Chapter 24

Peptidomics methods applied to the study of flower development

Raquel Álvarez-Urdiola[a], Eva Borràs[b,c], Federico Valverde[d], José Tomás Matus[a,e], Eduard Sabidó[b,c], and José Luis Riechmann[a,f,*]

[a] Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Edifici CRAG, Campus UAB, 08193 Cerdanyola del Vallès, Barcelona, Spain.

[b] Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology, 08003 Barcelona, Spain

[c] Universitat Pompeu Fabra, 08003 Barcelona, Spain

[d] Institute for Plant Biochemistry and Photosynthesis CSIC – University of Seville, 41092 Seville, Spain.

[e] Institute for Integrative Systems Biology (I2SysBio), Universitat de València-CSIC, Paterna, 46908, Valencia, Spain.

[f] Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

* Corresponding author: joseluis.riechmann@cragenomica.es

i. Running Head

Flower development peptidomics.

ii. Summary/Abstract

Understanding the global and dynamic nature of plant developmental processes requires not only the study of the transcriptome, but also of the proteome, including its largely uncharacterized peptidome fraction. Recent advances in proteomics and high-throughput analyses of translating RNAs (ribosome profiling) have begun to address this issue, evidencing the existence of novel, uncharacterized, and possibly functional peptides. To validate the accumulation in tissues of sORF-encoded polypeptides (SEPs), the basic set-up of proteomic analyses (i.e., LC-MS/MS) can be followed. However, the detection of peptides that are small (up to ~100 aa, 6-7 kDa) and novel (i.e., not annotated in reference databases) presents specific challenges that need to be addressed both experimentally and with computational biology resources. Several methods have been developed in recent years to isolate and identify peptides from plant tissues. In this chapter we outline two different peptide extraction protocols and the subsequent peptide identification by mass spectrometry using the database search or the *de novo* identification methods.

iii. Key Words

Peptidome, ultrafiltration, ammonium sulphate, reverse phase chromatography, C-18, Arabidopsis, mass spectrometry, database.

1. Introduction

Although a variety of peptides have been well documented in both animal and plant genomes, until recently the coding potential of eukaryotic short open reading frames (sORFs) at the genome-wide level had mostly been overlooked. One of the reasons behind this gap is the computational and experimental difficulties for their identification and functional characterization, and particularly for determining whether these sequences are in fact translated. However, it has become clear over the past few years that small peptides (usually defined as shorter than 100 amino acids in length) constitute an important part, largely still uncharacterized, of the eukaryotic proteome ***(1–13)***. Moreover, the massive and widespread transcription of the eukaryotic genome and the pervasive translation of long non-coding RNAs (lncRNAs) habilitate sORFs and the resulting small peptides as raw materials for *de novo* gene origin and evolution ***(14–19)***.

In plants, several peptides have been functionally characterized as key players in diverse signalling pathways of plant development, including flower formation and maturation, in Arabidopsis and other plant species (i.e., ***(20–22)***). Moreover, the presence of novel, uncharacterized Arabidopsis small peptides has been inferred from transcriptome data, in particular ribosome profiling (Poly-Ribo-Seq) experiments ***(23–25)***, leaving the door open for their identification through proteomics and peptidomics approaches. In fact, in studies with human cells and for selected SEPs identified from lncRNAs, primarily by Poly-Ribo-Seq, it was experimentally estimated that SEPs can be present in the cell at concentrations that are within the range of typical cellular proteins ***(26)***, that SEPs can exhibit different and specific subcellular localizations ***(27,28)***, and that they can carry out important biological functions (e.g., ***(29–33)***). Furthermore, in addition to transcriptomics, computational tools have also been used as a source of knowledge on

new potentially coding sORFs, in plants as well as in other eukaryotic organisms and bacteria (e.g., ***(34–36)***).

The sources of peptides that altogether would constitute the peptidome of a plant are several, and include: i) processing from larger functional or non-functional precursors, ii) additional short open reading frames (sORFs) in known protein-coding genes (up- or down-stream the main ORF, in introns, as short splice variants or in a different reading frame from that of the main ORF), and iii) sORFs in long non-coding RNAs (lncRNAs), transcripts of unknown function (TUFs), intergenic regions, junctions, and microRNA precursors ***(37–40)***. For instance, computational analyses suggested that several thousands of novel, potentially coding sORFs could exist in the intergenic regions of the *Arabidopsis* genome ***(35)***. In fact, it was found that when overexpressed, some of those novel sORFs could induce developmental alterations in plant size, leaf number and shape, fertility, or cause lethality, raising the possibility that (many) sORFs with coding potential but that are still uncharacterized in plant genomes might be associated with morphogenesis ***(37)*** and other developmental and physiological processes.

RNA-based methods are a very powerful tool to detect potentially translating sORFs, and the analysis of ribosome profiling data obtained from a variety of eukaryotic organisms provided strong support to the idea that lncRNAs are an important source of new peptides ***(41,42)***. Ribosome profiling has also been used to demonstrate extensive translation of open reading frames, including novel sORFs, in plant species such as *Arabidopsis* ***(23,25,43)***, maize ***(44)*** and tomato ***(45)***. The evaluation of the coding potential of the sequences identified through ribosome profiling is mostly computational but there are mass spectrometry (MS)-based methods able to detect peptides that are

translated from novel sORFs, thereby directly validating the protein-coding potential of the transcripts ***(27,38,44,46–53)***.

In parallel, the improvement of mass spectrometry and data interpretation bioinformatic algorithms have facilitated the analysis of complex protein mixtures. However, the detection of novel plant peptides derived from small ORFs that are not annotated in reference databases presents specific challenges that need to be addressed, both experimentally and with computational resources (**Figure 1**).

The first requirement is an efficient and high-quality extraction from abundant starting material, for which several methods have been developed and optimized. Most basic protocols used for protein extraction from plant tissue are trichloroacetic acid (TCA)-acetone and phenol-based methods. The optimal composition of the extraction buffer depends on the species and tissue of interest ***(54,55)***, but other aspects must be considered, such as heat treatment of the sample to diminish non-specific protease digestions ***(38,56,57)*** or the addition of protease inhibitors to avoid protein degradation ***(38,44,46,49,58,59)*** (**Table 1**). Besides, the processing and degradation of cellular proteins can generate peptidic fragments that increase the complexity of the peptidome sample, deteriorating the signal-to-noise ratio in the experiments. Therefore, strategies to separate larger proteins from peptides prior to LC-MS/MS analyses are crucial to improve the identification and sequence coverage of low-abundance peptides ***(55,60)***. Peptides can be separated and purified using different methods, such as electrophoresis gels ***(27,47,48,61)*** or molecular weight cut-off (MWCO) filters ***(38,52,54,58,59,62)*** (**Table 1**). Moreover, the optimal polypeptide size for detection by LC-MS/MS is approximately 10-20 amino acids, suggesting that trypsin (or trypsin + Lys-C) cleavage is crucial for high-sensitivity SEP detection. Nevertheless, smaller peptides that may be amenable to protease cleavage should be detectable as well ***(26)***.

An additional difficulty lies in undersampling (i.e., identification of only a subset of the peptides) by conventional data acquisition methods ***(63)***. According to a study to optimize a SEP discovery MS workflow using human samples ***(52)***, SEP detection is stochastic due to their size and expression characteristics. Therefore, to avoid undersampling and thus identify more SEPs, it is often more efficient to perform multiple technical and/or biological replicates (multiple runs on the MS platform) than, for example, introduce extensive fractionation methods before LC-MS/MS analyses (as in ***(26)***).

For peptide identification from tandem mass spectra there are two approaches that could be used: database search and *de novo* sequencing. In database search, all potential peptide sequences included in a specified database are retrieved for each spectrum, and each peptide-spectrum match is scored via a scoring function calculated by database search engines (such as SEQUEST ***(64)***, Mascot ***(65)***, Phenyx ***(66)***, X! Tandem ***(67)***, OMSSA ***(68)***, pFind ***(69)***, InsPecT ***(70)***, ByOnic ***(71)***, Comet ***(72)***, MS-GF+ ***(73)***, MaxQuant ***(74)*** or MSTracer ***(75)***). This guided approach is widely used for peptidomics and proteomics, and can be based on canonical (well-annotated) protein databases (e.g., UniProt) or customized databases containing putative SEPs identified by bioinformatic (e.g., sORFinder) ***(76)*** or transcriptomic analyses (i.e., RNA-sequencing or ribosome profiling).

The annotation of the genome of the organism under study is the first source for preparing the database for MS database search (that is, all proteins and peptides that are already known or identified from that genome). However, for the identification of novel SEPs in MS data, it is necessary to design more specific, expanded databases that should also include the potential novel coding sORFs. Current integrated peptidomics pipelines include different database creation strategies, from the use of ribosome

profiling data to the six-frame or three frame identification of sORFs at the genome or transcriptome level, respectively (**Tables 2 and 3**; see also the **Notes** section). For instance, a recent MS-based study identified over 1,000 novel human proteins derived from alternative ORFs identified by RNA-seq (mostly corresponding to SEPs, 57aa median length) ***(27)***. In plants, approximately 70,000 transcribed sORFs were detected in *Physcomitrella patents* (moss) using 'sORF Finder' ***(76)***, from which 828 distinct peptide sequences were identified by LC-MS/MS ***(49)***. Customized peptide databases can also be derived from the six-frame translation of genomic sequences, an approach that has been successfully used in microorganisms ***(62,77)***, humans ***(26,50,51)***, and recently also in both monocot and dicot plants, where a total of 1,993 and 1,860 SEPs were identified in maize and Arabidopsis respectively ***(38)***. Altogether, these and other studies illustrate the existence of a substantial, uncharted fraction of the eukaryotic proteome that is mainly composed of small proteins (peptidome) (**Table 1**).

In contrast to database search, for *de novo* peptide sequencing, peptide sequences are extracted directly from tandem mass spectra using specific algorithms, such as PEAKS ***(78)***, SPIDER ***(79)***, UniNovo ***(80)***, pNovo+ ***(81)***, Novor ***(82)***, DeepNovo ***(83)***, or DeepNovo-DIA ***(84)***. The *de novo* method is less powerful than database search, as many spectra cannot be unambiguously sequenced due to incomplete fragmentation. In addition, the *de novo* method is relatively slow when compared with the database-search engines, and the large search space of all possible amino acid sequences for each spectrum often leads to higher false discovery rates. Moreover, the complexity of tandem mass spectra can be significantly increased when post-translational modifications (PTMs) are considered as well ***(85)***. Some algorithms have been used for solving the *de novo* identification problems involving dynamic programming, integer linear programming, machine learning or other methods, and advances in mass

spectrometry instruments have improved *de novo* sequencing results ***(86)***. However, further optimizations of algorithms, particularly with respect to data confidence, are still necessary to turn the technique into an actual alternative to commonly used database search peptide identification methods. Despite the difficulties, *de novo* identification has been successfully implemented for SEP detection in several plant species ***(48,59,87–89)***. When combined with classic database search strategies, *de novo* approaches can help to provide more comprehensive results ***(59,90,91)*** (**Table 1**). In fact, several research groups have developed software that directly combines both, database search and *de novo* sequencing, for peptide identification from mass spectra ***(71,92)***.

Despite the advances in mass spectrometry and data interpretation, however, a problem still not fully addressed is the (high) number of unassigned spectra. New mass spectrometry sampling methods, such as data-independent acquisition (DIA ***(93)***), together with the development of new machine learning tools to predict peptide fragmentation ***(94–97)*** promise a bright and very exciting future in the peptidome field, in which a significant amount of information will be confidently recovered from the acquired data.

In this chapter, we provide two plant peptide extraction methods based on different extraction buffers and precipitation techniques, and describe an example of a LC-MS/MS pipeline, also introducing some suggestions for database design.

2. Materials

2.1 General

1. Protein low-binding microcentrifuge tubes (1.5 or 2 mL).

2. Mortar and pestle.

3. Liquid nitrogen.

2.2 Ultrafiltration

1. Extraction buffer: 1x phosphate-buffered saline (PBS), 1.5 M urea, 10 mM dithiothreitol (DTT), 2% v/v acetonitrile (ACN), 0.5% v/v trifluoroacetic acid (TFA), 10 μM MG-132 proteasome inhibitor, 1 tablet of Proteinase Inhibitor cocktail cOmplete (Roche) per each 50 mL of buffer, and 1 mM phenylmethylsulfonyl fluoride (PMSF) (*see* **Note 1**). Prepare fresh for each experiment (*see* **Note 2**).

2. 10× Phosphate Buffer Saline (PBS): 1.37 M NaCl, 0.027 M KCl, 80 mM Na_2HPO_4, 20 mM KH_2PO_4 pH 7 (NaOH). Prepare 1 L and autoclave it.

3. Ultra-0.5 mL 30-K centrifugal filter devices (Amicon®) (*see* **Note 3**).

2.3 Ammonium sulphate precipitation

1. Extraction buffer: 1x PBS, 2 M urea, 2% v/v acetonitrile, 10 mM DTT, 5% v/v trifluoroethanol (TFE), 50 mM Tris-HCl pH 7.6, 10 μM MG-132, 1 tablet of Proteinase Inhibitor cocktail cOmplete (Roche) per each 50 mL of buffer, and 1 mM PMSF. Prepare fresh for each experiment (*see* **Note 2**).

2. Ammonium sulphate (salt, EM/HPLC grade).

2.4 Reverse phase chromatography peptide extraction

1. C18 Spin columns, containing 8 mg of resin each (Pierce, Thermo Scientific).

2. Activation solution: 50% v/v ACN in distilled water (400 µL per sample).

3. Equilibration solution: 0.5% v/v TFA in 5% v/v ACN (400 µL per sample).

4. Sample buffer: 2% v/v TFA in 20% v/v ACN (1 µL for every 3 µL of sample) (*see* **Note 4**).

5. Wash solution: 0.5% v/v TFA in 5% v/v ACN (400-800 µL per sample) (*see* **Note 5**).

6. Elution buffer: 0.1% v/v formic acid in 70% v/v ACN (42 µL per sample) (*see* **Note 6**)

7. Qubit Protein Assay Kit.

8. Qubit fluorometer.

2.5 LC-MS/MS

1. DL-Dithiothreitol (DTT) (*see* **Note 7**)

2. Iodoacetamide.

3. Urea.

4. Ammonium bicarbonate.

5. Lysyl Endopeptidase.

6. Trypsin.

7. Formic acid.

8. MicroSpin C18 columns (The Nest Group, Inc)

9. Nano Trap C18 columns with an inner diameter of 100 μm packed with C18 particles of 5 μm particle size (Thermo Fisher Scientific) (Optional, depending on the setup of each laboratory).

10. Reverse-phase chromatography columns (C18, 2 μm, 15-50 cm length) (*see* **Note 8**).

11. Buffer A: 0.1% v/v formic acid in water.

12. Buffer B: 0.1% v/v formic acid in acetonitrile.

13. Bovine serum albumin (New England Biolabs cat # P8108S).

14. Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific) (*see* **Note 9**).

15. EASY-nLC 1000 (Thermo Fisher Scientific).

3. Methods

Below we provide two peptide extraction methods based on different extraction buffers and precipitation techniques (3.1 and 3.2), both of which are to be followed by a reverse phase chromatography (3.3) (**Figure 2**) and describe an example of a LC-MS/MS pipeline (3.4).

3.1 Ultrafiltration

1. Collect tissue of interest with clean material (*see* **Note 10**) and freeze directly in liquid nitrogen. Keep at -80ºC until required.

2. Using a different mortar and pestle for each sample, grind the tissue with liquid nitrogen until obtaining a whitish fine powder (*see* **Note 11**).

3. Collect 0.5 g of blended tissue distributed in two 2 mL Eppendorf tubes (*see* **Note 12**).

4. Add a total of 1.2 mL of extraction buffer to 0.5 g of tissue, vortex immediately and transfer to ice while preparing the rest of the samples (*see* **Note 13**).

5. Incubate the samples with continuous shaking for 1 h at 4ºC.

6. Spin the samples for 1 min at 4ºC in a microcentrifuge (max speed, ≥ 14,000 x g) to precipitate cellular debris and solid particles in suspension. Repeat as many times as necessary (*see* **Note 14**).

7. Insert each Amicon filter device in one of the provided microcentrifuge tubes.

8. Add up to 500 µL of the clean supernatant in the Amicon filter device and centrifuge at 14,000 x g for 10 min at 4ºC as indicated by manufacturer (*see* **Notes 15** and **16**). Repeat until all sample has passed through the same Amicon filter.

9. Keep the filtrate in the provided microcentrifuge tubes (flow-through) (*see* **Note 17**). Keep the samples on ice to immediately continue with the reverse phase chromatography or store them at -80ºC until use.

3.2 Ammonium sulphate precipitation

1. Collect tissue of interest with clean material (*see* **Note 10**) and freeze directly in liquid nitrogen. Keep at -80º C until required.

2. Using a different mortar and pestle for each sample, grind the tissue with liquid nitrogen until obtaining a whitish fine powder (*see* **Note 11**).

3. Collect 0.5 g of blended tissue distributed in two 2 mL Eppendorf tubes (*see* **Note 12**).

4. Add a total of 1.2 mL of extraction buffer to 0.5 g of tissue, vortex immediately and transfer to ice while preparing the rest of the samples (*see* **Note 13**).

5. Incubate the samples shaking for 30 min at 4ºC.

6. Spin the samples for 1 min at 4ºC in a microcentrifuge (max speed, ≥ 14,000 x g) to precipitate cellular debris and solid particles in suspension (*see* **Note 14**).

7. Add 75% (w/v) of ammonium sulphate to the supernatant to precipitate the proteins in solution at 4ºC. The salt must be added little by little pipetting slowly each time until proteins precipitate (*see* **Note 18**).

8. Centrifuge at maximum speed (≥ 14,000 x g) for 25 min at 4ºC.

9. Place the supernatant in a new low binding protein tube (smaller peptides will remain in the supernatant, whereas larger proteins precipitate). Keep the samples on ice to immediately continue with the reverse phase chromatography or store them at -80ºC until use.

3.3 Reverse phase chromatography peptide extraction

Prepare the reverse phase chromatography C18 columns as indicated by the manufacturer protocol. In brief:

Sample preparation:

1. Mix 3:1 parts of sample:sample buffer. The final sample mix will contain approximately 0.5% TFA in 5% ACN (*see* **Note 19**).

Column preparation:

2. Tap the column to settle the resin on the bottom of each column. Remove top and bottom caps (in that order). Place the column into a 2 mL receiver tube.

3. Add 200 µL of Activation Solution to wet the resin. Make sure to rinse the walls of the spin column (*see* **Note 20**).

4. Centrifuge at 1,000 x g for 1 min. Discard the flow-through and repeat steps 3 and 4.

5. Add 200 µL of Equilibration Solution, centrifuge at 1,500 x g for 1 min and discard the flow-through. Repeat this step once.

Sample binding:

6. Place the column into a receiver tube and load up to 150 µL of sample on top of resin bed (*see* **Note 21**).

7. Centrifuge at 1,500 x g for 1 min. Repeat steps 6 and 7 as many times as needed to load all the sample in the same column (*see* **Notes 22** and **23**).

Column Wash:

8. Add 200 µL of Wash Solution to the column and centrifuge at 1,500 x g for 1 min. Repeat this step once (*see* **Note 5**).

Elution:

9. Place the column in a new protein low binding receiver tube and add 21 µL of Elution Buffer to the top of the resin bed.

10. Centrifuge at 1500 x g for 1 min and repeat steps 9 and 10 with the same receiver tube.

11. Quantify the concentration and amount of total protein in each sample using a Qubit Protein Assay Kit: Mix 199 µL of Qubit buffer with 1 µL of Qubit reagent for each sample. Add 2 µL of sample to 198 µL of the reaction mixture, vortex and spin the tube. Incubate at room temperature for 15 min before measuring.

12. Store the samples at -80ºC until further analysis.

3.4 LC-MS/MS

3.4.1 Sample preparation

1. Prepare or dissolve samples in 6 M Urea, 200 mM ammonium bicarbonate.

2. Reduce the samples (10 µg of protein) with 30 nmols of dithiothreitol at 37ºC for 1 h.

3. Alkylate the samples (10 µg of protein) in the dark with 60 nmols of iodoacetamide at 25ºC for 30 min.

4. Dilute the sample extract to 2M urea with 200 mM ammonium bicarbonate for digestion with endoproteinase LysC (1:10 w:v), and incubate at 37ºC, o/n.

5. Dilute 2-fold with 200 mM ammonium bicarbonate for trypsin digestion (1:10 w:w), and incubate at 37ºC, 8h.

6. After digestion, add formic acid (10% v/v of the final volume) to acidify the peptide mix.

7. Desalt the samples with MicroSpin C18 columns prior to LC-MS/MS analysis, following manufacturer´s instructions.

3.4.2 Chromatographic and mass spectrometric analysis

1. Load the peptides onto the analytical column (C18, 2 μm, 15-50 cm length).

2. Separation of the peptides by reverse-phase chromatography with the corresponding columns.

3. Chromatographic gradients start at 93% buffer A and 7% buffer B with a flow rate of 250 nL/min for 5 minutes and gradually increase 65% buffer A and 35% buffer B in 60 min.

4. After each analysis, wash the column for 15 min with 10% buffer A and 90% buffer B.

5. Peptide eluates are dried in a vacuum centrifuge, and resuspended with Buffer A at a final concentration of 1 μg/μL prior to analysis by LC-MS/MS.

6. Operate the mass spectrometer to acquire peptide spectra (*see* **Note 24)**.

3.4.3 Data analysis for database-search peptide identification

1. Search the acquired spectra against the desired peptide database (*see* **Note 25**), plus a list of common contaminants (suggested: ***(98)***), and all the corresponding decoy entries.

2. Set the parameters accordingly to the experimental and mass spectrometric settings and, if appropriate, select variable post-translational modifications to be detected (*see* **Notes 26** and **27**).

3. Determine the peptide abundance estimation ***(99,100)***.

4. Add the information to the appropriate repositories (*see* **Note 28**).

4. Notes

1. Octyl-glucoside, a detergent, could be added (0.1% v/v) to the extraction buffer. The use of detergents is only necessary for the extraction and solubilization of hydrophobic peptides and proteins. However, the presence of detergents in peptide samples decreases chromatographic resolution in LC-MS/MS. Thus, they must be removed prior to MS analysis ***(101)***. As a general rule for MS/MS experiments, keep laboratory wear and high-quality chemicals separated from the rest of the laboratory materials, always use gloves and, if possible, disposable plastic material of the highest quality.

2. Prepare a new extraction buffer on every extraction day as protease-inhibitors could not work properly otherwise. MG-132 is available from several suppliers (we have routinely used MG-132 from Sigma-Aldrich). Proteinase Inhibitor cocktail cOmplete is from Roche. Different extraction buffers have been proposed in the recent years, and their final composition needs to be selected considering the final objective of the study and the type of analytes of interest (e.g., phosphopeptides), because its formulation may affect the final state of the peptides and proteins in the samples (**Table 1**).

3. The Amicon® Ultra-0.5 product line includes 5 different cut-offs depending on its Nominal Molecular Weight Limit (NMWL); 30-K (30 kDa filter) devices are recommended, as peptides would normally be below the 30 kDa cut-off.

4. ACN can be substituted for methanol in all sample preparation buffers, depending on the desired composition of the final elution buffer.

5. The required washing volume will be dependent upon amount and type of contaminants present in the samples. Samples already containing large amounts of urea or >100mM ammonium bicarbonate derived from the extraction buffer (**Table 1**) need to be washed one or two additional times.

6. The elution buffer used can be tailored to the downstream application. Acceptable buffers include 50-70% (v/v) ACN or methanol with or without 0.1% (v/v) TFA. For best results in LC-MS/MS analysis TFA is replaced with 0.1% (v/v) formic acid.

7. Reagents for LC-MS/MS can be obtained from several suppliers. As an example, we list here the specific products we use: Urea (GE Healthcare; Sigma-Aldrich, P/N 17-1319-01); Ammonium bicarbonate (BioUltra, ≥99.5% (T); Sigma-Aldrich, P/N 09830); Iodoacetamide (BioUltra; Sigma-Aldrich, P/N I1149); DL-Dithiothreitol (for electrophoresis, ≥99%; Sigma-Aldrich, P/N D9163); Formic acid for analysis EMSURE® (ACS Reag. Merck, P/N 1.00264.0100); Sequencing Grade Modified Trypsin (Promega, P/N V5111); Lysyl Endopeptidase (Wako Chemicals GmbH, P/N 129-02541).

8. Suitable reverse-phase chromatography columns that we have used are, for instance: 25-cm columns with an inner diameter of 75 μm, packed with 1.9 μm C18 particles (Nikkyo Technos Co.); and 50-cm columns with an inner diameter of 75 μm, packed with 2 μm C18 particles (EASY-Column, Thermo Fisher Scientific, ES903).

9. This is just a concrete example of a “modern high-resolution mass spectrometer”; other instruments could be used.

10. To reduce sample contamination with human proteins (i.e., keratins, collagen, etc.) during sample collection, the use of nitrile gloves and laboratory coats is recommended. Take precaution to avoid hair contamination. If flower organs or tissues are going to be dissected, cool tweezers and any other sampling instrument with liquid nitrogen.

11. Keep samples (before and after grinding) always frozen by pouring liquid nitrogen in the mortar sporadically. Cool collection spatulas before using them to collect homogenized tissue.

12. The extraction yields are around 1 mg of total protein for each 0.5 g of tissue. Peptides might represent about 1% of the total protein, and therefore the expected yield for these extraction methods would be 10-15 µg of peptides. For Arabidopsis inflorescences, a volume of 1 mL of blended tissue in a 2 mL Eppendorf tube is equivalent to approximately 0.5 g of tissue. Dividing the sample in different tubes facilitates its dissolution in the extraction buffer, i.e., using tubes with only 0.25 g (equivalent to 0.5 mL of volume) of blended tissue. After finishing the entire extraction protocol (including the reverse phase chromatography with C18 columns), 10-30 µg of total peptides are obtained when using 30-K filters. The efficiency of the ammonium sulphate precipitation method may be lower (~6 µg of total peptides) (**Figure 2**).

13. If total sample has been divided in two tubes, add approximately 0.6 mL of extraction buffer to each tube (with 0.25 g of blended tissue).

14. After each 1 min spin, transfer the supernatant to a new tube. Be careful to avoid both the pellet and remaining particles in suspension. Repeat the spin in a new tube as many times as needed until supernatant is clear (2 or 3 times should be enough).

15. When the sample has been divided in two tubes, the efficiency of using one single Amicon filter for all sub-samples and the same collection tube is sufficient to achieve a suitable yield.

16. The required centrifugation time may vary according to the NMWL of the columns used. This protocol is defined for 30-K (or upper) devices, yet a higher centrifugation time is necessary for 10-K or 3-K devices (15 and 30 min respectively).

17. The filtrate contains the smallest peptides depending on the weight limit of the filter device. However, if needed, it is possible to recover the concentrated solute by placing the filter device upside down in a clean microcentrifuge tube and centrifuging at 1,000 x g for 2 min at 4ºC. For optimal recovery, it is important to perform the reverse spin immediately after filtrating. Besides, desalting, buffer exchange or diafiltration of this concentrated solute can be accomplished before eluting it by reconstituting the concentrate retained in the column to the original sample volume with the desired solvent and repeating the ultrafiltration process from the beginning to the concentrated solute elution.

18. Ammonium sulphate calculator from EnCor Biotechnology Inc. (http://www.encorbio.com/protocols/AM-SO4.htm) (selecting 4ºC temperature) can be used to calculate the needed amount of ammonium sulphate for each specific sample. The salt addition will increase the sample volume, which should be considered for the reverse phase chromatography. For smallest peptides, ammonium sulphate could be added up to 80-85%.

19. The final exact concentrations of TFA and ACN will vary according to the Extraction Buffer, e.g., ultrafiltration or ammonium sulphate precipitation protocol. In these examples, the concentration of the sample:sample buffer mix prior to reverse

phase chromatography would be 6.5% (v/v) of ACN for both extraction methods, 0.875% (v/v) TFA for ultrafiltration and 0.5% (v/v) TFA for ammonium sulphate precipitation. Nevertheless, these slight variations do not appear to result in significant differences in the efficiency of the reverse phase chromatography process.

20. Add solutions carefully, especially in the activation step. Pour the solution through the walls of the column to avoid producing irregularities in the resin.

21. Each column can bind up to 30 µg of total peptide from 10 to 150 µL sample volumes.

22. In some cases, the extraction yield can be increased by recovering the flow-through and re-centrifuging it after each step.

23. Flow-through may be retained to confirm sample binding.

24. 1 to 2 µg of peptides are loaded onto an analytical column (25-cm C18 2µm particle size) using an autosampler device (e.g., EASY nLC 1000, Thermo Fisher Scientific) and the peptides are then separated by reverse-phase chromatography using a water-acetonitrile chromatographic gradient. Modern high-resolution mass spectrometers are recommended for data acquisition (e.g., Orbitrap or qTOF). The mass spectrometer is operated in data-dependent acquisition (DDA) mode, in which a full MS scan is recorded in each cycle, followed by the fragmentation of the 10-30 most intense precursor ions to obtain the fragment ion spectra.

25. Obtained raw data are analysed using a database search strategy. However, the results are susceptible to the characteristics of the reference database used for peptide identification. It is advisable to add the lists of putative SEPs to a database containing the canonical peptides and proteins of each organism (available in ENSEMBL, Uniprot

or other databases). The total number of sequences included in the database is also important, as an excessively large database (e.g., over 100,000 sequences) may lead to a higher false discovery rate in the identifications. There are several different approaches for the identification of novel potential SEP sequences to be included in the reference database.

One approach that has often been used is to make use of Ribo-seq data. There are multiple repositories which contain sORFs identified by Ribo-seq for plant species, such as GWIPS-viz (*Arabidopsis thaliana* and *Zea mays*) ***(102)***, RPFdb v2.0 (*A. thaliana*) ***(103)***, PsORF (35 plant species) ***(104)***, and uORFdb (*A. thaliana* and others) ***(105)*** (**Table 2**); as well as several tools for the analysis of Ribo-seq data and sORF identification such as RiboTaper ***(106)***, PRICE ***(107)***, RiboCode ***(108)***, RiboStreamR ***(109)***, RiboPlotR ***(110)*** or RiboNT ***(111)*** (**Table 3**).

An alternative (and complementary) approach for the identification of putative novel SEPs is to make use of the genome or lncRNA transcriptome sequences through sORF-prediction tools, such as SPADA ***(112)***, sORF Finder ***(35,76)***, PhyloCSF (Phylogenetic Codon Substitution Frequencies) ***(34)***, MiPepid ***(113)***, CPPred ***(114)***, lncPepid ***(115)***, or DeepCPP ***(116)*** (**Table 3**).

In addition, the putative peptide databases can also be derived from the six-frame translation of the corresponding genome sequence or from the three-frame translation of transcriptomic datasets (e.g., RNA-seq data, lncRNA), an approach referred to as peptidogenomics ***(117)***. It is a strategy that has been successfully implemented in micro-organisms ***(62,77)***, humans ***(26,51,52)***, and plants ***(38)***.

An additional consideration for the generation of the putative SEP database is whether the presence of translation initiation codons in the ORFs (the standard ATG or non-

canonical codons such as CTG or ACG; see ***(118–120)***) is a requirement or not, as both approaches have been used (e.g., ***(35,38)***).

26. Once the database has been constructed, the raw LC-MS/MS data needs to be interpreted using a database search engine (such as SEQUEST ***(64)***, Mascot ***(65)***, Phenyx ***(66)***, X! Tandem ***(67)***, OMSSA ***(68)***, pFind ***(69)***, InsPecT ***(70)***, ByOnic ***(71)***, Comet ***(72)***, MS-GF+ ***(73)***, MaxQuant ***(74)*** or MSTracer ***(75)***). As example, the Mascot search engine (v2.6) can be used, using the search parameters accordingly to the experimental and mass spectrometry settings. For peptide identification a precursor ion mass tolerance below 10-20 ppm is recommended, whereas the fragment ion mass tolerance can go from 10-20 ppm for high-resolution mass analyzers (Orbitrap, TOF) to 0.5 Da if a linear ion trap is used for the analysis of the tandem mass spectra. Common peptide modifications such as oxidation of methionine and N-terminal protein acetylation are used as variable modifications. False discovery rate (FDR) in peptide identification is set to a maximum of 1%.

27. Validation of (selected) identified peptides is highly recommended due to the intrinsic limitations of FDR estimation when working with large databases, although this cannot be done yet in a high-throughput manner. Peptide identifications that pass the FDR threshold can be further validated with the purchase and full LC-MS/MS characterization of synthetic peptides with the same identified sequence (e.g., ***(38)***), and/or by comparison with the fragmentation patterns and retention time predicted by the new machine learning algorithms (e.g., Prosit, MS^{2}PIP, etc.) (94,95,97).

28. Share data and results in a public repository. Data sharing in the public domain is the standard for omics research and a requirement for publication. For proteomics, the Proteomics IDEntifications (PRIDE) database (https://www.ebi.ac.uk/pride/) at the

European Bioinformatics Institute (EMBL-EBI, Hinxton, Cambridge, UK) has enabled public data deposition of MS data since 2004 and its archival component has become the largest repository for proteomics data sharing worldwide ***(121)***. The PRIDE database provides access to most of the experimental proteomics data described in MS-related scientific publications. Moreover, several repositories for sORFs and SEPs in plants have been developed with different purposes and using information from multiple *in silico* and experimental approaches (**Table 2**).

Acknowledgements

Our work on peptidomics was supported by grant BFU2014-58289-P (funded by MICIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe") and by grant 2017SGR718 (from the Agencia de Gestió d'Ajuts Universitaris I de Recerca) to JLR, and by institutional grant SEV-2015-0533 (funded by MCIN/AEI/10.13039/501100011033) and by the CERCA Programme / Generalitat de Catalunya. R.A. is supported by fellowship PRE2018-084278 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". The CRG/UPF Proteomics Unit is part of the Spanish Infrastructure for Omics Technologies (ICTS OmicsTech). We also acknowledge "Secretaria d'Universitats i Recerca del Departament d'Economia i Coneixement de la Generalitat de Catalunya" (2017SGR595) and support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme / Generalitat de Catalunya."

5. References

1. Tavormina P, De Coninck B, Nikonorova N, De Smet I, Cammuea BPA (2015) The plant peptidome: An expanding repertoire of structural features and biological functions. Plant Cell 27(8):2095–2118.

2. Hsu PY, Benfey PN (2018) Small but Mighty: Functional Peptides Encoded by Small ORFs in Plants. Proteomics 18:1700038.

3. Brunet MA, Leblanc S, Roucou X (2020) Reconsidering proteomic diversity with functional investigation of small ORFs and alternative ORFs. Exp Cell Res 393(1):112057.

4. Brunet MA, Levesque SA, Hunting DJ, Cohen AA, Roucou X (2018) Recognition of the polycistronic nature of human genes is critical to understanding the genotype-phenotype relationship. Genome Res 28(5):609–624.

5. Mudge JM, Ruiz-Orera J, Prensner JR, Brunet MA, Calvet F, Jungreis I et al (2022) Standardized annotation of translated open reading frames. Nat Biotechnol 40(7):994–999.

6. Lyapina I, Ivanov V, Fesenko I (2021) Peptidome: chaos or inevitability. Int J Mol Sci. 22:13128.

7. Hellens RP, Brown CM, Chisnall MAW, Waterhouse PM, Macknight RC (2016) The Emerging World of Small ORFs. Trends Plant Sci 21(4):317–328.

8. Takahashi F, Hanada K, Kondo T, Shinozaki K (2019) Hormone-like peptides and small coding genes in plant stress signaling and development. Curr Opin Plant Biol 51:88–95.

9. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. Nat Rev Genet. 2014;15(3):193–204.

10. Couso JP, Patraquim P (2017) Classification and function of small open reading frames. Nat Rev Mol Cell Biol 18(9):575–589.

11. Plaza S, Menschaert G, Payre F (2017) In Search of Lost Small Peptides. Annu Rev Cell Dev Biol 33:391–416.

12. Wright BW, Yi Z, Weissman JS, Chen J (2022) The dark proteome: translation from noncanonical open reading frames. Trends Cell Biol 32(3):243–258.

13. Orr MW, Mao Y, Storz G, Qian SB (2021) Alternative ORFs and small ORFs: Shedding light on the dark proteome. Nucleic Acids Res 48(3):1029–42.

14. Ruiz-Orera J, Hernandez-Rodriguez J, Chiva C, Sabidó E, Kondova I, Bontrop R, et al (2015) Origins of De Novo Genes in Human and Chimpanzee. PLoS Genet 11(12): e1005721.

15. Ruiz-Orera J, Verdaguer-Grau P, Villanueva-Cañas JL, Messeguer X, Albà MM (2018) Translation of neutrally evolving peptides provides a basis for de novo gene evolution. Nat Ecol Evol 2(5):890–896.

16. Ruiz-Orera J, Albà MM (2019) Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. Trends Genet 35(3):186–198.

17. Ruiz-Orera J, Villanueva-Cañas JL, Albà MM (2020) Evolution of new proteins from translated sORFs in long non-coding RNAs. Exp Cell Res 391(1):111940.

18. Blevins WR, Ruiz-Orera J, Messeguer X, Blasco-Moreno B, Villanueva-Cañas

JL, Espinar L et al (2021) Uncovering de novo gene birth in yeast using deep transcriptomics. Nat Commun 12(1):604.

19. Fesenko I, Shabalina SA, Mamaeva A, Knyazev A, Glushkevich A, Lyapina I et al (2021) A vast pool of lineage-specific microproteins encoded by long non-coding RNAs in plants. Nucleic Acids Res 49(18):10328–10346.

20. Goto H, Okuda S, Mizukami A, Mori H, Sasaki N, Kurihara D et al (2011) Chemical visualization of an attractant peptide, LURE. Plant Cell Physiol. 52(1):49–58.

21. Santiago J, Brandt B, Wildhagen M, Hohmann U, Hothorn LA, Butenko MA et al (2016) Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. eLife 5:e15075.

22. Covey PA, Subbaiah CC, Parsons RL, Pearce G, Lay FT, Anderson MA et al (2019) A Pollen-Specific RALF from Tomato That Regulates Pollen Tube Elongation. Plant Phys 153:703-715

23. Hsu PY, Calviello L, Wu HYL, Li FW, Rothfels CJ, Ohler U et al (2016) Super-resolution ribosome profiling reveals unannotated translation events in Arabidopsis. Proc Natl Acad Sci U S A 113(45):E7126–E7135.

24. Juntawong P, Girke T, Bazin J, Bailey-Serres J (2014) Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. Proc Natl Acad Sci U S A 111(1):E203–E212.

25. Bazin J, Baerenfaller K, Gosai SJ, Gregory BD, Crespi M, Bailey-Serres J (2017) Global analysis of ribosome-associated noncoding RNAs unveils new modes of

translational regulation. Proc Natl Acad Sci U S A 114(46):E10018–E10027.

26. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ et al (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. Nat Chem Biol 9(1):59–64.

27. Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, et al (2013) Direct Detection of Alternative Open Reading Frames Translation Products in Human Significantly Expands the Proteome. PLoS One 8(8):e70698

28. Aspden JL, Eyre-Walker YC, Phillips RJ, Amin U, Mumtaz MAS, Brocard M et al (2014) Extensive translation of small open reading frames revealed by poly-ribo-seq. eLife 3:e03528.

29. Huang JZ, Chen M, Chen D, Gao XC, Zhu S, Huang H et al (2017) A Peptide Encoded by a Putative lncRNA HOXB-AS3 Suppresses Colon Cancer Growth. Mol Cell 68(1):171-184.

30. Nelson BR, Makarewich CA, Anderson DM, Winders BR, Troupes CD, Wu F et al (2016) A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle. Science 351:271–275.

31. Makarewich CA, Baskin KK, Munir AZ, Bezprozvannaya S, Sharma G, Khemtong C et al (2018) MOXI Is a Mitochondrial Micropeptide That Enhances Fatty Acid β-Oxidation. Cell Rep 23(13):3701–3709.

32. Prensner JR, Enache OM, Luria V, Krug K, Clauser KR, Dempster JM et al (2021) Noncanonical open reading frames encode functional proteins essential for cancer cell survival. Nat Biotechnol 39(6):697–704.

33. Boix O, Martinez M, Vidal S, Giménez-Alejandre M, Palenzuela L, Lorenzo-Sanz L et al (2022) pTINCR microprotein promotes epithelial differentiation and suppresses tumor growth through CDC42 SUMOylation and activation. Nat Commun 13(1):6840.

34. Lin MF, Jungreis I, Kellis M (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics 27(13):i275–i282.

35. Hanada K, Zhang X, Borevitz JO, Li WH, Shiu SH (2007) A large number of novel coding small open reading frames in the intergenic regions of the Arabidopsis thaliana genome are transcribed and/or under purifying selection. Genome Res 17(5):632–40.

36. Miravet-Verde S, Ferrar T, Espadas-García G, Mazzolini R, Gharrab A, Sabido E et al (2019) Unraveling the hidden universe of small proteins in bacterial genomes. Mol Syst Biol 15(2):e8290.

37. Hanada K, Higuchi-Takeuchi M, Okamoto M, Yoshizumi T, Shimizu M, Nakaminami K et al (2013) Small open reading frames associated with morphogenesis are hidden in plant genomes. Proc Natl Acad Sci U S A 110(6):2395–2400.

38. Wang S, Tian L, Liu H, Li X, Zhang J, Chen X et al (2020) Large-Scale Discovery of Non-conventional Peptides in Maize and Arabidopsis through an Integrated Peptidogenomic Pipeline. Mol Plant 13(7):1078–1093.

39. Hazarika RR, De Coninck B, Yamamoto LR, Martin LR, Cammue BPA, Van Noort V (2017) ARA-PEPs: A repository of putative SORF-encoded peptides in

Arabidopsis thaliana. BMC Bioinformatics 18(1):37.

40. Couzigou J-M, Lauressergues D, Bécard G, Combier J-P, Ecard GB (2015) miRNA-encoded peptides (miPEPs): A new tool to analyze the roles of miRNAs in plant biology. RNA Biol 12:1178–1180.

41. Ruiz-Orera J, Messeguer X, Subirana JA, Alba MM (2014) Long non-coding RNAs as a source of new peptides. eLife 3:e03523.

42. Hartford CCR, Lal A (2020) When Long Noncoding Becomes Protein Coding. Mol Cell Biol 40(6):e00528-19.

43. Kurihara Y, Makita Y, Shimohira H, Fujita T, Iwasaki S, Matsui M (2020) Translational landscape of protein-coding and non-protein-coding RNAs upon light exposure in Arabidopsis. Plant Cell Physiol 61(3):536–545.

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

45. Wu HYL, Song G, Walley JW, Hsu PY (2019) The tomato translational landscape revealed by transcriptome assembly and ribosome profiling. Plant Physiol 181(1):367–380.

46. Mergner J, Frejno M, List M, Papacek M, Chen X, Chaudhary A et al (2020) Mass-spectrometry-based draft of the Arabidopsis proteome. Nature 579:409–414.

47. Wang P, Yao S, Kosami K-I, Guo T, Li J, Zhang Y et al (2020) Identification of endogenous small peptides involved in rice immunity through transcriptomics- and proteomics-based screening. Plant Biotechnology J. 18:415–428.

48. Jorge GL, Balbuena TS (2021) Identification of novel protein-coding sequences in Eucalyptus grandis plants by high-resolution mass spectrometry. Biochim Biophys Acta - Proteins Proteomics 1869:140594.

49. Fesenko I, Kirov I, Kniazev A, Khazigaleeva R, Lazarev V, Kharlampieva D et al (2019) Distinct types of short open reading frames are translated in plant cells. Genome Res 29(9):1464–1477.

50. Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet F et al (2021) Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat Biotechnol 40:209–217.

51. Chen J, Brunner AD, Cogan JZ, Nuñez JK, Fields AP, Adamson B et al (2020) Pervasive functional translation of noncanonical human open reading frames. Science 367:140–146.

52. Ma J, Ward CC, Jungreis I, Slavoff SA, Schwaid AG, Neveu J et al (2014) Discovery of human sORF-encoded polypeptides (SEPs) in cell lines and tissue. J Proteome Res 13(3):1757–1765.

53. Flower CT, Chen L, Jung HJ, Raghuram V, Knepper MA, Yang CR (2020) Genetic and Genomics Investigation of Structure and Function of the Kidney: An integrative proteogenomics approach reveals peptides encoded by annotated lincRNA in the mouse kidney inner medulla. Physiol Genomics 52(10):485.

54. Luo W, Xiao Y, Liang Q, Su Y, Xiao L (2019) Identification of potential auxin-responsive small signaling peptides through a peptidomics approach in arabidopsis thaliana. Molecules 24:3146.

55. Barashkova AS, Rogozhin EA (2020) Isolation of antimicrobial peptides from different plant sources: Does a general extraction method exist? Plant Methods 16:143.

56. Damerval C, De Vienne D, Zivy M, Thiellement H (1986) Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis 7(1):52–54.

57. Chatterjee M, Gupta S, Bhar A, Das S (2012) Optimization of an Efficient Protein Extraction Protocol Compatible with Two-Dimensional Electrophoresis and Mass Spectrometry from Recalcitrant Phenolic Rich Roots of Chickpea (Cicer arietinum L.). Int J Proteomics 2012:536963.

58. Shi Y, Li J, Li L, Lin G, Bilal AM, Smagghe G et al (2021) Genomics, transcriptomics, and peptidomics of Spodoptera frugiperda (Lepidoptera, Noctuidae) neuropeptides. Arch Insect Biochem Physiol 106:e21740

59. Culver KD, Allen JL, Shaw LN, Hicks LM (2021) Too Hot to Handle: Antibacterial Peptides Identified in Ghost Pepper. J Nat Prod 84:2200-2208.

60. Kuljanin M, Dieters-Castator DZ, Hess DA, Postovit L-M, Lajoie GA (2017) Comparison of sample preparation techniques for large-scale proteomics. Proteomics 17(1–2):1600337.

61. Flower CT, Chen L, Jung HJ, Raghuram V, Knepper MA, Yang C-R (2020) An integrative proteogenomics approach reveals peptides encoded by annotated lincRNA in the mouse kidney inner medulla. Physiol Genomics 52:485–491.

62. Cao S, Liu X, Huang Y, Yan Y, Zhou C, Shao C et al (2021) Proteogenomic

discovery of sORF-encoded peptides associated with bacterial virulence in Yersinia pestis. Commun Biol 4:1248.

63. Grossmann J, Roschitzki B, Panse C, Fortes C, Barkow-Oesterreicher S, Rutishauser D et al (2010) Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 73(9):1740–1746.

64. Eng JK, Mccormack AL, Yates JR (1994) An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. Am Soc Mass Spectrom 5:977–989.

65. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probablity-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3557.

66. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J (2003) OLAV: towards high-throughput tandem mass spectrometry data identification. Proteomics 3(8):1454–1463.

67. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra.cBioinformatics 20(9):1466–1467.

68. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM et al (2004) Open Mass Spectrometry Search Algorithm. J Proteome Res 3:958–964.

69. Fu Y, Yang Q, Sun R, Li D, Zeng R, Ling CX et al (2004) Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. Bioinformatics 20(12):1948–1954.

70. Tanner S, Shu H, Frank A, Wang LC, Zandi E, Mumby M et al (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. Anal Chem 77(14):4626–4639.

71. Bern M, Cai Y, Goldberg D (2007) Lookup Peaks: A Hybrid of de Novo Sequencing and Database Search for Protein Identification by Tandem Mass Spectrometry. Anal Chem 79(4):1393–1400.

72. Eng JK, Jahan TA, Hoopmann MR (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics 13(1):22–24.

73. Kim S, Pevzner PA (2014) MS-GF+ makes progress towards a universal database search tool for proteomics. Nat Commun 5(1):5277.

74. Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11(12):2301–2319.

75. Zeng X, Ma B (2021) MSTracer: A Machine Learning Software Tool for Peptide Feature Detection from Liquid Chromatography-Mass Spectrometry Data. J Proteome Res 20(7):3455–3462.

76. Hanada K, Akiyama K, Sakurai T, Toyoda T, Shinozaki K, Shiu S-H (2010) sORF finder: a program package to identify small open reading frames with high coding potential. Bioinformatics 26(3):399–400.

77. Yang X, Jensen SI, Wulff T, Harrison SJ, Long KS (2016) Identification and validation of novel small proteins in Pseudomonas putida. Environ Microbiol Rep 8(6):966–674.

78. Ma B, Zhang K, Hendrie C, Liang C, Li M, Doherty-Kirby A et al (2003)

PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Communnications Mass Spectrom 17:2337–2342.

79. Han Y, Ma B, Zhang K (2005) Spider: Software for protein identification from sequence tags with de novo sequencing error. J Bioinform Comput Biol 3(3):697–716.

80. Jeong K, Kim S, Pevzner PA (2013) UniNovo: a universal tool for de novo peptide sequencing. Bioinformatics 29(16):1953–1962.

81. Chi H, Chen H, He K, Wu L, Yang B, Sun R-X et al (2013) pNovo+: De Novo Peptide Sequencing Using Complementary HCD and ETD Tandem Mass Spectra. J Proteome Res 12:615–25.

82. Ma B (2015) Novor: Real-Time Peptide de Novo Sequencing Software. J Am Soc Mass Spectrom 26:1885–1894.

83. Tran NH, Zhang X, Xin L, Shan B, Li M (2017) De novo peptide sequencing by deep learning. Proc Natl Acad Sci U S A 114(31):8247–8252.

84. Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X et al (2019) Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. Nat Methods 16(1):63–66.

85. Pathan M, Samuel M, Keerthikumar S, Mathivanan S (2017) Unassigned MS/MS Spectra: Who Am I? In: Keerthikumar, S., Mathivanan, S. (eds) Proteome Bioinformatics. Methods in Molecular Biology, vol 1549. Humana Press, New York, NY. p 67–74.

86. Muth T, Renard BY (2018) Evaluating de novo sequencing in proteomics:

already an accurate alternative to database-driven peptide identification? Brief Bioinform 19(5):954–970.

87. Wu H, Johnson MC, Lu CH, Fritsche KL, Thomas AL, Lai Y et al (2015) Peptidomics study of anthocyanin-rich juice of elderberry. Talanta 131:640–644.

88. Gemperline E, Keller C, Jayaraman D, Maeda J, Sussman MR, Ané, J-MA et al (2016) Examination of Endogenous Peptides in Medicago truncatula Using Mass Spectrometry Imaging. J Proteome Res 15:4403–4411.

89. Gemperline E, Keller C, Li L (2016) Mass Spectrometry in Plant-omics. Anal Chem 88(7):3422–3434.

90. Ye X, Zhao N, Yu X, Han X, Gao H, Zhang X (2016) Extensive characterization of peptides from Panax ginseng C. A. Meyer using mass spectrometric approach. Proteomics 16:2788–2791.

91. Zhang K, Mckinlay C, Hocart CH, Djordjevic MA (2006) The Medicago truncatula Small Protein Proteome and Peptidome. J Proteome Res (12):3355–3367.

92. Wang X, Li Y, Wu Z, Wang H, Tan H, Peng J (2014) JUMP: A Tag-based Database Search Tool for Peptide Identification with High Sensitivity and Accuracy. Mol Cell Proteomics 13(12):3663–3673.

93. Röst HL, Rosenberger G, Navarro P, Gillet L, Miladinović SM, Schubert OT et al (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol 32:219–223.

94. Wilhelm M, Zolg DP, Graber M, Gessulat S, Schmidt T, Schnatbaum K et al

(2021) Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. Nat Commun 12:3346.

95. Gessulat S, Schmidt T, Zolg DP, Samaras P, Schnatbaum K, Zerweck J et al (2019) Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. Nat Methods 16:509–518.

96. Ekvall M, Truong P, Gabriel W, Wilhelm M, Käll L (2022) Prosit Transformer: A transformer for Prediction of MS2 Spectrum Intensities. J Proteome Res 21(5):1359–1364.

97. Gabriels R, Martens L, Degroeve S (2019) Updated MS^2PIP web server delivers fast and accurate MS^2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques. Nucleic Acids Res 47(W1):W295–W299.

98. Beer LA, Liu P, Ky B, Barnhart KT, Speicher DW (2017) Efficient quantitative comparisons of plasma proteomes using label-free analysis with MaxQuant. Methods Mol Biol 1619:339–352.

99. Gerster S, Kwon T, Ludwig C, Matondo M, Vogel C, Marcotte EM et al (2014) Statistical approach to protein quantification. Mol Cell Proteomics 13(2):666–77.

100. Fabre B, Lambour T, Bouyssié D, Menneteau T, Monsarrat B, Burlet-Schiltz O et al (2014) Comparison of label-free quantification methods for the determination of protein complexes subunits stoichiometry. EuPA Open Proteomics 4:82–86.

101. Yeung YG, Stanley ER (2010) Rapid Detergent Removal From Peptide Samples With Ethyl Acetate For Mass Spectrometry Analysis. Curr Protoc Protein Sci

16:16.12

102. Michel AM, Fox G, M. Kiran A, De Bo C, O'Connor PBF, Heaphy SM et al (2014) GWIPS-viz: development of a ribo-seq genome browser. Nucleic Acids Res 42:D859-D864.

103. Wang H, Yang L, Wang Y, Chen L, Li H, Xie Z (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. Nucleic Acids Res 47:D230-D234.

104. Chen Y, Li D, Fan W, Zheng X, Zhou Y, Ye H et al (2020) PsORF: a database of small ORFs in plants. Plant Biotechnol J 18:2158–2160.

105. Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A (2014) uORFdb--a comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res 42:D60-D67.

106. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M et al (2016) Detecting actively translated open reading frames in ribosome profiling data. Nat Methods 13(2):165–170.

107. Erhard F, Halenius A, Zimmermann C, L'Hernault A, Kowalewski DJ, Weekes MP et al (2018) Improved Ribo-seq enables identification of cryptic translation events. Nat Methods 15(5):363–366.

108. Xiao Z, Huang R, Xing X, Chen Y, Deng H, Yang X (2018) De novo annotation and characterization of the translatome with ribosome profiling data. Nucleic Acids Res 46(10):e61.

109. Perkins P, Mazzoni-Putman S, Stepanova A, Alonso J, Heber S (2019)

RiboStreamR: a web application for quality control, analysis, and visualization of Ribo-seq data. BMC Genomics 20:422.

110. Larry Wu H-Y, Yingshan Hsu P (2021) RiboPlotR: a visualization tool for periodic Ribo-seq reads. Plant Methods 17:124.

111. Song B, Jiang M, Gao L (2021) RiboNT: A Noise-Tolerant Predictor of Open Reading Frames from Ribosome-Protected Footprints. Life (Basel) 11(7):701.

112. Zhou P, Silverstein KAT, Gao L, Walton JD, Nallu S, Guhlin J et al (2013) Detecting small plant peptides using SPADA (Small Peptide Alignment Discovery Application). BMC Bioinformatics 14(1):335.

113. Zhu M, Gribskov M (2019) MiPepid: MicroPeptide identification tool using machine learning. BMC Bioinformatics 20(1):559

114. Tong X, Hong X, Xie J, Liu S (2020) CPPred-sORF: Coding Potential Prediction of sORF based on non-AUG. bioRxiv doi.org/10.1101/2020.03.31.017525.

115. Zhao S, Meng J, Luan Y (2022) LncRNA-Encoded Short Peptides Identification Using Feature Subset Recombination and Ensemble Learning. Interdiscip Sci Comput Life Sci 14(1):101–112.

116. Zhang Y, Jia C, Fullwood MJ, Kwoh CK (2021). DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. Brief Bioinform 22(2):2073–2084.

117. Kersten RD, Yang Y, Xu Y, Cimermancic P, Nam S-J, Fenical W et al (2011) A mass spectrometry-guided genome mining approach for natural product

peptidogenomics. Nat Chem Biol 7(11):794–802.

118. Cao X, Slavoff SA (2020) Non-AUG start codons: Expanding and regulating the small and alternative ORFeome. Exp Cell Res 391(1):111973.

119. Na CH, Barbhuiya MA, Kim MS, Verbruggen S, Eacker SM, Pletnikova O et al (2018) Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. Genome Res 28(1):25–36.

120. Li YR, Liu MJ. Prevalence of alternative AUG and non-AUG translation initiators and their regulatory effects across plants. Genome Res 30(10):1418–1433.

121. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S et al (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res 50(D1):D543–552.

122. Patel N, Mohd-Radzman NA, Corcilius L, Crossett B, Connolly A, Cordwell SJ et al (2018) Diverse peptide hormones affecting root growth identified in the Medicago truncatula secreted peptidome. Mol Cell Proteomics 17(1):160–174.

123. Chen YL, Lee CY, Cheng KT, Chang WH, Huang RN, Nam HG et al (2014) Quantitative peptidomics study reveals that a wound-induced peptide from PR-1 regulates immune signaling in tomato. Plant Cell 26(10):4135–4148.

124. Das D, Jaiswal M, Khan FN, Ahamad S, Kumar S (2020) PlantPepDB: A manually curated plant peptide database. Sci Rep 10(1):2194.

125. Szcześniak MW, Bryzghalov O, Ciomborowska-Basheer J, Makałowska I (2019)

CANTATAdb 2.0: Expanding the Collection of Plant Long Noncoding RNAs. In: Chekanova, JA, Wang, HLV (eds) Plant Long Non-Coding RNAs. Methods in Molecular Biology, vol 1933. Humana Press, New York, NY. p. 415–429.

126. Singh A, Vivek AT, Kumar S (2021) AlnC: An extensive database of long non-coding RNAs in angiosperms. PLoS One 16(4):e0247215.

127. Niu R, Zhou Y, Zhang Y, Mou R, Tang Z, Wang Z et al (2020) uORFlight: a vehicle toward uORF-mediated translational regulation mechanisms in eukaryotes. Database 2020:baaa007.

128. Niarchou A, Alexandridou A, Athanasiadis E, Spyrou G (2013) C-PAmP: large scale analysis and database construction containing high scoring computationally predicted antimicrobial peptides for all the available plant species. PLoS One 8(11):e79728.

129. Wang J, Yin T, Xiao X, He D, Xue Z, Jiang X et al (2018) StraPep: a structure database of bioactive peptides. Database 2018:bay038.

130. Shi G, Kang X, Dong F, Liu Y, Zhu N, Hu Y et al (2022) DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. Nucleic Acids Res 50(D1):D488–D496.

131. Boschiero C, Dai X, Lundquist PK, Roy S, de Bang TC, Zhang S et al (2020) MtSSPdb: The Medicago truncatula Small Secreted Peptide Database. Plant Physiol 183(1):399–413.

132. Lin X, Lin W, Ku YS, Wong FL, Li MW, Lam HM et al (2020) Analysis of Soybean Long Non-Coding RNAs Reveals a Subset of Small Peptide-Coding

Transcripts. Plant Physiol 182(3):1359–1374.

FIGURES

Figure 1. Workflow for peptide discovery and characterization based on mass spectrometry. Extraction method, and MS data analysis.

Figure 2. Schematic representation of the two peptide extraction methods described in this chapter. AS, Activation Solution; ES, Equilibration Solution; WS, Wash Solution; EB, Elution Buffer.

TABLES

Table 1. Overview of peptide extraction methods applied in LC-MS/MS studies in different plant species in recent years.

Table 2. Repositories for SEP-database generation.

Table 3. Tools for database design.

Table 1

Species	Sample type	Highlights of the extraction method characteristics	Identification Workflow	Identified peptides	Ref.
A. thaliana	Col-0 leaves (four-leaf stage)	- Heat treatment of the sample to diminish non-specific protease digestions - Trichloroacetic acid (TCA) precipitation - 10 kDa MWCO filter	1. Database: 6-frame translation of the complete *A. thaliana* genome 2. LC-MS/MS (DDA, Mascot)	1,860 novel SEPs	(38)
A. thaliana	Leaf tissue and leaf protoplast	- SDS-PAGE (10% gel) - 10 kDa MWCO filter	1. Canonical database (Araport11) 2. LC-MS/MS (DDA, ProteinLynx, ProteinPilot)	127 protein-derived auxin-responsive peptides	(54)
Medicago truncatula	Root cultures, xylem sap	- *o*-chlorophenol/acetone precipitation - Size exclusion chromatography	nano-LC-ESI-MS/MS (DDA, Proteome Discoverer, SEQUEST) matching spectra against three databases: i) 225 sequences from known members of different peptide families in *M. truncatula* identified using BLAST; ii) *M. truncatula* whole protein database; iii) ~940 C-Terminally Encoded Peptide (CEP) sequences from different species	12 peptide hormones	(122)
Z. mays	Inbred line B73 seeds	- Phenol extraction and ammonium sulphate precipitation	1. Identification of sORFs based on Ribo-seq data 2. LC-MS/MS (DIA, MaxQuant)	2,695 small peptides (up to 100aa)	(44)
Z. mays	Maize inbred line B73 leaves (three-leaf stage)	- Heat treatment of the sample to diminish non-specific protease digestions - Trichloroacetic acid (TCA) precipitation - 10kDa MWCO filter	1. Database: 6-frame translation of the complete *Z. mays* genome 2. LC-MS/MS (DDA, Mascot)	1,993 novel SEPs	(38)
Oryza sativa	Leaves of *O. sativa* L. ssp. *japonica* cv. Nipponbare and suspension cells derived from Nipponbare calli	- Anion-exchange chromatography and acetone precipitation - SDS-PAGE using (16.5% gel)	1. Database: rice genome (MSU7) considering and considering 3 different possible PTMs 2. LC-MS/MS (DDA, Mascot)	236 annotated and 52 unannotated novel small secreted proteins (SSPs).	(47)
Solanum lycopersicum	Tomato leaves	- Recovery of analytes (peptides) using Sep-Pak C18 Cartridges	1. Tomato hypothetical peptide database (TomHT database) and randomized databases (Ran Databases) 2. LC-MS/MS (DDA, Mascot)	46 unique peptides derived from 25 pre-proteins	(123)

Capsicum chinense x frutescens	Aerial tissue (~12-week-old plants)	- 30 kDa MWCO filter	1. Database generated using antimicrobial peptides (AMPs) prediction algorithms + Canonical *C- chinense* reference proteome 2. LC-MS/MS (DDA, Mascot)	14 AMPs	(59)
Panax ginseng	Ginseng radix (dry root)	- Methanol based extraction.	1. nano-LC-MS/MS (DIA, PEAKS) with canonical database search (Swissprot) 2. *De novo* nano-LC-MS/MS (DIA, SPIDER) peptide identification	308 peptides	(90)
Eucalyptus grandis	Plant stem material	- Organic solvent precipitation - SDS-PAGE	1. Database: 3-frame translation (Virtual Ribosome v.2.0) of mRNA, ncRNA and transcribed RNA sequences publicly available combined with *E. grandis* protein database 2. Database guided LC-MS/MS (DDA, PEAKS) + Novel peptide mapping and genomic classification (BLAST)	41 novel peptides	(48)
Physcomitrella patens	Protonemata	- Gel filtration	1. sORF database generated using sORFfinder at genome and transcriptome level 2. LC-MS/MS (DDA, MaxQuant)	828 peptide sequences.	(49)

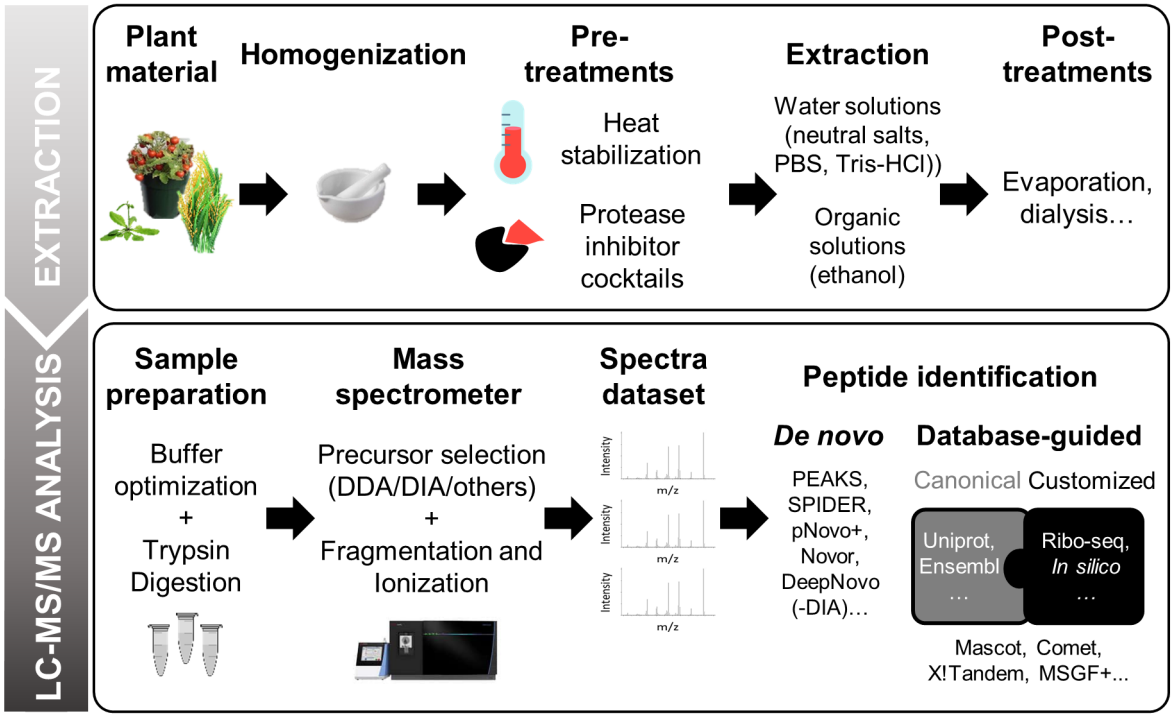
Table 2

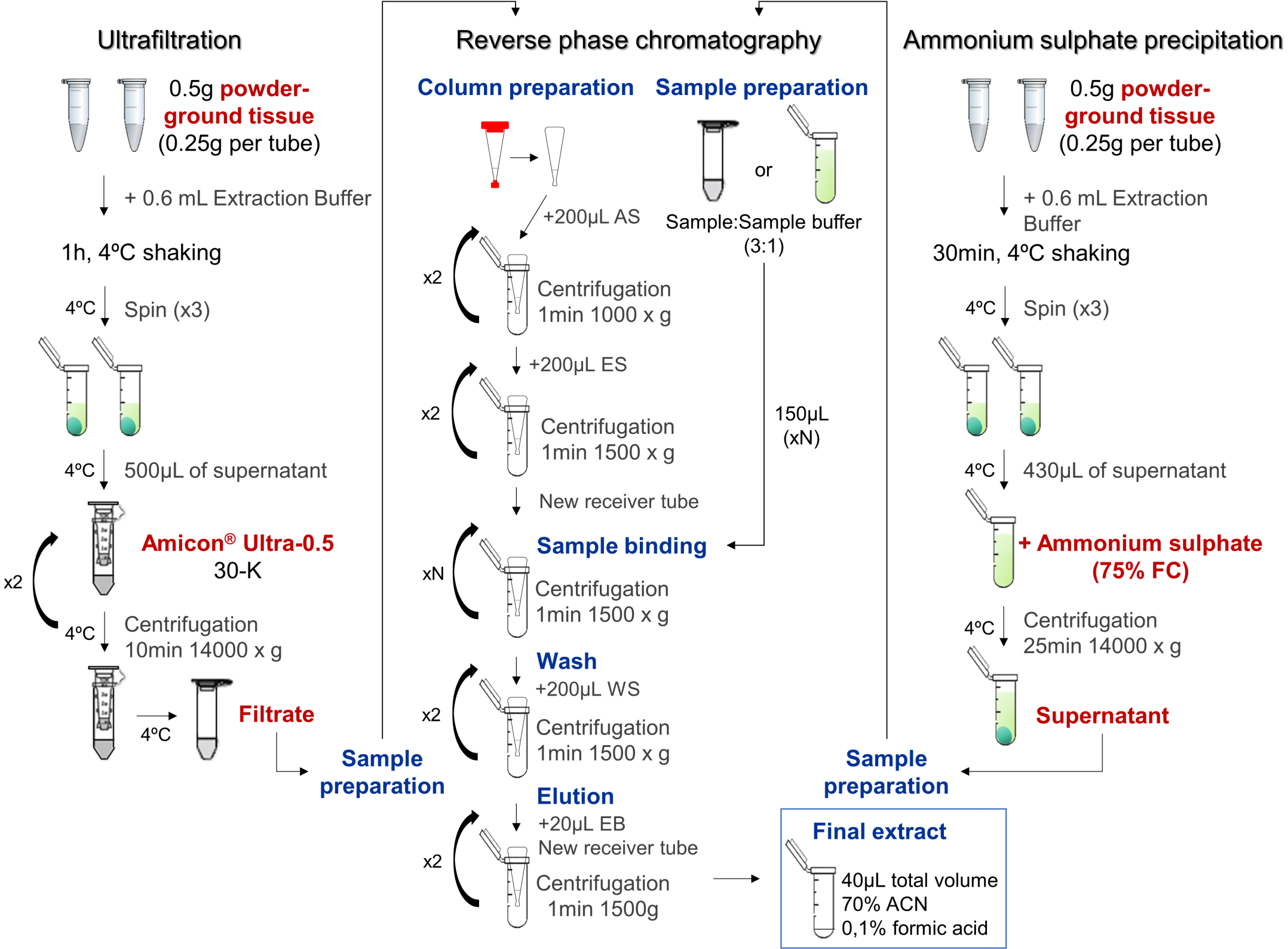
Database	Description	Collected data	Organism	Ref.
ARA-PEP	Putative peptides encoded by sORFs in the *A. thaliana* genome	Tiling arrays, RNA-seq data and other publicly available datasets	*A. thaliana*	(39)
PsORF	sORFs across different plant species	Genomic, transcriptomic, ribo-seq and MS data	35 plant species	(104)
PlantPepDB	Manually curated database of plant-derived peptides	Experimentally validated peptides, peptides with evidence at transcript level, based on computational predictions or inferred by homology	Several plant species including algae, bryophyte, angiosperms, and gymnosperms	(124)
RPFdb v2.0	Genome-wide information of translated mRNA	Ribo-seq samples	Plants: *A. thaliana* Others: 28 different species	(103)
CANTATAdb 2.0	lncRNA data from plant and algae	lncRNA identified computationally using publicly available RNA-seq data	39 plant species (including 3 algae)	(125)
AlnC	Angiosperm lncRNA Catalogue	lncRNA in angiosperms (1KP transcriptome data)	682 angiosperm plant species (809 tissues)	(126)
GWIPS-viz	Online visualization tool for ribo-seq data	Ribo-seq samples	Plants: *A.thaliana, Z. mays* Others: bacteria, animals, etc.	(102)
uORFlight	Database for the evaluation of uORF frequency among different accessions.	uORF identified in genome and transcriptome annotations	Plants: *A. thaliana*, *O. sativa, B. napus, G. max, G. raimondii, M. truncatula, S. lycopersicum, S. tuberosum, T. aestivum*, *Z. mays* Others: fungus, metazoan and vertebrate	(127)
uORFdb	Comprehensive literature database on eukaryotic uORFs.	uORF-related references; manually curated from all uORF-related literature listed at the PubMed database	Plants: *A. thaliana* Others: human, mouse, rat, virus, yeast, etc.	(105)
C-PAmP	Computationally predicted plant antimicrobial peptides	Selection of peptides included in the Antimicrobial Peptide Database (APD) and the Collection of Anti-Microbial Peptides (CAMP)	2,112 plant species in UniProtKB/Swiss-Prot	(128)
StraPep	Structure database of bioactive peptides.	Structural data collected from UniProtKB and PDB	452 different species including bacteria, yeast, animals, humans, and plants	(129)
DRAMP 3.0	Manually curated Data Repository of Antimicrobial Peptides	Peptides retrieved from Pubmed, Swiss-prot and Lens	Variety of organisms, including bacteria, archaea, protists, fungi, animals, and plants	(130)

Table 3

Tool	Description	Method	Example	Ref.
SPADA	Small peptide alignment discovery application. Free software tool that identifies and predicts the gene structure for short peptides with one or two exons	Sequence similarity	Creation of an *M. truncatula* small secreted peptide database (MtSSPdb) using SPADA and sORF Finder (131)	(112)
sORF Finder	Program package for the identification of sORFs with high coding potential	Codon pattern, codon substitution and cross-species conservation	51 new sORFs identified using sORF finder and the ARA-PEP repository (LC-MS/MS results) (46)	(35,76)
PhyloCSF	Phylogenetic Codon Substitution Frequencies: method to determine whether a multi-species nucleotide sequence alignment is likely to represent a protein-coding region	Codon pattern, codon substitution and cross-species conservation	Identification of small peptide-coding "long non-coding" RNAs in soybean (132)	(34)
MiPepid	RNA-seq sORF annotation in mammalian species	Machine learning	Identification of 82 novel species-specific translated sORFs (LC-MS/MS) from lncRNA (database generated using MiPepid) (19)	(113)
lncPepid	RNA-seq sORF annotation in plants. A discovery tool trained using maize and Arabidopsis data that considers sequence composition and physicochemical properties.	Machine learning		(115)
CPPred-sORF	Predicts the coding potential of sORFs based on non-AUG initiation of translation	Machine learning	sORF finder, miPepid, CPPred and DeepCPP used as control groups (115)	(114)
DeepCPP	Optimization of CPPred.	Deep learning	sORF finder, miPepid, CPPred and DeepCPP used as control groups (115)	(116)
RiboTaper	Statistical approach that identifies translated regions based on the characteristic three-nucleotide periodicity of ribo-seq data	Ribo-seq	Identification of uORFs, dORFs and altORFs in *A. thaliana* (23)	(106)
PRICE	Computational method that models experimental noise to resolve overlapping sORFs and non-canonical translation initiation in an accurate manner	Ribo-seq	Validation of the method using major histocompatibility complex class I (MHC I) peptidomics (107)	(107)
RiboCode	Unbiased method to recover the signal of active translation from the ribo-seq data	Ribo-seq	Identification of 9,388 sORF encoding peptides (2-100aa) in maize, from which 2,695 SEPs were verified by MS data (44)	(108)
RiboStreamR	Quality control platform for Ribo-seq data in the form of an R shine web application	Ribo-seq		(109)

RiboPlotR	Visualization package written in R. Representation of RNA-seq coverage and Ribo-seq reads in genomic coordinates for all annotated transcript isoforms of a gene	Ribo-seq	RiboPlotR combines transcriptome annotation files, standard RNA-seq bam files and Ribo-seq P-site position/count files to plot RNA-seq and Ribo-seq data with genomic coordinates for each isoform. Tested in Arabidopsis and tomato (110)	(110)
RiboNT	Noise tolerant sORF predictor that can use RPFs with poor periodicity	Ribo-seq	Identification of sORFs in Arabidopsis seedlings which are evolutionary conserved in diverse plant species (111)	(111)


EXTRACTION
Plant material
Homogenization
Pre-treatments
Heat stabilization
Protease inhibitor cocktails
Extraction
Water solutions (neutral salts, PBS, Tris-HCl))
Organic solutions (ethanol)
Post-treatments
Evaporation, dialysis…
LC-MS/MS ANALYSIS
Sample preparation
Buffer optimization + Trypsin Digestion
Mass spectrometer
Precursor selection (DDA/DIA/others) + Fragmentation and Ionization
Spectra dataset
Intensity
m/z
Peptide identification
De novo
PEAKS, SPIDER, pNovo+, Novor, DeepNovo (-DIA)…
Database-guided
Canonical Customized
Uniprot, Ensembl …
Ribo-seq, In silico …
Mascot, Comet, X!Tandem, MSGF+...


Ultrafiltration
0.5g powder-ground tissue (0.25g per tube)
+ 0.6 mL Extraction Buffer
1h, 4ºC shaking
4ºC
Spin (x3)
4ºC
500µL of supernatant
x2
Amicon® Ultra-0.5 30-K
4ºC
Centrifugation 10min 14000 x g
4ºC
Filtrate
Sample preparation
Reverse phase chromatography
Column preparation
Sample preparation
or
+200µL AS
Sample:Sample buffer (3:1)
x2
Centrifugation 1min 1000 x g
+200µL ES
x2
Centrifugation 1min 1500 x g
150µL (xN)
New receiver tube
Sample binding
xN
Centrifugation 1min 1500 x g
Wash
+200µL WS
x2
Centrifugation 1min 1500 x g
Elution
+20µL EB
New receiver tube
x2
Centrifugation 1min 1500g
Ammonium sulphate precipitation
0.5g powder-ground tissue (0.25g per tube)
+ 0.6 mL Extraction Buffer
30min, 4ºC shaking
4ºC
Spin (x3)
4ºC
430µL of supernatant
+ Ammonium sulphate (75% FC)
4ºC
Centrifugation 25min 14000 x g
Supernatant
Sample preparation
Final extract
40µL total volume
70% ACN
0,1% formic acid