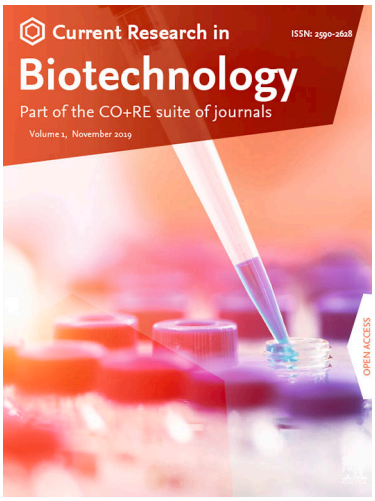
Journal Pre-proofs



FEEDS, the Food wastE biopEptiDe claSsifier: from microbial genomes and substrates to biopeptides function

Victor Borin Centurion, Edoardo Bizzotto, Stefano Tonini, Pasquale Filannino, Raffaella Di Cagno, Guido Zampieri, Stefano Campanaro

PII: S2590-2628(24)00012-1
DOI: https://doi.org/10.1016/j.crbiot.2024.100186
Reference: CRBIOT 100186

To appear in: *Current Research in Biotechnology*

Received Date: 1 November 2023
Revised Date: 27 January 2024
Accepted Date: 30 January 2024

Please cite this article as: V. Borin Centurion, E. Bizzotto, S. Tonini, P. Filannino, R. Di Cagno, G. Zampieri, S. Campanaro, FEEDS, the Food wastE biopEptiDe claSsifier: from microbial genomes and substrates to biopeptides function, *Current Research in Biotechnology* (2024), doi: https://doi.org/10.1016/j.crbiot.2024.100186

FEEDS, the Food wastE biopEptiDe claSsifier: from microbial genomes and substrates to biopeptides function

Victor Borin Centurion[1a*], Edoardo Bizzotto[1b], Stefano Tonini[2c], Pasquale Filannino[3d], Raffaella Di Cagno[2e], Guido Zampieri[1f], Stefano Campanaro[1g*]

[1] Department of Biology, University of Padua, via U. Bassi 58/b, 35131 Padova, Italy

[2]Faculty of Agricultural, Environmental and Food Sciences, University of Bolzano, piazza Università, 5, 39100, Bolzano, Padova, Italy

[3]Department of Soil, Plant and Food Sciences, University of Bari Aldo Moro, Piazza Umberto I, 1, 70121 Bari, Italy

[a]victor.borincenturionbiruel@unipd.it,[b]edoardo.bizzotto@studenti.unipd.it,[c]stefano.tonini@student.unibz.it,[d]pasquale.filannino1@uniba.it,[e]raffaella.dicagno@unibz.it,[f]guido.zampieri@unipd.it,[g]stefano.campanaro@unipd.it

Corresponding authors: victor.borincenturionbiruel@unipd.it, stefano.campanaro@unipd.it

Abstract

The production of biopeptides from food waste through microbial fermentation faces challenges arising from the diverse proteolytic abilities of microorganisms and substrate variability, impacting both the quality and yield of generated biopeptides. To address these challenges, preliminary in-silico bioinformatics analyses play a crucial role in evaluating suitable substrates and proteases for the fermentation process. However, existing tools lack comprehensive predictive capabilities for relevant proteases, substrate performance assessment, and final biopeptide family characterization. To overcome these limitations, we developed FEEDS (Food wastE biopEptiDe claSsifier), a novel biopeptide prediction and classification tool. FEEDS predicts biopeptide families based on microbial genome protease profiles and substrate composition during proteolysis. The tool also employs a machine learning approach for functional biopeptide classification. Results from testing on 1000 microbial genomes demonstrate the effectiveness of biopeptide classification, particularly in categorizing peptides derived from substrates like *Hordeum vulgare* and *Vitis vinifera* seed storage proteins. In addition to biopeptide classification, our study delves into the distinctive protease profiles of bacteria and yeast genomes. Bacterial genomes exhibited 60 to 100 proteases across 40 to 55 families. Contrastingly, yeast genomes displayed a more evenly distributed pattern with 150 to 160 protease-encoding genes across 60 to 67 families, surpassing bacterial counts. Remarkably, a substantial portion of yeast proteases (~66%) was secreted. Moreover, our integration of a machine learning methodology within the FEEDS pipeline proved highly effective, achieving over 80% accuracy in predicting the function of peptides derived from seed storage proteins. Notably, longer peptide sequences exceeding 20 amino acids consistently displayed a higher probability of correct assignment compared to shorter counterparts.

Keywords: Bioactive peptides, proteases, digestion, food waste, bacteria, yeasts.

1. Introduction

Biopeptides are short chains of amino acids that occur naturally in living organisms or are generated through fermentation processes. These small molecules have many applications and are used in various fields [1]. They have been found to exhibit various pharmacological activities such as antimicrobial, anti-inflammatory, antioxidant, and antihypertensive effects [1,2]. Biopeptides have gained popularity due to their multiple functions, including flavor enhancement, food preservation, and due to their physiological effects, such as improving digestion and lowering cholesterol levels [2,3]. In

agriculture, biopeptides are used as plant growth promoters and biopesticides, which are safer alternatives to traditional compounds obtained through chemical synthesis [4]. Biopeptides have also been utilized in cosmetics as active ingredients in anti-aging creams and other skincare products due to their ability to stimulate collagen and elastin synthesis, reduce the appearance of wrinkles, and improve skin hydration [5]. Overall, the versatility of biopeptides makes them valuable in many different fields and their potential applications are still being explored.

Microbial fermentation has emerged as an innovative method for biopeptide production. This approach harnesses the metabolic capabilities of microorganisms to synthesize biopeptides efficiently. The significance of microbial fermentation lies in its scalability, cost-effectiveness, and the ability to tailor the production process for specific biopeptide sequences. This novel production method not only enhances the yield and purity of biopeptides, but also opens avenues for exploring new bioactive compounds with diverse applications [1,3,6]. Expanding beyond biopeptides, microbial fermentation of food waste and byproducts involves microorganisms like bacteria and yeasts. These microorganisms utilize byproducts as growth substrates, digesting organic matter and producing enzymes that break down complex molecules. This comprehensive process underscores the versatility of microbial fermentation in transforming various substrates into valuable end products such as biofuels, bioplastics, and enzymes [6]. The fermentation process can occur in a batch or in continuous mode and can be performed under various conditions, such as aerobic or anaerobic, and at different pH and temperature ranges [7]. The selection of fermentation conditions depends on the specific microorganisms used and on the target compounds to be produced. Some microbial species of the genera *Lactobacillus*, *Bacillus*, *Streptococcus*, *Saccharomyces* and *Candida* have been found to produce biopeptides with various bioactive properties, including antimicrobial, antifungal, antioxidant, and immunomodulatory activities [6,8].

While some studies have explored the use of food-derived proteins as a source of biopeptides [9–11], few have focused the attention on the production of biopeptides specifically from food waste and byproducts [12]. One of the main advantages of using this matrix as a biopeptide source is that it represents a sustainable and low-cost alternative to traditional protein sources. According to a report by the United Nations Food and Agriculture Organization (FAO), the disposal of vegetable waste can substantially impact the environment since it has a substantial carbon footprint [13]. Furthermore, FAO has estimated that approximately one-third of all food produced worldwide is wasted or lost, with a significant proportion occurring in the fruit, vegetable, and seafood industries. Additionally, contaminants such as heavy metals and pesticides in some food waste streams can pose potential risks to human health and the environment [14]. This highlights the need for greater efforts to reduce food waste and increase global sustainability in food production and consumption practices. The recovery of food waste through the extraction of valuable compounds is an attractive approach, however, the generation of some products including biopeptides from food waste presents several challenges. For example, the composition of food waste can vary widely depending on the source, which can affect the quality and yield of the biopeptides produced.

To help encounter these challenges, preliminary in-silico analyses can help the evaluation of what substrates and protease enzymes should be used in the fermentation process. These tools use bioinformatics techniques to screen protein databases and identify potential biopeptide candidates based on specific criteria such as sequence length, physicochemical properties, and known bioactive regions [15]. One commonly used tool for biopeptide prediction is PeptideRanker, which uses machine learning techniques to predict the binding affinity of peptides to major histocompatibility complex (MHC) molecules [16]. Another tool is the food-derived bioactive peptides database (DFBP), which contains a comprehensive collection of experimentally validated bioactive peptides for peptidomics research [17]. In addition to these tools, there are also several software and databases available for predicting and designing biopeptides with specific functions, such as antimicrobial or antitumor activity [18–20]. However, none of these tools is able to combine a range of functions including prediction of proteases in microbial genomes, identification of their targets in the proteins of the substrate, and identification of the biopeptide families generated. Furthermore, all the available tools are online platforms with restrictions regarding the number of sequences for annotations. To address these limitations, we

introduce FEEDS, a food biopeptide classifier tool capable of efficiently predicting the biopeptide families generated through the cleavage site profiles of proteases derived from microorganism genome annotations and the substrate proteins from food sources.

2. Material and Methods

This section provides a detailed description of all the steps involved, while a summary of the FEEDS tool's process is presented in the flowchart depicted in Figure 1. The tool was developed using Python, and installation instructions, as well as all the features of the tool, can be accessed at the following link: https://github.com/vborincenturion/feeds.

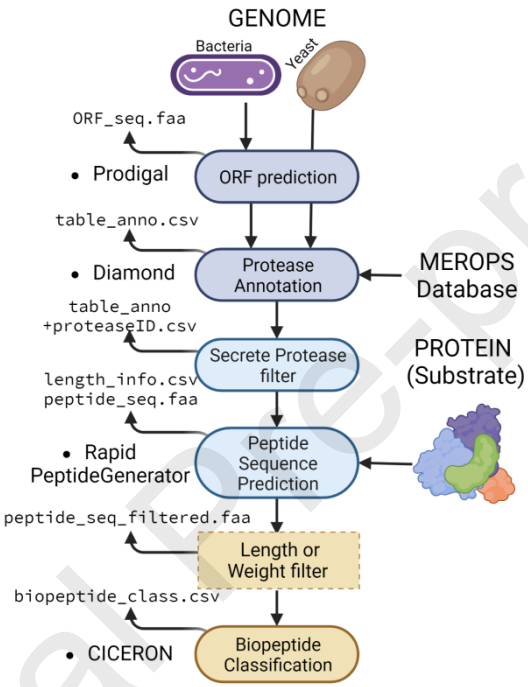


Figure 1. Flowchart of FEEDS tool pipeline. Colors are used to highlight the three main functions of the tool: proteases characterization (purple), simulation of proteases activity (blue) and biopeptides classification (beige). The dashed chart step is optional. Output files are represented by arrows coming out of the chart. External tools implemented in FEEDS are reported on the left and represented by bullet points.

2.1 Step1: Proteases annotation

The first step involves annotation of protein files obtained from bacterial genomes and yeasts for proteases prediction. To achieve this, the protease sequences from the MEROPS database, which employs a hierarchical and structure-based classification system, were used to generate a database using the “makedb” function of Diamond v2.1.4 [21]. For bacteria, the protein-encoding sequences were recovered from genomes using Prodigal v2.6.3 [22]. In the case of yeasts, users must provide the open reading frames (ORFs) prediction file since Prodigal is only applicable to prokaryotic genomes, and other tools like GeneMark [23] can provide reliable predictions only for some yeast species. Moreover, ORF prediction for yeast requires a specific frame rule for each genus and lacks a comprehensive gene-finding tool [24], however FEEDS can flexibly incorporate predictions from new tools. Next, the "blastp" function of the Diamond tool was used with the parameters "--more-sensitive -k 1 -f 6 qseqid

sseqid pident --id 90 --query-cover 85 --subject-cover 85" to align the query sequences to the database previously generated using the sequences collected from MEROPS. The strict identity used in the alignment, and the coverage criteria, ensured that only those proteases with highly significant similarity levels were annotated. Starting from the alignment results obtained, the main proteases information, including function, family, and also the ID number of the cleavage site for the secreted enzymes obtained from the RapidPeptideGenerator (RPG) v2.0.0 (Maillet, 2020) were recovered using Pandas library.

Starting from the entire list of proteases identified, only those potentially secreted were considered for the next step, as they are those with the highest probability to act on the substrates and, for this reason, those having the highest biotechnological relevance [26,27]. To identify the secreted proteases, all the proteins annotated as proteases from the Diamond output were clustered with CD-HIT v4.8.1 [28] with a 90% identity threshold. A single representative sequence from each cluster was analyzed with the BUSCA web server (http://busca.biocomp.unibo.it) [29], using the taxonomic selection of "Prokarya - Gram-positive" or "Eukarya - Fungi". Only protease families including in their annotation the term “extracellular space” compartments were considered as secreted proteases and this result was associated to all the proteases in the cluster. Further details on protease families can be found in Supplementary material 1.

2.2 Step2: Peptide sequence prediction

For each family of secreted proteases the cleavage sites were recovered from MEROPS database (https://www.ebi.ac.uk/merops/cgi-bin/protsearch.pl). After recovering all the cleavage sites, this information was added (if not already present) to the RPG tool using the "rpg -a" function. Peptide predictions can be generated by using either the “sequential” or “concurrent” digestion mode of RPG. “Concurrent” mode simulates the substrate hydrolysis using all enzymes at once, while “sequential” mode performs the simulation utilizing all the enzymes one by one in a sequential order. The mode setting can be defined by setting the "-d" option in FEEDS. The protein substrate files used for RPG digestion are provided by the user as a file in the "substrate" folder. FEEDS has two filtering functions that enable the selection of the predicted peptide sequences generated based on their length or according to their molecular weight (-f_length and -f_mol, respectively). In the final part of this step, a table with the peptide length or molecular weight information is generated. Furthermore, Supplementary material 2 includes a list of all the secreted proteases potentially used by FEEDS in the simulated substrate proteolysis, the cleavage site counts, and the corresponding IDs to be imputed during the processing with the "-e" function of RPG.

2.3 Step3: Functional Biopeptide Classification

The peptide sequences obtained in the previous step are classified using the CICERON tool [30], a novel machine learning method to identify functions of biopeptides obtained from hydrolysis of food protein substrates. This tool employs various methods for the classification of biopeptides including similarity and motif search against a database focused on microbial peptides, and several machine learning methods such as Logistic Regression, Random Forest, K-Nearest Neighbour and Neural Networks. CICERON automatically selects the most accurate method for each of nine functional classes based on prior systematic benchmark evaluations, including peptides with positive effects on vascular circulation, antidiabetic, antihypertensive, antimicrobial, antioxidant, celiac-disease-associated, and immunomodulatory peptides, neuropeptides, and opioids [30]. The underlying models were trained on peptides having a maximum length of 100 amino acids (AA), hence, for optimal results it is highly advisable to utilize this value as maximum length. The different possible functional classification of biopeptides are reported in the Supplementary material 3

2.4 In vitro experiments for validation of FEEDS.

FEEDS was validated comparing the results with those obtained in a "real" fermentation process performed using brewer spent grains (*Hordeum vulgare*) as substrate. In this experiment peptides were identified using mass spectrometry. Prior to inoculation and fermentation, the substrates underwent UV treatment to reduce the contaminating microbial load. This approach was preferred over thermal treatments to avoid damage and denaturation of the proteins. The selected microorganisms for the test were *Enterococcus faecalis* AVEL13, *Lactococcus lactis* WSL2, *Schizosaccharomyces pombe* J13151G1 and *Saccharomyces cerevisiae* KFAY3 and the fermentation time was set to 72 hours. Peptides with a molecular weight lower than less than 90 kilodaltons (~ 90 AA) and containing more than 6 AA were selected for UHPLC/HR-MS2 (UHPLC Ultimate 3000, Thermo Scientific, San Jose, CA, USA; Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer, Thermo Scientific, San Jose, CA, USA) equipped with a C18 column (Acquity UPLC-C18 Reversed-phase, 2.1 × 100 mm, 1.8 µm particle size, Waters Corporation, Milford, MA, USA) mass spectrometry (MS) analysis conducted after 72 hours apart from fermentation using the label-free quantification method. Peptides with fewer than 6 amino acids were excluded from this analysis due to limitations of software and a high-confidence peptide threshold from Proteome Discover 2.3 (Thermo Fisher Scientific, Dreieich, Germany). Based on the ion intensity, in-vitro peptides sequences were predicted using Proteome Discoverer 2.3 coupled with Matrix software (Matrix Science, Boston, MA, USA) for peptide sequencing and identification. The main parameters used for the identification process were: enzyme, no-enzyme; peptide mass tolerance, ±5 ppm; fragment mass tolerance, ±0.1 Da; variable modification, Demetilation (NQ), oxidation (M) and phosphorylation (ST). Peptide and protein identification results were exported after filtering with the Peptide and Protein Validator to achieve a false discovery rate (FDR) below 0.01. The obtained in-vitro proteins were subjected to in-silico digestion using sequential mode of RPG tool and the following microorganisms: *Enterococcus faecalis strain* AT22, *Lactococcus lactis subsp. lactis* LEY7, *Saccharomyces cerevisiae* YJM984, and *Schizosaccharomyces pombe* 972h-. The comparison between the peptides obtained and those predicted using FEEDS was performed using Diamond, with the in-vitro peptide sequences serving as the database and the predicted in-silico peptide sequences as the queries. A minimum threshold of 90% identity (± 4 mismatches) was applied, and the aligned sequences were considered as true matches. The in-silico and in-vitro peptides match sequences can be checked in the Supplementary material 7.

3. Results and Discussion

To gain insights into the most common protease families involved in biopeptide production, we selected 1,182 complete genomes of gram-positive bacteria from the RefSeq database [31]. For yeasts, 157 genomes having associated protein sequence files in the GenBank database were included [32]. The bacterial genera selected were *Enterococcus*, *Fructobacillus*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus*, and *Streptococcus*, all belonging to the Firmicutes phylum. The yeasts genera included were *Candida*, *Debaryomyces*, *Hanseniaspora*, *Kazachstania*, *Metschnikowia*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, and *Zygosaccharomyces*, all belonging to the Ascomycota phylum, with *Rhodotorula* being the only genus from the Basidiomycota phylum. In the next sections these fungal genera will be reported simply as "yeasts". To test FEEDS, the seed storage glutelin, legumin, vicilin, cruciferin, globulin and albumin proteins of *Vitis vinifera*, and the seed storage avenin, B-D-Y hordein, and glutelin proteins of *Hordeum vulgare* were used as substrates. Additional information on all the species, strains and protein substrates can be found in Supplementary material 3.

3.1 Bacteria and yeasts proteases profiles

Most of the selected bacterial genomes encoded from 60 to 100 different proteases, while the number of protease families ranged from 40 to 55 (Figure 2). Regarding the secreted proteases, most genomes encoded from 30 to 60 genes, which were included into 15 to 25 families. The genera with the higher number of proteases were *Lactobacillus* and *Enterococcus*, which are extensively used in the production of biopeptides [33,34]. In yeasts, the number of proteases-encoding genes was more evenly distributed and ranging from 150 to 160, while the protease families ranged from 60 to 67 (Figure 2). The number of protease-encoding genes in yeasts was higher than in bacteria, and ranged between 100 to 110 proteases and 60 to 67 protease families, with ~66% of them being secreted. According to these findings, despite the yeasts genera considered in the analysis were more distantly related than those of bacteria, some have a similar number of proteases. As asserted by Mirzaei et al. (2022), few studies have reported the use of yeasts as pure culture or in co-culture with bacteria to produce biopeptides. The high number of proteases, most of them secreted, makes yeast a more suitable candidate for biopeptides production than bacteria, however, other relevant aspects should be considered during the selection of the microbial species, including the range of biologically active peptides generated and the intra-species variability of the proteolytic activities. The number of proteases is directly correlated with the number of proteases families, and the results obtained for bacteria and yeasts were similar (R^2 between 0.93 and 0.97), while considering the secreted proteases the correlation was slightly lower (R^2 between 0.85 and 0.96) (Figure 2). Secreted proteases often play specific roles in extracellular environments, including nutrient acquisition and cell communication [26]. The slightly lower correlation may be reflective among secreted proteases compared to the broader spectrum of intracellular and membrane-associated proteases.

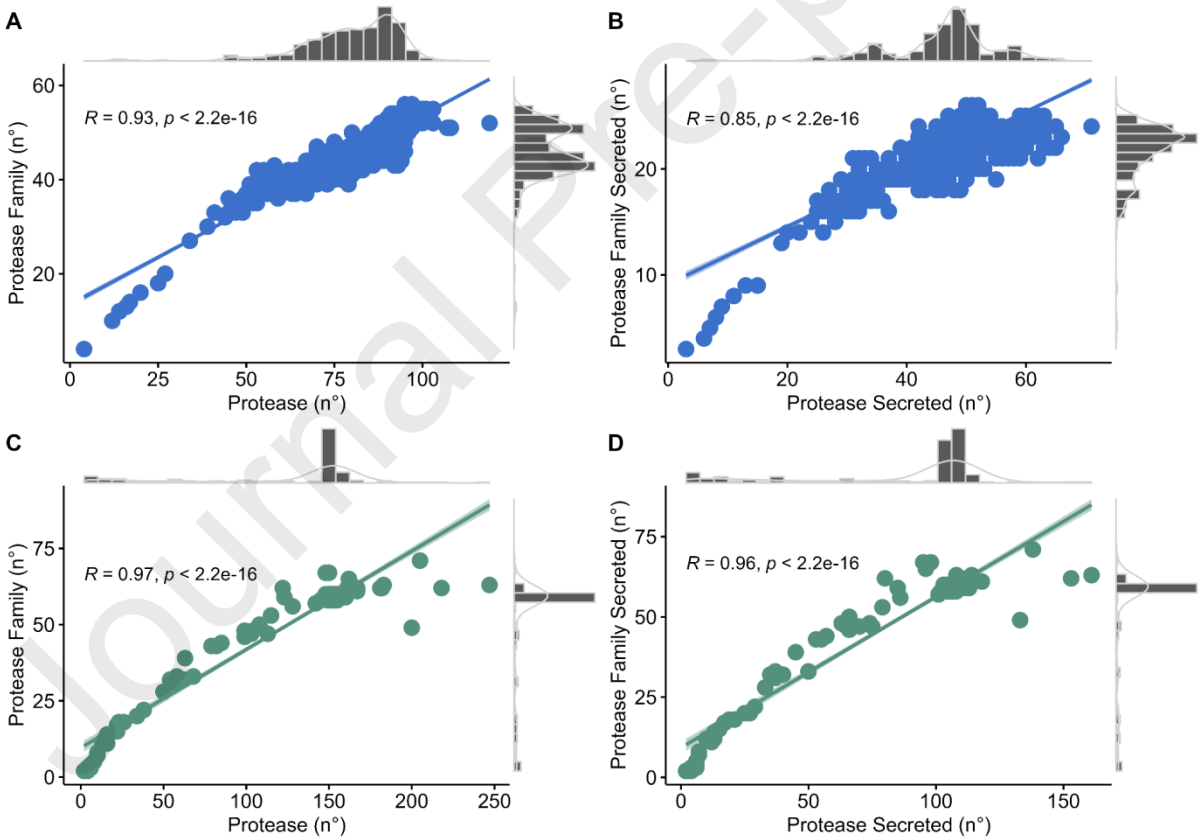


Figure 2. Proteases distribution in bacteria and yeasts. Scatterplots showing the Pearson correlation between the number of proteases and protease families in bacteria (blue) and yeasts (green). A and C show the trends for the total number of proteases, while in B and D the analysis was focused only on secreted proteases.

The most widely distributed bacterial secreted protease was gamma-glutamyl transferase (C26), which is a conserved enzyme found in bacteria, yeasts, plants, and animals (Figure 3 A, B). In *Bacillus*, it is involved in the degradation of poly γ-glutamic acid (PGA) into glutamate during nutrient starvation [35]. For yeasts, the most widely distributed protease was represented by the eukaryotic ubiquitin proteasome system (UPS; T01A) which is related to the Archaeal proteasome (Figure 3C) [36]. T01A is a highly conserved peptidase that regulates protein homeostasis [37]. Pepsin A (A01A) emerges as one of the most common secreted proteases in the examined yeasts (Figure 3 D); this family of proteins plays a crucial role in numerous physiological processes of *S. cerevisiae*, including the response to nutritional stress, regulation of the sporulation process, and growth under vegetative conditions [38].

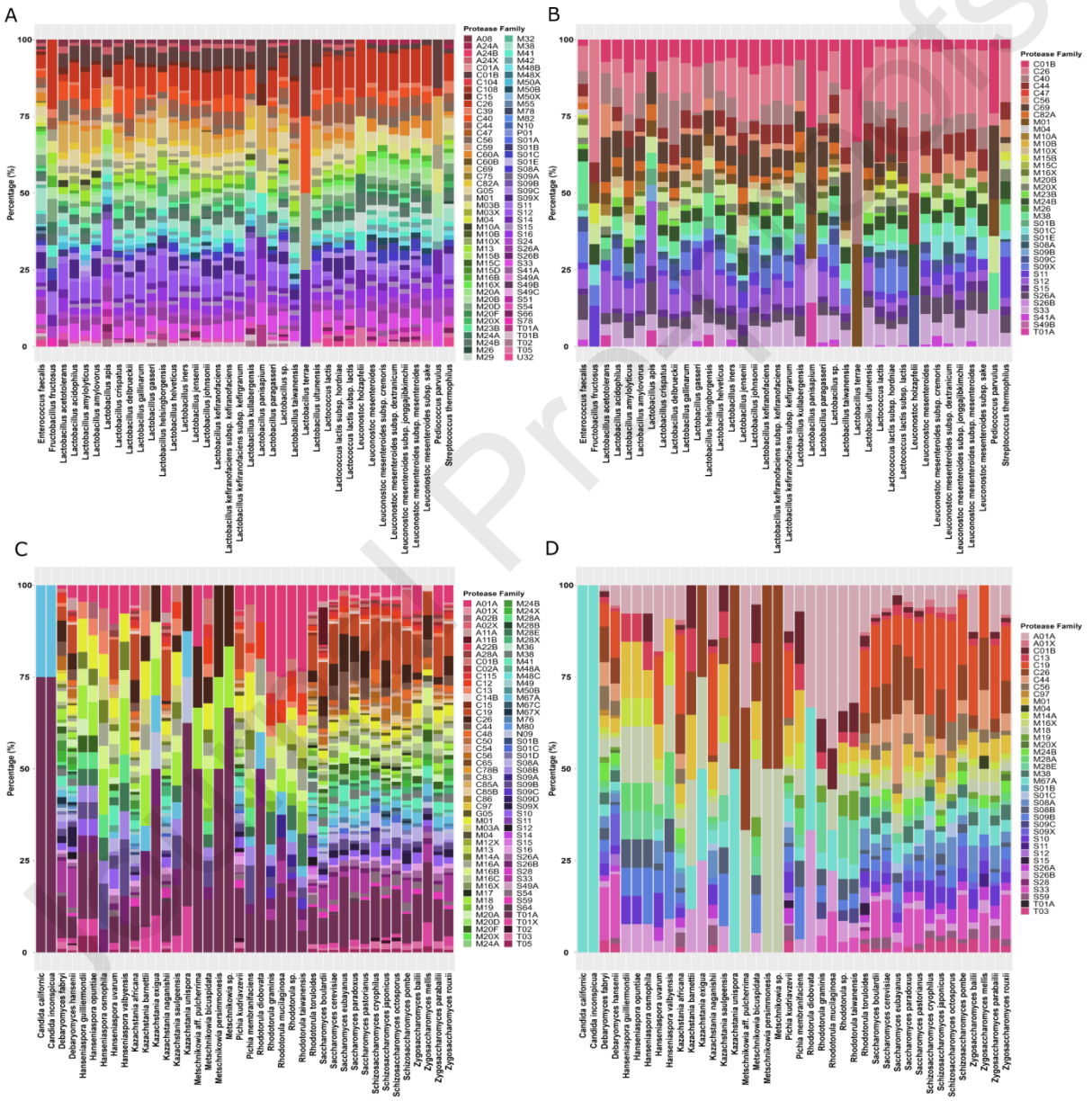


Figure 3. Number of intracellular and secreted proteases in the genomes. The barplot illustrates the number of intracellular and secreted protease families identified in all the bacterial (A, B) and yeast species (C, D) under investigation. Family names are derived from the MEROPS database and can be verified in Supplementary material 1.

3.2 In-silico generation of peptides

Functional foods offer additional health benefits beyond their basic nutritional value and their additional health benefits are often attributed to specific bioactive compounds, including biopeptides, or they can be fermented to reduce the allergenicity level [39]. Due to the great relevance for human health of the products obtained from the fermentation, highly abundant proteins present in seeds were used to test the FEEDS peptide prediction tool in an attempt to characterize the presence of bioactive peptides in the final product. The test was performed with five different genera of bacteria and yeasts taking into account the number, families and frequency of secreted proteases identified in the previous section. The two modes of the RPG tool (concurrent and sequential) were used to determine differences in the peptide length distribution, and all the results were reported in Supplementary material 5. The results obtained evidenced that most of the peptides generated by *Enterococcus faecalis AT22* using the "concurrent mode" were shorter than 20 amino acids and were matching the expected length of functional biopeptides [40] (Figure 4 A, B), suggesting that this species has a higher potential in generating bioactive peptides. In literature, *Enterococcus faecalis* strains were previously reported to produce ACE-inhibitory peptides from bovine skim milk [41]. *Enterococcus faecalis AT22* is distinct from the other species under investigation due to the presence of enzymes similar to thermolysin metallopeptidase M04 family (Supplementary material 5). One example of thermolysin M04 enzyme is the gelatinase, a protease responsible for biofilm formations [42]. According to the MEROPS database, thermolysin M04 family has a larger number of cleavage sites in comparison to other families, and this likely explains the higher number of peptides produced during the simulation performed with *Enterococcus faecalis AT22*. All the other bacterial species tested generated similar profiles, more specifically the number of peptides generated was lower and, as a consequence, length distribution was ranging from 21 to 100 amino acids (Figure 4).

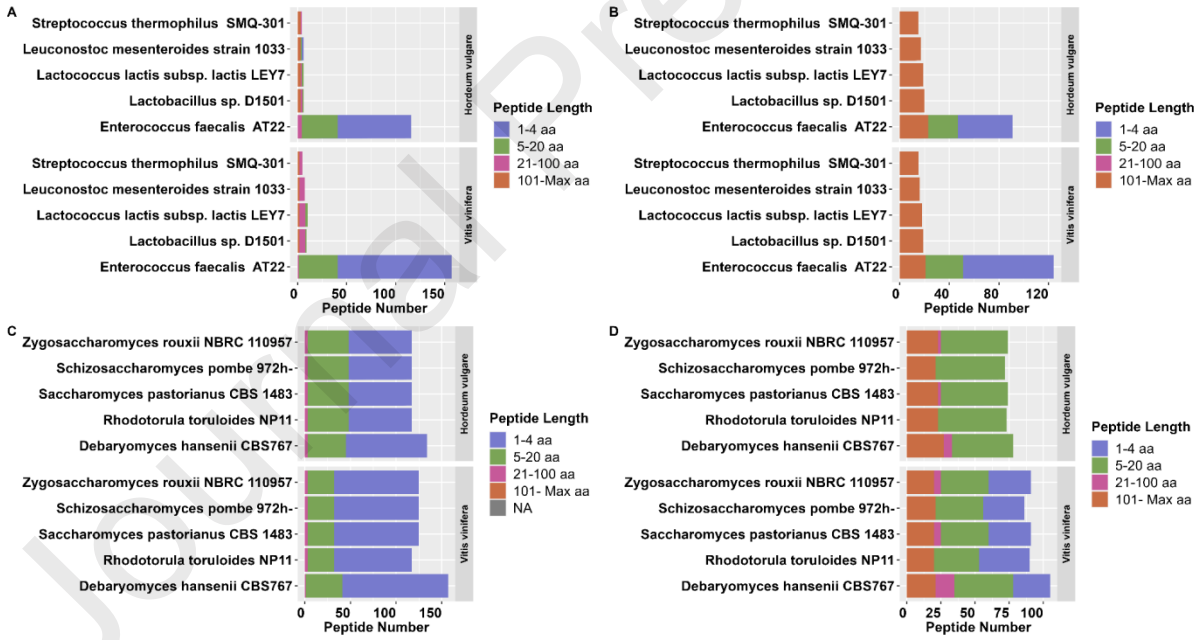


Figure 4. Length distribution of peptides generated from simulated hydrolysis. Number of peptides and length distribution obtained from in-silico tests performed with bacteria (A, B) and yeast strains (C, D) in concurrent (A, C) and sequential (B, D) mode. Two different protein substrates were used in the tests, one from *Hordeum vulgare* and one from *Vitis vinifera*. The first mix of proteins used as protein substrate included avenin, B1 hordein, B3 hordein, D hordein, type 1 glutelin, type 2 glutelin, Y1 hordein, and Y3 hordein from *Hordeum vulgare*. The second mix included 2S albumin, Cruciferin, 11S globulin, D1 glutelin, A3 glutelin, A legumin and Vicilin from *Vitis vinifera*.

The simulations performed revealed that the peptides generated from different yeast species were more closely related than those obtained from different bacteria (Supplementary material 5); as a confirmation, the comparison of the length profiles of the peptides obtained using different yeast species did not reveal significant differences (Figure 4). However, *Debaryomyces hansenii CBS767* analyzed in the concurrent mode showed a slightly higher number of short peptides with length between 1 and 4 AA. The shorter average length obtained is related to the presence of a glutamyl transpeptidase S01 serine family protease in *Debaryomyces hansenii CBS767*. This enzyme, similarly to the bacterial thermolysin, and according to the MEROPS database, can recognize a high number of cleavage sites. According to the literature, *Debaryomyces hansenii* is one of the most prevalent yeast species in dairy foods and, as previously reported, it can produce antihypertensive biopeptides from casein [43].

The lack of detailed information regarding the proteases cleavage sites in some species can result in a high similarity of the peptides pattern identified from different yeast species and this can limit the quality of the results obtained from FEEDS. Since the tool is strongly influenced by the quality of the information recovered from MEROPS, it should be used with caution when analysis data from poorly characterized species, but it can still provide a first glimpse on the digestion peptide predictions for each enzyme family and help to select among a range of different microorganism species. The quality of the predictions will certainly increase with the addition of new details regarding the proteases cleavage sites in yeasts.

3.3 Machine Learning Functional Prediction of Biopeptides

Biopeptides generated by the hydrolytic activity of fungal and bacterial species described in the previous section were examined with machine learning approaches in order to predict their potential function. Since the machine learning models were trained using 100 AA as maximum peptide length, this was also set as the maximum threshold for the next analyses. According to the probability of correct functional prediction probability, the peptides were assigned to three classes: biopeptides with a low probability of functional classification (lower than 50%), those with medium probability (between 50 and 70%), and those with a high probability (>70%). The peptides with low probability of classification were discarded. Only the biopeptides with medium and high were considered for further analyses (Supplementary material 6).

The results showed that the five bacterial species considered produced a panel of biopeptides characteristic for each microbial species, with only a fraction of identical sequences among species ranging from 0.3 (concurrent mode) to 1.6% (sequential mode) (Supplementary material 4 - Figure S1). The low percentage of identical peptides obtained in the comparison among the species is primarily attributed to *Enterococcus faecalis AT22*, which exhibited more than 80% specific biopeptides. Since the bioactive peptides tend to be highly different among bacterial species, the possibility to apply a preliminary bioinformatic screening can facilitate the species selection and provide a prediction of the process potentially related to the production of biopeptides with characteristic functions. As previously mentioned, some enzymes may contain additional information from the cleavage site, leading to redundant results. According to the simulations, the use of unicellular fungi for the simulated proteolysis lead to a higher fraction of identical peptides among species (ranging from ~54 to 61%). This is obviously due to the similarity in protease profiles among the fungal species investigated. Despite the high fraction of identical peptides, when setting the "sequential mode" *Debaryomyces hansenii CBS767* exhibits ~17% of specific sequences compared to the other strains.

The machine learning approach implemented in the FEEDS pipeline was able to predict the function of over 80% of the peptides derived from seed storage proteins with a probability of correct assignment higher than 50% (Table 1). It was observed peptide sequences longer than 20 AA frequently had a higher probability of correct assignment in comparison to the short ones. This trend is evident in analyzing the results obtained from *Enterococcus faecalis AT22*, which tends to produce short peptides

(Figure 4) compared to the other species (Table 1). On contrary, many peptides produced by *Lactobacillus sp. D1501*, *Leuconostoc mesenteroides FDAARGOS_1033*, and *Streptococcus thermophilus SMQ-301* were longer than 21 AA and were classified with high confidence as antimicrobials, a class of biopeptides frequently composed by long sequences [30].

Table 1. Functional annotation of biopeptides. Functional annotation of biopeptides produced from bacteria and fungi and shorter than 100 AA. Results were obtained using the "concurrent" and "sequential" modes and were separated according to the "medium (medium pr.)" or "high (high pr.)" probability of functional assignment.

Bacteria					**Yeast**				
Streptococcus thermophilus strain SMQ-301					**Schizosaccharomyces pombe 972h-**				
	Digestion Mode					**Digestion Mode**			
Biopeptide Family	**Concurrent**		**Sequential**		**Biopeptide Function**	**Concurrent**		**Sequential**	
	Medium pr.	**High pr.**	**Medium pr.**	**High pr.**		**Medium pr.**	**High pr.**	**Medium pr.**	**High pr.**
Antimicrobial	6	31	6	28	Antimicrobial	8	3	36	99
Opioid	0	0	0	0	Opioid	0	0	0	0
Antidiabetic	1	0	1	0	Antidiabetic	2	0	5	0
Antihypertensive	9	1	9	1	Antihypertensive	144	15	154	17
Antioxidant	9	2	8	1	Antioxidant	248	20	251	19
Cardiovascular	2	0	1	0	Cardiovascular	2	1	13	2
Celiac	2	10	1	10	Celiac	31	6	11	60
Immunomodulatory	0	0	0	0	Immunomodulatory	0	1	0	1
Neuropeptides	0	0	0	0	Neuropeptides	0	1	0	1
Total	29	44	26	40	Total	435	47	470	199

%	37.18	56.41	39.39	60.61	%	77.26	8.35	59.12	25.03
Lactobacillus sp. D1501					**Debaryomyces hansenii CBS767**				
Antimicrobial	7	48	7	49	Antimicrobial	2	2	63	180
Opioid	0	0	0	0	Opioid	0	0	0	0
Antidiabetic	1	0	1	0	Antidiabetic	0	2	5	0
Antihypertensive	10	1	10	1	Antihypertensive	157	19	178	20
Antioxidant	10	2	8	2	Antioxidant	293	19	309	31
Cardiovascular	2	0	1	0	Cardiovascular	2	1	20	2
Celiac	4	18	2	15	Celiac	7	31	13	86
Immunomodulatory	0	0	0	0	Immunomodulatory	0	1	1	1
Neuropeptides	0	0	0	0	Neuropeptides	0	1	1	1
Total	34	69	29	67	Total	461	76	590	321
%	30.91	62.73	32.95	76.14	%	74.60	12.30	52.91	28.79
Leuconostoc mesenteroides FDAARGOS_1033					**Rhodotorula toruloides NP11**				
Antimicrobial	4	30	4	23	Antimicrobial	3	2	47	170
Opioid	0	0	0	0	Opioid	0	0	0	0
Antidiabetic	0	0	0	0	Antidiabetic	2	0	4	0
Antihypertensive	6	0	7	0	Antihypertensive	157	19	173	22
Antioxidant	6	1	3	1	Antioxidant	289	19	273	19
Cardiovascular	3	0	1	0	Cardiovascular	2	1	20	2

Celiac	0	3	1	2	Celiac	7	31	17	108
Immunomodulatory	0	0	0	0	Immunomodulatory	0	1	0	1
Neuropeptides	0	0	0	0	Neuropeptides	0	1	1	1
Total	19	34	16	26	Total	460	74	535	323
%	34.55	61.82	39.02	63.41	%	74.92	12.05	53.93	32.56

Lactococcus lactis subsp. lactis LEY7					**Zygosaccharomyces rouxii 110957**				
Antimicrobial	5	42	7	47	Antimicrobial	8	3	33	129
Opioid	0	0	0	0	Opioid	0	0	0	0
Antidiabetic	1	0	1	0	Antidiabetic	2	0	5	0
Antihypertensive	12	1	12	1	Antihypertensive	147	16	150	17
Antioxidant	13	3	8	2	Antioxidant	252	19	247	18
Cardiovascular	2	0	1	0	Cardiovascular	2	1	16	2
Celiac	2	15	2	12	Celiac	6	31	12	66
Immunomodulatory	0	0	0	0	Immunomodulatory	0	1	0	1
Neuropeptides	0	0	0	0	Neuropeptides	0	1	1	1
Total	35	61	31	62	Total	417	72	464	234
%	31.82	55.45	34.83	69.66	%	73.81	12.74	56.24	28.36

Enterococcus faecalis strain AT22					**Saccharomyces pastorianus CBS_1483**				
Antimicrobial	3	4	37	83	Antimicrobial	8	3	36	138
Opioid	0	0	0	0	Opioid	0	0	0	0

Antidiabetic	1	0	1	0	Antidiabetic	2	0	5	0
Antihypertensive	182	33	187	30	Antihypertensive	145	16	160	19
Antioxidant	358	30	327	44	Antioxidant	252	18	265	19
Cardiovascular	3	4	10	5	Cardiovascular	2	1	15	1
Celiac	11	36	9	60	Celiac	6	31	13	66
Immunomodulatory	1	1	2	1	Immunomodulatory	0	1	0	1
Neuropeptides	0	0	0	0	Neuropeptides	0	1	1	1
Total	559	108	573	223	Total	415	71	495	245
%	84.57	16.34	65.26	25.40	%	73.84	12.63	56.00	27.71

In addition to antimicrobial peptides, antihypertensive, antioxidant, and celiac peptides were frequently predicted (table 1). Previous analyses performed using trypsin hydrolysis on *Hordeum vulgare* B-C-D hordein and globulin proteins (Tok et al., 2021) revealed also the presence of biopeptides having antihypertensive and antioxidant functions. Opioid, immunomodulatory and neuropeptides presented only one or any sequence, however, it should be noted that these peptide families were found to be associated with precursor proteins families not included in the present analysis, such as gliadin, and cliotide [45], as well as proteins from *Zea mays*, *Glycine max* [45], and insect proteins [46].

3.4 In-silico peptide sequence prediction validation

To validate the in-silico step of protein digestion, an in-vitro fermentation of brewer spent grains (*Hordeum vulgare*) was conducted in batch reactors and performing four independent tests using *Enterococcus faecalis* AVEL13, *Lactococcus lactis* WSL2, *Schizosaccharomyces pombe* J13151G1 and *Saccharomyces cerevisiae* KFAY3.

Peptides obtained from the fermentation were analyzed using mass spectrometry as described in the materials and methods and the results were compared with those obtained in-silico with FEEDS. The in-vitro analysis performed using the same four species yielded from 441 to 494 peptide sequences. In contrast to the in-silico analysis of protein sequence digestion, the in-vitro peptide sequences exhibited remarkable consistency among microorganism species, with approximately 20% of the peptide sequences being unique to each species (Supplementary material 4 - Figure S2). Furthermore, the number of peptides generated for each species remained consistent, on average 466 distinct peptides

produced by each microorganism. In-vitro studies involving the fermentation of goat milks utilizing *Lactiplantibacillus*, *Lactobacillus*, and *Streptococcus* and quantified by ultra-high performance liquid chromatography-quadrupole-time-of-flight tandem mass spectrometry (UPLC-Q-TOF-MS/MS) and high performance liquid chromatography-ion trap (HPLC-IT-MS/MS) provided similar results and evidenced similar profiles of peptide production between different genera of bacteria [47].

Comparative analysis between in-silico and in-vitro protein digestion revealed that 2.9% of in-silico sequences matched with those obtained in-vitro (Table 2). The low percentage can be due to the fact that in-vitro protein fermentation experiments are conducted under specific laboratory conditions, which may vary from study to study. Factors such as pH, temperature, microbial strains, and fermentation time can all influence the outcomes of the experiment. Certain peptide bonds could be resistant to cleavage due to their specific amino acid sequence or structural context, leading to missed cleavage events [48]. In contrast, FEEDS in-silico simulations use fixed parameters or assumptions, which may not align with the specific conditions of a particular in-vitro study.

To investigate the missed cleavage sites, we performed an in-silico digestion using eight enzymes and a variable percentage of miss cleavage events was set (Supplementary material 7). The findings demonstrated that when the in-silico analysis incorporated a 30% miss cleavage rate, there was an increased number of matches (3.7%) with the in-vitro digestion results (Table 2) suggesting that this can be an important parameter to consider in the future development of the tool.

Table 2. In-Silico peptide sequences matches with In-Vitro analysis. Number of peptide sequences of in-silico analysis that matches with in-vitro analysis considering identity of >90%.

Microorganism NCBI ID	Taxonomy	Number of Peptide Prediction Matches	
		Brewer Spent Grains (no miscleavage)	Brewer Spent Grains (miscleavage 30%)
GCA_000002945.2	*Schizosaccharomyces pombe* 972h-	29	30
GCA_000976545.2	*Saccharomyces cerevisiae* YJM984	36	36
GCF_018195835.1	*Lactococcus lactis subsp. lactis* LEY7	3	2
GCF_023299685.1	*Enterococcus faecalis* AT22	20	19
Cumulative count of distinct sequence matches		55	69

%	2.9	3.7

4. Conclusions

It was demonstrated here the possibility to develop a tool for the prediction of biopeptides composition and function by means of a simulated proteolytic digestion performed on protein sequences provided by the user. The results demonstrated that the predicted biopeptides show distinctive characteristics depending on the microbial species, while some proteases have the potential of providing more specificity to the generated profile. It was also demonstrated the capability of classifying the majority of peptides derived from seed storage proteins of *Hordeum vulgare* and *Vitis vinifera*. However, the authors envisage that further studies on bacterial and yeast protease cleavage sites will provide valuable information to enhance the reliability of the in-silico protein digestion step. The FEEDS tool is user-friendly, fast, and not only categorizes the formed peptides, but can also be utilized for the classification of proteases within the genomes of bacteria and yeasts. Furthermore, users can easily add new cleavage rules through the rapid peptide generator integrated into the tool. The tool has the potential to be developed to reinforce its capabilities, for example by incorporating a learning mechanism to refine its predictions based on user-generated data and false predictions. This adaptive approach holds promise for enhancing FEEDS' accuracy, particularly in scenarios involving complex food waste compositions with varied protein sources and non-protein materials. The possibility of performing bioinformatics pre-screening could pave the way to a faster and cheaper analysis of the most promising microbial candidates and protein substrates to be used for biopeptides production. Moreover, the FEEDS tool will allow a prediction of the potential bioactive compounds, leading to new approaches for mining valuable compounds in food waste.

Acknowledgements

This study was supported by the MIUR Research Projects of National Relevance (PRIN) “Future-proof bioactive peptides from food by-products: an eco-sustainable bioprocessing for tailored multifunctional foods” (PROACTIVE) – Prot. 2020CNRB84 and by the MIUR FFO-2022 *“L’innovazione delle biotecnologie nell’era della medicina di precisione, dei cambiamenti climatici e dell’economia circolare” (Una piattaforma biotecnologica per lo sfruttamento dei sottoprodotti agroindustriali nella coltivazione di microrganismi benefici)*.

Authorship contribution statement

Conceptualisation, V.B.C, E.B, and S.C; methodology, V.B.C and E.B; software, V.B.C and E.B; formal analysis, V.B.C and E.B; investigation, V.B.C and E.B; data curation, V.B.C; writing—original draft, V.B.C and E.B; writing—review and editing, V.B.C, S.C., L.T. and R.C.; visualization, V.B.C; supervision, R.C, S.C. and L.T.; fundings R.C. and S.C.; the authors read and approved the final manuscript.

Data Availability

MEROPS proteases database, diamond MEROPS database (.dmnd file), RPG id characterization and CICERON models are available on https://doi.org/10.6084/m9.figshare.22194535.v7. Installation instructions and usage information are available on https://github.com/vborincenturion/feeds.

References

[1] Zaky AA, Simal-Gandara J, Eun J-B, Shim J-H, Abd El-Aty AM. Bioactivities, Applications, Safety, and Health Benefits of Bioactive Peptides From Food and By-Products: A Review. Front Nutr 2022;8:815640. https://doi.org/10.3389/fnut.2021.815640.

[2] Chatterjee C, Gleddie S, Xiao C-W. Soybean Bioactive Peptides and Their Functional Properties. Nutrients 2018;10:1211. https://doi.org/10.3390/nu10091211.

[3] Peighambardoust SH, Karami Z, Pateiro M, Lorenzo JM. A Review on Health-Promoting, Biological, and Functional Aspects of Bioactive Peptides in Food Applications. Biomolecules 2021;11:631. https://doi.org/10.3390/biom11050631.

[4] Zhang Y-M, Ye D-X, Liu Y, Zhang X-Y, Zhou Y-L, Zhang L, et al. Peptides, new tools for plant protection in eco-agriculture. Adv Agrochem 2023;2:58–78. https://doi.org/10.1016/j.aac.2023.01.003.

[5] Aguilar-Toalá JE, Hernández-Mendoza A, González-Córdova AF, Vallejo-Cordoba B, Liceaga AM. Potential role of natural bioactive peptides for development of cosmeceutical skin products. Peptides 2019;122:170170. https://doi.org/10.1016/j.peptides.2019.170170.

[6] Mirzaei M, Shavandi A, Mirdamadi S, Soleymanzadeh N, Motahari P, Mirdamadi N, et al. Bioactive peptides from yeast: A comparative review on production methods, bioactivity, structure-function relationship, and stability. Trends Food Sci Technol 2021;118:297–315. https://doi.org/10.1016/j.tifs.2021.10.008.

[7] Bajić B, Vučurović D, Vasić Đ, Jevtić-Mučibabić R, Dodić S. Biotechnological Production of Sustainable Microbial Proteins from Agro-Industrial Residues and By-Products. Foods 2022;12:107. https://doi.org/10.3390/foods12010107.

[8] Ying X, Agyei D, Udenigwe C, Adhikari B, Wang B. Manufacturing of Plant-Based Bioactive Peptides Using Enzymatic Methods to Meet Health and Sustainability Targets of the Sustainable Development Goals. Front Sustain Food Syst 2021;5:769028. https://doi.org/10.3389/fsufs.2021.769028.

[9] Chakrabarti S, Guha S, Majumder K. Food-Derived Bioactive Peptides in Human Health: Challenges and Opportunities. Nutrients 2018;10:1738. https://doi.org/10.3390/nu10111738.

[10] Karami Z, Akbari-adergani B. Bioactive food derived peptides: a review on correlation between structure of bioactive peptides and their functional properties. J Food Sci Technol 2019;56:535–47. https://doi.org/10.1007/s13197-018-3549-4.

[11] Rutherfurd-Markwick KJ. Food proteins as a source of bioactive peptides with diverse functions. Br J Nutr 2012;108:S149–57. https://doi.org/10.1017/S000711451200253X.

[12] Lu J, Guo Y, Muhmood A, Zeng B, Qiu Y, Wang P, et al. Probing the antioxidant activity of functional proteins and bioactive peptides in Hermetia illucens larvae fed with food wastes. Sci Rep 2022;12:2799. https://doi.org/10.1038/s41598-022-06668-9.

[13] Food wastage footprint: impacts on natural resources: summary report. Rome: FAO; 2013.

[14] Schleiffer M, Speiser B. Presence of pesticides in the environment, transition into organic food, and implications for quality assurance along the European organic food chain – A review. Environ Pollut 2022;313:120116. https://doi.org/10.1016/j.envpol.2022.120116.

[15] Minkiewicz, Iwaniak, Darewicz. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. Int J Mol Sci 2019;20:5978. https://doi.org/10.3390/ijms20235978.

[16] Qeli E, Omasits U, Goetze S, Stekhoven DJ, Frey JE, Basler K, et al. Improved prediction of peptide detectability for targeted proteomics using a rank-based algorithm and organism-specific data. J Proteomics 2014;108:269–83. https://doi.org/10.1016/j.jprot.2014.05.011.

[17] Qin D, Bo W, Zheng X, Hao Y, Li B, Zheng J, et al. DFBP: a comprehensive database of food-derived bioactive peptides for peptidomics research. Bioinformatics 2022;38:3275–80. https://doi.org/10.1093/bioinformatics/btac323.

[18] Chaudhary A, Bhalla S, Patiyal S, Raghava GPS, Sahni G. FermFooDb: A database of bioactive peptides derived from fermented foods. Heliyon 2021;7:e06668. https://doi.org/10.1016/j.heliyon.2021.e06668.

[19] Ma Y, Guo Z, Xia B, Zhang Y, Liu X, Yu Y, et al. Identification of antimicrobial peptides from the human gut microbiome using deep learning. Nat Biotechnol 2022;40:921–31. https://doi.org/10.1038/s41587-022-01226-0.

[20] Minkiewicz P, Iwaniak A, Darewicz M. BIOPEP-UWM Virtual—A Novel Database of Food-Derived Peptides with In Silico-Predicted Biological Activity. Appl Sci 2022;12:7204. https://doi.org/10.3390/app12147204.

[21] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12:59–60. https://doi.org/10.1038/nmeth.3176.

[22] Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119. https://doi.org/10.1186/1471-2105-11-119.

[23] Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genomics Bioinforma 2020;2:lqaa026. https://doi.org/10.1093/nargab/lqaa026.

[24] Luo L, Li H, Zhang L. ORF Organization and Gene Recognition in the Yeast Genome. Comp Funct Genomics 2003;4:318–28. https://doi.org/10.1002/cfg.292.

[25] Maillet N. Rapid Peptides Generator: fast and efficient in silico protein digestion. NAR Genomics Bioinforma 2020;2:lqz004. https://doi.org/10.1093/nargab/lqz004.

[26] Cezairliyan B, Ausubel FM. Investment in secreted enzymes during nutrient-limited growth is utility dependent. Proc Natl Acad Sci 2017;114. https://doi.org/10.1073/pnas.1708580114.

[27] Razzaq A, Shamsi S, Ali A, Ali Q, Sajjad M, Malik A, et al. Microbial Proteases Applications. Front Bioeng Biotechnol 2019;7:110. https://doi.org/10.3389/fbioe.2019.00110.

[28] Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics 2001;17:282–3. https://doi.org/10.1093/bioinformatics/17.3.282.

[29] Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. BUSCA: an integrative web server to predict subcellular localization of proteins. Nucleic Acids Res 2018;46:W459–66. https://doi.org/10.1093/nar/gky320.

[30] Bizzotto, Edoardo, Zampieri, Guido, Treu, Laura, Filannino, Pasquale, Di Cagno, Rafaella, Campnaro, Stefano. Classification of bioactive peptides: a comparative analysis of models and encodings. bioRxiv n.d. https://doi.org/10.1101/2023.10.04.560809.

[31] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 2016;44:D733–45. https://doi.org/10.1093/nar/gkv1189.

[32] Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2016;44:D67–72. https://doi.org/10.1093/nar/gkv1276.

[33] Singh BP, Vij S. Growth and bioactive peptides production potential of Lactobacillus plantarum strain C2 in soy milk: A LC-MS/MS based revelation for peptides biofunctionality. LWT 2017;86:293–301. https://doi.org/10.1016/j.lwt.2017.08.013.

[34] Worsztynowicz P, Białas W, Grajek W. Integrated approach for obtaining bioactive peptides from whey proteins hydrolysed using a new proteolytic lactic acid bacteria. Food Chem 2020;312:126035. https://doi.org/10.1016/j.foodchem.2019.126035.

[35] Hsueh Y-H, Huang K-Y, Kunene S, Lee T-Y. Poly-γ-glutamic Acid Synthesis, Gene Regulation, Phylogenetic Relationships, and Role in Fermentation. Int J Mol Sci 2017;18:2644. https://doi.org/10.3390/ijms18122644.

[36] Maupin-Furlow JA. Archaeal Proteasomes and Sampylation. In: Dougan DA, editor. Regul. Proteolysis Microorg., vol. 66, Dordrecht: Springer Netherlands; 2013, p. 297–327. https://doi.org/10.1007/978-94-007-5940-4_11.

[37] Wendler P, Enenkel C. Nuclear Transport of Yeast Proteasomes. Front Mol Biosci 2019;6:34. https://doi.org/10.3389/fmolb.2019.00034.

[38] Zhang H, Zhang H, Chen Q, Ruan H, Fu M, He G. Effects of proteinase a on cultivation and viability characteristics of industrial Saccharomyces cerevisiae WZ65. J Zhejiang Univ Sci B 2009;10:769–76. https://doi.org/10.1631/jzus.B0920057.

[39] Jain A. Seed Storage Protein, Functional Diversity and Association with Allergy. Allergies 2023;3:25–38. https://doi.org/10.3390/allergies3010003.

[40] Bell MJ, Burrows JM, Brennan R, Miles JJ, Tellam J, McCluskey J, et al. The peptide length specificity of some HLA class I alleles is very broad and includes peptides of up to 25 amino acids in length. Mol Immunol 2009;46:1911–7. https://doi.org/10.1016/j.molimm.2008.12.003.

[41] Ramakrishna R, Sarkar D, Schwarz P, Shetty K. Phenolic linked anti-hyperglycemic bioactives of barley (Hordeum vulgare L.) cultivars as nutraceuticals targeting type 2 diabetes. Ind Crops Prod 2017;107:509–17. https://doi.org/10.1016/j.indcrop.2017.03.033.

[42] Del Papa MF, Hancock LE, Thomas VC, Perego M. Full Activation of *Enterococcus faecalis* Gelatinase by a C-Terminal Proteolytic Cleavage. J Bacteriol 2007;189:8835–43. https://doi.org/10.1128/JB.01311-07.

[43] García-Tejedor A, Sánchez-Rivera L, Recio I, Salom JB, Manzanares P. Dairy Debaryomyces hansenii strains produce the antihypertensive casein-derived peptides LHLPLP and HLPLP. LWT - Food Sci Technol 2015;61:550–6. https://doi.org/10.1016/j.lwt.2014.12.019.

[44] Tok K, Moulahoum H, Kocadag Kocazorbaz E, Zihnioglu F. Bioactive peptides with multiple activities extracted from Barley (*Hordeum vulgare* L.) grain protein hydrolysates: Biochemical analysis and computational identification. J Food Process Preserv 2021;45. https://doi.org/10.1111/jfpp.15024.

[45] Pavlicevic M, Marmiroli N, Maestri E. Immunomodulatory peptides—A promising source for novel functional food production and drug discovery. Peptides 2022;148:170696. https://doi.org/10.1016/j.peptides.2021.170696.

[46] Quah Y, Tong S-R, Bojarska J, Giller K, Tan S-A, Ziora ZM, et al. Bioactive Peptide Discovery from Edible Insects for Potential Applications in Human Health and Agriculture. Molecules 2023;28:1233. https://doi.org/10.3390/molecules28031233.

[47] Moreno-Montoro M, Jauregi P, Navarro-Alarcón M, Olalla-Herrera M, Giménez-Martínez R, Amigo L, et al. Bioaccessible peptides released by in vitro gastrointestinal digestion of fermented goat milks. Anal Bioanal Chem 2018;410:3597–606. https://doi.org/10.1007/s00216-018-0983-0.

[48] Aki K, Okamura E. D-β-aspartyl residue exhibiting uncommon high resistance to spontaneous peptide bond cleavage. Sci Rep 2016;6:21594. https://doi.org/10.1038/srep21594.

Supplementary Information

Supplementary material 1: Summary of Bacteria and Yeast strains, substrate and biopeptide families used in the FEEDS tool.

Supplementary material 2: MEROPS proteases families information.

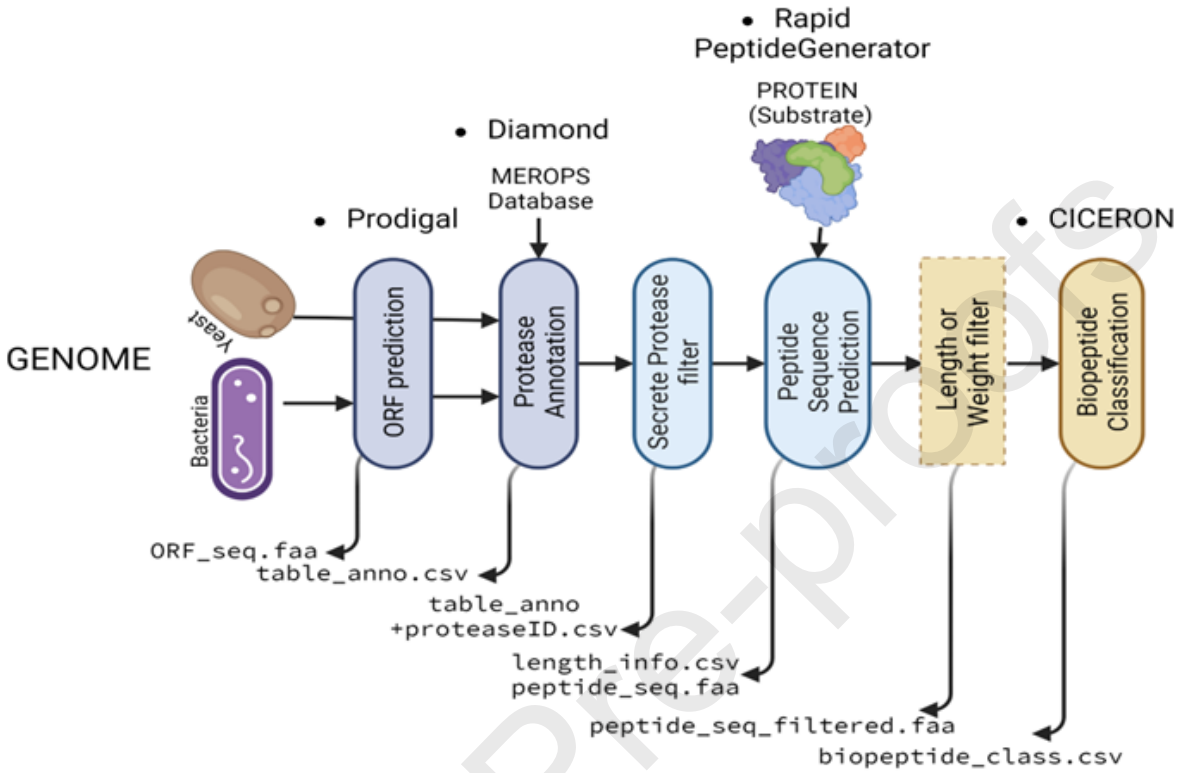
Supplementary material 3: Secreted protease, cleavage sites and RPG ID information

Supplementary material 4: Supplementary Figure 1, Venn diagram in-silico analysis; **Supplementary Figure 2,** Venn diagram in-vitro analysis

Supplementary material 5: Protease profiles of selected Bacteria and Yeast strains.

Supplementary material 6: CICERON sequential and concurrent of selected Bacteria and Yeast results information

Supplementary material 7: In-silico and In-vitro biopeptides sequences matches



Highlights

- FEEDS is a novel biopeptide prediction and classification tool
- Bioinformatics pre-screening will allow a faster and cheaper biopeptide production
- Biopeptide profiles show distinctive characteristics depending on the microbial species
- FEEDS performs functional classification for most peptides from seed storage proteins

Credit Author Statement

Conceptualisation, V.B.C, E.B, and S.C; methodology, V.B.C and E.B; software, V.B.C and E.B; formal analysis, V.B.C and E.B; investigation, V.B.C and E.B; data curation, V.B.C; writing—original draft, V.B.C and E.B; writing—review and editing, V.B.C, S.C. and R.C.; visualization, V.B.C; supervision, R.C. and S.C.; fundings R.C. and S.C.; the authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.