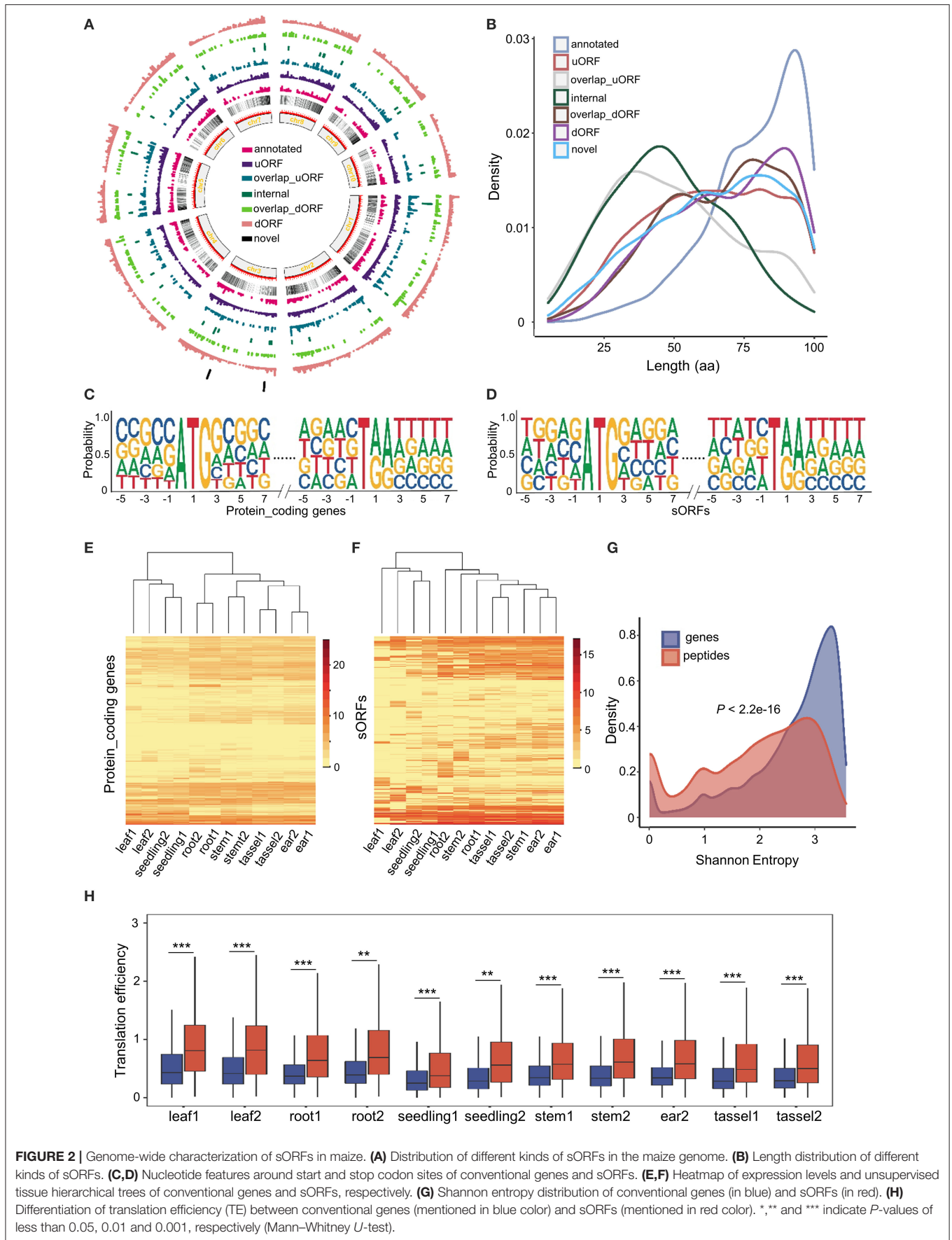
Genome-Wide Classification and Characterization of sPeptides in Maize

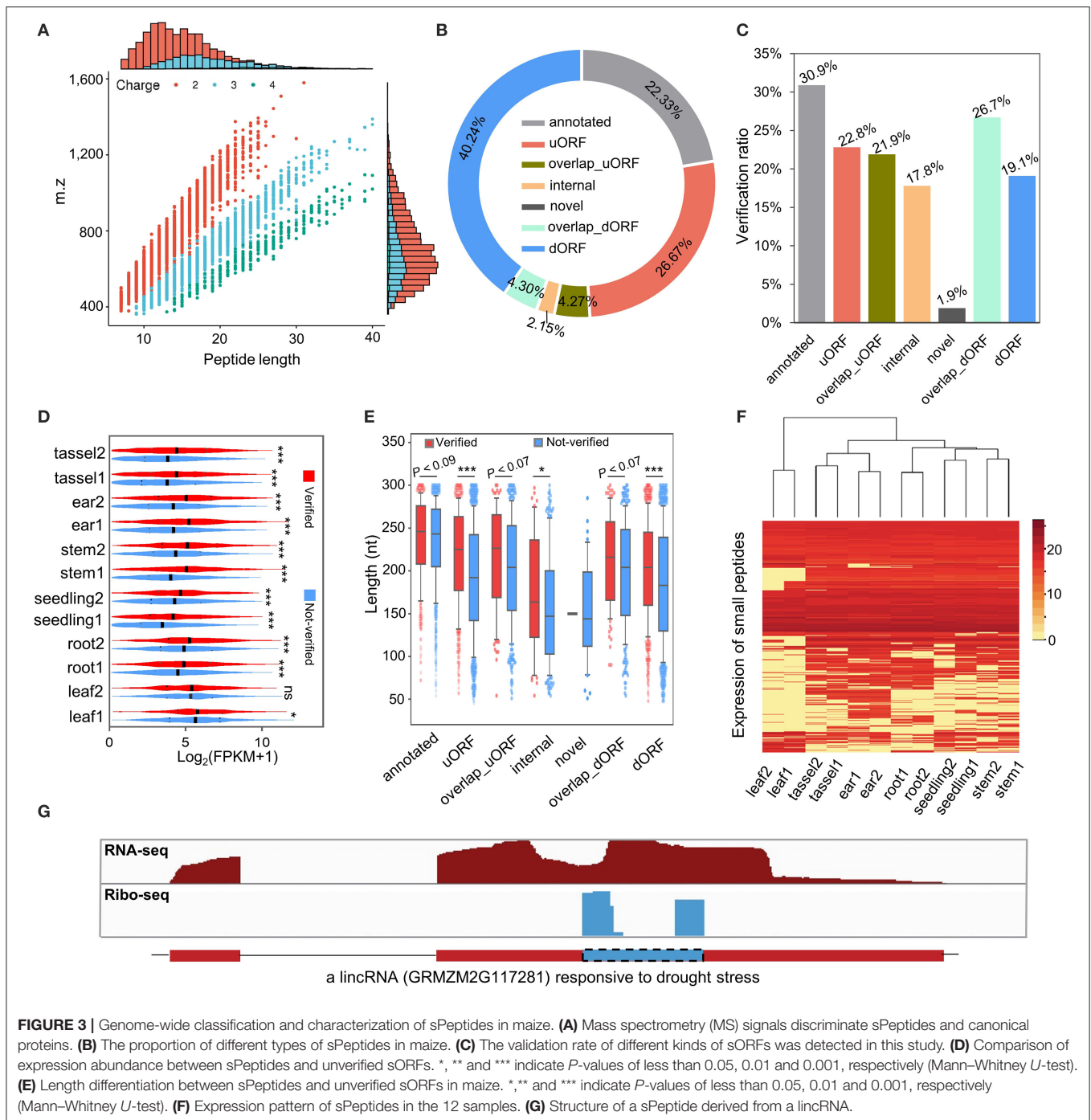
To evaluate whether sORFs are able to translate stable peptides in maize, we performed a proteogenomic analysis by searching the aa sequences of sORFs against the MS data (Walley and Briggs, 2015). Consistent with the definition of sPeptides as <100 aa, the sPeptides were mainly concentrated in the range of 0–10 kDa in the MS validation, which corresponds to the length of 5–100 aa (Figure 3A and Supplementary Figure 1A). By using the MS analysis, we verified 2,596 sPeptides of the 9,388 sORFs identified from Ribo-seq data. Of these sPeptides, ~85% were detected in more than five tissues (Supplementary Figure 1B). Unexpectedly, only about 20% of sPeptides were derived from annotated transcripts of the reference genome, suggesting that the number of sPeptides in maize is largely underestimated (Figure 3B). However, the comparison of the validation ratio of sPeptides derived from different types of sORFs demonstrated that sORFs annotated previously in the reference genome are most likely to be validated, while those from novel sORFs are least likely to be validated (Figure 3C).

The comparison of Shannon entropy between different kinds of sPeptides and unverified sORFs indicated that sPeptides with MS evidence were more uniformly expressed in the 12 samples (Supplementary Figure 1C). Notably, sPeptides exhibit higher transcriptome abundance and longer CDS in corresponding sORFs than those of the unverified sORFs, which may be due to the limitation of the detection power of MS (Figures 3D,E). Then, we clustered the 12 samples based on the expression-level variation of all detectable sPeptides and found that the sPeptides can robustly discriminate tissues, and the replicates of the same tissue were also clustered (Figure 3F). Interestingly, we found that many sPeptides are derived from putative-defined lincRNA regions. For example, a long intergenic non-coding RNA (lincRNA) *GRMZM2G117281_T01* annotated in the reference genome of maize was detected to be translated to a sPeptide (Figure 3G). A previous study proposed that *GRMZM2G117281* could be responsive to drought stress (Zhang et al., 2014), suggesting a potential functional role of this sPeptide. All these results demonstrate that sPeptides are abundant in the maize genome, exhibit distinct genomic features, and may function in different biological processes.

Annotated sPeptides Are Correlated With Their Corresponding Parental Genes and Might Function in Multiple Processes

A fraction of the ORFs located within annotated genes was shorter than 300 nucleotides. Some of them have been

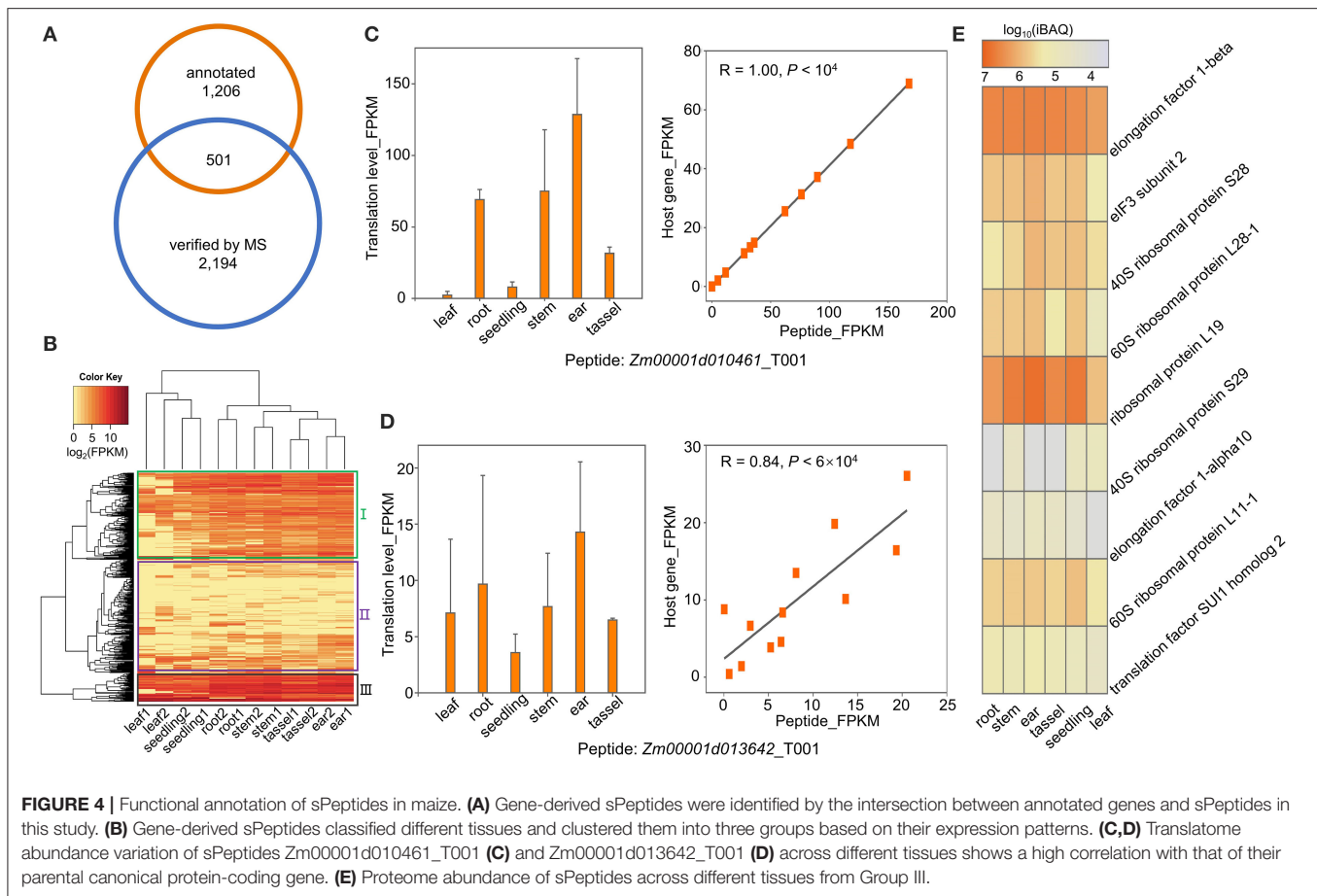




demonstrated to be capable of encoding sPeptides with biological functions such as signal peptides (Hsu and Benfey, 2018; Ruiz-Olera and Albà, 2019). In this study, we compared the sPeptides with the functional annotation in the maize B73 RefGen v4 reference genome. A total of 501 out of 2,596 sPeptides identified by Ribo-seq and MS were encoded by sORFs annotated in the reference genome (Figure 4A). Based on the expression of the sPeptides quantified by Ribo-seq, we clustered 501 genes into three groups (Figure 4B). Furthermore, the translational abundance of 501 sPeptides

was positively correlated with that of their host genes (Supplementary Figure 2D), suggesting that sPeptides might function by orchestrating parental gene expression.

To explore the potential functions of sPeptides, we performed the GO enrichment analysis for the three parental gene sets that correspond to the three different sPeptide groups. Group I parental genes were significantly enriched in the RNA splicing process (Supplementary Figure 2A). *Zm00001d010461*, *Zm00001d042725*, and *Zm00001d024593*, enriched in the GO term “RNA splicing,” are annotated to encode small nuclear



ribonucleoprotein family-like (LSM) proteins. Generally, LSM proteins are associated with development, response to stress, and abscisic acid signaling *via* mRNA splicing and processing in *Arabidopsis* (Xiong et al., 2001; Perea-Resa et al., 2012; Golisz et al., 2013; Cui et al., 2014; Okamoto et al., 2016). The abundance of the sPeptide, Zm00001d010461_T001 varied dramatically in different tissues and showed a highly positive correlation with the expression of Zm00001d010461 ($R = 1$, $p < 1e-04$) (Figure 4C). Therefore, we speculated that the sPeptide, Zm00001d010461_T001 could function in stress tolerance and development in maize.

For the second group of parental genes, the GO enrichment showed that these genes are related to organonitrogen compound biosynthetic and metabolic processing (Supplementary Figure 2B). Nitrogen is essential for plant growth and development and is also helpful for tolerance against biotic and abiotic stresses (Oh et al., 2017; Zipfel and Oldroyd, 2017; Arora et al., 2020). Notably, Zm00001d034602 is annotated to function in the organonitrogen compound metabolic process, and its abundance in the translome was significantly associated with that of the sPeptide Zm00001d034602_T001 (Figure 4D). AT2G22425, the ortholog of Zm00001d013642 in *Arabidopsis*, was reported to function in signal peptide processing during *Cabbage leaf curl virus* infection, indicating that the sPeptide Zm00001d013642_T001 may be involved in pathogen response

through organonitrogen compound biosynthetic and metabolic processing (Ascencio-Ibáñez et al., 2008).

The sPeptide parental genes in Group III were uniformly expressed across different tissues and showed high abundance at both the translome and proteome levels (Figures 4B,E). These genes are mainly enriched in translation and peptide biosynthetic processes (Supplementary Figure 2C). Nine genes with the most significant GO enrichment of translation are annotated as “subunit of ribosome and translation-regulatory factors.” These peptides are likely to participate in the fundamental biological pathways and affect multiple agronomic traits.

DISCUSSION

As a new frontier in the study of molecular players in life science, sORFs and sPeptides have been reported to be involved in several biological processes in plants. In this study, we collected RNA-seq, Ribo-seq, and MS data from six tissues of the maize reference inbred B73 with two replicates and performed a genome-wide *de novo* scan of translational elements using the RiboCode software (Xiao et al., 2018). We extracted ORFs encoding peptides with length ranging from 2 to 100 aa and identified 9,388 sORFs potentially encoding sPeptides. Then, we confirmed over 2,000 sPeptides by searching against MS data. Comparison of sORFs/sPeptides with canonical proteins demonstrated the

distinct features of these non-canonical molecules. Finally, the functional GO analyses indicated that these sPeptides are likely to be involved in the response to abiotic stress and plant development (Couso and Patraquim, 2017). This study presents a biological pipeline combining multi-omics data for the genome-wide scan of sORFs and sPeptides in maize, which paves the way for further functional study of sPeptides in maize.

A Robust Bioinformatics Pipeline for the Genome-Wide Scan sORFs and sPeptides Combining Multi-Omics Data of RNA-Seq, Ribo-Seq, and MS

lncRNAs have a low probability of encoding proteins and are more likely to encode sPeptides. lncRNAs have been widely uncovered by RNA-seq in maize (Li et al., 2014). The sORFs or sPeptides are detected by the bioinformatics analyses solely based on RNA-seq data lack representation at the translational level. Ribosome profiling (Ribo-seq), an RNA sequencing technique focused on the translational level, has been developed and used to monitor translation in real-time during protein biosynthesis (Ingolia et al., 2009). Since ribosomes directly decipher mRNA every three nucleotides, the periodic feature of ribosome footprints can be used to examine unannotated ORFs (Calviello et al., 2016; Hsu et al., 2016). Therefore, sORFs potentially encoding peptides that may not be detected by proteomics-based methods could be uncovered by combining ribosome profiling and bioinformatics analyses (Couso and Patraquim, 2017; Olexiouk et al., 2018; Ruiz-Orera and Albà, 2019).

Due to the presence of reads from the elongating and initiating stage in Ribo-seq experiments, non-AUG start site prediction becomes difficult and false-positive results are introduced (Olexiouk et al., 2018). Wang et al. generated an integrated proteomic pipeline using high-throughput MS to probe a customized six-frame translation database and applied it to a large-scale detection of non-conventional peptides in plants (Wang S. et al., 2020). In total, 1,993 and 1,860 non-conventional peptides were identified in maize and *Arabidopsis*, respectively. However, limited by the systematic error of the technology, in which proteins with higher abundance or larger size have better chances of being detected by the MS, the approaches used to identify peptides based on their mass spectra are likely to have high rates of false negatives (Slavoff et al., 2013; Olexiouk et al., 2018).

Thus, we generated a pipeline for the genome-wide scan of sORFs and sPeptides in maize combining sensitive and real-time Ribo-seq monitoring data as well as less-sensitive but direct MS data. In this study, the annotated sORFs had the highest verification ratio (Figure 3C), which may be associated with the fact that annotated sORFs generally have longer length than the other types of sORFs (Figure 2B). The corresponding peptides of annotated sORFs are more easily identified by the MS because of their larger molecular mass. We confirmed 2,695 sPeptides by the MS, which accounted for 28.71% of all identified sORFs (9,388). Therefore, Ribo-seq provides an unprecedented chance to detect more

potential sPeptides likely with low abundance or smaller relative molecular mass that are difficult to detect by the proteomics-based methods.

To further validate the capacity of pipeline of authors to detect sPeptides, we downloaded and analyzed the MS data from the previous study which identified non-conventional peptides in maize-based on a peptidogenomic method (Wang S. et al., 2020). All sORFs identified from the Ribo-seq data of six different maize tissues were searched against the downloaded MS data that were obtained only from maize leaves at the V3 stage, and 158 sORFs showed peptides evidence. Of these validated sORFs by Wang's MS data, 66 sORFs were also verified in the total MS data (Supplementary Figure 3). Although the number and types of tissues are different, there is a proportion of overlapped sPeptides (~42%) between this study and the previous study based on the non-digestion MS data, reflecting a certain degree of reliability in pipeline of authors. Moreover, the sPeptides, identified in this study, are complementary to previous studies to extend the knowledge of sPeptides in maize.

Potential Functional Roles of sPeptides

Numerous studies have reported that small secreted peptides are involved in different physiological processes including plant growth, development, reproduction, and stress responses (De Coninck and De Smet, 2016). Cysteine-rich peptides, a type of signal peptide, play important roles in developmental patterning as well as in plant-pathogen responses and symbiosis (Hemu et al., 2018). Based on a proteomics method, 1,993 unannotated peptides were identified in maize leaves, which were significantly enriched in regions identified from genome-wide association studies of agronomic traits and appear to be under domestication selection (Wang S. et al., 2020). Thus, the sPeptides identified in this study are likely to function in multiple pathways in maize.

Based on the expression patterns between tissues, the 501 gene-locus-encoded sPeptides annotated in the reference genome and verified by the MS were clustered into three groups through a hierarchical clustering algorithm (Figure 4B). Genes in Group I were significantly enriched in the pathway related to mRNA splicing and processing. A large number of studies reported that alternative splicing plays vital roles in growth, development, and responses to stress (de Francisco Amorim et al., 2018; Szakonyi and Duque, 2018; Li et al., 2021). In this study, the sPeptide-associated parental genes, *Zm00001d010461*, *Zm00001d042725*, and *Zm00001d024593*, which are annotated as LSM protein family genes, belong to Group I. In *Arabidopsis*, LSM proteins participate in mRNA splicing and degradation and thus regulate the development and tolerance to stress (Perea-Resa et al., 2012; Golisz et al., 2013; Cui et al., 2014; Okamoto et al., 2016). Moreover, *SAD1*, encoding a polypeptide similar to multifunctional LSM proteins, modulates abscisic acid signal transduction and biosynthesis in *Arabidopsis* through mRNA metabolism (Xiong et al., 2001). The second group (Group II) of sPeptide parental genes was enriched in organonitrogen compound biosynthetic and metabolic processes (Supplementary Figure 2B). Signal

molecules, such as reactive oxygen species, calcium, reactive nitrogen species, salicylic acid, and ethylene, are important in plant development and pathogen infection (Oh et al., 2017; Zipfel and Oldroyd, 2017). The form of nitrogen, such as nitrate or ammonium, plays a vital role in the production of these signal molecules (Oh et al., 2017; Arora et al., 2020). The parental genes of the Group II sPeptide, *Zm00001d013642_T001* and *Zm00001d013642*, were annotated in “organonitrogen compound metabolic processing,” and its ortholog *AT2G22425* was differentially expressed in *Cabbage leaf curl virus*-infected *Arabidopsis* leaves (Ascencio-Ibáñez et al., 2008). These results imply that *Zm00001d013642_T001* may be involved in signal transduction during the pathogen response through organonitrogen compound biosynthetic and metabolic processes. The last group (Group III) of genes were uniformly expressed in different tissues and with relatively high abundance at the translome level (Figure 4B). There were nine sPeptide parental genes annotated as “subunit of translation start and elongation factors” in Group III. Overall, sPeptides might participate in multiple pathways in different manners to influence the plant life cycle.

Precursors of miRNAs May Encode sPeptides

Some lncRNAs and circular RNAs were detected *in vivo* with evidence for the production of peptides in the shotgun MS data (van Heesch et al., 2019), which implies that non-coding transcripts are likely to play both coding and non-coding roles. A previous study reported that microRNAs (miRNAs) are capable of encoding peptides through the transcripts of their corresponding precursors (Chen et al., 2020). Primary miRNAs have been reported to encode regulatory peptides in *Arabidopsis*, grapevine (*Vitis vinifera*), soybean (*Glycine max*), and *Medicago* sp.; these peptides are named miRNA-encoded peptides (miPEPs) (Ren et al., 2021). For example, miPEP171d1 plays a regulatory role in adventitious root formation and response to stress in plants (Ma et al., 2014; Gao et al., 2019; Chen et al., 2020).

In this study, we aligned the aa sequences of identified sPeptides against the reference sequences of the precursors of miRNAs using BLAST+ (version 2.7.1). Two sPeptides validated by the MS are encoded by pre-miRNAs: ENSRNA049997513-T1 and ENSRNA049997089-T1. In a previous study, *zma-miR159d*, the mature product of ENSRNA049997513-T1, was predicted to target genes encoding MYB transcription factors (Samad et al., 2017). Moreover, *zma-miR159d* was found to be involved in the degradation of chlorophyll that induced earlier leaf senescence between different maize inbred lines (Wu et al., 2016). Considering the evidence of their translome and proteome study, we speculated that *zma-miR159d* could produce a sPeptide. The other sPeptide is encoded by ENSRNA049997089-T1, which is the primary transcript of *zma-MIR2275d*. *zma-MIR2275*, with a maximum expression in the fertile maize

anther, plays an important role in anther development and thereby influencing male reproduction in maize (Zhai et al., 2015; Huang et al., 2019). Furthermore, *zma-miR2275* also affects the drought tolerance by directly targeting drought-related mRNAs. The identification of the sPeptide encoded by the pre-miRNA ENSRNA049997089-T1 provides a perspective on the manner of *zma-MIR2275* function as a regulator of development and stress tolerance in maize.

DATA AVAILABILITY STATEMENT

The MS proteomics data for this project have been deposited at the ProteomeXchange Consortium with the dataset identifier PXD025997.

AUTHOR CONTRIBUTIONS

LL designed and supervised this study. YL, WZ, SC, and JQ performed the data analysis. YL, WZ, and LL prepared the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Natural Science Foundation of China (92035302; 31922068) and the Huazhong Agricultural University Scientific & Technological Self-Innovation Foundation (2662020ZKPY017).

ACKNOWLEDGMENTS

We thank all the contributors in the maize research community for the generosity of sharing the reference maize inbred, B73.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.695439/full#supplementary-material>

Supplementary Figure 1 | Size distribution of peptides in MS and comparison between verified and unverified sORFs. **(A)** Statistics of the mass of peptides. **(B)** Percentage of verified sORFs and their composition. **(C)** Comparison of Shannon entropy between different kinds of verified and unverified sORFs.

Supplementary Figure 2 | The results of GO enrichment analysis of three groups of sPeptide-associated parental genes and correlation of abundance between the 501 sPeptides and their parental genes. **(A–C)** Results of GO enrichment analysis of three clusters of parental genes, respectively. **(D)** The 501 sPeptides positively correlated with their parental genes in the translome.

Supplementary Figure 3 | The overlapped sPeptides validated by MS data of this study and Wang’s study.

Supplementary Table 1 | Tissues used in this study.

Supplementary File 1 | Genome annotation of sPeptides in GTF format.

Supplementary File 2 | Information of sORFs and sPeptides.

REFERENCES

- Archer, S. K., Shirokikh, N. E., Beilharz, T. H., and Preiss, T. (2016). Dynamics of ribosome scanning and recycling revealed by translation complex profiling. *Nature* 535, 570–574. doi: 10.1038/nature18647
- Arora, R., Singh, P., Kumari, A., Pathak, P. K., and Gupta, K. J. (2020). Using foldscope to monitor superoxide production and cell death during pathogen infection in arabidopsis under different nitrogen regimes. *Methods Mol. Biol.* 2057, 93–102. doi: 10.1007/978-1-4939-9790-9_9
- Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., et al. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* 148, 436–454. doi: 10.1104/pp.108.121038
- Basrai, M. A., Hieter, P., and Boeke, J. D. (1997). Small open reading frames: beautiful needles in the haystack. *Genome Res.* 7, 768–771. doi: 10.1101/gr.7.8.768
- Bazin, J., Baerenfaller, K., Gosai, S. J., Gregory, B. D., Crespi, M., and Bailey-Serres, J. (2017). Global analysis of ribosome-associated noncoding RNAs unveils new modes of translational regulation. *Proc. Natl. Acad. Sci. U.S.A.* 114, E10018–E10027. doi: 10.1073/pnas.1708433114
- Brar, G. A., and Weissman, J. S. (2015). Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664. doi: 10.1038/nrm4069
- Calviello, L., Mukherjee, N., Wyler, E., Zauber, H., Hirsekorn, A., Selbach, M., et al. (2016). Detecting actively translated open reading frames in ribosome profiling data. *Nat. Methods* 13, 165–170. doi: 10.1038/nmeth.3688
- Chen, Q. J., Deng, B. H., Gao, J., Zhao, Z. Y., Chen, Z. L., Song, S. R., et al. (2020). A miRNA-encoded small peptide, vvi-miPEP171d1, regulates adventitious root formation. *Plant Physiol.* 183, 656–670. doi: 10.1104/pp.20.00197
- Claverie, J. M. (1997). Computational methods for the identification of genes in vertebrate genomic sequences. *Hum. Mol. Genet.* 6, 1735–1744. doi: 10.1093/hmg/6.10.1735
- Couso, J. P., and Patraquim, P. (2017). Classification and function of small open reading frames. *Nat. Rev. Mol. Cell Biol.* 18, 575–589. doi: 10.1038/nrm.2017.58
- Cui, P., Zhang, S., Ding, F., Ali, S., and Xiong, L. (2014). Dynamic regulation of genome-wide pre-mRNA splicing and stress tolerance by the Sm-like protein LSM5 in Arabidopsis. *Genome Biol.* 15:R1. doi: 10.1186/gb-2014-15-1-r1
- Cui, Y., Li, M., Yin, X., Song, S., Xu, G., Wang, M., et al. (2018). OsDSSR1, a novel small peptide, enhances drought tolerance in transgenic rice. *Plant Sci.* 270, 85–96. doi: 10.1016/j.plantsci.2018.02.015
- De Coninck, B., and De Smet, I. (2016). Plant peptides - taking them to the next level. *J. Exp. Bot.* 67, 4791–4795. doi: 10.1093/jxb/erw309
- de Francisco Amorim, M., Willing, E. M., Szabo, E. X., Francisco-Mangilet, A. G., Droste-Borel, I., Maček, B., et al. (2018). The U1 snRNP subunit LUC7 modulates plant development and stress responses via regulation of alternative splicing. *Plant Cell* 30, 2838–2854. doi: 10.1105/tpc.18.00244
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Gao, Z., Li, J., Luo, M., Li, H., Chen, Q., Wang, L., et al. (2019). Characterization and cloning of grape circular RNAs identified the cold resistance-related Vv-circATS1. *Plant Physiol.* 180, 966–985. doi: 10.1104/pp.18.01331
- Golisz, A., Sikorski, P. J., Kruszka, K., and Kufel, J. (2013). Arabidopsis thaliana LSM proteins function in mRNA splicing and degradation. *Nucleic Acids Res.* 41, 6232–6249. doi: 10.1093/nar/gkt296
- Guillén, G., Diaz-Camino, C., Loyola-Torres, C. A., Aparicio-Fabre, R., Hernández-López, A., Díaz-Sánchez, M., et al. (2013). Detailed analysis of putative genes encoding small proteins in legume genomes. *Front. Plant Sci.* 4:208. doi: 10.3389/fpls.2013.00208
- Guo, P., Yoshimura, A., Ishikawa, N., Yamaguchi, T., Guo, Y., and Tsukaya, H. (2015). Comparative analysis of the RTFL peptide family on the control of plant organogenesis. *J. Plant Res.* 128, 497–510. doi: 10.1007/s10265-015-0703-1
- Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., and Lander, E. S. (2013). Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240–251. doi: 10.1016/j.cell.2013.06.009
- Guydosh, N. R., and Green, R. (2014). Dom34 rescues ribosomes in 3' untranslated regions. *Cell* 156, 950–962. doi: 10.1016/j.cell.2014.02.006
- Hemu, X., Serra, A., Darwis, D. A., Cornvik, T., Sze, S. K., and Tam, J. P. (2018). Peptidomic identification of cysteine-rich peptides from plants. *Methods Mol. Biol.* 1719, 379–393. doi: 10.1007/978-1-4939-7537-2_26
- Hinnebusch, A. G., Ivanov, I. P., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416. doi: 10.1126/science.aad9868
- Hsu, P. Y., and Benfey, P. N. (2018). Small but mighty: functional peptides encoded by small ORFs in plants. *Proteomics* 18:e1700038. doi: 10.1002/pmic.20170038
- Hsu, P. Y., Calviello, L., Wu, H. L., Li, F. W., Rothfels, C. J., Ohler, U., et al. (2016). Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7126–E7135. doi: 10.1073/pnas.1614788113
- Huang, K., Baldreich, P., Meyers, B. C., and Caplan, J. L. (2019). sRNA-FISH: versatile fluorescent *in situ* detection of small RNAs in plants. *Plant J.* 98, 359–369. doi: 10.1111/tj.14210
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., and Weissman, J. S. (2009). Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. doi: 10.1126/science.1168978
- Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., et al. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics* 62, e51. doi: 10.1002/cpbi.51
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lease, K. A., and Walker, J. C. (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiol.* 142, 831–838. doi: 10.1104/pp.106.086041
- Li, L., Eichten, S. R., Shimizu, R., Petsch, K., Yeh, C. T., Wu, W., et al. (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 15:R40. doi: 10.1186/gb-2014-15-2-r40
- Li, Y., Guo, Q., Liu, P., Huang, J., Zhang, S., Yang, G., et al. (2021). Dual roles of the serine/arginine-rich splicing factor SR45a in promoting and interacting with nuclear cap-binding complex to modulate the salt-stress response in Arabidopsis. *New Phytol.* 230, 641–655. doi: 10.1111/nph.17175
- Ma, C., Lu, Y., Bai, S., Zhang, W., Duan, X., Meng, D., et al. (2014). Cloning and characterization of miRNAs and their targets, including a novel miRNA-targeted NBS-LRR protein class gene in apple (Golden Delicious). *Mol. Plant* 7, 218–230. doi: 10.1093/mp/sst101
- Marshall, E., Costa, L. M., and Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J. Exp. Bot.* 62, 1677–1686. doi: 10.1093/jxb/err002
- Oh, Y., Robertson, S. L., Parker, J., Muddiman, D. C., and Dean, R. A. (2017). Comparative proteomic analysis between nitrogen supplemented and starved conditions in Magnaporthe oryzae. *Proteome Sci.* 15:20. doi: 10.1186/s12953-017-0128-y
- Ohyama, K., Ogawa, M., and Matsubayashi, Y. (2008). Identification of a biologically active, small, secreted peptide in Arabidopsis by *in silico* gene screening, followed by LC-MS-based structure analysis. *Plant J.* 55, 152–160. doi: 10.1111/j.1365-313X.2008.03464.x
- Okamoto, M., Matsui, A., Tanaka, M., Morosawa, T., Ishida, J., Iida, K., et al. (2016). Sm-like protein-mediated RNA metabolism is required for heat stress tolerance in arabidopsis. *Front. Plant Sci.* 7:1079. doi: 10.3389/fpls.2016.01079
- Olexiouk, V., Van Crielinge, W., and Menschaert, G. (2018). An update on sORFs.org: a repository of small ORFs identified by ribosome profiling. *Nucleic Acids Res.* 46, D497–D502. doi: 10.1093/nar/gkx1130
- Perea-Resca, C., Hernández-Verdeja, T., López-Cobollo, R., del Mar Castellano, M., and Salinas, J. (2012). LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal Arabidopsis development. *Plant Cell* 24, 4930–4947. doi: 10.1105/tpc.112.103697
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Ren, Y., Song, Y., Zhang, L., Guo, D., He, J., Wang, L., et al. (2021). Coding of non-coding RNA: insights into the regulatory functions of Pri-MicroRNA-encoded peptides in plants. *Front. Plant Sci.* 12:641351. doi: 10.3389/fpls.2021.641351

- Ruiz-Orera, J., and Albà, M. M. (2019). Translation of small open reading frames: roles in regulation and evolutionary innovation. *Trends Genet.* 35, 186–198. doi: 10.1016/j.tig.2018.12.003
- Ruiz-Orera, J., Verdaguier-Grau, P., Villanueva-Cañas, J. L., Messeguer, X., and Albà, M. M. (2018). Translation of neutrally evolving peptides provides a basis for *de novo* gene evolution. *Nat. Ecol. Evol.* 2, 890–896. doi: 10.1038/s41559-018-0506-6
- Samad, A. F. A., Sajad, M., Nazaruddin, N., Fauzi, I. A., Murad, A. M. A., Zainal, Z., et al. (2017). MicroRNA and transcription factor: key players in plant regulatory network. *Front. Plant Sci.* 8:565. doi: 10.3389/fpls.2017.00565
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534
- Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., et al. (2013). Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* 9, 59–64. doi: 10.1038/nchembio.1120
- Szakonyi, D., and Duque, P. (2018). Alternative splicing as a regulator of early plant development. *Front. Plant Sci.* 9:1174. doi: 10.3389/fpls.2018.01174
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J. F., Adami, E., Faber, A. B., et al. (2019). The translational landscape of the human heart. *Cell* 178, 242–260.e229. doi: 10.1016/j.cell.2019.05.010
- Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647. doi: 10.1093/bioinformatics/btx469
- Walley, J. W., and Briggs, S. P. (2015). Dual use of peptide mass spectra: protein atlas and genome annotation. *Curr. Plant Biol.* 2, 21–24. doi: 10.1016/j.cpb.2015.02.001
- Wang, P., Yao, S., Kosami, K. I., Guo, T., Li, J., Zhang, Y., et al. (2020). Identification of endogenous small peptides involved in rice immunity through transcriptomics- and proteomics-based screening. *Plant Biotechnol. J.* 18, 415–428. doi: 10.1111/pbi.13208
- Wang, S., Tian, L., Liu, H., Li, X., Zhang, J., Chen, X., et al. (2020). Large-scale discovery of non-conventional peptides in maize and arabidopsis through an integrated peptidogenomic pipeline. *Mol. Plant* 13, 1078–1093. doi: 10.1016/j.molp.2020.05.012
- Wen, J., Lease, K. A., and Walker, J. C. (2004). DVL, a novel class of small polypeptides: overexpression alters Arabidopsis development. *Plant J.* 37, 668–677. doi: 10.1111/j.1365-313X.2003.01994.x
- Wu, H. L., Song, G., Walley, J. W., and Hsu, P. Y. (2019). The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. *Plant Physiol.* 181, 367–380. doi: 10.1104/pp.19.00541
- Wu, X., Ding, D., Shi, C., Xue, Y., Zhang, Z., Tang, G., et al. (2016). microRNA-dependent gene regulatory networks in maize leaf senescence. *BMC Plant Biol.* 16:73. doi: 10.1186/s12870-016-0755-y
- Xia, J., Yamaji, N., and Ma, J. F. (2013). A plasma membrane-localized small peptide is involved in rice aluminum tolerance. *Plant J.* 76, 345–355. doi: 10.1111/tpj.12296
- Xiao, Z., Huang, R., Xing, X., Chen, Y., Deng, H., and Yang, X. (2018). *De novo* annotation and characterization of the translome with ribosome profiling data. *Nucleic Acids Res.* 46:e61. doi: 10.1093/nar/gky179
- Xiong, L., Gong, Z., Rock, C. D., Subramanian, S., Guo, Y., Xu, W., et al. (2001). Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in Arabidopsis. *Dev. Cell* 1, 771–781. doi: 10.1016/S1534-5807(01)00087-9
- Zhai, J., Zhang, H., Arikiti, S., Huang, K., Nan, G. L., Walbot, V., et al. (2015). Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc. Natl. Acad. Sci. U.S.A.* 112, 3146–3151. doi: 10.1073/pnas.1418918112
- Zhang, W., Han, Z., Guo, Q., Liu, Y., Zheng, Y., Wu, F., et al. (2014). Identification of maize long non-coding RNAs responsive to drought stress. *PLoS ONE* 9:e98958. doi: 10.1371/journal.pone.0098958
- Zhou, P., Silverstein, K. A., Gao, L., Walton, J. D., Nallu, S., Guhlin, J., et al. (2013). Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics* 14:335. doi: 10.1186/1471-2105-14-335
- Zhu, W., Xu, J., Chen, S., Chen, J., Liang, Y., Zhang, C., et al. (2021). Large-scale translome profiling annotates functional genome and reveals the key role of genic 3and#x2b9; untranslated regions in translomic variation in plants. *Plant Commun.* 2, 100181. doi: 10.1016/j.xplc.2021.100181
- Zipfel, C., and Oldroyd, G. E. (2017). Plant signalling in symbiosis and immunity. *Nature* 543, 328–336. doi: 10.1038/nature22009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Liang, Zhu, Chen, Qian and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.