

## Methods Paper

**Cite this article:** Saccaram C, Bernay B, Brosse C, Collet B, Peng S, Sourdeval D, François T, Corso M, Rajjou L (2025). Development of an extraction method for the identification of peptides in the spermosphere of common bean (*Phaseolus vulgaris* L.). *Seed Science Research* 1–11. <https://doi.org/10.1017/S0960258525000030>

Received: 20 June 2024

Revised: 5 October 2024

Accepted: 2 December 2024

**Keywords:**




common bean; germination; peptides; seeds; spermosphere

**Corresponding author:**

Loïc Rajjou;

Email: [loic.rajjou@agroparistech.fr](mailto:loic.rajjou@agroparistech.fr)

# Development of an extraction method for the identification of peptides in the spermosphere of common bean (*Phaseolus vulgaris* L.)

Chandrodhay Saccaram<sup>1</sup> , Benoit Bernay<sup>2</sup> , Céline Brosse<sup>1</sup>, Boris Collet<sup>1</sup>, Shuang Peng<sup>1</sup>, Delphine Sourdeval<sup>1</sup>, Tracy François<sup>1</sup>, Massimiliano Corso<sup>1</sup> and Loïc Rajjou<sup>1</sup> 

<sup>1</sup>Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin for Plant Sciences (IJPB), Versailles 78000, France and <sup>2</sup>Plateforme Proteogen, US EMerode, Université de Caen Normandie, Caen 14000, France

**Abstract**

Seeds are crucial for plant reproduction, dispersal and agriculture. Seed quality and vigour greatly impact crop production by enabling rapid and uniform germination under various environmental conditions. This leads to healthy seedlings that can withstand both biotic and abiotic stresses, which are particularly important in the context of the accentuation of global climate change. Upon imbibition during germination *sensu stricto*, seeds release exudates, complex mixtures of organic and inorganic molecules, into the microenvironment surrounding them, known as the spermosphere. These exudates play a pivotal role in seedling development and overall plant fitness by influencing microbial selection, growth and interactions in the spermosphere, ultimately shaping the plant's microbiome. Proteins such as enzymes with protection properties have previously been demonstrated to be released by the seeds in their exudates. However, limited information is available pertaining to peptides in seed exudates. Here, we developed an experimental protocol to extract and identify peptides in the spermosphere of germinating common bean seeds. We showed that our methodology was successful in identifying a broad spectrum of peptides and that extraction solvent choice impacts peptide identification both qualitatively and quantitatively. We also show the possibility of using online prediction tools to predict the properties of identified peptides based on their amino acid sequence. We propose that this approach may be used to identify potential molecules that could be used as candidates for developing strategies to enhance seed quality and improve crop productivity.

**Introduction**

The early stages of the plant life cycle, including seed germination and seedling establishment, are key events for seed plant propagation. However, these stages are highly susceptible to environmental conditions, both biotic and abiotic factors. The ability of the seed embryo to restart its metabolic activity in a coordinated and sequential manner is defined as germination vigour (Rajjou et al., 2012; Galland et al., 2014). This is a highly sought trait in agriculture as it is directly correlated to eventual total biomass and yield production. Germination *sensu stricto* specifically refers to the rehydration of the quiescent and non-dormant dry seed and the subsequent protrusion of the embryonic axis through the surrounding structures, which are driven by localized cell elongation (Bewley and Black 1994; Sliwiska et al., 2009).

Upon imbibition, the seeds release in their surrounding microenvironment, known as the spermosphere, a complex mixture of organic and inorganic molecules (Martins et al., 2018; Nelson 2018). The spermosphere is the transient region of microbial and metabolic activities around a germinating seed, which are driven by seed exudate compounds, that exists specifically during the germination *sensu stricto* phase before transitioning to the rhizosphere. These exudate molecules shape and drive the selection, growth and interactions of microorganisms, which ultimately play an important role in seedling establishment and plant fitness by contributing to the future plant microbiome (Hill et al., 2012; Nelson 2018). The composition and functions of plant exudates have been relatively well studied in the phyllosphere (Xu et al., 2022) and rhizosphere (Sasse et al., 2018; Sun et al., 2021). In contrast, the spermosphere is less well described, though this area is advancing and garnering increasing attention, especially with the development of high-throughput techniques (Schiltz et al., 2015). The seed-derived peptides exuded in the spermosphere are expected to play crucial roles in shaping microbial communities and influencing plant–microbe interactions, as previously described for other plant-derived peptides (Hu et al., 2018; Pantigoso et al., 2022). These peptides can serve as nutrient sources, providing carbon and nitrogen to microorganisms and supporting their growth during seed germination. They may act as signalling molecules, affecting microbial colonization and seed defence. Several studies have demonstrated the release of a diverse range of proteins from seed exudates across

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

various plant species, which are often involved in defence mechanisms (Fuerst *et al.*, 2014; Rocha *et al.*, 2015; Raviv *et al.*, 2017; Hubert *et al.*, 2024). Despite the limited focus on the seed peptidome, seeds accumulate significant amounts of storage proteins, serving as reservoirs for potential peptides (Ayala-Niño *et al.*, 2019). For example, *Macadamia integrifolia* seeds have been shown to release antimicrobial peptides during germination, illustrating the presence of seed-exuded peptides in the spermosphere (Marcus *et al.*, 1999).

Peptides are polymers of amino acids. Although no definitive molecular size criterion formally distinguishes peptides from proteins, peptides are conventionally defined as chains composed of less than 50 amino acid residues, contrasting with larger protein molecules (Farrokhi *et al.*, 2008; Marmiroli and Maestri, 2014). Peptide roles in signalling and defence have been documented in the plant (Marmiroli and Maestri, 2014), but there is literature pertaining to their functions and roles in seeds or their exudation during seed germination. Interestingly, recent advancements in integrated peptidogenomic pipelines have allowed for the large-scale identification of non-conventional peptides (NCPs) in plants, emphasizing their significant roles in essential biological processes (Wang *et al.*, 2020).

Our study aims to apply untargeted peptidomics to investigate peptides exuded by germinating seeds in the spermosphere, addressing a significant gap in current research. Seed exudates contain a complex proteinaceous fraction that includes both high- and low-molecular-weight peptides (Scarafoni *et al.*, 2013; Rocha *et al.*, 2015). Thus, developing an optimized extraction method is essential for effective peptide isolation. We evaluated various solvents, including a commonly used weak acid, a strong acid, and an acidified organic solvent known for their peptide extraction efficacy (Vitorino *et al.*, 2012; Yu *et al.*, 2017; Chen *et al.*, 2018; Barashkova and Rogozhin, 2020). This assessment aims to identify the most efficient method for capturing a diverse range of peptides from the spermosphere.

To achieve this, we present a novel methodology specifically designed to study exuded peptides during seed germination using common bean seeds. Common bean was chosen as a potential reservoir of bioactive peptides due to their high protein content (Mojica and de Mejía, 2015; Mojica *et al.*, 2017). Our approach involved isolating peptides from the spermosphere using three selected extraction solvents, followed by mass spectrometry for peptide identification. Additionally, we used an online tool to predict various properties of the identified peptides based on their sequences.

## Material and methods

### Biological material and germination assay

Seeds were harvested in 2020 from the common bean genotype 'CON', a French commercial cultivar of *Phaseolus vulgaris* L. cv. Contender, provided by Vilmorin (<https://www.vilmorinmikado.fr/>). The seeds were cultivated from the FNAMS field experimental station in Maine-et-Loire (GPS, Global Positioning System coordinates: 47.470532, -0.394526) in the Loire Valley, France. Before testing, the seeds underwent micro-cleaning, and a purity test was performed by GEVES (the French official seed testing organization) in accordance with the International Seed Testing Association (ISTA) recommendations. The thousand seed weight was measured at 286.33 g, and the moisture content was 12.9%. To assess seed vigour, germination tests were conducted at the

Institut Jean-Pierre Bourgin (IJPB, INRAE-AgroParisTech) in Versailles, France. Fifty seeds were placed on SARTORIUS Pleated Seed Testing Papers (Grade 20, 119 g/m<sup>2</sup>) from Fischer Scientific and positioned in polypropylene germination boxes (55 mm × 120 mm × 180 mm, known as GEVES-type boxes, Loire Plastic, France), each containing 60 mL of Milli-Q water. These boxes were incubated in a Phytotron SANYO MLR-359, which was equipped with fluorescent white LED lamps (FL-40SS-W/37) set to maximum brightness (Light Step unit 5) to provide continuous daylight while maintaining a constant temperature of 25°C. Germination was monitored by counting radicle protrusion at 0, 24, 48, 72 and 96 h. The cumulative germination percentage was calculated, and a germination curve was generated using GerminaQuant in R (Lozano-Isla *et al.*, 2019). The seed lot exhibited high germination potential, exceeding 99% (Fig. 1A–C).

### Spermosphere collection

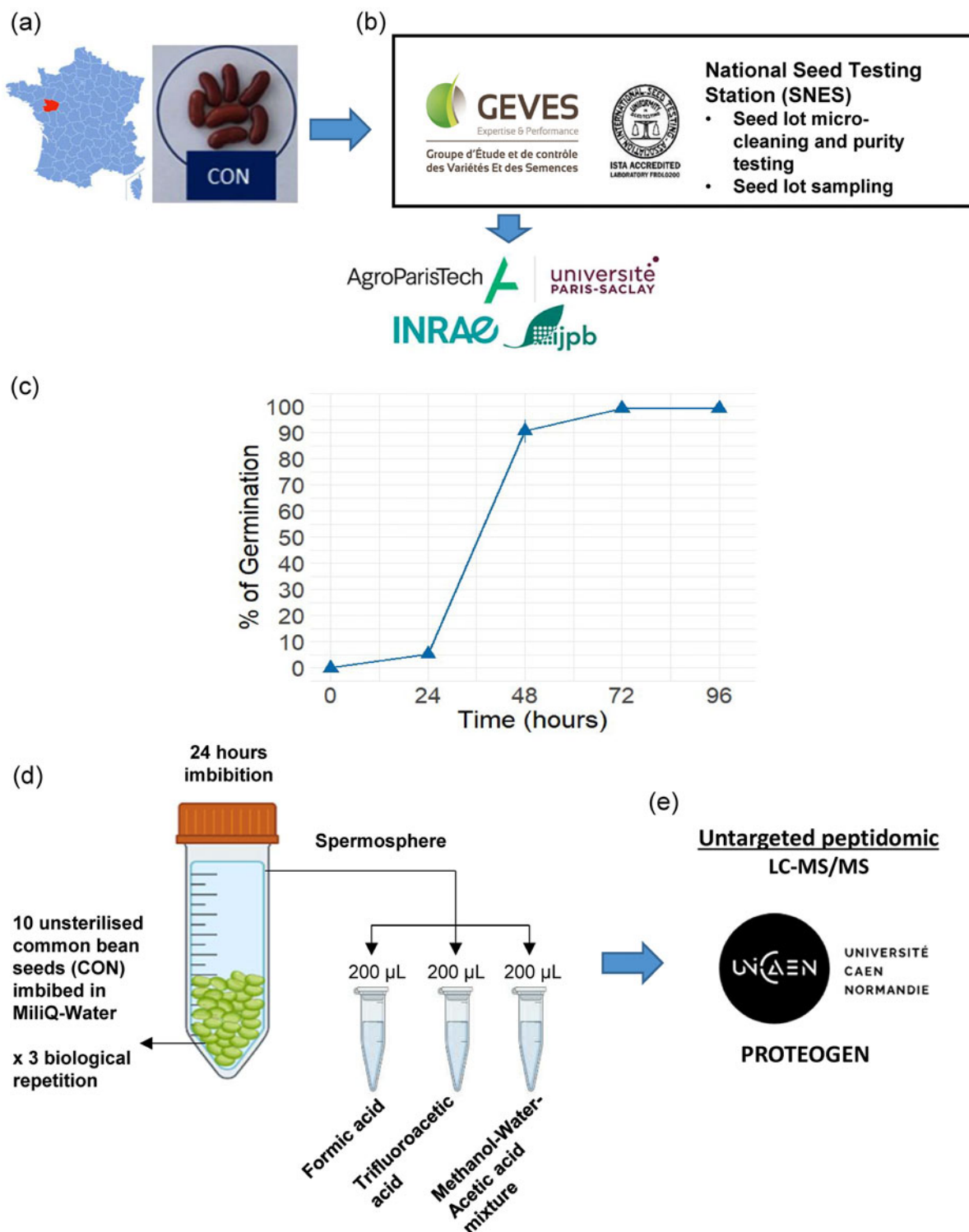
In this study, we assessed the composition of seed-exuded peptides in the spermosphere (Fig. 1D). To ensure the reliability of our results, three independent biological repetitions were considered in this study. To collect the spermosphere, 10 seeds were imbibed for 24 h, meaning at the end of phase II of the germination process, since the first radicle protrusions were observed (Fig. 1C) (Bewley, 1997). For each sample, the collective weight of the 10 seeds was determined. According to previous studies (Mendoza *et al.*, 2018; Vidak *et al.*, 2022), the absorption capacity of the seeds was measured beforehand and reached a plateau at 110% of their dry mass till radicle protrusion. Therefore, to provide an excess of sterile Milli-Q water (pH 6.5) for sampling the spermosphere without fully submerging the seeds, we determined that using a water mass corresponding to 150% of the dry mass is appropriate for our analysis. Imbibed seeds were shaken at 70 rpm on a platform shaker and kept at 25°C (Novotron INFORS AG CH-4103 Bottmingen) under a light intensity of 1000 lm/m<sup>2</sup>. After checking and finding that none of the 10 seeds had germinated, the water surrounding the seeds, including the spermosphere, was collected after 24 h.

### Extraction of peptides from the spermosphere and untargeted LC-MS/MS peptidomic analysis

#### Peptide extraction

Three biological replicates of the common bean spermosphere were subjected to peptide extraction using three different solvents. In particular, formic acid (FA, 0.1%), trifluoroacetic acid (TFA, 0.1%) and methanol:water:acetic acid (MeOH–H<sub>2</sub>O–AcOH, 90:9:1 v/v/v) were used separately for peptide solubilization and extraction as previously described (Kline *et al.*, 2003; Husson *et al.*, 2005; Hoofnagle *et al.*, 2016). Briefly, 200 µL of spermosphere samples from each biological replicate were distributed into three separate 2 mL Eppendorf tubes for solvent extraction (Fig. 1D). Subsequently, 1.8 mL of each solvent was added separately to the tubes containing the spermosphere samples. The samples with their respective extraction solvents were agitated at 24 rpm on a rotator (Mini LabRoller™ Rotator, LabNet) at 4°C for 30 min. The samples were then centrifuged at 10,000 g to separate the soluble peptides, which were found in the supernatant, from the insoluble components (precipitate). The peptides extracted in the supernatant were subsequently dried using a SpeedVac vacuum concentrator (Thermo Scientific™ SPD 111V).

## Seed origin and experimental design



**Figure 1.** Workflow of experimental design for spermosphere collection and analysis. (A) Seeds of the genotype 'CON' (*Phaseolus vulgaris* L.), a French commercial cultivar, were harvested in 2020 from the FNAMS field experimental station in Maine-et-Loire, France. (B) Harvested seeds were sent to the National Seed Testing Station (SNES) at GEVES, where they underwent micro-cleaning, purity testing, and sampling according to the ISTA guidelines. (C) Sampled seeds were then sent to the IJPB. Germination tests were conducted in three biological replicates on seed lots of the 'CON' genotype harvested in Maine-et-Loire. (D) Spermosphere collection from germinating seeds involved soaking 10 unsterilized seeds from the genotype CON seed lot harvested in Maine-et-Loire. To ensure the reliability of our results, three independent biological repetitions were considered in this study. The total weight of the seeds was measured before placing them in a 50 mL Falcon tube. The seeds were then imbibed with 1.5 times their total weight in sterile Milli-Q water. After verifying that none of the 10 seeds had germinated, the spermosphere was collected 24 h after imbibition. This spermosphere was divided into three Eppendorf tubes (200  $\mu$ L each) for peptide extraction with different solvents (FA, formic acid: 0.1% or TFA, trifluoroacetic acid: 0.1% or MeOH-H<sub>2</sub>O-AcOH, methanol:water:acetic acid: 90:9:1 v/v/v). (e) The untargeted peptidomic analyses using mass spectrometry were conducted at the Proteogen platform in Caen (France).

### Nano-LC-MS/MS analysis

The untargeted peptidomic analyses using mass spectrometry were conducted at the Proteogen platform in Caen (France) (Figs. 1E and 2). For nano-LC fragmentation, peptide samples were first desalted and concentrated onto a  $\mu$ C18 Omix (Agilent) before analysis. Dried peptide samples were resuspended in 40  $\mu$ L of FA (0.1%). The resuspended peptide samples were loaded to a  $\mu$ C18 Omix (Agilent) column for desalting and concentration and subsequently eluted with acetonitrile (80%) containing FA (0.1%). The eluted peptides are dried using a SpeedVac vacuum concentrator. The resulting concentrated peptide samples were then resuspended in 10  $\mu$ L of 0.1% FA. Finally, 1  $\mu$ L of the resuspended peptide extract was injected for LC-MS/MS analysis.

The chromatography step was performed on a NanoElute (Bruker Daltonics) ultra-high-pressure nanoflow chromatography system. Peptide samples were first concentrated onto a C18 PepMap 100 (5 mm  $\times$  300  $\mu$ m i.d.) precolumn (Thermo Scientific) and then separated at 50°C onto a reversed-phase Reprosil column (25 cm  $\times$  75  $\mu$ m i.d.) packed with 1.6  $\mu$ m C18-coated porous silica beads (IonOpticks). The mobile phases consisted of 0.1% FA, 99.9% water (v/v) (solvent A) and 0.1% FA in 99.9% acetonitrile (v/v) (solvent B). The nanoflow rate was set at 300 nL/min, and the gradient profile was as follows: from 2 to 15% solvent B within 15 min, followed by an increase to 25% solvent B within 10 min and to 37% solvent B within 12 min, followed by a 95% solvent B washing step before reequilibration.

Mass spectrometry experiments were carried out on a Trapped Ion Mobility Spectrometry – Time of Flight Pro mass spectrometer (Bruker Daltonics) with a modified CaptiveSpray Nano Electrospray Ion Source (ESI, Bruker Daltonics). The system was calibrated weekly, and the mass accuracy was better than 1 ppm. A 1600 spray voltage with a capillary temperature of 180°C was typically employed for ionization. MS spectra were acquired in the positive mode in the mass range from 100 to 1700 m/z. In the experiments described here, the mass spectrometer was operated in a parallel accumulation–serial fragmentation (PASEF) mode without the exclusion of single-charged peptides. A number of 10 PASEF MS/MS scans were performed for 1.25 s.

### Database searching for peptide identification

Peptide identification was performed using PEAKS XPro software (Bioinformatic Solutions Inc., CA) against an updated version of the UniProt *Phaseolus vulgaris* database (Proteome ID: UP000000226; The UniProt Consortium, 2022). The variable modifications allowed were as follows: methionine oxidation, acetylation (N-term), deamidation and carbamylation. The parameter ‘no enzyme’ was selected to permit the identification of peptides irrespective of any specific enzyme digestion. The mass accuracy settings were established at 20 ppm and 0.05 Da for MS and MS/MS mode, respectively. These settings were optimized based on the calibration of the mass spectrometer to ensure high-fidelity measurements of peptide masses. Identified peptides were filtered according to a false discovery rate (FDR) of 5%. Additionally, we implemented a redundancy elimination strategy based on proteins being evidenced by the same set or a subset of peptides to ensure the uniqueness of identified proteins.

### Data pre-treatment

Unique identifiers were assigned to each detected peptide based on its sequence, mass, retention time (RT) and presence of post-

translational modifications (PTMs), namely methionine oxidation, acetylation (N-term), deamidation and carbamylation. Subsequently, redundant peptides were removed.

### Statistical analysis

Principal component analysis (PCA) was conducted on the peptidomic dataset using R (v.4.1.1) to assess the relationships among peptides identified across samples. Only commonly identified peptides across all biological repetitions and all solvents (commonly identified peptides across all nine samples) were analysed. Peptide intensities of commonly identified peptides were normalized by the weight of 10 seeds for PCA.

### Peptide properties prediction

Following a selection process, the research focused on a specific set of peptides retained for further analysis. To gain insights into their behaviour, the physico-chemical properties of these retained peptides were predicted using PepAnalyser (Rustagi *et al.*, 2023). This online tool uses the amino acid sequences of the peptides to predict various characteristics, including molecular weight, isoelectric point (pI), net charge at physiological pH, hydrophobicity, half-life in different environments, instability index, aliphatic index and grand average of hydrophobicity (GRAVY).

## Results

### Diversity of distinct peptides in the spermosphere

Using three distinct solvents (FA, TFA, MeOH–H<sub>2</sub>O–AcOH), a total of 316 unique peptides were identified across three biological replicates of the common bean spermosphere (Supplementary Table S1). Unique codes were assigned to each peptide based on its amino acid sequence, mass, and any PTMs present. The peptide sequences ranged in length from 5 to 29 amino acids, and their masses spanned a range of 546.34 to 2738.31 Da (Fig. 3A, Supplementary Table S1).

The analysis of identified peptides revealed a distinct distribution in their lengths with peptides composed of 9, 10 or 11 amino acids being the most represented. The vast majority (98%) of all identified peptides fell within a broader range of 6–22 amino acids (Fig. 3A).

As illustrated in Fig. 3B, 31 out of the 316 identified peptides in the germinating common bean spermosphere (approximately 10%) exhibited PTMs. Among the four PTMs investigated on the identified peptides, acetylation (48.4%) was the most abundant modification, followed by deamidation (29.0%), carbamylation (19.4%) and methionine oxidation (12.9%) (Fig. 3C).

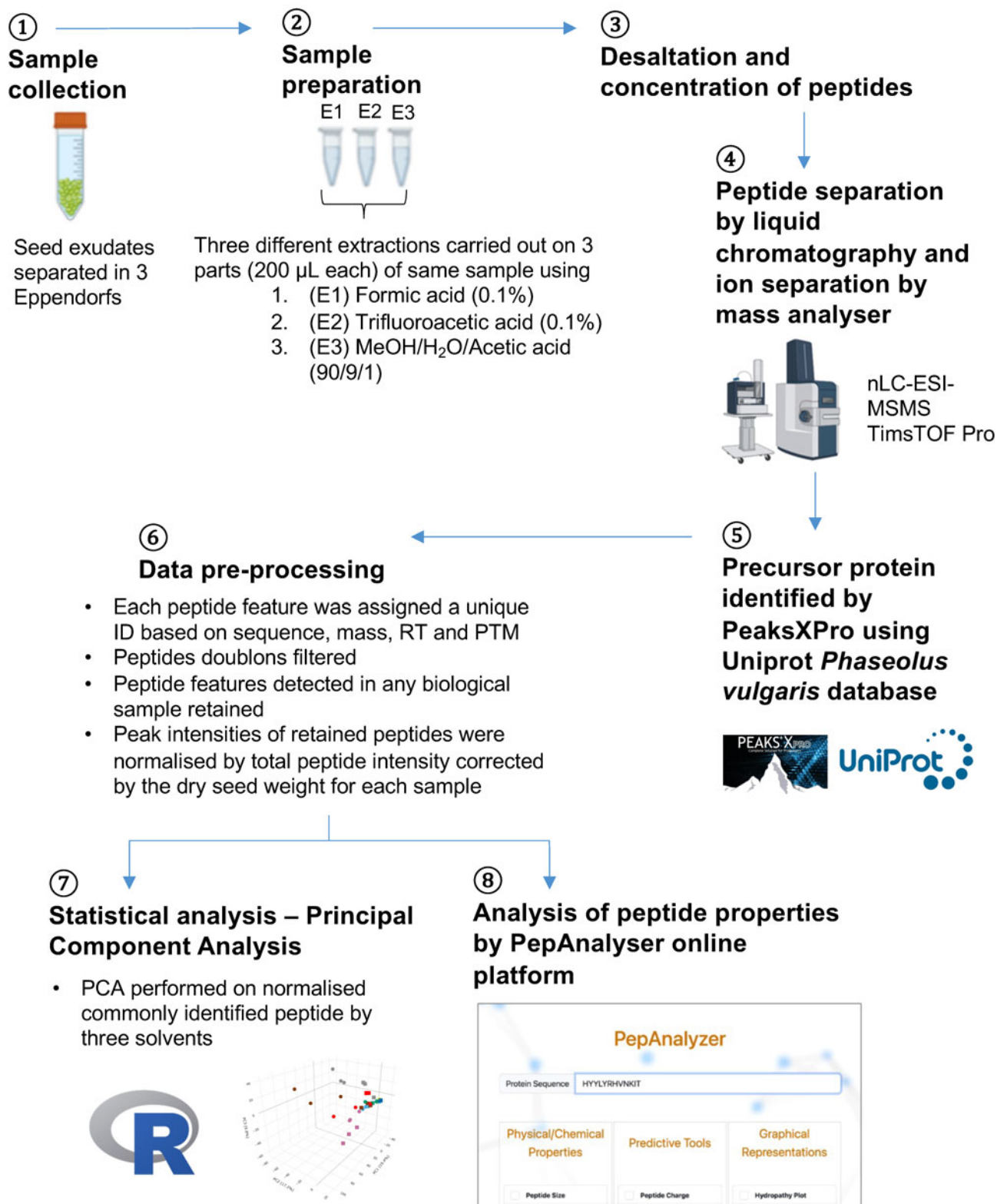
### Solvent choice impacts peptide identified

The choice of solvent significantly impacted peptide yield during extraction from the spermosphere. Among the three solvents tested (FA, TFA and MeOH–H<sub>2</sub>O–AcOH), FA emerged as the most efficient solvent, accounting for 85% of the identified peptides. MeOH–H<sub>2</sub>O–AcOH and TFA exhibited moderate efficiency, extracting 64 and 54% of the peptides, respectively (Fig. 4A).

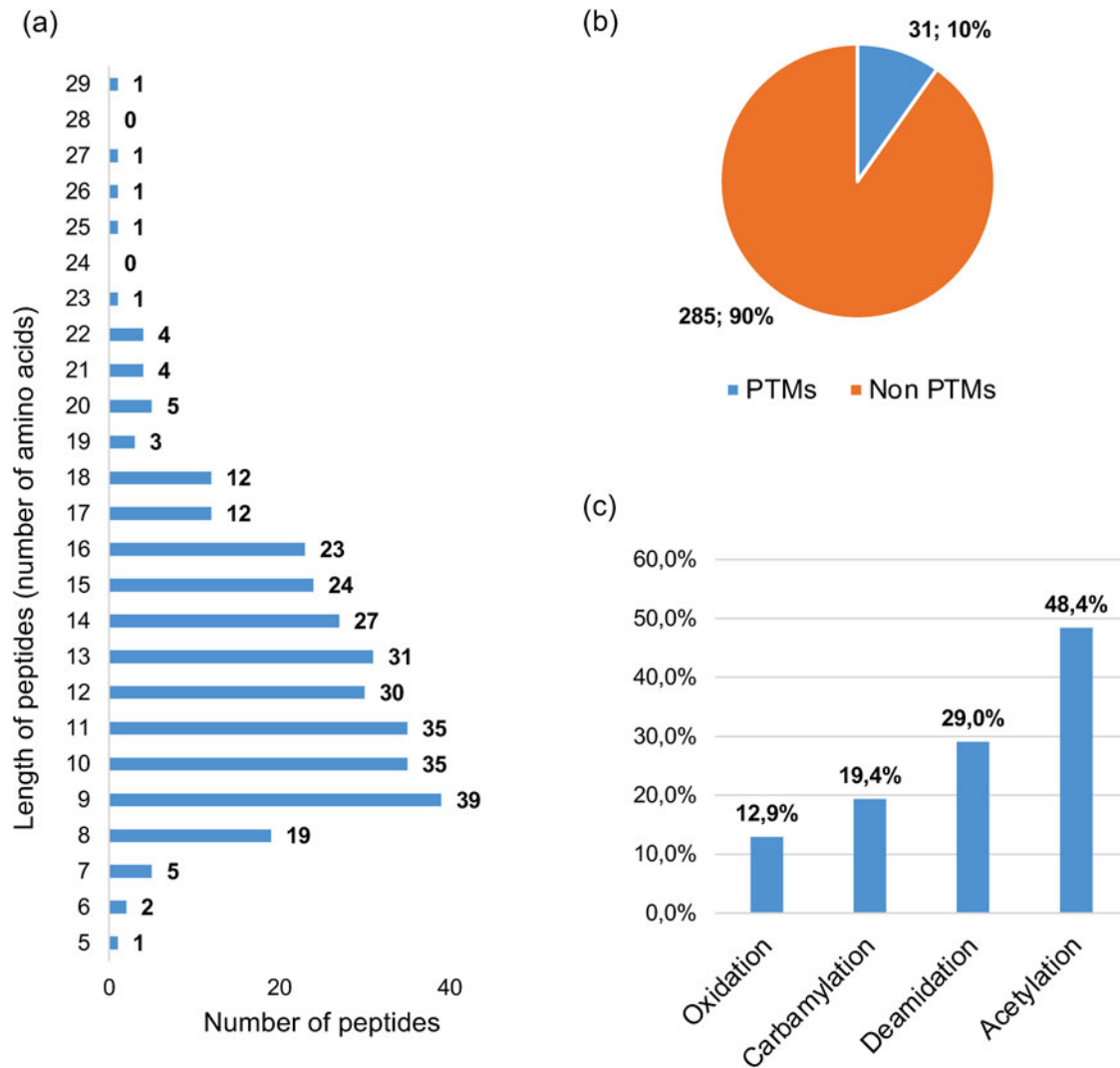
To account for biological variation, our experimental design incorporated three biological replicates for each solvent used in



## Peptidomic analysis workflow



**Figure 2.** Peptidomic analysis workflow. This pipeline illustrates an overview of the steps involved in analysing the peptidome from the spermosphere of germinating common bean seeds.



**Figure 3.** Characteristics of peptides identified in the spermosphere of common bean of genotype CON. (A) A total of 316 distinct peptides, with lengths ranging from 5 to 29 amino acids, were identified across the three biological repetitions using three solvents: formic acid (FA) at 0.1%, trifluoroacetic acid (TFA) at 0.1%, and a methanol–water–acetic acid mixture (90:9:1 v/v/v). (B) The distribution of identified peptides with PTMs, including methionine oxidation, acetylation (N-term), deamidation, and carbamylation, among the total of 316 identified peptides indicated that 31 peptides exhibited at least one PTM. (C) Percentage distribution of the various types of PTMs among the 31 identified peptides.

peptide extraction. The repeatability of peptide identification across these replicates was evaluated by analysing commonly identified peptides. Notably, using FA, 269 different peptides were identified, with 100 peptides consistently detected in all three biological replicates. Using MeOH–H<sub>2</sub>O–AcOH, we identified 203 peptides in the samples and 87 of them were consistently identified across the three biological repetitions. Lastly, 172 peptides were identified with TFA in the samples, with 71 peptides being consistently identified across the three biological repetitions. For each solvent, we determined the ratio of consistently identified peptides across the three biological repetitions to the total number of peptides identified by the solvent expressed in percentage to evaluate the repeatability in peptide extraction and identification (Fig. 4B).

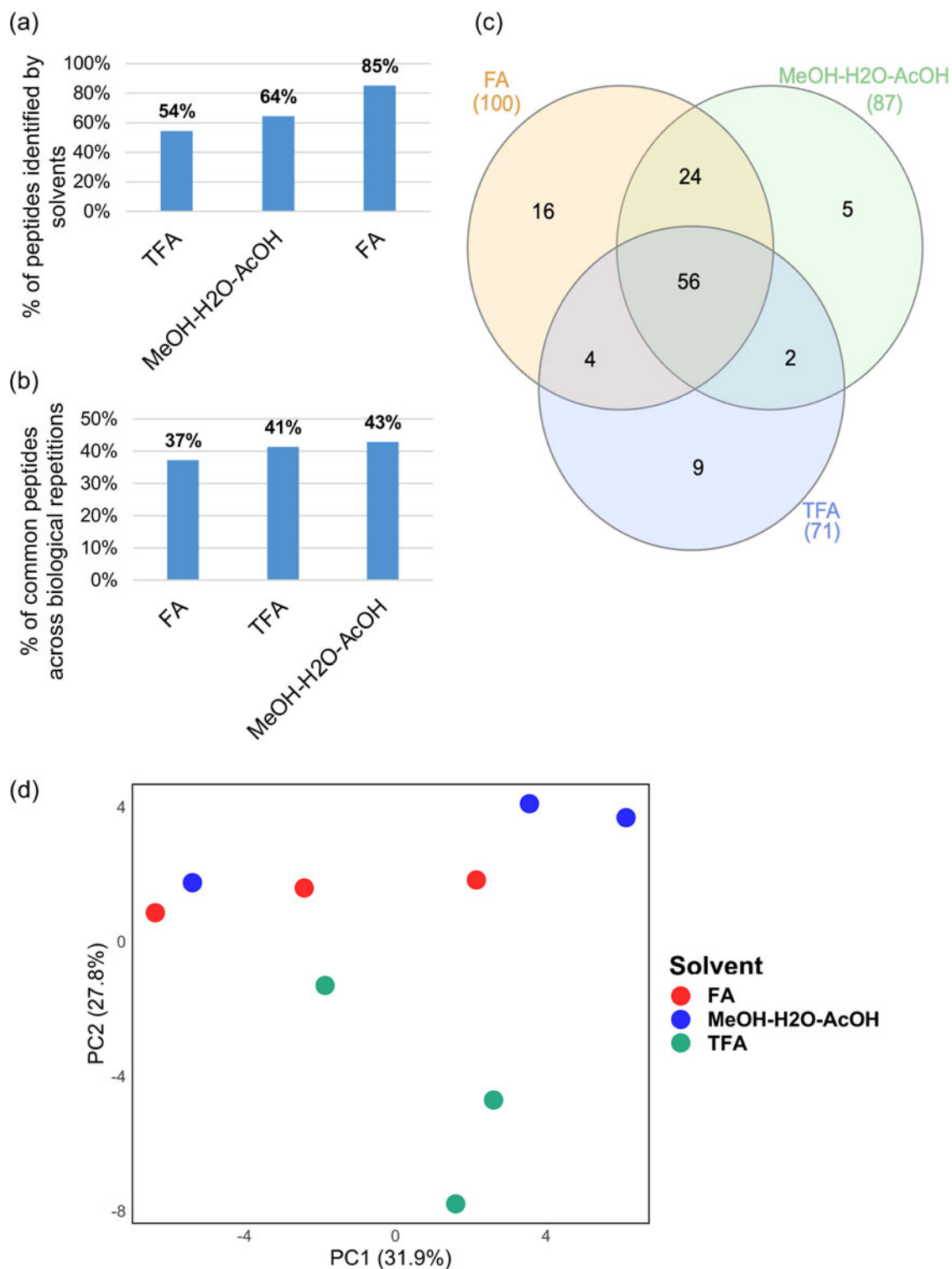
We compared the consistently identified peptides across the three biological repetitions (116 peptides) for each solvent by means of a Venn diagram (Heberle et al., 2015). We revealed 56 commonly identified peptides by the three different solvents

(Fig. 4C). Unique peptides were also identified according to each solvent. FA, MeOH–H<sub>2</sub>O–AcOH and TFA were responsible for the identification of 16, 5 and 9 unique peptides, respectively.

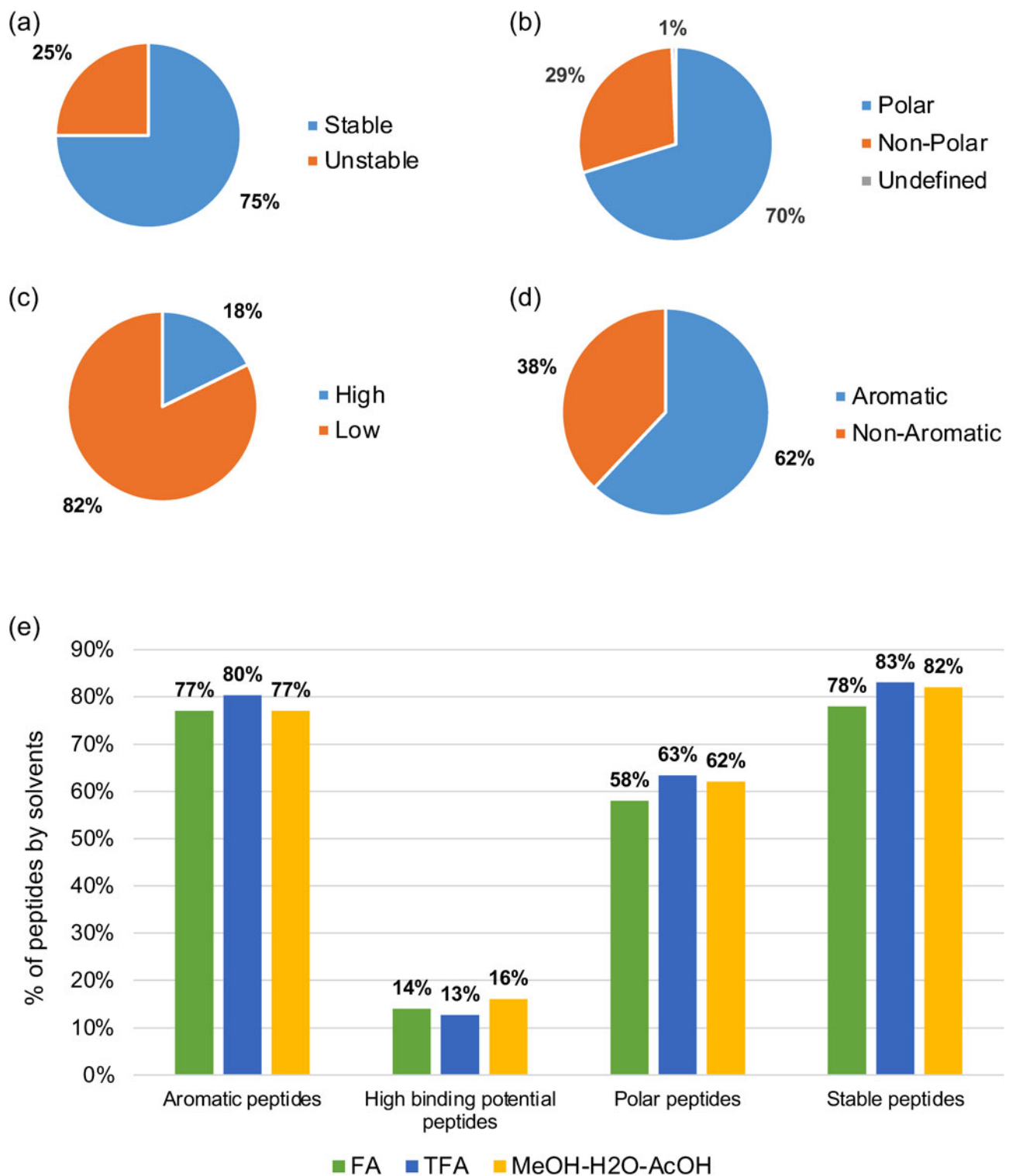
PCA based on peptide intensity data of the 56 commonly identified peptides grouped samples according to extraction solvents (Fig. 4D). Analysis of variance (ANOVA) of peptide intensities of the 56 common peptides across the three distinct solvents revealed only four differential peptides (FDR, *P*-value 0.05) (Supplementary Table S2).

#### Characteristics of identified peptides by the three different solvents

Using PepAnalyzer (<http://www.iksmbrlabdu.in/>), an online user-friendly tool that predicts several properties of a known peptide sequence (Rustagi et al., 2023), we identified several properties pertaining to the 316 identified peptides (Supplementary Table S1, Fig. 5).



**Figure 4.** Comparison of extraction solvents for peptide extraction. (A) Proportion of the total peptides identified by using the different extraction solvents (FA, formic acid: 0.1% or TFA, trifluoroacetic acid: 0.1% or MeOH-H<sub>2</sub>O-AcOH, methanol:water:acetic acid: 90:9:1 v/v/v) represented as percentages. (B) The repeatability of each solvent is represented as the percentage of commonly identified peptides across the three biological repetitions for each solvent, respectively. The commonly identified peptides across the three biological repetitions for each solvent were represented as a proportion of the total identified peptides from the solvents, respectively. (C) Comparison of the extraction solvents for the identification of peptides. The commonly identified peptides across the three biological repetitions for each solvent, respectively, were used to compute the Venn diagram. (D) PCA of commonly identified peptides across any three biological repetitions. PCA was applied to the normalized peptide intensity data. The figure illustrates the distribution of samples in the reduced-dimensional space defined by the first two principal components (PC1 and PC2). Each point represents an individual sample, and colours indicate the different solvents.



**Figure 5.** Predicted properties of 316 identified peptides in the spermosphere. The physico-chemical properties of peptides identified in the spermosphere were predicted based on their amino acid sequence using PepAnalyser (Rustagi *et al.*, 2023). Each property is shown as a percentage of the total dataset, giving insights into the relative prevalence of these properties among the peptides identified in the spermosphere. (A) Distribution of peptide stability. (B) Distribution of peptide polarity. (C) Distribution of peptide-binding potential. (D) Distribution of peptide aromaticity. (E) Comparison of extraction solvents (FA, formic acid: 0.1% or TFA, trifluoroacetic acid: 0.1% or MeOH-H<sub>2</sub>O-AcOH, methanol:water:acetic acid: 90:9:1 v/v/v) on predicted peptide properties.

The stability of the peptides was evaluated using an algorithm developed by Guruprasad *et al.* (1990) to calculate an instability index (Supplementary Table S1). According to this algorithm, a

score below 40 suggests a stable peptide, while values above 40 indicate potential instability. In accordance with this criterion, our results (Fig. 5A) indicate that 75% of the peptides were predicted as stable.



Kyte and Doolittle's (1982) method was used to assess peptide polarity through the grand average of hydropathy (GRAVY) index (Supplementary Table S1). This method assigns a hydropathy value to each amino acid in the sequence. To assess local preferences, it calculates the average hydropathy over a window of a pre-defined size that slides along the sequence from the N-terminus to the C-terminus. Peptides with a positive GRAVY index are predicted to be more hydrophilic (polar), while negative values indicate a more hydrophobic character. Our analysis revealed that a majority (70%) of the identified peptides exhibited a preference for hydrophilic interactions (Fig. 5B).

The propensity of the peptides to interact with proteins or membranes was evaluated using the Boman index (Supplementary Table S1). This index, also known as the potential protein interaction index, was developed by Boman (2003) to estimate binding potential. It is an estimate of the potential of a peptide to bind to membranes or other proteins as receptors (Boman et al. 1989; Boman 2003). The Boman index is determined by summing the solubility values of all residues in a sequence and then normalizing this sum by dividing it by the number of residues. A binding potential of greater than 2.48 indicates that a peptide has a strong binding potential (Boman 2003; Karnati et al., 2022). Only a small proportion of 316 identified peptides (18%) were defined as having a high binding potential (Fig. 5C).

Aromaticity describes whether a peptide sequence contains at least one of the aromatic amino acids (phenylalanine, tryptophan or tyrosine). The 316 peptides identified by the three distinct solvents (Supplementary Table S1) had a higher proportion of aromatic peptides (62%) (Fig. 5D).

We compared the properties of peptides extracted by different solvents by focusing on the peptides commonly identified in all three biological replicates for each solvent (FA-100, MeOH-H<sub>2</sub>O-AcOH-87 and TFA-71). We expressed the number of peptides with specific properties (aromaticity, binding potential, polarity and stability) as a proportion of the total commonly identified peptides for each solvent (Fig. 5E).

## Discussion

Peptides are increasingly recognized for their diverse bioactivities, playing crucial roles in plant-microbe interactions, signalling, defence mechanisms and nutrition (Hill et al., 2012; Hu et al., 2018; Li et al., 2021; Pantigoso et al., 2022; Furumizu and Shinohara, 2024). Despite this growing interest, research on peptides exuded by germinating seeds, particularly within the spermosphere, remains limited. The spermosphere represents a highly dynamic microenvironment where plant-microbe interactions can significantly influence seed vigour and protection. The need for further functional studies on seed-derived peptides in this context is crucial, given their potential to improve seed resilience in the face of varying biotic and abiotic stresses (Schiltz et al., 2015; Feng et al., 2023). Given the complex composition of seed exudates and the diverse physico-chemical characteristics of peptides, peptide isolation poses significant challenges, necessitating optimized extraction methods (Farrokhi et al., 2008; Finoult et al., 2011; Pinkse et al., 2014; Schiltz et al., 2015). This study represents a novel attempt to address these challenges by extracting and analysing seed-derived peptides from the spermosphere in an untargeted manner.

In the broader context of plant peptide extraction, solvents are generally classified into two categories: water-based solutions, such as acids and buffers, and organic-based solutions, like

ethanol-water mixtures (Barashkova and Rogozhin, 2020). In this study, we employed two commonly used water-based acidic solvents, TFA and FA, along with an acidified organic solvent, MeOH-H<sub>2</sub>O-AcOH. These solvents are widely used for their capacity to selectively precipitate and extract peptides in a single step (Vitorino et al., 2012; Yu et al., 2017; Chen et al., 2018). Given their differing properties, we hypothesized that each solvent would influence the nature of the extracted peptides. Among the 316 peptides identified, FA demonstrated the highest extraction efficiency, yet each solvent uniquely identified specific peptides. While the solvents efficiently identified a shared set of 56 peptides, no consistent physico-chemical characteristics were favoured by any particular solvent (Fig. 5E). A PCA of peptide peak intensities further revealed differences in the extraction efficiency of the common peptides, underscoring the significant influence of solvent selection (Fig. 4D).

A notable limitation of our experimental design was the use of a small seed sample size (10 seeds per biological replicate), which likely limited our ability to effectively minimize biological variability (Barashkova and Rogozhin, 2020). This, combined with technical variability associated with solvent extraction, may have influenced the reproducibility of the results. To better account for biological variability, future studies should increase the number of seeds per sample and examine multiple genotypes to capture a broader spectrum of peptide diversity (Vitorino et al., 2012). Moreover, given the dynamic nature of the spermosphere, focusing on a single time point (24 h after imbibition) may have limited our ability to capture the full temporal dynamics of peptide release. It is also important to note that factors such as the seed's water content prior to imbibition, as well as the temperature during imbibition, significantly influence both the quantity and composition of seed exudates, particularly in larger seeds like beans (Schroth et al., 1966; Toledo et al., 2010). Studying the peptidome at multiple time points and diverse environments would provide a more comprehensive understanding of the changes in peptide composition throughout seed germination (Nelson, 2004; Schiltz et al., 2015).

Our untargeted peptidome analysis successfully identified peptides ranging from 5 to 29 amino acids in length. However, a well-recognized challenge in mass spectrometry-based peptidomics is the identification of very short peptides (<5 amino acids) (Nongonierma and FitzGerald, 2017). These peptides are difficult to detect due to poor fragmentation patterns and the limited structural diversity that complicates their assignment (Fricker, 2015; Hollebrands et al., 2023). Moreover, existing protein databases often exclude short peptides, increasing the likelihood of false discoveries (Miller and Smith, 2023). To overcome these challenges, future research should consider incorporating De Novo sequencing, specialized RT prediction models or alternative workflows better suited for short peptide detection and identification (Piovesana et al., 2018; Hollebrands et al., 2023).

Our methodology not only allowed for the identification of peptide sequences but also provided insights into their predicted properties, such as stability, polarity and aromaticity (Rustagi et al., 2023). These characteristics suggest significant potential for biological activity, aligning with recent research into the functional roles of peptides. The findings of Wang et al. (2020) further underscore the potential functional diversity of peptides in plants, indicating that many NCPs may play significant roles in key biological processes. This includes the proposed functions of seed-derived peptides in influencing and shaping microbial communities within the spermosphere.

## Conclusion

Our study represents a pioneering effort to develop a methodology for analysing peptide diversity in the seed spermosphere, focusing on establishing an untargeted approach to detect the presence of peptides and evaluate the impact of different solvents on extraction efficiency. This work lays the groundwork for future investigations by demonstrating the importance of solvent choice in peptide recovery and characterization. For more comprehensive studies, future research should consider pooling peptides extracted with various solvents and including a broader range of genotypes to capture a wider diversity of peptides. Experimental protocols and peptidomic analysis pipelines should be tailored to address specific biological questions, such as the roles of these peptides in plant–microbe interactions and their potential applications in seed technology. Furthermore, although factors such as sample size, temporal dynamics, genetic diversity and environmental seed conditions were not evaluated in this study, they are crucial considerations for optimizing peptide analysis outcomes and should be informed by the specific biological context of future investigations.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0960258525000030>.

**Data availability statement.** The data supporting the findings of this study are openly available in online repositories. Peptidomics data are available on iProX at <https://www.iprox.cn/> and are referenced under the Project ID: IPX0008926001.

**Author contributions.** C.S., B.B., C.B., and B.C. conducted experiments. C.S., B.B., S.P., and L.R. analysed the results. C.S. and L.R. wrote the original draft preparation. C.S., B.B., M.C., and L.R. wrote, reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding statement.** This work was part of the 3rd Programme for Future Investments (France2030). It is operated by the SUCSEED project (ANR-20-PCPA-0009) and funded by the ‘Growing and Protecting Crops Differently’ French Priority Research Program (PPR-CPA), part of the national investment plan operated by the French National Research Agency (ANR). The IJPB benefits from the support of Saclay Plant Sciences – SPS (ANR-17-EUR-0007).

**Competing interests.** The authors declare no conflicts of interest.

## References

- Ayala-Niño A, Rodríguez-Serrano GM, González-Olivares LG, Contreras-López E, Regal-López P and Cepeda-Saez A (2019) Sequence identification of bioactive peptides from Amaranth seed proteins (*Amaranthus hypochondriacus* spp.). *Molecules* **24**, 3033.
- Barashkova AS and Rogozhin EA (2020) Isolation of antimicrobial peptides from different plant sources: does a general extraction method exist? *Plant Methods* **16**, 143.
- Bewley JD (1997) Seed germination and dormancy. *The Plant Cell* **9**, 1055.
- Bewley JD and Black M (1994) *Seeds: Physiology of Development and Germination*, 2nd Edn. New York: Springer, Plenum Press.
- Boman HG (2003) Antibacterial peptides: basic facts and emerging concepts. *Journal of Internal Medicine* **254**, 197–215.
- Boman HG, Wade D, Boman IA, Wählin B and Merrifield RB (1989) Antibacterial and antimalarial properties of peptides that are cecropin-melittin hybrids. *FEBS Letters* **259**, 103–106.
- Chen W, Hwang YY, Gleaton JW, Titus JK and Hamlin NJ (2018) Optimization of a peptide extraction and LC–MS protocol for quantitative analysis of antimicrobial peptides. *Future Science OA* **5**, FSO348.
- Farrokhi N, Whitelegge JP and Brusslan JA (2008) Plant peptides and peptidomics. *Plant Biotechnology Journal* **6**, 105–134.
- Feng YZ, Zhu QF, Xue J, Chen P and Yu Y (2023) Shining in the dark: the big world of small peptides in plants. *Abiotech* **4**, 238–256.
- Finoulst I, Pinkse M, Van Dongen W and Verhaert P (2011) Sample preparation techniques for the untargeted LC–MS-based discovery of peptides in complex biological matrices. *BioMed Research International* **2011**, 245291.
- Fricker LD (2015) Limitations of mass spectrometry-based peptidomic approaches. *Journal of the American Society for Mass Spectrometry* **26**, 1981–1991.
- Fuerst EP, Okubara PA, Anderson JV and Morris CF (2014) Polyphenol oxidase as a biochemical seed defense mechanism. *Frontiers in Plant Science* **5**, 689.
- Furumizu C and Shinohara H (2024) Land plant peptide signaling: what we know—and don't know—about its evolution. *Physiologia Plantarum* **176**, e14172.
- Galland M, Huguet R, Arc E, Cueff G, Job D and Rajjou L (2014) Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during *Arabidopsis* seed germination. *Molecular & Cellular Proteomics* **13**, 252–268.
- Guruprasad K, Reddy BV and Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering* **4**, 155–161.
- Heberle H, Meirelles GV, da Silva FR, Telles GP and Minghim R (2015) InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169.
- Hill PW, Farrell M and Jones DL (2012) Bigger may be better in soil N cycling: does rapid acquisition of small L-peptides by soil microbes dominate fluxes of protein-derived N in soil? *Soil Biology and Biochemistry* **48**, 106–112.
- Hollebrands B, Hageman JA, van de Sande JW, Albada B and Janssen HG (2023) Improved LC–MS identification of short homologous peptides using sequence-specific retention time predictors. *Analytical and Bioanalytical Chemistry* **415**, 2715–2726.
- Hoofnagle AN, Whiteaker JR, Carr S, Kuhn E, Liu T, Massoni S, Thomas SN, Townsend RR, Zimmerman LJ, Boja E, Chen J, Crimmins DL, Davies SR, Gao Y, Hiltke T, Ketchum KA, Kinsinger CR, Mesri M, Meyer MR, Qian W, Schoenherr RM, Scott MG, Shi T, Whiteley GR, Wrobel JA, Wu C, Ackermann BL, Aebersold R, Barnidge DR, Bunk DM, Clarke NJ, Fishman JB, Grant RP, Kusebauch U, Kushnir MM, Lowenthal MS, Moritz RL, Neubert H, Patterson SD, Rockwood AL, Rogers JC, Singh RJ, Van Eyk JE, Wong SH, Zhang S, Chan DW, Chen X, Ellis MJ, Liebler DC, Rodland KD, Rodriguez H, Smith R, Zhang Z, Zhang H and Paulovich AG (2016) Recommendations for the generation, quantification, storage, and handling of peptides used for mass spectrometry-based assays. *Clinical Chemistry* **62**, 48–69.
- Hu Z, Zhang H and Shi K (2018) Plant peptides in plant defense responses. *Plant Signaling & Behavior* **13**, e1475175.
- Hubert B, Marchi M, Ly Vu J, Tranchant C, Tarkowski ŁP, Leprince O and Buitink J (2024) A method to determine antifungal activity in seed exudates by nephelometry. *Plant Methods* **20**, 16.
- Husson SJ, Clynen E, Baggerman G, De Loof A and Schoofs L (2005) Peptidomics of *Caenorhabditis elegans*: in search of neuropeptides. *Communications in Agricultural and Applied Biological Sciences* **70**, 153–156.
- Karnati P, Gonuguntala R, Barbadikar KM, Mishra D, Jha G, Prakasham V, Chilumula P, Shaik H, Pesari M, Sundaram RM and ChinnaSwami K (2022) Performance of novel antimicrobial protein Bg\_9562 and in silico predictions on its properties with reference to its antimicrobial efficiency against *Rhizoctonia solani*. *Antibiotics (Basel, Switzerland)* **11**, 363.
- Kline TR, Pang J, Hefta SA, Opitck GJ, Kiefer SE and Scheffler JE (2003) A high-yield method to extract peptides from rat brain tissue. *Analytical Biochemistry* **315**, 183–188.
- Kyte J and Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* **157**, 105–132.
- Li J, Hu S, Jian W, Xie C and Yang X (2021) Plant antimicrobial peptides: structures, functions, and applications. *Botanical Studies* **62**, 5.

- Lozano-Isla F, Benites-Alfaro OE and Pompelli MF** (2019) *GerminaR: An R Package for Germination Analysis with the Interactive Web Application 'GerminaQuant for R'*, Vol. 34 Hoboken, USA: John Wiley & Sons, Inc., pp. 339–346.
- Marcus JP, Green JL, Goulter KC and Manners JM** (1999) A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *The Plant Journal* **19**, 699–710.
- Marmioli N and Maestri E** (2014) Plant peptides in defense and signaling. *Peptides* **56**, 30–44.
- Martins SJ, Medeiros FHV, Lakshmanan V and Bais HP** (2018) Impact of seed exudates on growth and biofilm formation of *Bacillus amyloliquefaciens* ALB629 in common bean. *Frontiers in Microbiology* **8**, 2631.
- Mendoza FA, Wiesinger JA, Lu R, Nchimbi-Msolla S, Miklas PN, Kelly JD and Cichy KA** (2018) Prediction of cooking time for soaked and unsoaked dry beans (*Phaseolus vulgaris* L.) using hyperspectral imaging technology. *The Plant Phenome Journal* **1**, 1–9.
- Miller RM and Smith LM** (2023) Overview and considerations in bottom-up proteomics. *Analyst* **148**, 475–486.
- Mojica L and de Mejía EG** (2015) Characterization and comparison of protein and peptide profiles and their biological activities of improved common bean cultivars (*Phaseolus vulgaris* L.) from Mexico and Brazil. *Plant Foods for Human Nutrition* **70**, 105–112.
- Mojica L, Luna-Vital DA and González de Mejía E** (2017) Characterization of peptides from common bean protein isolates and their potential to inhibit markers of type-2 diabetes, hypertension and oxidative stress. *Journal of the Science of Food and Agriculture* **97**, 2401–2410.
- Nelson EB** (2004) Microbial dynamics and interactions in the spermosphere. *Annual Review of Phytopathology* **42**, 271–309.
- Nelson EB** (2018) The seed microbiome: origins, interactions, and impacts. *Plant and Soil* **422**, 7–34.
- Nongonierma A and FitzGerald R** (2017) Strategies for the discovery and identification of food protein-derived biologically active peptides. *Trends in Food Science & Technology* **69**, 289–305.
- Pantigoso HA, Newberger D and Vivanco JM** (2022) The rhizosphere microbiome: plant–microbial interactions for resource acquisition. *Journal of Applied Microbiology* **133**, 2864–2876.
- Pinkse M, Evaristo G, Pieterse M, Yu Y and Verhaert P** (2014) MS approaches to select peptides with post-translational modifications from amphibian defense secretions prior to full sequence elucidation. *EuPA Open Proteomics* **5**, 32–40.
- Piovesana S, Capriotti AL, Cavaliere C, La Barbera G, Montone CM, Zenezini Chiozzi R and Laganà A** (2018) Recent trends and analytical challenges in plant bioactive peptide separation, identification and validation. *Analytical and Bioanalytical Chemistry* **410**, 3425–3444.
- Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C and Job D** (2012) Seed germination and vigor. *Annual Review of Plant Biology* **63**, 507–533.
- Raviv B, Aghajanyan L, Granot G, Makover V, Frenkel O, Gutterman Y and Grafi G** (2017) The dead seed coat functions as a long-term storage for active hydrolytic enzymes. *PLoS One* **12**, e0181102.
- Rocha RO, Morais JKS, Oliveira JTA, Oliveira HD, Sousa DOB, Souza CEA, Moreno FB, Monteiro-Moreira ACO, de Souza Júnior JDA, de Sá MFG and Vasconcelos IM** (2015) Proteome of soybean seed exudates contains plant defense-related proteins active against the root-knot nematode *Meloidogyne incognita*. *Journal of Agricultural and Food Chemistry* **63**, 5335–5343.
- Rustagi V, Gupta S, Bajaj M, Singh A and Singh I** (2023) PepAnalyzer: predicting peptide properties using its sequence. *Amino Acids* **55**, 1371–1379.
- Sasse J, Martinoia E and Northen T** (2018) Feed your friends: do plant exudates shape the root microbiome? *Trends in Plant Science* **23**, 25–41.
- Scarafoni A, Ronchi A, Prinsi B, Espen L, Assante G, Venturini G and Duranti M** (2013) The proteome of exudates from germinating *Lupinus albus* seeds is secreted through a selective dual-step process and contains proteins involved in plant defence. *The FEBS Journal* **280**, 1443–1459.
- Schiltz S, Gaillard I, Pawlicki-Julian N, Thiombiano B, Mesnard F and Gontier E** (2015) A review: what is the spermosphere and how can it be studied? *Journal of Applied Microbiology* **119**, 1467–1481.
- Schroth MN, Weinhold AR and Hayman DS** (1966) The effect of temperature on quantitative differences in exudates from germinating seeds of bean, pea, and cotton. *Canadian Journal of Botany* **44**, 1429–1432.
- Sliwinska E, Bassel GW and Bewley JD** (2009) Germination of *Arabidopsis thaliana* seeds is not completed as a result of elongation of the radicle but of the adjacent transition zone and lower hypocotyl. *Journal of Experimental Botany* **60**, 3587–3594.
- Sun H, Jiang S, Jiang C, Wu C, Gao M and Wang Q** (2021) A review of root exudates and rhizosphere microbiome for crop production. *Environmental Science and Pollution Research International* **28**, 54497–54510.
- The UniProt Consortium** (2022) UniProt: the universal protein knowledge-base in 2023. *Nucleic Acids Research* **51**, D523–D531.
- Toledo MZ, Cavarani C, França-Neto JDB and Nakagawa J** (2010) Imbibition damage in soybean seeds as affected by initial moisture content, cultivar and production location. *Seed Science and Technology* **38**, 399–408.
- Vidak M, Lazarević B, Javornik T, Šatović Z and Carović-Stanko K** (2022) Seed water absorption, germination, emergence and seedling phenotypic characterization of the common bean landraces differing in seed size and color. *Seeds* **1**, 324–339.
- Vitorino R, Barros AS, Caseiro A, Ferreira R and Amado F** (2012) Evaluation of different extraction procedures for salivary peptide analysis. *Talanta* **94**, 209–215.
- Wang S, Tian L, Liu H, Li X, Zhang J, Chen X, Jia X, Zheng X, Wu S, Chen Y, Yan J and Wu L** (2020) Large-scale discovery of non-conventional peptides in Maize and Arabidopsis through an integrated peptidogenomic pipeline. *Molecular Plant* **13**, 1078–1093. doi:10.1016/j.molp.2020.05.012.
- Xu N, Zhao Q, Zhang Z, Zhang Q, Wang Y, Qin G, Ke M, Qiu D, Peijnenburg WJGM, Lu T and Qian H** (2022) Phyllosphere microorganisms: sources, drivers, and their interactions with plant hosts. *Journal of Agricultural and Food Chemistry* **70**, 4860–4870.
- Yu Q, Canales A, Glover MS, Das R, Shi X, Liu Y, Keller MP, Attie AD and Li L** (2017) Targeted mass spectrometry approach enabled discovery of O-glycosylated insulin and related signaling peptides in mouse and human pancreatic islets. *Analytical Chemistry* **89**, 9184–9191.