Identification and Functional Investigation of Genome-Encoded, Small, Secreted Peptides in Plants

Clarissa Boschiero,[1,4] Peter K. Lundquist,[2,3,4] Sonali Roy,[1,4] Xinbin Dai,[1,4] Patrick X. Zhao,[1] and Wolf-Rüdiger Scheible[1,5]

[1]Noble Research Institute, LLC, Ardmore, Oklahoma
[2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan
[3]Plant Resilience Institute, Michigan State University, East Lansing, Michigan
[4]These authors contributed equally to this work.
[5]Corresponding author: *wrscheible@noble.org*

Hundreds to thousands of small secreted peptides (SSPs) are encoded in plant genomes but have been overlooked, and most remain unannotated and unstudied. Despite their low profile, they have been found to confer dramatic effects on growth and development of plants. With the growing appreciation of their significance, the development of appropriate methods to identify and functionally assess the myriad SSPs encoded in plant genomes has become critical. Here, we provide protocols for the computational and physiological analysis of SSPs in plant genomes. We first describe our methodology successfully used for genome-wide identification and annotation of SSP-coding genes in the model legume *Medicago truncatula*, which can be readily adapted for other plant species. We then provide protocols for the functional analysis of SSPs using various synthetic peptide screens. Considerations for the design and handling of peptides are included.

Keywords: annotation • bioinformatics • genome • peptide • root growth

How to cite this article:
Boschiero, C., Lundquist, P. K., Roy, S., Dai, X., Zhao, P. X., & Scheible, W.-R. (2019). Identification and functional investigation of genome-encoded, small, secreted peptides in plants. *Current Protocols in Plant Biology*, *4,* e20098. doi: 10.1002/cppb.20098

INTRODUCTION

Small secreted peptides (SSPs) encoded in plant genomes, also termed peptide hormones, possess crucial roles in plant growth and development (Breiden & Simon 2016; Murphy, Smith, & De Smet, 2012; Patel et al., 2018). Plant SSPs have also been shown to be important regulators of stress tolerance (Nakaminami et al., 2018; Takahashi et al., 2018) and nutrient acquisition (Grillet, Lan, Li, Mokkapati, & Schmidt, 2018; Tabata et al., 2014). In light of their diverse and powerful effects on plant development and stress physiology, SSPs are of great interest for fundamental plant biology and may hold potential for yield improvements in agriculture.

SSPs share certain characteristics that provide some guidance for their identification. First, they will usually contain an N-terminal encoded signal peptide directing the protein to the secretory pathway. Second, the bioactive small peptide will be often encoded near the C-terminus of its precursor and is proteolytically cleaved during maturation.

CURRENT PROTOCOLS
A Wiley Brand
Current Protocols in Plant Biology e20098, Volume 4
Published in Wiley Online Library (wileyonlinelibrary.com).
doi: 10.1002/cppb.20098

This means that the C-terminal part of the open reading frame (ORF) will contain the conserved residues, diagnostic of an SSP family.

This article is broken into four basic protocols describing how to identify SSPs in plants using genomic data and bioinformatics tools and subsequently perform functional assays based on synthetic peptide screens, as previously employed (de Bang et al., 2017). The first two protocols are bioinformatics based and rely on in silico analyses to identify SSP-encoded genes in a plant genome. The prediction of SSP-coding genes from genomic sequences is described in Basic Protocol 1. Then, the second stage is to identify and annotate SSPs, outlined in Basic Protocol 2. It is important to note that genes identified on the basis of bioinformatics approaches alone require subsequent functional identification by biochemical or genetic approaches. In this respect, a straightforward approach is the testing of gene-derived synthetic peptides for visible and/or molecular phenotypes. In Basic Protocol 3 the screening of *Medicago truncatula* for peptide-induced root traits is described, including the setup of growth systems and image analysis steps, and in Basic Protocol 4 the screening of *Arabidopsis* roots for peptide-induced calcium bursts is outlined. Collectively, this article provides successful methods for the identification and functional study of SSPs from *M. truncatula*, which can be adapted to other plant species for the advancement of SSP biology.

BASIC PROTOCOL 1

SSP-CODING GENE DISCOVERY IN PLANTS

This protocol describes a bioinformatics pipeline to identify plant SSP-coding genes. It aims to improve the identification of SSPs by the discovery of small genes in plant genomes because they are frequently overlooked or unannotated. The principle is to reannotate a plant genome of your choice and obtain new gene models, particularly of small gene products (often <200 residues) which are more likely to encode SSPs (Fig. 1A). The identification of new or small genes will help the identification of SSPs (Basic Protocol 2). Here, we describe genome reannotation, focusing on identification of SSP-coding genes via two software programs: MAKER and SPADA (Small Peptide Alignment Discovery Application). MAKER is a well-known and easy-to-use general genome annotation tool designed to produce ab initio gene predictions with evidence-based quality scores (Cantarel et al., 2008) and can utilize protein expressed sequence tags (ESTs) and RNA sequencing (RNA-seq) evidence to guide identification. SPADA software is designed specifically for the identification of SSP-coding genes. It is a homology-based gene prediction software that uses multiple similarity search tools including BLAST and HMMER (Zhou et al., 2013).

To facilitate the explanation and demonstration of Basic Protocols 1 and 2 and the dissemination of our customized bioinformatics analysis pipelines, we have generated an online open-source document on GitHub (*https://github.com/ZhaoBioinformaticsLab/PlantSSPProtocols/*) to present the information in an up-to-date and readily accessible manner.

We further provide a Docker image on the Docker hub (*https://hub.docker.com/r/noblebioinfo/sspgene*), in which we pack all preinstalled and properly configured or customized software, including MAKER, SPADA, RNA-seq analysis tools, and other Linux scripts for data analysis. End users can download and run the Docker image on a local computer that supports a program called "docker." When the Docker image is loaded by the "docker" program, it creates an instance termed as "container," providing the user a virtual operating system along with preinstalled, ready-to-use software and all its dependencies.

Users are recommended to read our online document for details about Docker installation, Linux commands, and our demo analysis and further follow the presented practices using

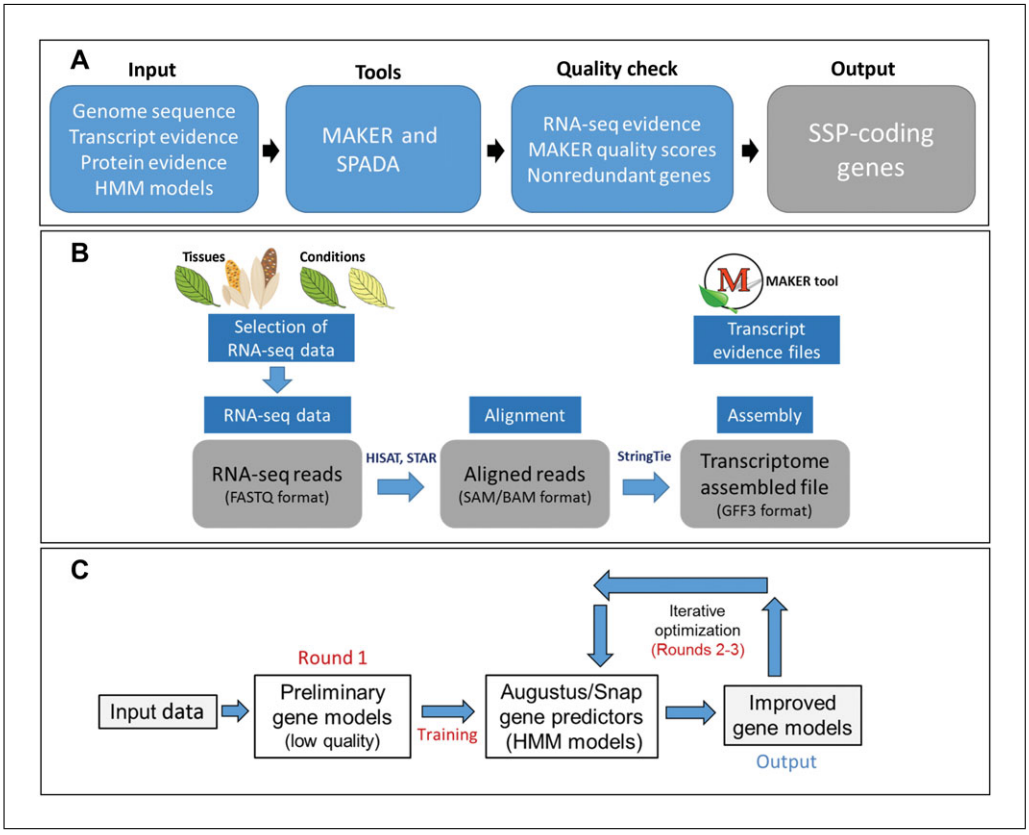

Figure 1 Workflow for Basic Protocol 1. (**A**) Pipeline overview of SSP-coding gene discovery with MAKER and SPADA tools. (**B**) Transcriptome workflow for file generation of transcript evidence to be used by MAKER including alignment, assemble, and GFF3 file process. (**C**) Genome annotation workflow using MAKER software.

the provided tools in the Docker image. A basic knowledge and ability to use Linux is required.

Materials

Hardware

High-performance computer (i7/Xeon processor and >16 GB RAM) or virtual machine with Linux system (CentOS 7, Ubuntu 16.04 or higher)

Software

HISAT2 (Kim, Langmead, & Salzberg, 2015)
Stringtie (Pertea et al., 2015)
Docker image, available at *https://hub.docker.com/r/noblebioinfo/sspgene*

Files

For MAKER (see Table 1):
Genomic sequences in FASTA format
Reference annotations for the same genomic sequences in GFF3 format if available
RNA-seq data in compressed FASTQ format
Protein sequences of known SSP-coding genes in FASTA format
Other related protein sequences in FASTA format
EST/transcript sequences from the same species

For SPADA (see Table 1):
Genomic sequence
Hidden Markov Model (HMM) files for known SSP families

Preparation of SSP gene expression evidence for MAKER pipeline

General genome annotation procedures can be optimized for SSP-coding gene identification by including diverse RNA-seq samples and known SSP protein sequences. The following steps describe how to align and assemble RNA-seq reads on genomic sequences and generate assembled results in GFF3 format (Fig. 1B). The GFF3 file containing assembled transcriptome data will be used subsequently.

1. Collect RNA-seq data from diverse experiments including different tissues, developmental stages, and treatments. Consider including tissues or treatments related to the biological process of interest. For example, use RNA-seq data generated from tissue exposed to nutrient-deficient conditions when seeking SSPs involved in nutrient stress regulation.

2. Download genome sequences from Phytozome (*https://phytozome.jgi.doe.gov/pz/portal.html*) or other public repository, and compile sequences using HISAT2 (Kim et al., 2015). For example, use the following command that can be tested in the demo data folder of the Docker container:

```
hisat2-build data/genome.fa genome_hisat2
```

3. Extract splicing sites (if reference annotation is available) using HISAT2:

```
gffread data/ref.gff3 -T -o ref.gtf
hisat2_extract_splice_sites.py ref.gtf >
  splicesites.txt
```

In the above example, `data/ref.gff3` is the GFF3 file that contains reference annotations, and `splicesites.txt` is the text file with the splice site information extracted from the reference annotation for the next step.

4. Map RNA-seq reads onto the reference genomic sequence using HISAT2. The output of the alignment is a BAM file, which is subsequently sorted using sambamba (*http://lomereiter.github.io/sambamba/*). If running Linux, use the following commands (also see GitHub document sections 1.2.4 and 1.2.5):

```
hisat2 -x genome_hisat2 --known-splicesite-infile
  splicesites.txt --dta --dta-cufflinks -1
  root_R1.fq.gz -2 root_R2.fq.gz | samtools view -bS
  - > all_runs.bam
sambamba sort -m 40G --tmpdir tmp/ -o
  all_runs.sorted.bam -p -t 20 all_runs.bam
```

In the above example, `root_R1.fq.gz` and `root_R2.fq.gz` are the FASTQ files that contain RNA-seq reads; `all_runs.bam` is the mapping result file; and `all_runs.sorted.bam` is the sorted BAM file.

5. Assemble mapped RNA-seq reads (in BAM file) to transcript using Stringtie (Pertea et al., 2015). A GFF3 file will be generated as the input of MAKER pipeline. Use the commands:

```
stringtie all_runs.sorted.bam -o
  transcriptome_models.gtf -p 20
cufflinks2gff3 transcriptome_models.gtf >
  transcriptome_models.gff3
```

Table 1 Input Files Used for MAKER and SPADA

Program	File	Format	Description
MAKER	Genome assembly	FASTA or GFF3[a]	Sequence can be found at NCBI or JGI[b] databases
	Protein evidence	FASTA	Sequences from a species of interest or from a closely related species; we recommend using a curated database such as UniProt/SwissProt[c] or SSP plant curated proteins[d]
	Transcript evidence	FASTA or GFF3[a]	Includes (1) assembled mRNA-seq transcripts or ESTs or (2) aligned ESTs or mRNA-seq
	Repeat library	FASTA	Optional
SPADA	Genome sequence	FASTA	Sequence can be found at NCBI or JGI[b] databases
	Gene annotation	GFF3[a]	Optional; can be found at NCBI or JGI[b] databases
	HMM library	HMM	Library with HMM files from SPADA or a custom library

[a]GFF3 (General Feature Format) is a standard, tab-delimited file format for storing genomic features (*http://gmod.org/wiki/GFF3*).
[b]*https://phytozome.jgi.doe.gov/pz/portal.html*.
[c]*https://www.uniprot.org/*.
[d]*http://mtsspdb.noble.org/database/download.*
EST, expressed sequence tag; HMM, hidden Markov model; JGI, Joint Genome Institute; mRNA-seq; mRNA sequence; NCBI, National Center for Biotechnology Information; SSP, small secreted peptides.

Optimized procedure for mining SSP genes from genomic sequences using the MAKER pipeline

MAKER is a portable and easily configurable genome annotation pipeline. This pipeline is preinstalled and available in Docker image. It can be freely downloaded under the GNU General Public License. Details about installing and running the Docker service and image are available on our Github online document (section 1.1.6). The procedure for MAKER annotation has been well documented (Current Protocols article: Campbell, Holt, Moore, & Yandell, 2014). The following steps are executed for genome annotation with MAKER (Fig. 1C). The MAKER commands and data processing are further detailed in the GitHub online document section 1.3. This protocol requires different input files (see Table 1) for MAKER and SPADA analysis.

6. In the terminal of the Docker container, create three control files that are the files with the necessary configuration and parameters for the analysis (maker_opts.ctl, maker_bopts.ctl, and maker_exe.ctl) using the command `maker -CTL`.
7. Revise maker_opts.ctl to contain the file path of the input files, including genomic sequence, GFF3 file for reference annotation, related protein/EST sequences, GFF3 generated by step 5, and known SSP protein sequences.
8. Run the following MAKER command:

```
mpiexec -n 20 maker -fix_nucleotides
```

Here, 20 is the number of CPU cores in the user's computer.

9. Use the generated files to generate and optimize gene models for SNAP (see Korf, 2004).
10. Revise maker_opts.ctl to include SNAP models and also enable the gene prediction option.
11. Run MAKER command again.
12. Repeat steps 9, 10, and 11 to generate the final GFF3 file.

Genome annotation with SPADA

SPADA (Zhou et al., 2013) typically uses conserved gene family domains in HMM format to identify related (SSP) gene models from genomic sequences. HMM is useful for identification of conserved sequence patterns. If unique patterns are known, an HMM search can help identify novel sequences matching that motif under study. SPADA has also been installed in Docker image and is available for download under the Apache License. The pipeline includes prebuilt HMMs for plant cysteine-rich peptides. To cover more SSP families, the installation in Docker image also includes HMMs from PlantSSPDB (Ghorbani et al., 2015) and our published efforts (de Bang et al., 2017).

The HMM files for known SSP families required by SPADA for SSP gene annotation have been packed into the Docker image as part of the SPADA installation (GitHub online document sections 1.4 and 2.4.1).

Users can extract the protein sequences of known SSP-coding genes from the demo dataset (see GitHub online document sections 1.1.5) to regenerate or further modify the HMM files.

If you have protein sequences from newly identified SSP families, continue to step 13. Otherwise, skip to step 15 to use prebuilt HMMs.

13. Use MUSCLE multiple sequence alignment (*https://www.ebi.ac.uk/Tools/msa/muscle/*) to generate an alignment (aln) file from the member proteins of a new family.
14. Run build_profile.pl to convert aln files into SPADA HMM profile.
15. Run SPADA in Docker image as follows:

```
perl /opt/spada_soft/spada/spada.pl --cfg
  /opt/spada_soft/spada/cfg.txt -d sspanno -p
/opt/spada_soft/spada/CRP_PlantSSPv1_Noble -f
  genome.fa -t 20 -o arabidopsis
```

In the sample command, `/opt/spada_soft/spada/CRP_PlantSSPv1_Noble` is the prebuilt directory containing HMM profiles from the prebuilt CRP family, families downloaded from PlantSSPDB, and our published paper; `sspanno` is the folder for the annotation results.

16. Save 61_final.gff under sspanno/31_model_evaluation directory as the SPADA annotation result using the command:

```
cat /31_model_evaluation/61_final.gff > spada.gff
```

Merge annotation results from MAKER and SPADA

Duplicate gene models between MAKER and SPADA annotations can be removed through the following steps/commands. Also refer to our GitHub online document section 1.5 for the detailed Linux commands.

17. Prepare coding sequence (CDS; coding region of a gene) for each annotation:

```
gffread spada.gff -g data/genome.fa -x spada_cds.fa
```

18. Run NCBI BLASTN (*https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch*) between the two generated CDS files, and only keep the query-hit pairs with >50% overlapping regions.

19. Check the coordinates of the query-hit pairs in the GFF file. If both genes share overlapping regions and have the same chromosomal orientation, mark one of them as duplicate.

20. Remove duplicate genes from corresponding GFF file.

21. Merge two GFF files.

22. Generate protein and transcript files from merged GFF files using the following gffread commands:

```
gffread all.gff -g data/genome.fa -y all_protein.fa
gffread all.gff -g data/genome.fa -w all_transcript.
  fa
```

ANNOTATION OF PLANT SSP-CODING GENES

BASIC PROTOCOL 2

We developed a bioinformatics pipeline for SSP gene discovery and SSP classification (Fig. 2). In this pipeline, we apply multiple criteria to identify plant SSP coding genes from candidates. The criteria include: (1) maximum peptide length <250; (2) presence of a signal peptide cleavage site predicted with SignalP (*http://www.cbs.dtu.dk/services/SignalP-4.1/*; Fig. 3; Petersen, Brunak, von Heijne, & Nielsen, 2011); (3) sequence homology matching with any known SSP-coding genes using the Smith-Waterman search

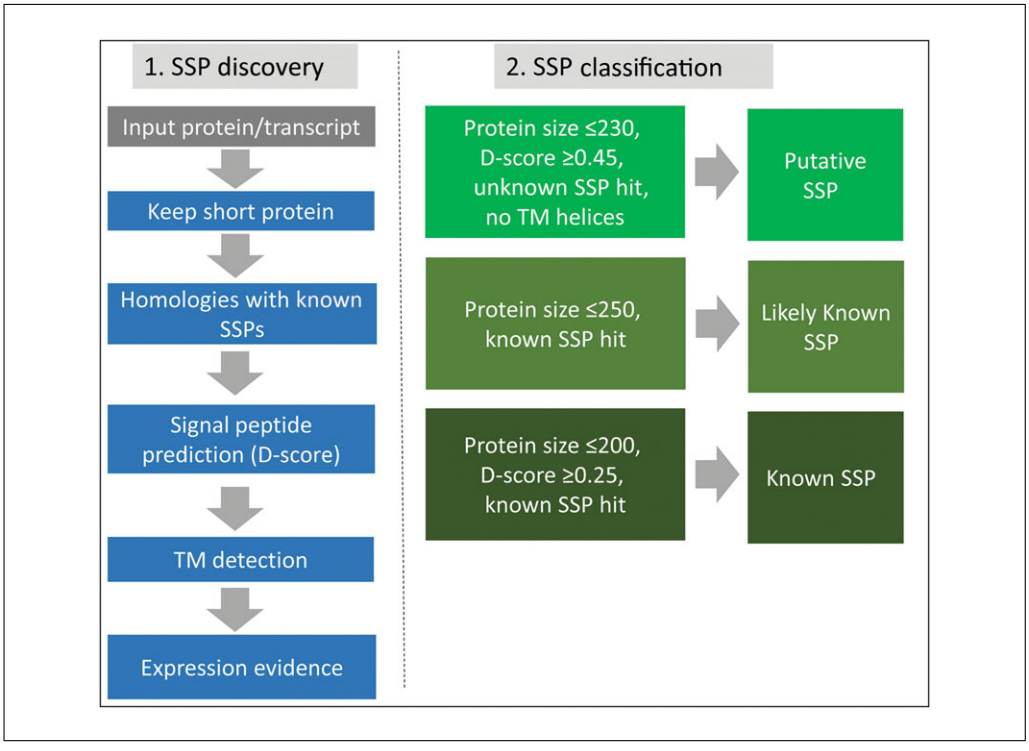

Figure 2 Overview of the SSP prediction pipeline and classification criteria for different SSP types (known, likely known, and putative SSPs).This pipeline was recently used to identify SSPs in the legume model *Medicago truncatula* (de Bang et al., 2017).

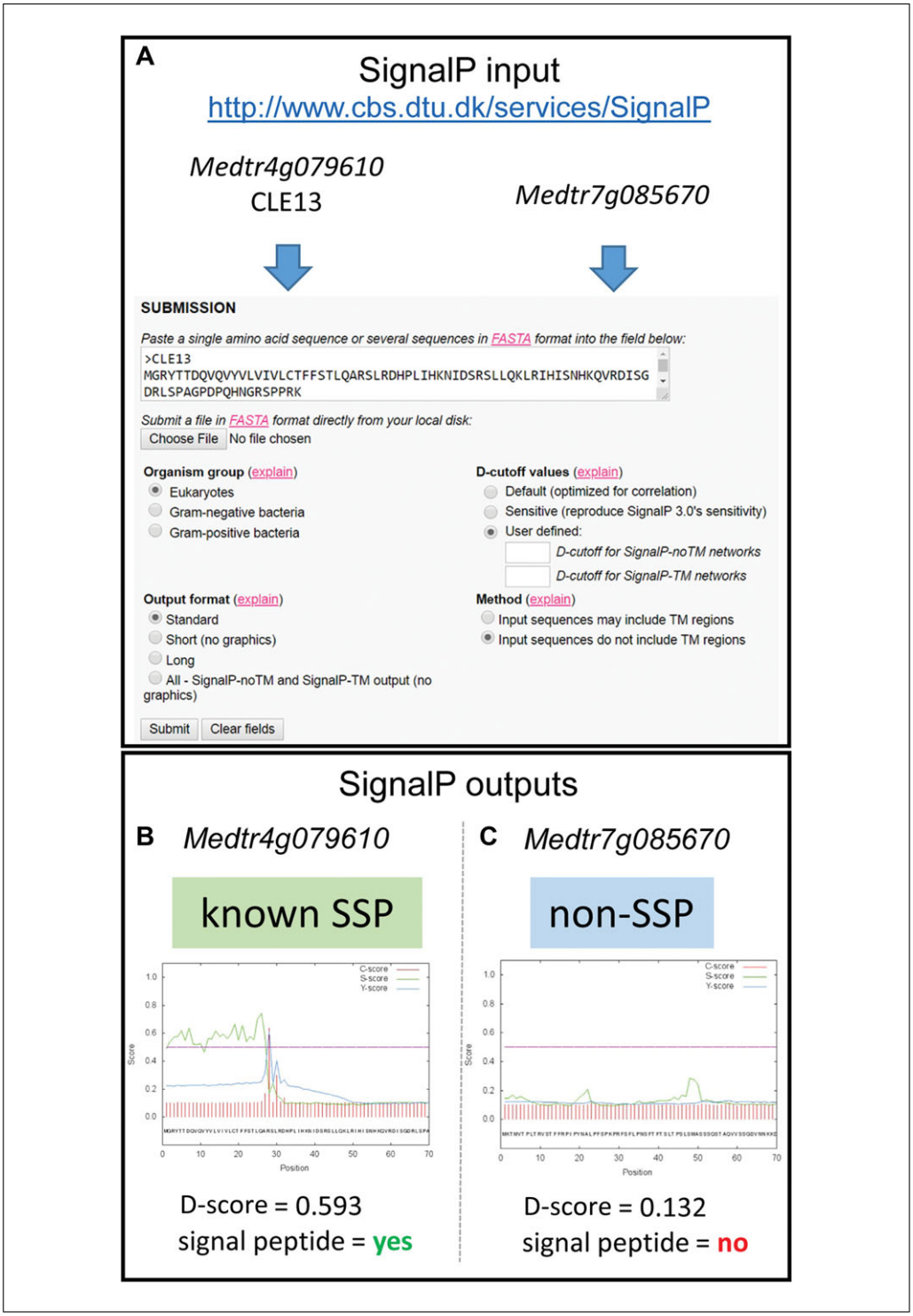

Figure 3 (**A**) SignalP 4.1 input information of two protein sequences in *M. truncatula* and its respective outputs of a (**B**) known SSP and (**C**) non-SSP. The graphical output shows different scores including C, S, and Y. D-score is used to discriminate signal peptides from nonsignal peptides.

algorithm (Fig. 4A; Current Protocols article: Ropelewski, Nicholas, & Deerfield, 2004); (4) matching with any known SSP family HMM profiles using HMMER (Fig. 4B; Finn, Clements, & Eddy, 2011); (5) absence of any transmembrane helices (TMH) predicted by TMHMM Server v.2.0 (*http://www.cbs.dtu.dk/services/TMHMM/*; Krogh, Larsson, von Heijne, & Sonnhammer, 2001); (6) absence of a C-terminal endoplasmic reticulum

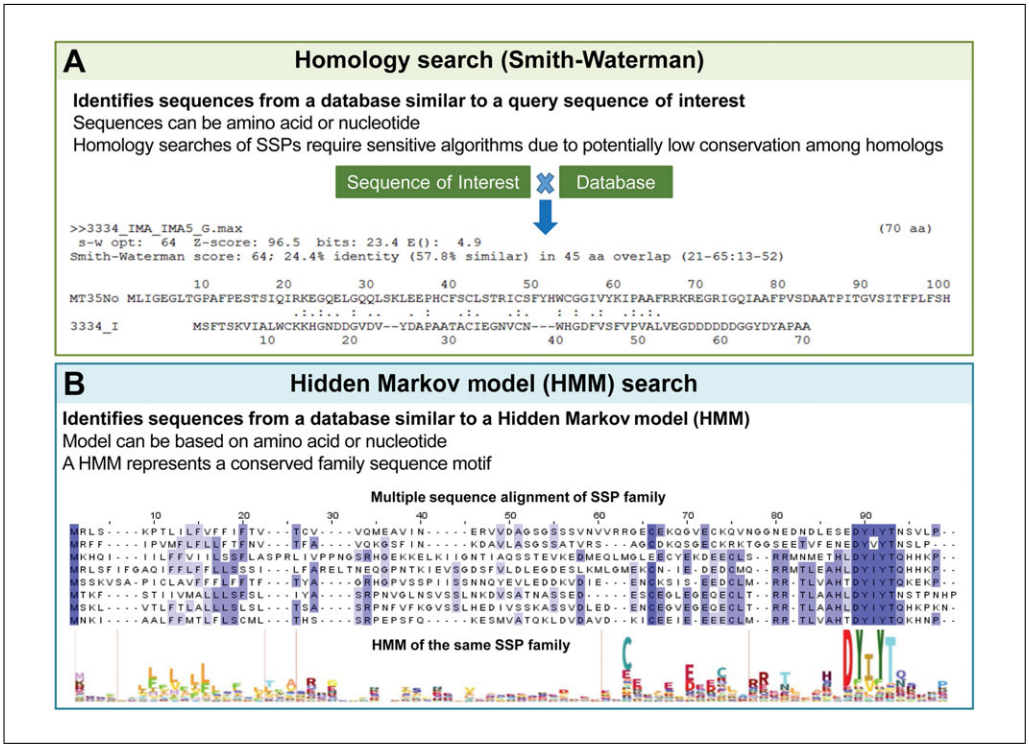

Figure 4 Types of homology searches used to predict SSPs (adapted from Roy et al., 2018). (**A**) Homology search (Smith-Waterman) with SSearch tool and (**B**) HMM profile search with HMMER tool.

(ER)-retention signal; and (7) presence of gene expression evidence in RNA-seq data. The putative SSPs are further classified into families using the Markov Cluster (MCL) algorithm (Enright, Van Dongen, & Ouzounis, 2002). Similar filtering criteria have also been applied for the discovery of SSPs in other plant species (Ghorbani et al., 2015).

This pipeline was recently used to identify SSPs in the model legume *Medicago truncatula* (de Bang et al., 2017). These SSPs can be found in our recently developed *Medicago truncatula* Small Signaling Peptide Database (MtSSPdb; *http://mtsspdb.noble.org*), which hosts SSP gene sequences, detailed SSP gene function annotation, family information and HMM profiles (see Fig. 5 for an example), and their expression profiles. It is worth mentioning that we have developed a plant SSP prediction tool that implements all the described filtering steps below (except RNA-seq expression analysis) to provide plant SSP gene identification and an annotation online service through the MtSSPdb.

Materials *(also see Basic Protocol 1)*

Files

Protein sequences (with gene ID as protein ID) of SSP gene candidates in FASTA format
Transcript sequences of SSP candidate genes
Diverse RNA-seq data in compressed FASTQ format (Table 2)

Filter protein sequence and keep proteins shorter than 250 amino acids

SSPs are frequently encoded within a longer precursor protein of <250 amino acids (Breiden & Simon, 2016). We provide a Linux script to only keep short sequences.

1. Use the command `keepshortseq seq.fa 250 > short-seq.fa` to select sequences shorter than 250 amino acids.

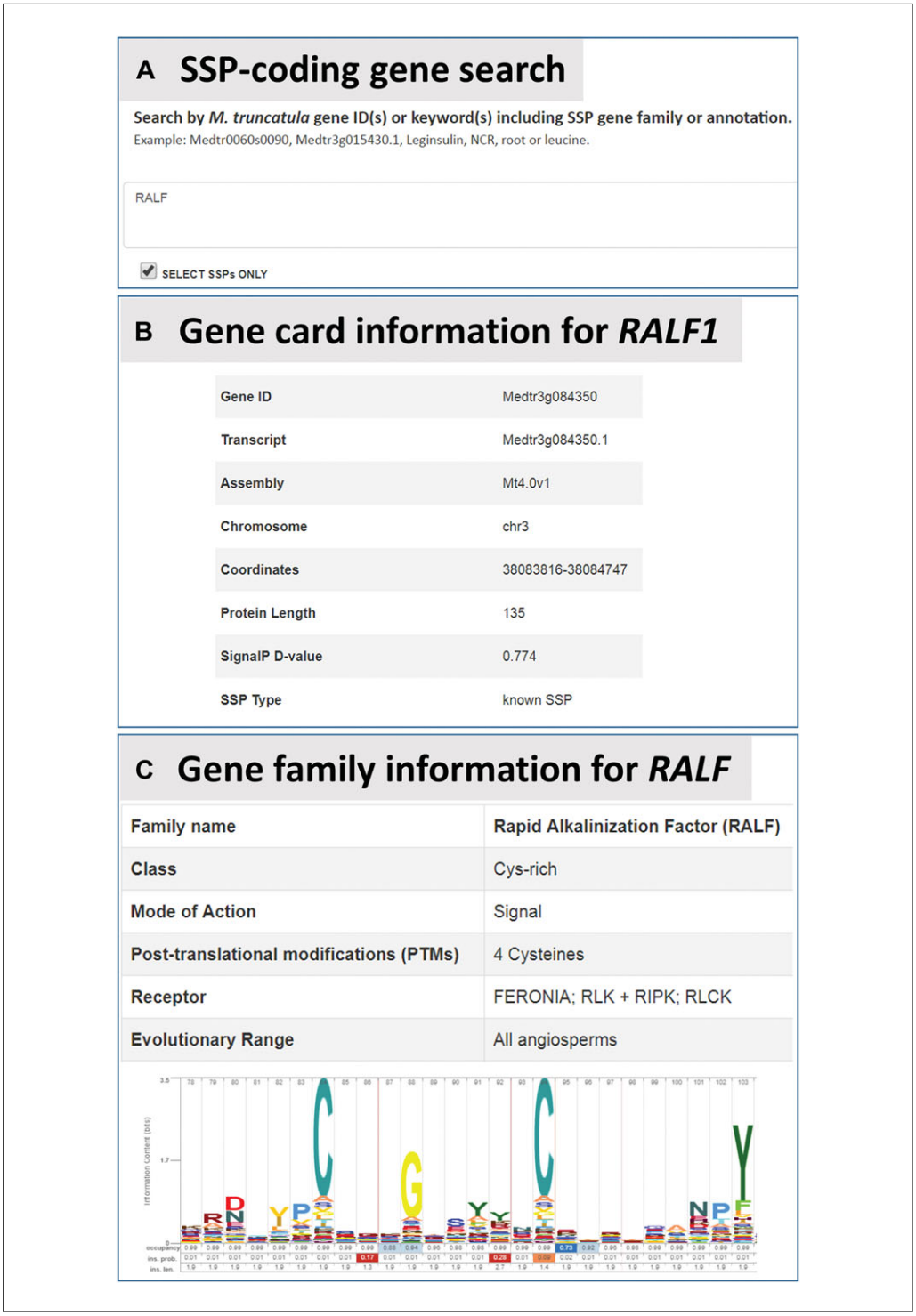

Figure 5 MtSSPdb usage example. (**A**) Search for rapid alkalinization factor (RALF) SSP genes. (**B**) Gene card information for *RALF1* (*Medtr3g084350*). (**C**) Overview of RALF family with gene family summary information and HMM profile logo showing the C-terminal region of the protein.

Run Smith-Waterman search against known SSP proteins

The Smith-Waterman algorithm performs sequence similarity searches that are more reliable for short sequences (Fig. 4A). Known SSP sequences data/ssp_family.fa is available in the demo dataset (see GitHub online document section 1.1.5).

2. Perform Smith-Waterman search between short protein sequences (generated in step 1) against known SSP proteins using the command:

Table 2 Input Files Used to Identify and Annotate Candidate Plant SSPs

Program	File	Format	Description
SignalP and TMHMM server	Protein sequence(s)	FASTA	Protein sequence(s) from a species of interest
HMMER	Protein sequence(s)	FASTA	Protein sequence(s) from a species of interest
	HMM-profile database	HMM	Library file containing all HMM profiles for each SSP family
SSearch	Protein sequence(s)	FASTA	Protein sequence(s) from a species of interest
	Protein sequence(s)	FASTA	Protein sequence(s) of known SSPs
MCL	Protein sequence(s)	FASTA	Last 50-70 amino acids of the protein sequences

HMM, hidden Markov model; MCL, Markov Cluster; SSP, small secreted peptides.

```
swsearch short_seq /work/ssp/data/ssp_family.fa 0.01
  > sw.txt
```

The script will automatically invoke the Ssearch program from the FASTA package (e-value <0.01) and parse the output. The results file sw.txt will include sequence name and SSP family name. See GitHub online document section 2.3 for more details.

Perform HMM search against known SSP families

To search protein sequences against HMM models of known SSP families (Fig. 4B), first build a library to host the models.

3. Generate HMM library for all known SSP families from SPADA installation:

```
cat /opt/spada_soft/spada/CRP_PlantSSPv1_Noble/15_
  hmm/*.hmm > all.hmm
```

4. Compile HMM library:

```
/opt/spada_soft/hmmer/bin/hmmpress all.hmm
```

5. After the library is generated, use the following command to search against the HMM library:

```
hmmscan --cpu 4 -E 0.01 --tblout hmm_output.txt all.
  hmm short-seq.fa > /dev/null
```

In this sample command, the expectation cutoff is 0.01, `all.hmm` is the HMM library, and the output file is hmm_output.txt.

Also refer to GitHub online document sections 2.4.1 and 2.4.2 for this step.

Run signal peptide prediction analysis using SignalP server

We chose the SignalP tool (*http://www.cbs.dtu.dk/services/SignalP/*) to predict N-terminal signal peptides from SSP candidate peptides (Fig. 3).

6. Run SignalP using:

```
signalp -t euk -f long -s notm short-seq.fa >
  signalp_long.txt
```

We recommend using `notm` and long output format. D-score thresholds are usually 0.45 or 0.50, depending on the type of network chosen (with or without transmembrane segments). Here we recommend using a D-score of ≥0.25 for known SSPs or ≥0.45 for putative SSPs (Fig. 2; de Bang et al., 2017).

7. Convert the SignalP output into a simple tab-delimited file with protein ID and D-score using:

```
cat signalp_long.txt | singalP_parser > sp.txt
```

Perform MCL analysis for clustering putative SSP genes

SSP candidates with SignalP D-score >0.45 can be further clustered into candidate SSP families using the MCL algorithm. MCL is a well-known software that can be used to identify clusters of protein sequences (Enright et al., 2002) and has recently been used to help the discovery of SSP gene families (de Bang et al., 2017; Ghorbani et al., 2015). Candidate peptides should be <230 amino acids, and the analysis procedure should be performed on the last 50 amino acids of the peptide.

8. Only keep proteins with a D-score >0.45 using the commands:

```
cdbfasta short-seq.fa
cat sp.txt | awk '{if($4>0.45) print $1}' | cdbyank
  short-seq.fa.cidx > all_putative_ssp.fa
```

9. Use the following script to keep only sequences shorter than 230 amino acids and take only the last 50 amino acids for the next step:

```
shortseqtail all_putative_ssp.fa 230 50 >
  peptide-tail.fa
```

10. Generate a protein-versus-protein relationship file with e-values <0.01. We recommend a Smith-Waterman search to generate a protein-versus-protein relationship measured with e-value, using the commands:

```
/opt/bin/ssearch35_t -T 20 -Q -H -m 9 -b 100 -d 100
  peptide-tail.fa peptide-tail.fa >
  sw_for_peptidetail
bioparser -t ssearch -m sw_for_peptidetail | awk
  '{print $3,$6,$14}' FS="\t" | sort | uniq | awk
  '{if($1!=$2&&$3 < 0.01) print $0;}' FS=" " >
  protein-protein-rel.txt
```

The commands above will generate a results file "protein-protein-rel.txt" of three columns with two protein/gene IDs and their relation in e-values. All protein-protein pairs with e-value >0.01 will be removed. Also see GitHub online document section 2.6.4 for details.

11. Cluster proteins using the MCL approach. Use MCL software to cluster proteins referring to protein-protein relationship (generated in step 10):

```
mcxload -abc protein-protein-rel.txt --stream-mirror
  --stream-neg-log10 -stream-tf 'ceil(200)' -o
  protein-protein.mci -write-tab protein-protein.tab
mcl last50seq.mci -I 1.4 -use-tab
  protein-protein.tab
```

The commands will generate a results file with the name "out.last50seq.mci.I14" in which all genes/proteins belonging to the same cluster will be put in one line.

12. Generate a cluster name versus gene/protein ID table file from the above MCL result:

```
cat out.protein-protein.mci.I14 | awk '{print
  "Cluster_" NR "\t" $0}' |awk '{OFS="\n" $1
  "\t";$1="";print $0;}'|grep -Pv '^\s*$' >
  mclcluster_protein.txt
```

The "mclcluster_protein.txt" file is the results file containing cluster-protein mapping.

Gene expression analysis of RNA-seq data

13. Follow the previously published protocol (Li & Dewey, 2011) to generate a gene expression table using RNA-seq data. The input data includes transcript sequences and mapping between gene ID and transcript ID. In the generated table, each gene will occupy one row and each RNA-seq sample will take one column. We recommend the transcript per million transcripts (TPM) value as the expression value in this table. For example:

```
rsem-calculate-expression --output-genome-bam --star
  -star-gzipped-read-file --paired-end
  root_R1.fastq.
  gz root_R2.fastq.gz /index_star/ Root
```

In the example, `root_R1.fastq.gz` and `root_R2.fastq.gz` are the FASTQ files that contain RNA-seq reads for root.

Perform transmembrane helix prediction

TMH prediction is a criterion used to classify putative SSPs (de Bang et al., 2017) because a gene harboring TMHs cannot be considered as an SSP.

14. Remove the N-terminal signal peptide from the input protein sequence(s). The signal peptide tends to display high hydrophobicity and often results in a false positive TMH prediction. The processed protein sequences (processed_putative.fa) will be used in next step. Run the following command to generate the putative sequences without the signal peptide regions from the putative SSP input protein sequences:

```
/opt/spada_soft/signalp-4.1/signalp -t euk -f short
  -m processed_putative.fa -u 0.5 -s notm
  all_putative_ssp.fa > signalp_putative.txt
```

15. Perform TMH prediction.

We recommend the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/) because it is easy to use and accurate due to its HMM search algorithm (Krogh et al., 2001). TMHMM is a user-friendly web-based analysis tool. Just submit your processed protein sequence(s) as the input file, and select the output format with graphics. The output result provides a list of predicted TMHs and their locations. Positive TMH prediction leads to rejection of a protein sequence as putative SSP.

Perform ER-retention signal search in the C-terminal region

SSPs act as intracellular, intercellular, or inter-organ signals and by definition should not be retained in the ER. Hence, sequences with a C-terminal ER-retention sequence (i.e., KDEL, HDEL, or KQEL; Pagny, Lerouge, Faye, & Gomord, 1999) should be removed. This additional filter was previously applied during the analysis of putative SSPs from *Medicago truncatula* (de Bang et al., 2017). However, only very few sequences (4 in 2690) displayed a C-terminal ER-retention signal.

16. Search ER-retention signals using the file generated in step 9 as input and this command:

```
grep -i 'KDEL\|HDEL\|KQEL' peptide-tail.fa
```

Generate a comprehensive gene annotation table

17. Generate a comprehensive annotation for SSP gene candidates.

A series of tab-delimited result files for the genes of interest will be generated from all above steps. These files include annotation information for these genes, including the genes' RNA-seq expression value, family name from Smith-Waterman search and HMM search, D-score from signalP, and cluster ID from MCL analysis. The results files should be merged into a comprehensive data table in which each gene will use one row and each annotation information take one column. The Power Query or Merge Tables Wizard in Microsoft Excel is a good tool to perform the merging operation.

18. Curate known and screen putative SSP genes using the comprehensive annotation table.

Figure 2 illustrates major criteria that can be applied for SSP gene curation. Briefly, the genes with Smith-Waterman or HMM hits will be considered as either known SSP genes if their D-score is larger than 0.25 and length is shorter than 200 amino acids, or likely known SSP genes if their length is shorter than 250 amino acids. Other genes with D-score ≥0.45, no TMH report, and shorter than 230 amino acids will be considered as putative SSP genes. The putative genes in the same MCL cluster can be aligned to generate a conserved domain. The gene expression value is also helpful to identify high-confidence genes from the above known or putative genes.

BASIC PROTOCOL 3

SCREENING SYNTHETIC PEPTIDES FOR BIOLOGICAL ACTIVITY ON *MEDICAGO TRUNCATULA* ROOTS IN A "FILTER PAPER SANDWICH" PLATE SYSTEM

The use of synthetic peptides represents a straightforward, nontransgenic approach to survey SSP activity and probe SSP gene function. Custom peptides of varying degrees of purity are available as lyophilized powders from numerous providers with a typical turnaround time of 3 to 4 weeks. Upon application to plants, these peptides are often recognized by receptors in planta and therefore retain biological activity (see Table 3 for examples and references). Although the use of custom peptides can quickly identify molecules with (strong) biological effects and potentially inform about SSP gene function, the approach is not a full or adequate substitute for genetic approaches to investigate gene function such as gene overexpression, insertion mutant analysis, or CRISPR-Cas9 gene editing. Hence, biological effects of synthetic peptides ultimately need to be verified by alternative means to elucidate the biological function of an SSP-coding gene.

Here, we provide some general guidelines for synthetic peptide selection and handling and describe tested protocols for screening of the biological effects of peptides on *Medicago truncatula* roots. In this regard the choice of medium for plant growth, together with the selection of the growth system, is critical for experimental success. Treatments with simple, linear peptides are most effective in contained, sterile growth systems such as Petri dishes and pouches. Based on treatment duration, the most appropriate shape and size of plates or pouches can be selected. Selecting a medium that is right for the assay must take into account the duration of the experiment since different media can alter plant growth rates. The pH of the medium can also affect peptide activity (Hou et al., 2014). Peptide applications can affect plants at the cellular and molecular level (e.g., gene expression, respiratory oxidative bursts, changes in cytosolic calcium) within minutes (Haruta, Monshausen, Gilroy, & Sussman, 2008; Hou et al., 2014; Ma, Zhao, Walker, & Berkowitz, 2013), whereas much longer treatments (2 days to 2 weeks)

Table 3 Examples of Synthetic Peptides Used to Unravel In Vivo Functions of SSP-Coding Genes

Peptide family	Peptide sequence (with modifications)	Plant species tested	Reference
CAP-derived peptide (CAPE)	PVGNWIGQRPY	*Solanum lycopersicum*	Chen et al. (2014)
CLAVATA3/endosperm surrounding region (CLE)	RTVPSGP(Hyp-Ara_3) DPLHH	*Arabidopsis thaliana*	Xu et al. (2015)
C-terminally encoded peptide (CEP)	DFR(Hyp)TNPGNS (Hyp)GVGH	*Medicago truncatula*	Imin et al. (2013)
GOLVEN/root growth factor (GLV/RGF)	DY(SO3H)PQPHRKPP (Hyp)IHNE	*Arabidopsis thaliana*	Whitford et al. (2012)
Inflorescence dehiscence in abscission (IDA)	FGYLPKGVPIPPSAPS KRHN	*Arabidopsis thaliana*	Stenvik et al. (2008)
PAMP-induced secreted peptides (PIP)	RLASG[Hyp]SPRGPGH	*Arabidopsis thaliana*	Hou et al. (2014)
Peptide suppressing nodulation (PSN)	D(Hyp)RDHH(sY)FH HNP	*Medicago truncatula*	de Bang et al. (2017)
Phytosulfokine (PSK)	Y(SO3H)IY(SO3H)TQ	*Asparagus officinalis, Arabidopsis thaliana, Lotus japonicus*	Matsubayashi & Sakagami (1996); Zhang et al. (2018); Wang et al. (2015)
Plant elicitor peptide (PEP)	TRTPPWPPCPPEEGS GGNGGSHN	*Zea mays*	Huffaker et al. (2013)
Rapid alkalinization factor (RALF)	ATTKYISYQSLKRNSVP CSRRGASYYNCQNGAQ ANPYSRGCSKIARCRS	*Arabidopsis thaliana*	Du et al. (2016)

Ara, hydroxyproline arabinosylation; Hyp, proline hydroxylation; SO3H, tyrosine sulfation.

are necessary to detect visible phenotypic changes, such as changes in root growth or architecture. Also, adding a secondary experimental variable, such as a pathogenic or symbiotic microorganism, may require pretreatment with peptides to see visible effects. Determining the optimal duration of pretreatment is crucial before proceeding with large scale screening.

Candidate SSP genes can be selected based on their expression profiles in single cell types, tissues, or treatments, which can then be used to test hypothesized functions, typically derived from expression patterns (de Bang et al., 2017). Choosing the peptide sequence to be synthesized can be a difficult task, especially for novel SSP genes that do not belong to established and previously investigated families. In vivo peptide maturation (i.e., peptidase processing and addition of side chain modifications) in the ER (i.e., the secretory pathway) is generally poorly understood (Olsson et al., 2018; Stührwohldt & Schaller, 2019; Stührwohldt, Schardon, Stintzi, & Schaller, 2017). For a novel SSP gene, it is therefore advisable to initially select and synthesize several staggered peptides based on gene alignments and/or HMMs (Fig. 4B). For a novel SSP gene member of an established family sequence, selection should be relatively straightforward, and even common modifications (e.g., proline hydroxylation) of other family SSPs may be known. In some cases, cleavage by peptidases at conserved motifs such as the DY or the RxLR motif can help guide the prediction of the mature peptide sequence

(Schardon et al., 2016; Srivastava, Liu, & Howell, 2008). However, further proteolytic processing steps cannot be ruled out.

Once candidate peptides have been selected and synthesized with desired modifications and requisite purity (usually 90% or more), they can be used for biological assays. Peptides must be in contact with the organism of interest to influence their growth and development. Since peptides can act over long distances (Okamoto, Suzuki, Kawaguchi, Higashiyama, & Matsubayashi, 2015), it is possible that treatment of a single organ system can exert whole-plant effects (Takahashi et al., 2018). In this protocol, however, we focus on changes in root traits, such as architecture and development, induced by synthetic peptide application.

Materials

Lyophilized peptide of interest
Sterile solvent appropriate for peptide
Growth medium (e.g., Fahraeus medium, Broughton and Dilworth [B&D] medium [with dropped nutrients of interest, pH 6.8]; see recipe)
Medicago truncatula seedlings (see Support Protocol 1)

Laminar flow hood
50-ml conical tube
120 × 120–mm (~5 × 5–in.) Petri dishes (e.g., Greiner Bio-One) or sterile growth pouches
Scalpel, sterile
Forceps, sterile
Filter paper, sterile (e.g., Whatman, cat. no. 10334365)
Micropore tape
Black plastic wrap (e.g., ULINE, cat. no. S-17968)
Growth room
Camera
Lightbox
Ruler

Prepare filter paper sandwich plates containing synthetic peptides

It is important to consider the purpose of an experiment before selecting the medium. For example, when screening for peptides that are induced in nitrate-deficient conditions, one might consider using low nitrate in the growth medium (enough to allow plant growth without interfering with peptide activity). A nutrient-rich growth medium such as Fahraeus medium is used routinely for *M. truncatula* growth. The following "filter paper sandwich" growth system was adapted from Breakspear et al., 2014.

1. Dissolve 1-mg aliquots of lyophilized peptides to a concentration of 1 mM (i.e., stock).

   To dilute a peptide to a required working concentration, the following formula can be used: volume (ml) = X / (MW × C), where X is the amount of peptide in mg, MW is the molecular weight in g/mol, and C is the intended final concentration in mol/liter.

   The GRAVY (grand average of hydropathy) value (Kyte & Doolittle, 1982) calculates the overall hydrophobicity of a peptide, and the isoelectric point is a measure of the acidity/alkalinity of a peptide. These values are important to determine the appropriate solvent for the lyophilized peptide.

   Store unused working stocks at −20°C for up to 2 weeks. Avoid repeated freeze-thaw cycles that can result in progressive peptide degradation, reduce biological activity, and compound problems of reproducibility between experiments. A good practice is to freeze small enough aliquots of stocks that will be used in one or two experiments. Store

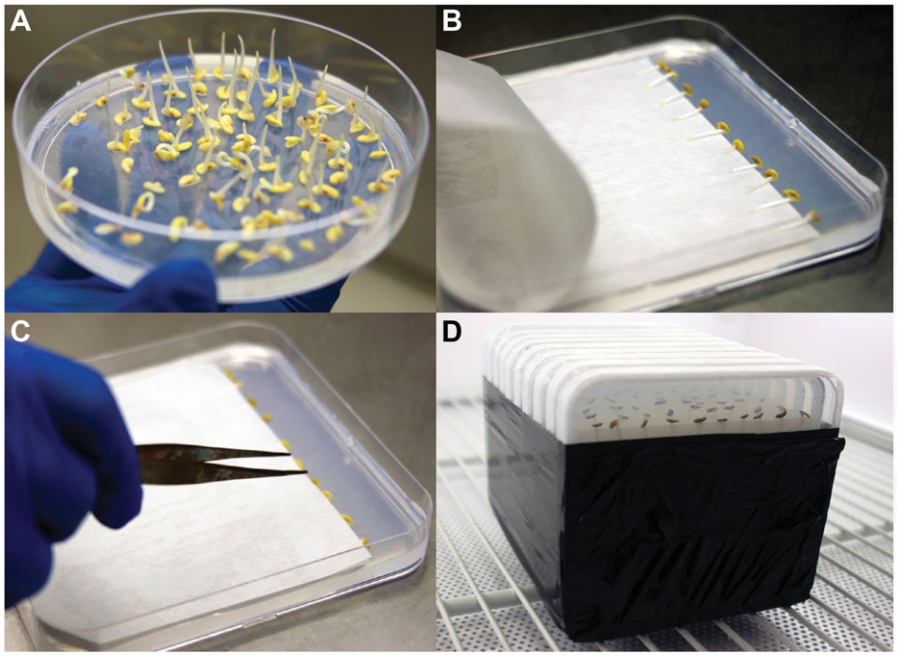

Figure 6 Filter paper sandwich system for peptide application to *M. truncatula* roots. (**A**) Deep dish Petri plate showing emerged radicles after germination overnight. (**B**) Evenly placed seedlings on plates containing peptides of interest. (**C**) Seedlings "sandwiched" between top and bottom filter paper on plates to hold them in place. (**D**) Sealed upright plates placed in a controlled growth chamber with the lower three-quarters covered in black plastic film to simulate below-ground conditions.

lyophilized peptide aliquots indefinitely at −20°C or −70°C in airtight tubes surrounded by desiccant.

2. In a laminar flow hood, under sterile conditions, add peptide at appropriate concentration to a 50-ml conical tube containing appropriate medium that has cooled to the touch but has not yet started solidifying.
3. Mix by inverting gently, and pour into square Petri dishes (~50 ml).
4. Once the gel has set, use a sterile scalpel blade to make an incision ~1 in. from the top of the Petri dish, and discard the upper ~1-in. × 5-in. gel slice using sterile forceps.

Since this is a screen for effects of peptides on root growth, this step ensures that the shoot system is not directly in contact with the medium.

5. Use sterile forceps to place sterile filter paper, cut to the size of the remaining gel, on the solidified medium, and allow it to soak moisture from the plate and adhere to the surface.

Complete contact between the filter paper and the agar medium is critical to ensure wicking of peptide during subsequent growth of the seedlings on the plate (as in steps 10 and 11).

Place Medicago truncatula seedlings on plates

6. Gently remove *M. truncatula* seed coat without damaging the seedlings to minimize growth differences, and place the required number of seedlings equidistant from one another near the top of the Petri dish (Fig. 6A,B).

For germinating and sterilizing M. truncatula seeds, see Support Protocol 1.

When measuring root architecture, it is a good idea to use only 2 or 3 seedlings on a single (120 × 120 × 17 mm) Petri dish to allow unrestricted growth of primary and secondary roots.

Preselecting healthy seedlings with straight emerging roots ensures reproducibility between samples and ensures proper development of seedlings. Any damage to the root tip prevents root growth; therefore seedlings showing any damage should not be used in the assay system. It is recommended to add sterile water to seedlings while working with them to prevent drying.

7. Place filter paper squarely on top of the first filter paper, covering the seedling roots without moving the seedlings.

   Wicking of moisture into the top filter paper ensures complete contact between filter papers and root tissue, forming a tight "sandwich" (Fig. 6 C).

8. Seal plate with micropore tape to allow gaseous exchange while preventing entry of air-borne pathogens.

9. Label with date, medium, and other identifying information, and partially wrap plates with black plastic film to simulate below-ground root growth conditions (Fig. 6 D).

10. Place stacked plates upright (vertically at an angle of ~80°) in a growth room with appropriate environmental conditions.

    For M. truncatula, we use a controlled plant growth chamber with light flux 125 µmol/m^2/sec, 16 hr daylight and 8 hr dark, humidity 80%, and temperature of 24°C.

11. Allow seedlings to grow for the desired length of time (typically 7 to 10 days or just before roots touch the bottom of the plate).

12. Image seedlings with a digital camera by removing the plate cover and placing the agar plate on a lightbox, to provide backlighting and contrast for the roots. Include a ruler in the image for length calibration.

13. Analyze digital images for root growth traits as described in Support Protocol 2.

ALTERNATE PROTOCOL

SCREENING ADDITIONAL ROOT GROWTH PARAMETERS UNDER SYMBIOTIC CONDITIONS

Basic Protocol 3 can be modified to include a secondary variable, such as infection by a pathogenic or mutualistic microbe. The *M. truncatula–Sinorhizobium meliloti* interaction is a model pathosystem for understanding fundamental concepts about root nodule symbiosis. Addition of rhizobia leads to de novo formation of lateral root organs called nodules. Peptides can therefore be screened for effects related to rhizobial symbiosis, such as symbiotic root hair curling, infection thread formation, nodule number, nodule density, nodule position, and zone of infection. To date, peptides from the C-terminally encoded peptide (CEP), phytosulfokine (PSK), and CLAVATA3/endosperm surrounding region-related (CLE) families have been conclusively shown to play roles in root nodule symbiosis (Imin, Mohd-Radzman, Ogilvie, & Djordjevic, 2013, Mortier et al., 2010, Wang et al., 2015). This protocol describes a method for infection on a plate system using the *S. meliloti* strain Sm2011. Using transgenic variant Sm2011 harboring the *hemA*::LacZ construct allows the additional option of screening for infection threads if desired (Pichon et al., 1994).

Materials

Compatible rhizobial strain (*S. meliloti* strain Sm2011 or *S. medicae*)
Tryptone yeast extract (TY) medium (see recipe)

1 M $CaCl_2$, filter sterilized
Seedlings pretreated with peptide (see Basic Protocol 3)

28°C incubator with shaking
Spectrophotometer
Refrigerated centrifuge (e.g., Sorvall Legend RT Plus or equivalent)
Forceps, sterile
Laminar flow hood
Petri dish
Black plastic wrap (e.g., ULINE, cat. no. S-17968)
Growth room
Stereomicroscope
Camera or scanner
Ruler

1. Set up overnight cultures of Sm2011 in 10 ml TY medium supplemented with 10 mM $CaCl_2$ at 28°C under constant shaking (200 to 250 rpm). Allow bacteria to grow until an optical density at λ=600 nm (OD_{600}) of 1.0 is reached (usually 16 to 18 hr), which corresponds to the exponential growth phase of the bacteria.

2. To prepare inoculum, centrifuge cultures 15 min at 3600 × *g* (~4000 rpm), 4°C. Discard supernatant and resuspend pellet in 10 ml sterile water. Dilute resultant bacteria suspension to an OD_{600} of 0.05 in sterile water, reserving ~1 ml inoculum per plate.

3. Inoculate seedlings that have been pretreated with peptides for the desired duration (24 hr in our system). On the day of inoculation, remove and discard the top filter paper using sterile forceps in a laminar flow hood. In a separate plate, pipette 700 µl rhizobial inoculum (from step 2) onto fresh filter paper, and transfer onto seedlings.

   Between 5 and 10 seedlings can be used per plate.

4. Wrap plates with black plastic film as described in Basic Protocol 3 step 9 (Fig. 6D), and allow plants to grow for an additional 7 to 10 days before scoring nodulation under a stereomicroscope.

   Growth parameters that can be screened are shown in Figure 7.

5. In addition to scoring root growth parameters manually under the microscope, document the experiment by collecting digital images or scanning the plates with a digital scanner. Include a ruler for calibrating the distances in image analysis software such as SmartRoot or WinRhizo (see Support Protocol 2).

PREPARATION OF GERMINATING *MEDICAGO TRUNCATULA* SEEDS

SUPPORT PROTOCOL 1

All experiments require seedlings at the same growth stage for use as replicates. Inbreeding species such as *Medicago truncatula* are valuable for research purposes as progeny from a single parent can be considered as genetically almost identical. Still, slight differences in germination and growth exist in such sibling populations, which can affect and confound results. To ensure sufficient near-identical plant material for an experiment, seed scarification, sterilization, and cold pretreatment are key. Seed scarification is required to weaken the seed coat and allow seeds to imbibe water, which raises turgor pressure enough to allow seedling emergence. Surface sterilization is essential for avoiding the introduction of bacterial and fungal pathogens. A cold pretreatment at 4°C ensures more synchronous germination of seeds upon transfer to germination-permissive temperatures.

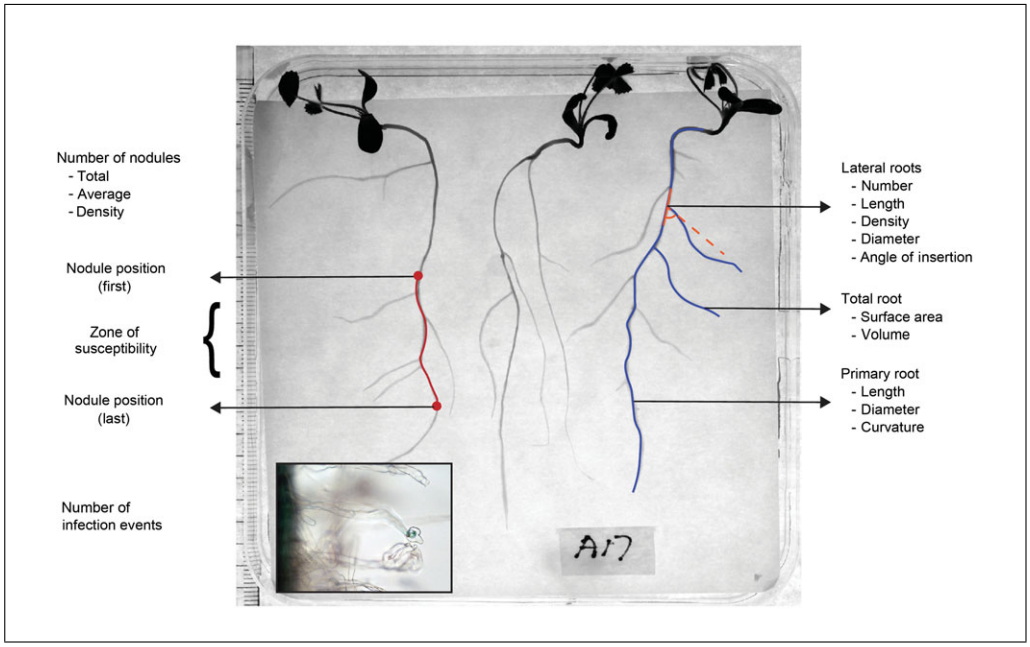

Figure 7 Root system architecture traits under consideration. Image showing root growth parameters (blue) and nodulation traits (red) that can be scored in the described plate setup using peptides. Inset image shows a developing infection thread stained with X-gal (blue).

Materials

Medicago truncatula seeds (e.g., A17 or R108)
H_2SO_4, HPLC grade
10% bleach
Deep-dish petri plate (e.g., Sigma-Aldrich, cat. no. P7741) containing B&D medium or Fahraeus medium with 1% agarose (see recipes)

Conical tube or glass tube, sterile
Aluminum foil or black plastic wrap (e.g., ULINE, cat. no. S-17968)

Seed scarification

1. Count out number of seeds required, and place seeds in a capped conical tube or sterile glass tube.

   For Medicago, a general rule is 0.4 g for 100 seeds. It is not recommended to scarify more than 5 g of seeds in a single tube or vial.

2. Cover seeds with H_2SO_4 (~5 ml for 1000 seeds). Treat for 10 min or until brown flecks appear on a majority of the seeds.

   Treatment with H_2SO_4 for longer than 12 min will damage the seeds and strongly reduce germination percentage.

3. Carefully siphon off acid under a fume hood, and discard into a large volume of water. Wash seeds with sterile distilled water ~5 times (2 to 5 min per wash) by inverting tubes gently by hand.

   It is recommended to use ice-cold water for the first wash since adding water to acid is an exothermic reaction that generates heat, which might be detrimental to seed germination. Alternatively, discard the first wash rapidly.

   Neutralize acidified water with commercial baking soda before disposal.

   An alternate method for seed scarification utilizes sandpaper. Usually fine to medium-coarse paper is enough to form gentle abrasions on the seed surface.

Surface sterilization and germination of seeds

Perform the remaining steps in a laminar flow hood.

4. To scarified seeds, add 10% bleach to just cover the seeds. Do not allow treatment with bleach to proceed beyond 2 to 3 min.
5. Use sterile pipette tips to remove bleach, and discard into a designated disposal container. Dilute with tap water before disposing.
6. Wash seeds repeatedly by adding sterile water (2 to 5 min per wash) and inverting tubes gently by hand before decanting and disposing of the washes.

   Since considerable damage to seeds, and consequently their germination percentage, can occur at this stage, it is important to handle seeds very carefully. Any residual bleach, if present, can be determined by the chlorine-containing odor of the water in the last wash and can severely diminish seed germination rates.

7. Leave seeds in water for at least 1 hr or until they imbibe water and swell.

   Noticeably, the seeds change color from brown to a dark yellow.

8. Place seeds 1 by 1 on a deep-dish agar plate (50 to 100 seeds per plate), and store at 4°C for 3 days. Wrap plates in aluminum foil or black plastic film to mimic below-ground dark conditions.

   Ensure that there is enough airspace within the Petri plate to allow radicles to emerge at least 1 to 2 cm.

9. Germinate seeds by inverting the plate in the growth room overnight.

   Plates should still be wrapped in foil or film at this stage.

IMAGE ANALYSIS OF ROOT GROWTH TRAITS

SUPPORT PROTOCOL 2

The past 15 years have seen a rapid emergence of bioinformatics tools for the analysis of root system architecture. This growth has paralleled the growing interest in breeding for beneficial root growth and architectural traits and reflects a desire to rapidly screen germplasm, environmental conditions, or chemical treatments for impacts on the root. These tools work with digital images of roots captured by an investigator and assist in the tracing and measuring of root parameters, including lengths, widths, and angles. Depending on the software, the tool may be automated or semiautomated and may require the purchase of a license (e.g., WinRhizo; Arsenault, Poulcur, Messier, & Guay, 1995) or be freely available as open-source (e.g., SmartRoot, ARIA, RootNav, or EZ-Rhizo; Lobet, Pagès, & Draye, 2011; Pace, Lee, Naik, Ganapathysubramanian, & Lübberstedt, 2014; Pound et al., 2013). The possible trait parameters outputted by a software program can also vary widely depending on the complexity of the software; however, virtually all share an ability to output certain common root traits such as length, diameter, angles, or number. We are interested in studying the effect of SSP gene–derived synthetic peptides on root growth traits in *Medicago truncatula* and other plant species. For our purposes, we favor the open-source SmartRoot software, which has become an indispensable step in our peptide screening pipeline. SmartRoot is a semiautomated tool providing an extensive set of measured traits for downstream analysis. The software is platform independent, and metadata is saved in RootSystemML file format (Lobet et al., 2015) for data portability across software tools and interoperability. Here, we present the analysis steps for measuring root traits using the SmartRoot software program, as applied in de Bang et al. (2017).

This protocol presents the basic tasks essential for analyzing root architecture with the SmartRoot program. However, there are many additional capabilities. We direct the reader

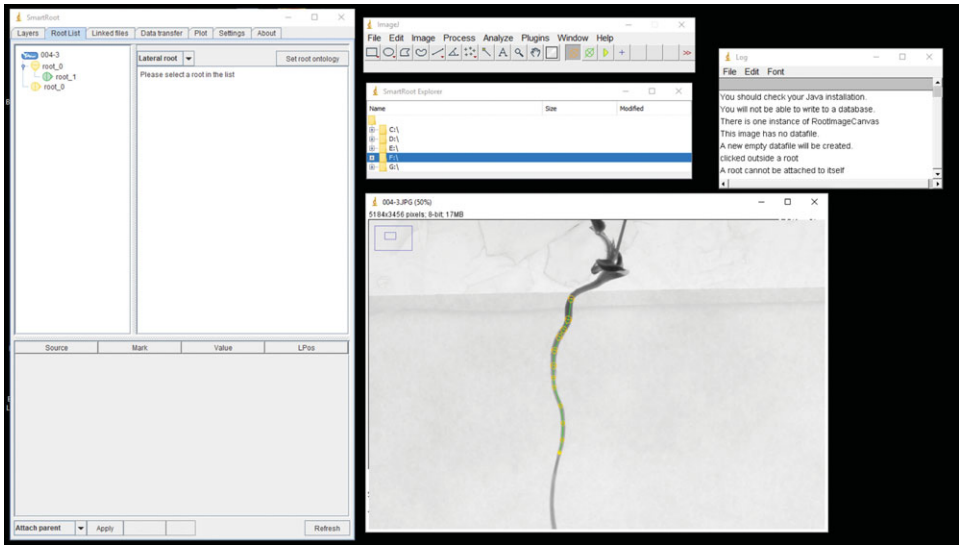

Figure 8 Screenshot of the SmartRoot software. The program operates as a plugin in ImageJ and is operated through four windows. The SmartRoot window is opened to the Root List tab where the primary roots for two plants are seen in yellow (second plant not shown).The image window, showing the image currently being analyzed, has been partially traced. Nodes are identified as yellow circles and the estimated center of the root as a green line.

to the SmartRoot User Guide for detailed explanations of capabilities and instructions. See the same User Guide for directions to install SmartRoot.

We find the most orderly way to organize files prior to analysis is for each experiment to have its own single folder. In this way, all files to be analyzed and compared are found together in a single folder. The software will work equally well with other approaches. However, it must be kept in mind that when measurements are outputted into .csv files using the batch export tool for downstream statistical analysis, all image files present within the same folder will be compiled into one .csv file.

Materials

ImageJ software, available at *https://imagej.net/*
SmartRoot v4.1 software plugin for ImageJ, available at *https://smartroot.github.io/*
.tif or .jpg images to be analyzed, converted to 8-bit grayscale

1. To open SmartRoot, first open ImageJ, and then navigate to the Plugins tab, then SmartRoot, and then SR Explorer.

   The SmartRoot plugin opens four different windows (Fig. 8).

2. Open an image to analyze within SmartRoot by navigating through the file directory within the SmartRoot Explorer window and double clicking on the desired file.

   SmartRoot only processes 8-bit grayscale images. A warning message will appear if the image has not already been converted.

3. Image resolution, for the scale of the image, can be set before or after tracing and can be updated or corrected at any time. To set the scale, select the line tool found in the ImageJ window, and draw a line between two points of known distance. Click the Get Line button, and enter the corresponding physical length in either mm, cm, or in. Finally, click the Apply button to automatically update measurements from pixel units to the physical length units selected previously.

4. Central to the accurate measurement of root traits is the proper tracing of the roots. Tracing can be done either manually or semiautomatically. First, select the trace button from the ImageJ window. To trace manually, place the cursor at the beginning of the root, and click to place a node. Continue placing nodes successively along the length of the root until you reach the root tip. To mark the final node of the root at its tip, double click while placing the node. For semiautomated tracing, first place a node at the beginning of a root, as for the manual tracing procedure. Then, while placing a second node hold down the Alt key on the keyboard. The software will place the second node as well as the nodes for the remainder of the root, according to the predicted limits of the root. It is possible, and sometimes advisable, to manually trace multiple nodes of the root before employing the semiautomated tracing ability.

   To see what you are tracing, make sure the Display Nodes and Display Axis options are selected from the Layers tab of the SmartRoot window. When a complete root has been traced, you have the option to name the root. It is also possible to manually adjust the placement of single nodes after a root has been traced by clicking and dragging the node.

5. Once all roots of an individual have been traced, it is necessary to define the relationship of parent roots and child (or lateral) roots. Within the RootList tab of the SmartRoot window is a list of each traced root. Select one or more lateral roots from the list, and activate the Attach Parent Root tool at the bottom of the tab. Then, select the parent root from the list, and confirm the selection by clicking ok. Repeat this process until all roots have been assigned a parent root (except the taproot).

   Relationships are displayed using hierarchical tree format in the root list. Be sure to click the Refresh button in the bottom right corner of the tab to make sure the relationships are up to date.

6. Measurement results can be viewed from the same Root List tab within the SmartRoot window. Select a root from the list, and view a number of relevant parameters in the sub-window on the right of the tab. Parameters include root length, number of lateral roots, mean diameter, and surface area, which along with volume, is inferred from the length and diameter measurements of the root.

7. To extract the dataset from an experiment to a .csv file, navigate to the Data Transfer tab of the SmartRoot window. Select the Send to CSV file checkbox. Choose both the source folder (i.e., the folder containing the analyzed images of your experiment) and an appropriate location to save your outputted .csv file. Click the Run batch export button.

   Exported data files can be further analyzed using a number of approaches including Excel or RStudio (RStudio Team, 2015).

CALCIUM BURST ASSAY FOR *ARABIDOPSIS* ROOTS

BASIC PROTOCOL 4

Free Ca^{2+} ion pools are maintained in subcellular compartments of cells, such as the ER, and are a common second messenger to convey signals originating from outside of the cell. Upon perception of an appropriate stimulus, Ca^{2+} ion pools are released into the cytosol inducing a transient increase in calcium concentration ($[Ca^{2+}]$), which is known as a calcium burst or calcium spike. These calcium bursts are deciphered by calcium-binding proteins that further convey the signal to downstream processes, leading to suitable physiological responses, such as transcriptional changes in the nucleus. Interestingly, transient changes in $[Ca^{2+}]$ will take on a waveform over a time frame of seconds to minutes. Variations in the frequency or amplitude of the changing $[Ca^{2+}]$ wave can be distinguished and produce differing downstream responses.

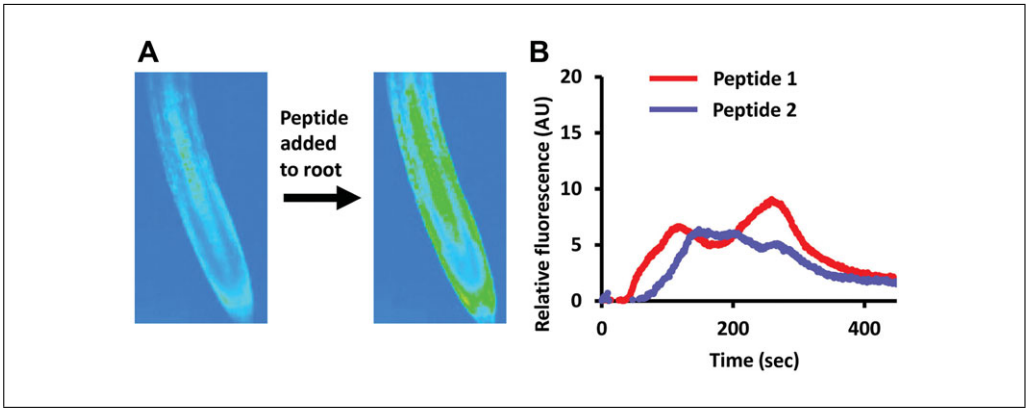

Figure 9 Calcium spike assay results. (**A**) Time points collected from video collection illustrate increased fluorescence in response to addition of peptide. (**B**) Time course of the fluorescence intensity over time illustrates the rate of induction, amplitude, and duration of the calcium spike. Plots of two different peptides are presented. The 0-sec time point represents the time the peptide was administered. AU, arbitrary unit.

Perception of certain hormones, including several peptide hormones (Haruta et al., 2008), have been shown to induce calcium burst signals that are required for their downstream biological effect. It is likely that there are additional peptide hormones that similarly use [Ca^{2+}] burst signals to stimulate biological effects, and an effective method to screen for these effects is needed. Several genetically encoded fluorescent reporters for Ca^{2+} have been developed that produce fluorescent signal upon binding of Ca^{2+} (Baubet et al., 2000; Nagai, Sawano, Park, & Miyawaki, 2001; Nakai, Ohkura, & Imoto, 2001). Here we describe a moderate-throughput approach to test synthetic peptides for ability to induce transient [Ca^{2+}] bursts in root tissue of *Arabidopsis thaliana* seedlings. This protocol employs an *A. thaliana* line stably expressing the calcium reporter GCaMP (Nakai et al., 2001) and uses direct application of peptide while under the confocal microscope to monitor fluorescence over a time course (Fig. 9; Supporting Information Video). This protocol permits only a relative comparison of fluorescence changes in response to peptide application but is higher throughput, lending itself well to screens. Other fluorescent reporters are available which can be calibrated to provide measurement of absolute concentrations in the cell.

Materials

Growth medium with agarose (GMA; see recipe)
A. thaliana seed stably expressing GCaMP fluorescent reporter
Synthetic peptide of interest
Growth medium (GM; see recipe)

Glass coverslips (7 × 4–cm)
Disposable L-shaped spreader, sterile
Square Petri dishes (15 × 100–mm)
Toothpicks, sterile
Parafilm
Growth chamber
Confocal microscope equipped with a GFP filter set and video camera

Sowing of A. thaliana seeds (day 1)

1. In a sterile fume hood, pipette 2 ml melted GMA onto a coverslip, and gently spread over the coverslip with an L-shaped spreader.

The volume of medium has been optimized for this size of coverslip and should cover the entire coverslip to a proper thickness. Be careful to prevent medium from spilling off the coverslip, relying on the surface tension between coverslip and media.

2. Transfer coverslip to a Petri dish to solidify.
3. Using a moistened, sterile toothpick, sow 5 sterilized seeds on each coverslip, leaving 1.5 cm from the top of the coverslip and an even spacing laterally.

   A. thaliana seed can be prepared in bulk and stored at room temperature for at least 1 year. Sterilize using a standard ethanol and bleach protocol and deposit on filter paper in Petri dishes. Do not stratify seeds (step 5) until they have been sowed in preparation for the assay.

   The even spacing should give about 1 to 2 cm between each seed, which is important to ensure application of peptide to only one individual seed at a time. The 1.5-cm spacing from the top provides sufficient space for emergence and growth of cotyledon while maximizing the space for root growth.

4. Cover Petri plate, and seal tightly with two layers of Parafilm.

   It is important to fully seal the plate to prevent rapid dehydration of the medium during stratification and germination.

5. Stratify seeds by placing plates at 4°C for at least 48 hr.
6. Germinate seeds by incubating plates in a growth chamber for 3 days at 22°C, 120 µE light intensity, and 16 hr photoperiod.

   It is beneficial to have the plates tilted backwards at a slight angle such that the force of gravity pushes the roots continually into the agar. This is to keep the roots growing tightly against the medium. The subsequent peptide treatment will give spurious results wherever root tissue is growing through air.

Peptide treatment and fluorescence microscopy (day 4)

7. Resuspend synthetic peptide in GM to a final concentration of 15 µM.

   A total of 15 µM is an effective concentration for screening purposes to distinguish between true and spurious (i.e., nonphysiological) elicitation.

   This peptide concentration is artificially high for screening. It will be appropriate to adjust the concentration for other purposes, such as follow-up studies on a peptide of interest or dose-response curves.

8. Bring the plates and peptide solutions to the microscopy room 1 hr prior to beginning assays to allow plates to equilibrate to room temperature.

   Bringing the seeds, agar medium, and peptide solutions to a constant and equal temperature prior to assaying is critical. Calcium bursts in plant tissue are highly sensitive to temperature changes (as well as osmotic or pH changes).

9. Open one plate, and place the coverslip securely on the microscope stage. Focus the objective on a single root tip. Wait 5 min before continuing to peptide application to allow the plant tissue to adjust to the movement and change in humidity that results from removal of the plant from the plate and its placement on the stage.
10. Using the appropriate GFP excitation and emission settings, begin collecting video, and then carefully spot 10 µl peptide solution onto the root tip, keeping the pipette tip ~1 cm above the root tip when depositing the peptide.

    It is advised to apply the peptide at a consistent time after starting the video for all recordings. This facilitates comparison of multiple recordings during data analysis. Take care to not bump the root with the pipette, which will induce a calcium burst.

11. Collect a 10-min video to monitor changes in GFP fluorescence.

Application of the negative control should not produce additional fluorescence above background levels. In contrast, application of AtRALF should induce a single, extended burst of Ca^{2+} beginning about 1 min after application and slowly tapering over the course of the 10-min video (see Supporting Information Video).

12. Quantify and plot changes in $[Ca^{2+}]$ over the time course of the collected video using the confocal instrument software. Quantify fluorescence intensity within a region of interest for each frame.

This data can be exported as a tabular file, such as a comma separated values (csv) file and plotted as the change in fluorescence intensity (y axis) over time (x axis) in a software program such as Microsoft Excel.

It may be most useful to select a specific region of tissue for quantification, rather than from the entire field of view. In this case, it is essential to maintain the same region size for all individuals for proper comparison.

REAGENTS AND SOLUTIONS

Broughton and Dilworth (B&D) medium

To 1 liter double-distilled water add:
1.0 ml 500 mM KH_2PO_4 (0.5 mM final)
1.0 ml 250 mM K_2SO_4 (0.25 mM final)
1.0 ml 250 mM $MgSO_4 \cdot 7H_2O$ (0.25 mM final)
2.0 ml 1 M KNO_3 (2 mM final)
2.0 ml 1 M NH_4NO_3 (2 mM final)
1.0 ml 1 M $CaCl_2 \cdot 2H_2O$ (1 mM final)
1.0 ml 1.0 mM $MnSO_4 \cdot 7H_2O$ (1 μM final)
1.0 ml 2.0 mM H_3BO_3 (2 μM final)
1.0 ml 0.5 mM $ZnSO_4 \cdot 7H_2O$ (0.5 μM final)
1.0 ml 0.2 mM $CuSO_4 \cdot 5H_2O$ (0.2 μM final)
1.0 ml 100 μM $CoSO_4 \cdot 7H_2O$ (0.1 μM final)
1.0 ml 100 μM $Na_2MoO_4 \cdot 2H_2O$ (0.1 μM final)
1.0 ml 10 mM Fe-citrate (10 μM final)
Store at 4°C for up to 6 months

For nodulation assays, add KNO_3 to a final concentration of 500 μM, and do not add NH_4NO_3.

For the micronutrients (Mn, B, Zn, Cu, Co, Mo, Fe) a combined 1000× stock solution can be prepared for convenience if B&D medium is used frequently.

Adapted from Broughton & Dilworth (1971).

Fahraeus medium

0.5 ml 1000 mM $MgSO_4 \cdot 7H_2O$ (0.5 mM final)
1.5 ml 500 mM KH_2PO_4 (0.75 mM final)
1.5 ml 500 mM $Na_2HPO_4 \cdot 7H_2O$ (0.75 mM final)
1 ml 1000 mM NH_4NO_3 (1 mM final)
1 ml 1000 mM $CaCl_2 \cdot 2H_2O$ (1 mM final)
400 μl 2500× micronutrients solution (see recipe)
20 g gellan gum (e.g., Gelzan, Caisson; 2% [w/v] final)
Bring volume to 1 liter with sterile double-distilled water
Store at 4°C for up to 6 months

Growth medium (GM)

0.215% (w/v) Murashige and Skoog basal salts with vitamins
2 mM 2-(N-morpholino)ethanesulfonic acid (MES) sodium salt

1% (w/v) sucrose
Adjust pH to 5.7 with 1 M KOH
Aliquot and store at −20°C for up to 2 years

Growth medium with agarose (GMA)

GM (see recipe) supplemented with:
0.5% (w/v) low-melting-temperature agarose (e.g., NuSieve GTG, Lonza)
Autoclave to dissolve agarose
Store at room temperature for up to 12 months

Microwave before use.

Micronutrients solution, 2500×

To 100 ml double-distilled water add:
927 mg H_3BO_3 (150 mM final)
530 mg $MnSO_4$ (35 mM final)
72 mg $ZnSO_4 \cdot 7H_2O$ (2.5 mM final)
24 mg $CuSO_4 \cdot 5H_2O$ (1.5 mM final)
13 mg $NiCl_2$ (1.0 mM final)
12 mg $HMoO_4$ (0.75 mM final)
0.7 mg $CoCl_2 \cdot 6H_2O$ (50 μM final)
3700 mg ferric sodium EDTA (100 mM final)
Store at –20°C for up to 12 months

Adopted from Scheible et al. (2004).

Tryptone yeast (TY) medium

1 liter water
5 g tryptone
3 g yeast extract
Store at room temperature for up to 3 months

COMMENTARY

Background Information

Small secreted peptides (SSPs) are now known to not only regulate many important aspects of plant development but also plant nutrient acquisition and abiotic and biotic stress adaptation (Olsson et al., 2018; Roy, Lundquist, Udvardi, & Scheible, 2018). SSPs (5 to 60 amino acids long) are also called peptide hormones because they display many characteristics of the classical phytohormones. This includes their ability to affect biological processes at very low (nanomolar) concentrations (Matsubayashi, Ogawa, Kihara, Niwa, & Sakagami, 2006). The first described plant peptide was called systemin and was reported to induce proteinase inhibitor proteins in tomato leaves upon wounding (Pearce, Strydom, Johnson, & Ryan, 1991; Ryan, 1974). Several years later, a small sulfated peptide, phytosulfokine, was reported to induce proliferation of asparagus mesophyll cells (Matsubayashi & Sakagami, 1996). Another milestone study in plant peptide research was the discovery of CLAVATA3 (CLV3), a small, extracellular peptide ligand of the CLAVATA1 receptor kinase that signals cell fate decisions in *Arabidopsis* shoot meristems (Fletcher, Brand, Running, Simon, & Meyerowitz, 1999).

Faster progress in plant peptide and SSP biology, however, has been stymied for years by the fact that early genome annotation pipelines were biased against small open reading frames (sORFs), which were most likely to encode small peptides, because they were difficult to distinguish from random sORFs (Lease & Walker, 2006; Olsen, Mundy, & Skriver, 2002), the number of which increases exponentially with decreasing length. Moreover, the small size of peptide-coding genes also reduced the probability of recovering mutants by gene knockdown or knockout methodologies such as chemical (e.g., ethyl methanesulfonate [EMS]) or biological (e.g., T-DNA) mutagenesis. Not being (well) annotated and known,

peptide-coding genes were not appropriately studied (Lease & Walker, 2006).

Despite some evidence from cDNA libraries or expressed sequence tag (EST) collections (e.g., Riaño-Pachón, Dreyer, & Mueller-Roeber, 2005), the realization of the abundance of small expressed genes, including those encoding SSPs, only increased in recent years with the ability to perform powerful comparative genomics on short proteins with a signal peptide in plant genomes with superior annotations (Ghorbani et al., 2015) and with the amassment of RNA sequencing (RNA-seq) datasets in the public domain. RNA-seq provides the strongest argument for the existence and expression of unannotated small genes and sORFs with low expression, such as those encoding SSPs. Determining which of the expressed sORFs encode potential SSP-coding genes or simply are novel members of established SSP gene families is another challenge that requires additional filtering steps or approaches as outlined here (Basic Protocols 1 and 2) and elsewhere (e.g., de Bang et al., 2017; Ghorbani et al., 2015; Hastwell, de Bang, Gresshoff, & Ferguson, 2017; Murphy et al., 2012; Olsson et al., 2018).

Conservation within a given SSP family is typically limited to short stretches of amino acid residues, reducing the effectiveness of traditional homology search strategies, such as NCBI BLAST. Furthermore, the rules of peptidase processing and SSP maturation, including the addition of post-translational modifications in vivo, are poorly defined (Matsubayashi, 2014; Olsson et al., 2018; Stührwohldt & Schaller, 2019). These latter issues can significantly complicate the selection of sequences for chemical peptide synthesis (i.e., solid-phase Merrifield synthesis) or even prevent synthesis of a bioactive peptide, as exemplified by the importance of the tri-arabinosylation modification of some CLAVATA3/endosperm surrounding region-related (CLE) peptides (Imin, Patel, Corcilius, Payne, & Djordjevic, 2018). Nonetheless, gene-derived synthetic peptides are an important and straightforward approach to probe SSP gene function or to identify novel effectors of specific traits of interests (e.g., root architectural traits; Patel et al., 2018; also see Table 4) with potential use in agriculture (Lee, Huffaker, Crippen, Robbins, & Goggin, 2018). In this regard, gene-derived synthetic peptide libraries in the public domain would represent highly useful tools for biochemical genomics and leverage the cost for their synthesis. Novel databases, such as MtSSPdb, that (1) descriptively and comprehensively display identified SSP genes, (2) provide RNA-seq-based expression histories, and (3) collect and merge experimental data (such as phenotypes induced by derived synthetic peptides) will be highly instrumental in advancing SSP biology.

Critical Parameters

The interdisciplinary and, in part, complex procedures described are subject to a number of critical parameters that naturally are very different for bioinformatics and laboratory-based protocols. Here we identify parameters with reference to the respective protocol.

Basic Protocol 1: The performance of SSP-coding gene identification by MAKER strongly relies on evidence of gene expression and protein products. A vast amount of RNA-seq expression data, ideally from different tissues, developmental stages, and (stress) treatments, is needed for identification of a near-complete complement of SSP-coding genes from genomic sequences. This prerequisite is further stressed by the fact that SSP-coding genes show low average expression. Publicly available RNA-seq data (FASTQ files) for many species can be found (e.g., NCBI sequence read archive [SRA] database).

Basic Protocol 1: At the end of the third round for MAKER genome annotation, the final GFF3 format file contains gene models and evidences. The quality of predicted gene models can be evaluated by annotation edit distance (AED) scores in the GFF3 file. AED scores range from 0 to 1, and the lower the AED, the higher the confidence of a gene model. Gene models with AED <0.5 are considered high confidence.

Basic Protocol 1: To remove duplicate gene models between MAKER and SPADA pipelines, we search coding sequences (CDS) of gene models from MAKER against the CDS from SPADA using NCBI BLASTN. Gene models with overlapping coordinates on the same chromosome and at least 50% CDS identity will be considered duplicates. The 50% cutoff can be decreased to remove more duplicate gene models.

Basic Protocol 2: The SignalP algorithm is based on neural network and hidden Markov model (HMM) algorithms (Petersen et al., 2011). The SignalP server generates five different scores: C, D, S, S-mean, and Y scores. The graphical output reports on the C (raw cleavage site score), S (signal peptide score), and Y (combined cleavage site score) scores.

Table 4 Phenotypes Affected by Synthetic Peptides

Peptide	Screen parameter	Reference
AtPIP	Stomatal aperture size, reactive oxygen species burst, callose deposition, root growth inhibition, marker gene induction, MAP kinase assay	Hou et al. (2014)
AtRALF1	Cytosolic calcium burst	Haruta et al. (2008)
AtPEP, GmPEP	Cytosolic calcium burst, nitric oxide generation, nematode infestation	Ma et al. (2013); Lee et al. (2018)
EPIP	Percentage of flowers abscised	Stenvik et al. (2008)
CLE	Shoot apical meristem size	Xu et al. (2015)

The D-score is used to discriminate signal peptides from non-signal peptides, and it combines the S-mean and Y scores. A D-score ≥ 0.25 is a criterion for putative SSP candidates.

Basic Protocol 2: The Markov cluster (MCL) algorithm allows a single parameter to control the granularity of the output clustering, namely the –I inflation option. In this protocol, we choose –I = 1.4 to generate coarse-grained clusterings (i.e., relatively large clusters). You may increase this value to generate smaller clusters. A good set of starting values for granularity is 1.4, 2, 4, and 6 per the suggestion from the MCL manual (*https://micans.org/mcl/*).

Basic Protocol 2: The e-value cutoff for Smith-Waterman and HMM searches was set to 0.01, which ensures discovery of more potential homologs of known SSPs. Lowering the cutoff value (e.g., 1e-3 or 1e-4) will generate more stringent results. Refer to GitHub online document sections 2.3 and 2.4.3 for e-value adjustments.

Basic Protocol 3: The choice and handling of synthetic peptides is critical for subsequent applications. Consider peptides with high purity (preferably 90% or more; i.e., peptides largely devoid of truncated side products produced during synthesis) and with good quality control documentation. Order from a supplier that provides several precisely weighted aliquots of lyophilized peptide rather than as a single sample. For additional tips and guidelines, see the protocol introduction for Basic Protocol 3.

Basic Protocol 3: A peptide's effective concentration and activity can be influenced by its binding affinity towards surfaces. Using non-reactive Petri dishes and germination paper that do not sequester the peptide ensures maximum peptide activity. In our experience, using thicker germination paper (blue) dampens the plant response to peptides.

Basic Protocol 4: Be sure to provide time for the seedling to acclimate on the microscope before beginning the assay. Calcium bursts are highly sensitive to changes in temperature, humidity, and movement and will be induced following removal from the dish and placement on the microscope slide. Waiting several minutes allows intracellular calcium to return to basal levels critical to an accurate assessment of peptide effects.

Basic Protocol 4: The use of suitable positive and negative controls is crucial. An appropriate negative control is application of GM medium prepared without the addition of a peptide. This controls for temperature or osmotic effects on calcium burst. Other negative controls are scrambled or point-mutated versions of a given bioactive peptide of interest. These, most likely, have no biological activity. The AtRALF1 peptide (Haruta et al., 2008), with a well-characterized effect on calcium release, is an effective positive control to ensure proper peptide perception and GCaMP fluorescence emission in *A. thaliana* lines.

Alternate Protocol 4: Use freshly streaked bacteria to set up cultures, and do not grow bacterial cultures beyond $OD_{600} = 1.0$. Most differences in infection efficiency, and therefore experimental outcome, depend on the growth phase of the bacteria. If the culture's OD_{600} extends beyond 1.2, or if the bacteria haven't grown enough, infection rates will be poor.

Troubleshooting

MAKER or SPADA usage errors can be found at *https://groups.google.com/forum/#!forum/maker-devel* or *https://groups.google.com/forum/#!forum/SPADA*, respectively.

The MAKER tool has many parameters (see Internet Resources) that can be

customized and that will affect the results. It is advisable to carefully check three configuration files, especially the file named maker_opts.ctl. Note the maker_opts.ctl file for the first round is different from the one for the second and third rounds in the Gene Prediction section.

The Docker image developed for Basic Protocols 1 and 2 were only tested using Linux CentOS 7 and Ubuntu 16.04 LTS. We do not recommend using the program with Windows and Mac, although it can be installed on Windows and Mac systems. As a Linux user, installing Docker support software and starting backend service require root or sudo privileges. Downloading Docker image and starting a container for the image only requires the user to be a member of the Docker user group or to have root or sudo privileges. Contact your Linux administrator if you are using a virtual Linux machine without root or sudo privileges in the Data center and having permission or privileges issues when running the Docker container.

HMMER software is used for the HMM search in Basic Protocols 1 and 2. Be advised that the HMM libraries compiled by different versions of HMMER are incompatible each other. The installed HMMER in Docker image is version 3.0. If user plans to compile an HMM library using their SSP family data, use the same version.

Frequently asked questions regarding SignalP can be found at *http://www.cbs.dtu.dk/services/SignalP/faq.php*. TMHMM instruction information can be found at *http://www.cbs.dtu.dk/services/TMHMM/TMHMM2.0b.guide.php*. The HMMER help page can be found at *https://www.ebi.ac.uk/Tools/hmmer/help*. Frequently asked questions regarding the MCL algorithm can be found at *https://micans.org/mcl/*.

The following comments apply to laboratory protocols using synthetic peptides and image analysis of plate-grown plants.

Effects are not reproducible between experiments: Use a fresh aliquot of the peptide stored at −80°C. Repeated freeze-thaw cycles likely degrade the peptide and thereby reduce its activity. Also check that the peptide is completely soluble in the chosen solvent. Incomplete solubilization and/or precipitation reduces peptide activity.

Predicted peptide does not show activity: If the peptide was synthesized based on in silico predictions alone, it is possible that some secondary modifications might exist that have not been included in the synthetic version of the peptide. Additionally, cysteine-rich peptides that are not correctly folded might not show activity when only the peptide chain is synthesized (Takeuchi & Higashiyama, 2012). Consider purifying the biologically active peptides from lines overexpressing the peptide-coding gene and detecting the secondary modifications using mass spectrometry. Alternatively, correctly folded peptides can be purified from bacteria expressing the peptides of interest. It is also possible that the correct sequence was not chosen for peptide synthesis. More than one SSP can be derived from a single precursor polypeptide chain (Kinoshita et al., 2007), and not selecting the key SSP-encoding gene family member or the derived peptide can affect the outcome of the experiment.

Large quantities of peptide are required: Perform a serial dilution to determine the lowest concentration at which the peptide shows activity.

Software misidentifies plate edges as roots: Extensive manual curation may be required to eliminate misidentification of roots by the image analysis software. Seedlings that grow too close to the edge not only show different growth characteristics and are not truly representative of root growth but also cannot be easily distinguished by the software. Distribute seedlings at least 1 in. from each other and from plate edges.

Nodulation efficiency is low on plates: Adding a secondary factor such as a pathogen or a symbiotic microbe can add another level of variability to the experiment, which affects its reproducibility. The accumulation of ethylene, a negative regulator of rhizobial infections, in plate systems likely affects the number of nodules formed (Oldroyd, Engstrom, & Long, 2001). To improve efficiency of nodulation, ensure freshly streaked bacteria are used as inoculum in addition to harvesting the bacterial cultures only in the log phase ($OD_{600} = 1.0$). Although, AVG (2-aminoethoxyvinyl glycine) is a common ethylene inhibitor that can be added to improve infection rates on plates, it is not recommended for screening purposes since it might interfere with the activity of the peptide itself. Lastly, consider using high-efficiency rhizobia strains such as Sm2011 rather than Sm1021 for screening.

High biological variability: If a peptide treatment gives conflicting results from replicate to replicate in the calcium burst assay, plants may not have had sufficient time to acclimate to the environment of the microscope (see Critical Parameters). Provide at least 5

min before beginning treatments of plants. Additionally, it should be considered that roots occasionally may not maintain contact with the agar, which will invariably lead to induction of calcium burst upon application of a solution (even the medium without peptide). If germination and growth of the seedlings are performed with the plates tilted backward at a slight angle, this should be minimized. Finally, it is crucial that peptides are applied consistently to the same region of the root tissue, as different responses can be expected in different zones.

Statistical Analyses

All statistical analyses can be performed using Microsoft Excel. A sample size of 10 to 20 seedlings per treatment is generally enough for observing statistically significant differences using an analysis of variance or a Student's *t*-test. Ensure experiments follow a randomized block design by using a minimum of two plates per treatment.

Understanding Results

Synthetic peptides can be potent morphogens and provide a quick method for determining whether a corresponding peptide-encoding gene is involved in a biological process or not. However, interpretations about function should be made with caution. For example, suppression of organogenesis by a peptide might not necessarily indicate a negative role of the gene in that process. As an additional example, IDA (inflorescence deficient in abscission) peptides control emergence of lateral roots by regulating cell wall genes. However, overexpression of IDA peptides using the constitutive 35S promoter in *Arabidopsis* (*35S:AtIDA)* causes restricted root growth and development of seedlings that produce very few lateral roots (Kumpf et al., 2013).

After 10 days in the plate system described in the protocols, *M. truncatula* A17 develops, on average, 5 lateral roots with an average root length of 8 to 9 cm. An average of ~2 to 4 nodules can be detected on *M. truncatula* grown on plates 7 days postinoculation with *Sinorhizobium meliloti* 2011.

In *A. thaliana*, calcium bursts from peptides or other inducers can take multiple forms, with variable rates of induction, amplitude, and duration. These variable characteristics give peptides distinct signatures that can be interpreted differentially by the plant. In our experience we have frequently seen calcium waves travel through the root in the direction of base to tip during the collection of a video. Bursts typically commence within 50 to 60 sec of peptide application; however, to ensure identification of all peptides positive for induction, we recommend collecting videos for 10 min.

Time Considerations

Reannotation of a plant genome for peptide-coding genes can take about 2 to 3 weeks, depending on the plant genome size, informatics resources available, and number of transcript/protein evidence files used.

Identification and annotation of candidate plant SSPs can be done with the online SSP prediction tool from the MtSSP database (*http://mtsspdb.noble.org/prediction/*). This tool allows users to submit proteins on genome scale and predicts if a given protein sequence is an SSP based on multiple criteria within minutes to hours.

Depending on the length and complexity (modifications, number of cysteines, purity, synthesis scale) of peptides, it can take anywhere between 3 and 8 weeks to obtain requisite peptides from competent manufacturers.

From start to finish, Basic Protocol 3 can be completed in 2 weeks. Additional time is required for image analyses and processing the collected data. Since the plate size limits plant growth, peptide treatment on plates beyond 2 weeks has not been tested using this system.

Preparation of plates for the calcium spike assay, as in Basic Protocol 4, requires about an hour. Subsequent collection of videos at the microscope can take 0.5 to 1 hr per peptide to be assayed, depending on the number of replicates to be collected. In a full day at the microscope, about 10 to 12 peptides, plus positive and negative controls, can be assayed.

Acknowledgements

We thank Dr. Elison Blancaflor and J. Alan Sparks for their help in demonstrating the calcium spike assay and sharing GCaMP *A. thaliana* lines and Shulan Zhang for valuable assistance with the peptide screening assay. Financial support was received from the Noble Research Institute, Oklahoma Center for the Advancement of Science and Technology (OCAST grant no. PS18-012) and the National Science Foundation (Division of Integrative Organismal Systems grant no. 1444549). As this work was funded by the NSF, NSF-PAR deposit is required.

Literature Cited

Arsenault, J.-L., Poulcur, S., Messier, C., & Guay, R. (1995). WinRHlZO™, a root-measuring system with a unique overlap correction method. *HortScience*, *30*, 906. doi: 10.21273/HORTSCI.30.4.906D.

Baubet, V., Le Mouellic, H., Campbell, A. K., Lucas-Meunier, E., Fossier, P., & Brúlet, P. (2000). Chimeric green fluorescent protein-aequorin as bioluminescent Ca^{2+} reporters at the single-cell level. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 7260–7265. doi: 10.1073/pnas.97.13.7260.

Breakspear, A., Liu, C., Roy, S., Stacey, N., Rogers, C., Trick, M., ... Murray, J. D. (2014). The root hair "infectome" of *Medicago truncatula* uncovers changes in cell cycle genes and reveals a requirement for auxin signaling in rhizobial infection. *The Plant Cell*, *26*, 4680–4701. doi: 10.1105/tpc.114.133496.

Breiden, M., & Simon, R. (2016). Q&A: How does peptide signaling direct plant development? *BMC Biology*, *14*, 58. doi: 10.1186/s12915-016-0280-3.

Broughton, W. J., & Dilworth, M. J. (1971). Control of leghemoglobin synthesis in snake beans. *Biochemical Journal*, *125*, 1075–1080. doi: 10.1042/bj1251075.

Campbell, M. S., Holt, C., Moore, B., & Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, *48*, 4.11.1–39. doi: 10.1002/0471250953.bi0411s48.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., ... Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*, 188–196. doi: 10.1101/gr.6743907.

Chen, Y. L., Lee, C. Y., Cheng, K. T., Chang, W. H., Huang, R. N., Nam, H. G., & Chen, Y. R. (2014). Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato. *The Plant Cell*, *26*, 4135–4148. doi: 10.1105/tpc.114.131185.

de Bang, T. C., Lundquist, P. K., Dai, X., Boschiero, C., Zhuang, Z., Pant, P., ... Scheible, W.-R. (2017). Genome-wide identification of *Medicago* peptides involved in macronutrient responses and nodulation. *Plant Physiology*, *175*, 1669–1689. doi: 10.1104/pp.17.01096.

Du, C., Li, X., Chen, J., Chen, W., Li, B., Li, C., ... Yu, F. (2016). Receptor kinase complex transmits RALF peptide signal to inhibit root growth in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E8326–E8334. doi: 10.1073/pnas.1609626113.

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*, 1575–1584. doi: 10.1093/nar/30.7.1575.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, *39*, W29–W37. doi: 10.1093/nar/gkr367.

Fletcher, J. C., Brand, U., Running, M. P., Simon, R., & Meyerowitz, E. M. (1999). Signaling of cell fate decisions by CLAVATA3 in *Arabidopsis* shoot meristems. *Science*, *283*, 1911–1914. doi: 10.1126/science.283.5409.1911.

Ghorbani, S., Lin, Y. C., Parizot, B., Fernandez, A., Njo, M. F., Van de Peer, Y., ... Hilson, P. (2015). Expanding the repertoire of secretory peptides controlling root development with comparative genome analysis and functional assays. *Journal of Experimental Botany*, *66*, 5257–5269. doi: 10.1093/jxb/erv346.

Grillet, L., Lan, P., Li, W., Mokkapati, G., & Schmidt, W. (2018). IRON MAN is a ubiquitous family of peptides that control iron transport in plants. *Nature Plants*, *4*, 953–963. doi: 10.1038/s41477-018-0266-y.

Hastwell, A. H., de Bang, T. C., Gresshoff, P. M., & Ferguson, B. J. (2017). CLE peptide-encoding gene families in *Medicago truncatula* and *Lotus japonicus*, compared with those of soybean, common bean and *Arabidopsis*. *Scientific Reports*, *7*, 9384. doi: 10.1038/s41598-017-09296-w.

Haruta, M., Monshausen, G., Gilroy, S., & Sussman, M. R. (2008). A cytoplasmic Ca2+ functional assay for identifying and purifying endogenous cell signaling peptides in *Arabidopsis* seedlings: Identification of AtRALF1 peptide. *Biochemistry*, *47*, 6311–6321. doi: 10.1021/bi8001488.

Hou, S., Wang, X., Chen, D., Yang, X., Wang, M., Turrà, D., ... Zhang, W. (2014). The secreted peptide PIP1 amplifies immunity through receptor-like kinase 7. *PLoS Pathogens*, *10*, e1004331. doi: 10.1371/journal.ppat.1004331.

Huffaker, A., Pearce, G., Veyrat, N., Erb, M., Turlings, T. C., Sartor, R., ... Teal, P. E. (2013). Plant elicitor peptides are conserved signals regulating direct and indirect antiherbivore defense. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 5707–5712. doi: 10.1073/pnas.1214668110.

Imin, N., Mohd-Radzman, N. A., Ogilvie, H. A., & Djordjevic, M. A. (2013). The peptide-encoding CEP1 gene modulates lateral root and nodule numbers in *Medicago truncatula*. *Journal of Experimental Botany*, *64*, 5395–5409. doi: 10.1093/jxb/ert369.

Imin, N., Patel, N., Corcilius, L., Payne, R. J., & Djordjevic, M. A. (2018). CLE peptide tri-arabinosylation and peptide domain sequence composition are essential for SUNN-dependent autoregulation of nodulation in *Medicago truncatula*. *New Phytologist*, *218*, 73–80. doi: 10.1111/nph.15019.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*, 357–360. doi: 10.1038/nmeth.3317.

Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H., & Sawa, S. (2007). Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in Arabidopsis thaliana and Oryza sativa. *Plant and Cell Physiology*, *48*, 1821–1825. doi: 10.1093/pcp/pcm154.

Korf, I. (2004). Gene finding in novel Genomes. *BMC Bioinformatics*, *5*, 59. doi: 10.1186/1471-2105-5-59.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, F. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, *305*, 567–580. doi: 10.1006/jmbi.2000.4315.

Kumpf, R. P., Shi, C. L., Larrieu, A., Stø, I. M., Butenko, M. A., Péret, B., ... Aalen, R. B. (2013). Floral organ abscission peptide IDA and its HAE/HSL2 receptors control cell separation during lateral root emergence. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 5235–5240. doi: 10.1073/pnas.1210835110.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*, 105–132. doi: 10.1016/0022-2836(82)90515-0.

Lease, K. A., & Walker, J. C. (2006). The Arabidopsis unannotated secreted peptide database, a resource for plant peptidomics. *Plant Physiology*, *142*, 831–838. doi: 10.1104/pp.106.086041.

Lee, M. W., Huffaker, A., Crippen, D., Robbins, R. T., & Goggin, F. L. (2018). Plant elicitor peptides promote plant defences against nematodes in soybean. *Molecular Plant Pathology*, *19*(4), 858–869. doi: 10.1111/mpp.12570.

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323. doi: 10.1186/1471-2105-12-323.

Lobet, G., Pagès, L., & Draye, X. (2011). A novel image-analysis toolbox enabling quantitative analysis of root system architecture. *Plant Physiology*, *157*, 29–39. doi: 10.1104/pp.111.179895.

Lobet, G., Pound, M. P., Diener, J., Pradal, C., Draye, X., Godin, C., ... Schnepf, A. (2015). Root system markup language: Toward a unified root architecture description language. *Plant Physiology*, *167*, 617–627. doi: 10.1104/pp.114.253625.

Ma, Y., Zhao, Y., Walker, R. K., & Berkowitz, G. A. (2013). Molecular steps in the immune signaling pathway evoked by plant elicitor peptides: Ca^{2+}-dependent protein kinases, nitric oxide, and reactive oxygen species are downstream from the early Ca^{2+} signal. *Plant Physiology*, *163*, 1459–1471. doi: 10.1104/pp.113.226068.

Matsubayashi, Y., & Sakagami, Y. (1996). Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis* L. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 7623–7627. doi: 10.1073/pnas.93.15.7623.

Matsubayashi, Y., Ogawa, M., Kihara, H., Niwa, M., & Sakagami, Y. (2006). Disruption and overexpression of *Arabidopsis* phytosulfokine receptor gene affects cellular longevity and potential for growth. *Plant Physiology*, *142*, 45–53. doi: 10.1104/pp.106.081109.

Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annual Review of Plant Biology*, *65*, 385–413. doi: 10.1146/annurev-arplant-050312-120122.

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'haeseleer, K., ... Goormachtig, S. (2010). CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiology*, *153*, 222–237. doi: 10.1104/pp.110.153718.

Murphy, E., Smith, S., & De Smet, I. (2012). Small signaling peptides in *Arabidopsis* development: How cells communicate over a short distance. *The Plant Cell*, *24*, 3198–3217. doi: 10.1105/tpc.112.099010.

Nagai, T., Sawano, A., Park, E. S., & Miyawaki, A. (2001). Circularly permuted green fluorescent proteins engineered to sense Ca^{2+}. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 3197–3202. doi: 10.1073/pnas.051636098.

Nakai, J., Ohkura, M., & Imoto, K. (2001). A high signal-to-noise Ca^{2+} probe composed of a single green fluorescent protein. *Nature Biotechnology*, *19*, 137–141. doi: 10.1038/84397.

Nakaminami, K., Okamoto, M., Higuchi-Takeuchi, M., Yoshizumi, T., Yamaguchi, Y., Fukao, Y., ... Shinozaki, K. (2018). AtPep3 is a hormone-like peptide that plays a role in the salinity stress tolerance of plants. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 5810–5815. doi: 10.1073/pnas.1719491115.

Okamoto, S., Suzuki, T., Kawaguchi, M., Higashiyama, T., & Matsubayashi, Y. (2015). A comprehensive strategy for identifying long-distance mobile peptides in xylem sap. *The Plant Journal*, *84*, 611–620. doi: 10.1111/tpj.13015.

Oldroyd, G. E., Engstrom, E. M., & Long, S. R. (2001). Ethylene inhibits the Nod factor signal transduction pathway of *Medicago truncatula*. *The Plant Cell*, *13*, 1835–1849. doi: 10.1105/tpc.13.8.1835.

Olsen, A. N., Mundy, J., & Skriver, K. (2002). Peptomics, identification of novel cationic *Arabidopsis* peptides with conserved sequence motifs. *In Silico Biology*, *2*, 441–451.

Olsson, V., Joos, L., Zhu, S., Gevaert, K., Butenko, M. A., & De Smet, I. (2018). Look closely, the beautiful may be small: Precursor-derived peptides in plants. *Annual Review of Plant Biology*, *70*, 153–186. doi: 10.1146/annurev-arplant-042817-040413.

Pace, J., Lee, N., Naik, N. S., Ganapathysubramanian, B., & Lübberstedt, T. (2014). Analysis of maize (Zea mays L.) seedling roots with the high-throughput image analysis tool ARIA (Automatic Root Image Analysis). *PLoS One*, *9*, e108255. doi: 10.1371/journal.pone.0108255.

Pagny, S., Lerouge, P., Faye, L., & Gomord, V. (1999). Signals and mechanisms for protein retention in the endoplasmic reticulum.

Journal of Experimental Botany, *50*, 157–164. doi: 10.1093/jxb/50.331.157.

Patel, N., Mohd-Radzman, N. A., Corcilius, L., Crossett, B., Connolly, A., Cordwell, S. J., ... Djordjevic, M. A. (2018). Diverse peptide hormones affecting root growth identified in the *Medicago truncatula* secreted peptidome. *Molecular & Cellular Proteomics*, *17*, 160–174. doi: 10.1074/mcp.RA117.000168

Pearce, G., Strydom, D., Johnson, S., & Ryan, C. A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science*, *253*, 895–897. doi: 10.1126/science.253.5022.895.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*, 785–786. doi: 10.1038/nmeth.1701.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*, 290–295. doi: 10.1038/nbt.3122.

Pichon, M., Journet, E. P., Billy, F. D., Dedieu, A., Huguet, T., Truchet, G., & Barker, D. G. (1994). ENOD12 gene expression as a molecular marker for comparing *Rhizobium*-dependent and-independent nodulation in alfalfa. *Molecular Plant-Microbe Interactions*, *7*, 740–747. doi: 10.1094/MPMI-7-0740.

Pound, M. P., French, A. P., Atkinson, J. A., Wells, D. M., Bennett, M. J., & Pridmore, T. (2013). RootNav: Navigating images of complex root architectures. *Plant Physiology*, *162*, 1802–1814. doi: 10.1104/pp.113.221531.

Riaño-Pachón, D. M., Dreyer, I., & Mueller-Roeber, B. (2005). Orphan transcripts in *Arabidopsis thaliana*: Identification of several hundred previously unrecognized genes. *The Plant Journal*, *43*, 205–212. doi: 10.6851111/j.1365-313X.2005.02438.x.

Ropelewski, A. J., Nicholas, H. B. Jr, & Deerfield, D. W. 2nd (2004). Mathematically complete nucleotide and protein sequence searching using Ssearch. *Current Protocols in Bioinformatics*, *4*, 3.10.1–3.10.12. doi: 10.1002/0471250953.bi0310s04

Roy, S., Lundquist, P. K., Udvardi, M. K., & Scheible, W.-R. (2018). Small and mighty: Peptide hormones in plant biology – Teaching guide. *The Plant Cell*. Retrieved from *http://www.plantcell.org/content/30/8/tpc.118.tt0718*. doi: 10.1105/tpc.118.tt0718.

RStudio Team. (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio, Inc. Retrieved from *http://www.rstudio.com*.

Ryan, C. A. (1974). Assay and biochemical properties of the proteinase inhibitor-inducing factor, a wound hormone. *Plant Physiology*, *54*, 328–332. doi: 10.1104/pp.54.3.328.

Schardon, K., Hohl, M., Graff, L., Pfannstiel, J., Schulze, W., Stintzi, A., & Schaller, A. (2016). Precursor processing for plant peptide hormone maturation by subtilisin-like serine proteinases. *Science*, *354*, 1594–1597. doi: 10.1126/science.aai8550.

Scheible, W.-R., Morcuende, R., Czechowski, T., Fritz, C., Osuna, D., Palacios-Rojas, N., ... Stitt, M. (2004). Genome-wide reprogramming of primary and secondary metabolism, protein synthesis, cellular growth processes, and the regulatory infrastructure of Arabidopsis in response to nitrogen. *Plant Physiology*, *136*, 2483–2499. doi: 10.1104/pp.104.047019.

Stenvik, G. E., Tandstad, N. M., Guo, Y., Shi, C. L., Kristiansen, W., Holmgren, A., ... Butenko, M. A. (2008). The EPIP peptide of INFLORESCENCE DEFICIENT IN ABSCISSION is sufficient to induce abscission in *Arabidopsis* through the receptor-like kinases HAESA and HAESA-LIKE2. *The Plant Cell*, *20*, 1805–1817. doi: 10.1105/tpc.108.059139.

Stührwohldt, N., Schardon, K., Stintzi, A., & Schaller, A. (2017). A toolbox for the analysis of peptide signal biogenesis. *Molecular Plant*, *10*, 1023–1025. doi: 10.1016/j.molp.2017.07.005.

Stührwohldt, N., & Schaller, A. (2019). Regulation of plant peptide hormones and growth factors by post-translational modification. *Plant Biology*, *21*, 49–63. doi: 10.1111/plb.12881.

Srivastava, R., Liu, J. X., & Howell, S. H. (2008). Proteolytic processing of a precursor protein for a growth-promoting peptide by a subtilisin serine protease in Arabidopsis. *The Plant Journal*, *56*, 219–227. doi: 10.1111/j.1365-313X.2008.03598.x.

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H., & Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. *Science*, *346*, 343–346. doi: 10.1126/science.1257800.

Takahashi, F., Suzuki, T., Osakabe, Y., Betsuyaku, S., Kondo, Y., Dohmae, N., ... Shinozaki, K. (2018). A small peptide modulates stomatal control via abscisic acid in long-distance signaling. *Nature*, *556*, 235–238. doi: 10.1038/s41586-018-0009-2.

Takeuchi, H., & Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biology*, *10*, e1001449. doi: 10.1371/journal.pbio.1001449.

Wang, C., Yu, H., Zhang, Z., Yu, L., Xu, X., Hong, Z., & Luo, L. (2015). Phytosulfokine is involved in positive regulation of *Lotus japonicus* nodulation. *Molecular Plant-Microbe Interactions*, *28*, 847–855. doi: 10.1094/MPMI-02-15-0032-R.

Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z., Chu, Y. H., Jiang, K., ... Lippman, Z. B. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics*, *47*, 784–792. doi: 10.1038/ng.3309.

Whitford, R., Fernandez, A., Tejos, R., Pérez, A. C., Kleine-Vehn, J., Vanneste, S., ... Hilson, P. (2012). GOLVEN secretory peptides regulate auxin carrier turnover during plant gravitropic responses. *Developmental Cell*, *22*, 678–685. doi: 10.1016/j.devcel.2012.02.002.

Zhang, H., Hu, Z., Lei, C., Zheng, C., Wang, J., Shao, S., ... Shi, K. (2018). A plant phytosulfokine peptide initiates auxin-dependent immunity through cytosolic Ca^{2+} signaling in tomato. *The Plant Cell*, *30*, 652–667. doi: 10.1105/tpc.17.00537.

Zhou, P., Silverstein, K. A., Gao, L., Walton, J. D., Nallu, S., Guhlin, J., & Young, N. D. (2013). Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). *BMC Bioinformatics*, *14*, 335. doi: 10.1186/1471-2105-14-335.

Key References

Cantarel et al. (2008). See above.
Original publication for the MAKER genome annotation pipeline.

de Bang et al. (2017). See above.
Original publication for which most of the described protocols were developed and used.

Ghorbani et al. (2015). See above.
Large-scale identification of small secretory peptide–coding genes in numerous plant species based on comparative genomics analysis, including signal peptide prediction, MCL clustering, and HMM profiles.

Matsubayashi et al. (2014). See above.
Review including historical background on peptide identification, structural characteristics, post-translational modifications, proteolytic processing, and functions of post-translationally modified peptides.

Olsson et al. (2018). See above.
Review on precursor-derived plant peptides, including an overview of the steps involved in producing a biologically active peptide, a discussion of how peptide-receptor binding leads to specific cellular outputs, experimental approaches to match peptide ligands with their receptor, and more.

Petersen et al. (2011). See above.
Description of the SignalP 4.0 server that predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms.

Roy et al. (2018). See above.
Starting point for students, professors, and researchers to quickly learn about plant peptide hormones, consisting of lecture notes, clear and graphical PowerPoint slides, and a teaching guide.

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I., & Cammue, B. P. (2015). The plant peptidome: An expanding repertoire of structural features and biological functions. *The Plant Cell*, *27*, 2095–2118. doi: 10.1105/tpc.15.00440.
Review of major classes of plant peptides, including mechanistic diversity of their biosynthesis, functional diversity, key features, and classification system.

Zhou et al. (2013). See above.
Original publication for SPADA, the Small Peptide Alignment Discovery Application.

Internet Resources

https://hub.docker.com/r/noblebioinfo/sspgene
Docker image hosting all software.

https://github.com/ZhaoBioinformaticsLab/PlantSSPProtocols/
Online document describing Basic Protocols 1 and 2 in detail.

https://bioinfo.noble.org/manuscript-support/ssp-protocol/ssp-demo.tar.gz
Demo data used in the Docker image and online document.

https://ccb.jhu.edu/software/hisat2/index.shtml
HISAT2, a fast and sensitive alignment program for mapping next-generation sequencing reads (both DNA and RNA) to genomes.

http://hmmer.org
HMMER, biosequence analysis using profile hidden Markov models.

http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/Main_Page
MAKER genome annotation analysis tutorial.

http://weatherby.genetics.utah.edu/MAKER/wiki/index.php/The_MAKER_control_files_explained
Details of MAKER control files.

https://micans.org/mcl/
Markov Cluster (MCL) algorithm.

http://mtsspdb.noble.org/
Medicago truncatula Small Signaling Peptide Database (MtSSPdb), hosts large-scale genomics and transcriptomics data for M. truncatula and provides multiple functions to search, analyze, and visualize different datasets. MtSSPdb is the first plant SSP database that integrates gene expression, an SSP prediction online tool, and synthetic peptide induced phenotype information.

https://www.ebi.ac.uk/Tools/msa/muscle/
MUSCLE (MUltiple Sequence Comparison by Log-Expectation) multiple sequence alignment.

http://bioinformatics.psb.ugent.be/webtools/PlantSSP/
Plant Small Secretory Peptides Database (PlantSSPdb), which hosts a collection of small peptides from 32 plant species.

http://lomereiter.github.io/sambamba/
Sambamba for processing of BAM data.

http://www.cbs.dtu.dk/services/SignalP/
SignalP server for signal peptide prediction.

https://github.com/orionzhou/SPADA/wiki/Usage
SPADA profile alignments generation.

https://www.ncbi.nlm.nih.gov/sra
SRA (Sequence Read Archive), the National Institute of Health's primary archive of high-throughput sequencing data.

http://www.biology.wustl.edu/gcg/ssearch.html
SSearch tool (Smith-Waterman search).

https://ccb.jhu.edu/software/stringtie/
StringTie, a fast and highly efficient assembler of RNA-seq alignments into potential transcripts.

http://www.cbs.dtu.dk/services/TMHMM/
TMHMM server for transmembrane helix prediction.