



**HAL**  
open science

## **A mass spectrometry-based peptidomic dataset of the spermosphere in common bean (*Phaseolus vulgaris* L.) seeds**

Chandrodhay Saccaram, Céline Brosse, Boris Collet, Delphine Sourdeval, Tracy François, Benoît Bernay, Massimiliano Corso, Loïc Rajjou

### ► **To cite this version:**

Chandrodhay Saccaram, Céline Brosse, Boris Collet, Delphine Sourdeval, Tracy François, et al.. A mass spectrometry-based peptidomic dataset of the spermosphere in common bean (*Phaseolus vulgaris* L.) seeds. *Scientific Data*, 2024, 11 (1), pp.1202. <10.1038/s41597-024-04044-y>. <hal-05166334>

**HAL Id: hal-05166334**

**<https://hal.science/hal-05166334v1>**

Submitted on 16 Jul 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# A mass spectrometry-based peptidomic dataset of the spermosphere in common bean (*Phaseolus vulgaris* L.) seeds

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

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

Corresponding author(s):

Chandrodhay Saccaram & Loïc Rajjou

[chandrodhay.saccaram@inrae.fr](mailto:chandrodhay.saccaram@inrae.fr) & [loic.rajjou@agroparistech.fr](mailto:loic.rajjou@agroparistech.fr)

## Abstract (141 words)

The spermosphere, a dynamic microenvironment surrounding germinating seeds, is shaped by the complex interactions between natural compounds exuded by seeds and seed-associated microbial communities. While peptides exuded by plants are known to influence microbiota diversity, little is known about those specifically exuded by seeds. In this study, we characterised the peptidome profile of the spermosphere for the first time using seeds from eight genotypes of common bean (*Phaseolus vulgaris*) grown in two contrasting production regions. An untargeted LC-MS/MS peptidomic analysis revealed 3,258 peptides derived from 414 precursor proteins of common bean in the spermosphere. This comprehensive peptidomic dataset provides valuable insights into the characteristics of peptides exuded by common bean seeds in the spermosphere. It can be used to identify peptides with potential antimicrobial or other biological activities, advancing our understanding of the functional roles of seed-exuded peptides in the spermosphere.

## Background & Summary

During germination, seeds release a complex mixture of organic and inorganic compounds into their surrounding environment, leading to the formation of a localised microenvironment known as the spermosphere, which is defined as the zone of increased microbial activity around germinating seeds<sup>1,2</sup>. While plant exudates have been extensively characterised and their roles in plant-microorganism interactions explored in other plant compartments, such as the rhizosphere<sup>3,4</sup>, seed exudates have received comparatively less attention. Similar to the interactions in the rhizosphere driven by root exudates, seed exudates during germination are likely instrumental in shaping the spermosphere microbiome<sup>5</sup>. A previous study characterising the proteome of germinating *Lupinus albus* seeds found that seed-exuded proteins are involved in plant defense mechanisms<sup>6</sup>. Indeed, proteins have long been recognised as playing crucial roles in plant immunity<sup>7</sup>.

Similar to proteins, several studies have characterised peptides that possess antimicrobial activities<sup>8</sup>. Although the seed peptidome has received limited attention, storage proteins (important reservoirs of potential peptides) are abundantly accumulated in seeds<sup>9</sup>. One study demonstrated that anti-microbial peptides (AMPs) are released by *Macadamia integrifolia* seeds during germination<sup>10</sup>, confirming the presence of seed-exuded peptides in the spermosphere. Beyond their role as antimicrobial compounds, plant-exuded peptides may also serve as a nutritional source for microorganisms<sup>11</sup>. This raises important questions about the composition, activities, and functions of seed-exuded peptides in the spermosphere. The current study aims to characterise peptides exuded by germinating seeds in the spermosphere, using common bean as a model system.

58 Common bean (*Phaseolus vulgaris* L.) is a widely consumed grain legume known for its high  
59 nutritional value and protein content in seeds<sup>12,13</sup>. As a legume species, common bean engages  
60 in mutually beneficial relationships with symbionts, mediated by complex molecular signalling  
61 between the host and rhizobial microsymbionts<sup>14</sup>. While nodulation in leguminous plants and  
62 the signalling processes involved in shaping the rhizosphere microbiome have been relatively  
63 well studied<sup>14–17</sup>, there is limited information on the early phase of the plant life cycle, particularly  
64 during seed germination. Previous research has shown that the common bean seed  
65 microbiome undergoes shifts in both abundance and community structure throughout  
66 germination<sup>18</sup>. This suggests the hypothesis that seeds may selectively influence their  
67 microbiome, at least partially, through exudates released during germination. Given that  
68 antimicrobial peptides have been found in seed exudates<sup>10</sup> and considering that plant-exuded  
69 peptides can also serve as a nutrient source for microorganisms<sup>11</sup>, the protein-rich common  
70 bean seed provides an ideal model for exploring the diversity of peptides exuded in the  
71 spermosphere during germination.

72  
73 In this study, we investigated the peptide composition of the spermosphere during the  
74 germination of common bean seeds under in vitro conditions. Our focus was on the peptides  
75 exuded by seeds over a 24-hour period. We analysed the peptidome profile in the  
76 spermosphere of eight different common bean genotypes, sourced from plants cultivated in two  
77 distinct geographical regions in France. The objectives were to (i) explore the peptidome  
78 landscape in the spermosphere and (ii) highlight the differences among the genotypes. Seeds  
79 were imbibed in sterile water at four time intervals (2, 8, 16, and 24 hours), and the surrounding  
80 water for each time interval was collected and pooled. Untargeted LC-MS/MS peptidomic  
81 analysis was then performed to assess the diversity of peptides in the pooled spermosphere.

## 82 83 **Methods**

### 84 85 **Biological materials**

86  
87 Seeds were harvested in 2020 from eight different common bean genotypes (French  
88 commercial cultivars of *Phaseolus vulgaris* L.), cultivated at two FNAMS field experimental  
89 stations: Gers in the southwest of France (43°57'25.2"N 0°23'31.7"E) and Maine-et-Loire  
90 (47°28'13.9"N 0°23'40.3"W) in the Loire Valley. The selected genotypes (i.e. VEZ, CON, CAP,  
91 FAC, FLA, LIN, DEE, and VAN) were chosen to represent the highest diversity among breeding  
92 programs and qualitative seed traits in France, as determined by the seed company Vilmorin-  
93 Mikado SAS (Limagrain group) (Figure 1).

### 94 **Spermosphere production and collection**

95  
96 The experimental design for the collection of seed exudates in the spermosphere is illustrated  
97 in Figure 1. This study characterised the composition of peptides exuded by seeds in the  
98 spermosphere and evaluated the impact of seed production location and genetic background  
99 on these exuded peptides. Each seed lot corresponded to seeds from a single genotype  
100 produced at one location. Three independent biological replicates were performed for each  
101 seed lot, resulting in a total of 48 samples (i.e. three replicates of eight genotypes produced at  
102 two locations).

103 To collect spermosphere samples from germinating seeds, 30 unsterilised seeds were used for  
104 each sample and imbibed at four distinct time intervals: 2, 8, 16, and 24 hours. The weight of  
105 the 30 seeds was determined before placing them in a 50 mL Falcon tube. The seeds were  
106 then imbibed with 1.5 times their total weight in sterile Milli-Q water. Imbibition was carried out  
107 with the samples (including blanks) placed on a platform shaker, agitated at 70 rpm, and  
108 maintained at 25°C (Novotron INFORS AG, CH-4103 Bottmingen). The spermosphere,  
109 consisting of the water surrounding the germinating seeds, was collected at each time interval  
110 and pooled (250 µL per time interval) in 2 mL Eppendorf tubes for each sample. A blank sample  
111 of sterile Milli-Q water in a 50 mL Falcon tube was also included in the analyses.

112  
113 The pooled spermosphere samples collected during the germination time course represent the  
114 interface between the seed and its environment. Each sample (1 mL) along with the blank was  
115 immediately flash-frozen in liquid nitrogen and stored at -80°C for subsequent analysis.

116 Add Figure 1 here.

## 117 Extraction of peptides from spermosphere and untargeted LC-MS/MS 118 peptidomic analysis

119 The frozen pools of each spermosphere samples were thawed at 4°C. Each sample (parent  
120 samples) were subsequently separated into three equal parts of 150 µL (daughter samples) in  
121 2 mL Eppendorf tubes. To maximise peptide extraction diversity, three different extraction  
122 solvents were used for each daughter sample. Specifically, 1350 µL of the following solvents  
123 were added separately to each daughter sample: (i) 0.1% formic acid, (ii) 0.1% trifluoroacetic  
124 acid, and (iii) a methanol-water-acetic acid mixture (90/9/1 v/v/v). The daughter samples with  
125 their respective extraction solvents were stirred at 24 rpm on a rotator (Mini LabRoller™ Rotator,  
126 LabNet) at 4°C for 30 minutes. Following extraction, the samples were centrifuged at 10,000 g  
127 to separate the soluble peptides, which were found in the supernatant, from the insoluble  
128 components (precipitate). The supernatants from the daughter samples were pooled to create  
129 a combined peptide extract for each parent sample, which was then dried using a SpeedVac  
130 vacuum concentrator (Thermo Scientific™ SPD 111V) (Figure 2).

131  
132 Before nano-LC fragmentation analysis, peptide samples were desalted and concentrated using  
133 a µC18 Omix (Agilent). Briefly, the dried peptide samples were resuspended in 40 µL of 0.1%  
134 formic acid. These resuspended samples were then loaded onto a µC18 Omix column for  
135 desalting and concentration and subsequently eluted with 80% acetonitrile containing 0.1%  
136 formic acid. The eluted peptides were dried using a SpeedVac vacuum concentrator. The dried,  
137 concentrated peptide samples were then resuspended in 10 µL of 0.1% formic acid. One  
138 microliter of this resuspended peptide extract was injected for LC-MS/MS analysis.  
139 Chromatographic separation was performed using a NanoElute (Bruker Daltonics) ultra-high-  
140 pressure nano flow chromatography system. Peptide samples were first concentrated on a C18  
141 PepMap 100 (5 mm x 300 µm i.d.) precolumn (Thermo Scientific) and then separated at 50°C  
142 on a reversed-phase Reprosil column (25 cm x 75 µm i.d.) packed with 1.6 µm C18-coated  
143 porous silica beads (Ionopticks). The mobile phases were composed of 0.1% formic acid in  
144 99.9% water (v/v) (Phase A) and 0.1% formic acid in 99.9% acetonitrile (v/v) (Phase B). The  
145 nanoflow rate was set at 300 nL/min, with the following gradient profile: from 2% to 15% Phase  
146 B within 15 minutes, increasing to 25% Phase B within 10 minutes, then to 37% Phase B within  
147 12 minutes, followed by a decrease to 9% Phase B within 7 minutes and re-equilibration.

148  
149 MS experiments were conducted on a TIMS-TOF Pro mass spectrometer (Bruker Daltonics)  
150 equipped with a modified nano electrospray ion source (CaptiveSpray, Bruker Daltonics).  
151 Ionisation was typically achieved with a spray voltage of 1600 V and a capillary temperature of  
152 180°C. MS spectra were acquired in positive mode over a mass range of 100 to 1700 m/z. The  
153 mass spectrometer operated in Parallel Accumulation–Serial Fragmentation (PASEF) mode  
154 without exclusion of singly charged peptides. A total of 10 PASEF MS/MS scans were  
155 performed over a 1.25-second cycle. The mass spectrometry proteomics raw spectral data have  
156 been deposited in the ProteomeXchange Consortium  
157 (<https://proteomecentral.proteomexchange.org>)<sup>19</sup> via the iProX partner repository<sup>20,21</sup>, with the  
158 dataset identifier **PXD051625**<sup>22</sup>.

## 159 160 Peptide identification

161  
162 Peptide identification was performed using PEAKS XPro software (Bioinformatic Solutions Inc.,  
163 CA) against an updated version of the UniProt *Phaseolus vulgaris* database (Proteome ID:  
164 UP000000226; The UniProt Consortium, 2022) (Figure 2). The following modifications were  
165 included: methionine oxidation, N-terminal acetylation, deamidation, and carbamylation. The  
166 "no enzyme" parameter was selected for the analysis. Mass accuracy was set to 20 ppm for  
167 MS mode and 0.05 Da for MS/MS mode. Data were filtered with a false discovery rate (FDR)  
168 of 5%, and protein redundancy was eliminated based on proteins being evidenced by the same  
169 set or a subset of peptides.

## 170 171 Data pre-treatment, normalisation and statistical analysis

172

173 Each detected peptide ion feature was assigned a unique identifier based on its sequence,  
174 mass, retention time (RT), and post-translational modifications (PTMs). Peptide duplicates were  
175 then filtered out. Only peptide features detected in all three biological replicates were retained  
176 for further analysis. Precursor proteins were filtered based on the retained peptide features.  
177 Peak intensities of the retained peptides and proteins were normalised either by the sum of all  
178 peptide intensities of each sample or by the sum of all peptide intensities and corrected by the  
179 dry weight of 30 seeds. (Figure 3). The complete lists of peptides, including both unfiltered and  
180 filtered datasets from the analysis, are publicly available in the *Recherche Data Gouv*  
181 multidisciplinary repository<sup>23</sup> at <https://entrepot.recherche.data.gouv.fr/dataverse/inrae>, with the  
182 unique DOI: <https://doi.org/10.57745/8CC18X<sup>24</sup>>.

183  
184 Principal Component Analysis (PCA) was performed on the peptidomic dataset using R  
185 (v.4.2.1). For peptide features that were detected but not quantified (peak intensities of 0),  
186 values were replaced by the smallest peak intensity in the dataset divided by 20. Peptide  
187 features that were not detected (empty cells) were assigned a peak intensity of 0. Two  
188 normalisation approaches were applied to the filtered peptide datasets. The first approach  
189 involved normalising each peptide peak intensity by the sum of all peptide intensities per  
190 sample. The second approach involved normalising each peptide intensity by the sum of all  
191 peptide intensities, with an additional correction for the aggregate dry weight of seeds per  
192 sample. Principal Component Analyses (PCA) were then generated from the normalised  
193 datasets (Figure 7a and 7b).

194  
195

196 **Add Figure 2 here**

197

## 198 **Data Records**

199

200 The mass spectrometry proteomics raw spectral data is available at the ProteomeXchange  
201 online platform (<https://proteomecentral.proteomexchange.org>) via the iProX partner repository,  
202 under the dataset identifier **PXD051625<sup>22</sup>**. The complete lists of peptides from the analysis are  
203 available in the *Recherche Data Gouv* multidisciplinary repository at  
204 <https://entrepot.recherche.data.gouv.fr/dataverse/inrae>, with the unique DOI:  
205 <https://doi.org/10.57745/8CC18X<sup>24</sup>>.

206

207 The peptidomic datasets of the spermosphere samples from eight common bean genotypes  
208 produced at two locations, analysed by the Proteogen platform (University of Caen Normandy;  
209 <https://www.unicaen.fr/laboratoire/proteogen/>), are available under the name  
210 "Common\_Bean\_Spermosphere\_Peptidomic\_Datasets"<sup>24</sup>. The zip archive includes eight files,  
211 as detailed in Table 1.

212

213 The "Raw peptide dataset.xlsx" file contains a list of all peptide features identified during the  
214 peptidomic analysis, while the "Raw precursor protein dataset.xlsx" file includes the precursor  
215 proteins (protein accessions from the UniProt database of *Phaseolus vulgaris* L.) corresponding  
216 to the identified peptides<sup>24</sup>. Details about the content of these files are provided in Table 1.

217

218 **Add Table 1 here**

219

220 The "Filtered peptide dataset.xlsx" and "Filtered precursor protein dataset.xlsx" files contain the  
221 filtered and processed data from the untargeted peptidomic analysis, respectively<sup>24</sup>. The  
222 manual preprocessing steps used to generate these files are summarised in Table 1. A total of  
223 3,258 peptide features were identified across the 48 common bean spermosphere samples,  
224 with substantial variability in the identified peptides observed across the eight genotypes  
225 produced in the two distinct locations (Figure 3a). The identified peptide sequences ranged in  
226 length from 5 to 48 amino acids (Figure 3b), and their masses ranged from 429.2587 Da to  
227 4493.0254 Da. These peptides were derived from 414 precursor proteins, with precursor protein  
228 masses ranging from 7,291 Da to 401,866 Da (Figure 4a). Most of the precursor proteins (60%)  
229 were annotated and had descriptions in the UniProt database for *Phaseolus vulgaris* (Proteome  
230 ID: UP000000226; The UniProt Consortium, 2022) (Figure 4b).

231

232 **Add Figure 3 here**

233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291

#### **Add Figure 4 here**

The “Filtered peptide normalised dataset 1.csv” and “Filtered peptide normalised dataset 2.csv” files contains the normalised peak intensities of the data from the “Filtered peptide dataset.xlsx”<sup>24</sup>. The normalisation steps are described in detail in Table 1. Principal Component Analysis (PCA) was performed on this normalised peptidomic dataset (Figure 7). The “Metadata.xlsx” file provides detailed information on the 48 common bean spermosphere samples analysed in the experiment<sup>24</sup>.

### **Technical Validation**

- **Peptides identified in all three biological replicates**

To ensure the robustness of the peptidomic dataset, only peptide features that were consistently identified in all three biological replicates of each genotype were included. This approach was adopted to enhance the reliability and reproducibility of the dataset by focusing exclusively on peptides that appeared consistently across all replicates, thereby minimising bias from technical variations (Figure 5).

#### **Add Figure 5 here**

- **Peptide and protein identification significance score assuring ensuring dataset robustness**

In mass spectrometry-based peptidomic and proteomic studies, the  $-10\log P$  score (a.k.a. the significance score) is used to assess the confidence of peptide identifications<sup>25</sup>. PeaksXPro calculates this score based on a p-value (probability score P), which helps in interpreting peptide scoring for identification purposes<sup>25</sup>. The p-value reflects the probability that a false identification in the search yields the same or a better matching score. The  $-10\log P$  score, which is the negative base-10 logarithm of this probability, quantifies the likelihood that a peptide-spectrum match (PSM) is a genuine match rather than a random occurrence. A higher  $-10\log P$  value indicates a more significant match. For example, a p-value of 0.05 corresponds to a  $-10\log P$  score of 13, while a p-value of 0.01 corresponds to a  $-10\log P$  score of 20. In our study, peptides and protein features identified in the spermosphere of germinating common bean seeds have  $-10\log P$  scores<sup>25</sup> greater than 15, indicating a high level of confidence in their identification and thus ensuring the robustness of our datasets (Figure 6a and b).

#### **Add Figure 6 here**

- **Principal Component Analysis (PCA)**

PCA based on normalised peptide intensities effectively grouped biological replicates together, demonstrating a consistent clustering across all samples independently of the normalisation approach (Figure 7a and 7b). Compared to other genotypes, the analysis revealed that the VAN, FLA and VEZ genotypes exhibited a more distinct peptidome composition between samples produced from the two locations (“Gers” and “Maine et Loire”).

#### **Add Figure 7 here**

### **Usage Notes**

During seed germination, molecules and macromolecules are released into the surrounding environment<sup>2</sup>. Given that proteins are highly accumulated in seeds<sup>26</sup>, it is anticipated that they are among the compounds exuded. However, there is limited research on proteins in this context<sup>27</sup>, and consequently, there is scant information regarding the peptides and their origins in the spermosphere.

292  
293  
294  
295  
296  
297  
298  
299  
300

Here, we have characterised for the first time the diversity of peptides in the spermosphere using eight different common bean genotypes produced in two distinct locations. Peptides are known to play diverse roles in plant growth, development, and stress responses<sup>28</sup>. Additionally, their specific structures are thought to confer a wide range of bioactivities<sup>28</sup>. Our dataset, which details the peptide diversity in the spermosphere, may contribute to the scientific community's understanding of how seeds interact with their environment during germination through peptide-mediated processes<sup>29,30</sup>.

## 301 **Code Availability**

302  
303  
304  
305  
306  
307

The Principal Component Analysis R script, named "PCA Script.R", is available at *Recherche Data Gouv* public repository with unique DOI: <https://doi.org/10.57745/8CC18X><sup>24</sup>. Analyses were performed using R v4.2.1.

## 308 **Acknowledgements**

309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319

This work was also part of the 3<sup>rd</sup> Programme for Future Investments (France2030) and conducted under the SUCSEED project (ANR-20-PCPA-0009), which is funded by the 'Growing and Protecting Crops Differently' French Priority Research Program (PPR-CPA), as part of the national investment plan managed by the French National Research Agency (ANR). Additionally, this study benefited from the support of the IJPB Plant Observatory platforms, including the Chemistry and Metabolism platform (PO-Chem) and the Biochemistry platform (PO-Bioch). The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

## 320 **Author contributions**

321  
322  
323  
324  
325  
326  
327  
328  
329  
330

L.R. and C.S. conceptualised the study. C.S., C.B., B.C., D.S., and T.F. performed the experimental work for sample preparation. B.B. was responsible for the peptidome mass spectrometry data acquisition. C.S. conducted the analysis, interpretation, and drafting of the manuscript under the supervision of B.B., M.C., and L.R. All authors approved the final version of the manuscript and are accountable for all aspects of the work, ensuring that any questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All persons designated as authors qualify for authorship, and all those who qualify for authorship are listed.

## 331 **Competing interests**

332  
333  
334

The authors declare no conflict of interests.

## 335 **Figures**

336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349

### 337 **Figure 1: Workflow of experimental design for spermosphere collection and analysis.**

Spermosphere collection involved imbibing 30 unsterilised seeds per sample at four distinct time intervals: 2, 8, 16, and 24 hours. The surrounding water of the germinating seeds, representing the spermosphere, was collected at each time point and pooled for subsequent peptidomic analysis.

### 343 **Figure 2: Peptidomic analysis workflow.**

This pipeline illustrates the steps involved in analysing the peptidome from the spermosphere of germinating common bean seeds.

### 347 **Figure 3: Characteristics of peptides in the spermosphere of germinating common beans.**

(a) Histograms displaying the number of peptides identified in the spermosphere for each

350 common bean genotype produced in two different regions. Only peptides identified across all  
351 three biological replicates were included for each genotype and location.  
352 **(b)** Histogram showing the distribution of peptide lengths in the spermosphere of eight common  
353 bean genotypes from two production regions.

354  
355 **Figure 4: Characteristics of precursor proteins in the spermosphere of germinating**  
356 **common beans.**

357 **(a)** Histogram showing the mass distribution of precursor proteins associated with identified  
358 peptides in the spermosphere of eight common bean genotypes from two production regions.  
359 **(b)** Proportion of characterised precursor proteins identified through a search in the *Phaseolus*  
360 *vulgaris* UniProt database.

361  
362 **Figure 5: Pre-processing of peptidome dataset.**  
363 Venn diagrams showing the overlap of identified peptides in biological replicates (B, C, and D)  
364 for each genotype at each production location. Only peptides identified in all three biological  
365 replicates at a given location were retained.

366  
367 **Figure 6: Peptides and protein identification scores.**  
368 Histograms showing the distribution of the  $-10\log P$  score (a.k.a. significance score), which  
369 measures the confidence in peptide and protein identifications.

370 **(a)** Histogram of significance scores for retained identified peptides.  
371 **(b)** Histogram of significance scores for retained protein precursors.

372  
373 **Figure 7: Principal component analysis (PCA) of peptide intensities.**  
374 PCA is based on normalised peak intensities of identified peptides. Different colours represent  
375 the eight common bean genotypes, while different shapes indicate the seed production  
376 locations.

377 **(a)** Single peptide intensity normalised by the sum of all peptide intensities per sample.  
378 **(b)** Single peptide intensity normalised by both the sum of all peptide intensities and  
379 aggregate dry weight of seeds per samples.

380  
381

## 382 Tables

383  
384 **Table 1: Data files from peptidomic data acquisition and pre-processing of spermosphere**  
385 **samples from germinating common beans.**

386 This table lists the six files associated with the peptidomic data acquisition and pre-processing  
387 for spermosphere samples from eight common bean genotypes produced in two locations.  
388 Details on the content of each file and the data processing steps applied (if any) are provided.

389

## 390 References

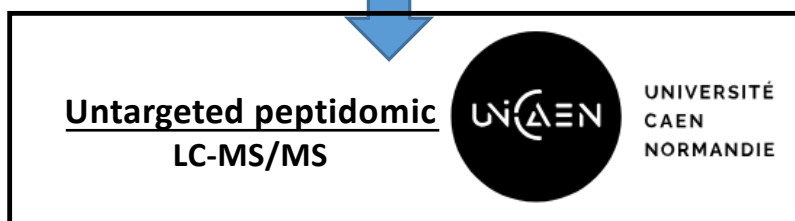
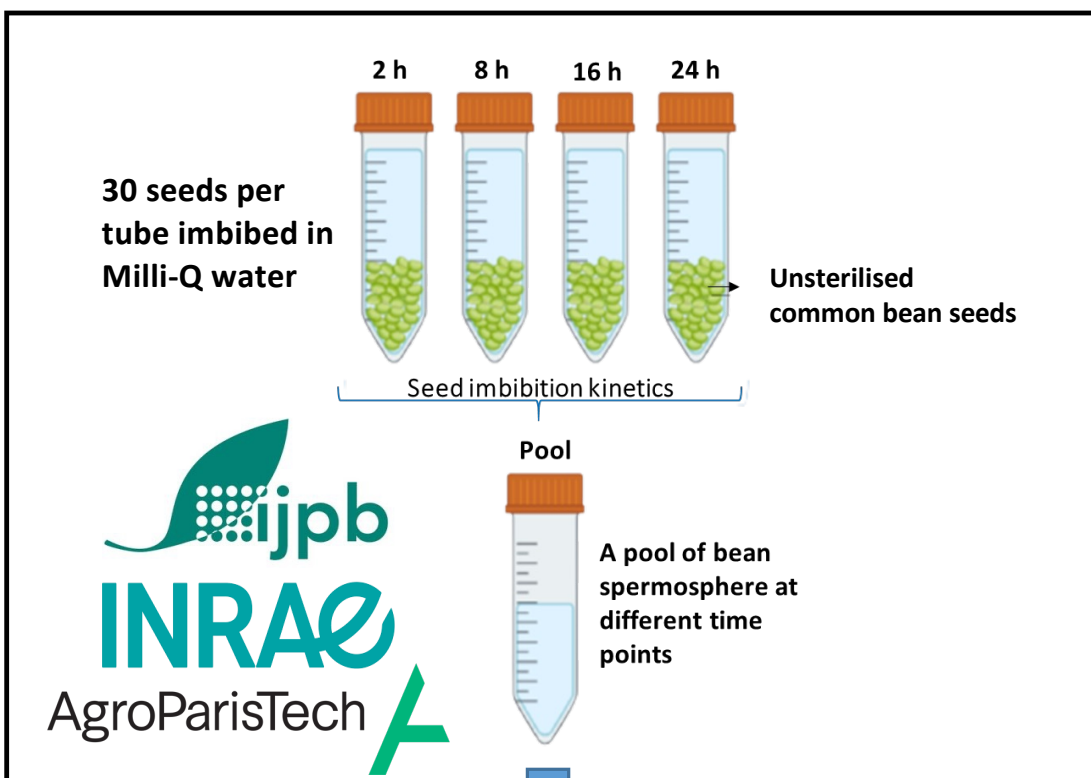
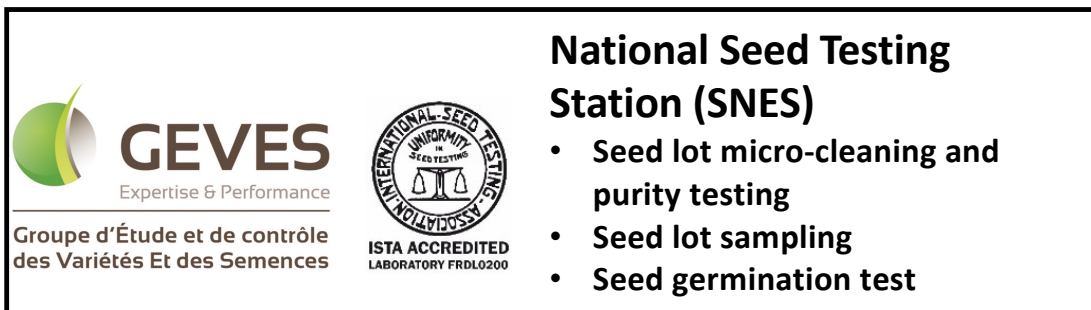
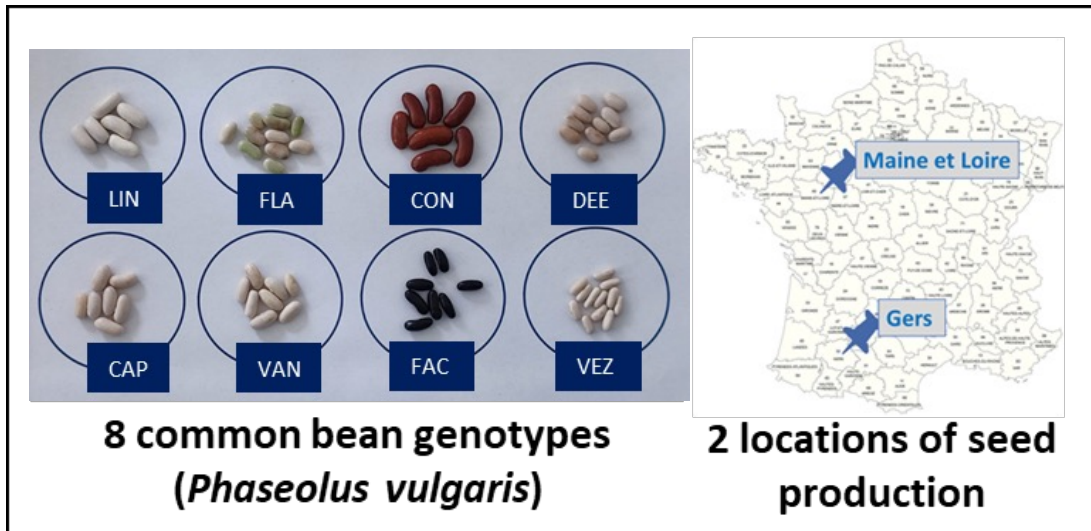
- 391  
392 1. Verona, O. [The spermosphere]. *Ann. Inst. Pasteur* **95**, 795–798 (1958).
- 393 2. Nelson, E. B. The seed microbiome: Origins, interactions, and impacts. *Plant Soil* **422**, 7–34  
394 (2018).
- 395 3. Sasse, J., Martinoia, E. & Northen, T. Feed Your Friends: Do Plant Exudates Shape the Root  
396 Microbiome? *Trends Plant Sci.* **23**, 25–41 (2018).
- 397 4. Bai, B. *et al.* The root microbiome: Community assembly and its contributions to plant fitness.  
398 *J. Integr. Plant Biol.* **64**, 230–243 (2022).
- 399 5. Yu, P. *et al.* Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize  
400 performance under nitrogen deprivation. *Nat. Plants* **7**, 481–499 (2021).

- 401 6. Scarafoni, A. *et al.* The proteome of exudates from germinating *Lupinus albus* seeds is  
402 secreted through a selective dual-step process and contains proteins involved in plant  
403 defence. *FEBS J.* **280**, 1443–1459 (2013).
- 404 7. Jain, M., Amera, G. M., Muthukumar, J. & Singh, A. K. Insights into biological role of plant  
405 defense proteins: A review. *Biocatal. Agric. Biotechnol.* **40**, 102293 (2022).
- 406 8. Li, J., Hu, S., Jian, W., Xie, C. & Yang, X. Plant antimicrobial peptides: structures, functions,  
407 and applications. *Bot. Stud.* **62**, 5 (2021).
- 408 9. Ayala-Niño, A. *et al.* Sequence Identification of Bioactive Peptides from Amaranth Seed  
409 Proteins (*Amaranthus hypochondriacus* spp.). *Molecules* **24**, 3033 (2019).
- 410 10. Marcus, J. P., Green, J. L., Goulter, K. C. & Manners, J. M. A family of antimicrobial  
411 peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia*  
412 kernels. *Plant J.* **19**, 699–710 (1999).
- 413 11. Hill, P. W., Farrell, M. & Jones, D. L. Bigger may be better in soil N cycling: Does rapid  
414 acquisition of small l-peptides by soil microbes dominate fluxes of protein-derived N in soil?  
415 *Soil Biol. Biochem.* **48**, 106–112 (2012).
- 416 12. Los, F. G. B., Zielinski, A. A. F., Wojeicchowski, J. P., Nogueira, A. & Demiate, I. M.  
417 Beans (*Phaseolus vulgaris* L.): whole seeds with complex chemical composition. *Curr. Opin.*  
418 *Food Sci.* **19**, 63–71 (2018).
- 419 13. Singh, N., Jain, P., Ujainwal, M. & Langyan, S. Escalate protein plates from legumes for  
420 sustainable human nutrition. *Front. Nutr.* **9**, 977986 (2022).
- 421 14. Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. The Rules of Engagement  
422 in the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).
- 423 15. Servín-Garcidueñas, L. E. *et al.* Symbiont shift towards *Rhizobium* nodulation in a  
424 group of phylogenetically related *Phaseolus* species. *Mol. Phylogenet. Evol.* **79**, 1–11  
425 (2014).
- 426 16. Bag, S., Mondal, A., Majumder, A., Mondal, S. K. & Banik, A. Flavonoid mediated  
427 selective cross-talk between plants and beneficial soil microbiome. *Phytochem. Rev.* **21**,  
428 1739–1760 (2022).
- 429 17. Lepetit, M. & Brouquisse, R. Control of the rhizobium–legume symbiosis by the plant  
430 nitrogen demand is tightly integrated at the whole plant level and requires inter-organ  
431 systemic signaling. *Front. Plant Sci.* **14**, 1114840 (2023).

- 432 18. Torres-Cortés, G. *et al.* Functional Microbial Features Driving Community Assembly  
433 During Seed Germination and Emergence. *Front. Plant Sci.* **9**, 902 (2018).
- 434 19. Deutsch, E. W. *et al.* The ProteomeXchange consortium at 10 years: 2023 update.  
435 *Nucleic Acids Res.* **51**, D1539–D1548 (2022).
- 436 20. Ma, J. *et al.* iProX: an integrated proteome resource. *Nucleic Acids Res.* **47**, D1211–  
437 D1217 (2019).
- 438 21. Chen, T. *et al.* iProX in 2021: connecting proteomics data sharing with big data. *Nucleic*  
439 *Acids Res.* **50**, D1522–D1527 (2021).
- 440 22. Bernay, B. & Rajjou, L. Characterization of the peptidome in the spermosphere of  
441 common bean (*Phaseolus vulgaris*). ProteomeXchange  
442 <https://identifiers.org/px:PXD051625> (2024).
- 443 23. Dzalé Yeumo, E. Data INRAE – The Networked Repository. in *Research Data Sharing*  
444 *and Valorization* 63–75 (John Wiley & Sons, Ltd, 2022). doi:10.1002/9781394163410.ch4.
- 445 24. Saccaram, C., Bernay, B. & Rajjou, L. Untargeted peptidomic (LC-MS/MS) analysis of  
446 the spermosphere of eight common bean genotypes (*Phaseolus vulgaris* L.). DataGOUV  
447 <https://doi.org/10.57745/8CC18X> (2024).
- 448 25. Zhang, J. *et al.* PEAKS DB: De Novo Sequencing Assisted Database Search for  
449 Sensitive and Accurate Peptide Identification. *Mol. Cell. Proteomics MCP* **11**, M111.010587  
450 (2012).
- 451 26. Fujiwara, T., Nambara, E., Yamagishi, K., Goto, D. B. & Naito, S. Storage proteins.  
452 *Arab. Book* **1**, e0020 (2002).
- 453 27. Rocha, R. O. *et al.* Proteome of Soybean Seed Exudates Contains Plant Defense-  
454 Related Proteins Active against the Root-Knot Nematode *Meloidogyne incognita*. *J. Agric.*  
455 *Food Chem.* **63**, 5335–5343 (2015).
- 456 28. Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I. & Cammue, B. P. A. The  
457 Plant Peptidome: An Expanding Repertoire of Structural Features and Biological Functions.  
458 *Plant Cell* **27**, 2095–2118 (2015).
- 459 29. Marmioli, N. & Maestri, E. Plant peptides in defense and signaling. *Peptides* **56**, 30–  
460 44 (2014).

461 30. Ren, G., Zhang, Y., Chen, Z., Xue, X. & Fan, H. Research Progress of Small Plant  
462 Peptides on the Regulation of Plant Growth, Development, and Abiotic Stress. *Int. J. Mol.*  
463 *Sci.* **25**, 4114 (2024).  
464  
465

# Workflow of seed origins and experimental design



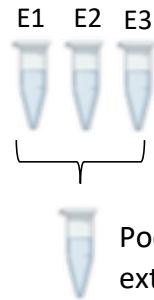
# Peptidomic analysis workflow

## ① Sample collection



Kinetic pooled seed exudates frozen in liquid nitrogen

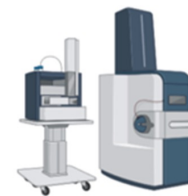
## ② Sample preparation



- Samples thawed at 4°C
- Each samples seperated in 3 parts (150 µL each)
- Blank: Milli-Q water
- Three different extractions carried out on 3 parts of same sample using
  1. (E1) Formic acid (0.1%)
  2. (E2) Trifluoroacetic acid (0.1%)
  3. (E3) Methanol-water-acetic acid (90/9/1 v/v/v)
- Pooling of different extraction solvents for same sample and drying using vaccum concentrator

## ③ Desalation and concentration of peptides

## ④ Peptide separation by liquid chromatography and ion separation by mass analyser



nLC-ESI-MS/MS  
TIMS-TOF Pro

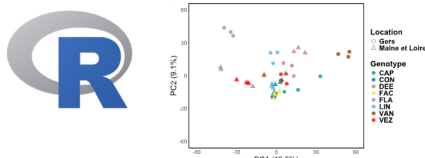
## ⑤ Precursor protein identified by PeaksXPro using UniProt *Phaseolus vulgaris* database



## ⑥ Data pre-treatment and normalisation

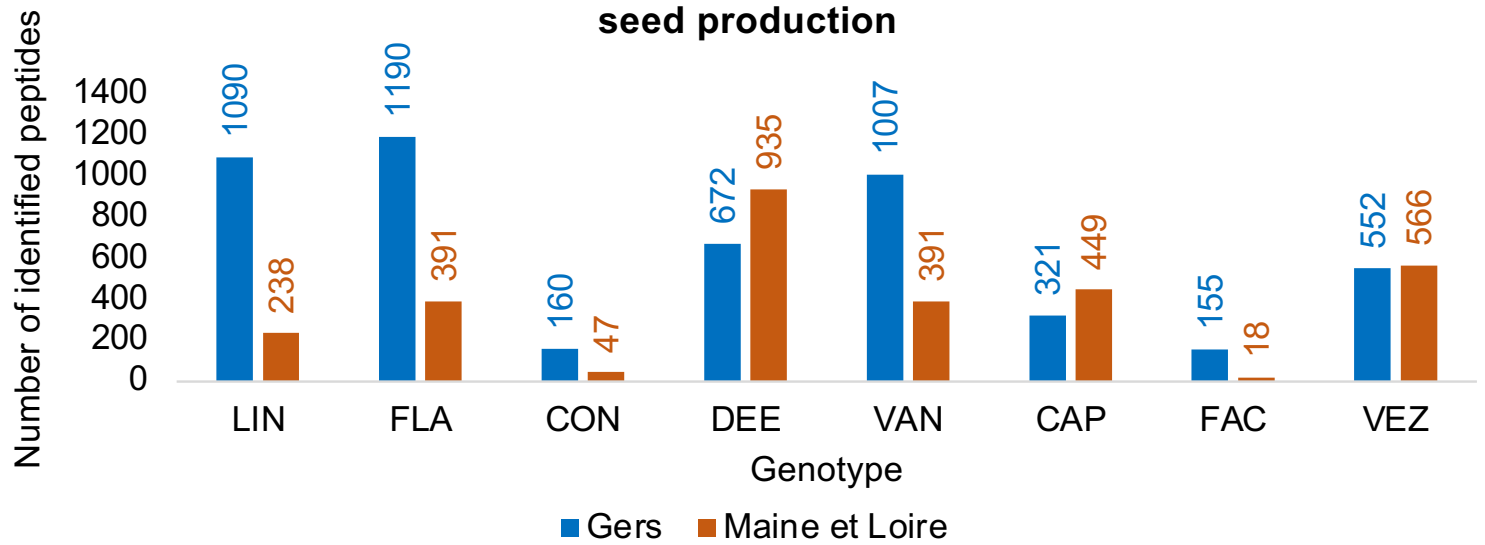
- Raw spectral data uploaded at ProteomeXchange
- Each peptide feature was assigned a unique ID based on sequence, mass, RT and PTM
- Peptides doublons filtered
- Peptide features detected in any 3 biological repetitions retained
- Precursor proteins filtered based on retained peptide features
- Peak intensities of retained peptides were normalised by total peptide intensity corrected or not by the dry seed weight for each sample

## ⑦ Data quality validation by Principal Component Analysis (PCA)



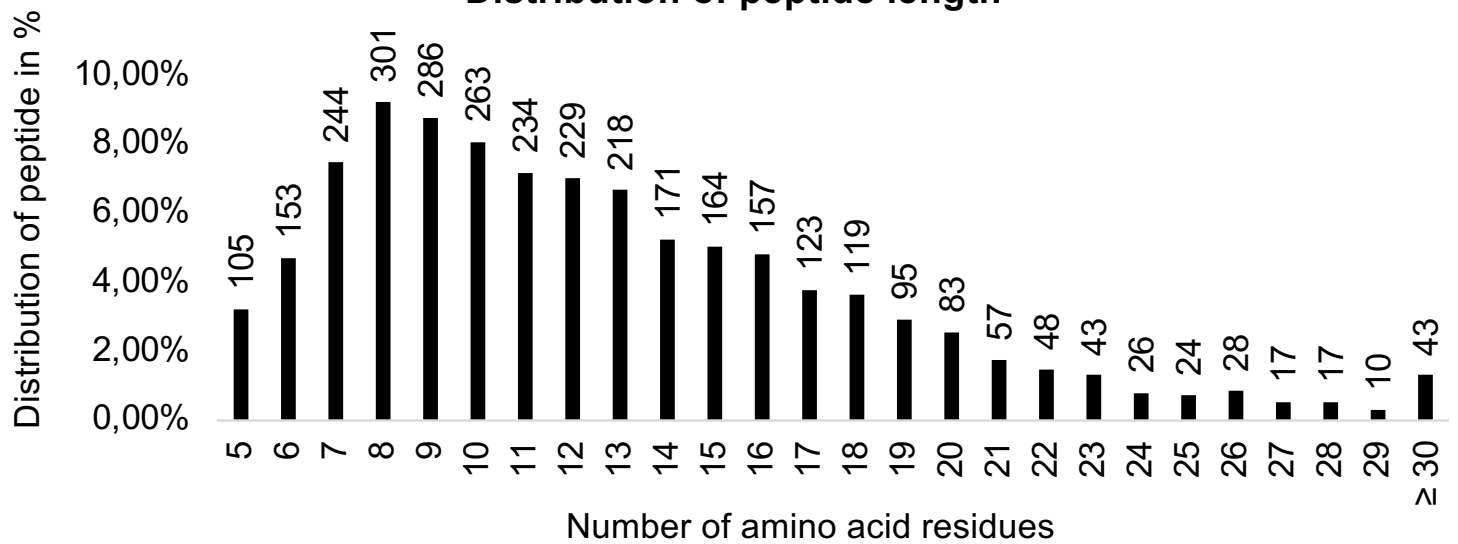
(a)

### Number of identified peptides per genotype in two locations of seed production

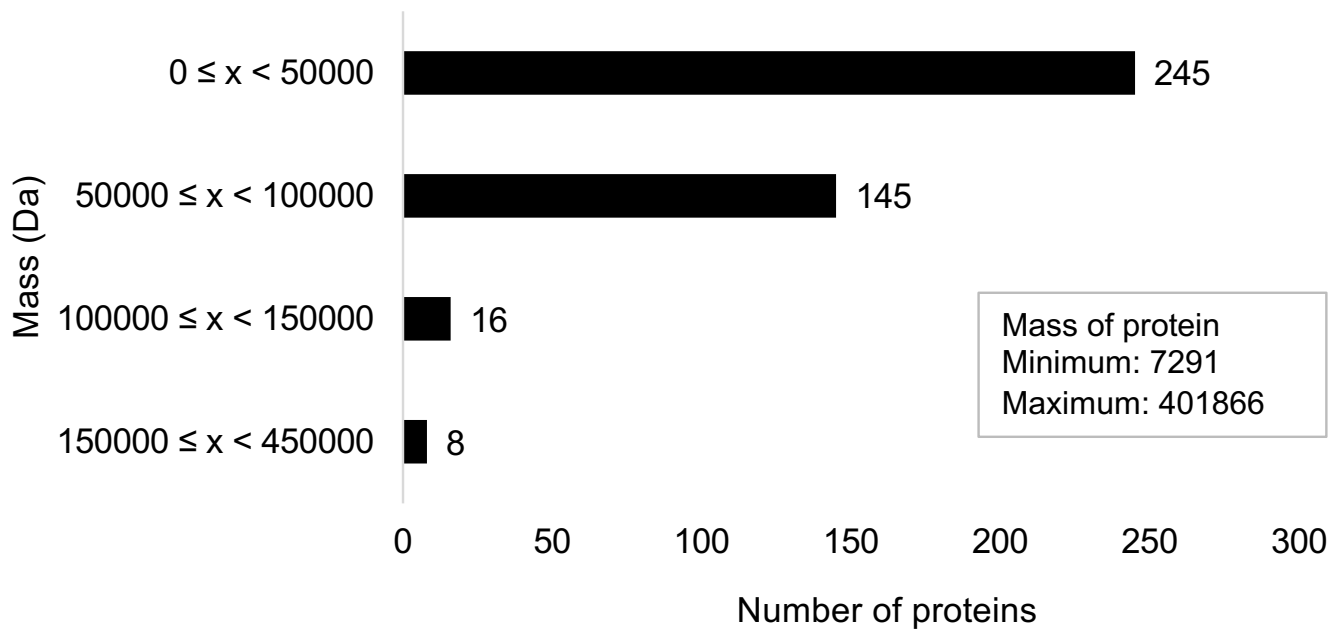


(b)

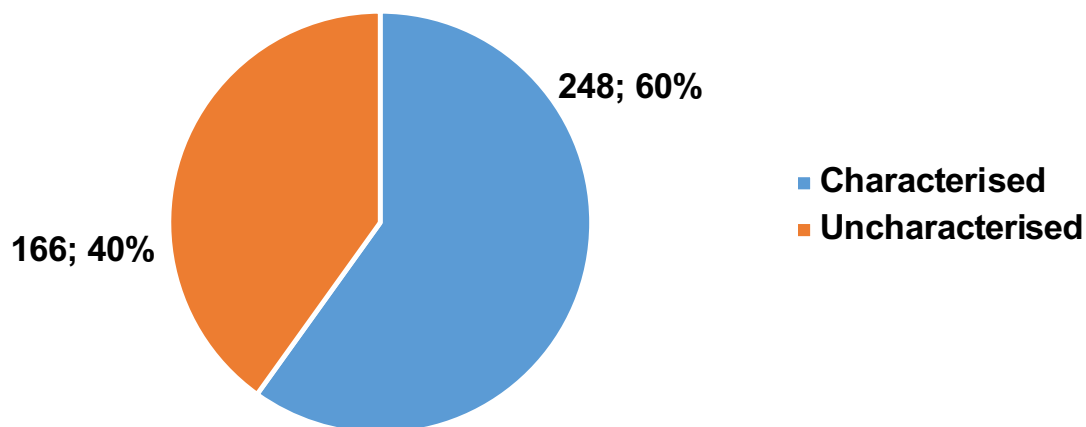
### Distribution of peptide length

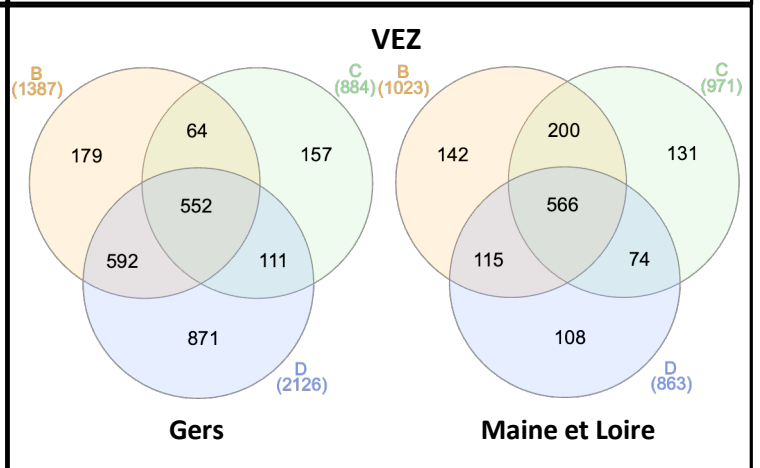
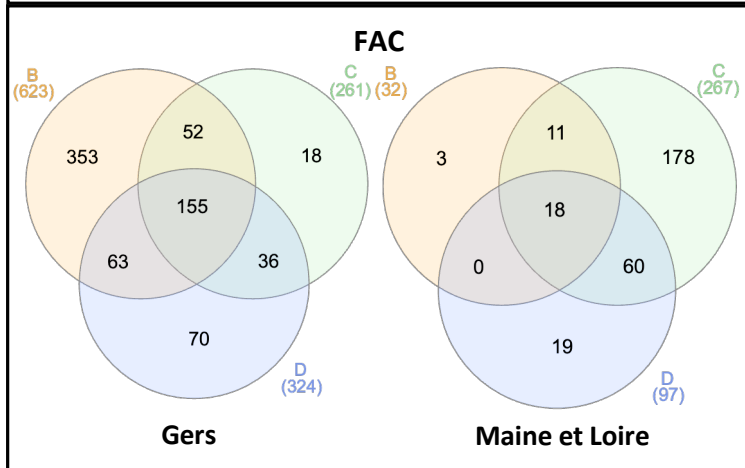
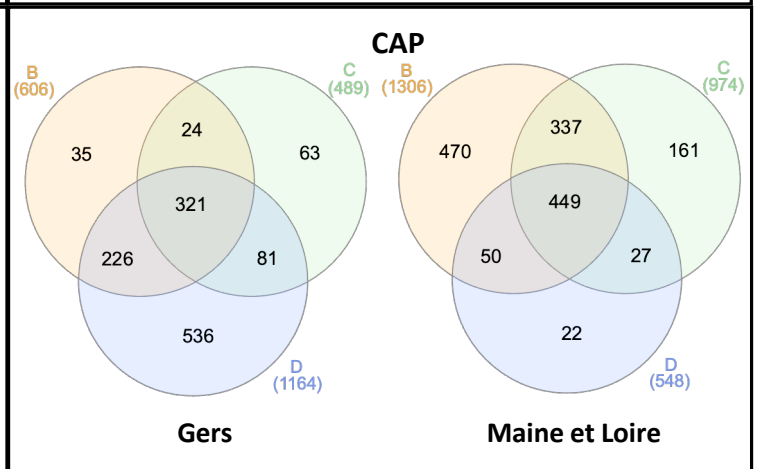
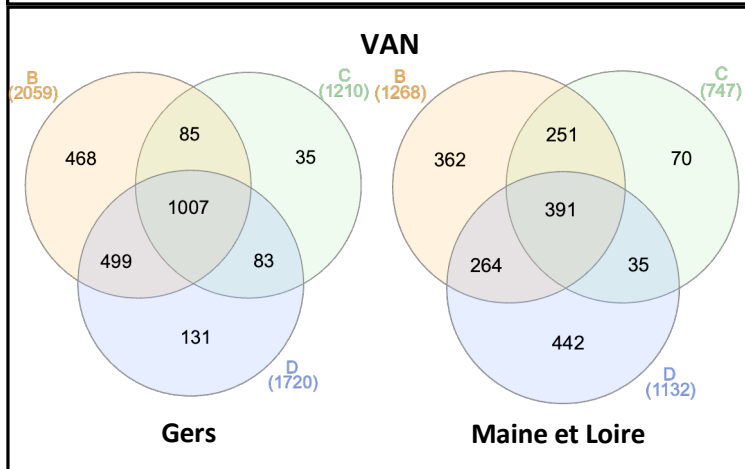
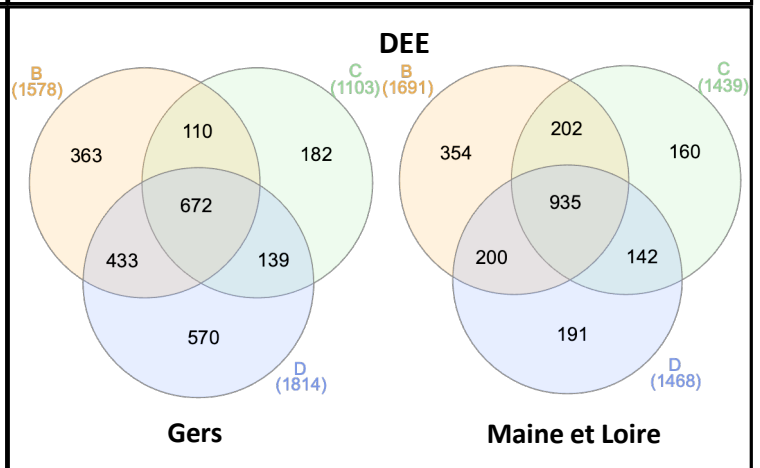
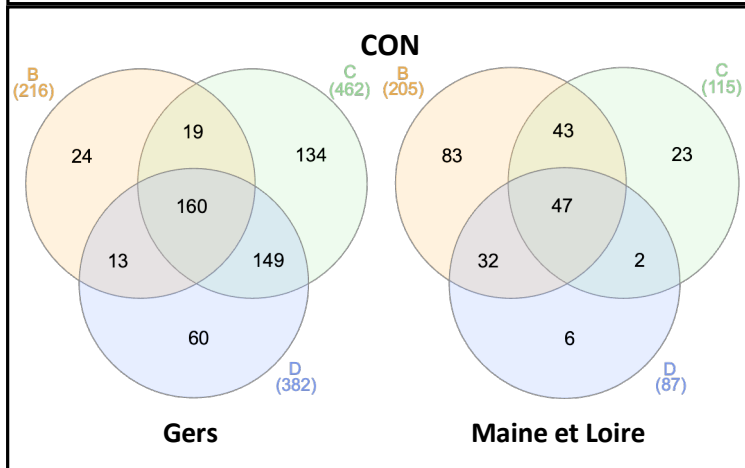
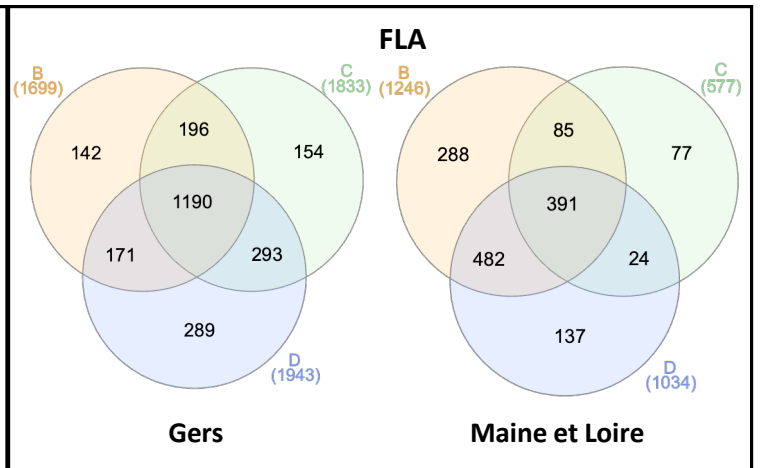
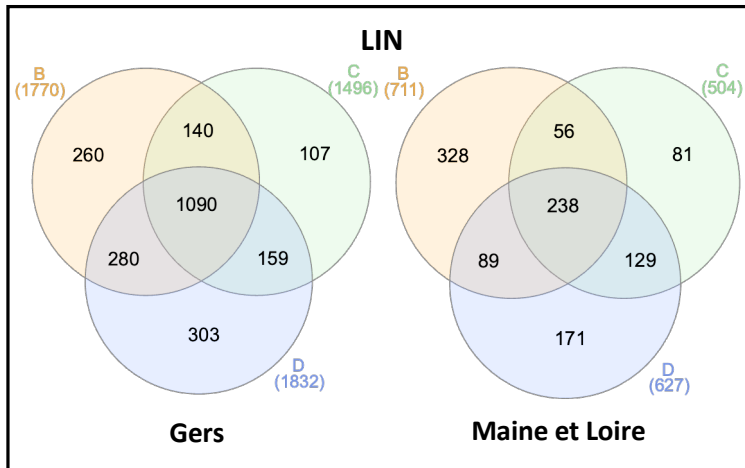


(a) **Distribution of mass (Da) of identified proteins**



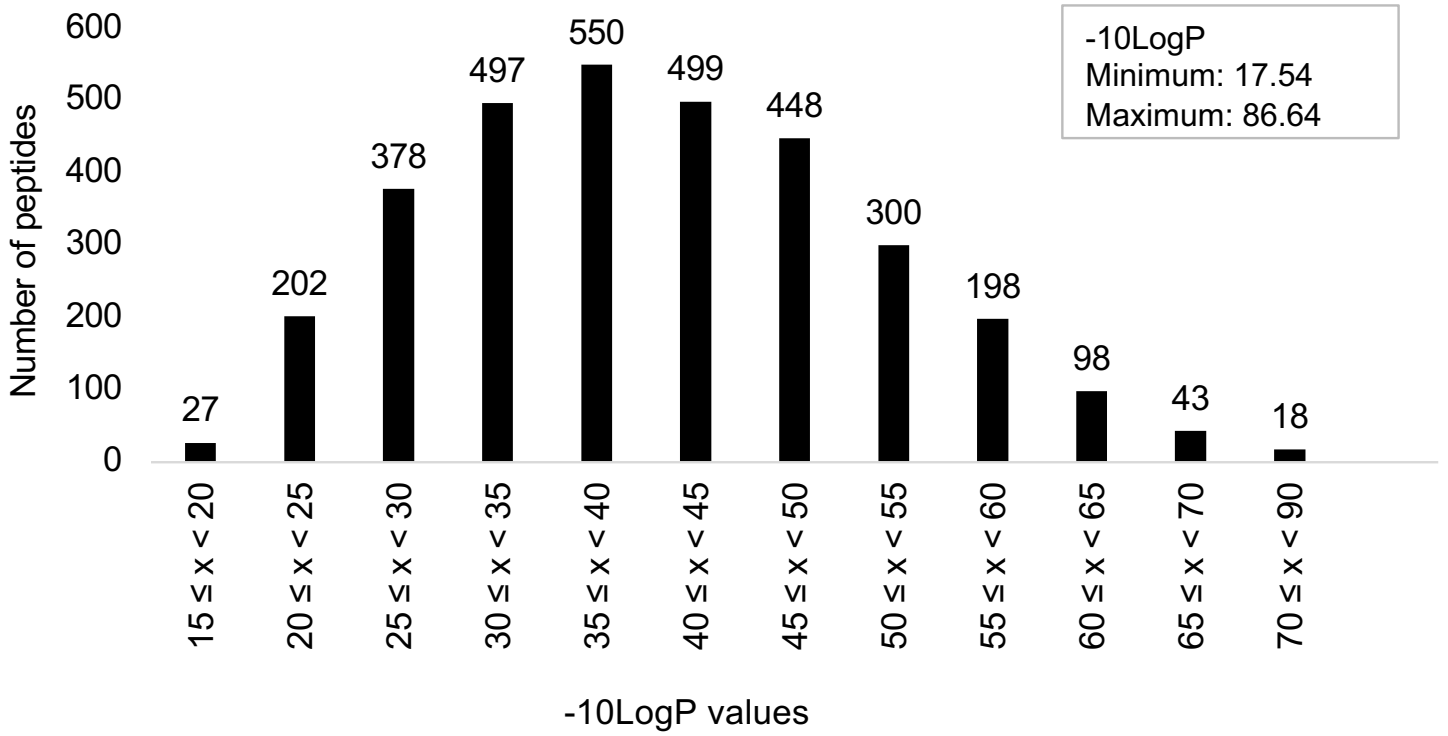
(b) **Description of identified proteins**





(a)

Distribution  $-10\text{LogP}$  values of identified peptides



(b)

Distribution  $-10\text{LogP}$  values of identified peptides

